Search for $Z' \to Zh \to llbb$ in pp Collisions at $\sqrt{s} =$ 8 TeV Using the CMS Detector at the LHC

by

Jun-Yi Wu

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Master of Science

at the

NATIONAL CENTRAL UNIVERSITY

June 2016

Author . . . .
Department of Physics
June 16, 2016

Certified by . . . .
Shin-Shan Yu
Associate Professor
Thesis Supervisor

Accepted by . . . .
Chiang-Mei Chen
Professor
Chairman, Thesis Committee

Search for $Z' \to Zh \to llbb$ in pp Collisions at $\sqrt{s} = 8$ TeV Using the CMS Detector at the LHC

by

Jun-Yi Wu

Submitted to the Department of Physics
on June 16, 2016, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

The result of a search for new particles decaying to Z and Higgs bosons with subsequent decay to a final state containing two leptons and two quarks, $Z' \to Zh \to llqq$, is reported. This analysis is based on the proton-proton collisions at $\sqrt{s} = 8$ TeV collected with the CMS detector at the LHC, corresponding to an integrated luminosity of 19.7/fb. Subjet b-tagging techniques are exploited for separating the signal from the SM background when the boost of the Higgs causes the two b quarks to merge into the same jet reconstructed in the detector, and the exclusion limit result based on two dimension shape method is presented.

Thesis Supervisor: Shin-Shan Yu
Title: Associate Professor

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Theory Overview

1.1 Introduction

This thesis presents the analysis details and the results of the search for heavy resonances decaying into a Z boson and a Higgs boson (h) at the center-of-mass energy of 8 TeV, using 19.7 fb^{-1} p-p collision data. In turn, the Z boson is identified through its leptonic decays (leptons often refer to e and μ only in experiments. $l = e, \mu$). The Higgs boson h is expected to hadronically decay into a pair of b-quarks. The investigated final states consist of two charged leptons which are identified in the detector and limit the presence of the background, and two b-quarks from the hadronic Higgs decay which collects the largest possible fraction of Higgs events.

This thesis is organised as follows. In the latter part of this chapter, the model that predicts heavy resonances is introduced, including the expected cross section and the specification of model parameters. In chapter 2, the LHC and the CMS experiment are described, including the information of each sub-detector and the trigger system of the CMS. The details of the analysis are shown in chapter 3. This chapter reveals the way to reconstruct physical objects in CMS. By adding some proper kinematic selections on those physics objects, the interested events in data collected by the CMS detector can be selected. Moreover, this chapter shows the comparison between data and simulation. In the last chapter, the results of the search and the conclusion are presented.

1.2 Theory Overview

Although the Higgs boson discovered by the ATLAS and CMS collaborations [3–5] imposes strong constraints on theories beyond the Standard Model(SM), the extreme fine tuning in quantum corrections required to have a light fundamental Higgs boson with mass close to 125 GeV [6–9] suggests that the Standard Model may be incomplete, and not valid beyond a scale of a few TeV. Various dynamical electroweak symmetry breaking scenarios which attempt to solve this naturalness problem, such as Minimal Walking Technicolor [10], Little Higgs [11–13], or composite Higgs models [14–16] predict the existence of new resonances decaying to a vector boson plus a Higgs boson.

1.2.1 Heavy Vector Triplet Model

Resonance searches are typically not sensitive to all the details and the free parameters of the underlying model, but only to those parameters or combinations of parameters that control the mass of the resonance and the interactions involved in its production and decay. Therefore, one can employ a simplified description of the resonance defined by a phenomenological Lagrangian where only the relevant couplings and mass parameters are retained. This model-independent strategy applies a Heavy Vector Triplet (HVT) [17] to the Standard Model group and reproduces a large class of explicit models. In Eq. (1.1), the mathematical form of the simplified Lagrangian is defined, where V_ν^a , a = 1,2,3, is a real vector with vanishing hypercharge in the adjoint representation of $SU(2)_L$, it decscribes one charged and one neutral heavy spin-1 particle with charge eigenstate fields, and $D_{[\mu}V_{\nu]}^a$ represents the covariant derivative.

$$\begin{aligned}\mathcal{L}_{\mathcal{V}} = &-\frac{1}{4}D_{[\mu}V_{\nu]}^a D^{[\mu}V^{\nu]a} + \frac{m_V^2}{2}V_\mu^a V^{\mu a}\\ &+ i g_V c_H V_\mu^a H^\dagger \tau^a \overset{\leftrightarrow}{D}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a \sum_f \bar{f}_L \gamma^\mu \tau^a f_L\\ &+ \frac{g_V}{2} c_{VVV} \epsilon_{abc} V_\mu^a V_\nu^b D^{[\mu}V^{\nu]c} + \text{quadrilinear terms}\end{aligned} \tag{1.1}$$

$$V_\mu^\pm = \frac{V_\mu^1 \mp i V_\mu^2}{\sqrt{2}} \ , \ V_\mu^0 = V_\mu^3 \tag{1.2}$$

$$D_{[\mu} V_{\nu]}^a = D_\mu V_\nu^a - D_\nu V_\mu^a \ , \ D_\mu V_\nu^a = \partial_\mu V_\nu^a + g \epsilon^{abc} W_\mu^b V_\nu^c \tag{1.3}$$

$$H = \begin{pmatrix} \phi^+ \\ \phi^- \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \\ \frac{1}{\sqrt{2}}(\phi_3 + i\phi_4) \end{pmatrix} \tag{1.4}$$

In these models, new heavy vector bosons $(V^\pm, V^0)$ that couple to the SM Higgs doublet (Eq. 1.4) and SM gauge bosons with the parameters c_H and g_V and to the fermions via the combination $(g^2/g_V)c_F$. The parameter g_V represents the strength of the new vector boson interaction, while c_H and c_F represent the couplings to the Higgs and the fermions respectively, and are expected to be of the order of unity in most models.

1.2.2 Basic Phenomenology

Masses and Mixings

After electro-weak symmetry breaking (EWSB), the only massless state is photon, which can be identified as the gauge field associated with the unbroken $U(1)_{em}$. The two other neutral mass eigenstates are the SM Z boson and one heavy vector of mass M_0 which are obtained by diagonalizing the mass matrix of the (Z, V^0) system by a rotation with angle θ_N

$$\begin{pmatrix} Z \\ V^0 \end{pmatrix} \rightarrow \begin{pmatrix} \cos\theta_N & \sin\theta_N \\ -\sin\theta_N & \cos\theta_N \end{pmatrix} \begin{pmatrix} Z \\ V^0 \end{pmatrix} . \tag{1.5}$$

The mass matrix is

$$\mathcal{M}_N^2 = \begin{pmatrix} \hat{m}_Z^2 & c_H \xi \hat{m}_Z \hat{m}_V \\ c_H \xi \hat{m}_Z \hat{m}_V & \hat{m}_V^2 \end{pmatrix} \text{ , where } \begin{cases} \hat{m}_Z = \frac{e\hat{v}}{2\sin\theta_W \cos\theta_W} \\ \hat{m}_V^2 = m_V^2 + g_V^2 c_{VVHH} \hat{v}^2 \\ \xi = \frac{g_V \hat{v}}{2\hat{m}_V} \end{cases} . \tag{1.6}$$

In the above equations $\hat{v}$ denotes the Vacuum Expetation Value (VEV) defined by $\langle H^\dagger H \rangle = \hat{v}^2/2$, and one should know the masses $\hat{m}_Z$ and $\hat{m}_V$ do not coincide with the physical Z boson and the masses of the new resonances of this model, although they do in the approximations later (Eq. 1.12). The mass eigenvalues and the rotation angles are easily obtained by inverting the relations

$$\begin{aligned} Tr[\mathcal{M}_N^2] &= \hat{m}_Z^2 + \hat{m}_V^2 = m_Z^2 + M_0^2 \ , \\ Det[\mathcal{M}_N^2] &= \hat{m}_Z^2 \hat{m}_V^2 (1 - c_H^2 \xi^2) = m_Z^2 M_0^2 \ , \\ \tan 2\theta_N &= \frac{2 c_H \xi \hat{m}_Z \hat{m}_V}{\hat{m}_V^2 - \hat{m}_Z^2} \ . \end{aligned} \tag{1.7}$$

Note that M_0 represents the real mass eigenvalue of the neutral heavy vector boson. Moreover, since we assume $\hat{m}_V > \hat{m}_Z$, the only variable controls the sign of the tangent function is c_H, which is model dependent. Once the sign of c_H is determined, the tangent can be uniquely inverted.

The situation is similar in the charged vector mass matrix of $(W^\pm, V^\pm)$ system, and $M_\pm$ denotes the real mass eigenvalue of charged states as well.

$$\mathcal{M}_C^2 = \begin{pmatrix} \hat{m}_W^2 & c_H \xi \hat{m}_W \hat{m}_V \\ c_H \xi \hat{m}_W \hat{m}_V & \hat{m}_V^2 \end{pmatrix} \text{ , where } \hat{m}_W = \frac{e\hat{v}}{2\sin\theta_W} = \cos\theta_W \hat{m}_Z \ , \tag{1.8}$$

where it is diagonalized by

$$\begin{aligned}
Tr[\mathcal{M}_C^2] &= \hat{m}_W^2 + \hat{m}_V^2 = m_W^2 + M_\pm^2 \ , \\
Det[\mathcal{M}_C^2] &= \hat{m}_W^2 \hat{m}_V^2 (1 - c_H^2 \xi^2) = m_W^2 M_\pm^2 \ , \\
\tan 2\theta_C &= \frac{2 c_H \xi \hat{m}_W \hat{m}_V}{\hat{m}_V^2 - \hat{m}_W^2} \ .
\end{aligned} \tag{1.9}$$

By checking Eq. (1.6) and Eq. (1.8), the charged and neutral mass matrices are connected by custodial symmetry, which can be shown in full generality to imply

$$\mathcal{M}_C^2 = \begin{pmatrix} \cos\theta_W & 0 \\ 0 & 1 \end{pmatrix} \mathcal{M}_N^2 \begin{pmatrix} \cos\theta_W & 0 \\ 0 & 1 \end{pmatrix} \ . \tag{1.10}$$

By taking the determinant of the above equation, or equivalently by comparing the charged and neutral determinants in Eq. (1.7) and Eq. (1.9), we obtain a generalized custodial relation among the physical masses

$$m_W^2 M_\pm^2 = \cos^2\theta_W m_Z^2 M_0^2 \ . \tag{1.11}$$

From the simple formula above, we can start to identify the physically reasonable region of the parameter space in this model. We aim at describing new vectors with masses at or above the TeV scale, but we also want the SM masses $m_{W,Z} \sim 100$ GeV to be reproduced. Therefore we require a hierachy in the mass relation of SM Z and W bosons versus the new vectors.

$$\frac{\hat{m}_{W,Z}}{\hat{m}_V} \sim \frac{m_{W,Z}}{M_{\pm,0}} \leq 10^{-1} \ll 1 \tag{1.12}$$

Use the limit above, we can expand the determinant formulas both in Eq. (1.7) and Eq. (1.9) to obtain simple approximations for m_W and m_Z

$$\begin{aligned}
m_Z^2 &= \hat{m}_Z^2 (1 - c_H^2 \xi^2)(1 + \mathcal{O}(\hat{m}_Z^2/\hat{m}_V^2)) \ , \\
m_W^2 &= \hat{m}_W^2 (1 - c_H^2 \xi^2)(1 + \mathcal{O}(\hat{m}_W^2/\hat{m}_V^2)) \ .
\end{aligned} \tag{1.13}$$

The parameter ξ can be either very small or of order unity. Both cases are realized in explicit models. While $\xi \ll 1$ is the most common situation, $\xi \sim 1$ only occurs in strongly coupled scenarios at very large g_V. In these approximations, SM tree-level experimental observation can be reproduced to percent accuracy.
Since $\hat{m}_W = \cos\theta_W \hat{m}_Z$, the W-Z mass ratio is thus given by

$$\frac{m_W^2}{m_Z^2} \simeq \cos^2\theta_W \ . \tag{1.14}$$

Eq. (1.14) has one important implication on the masses of the new vectors. When combined with the custodial relation Eq. (1.11), it tells us that the charged and neutral Vs are practically degenerate

$$M_\pm^2 = M_0^2(1 + \mathcal{O}(\%)) \ , \tag{1.15}$$

In the following, when working at the leading order in the limit Eq. (1.12), we can ignore the mass splitting and denote the mass of the charged and the neutral states collectively as M_V. It is easy to check that in that limit $M_V = \hat{m}_V$.

Decay Widths

Because of the hierachy in the mass matrices, the mixing angles are naturally small. By looking at Eqs. (1.7), (1.9) and (1.12) we can estimate

$$\theta_{N,C} \simeq c_H \xi \frac{\hat{m}_{W,Z}}{\hat{m}_V} \leq 10^{-1} \ , \tag{1.16}$$

and after rotating to the mass basis, the coupling of the neutral and charged resonances to left- and right-handed fermion chiralities can be written in a compact form for each fermion species $F = \{l, q, 3\}$.

$$\begin{cases} g_L^N = \frac{g^2}{g_V}\frac{c_F}{2}\cos\theta_C + (g_L^Z)_{SM}\sin\theta_N \simeq \frac{g^2}{g_V}\frac{c_F}{2}\ , \\ g_R^N = (g_R^Z)_{SM}\sin\theta_N \simeq 0 \end{cases}$$
$$\begin{cases} g_L^C = \frac{g^2}{g_V}\frac{c_F}{2}\cos\theta_C + (g_L^W)_{SM}\sin\theta_C \simeq \frac{g^2}{g_V}\frac{c_F}{2}\ , \\ g_R^C = 0 \end{cases} \tag{1.17}$$

In the above equation $(g_{L,R}^{W,Z})_{SM}$ denote the ordinary SM W and Z couplings (with the normalization given by $g_L^W = g/\sqrt{2}$).

Given that the rotation angles are small, the couplings further simplify, as also shown in the equation. We could see that V interact mainly with left-handed chiralities and that all the couplings for each fermion species are controlled by the parameter combination $g^2/g_V c_F$. This gives tight correlations among different channels

$$\Gamma_{V_\pm \to f\bar{f}'} \simeq 2\Gamma_{V_0 \to f\bar{f}'} \simeq N_C[f](\frac{g^2 c_F}{g_V})^2 \frac{M_V}{48\pi}\ , \tag{1.18}$$

where $N_C[f]$ is the number of colors (3 for the di-quark and 1 for the dilepton decays). The parameters $c_F = \{c_l, c_q, c_3\}$ control the relative BRs to leptons, light quarks and the third family.

In the case of di-boson decay width

$$\begin{aligned} \Gamma_{V_0 \to W_L^+ W_L^-} \simeq \Gamma_{V_\pm \to W_L^\pm Z_L} &\simeq \frac{g_V^2 c_H^2 M_V}{192\pi} \frac{(1 + c_H c_{VVV}\xi^2)^2}{(1 - c_H^2 \xi^2)^2} = \frac{g_V^2 c_H^2 M_V}{192\pi}[1 + \mathcal{O}(\xi^2)]\ , \\ \Gamma_{V_0 \to Z_L h} \simeq \Gamma_{V_\pm \to W_L^\pm h} &\simeq \frac{g_V^2 c_H^2 M_V}{192\pi} \frac{(1 - 4c_{VVHH}\xi^2)^2}{1 - c_H^2 \xi^2} = \frac{g_V^2 c_H^2 M_V}{192\pi}[1 + \mathcal{O}(\xi^2)]\ . \end{aligned} \tag{1.19}$$

Note that Eq. (1.19) is derived in the Equivalent Gauge [18] because the decay to transverse SM vectors is highly suppressed while to the longitudinal parts grows with the energy of the process, therefore the Unitary Gauge which is used in the original

Lagrangian is instead useful. The channels that are not shown in the above equations are either forbidden or suppressed like the decays to transverse polarizations.

From this section, a very simple picture emerges. At small ξ, all the decay widths are fixed with a given resonance mass M_V and the couplings $\{g^2 c_F/g_V,\, g_V c_H\}$ which control the BRs in all relevant channels. Paremeters c_{VVV}, c_{VVHH} and c_{VVW} are basically irrelevent. Thus, the basic phenomenology of this model is well described by a good approximation.

1.2.3 Explicit Models

Now the general picture is clear, we can get exact values of the widths and BRs from explicit models. Consider two benchmark models, A and B, which correspond to two explicit models describing the heavy vectors in Refs. [19] and [14] respectively. All the c parameters are fixed to specific values in these models and the only free parameters are the resonance mass M_V and coupling g_V. Moreover, model A is inspired by weakly coupled extensions of the SM gauge group while model B is by strongly coupled scenarios of EWSB, *i.e.* Composite Higgs models, we will consider them in different regions of g_V, relatively small $g_V \leq 3$ and relatively large $g_V \geq 3$.

Figure 1-1 shows the BRs as functions of the mass in model A and B. As expected from the previous discussion and according to Refs. [19], model A predicts

$$\begin{aligned} &c_H \simeq -g^2/g_V^2 \ , \ c_F \simeq 1 \ , \\ &g_V^2 c_H^2 \simeq g^4 c_F^2/g_V^2 \simeq g^2/g_V \ . \end{aligned} \tag{1.20}$$

Therefore Eq. (1.18) and (1.19) can be determined in the following form for V_0 in model A ($g_V = 1$),

$$\begin{aligned} &\Gamma_{V_0 \to f\bar{f}'} \simeq N_c[f]\frac{g^4 M_V}{96\pi} \\ &\Gamma_{V_0 \to W^+W^-} \simeq \Gamma_{V_0 \to Zh} \simeq \frac{g^4 M_V}{192\pi} \ . \end{aligned} \tag{1.21}$$

One can easily check either from the plot or the equation, a factor of two difference comparing the BRs between fermions and bosons. Due to the color factor, leptons and quarks also have a difference by a factor of three. Since the c_F term is universal both in A and B. The total width in model A decreases with increasing g_V because of the overall suppresion (g^2/g_V) in Eq. (1.20).
On the contrary, in model B the c_H term is unsuppressed

$$\begin{aligned} & c_H \simeq c_F \simeq 1 \ , \\ & g_V^2 c_H^2 \simeq g_V^2 \ , \ g^2 c_{c_F}/g_{g_V} \simeq g^2/g_V \ . \end{aligned} \tag{1.22}$$

Thus the determinate V_0 decay widths for model B ($g_V = 3$) are

$$\begin{aligned} & \Gamma_{V_0 \to f\bar{f}'} \simeq N_c[f] \frac{g^4 M_V}{342\pi} \\ & \Gamma_{V_0 \to W^+W^-} \simeq \Gamma_{V_0 \to Zh} \simeq \frac{3M_V}{64\pi} \ . \end{aligned} \tag{1.23}$$

For model B$_{g_V=3}$ the dominant BRs are into di-bosons and the fermionic decays are extremely suppressed. Moreover, the total width increases with increasing g_V since it is dominated by the di-boson width which grows with g_V as expected from Eq. (1.22). This model B is particularly interesting for the present search, since it predicts signal cross sections of the order of fb [17] [20] [Fig. 1-2], branching ratios to vector bosons close to unity, and thus being accessible at the LHC.

In the latter chapters, the mass eigenstate of the neutral heavy vector boson in model B scenario refers to the Z' particle. Theoretically, Z' particle is from the mechanism of new broken gauge symmetry, it is named after the SM Z boson which is similarly from the mechanism of the broken U(1) gauge symmetry. In the minimal composite Higgs model, the broken SO(5) symmetry is not a gauge symmetry. However, to be consistent with other similar analyses [20], we keep this name as our search target in this thesis.

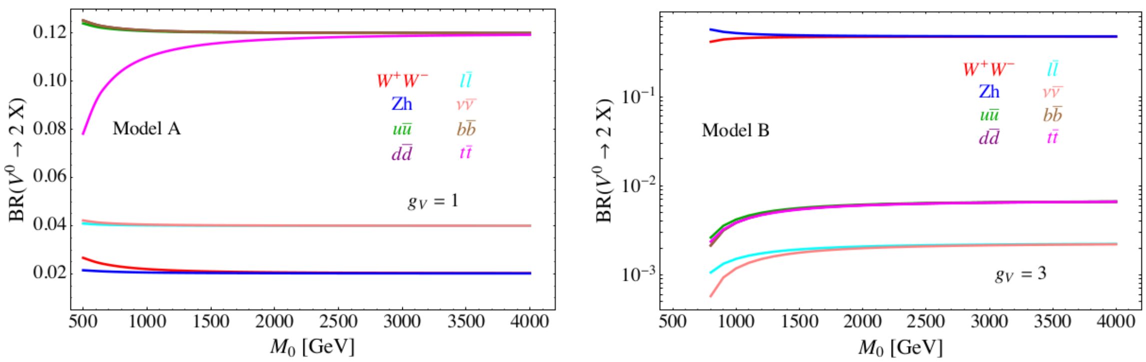


Figure 1-1: Branching ratios as a function of the resonance mass for the HVT benchmark model A(left) and model B(right).

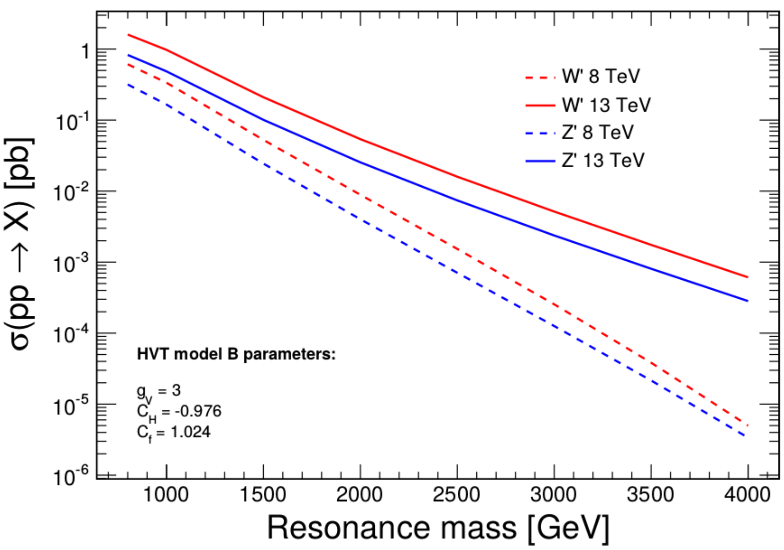


Figure 1-2: Theoretical production cross-section as a function of new resonance particles for HVT model B benchmark. Dash lines are 8 TeV predictions while solid lines are 13 TeV predictions.

Chapter 2

The LHC and the CMS Detector

This thesis is based on the data collected by the Compact Muon Solenid (CMS) detector at the Large Hadron Collider (LHC). CMS is one of the two largest detectors built at the LHC. This chapter will briefly introduce the LHC and the CMS detector.

2.1 Large Hadron Collider

The LHC is the world's most powerful hadron collider and the largest experimental facility ever. It was built by the European Organization for Nuclear Research (CERN) between 1998 and 2008 in collaboration with over 10,000 scientists and engineers from over 100 countries, as well as hundreds of universities and laboratories. It lies in a tunnel of 27 km in circumference, as deep as 175 m beneath the France−Switzerland border near Geneva. The designed maximum collison energy and highest luminosity of the LHC are 14 TeV and $10^{-34}\text{cm}^{-2}\text{s}^{-1}$, respectively.

Other accelerators that had been originally built at CERN for previous experiments serve as an injection chain for the LHC now (Fig. 2-2). The proton beam starts from LINAC, a small linear accelerator, where the energy of protons first reaches at 50 MeV. The proton beam then passes through a booster and goes to the PS, where it is accelerated up to 25 GeV. After that, it reaches 450 GeV in the SPS. The beam is finally injected in the LHC ring from the SPS, and it had been accelerated up to 4 TeV in 2012. In early 2015, the proton beam had been acceletated to 6.5 TeV, a value near its designed energy, before undergoing collision.

There are four collision points at the LHC, corresponding to four main experiments, CMS, ATLAS, LHCb and ALICE. The ALICE experiment is optimized to study heavy-ion (Pb-Pb nuclei) collisions and focusing on the physics of strongly interacting matter at extreme energy densities. LHCb is a specialized b-physics experiment, measuring the parameters of CP violation in the interactions of b-hadrons. Such studies can help to explain the matter-antimatter asymmetry of the universe. Last, CMS and ATLAS are two general purpose detectors. The aims of these two experiments are investigating a wide range of physics, including the search for the beyond standard model particles, extra dimensions, and dark matter.

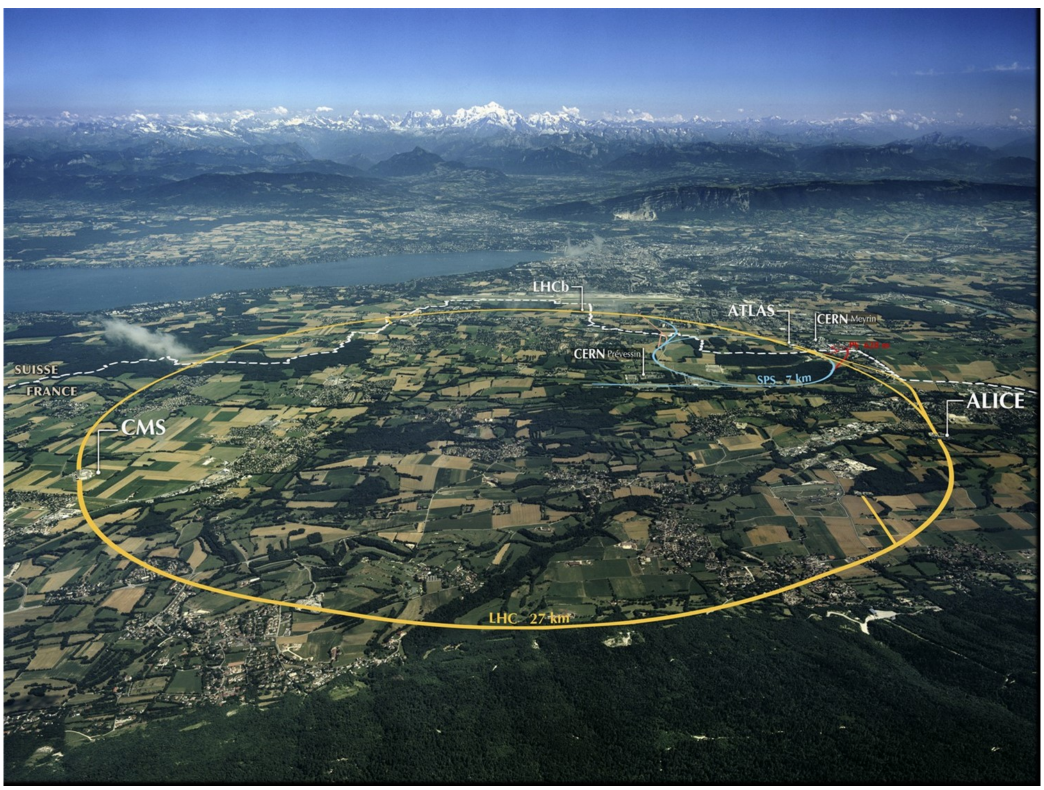


Figure 2-1: Overview of the LHC and relative location of the detectors.

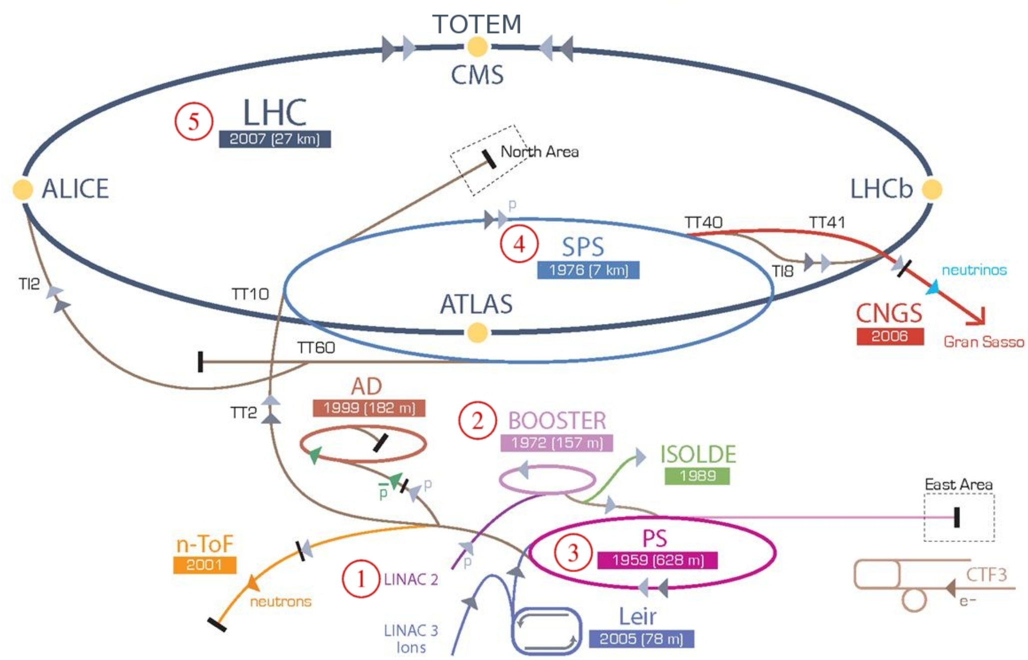


Figure 2-2: CERN accelerator complex.

2.2 Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is designed to cope very high rate of interactions expected to take place at the high LHC luminosity. It has the typical structure of detectors at hadron colliders: a central region (*barrel*) enclosed by two disks (*endcaps*). The structure of CMS can be seen in Fig. 2-3.

Solenoid and Sub-detectors

CMS features a powerful superconducting coil, generating a solenoidal magnetic field around 3.8 Tesla in a large volume which hosts different sub-detctors. The magnetic field lines close through steel yoke in the outer region. The distinct sub-detectors are designed in order to obtain the highest possible resolution and the largest acceptance for every kind of particles.

The innermost layer is a silicon-based tracker. Surrounding it is a scintillating crystal electromagnetic calorimeter (ECAL), which is itself surrounded with a sampling calorimeter for hadrons (HCAL). The tracker and the calorimeters are compact enough to fit inside the CMS Solenoid. Outside the magnet are the large muon chambers.

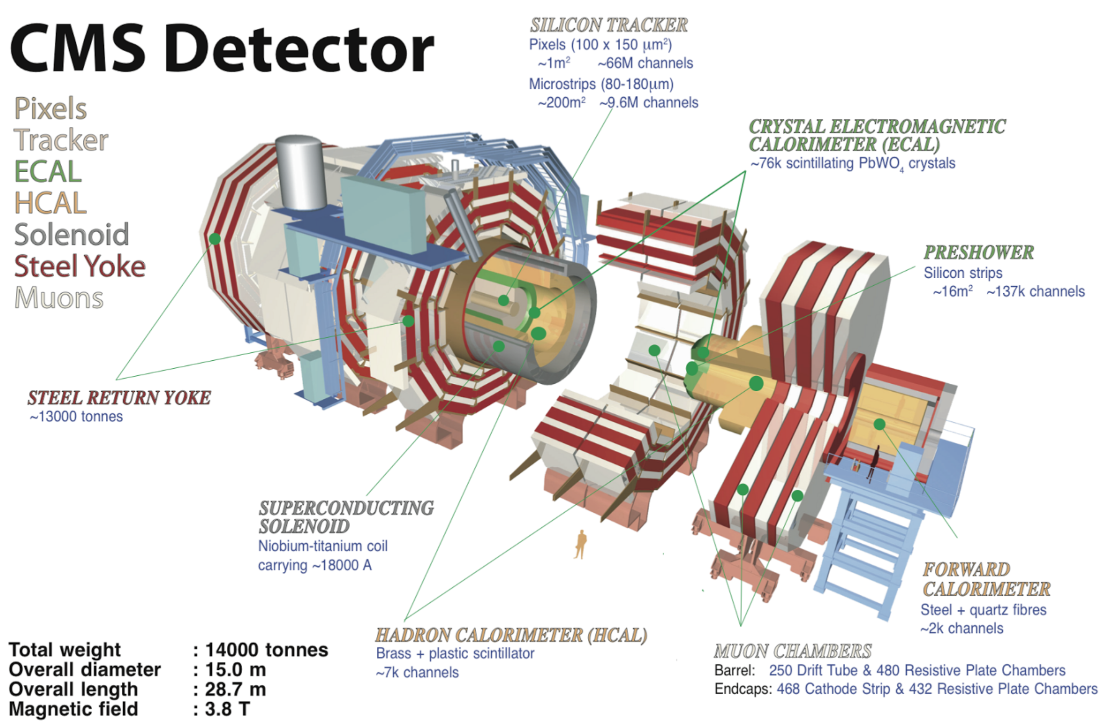


Figure 2-3: Structure overview of the CMS detector.

Coordinate System

The CMS coordinate system is oriented such that the x-axis points to the center of the LHC ring, the y-axis points vertically upward and the z-axis is in the direction of the beam. The azimuthal angle ϕ is measured from the x-axis in the $x-y$ plane and the radial coordinate in this plane is denoted by r. The polar angle θ is defined in the $r-z$ plane, while the pseudo-rapidity $\eta = -ln \tan(\theta/2)$. The momentum component transverse to the beam direction, denoted by p_T, is computed from the x- and y-components, and the transverse energy is defined as $E_T = E \sin\theta$.

2.2.1 Tracker

Tracker is the most inner part of CMS that records the productions of collisions in the first place. It traces the charged particles' trajectories. Physicists can reconstruct the vertices of the interaction and the momentum of charged particles by linking tracks to the collider's pipe and measuring the curves of particles under magnetic field.

The tracking system is composed of two kinds of detector, the pixel detector and silicon strip detector. The pixel detector is built from three barrel layers at r = 44, 73, 102 mm, and two endcap disks on each side at $z = \pm 345$, ± 465 mm.

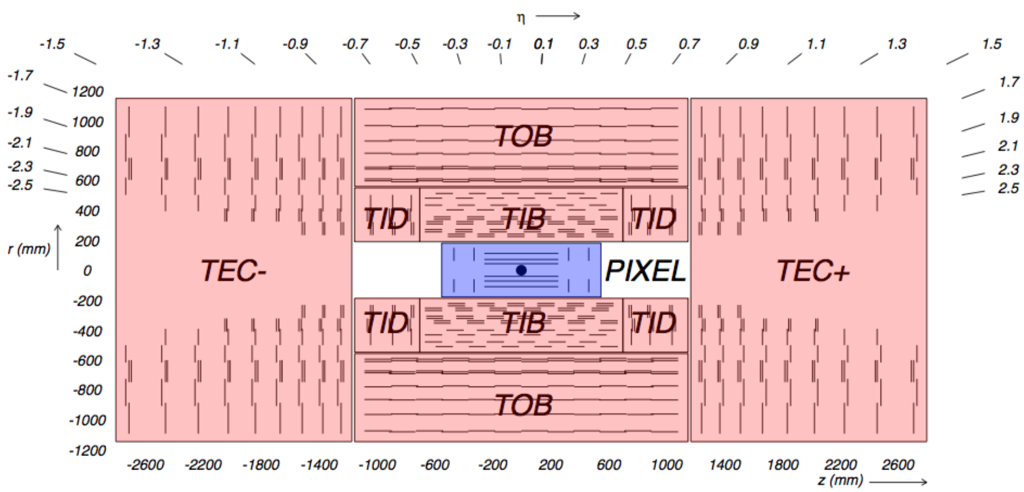


Figure 2-4: Schematic layout of tracker.

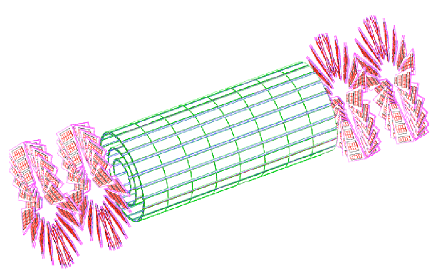


Figure 2-5: The pixel detector inside tracker.

The pixel detector, though about the size of a shoebox, consists of 1440 segmented silicon sensor modules with total 66 millon readout channels. Charge carriers are distributed over several pixels. The analog pulse height information can be used to calculate the center of certain charge distribution which could improve the hit information. The spatial resolution is measured to be about 10 μm for the $r - \phi$ plane or about 20 μm for z direction measurement.

Outside the pixel detector, there comes the silicon strip detector. The barrel region of silicon strip detector is divided into two parts, the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB). The former is composed of four layers of silicon sensors with a thickness of 320 μm and of strip pitches varying from 80 to 120 μm. The TOB is made of six layers. In this kind of sub-detector, thicker silicon sensors (500 μm) are employed, while the strip pitch varies from 120 to 180 μm. The endcap region ($|\eta| > 1.6$) is covered by the Tracker Inner Disks (TID) and the Tracker End Cap (TEC). The entire silicon strip detector is comprised of 15200 high-sensitivity modules consisting of detecting unit, supporting structure and readout electonic system.

2.2.2 ECAL

The Electromagnetic Calorimeter (ECAL) measures the energy of photons, electrons and positrons. It it is placed just outside the tracker, but still inside the solenoid. ECAL is made of 74848 lead-tungstate ($PbWO_4$) crystals. This material is characterized by a high density (8.28 g/cm^3), which gives the crystals a very compact form and makes them particularly suitable to be placed inside the magnetic coil. Another reason, this material has also a fast temporal response ($\sim$10 ns) and its radiation length (X_0) of 0.89 cm give ECAL the possibility to fully contain the expansion of the electromagnetic shower.

The arrangement of ECAL is shown in Fig. (2-6). The barrel crystals have a front face area of 2.2 $\times$ 2.2 cm^2 and a length of 23 cm. They are positioned at $r = 1.29$ m in pseudo-rapidity region $0 < |\eta| < 1.479$. The crystals in the endcaps have a 2.47 $\times$ 2.47 cm^2 front face, a 22 cm length and they are positioned at $z = 3.17$ m in $1.479 < |\eta| < 3.0$. A Preshower detector is placed in front of the endcaps crystals. The active elements of Preshower are two planes of silicon strips with a pitch of 1.9 mm, which lie behind disks of lead absorber at depths of $2X_0$ and $3X_0$. It allows the rejection of photon pairs from π^0 decays and improves the estimation of the direction of photons, to enhance the measurement of the two-photon invariant mass.

The energy resolution of the ECAL is given by three different contributions [21] (E in GeV),

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{2.1}$$

where the first term is statistical in nature, it also contains fluctuation in showering and in the amplification through photodiodes, the second one considers electronic noise and pile-up, the last term is mainly due to the calibration.

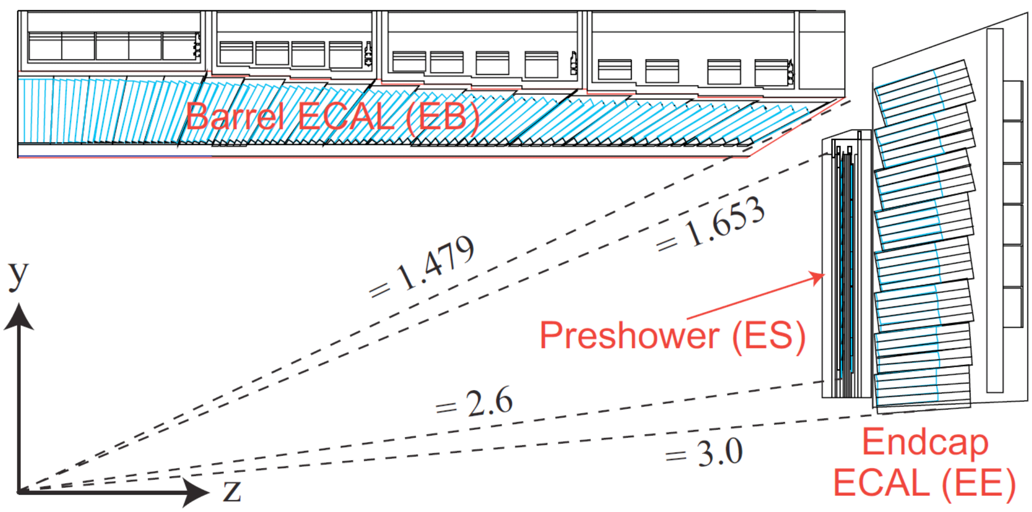


Figure 2-6: Schematic layout of the CMS ECAL.

2.2.3 HCAL

The hadron calorimeter (HCAL) is placed mainly between ECAL and the magnet coil. It measures the energy of hadrons and mesons. Additionally it provides indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos. The design is strongly influenced by these aims, hence an important requirement is the high hermeticity (the ability to capture every particle emerging from the collisions). This means the detector must cover the biggest possible portion of the solid angle.

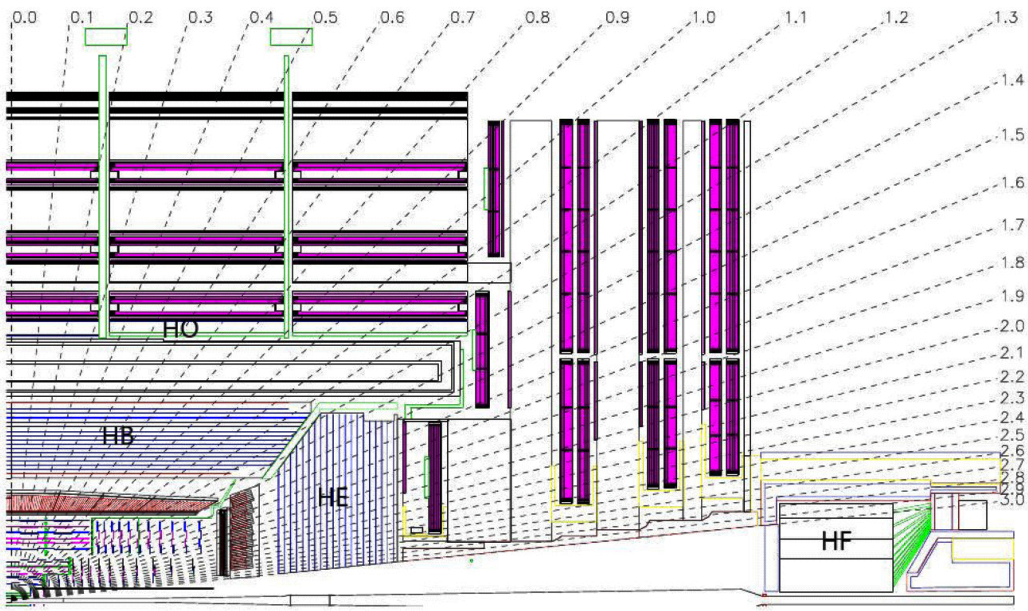


Figure 2-7: Longitudinal view of one quarter of CMS and the location of HB (hadron barrel calorimeter), HE (hadron endcap calorimeter), HF (hadron forward calorimeter) and HO (hadron outer calorimeter)

For this reason, a hadron forward calorimeter is required, which is placed outside the magnet return yokes, with a total coverage of $3 < |\eta| < 5.31$ at 11 m from the interaction point. Moreover, an outer hadronic calorimeter is placed in the first muon absorber layer in order to enhance the containment of high energy jets in the central region of the detector.

HCAL is a sampling calorimeter, whose active elements are plastic scintillators interleaved with brass absorber plates and read out by wavelength shifting fibers. Brass has been chosen as absorber material for its short interaction length and because it is non-magnetic. The thickness of the absorber layers is between 60 mm in the barrel and 80 mm in the endcaps. The barrel has 5.46 interaction lengths at $\eta = 0$ and 10.82 at $\eta = 1.3$, while the endcaps have an average of 11 interaction lengths [22].

The HCAL energy resolution (E in GeV and measured by pion) [23] is

$$\frac{\sigma_E}{E} \simeq \frac{a}{\sqrt{E}} \oplus 5\% \tag{2.2}$$

where a $\simeq$ 65% in the barrel, a $\simeq$ 85% in the endcaps and a $\simeq$ 100% in the HF.

2.2.4 Muon Chamber

The efficient detection of muons has primary importance, as muons represent a clear signature for a large number of processes. Muons can penetrate several meters of iron without interacting. Unlike most particles, they are not stopped by any of calorimeters in CMS. Therefore, chambers to detect muons are placed at the very edge of the experiment where they are the only particles likely to register a signal.

The muon system fulfills three purposes, muon identification, momentum measurement and triggering. Three different types of gaseous detectors are used for CMS muon system depending on the requirements.

Drift Tube

The drift tube (DT) system measures muon positions in the barrel part of the detector. Each DT chamber, on average 2 m $\times$ 2.5 m in size, consists of 12 aluminium layers, arranged in three groups of four segmentations, each with up to 60 4-cm-wide tubes that contain a stretched wire within each gas volume. The middle group measures the coordinate along the direction parallel to the beam and the two outside groups measure the perpendicular coordinate.

Cathod Strip Chamber

In the two endcaps, where the muon flux and the residual inhomogeneous magnetic field are higher, cathode strip chambers (CSC) are used. CSC is composed of anode wires and cathod strips in the gas volume. The chambers are arranged in 4 disks perpendicular to the beam, and in concentric rings (3 rings in the innermost station, 2 in the others) in each of the endcaps.

Resistive Plate Chambers

Resistive plate chambers (RPC) are fast gaseous detectors that provide a muon trigger system parallel with DTs and CSCs. Each RPC consists of two parallel plates, a positively charged anode and a negatively charged cathode, both made of a very high resistivity plastic material and separated by a gas volume.

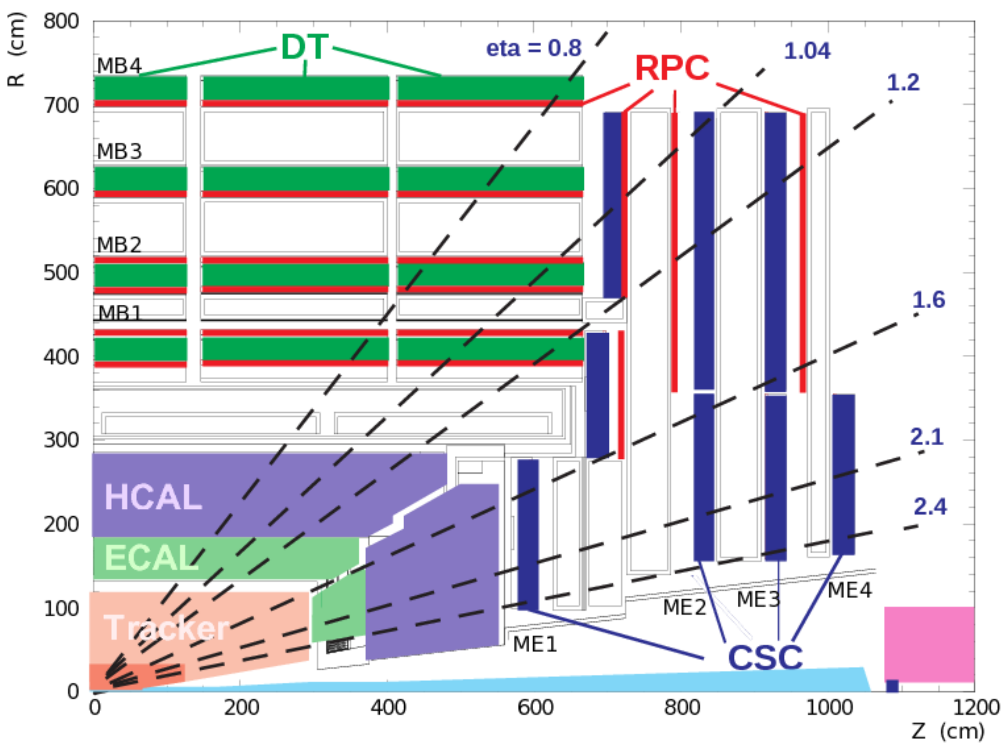


Figure 2-8: Slice view of one quarter of muon chamber system.

2.2.5 Trigger System

To have a good chance of producing rare particles, a very large number of collisions is required (LHC proton bunches collide every 25 ns). Most collision events in the detector are soft and do not produce interesting events. The amount of data from each crossing is approximately 1 megabytes, which at the 40 MHz crossing rate would result in 40 terabytes of data a second, an amount that the experiment cannot store. The task of the trigger system is to reduce the storage rate while keeping a high efficiency on the potentially interesting events. In CMS the input rate is reduced in two steps, Level-1 Trigger (L1T) and High-Level Trigger (HLT).

Level-1 Trigger

After the Level-1 Trigger selection, the event recording frequency is decreased to 100kHz [24], which is much smaller than the collision rate. The L1T objects are particles (such as photons, muons and elections), jet candidates, global transverse energy and missing transverse energy. Level-1 Trigger just chooses the event with E_T and P_T higher than the thresholds.

High Level Trigger

High Level Trigger is behind the readout buffers after Level-1 Trigger. It reduces the data output rate to 100Hz by using all the information from CMS including the sub-detectors. The reconstruction algorithms are the same as the off-line analysis. However, the triggering procedure doesn't need maximal precision, therefore these algorithms are modified to be faster even with lower resolution.

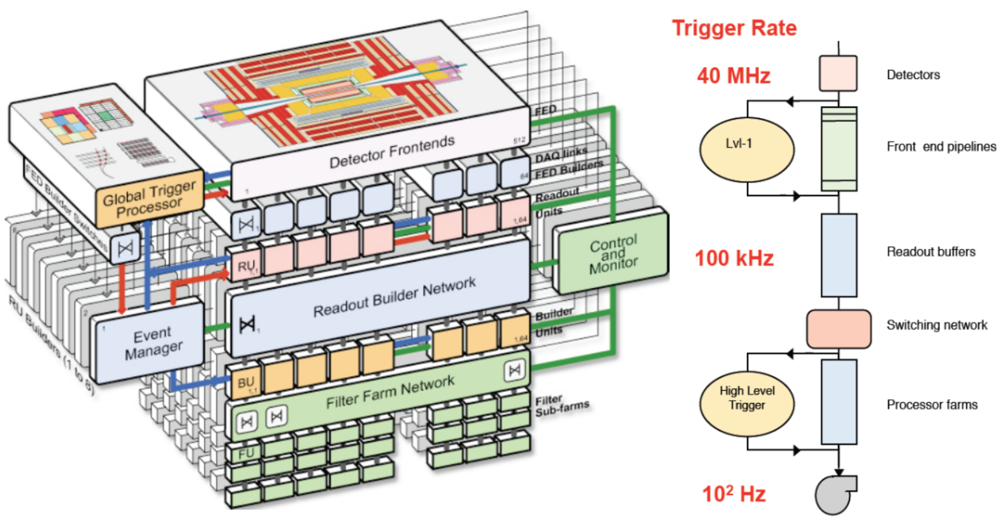


Figure 2-9: CMS triggering and data acquisition architecture.

Chapter 3

Analysis Procedures

In this chapter, the analysis procedures of the search for Z' decaying into Zh in $llbb$ final state are reported. The data sets and Monte Carlo (MC) samples we used in this analysis will be indicated. Physics objects reconstruction and event selections are also introduced. Moreover, background yields and the effects of systematic uncertainties will be discussed in the end of this chapter.

3.1 Monte Carlo Samples and Data sets

3.1.1 Signal MC

As introduced in section 1.2.3, the signal hypothesis is HVT model B benchmark. The heavy resonance (Z') is tested using a wide set of masses from 800 GeV to 2000 GeV, one masspoint every 100 GeV (Table 3.1). The signal is generated by MadGraph5_aMC@NLOv5.2.2.1 [25] in LO mode, as a narrow spin-1 neutral resonance and is forced to decay in the $Z' \rightarrow Zh \rightarrow llqq$ channel. Showering and hadronization are performed with PYTHIA6 [26].

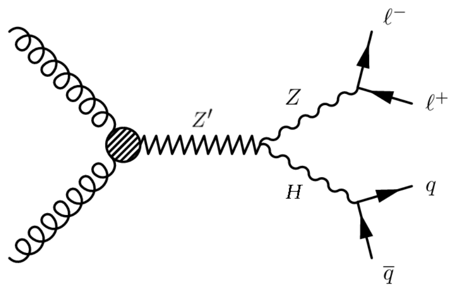


Figure 3-1: Feynman diagram for $Z' \to Zh \to 2l2q$.

Sample	Number of Processed Events	σ_{LO}(pb)
ZPrime_ZH_lljj_M800-MADGRAPH	10710	0.00685367
ZPrime_ZH_lljj_M900-MADGRAPH	10209	0.00485861
ZPrime_ZH_lljj_M1000-MADGRAPH	19997	0.003263
ZPrime_ZH_lljj_M1100-MADGRAPH	9370	0.00217483
ZPrime_ZH_lljj_M1200-MADGRAPH	10710	0.00145484
ZPrime_ZH_lljj_M1300-MADGRAPH	9369	0.000979745
ZPrime_ZH_lljj_M1400-MADGRAPH	10497	0.000664783
ZPrime_ZH_lljj_M1500-MADGRAPH	19999	0.000454339
ZPrime_ZH_lljj_M1600-MADGRAPH	8950	0.000312541
ZPrime_ZH_lljj_M1700-MADGRAPH	9369	0.000216282
ZPrime_ZH_lljj_M1800-MADGRAPH	10708	0.000150398
ZPrime_ZH_lljj_M1900-MADGRAPH	10498	0.000105039
ZPrime_ZH_lljj_M2000-MADGRAPH	19999	7.36377e-05

Table 3.1: Signal samples used in the analysis.

3.1.2 Background MC

Since we are looking for new resonances decaying in semi-leptonic final state, the background samples of this analysis are originated by all SM events with two leptons and at least one jet as final state. The dominant background contribution is the produciton of Z boson with jets. This Z+jets sample is produced by MADGRAPH. In the matrix element level, the Z boson is forced to decay into two leptons, and further this sample is divided into two samples depending on the Z p_T, higher than 100 GeV or between 70 and 100 GeV. The contribution of events with Z p_T less than 70 GeV is negligible due to further cut on the objects p_T in the selection criteria.

The second dominant source of background is $t\bar{t}$ production. Both of the two top quarks decay into the final state (top decays into a W boson and a b quark first) that gives two charged leptons, neutrinos and two b-jets. This sample is generated by POWHEGv1 [27].

Other sources of background considered are SM di-boson productions (WW, WZ and ZZ) generated by PYTHIA6. All the background samples are required to pass phase-space cuts, p_T^{ll} >60 GeV and 60< M_{ll} <120 GeV. Related statistics are reported in Table 3.2.

Sample	Number of Processed Events	σ_{NLO}(pb)
DYJetsToLL_PtZ-70To100	11764538	63.5
DYJetsToLL_PtZ-100	12511326	39.4
TTTo2L2Nu2B	10783509	25.8
WW	7759752	56.0
WZ	9910267	22.4
ZZ	9769891	7.6

Table 3.2: Background samples used in the analysis.

3.1.3 Data Samples

In this analysis, the full CMS data collected in 2012 are used, corresponding to the integrated luminosity of 19.7 fb^{-1} at the center-of-mass energy $\sqrt{s}$ =8 TeV. For each lepton channel, there are four datasets. All datasets are collected with a double muon or a double electron trigger, as explained in detail in the next section. The trigger algorithm employed for the electron samples doesn't use any information from the tracker but only the energy deposite in the ECAL. This expedient is implemented in order to avoid any possible inefficiencies due to the presence of two tracks very close to each other when the Z is highly boosted and its decay products are very collimated. Such a trigger is contained in the Photon/DoublePhotonHighPt dataset. The full dataset names are listed in Table 3.3.

AOD Sample	Luminosity (pb^{-1})
DoubleMu/Run2012A-22Jan2013-v1	876.225
DoubleMuParked/Run2012B-22Jan2013-v1	4409
DoubleMuParked/Run2012C-22Jan2013-v1	7017
DoubleMuParked/Run2012D-22Jan2013-v1	7369
Photon/Run2012A-22Jan2013-v1	876.225
DoublePhotonHighPt/Run2012B-22Jan2013-v1	4412
DoublePhotonHighPt/Run2012C-22Jan2013-v1	7055
DoublePhotonHighPt/Run2012D-22Jan2013-v1	7369

Table 3.3: Data sets used in this analysis.

3.2 Trigger

Since the final state contains two leptons of the same flavour and at least one jet, we perform this analysis on the DoubleMu and Photon/DoublePhotonHighPt datasets. The first dataset is triggered by two muons, the second one is triggered by two eletrons. These triggers are:

- HLT_Mu22_TkMu8* (for DoubleMu datasets)
- HLT_DoubleEle33_* (for Photon/DoublePhontonHighPt datasets)

The muon trigger has a double p_T threshold, requires leading muon p_T greater than 22 GeV and sub-leading muon p_T greater than 8 GeV. Differently, the double electron trigger requires a higher threshold of 33 GeV on both electrons. The trigger efficiencies are close to 1 in both cases.

3.3 Physics Objects

3.3.1 Muon

Reconstruction

The muon reconstruction algorithm at CMS takes advantage of the redundancy of detection methods. Muon tracks are first reconstructed independently in the inner tracker (tracker track) and in the muon system (standalone track). Based on these objects, two reconstruction approaches are used [28]:

- *Globol Muon* (outside-in): Starting from a standalone track, this algorithm finds a best tracker track to match the standalone track. Then, the fit of the track is repeated using the hits both in the tracker and in the muon system [29]. The resulting object is called a *Global Muon*. At large transverse momentum (p_t >200 GeV), the global muon fit can improve the momentum resolution compared to the tracker-only fit.

- *Tracker Muon* (inside-out): A tracker muon is reconstructed using an opposite direction of the direction for a global muon. In this approach, all tracker tracks with $p_T > 0.5$ GeV and the total momentum $p > 2.5$ GeV are considered as possible muon candidates. The extrapolation to the muon system takes into account the magnetic field, average expected energy losses, and multiple scattering in the detector material. If at least one muon segment matches the extrapolated track, the corresponding track track qualifies as a *Tracker Muon*. This algorithm is useful for low-p_T muons that do not fully penetrate the muon system, and therefore only register a few hits.

If no match is found when extrapolating outside-in, the standalone track is stored as a *Stanalone Muon*. This happens only for less than 1% of the muons produced in a collison, and the reconstruction efficiency is about 99% for the muon which carries enough high momentum within detector coverage [28].

Identification

We use both tracker muons and global muons in this analysis. To identify muons from the signal, the muons must pass one of these two off-line selections, high$-p_T$ muon ID or tracker-based muon ID [30]. The requirements are listed as follows:

High-p_T muon ID

- Muon identified as a *Global Muon*.
- Number of muon hits in the global track > 0.
- Number of matched muon stations > 1.
- Number of pixel hits > 0.
- Number of tracker layers with hits > 8.
- Transverse impact parameter $d_{xy} < 0.2$ cm.
- Longitudinal impact parameter $d_z < 0.5$ cm.
- Relative error on the track transverse momentum $\sigma_{p_T}/p_T < 0.3$.

In the tracker-based muon ID, the muon has to be identified as a *Tracker Muon*, and the requirement of muon hits in the global track is removed. Other requirements are the same.

An additional useful variable for lepton identification is the isolation. It is defined as the scalar sum of the p_T of the reconstructed objects within a cone (typical size is $\Delta R = 0.3$, $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$) space around the lepton track but excluding the p_T of the lepton itself. Moreover, the relative isolation is defined as isolation divided by the lepton p_T ($I_{rel} = Iso/p_T^{lep}$). The relative isolation is more frequently used in the modern analysis.

In this analysis, a modified isolation criteria is used. The two muons originated from boosted Z decay are close to each other, and consequently the presence of another muon in the isolation cone could break the function of this variable. In order to solve this problem we use a tracker-based isolation relative quantity, explicitely removing the momentum flow from any other muon passing our muon selection. Moreover, a tracker-based isolation is well motivated also by two additional aspects: it is more independent of pile up (pile-up tracks tipically do not match the primary vertex) and does not include possible muon radiation. Finally, the modified requirement is $I_{rel}^{mod} < 0.1$.

Variable	High-p_T	Tracker-based
Muon type	Global muon	Tracker muon
Muon hits in global track	≥ 1	-
Muon stations matched	≥ 2	≥ 2
d_{xy}	< 0.2 cm	< 0.2 cm
d_z	< 0.5 cm	< 0.5 cm
Pixel hits	≥ 1	≥ 1
Tracker layers	≥ 8	≥ 8
σ_{p_T}/p_T	< 0.3	< 0.3
I_{rel}^{mod}	< 0.1	< 0.1

Table 3.4: Summary of the muon ID selection criteria.

3.3.2 Electron

Reconstruction

Electrons are reconstructed from energy deposits in the ECAL matched to tracks reconstructed in the silicon tracker. They have less distinguishable signatures than muons in the detector due to the jet-induced background in ECAL. Therefore, to reconstruct an election, it's essential to find a briliant way to match the infomation acquired in both tracker and ECAL.

Clustering

For every single hit from an electron in ECAL, approximately 94% of energy is contained in 3×3 crystals and 97% in 5×5 crystals. To form a cluster, crystals with $E_T > 1$ GeV is picked as seeds. Then starting from seed crystal, dominos of size 1×3 or 1×5 are created in $\eta - \phi$ plane. If the energy of these dominos exceed 0.1 GeV threshold, thus added up the energy of dominos along the ϕ direction.

Moreover, the effects of conversion process and bremsstrahlung radiation must be taken into account to reconstruct the electron energy more precisely. The concept of electron superclustering is to collect the energy of photons from bremsstrahlung radiation along the ϕ direction with fixed η width of the seed crystal. The technical details of the supercluster can be found in [31].

Tracking

There are several steps to reconstruct electron tracks in CMS experiment. The hit on the pixel and the tracker are reconstructed in the first step. The second step is to find the seed of tracks by matching at least two hits in pixel detector. Lst step is to fit the trajetory starting from the seed. To perform this fit, the Gaussian-sum filter (GSF) algorithm is used [32]. In GSF algorithm, the energy loss probability density function is constructed by multiple Gaussian functions. Finally, the electrons are built by matching the superclusters to the GSF tracks. Reconstruction efficiencies for electrons $E_T > 20$ GeV are generally above 95% in EB and 90% in EE [33].

Identification

The electron identification used in this analysis is based on HEEPv4.1 [34]. As what we did for the muon id, we need to modify the isolation definition again because of the small ΔR between two electrons. The requirements are listed as follows:

Modified HEEPv4.1 electron ID

- Transversal supercluster energy > 40 GeV.
- Pseudorapidity of supercluster $|\eta_{\mathrm{SC}}| < 1.442$ for barrel electrons, or $1.556 < |\eta_{\mathrm{SC}}| < 2.5$ for endcap electrons.
- Have either $E_{2\times5}/E_{5\times5} > 0.94$ or $E_{1\times5}/E_{5\times5} > 0.83$.
- Ratio of HCAL deposit energy to ECAL deposit energy (Fig. 3-2) smaller than 0.5.
- Number of inner layer lost hits smaller than 2.
- Have $|d_{xy}| < 0.02$ cm for barrel electrons, or $|d_{xy}| < 0.05$ cm for endcap electrons.

As what we did for the muon id, by the same reason we need to modify the isolation definition again. In this case there are three isolation variables that have to be changed.

- Modified track isolation is required be to lower than 5 GeV. This variable is defined as the scalar p_T sum of the tracks within a $\Delta R = 0.3$ cone around the electron, and excluding the p_T of another electron which passes the above selections and its track is inside the cone.
- The electromagnetic calorimeter isolation I_{ECAL} is defined as the scalar sum of E_T of the crystals in a $\Delta R = 0.3$ cone around the particular electron (an

inner area of full-width 3 crystal), excluding a 4 crystals width around any other electron. The dimension of the ECAL crystals corresponds roughly at $\Delta R \sim$ 0.01 to 0.02. The threshold of I_{ECAL} is varying with the electron transverse energy.

- The hadronic calorimeter isolation I_{HCAL} is defined as the scalar sum of E_T of the HCAL caloTowers with a center in a $\Delta R = 0.3$ cone around the electron, excluding those lying within $\Delta R = 0.15$ of the electron itself and of any other electron [35]. The threshold of this variable also varies with the electron transverse energy.

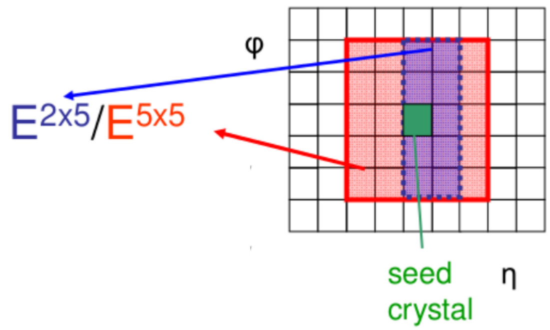


Figure 3-2: Example of the $E_{2\times5}/E_{5\times5}$ computation in the ECAL crystals. $E_{i\times j}$ is the energy contained in a $i \times j$ block around the seed crystal (defined as the highest deposit of the energy of the cluster).

Variable	Barrel	Endcap
E_T	> 40 GeV	> 40 GeV
$\lvert\eta_{SC}\rvert$	$\lvert\eta_{SC}\rvert < 1.442$	$1.556 < \lvert\eta_{SC}\rvert < 2.5$
H/E	< 0.05	< 0.05
$E_{2\times5}/E_{5\times5}$	> 0.94	-
	or $E_{1\times5}/E_{5\times5} > 0.83$	
Inner Layer Lost Hits	$<= 1$	$<= 1$
$\lvert d_{xy}\rvert$	< 0.02 cm	< 0.05 cm
Iso^{mod}_{Trk}	< 5 GeV	< 5 GeV
$Iso^{mod}_{Ecal,Hcal}$	< 2 GeV $+ 0.03E_T$	< 2.5 GeV for $E_T < 50$ GeV < 2.5 GeV $+ 0.03E_T$ for $E_T > 50$ GeV

Table 3.5: Summary of the modified HEEPv4.1 electron ID.

3.3.3 Jet

Since gluons and quarks cannot exist in free state due to the color confinement [36], they fragment into hadrons. The result of this fragmentation (or called the "hadronization") is a jet of particles depositing energy in the detectors.

Reconstruction

Jet clustering algorithms are among the main tools for analyzing data from hadronic collisions. In this section, an extensively used clustering algorithm called "sequential recombination algorithm" [37,38] will be introduced. At first, events are reconstructed using the particle flow algorithm (PF) [39], which attempts to identify and measure all the stable particles in a collision by combining information from all the sub-detectors. The PF algorithm categorizes all particles into five types: muons, electrons, photons, charged and neutral hadrons. The resulting particle flow candidates are passed to the anti-k$_T$ (AK) [37] or Cambridge/Aachen (CA) [38,40] jet clustering algorithms to create jets (AK and CA are two commonly used branches of sequential recombination algorithms).

The jet clustering algorithms are implemented as follows:

$$
\begin{aligned}
d_{iB} &= p_{T,i}^{2p} \\
d_{ij} &= min(p_{T,i}^{2p}, p_{T,j}^{2p})\frac{\Delta R_{i,j}^2}{R^2}
\end{aligned}
\tag{3.1}
$$

In Eq. (3.1), the pair of objects (i , j) denote two input candidate particles to the algorithm. ΔR_{ij} ($\Delta R_{ij} \equiv \sqrt{\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2}$) is the spatial distance between the two objects, and R is the typical cone size of the jet. The parameter p corresponds to different catagories of algorithms and will be introduced in the later paragraph.

The clustering proceeds by comparing the value of the two quantities, d_{iB} and d_{ij}. If $d_{ij} > d_{iB}$, the algorithm will look for another possible combination of (i , j). If $d_{ij} < d_{iB}$, the 4-momentum of object i will merge into j, and then the object j forms a pseudo-jet, but the algorithm will not stop looking for a new object until all particles are clustered into jet.

The parameter p can be chosen as follows:

- $p = 0$: Cambridge/Aachen algorithm;
- $p = -1$: Anti-k_T algorithm;

The difference between CA/AK algorithms is clear, CA algorithm is only considering the spatial distance of the two candidate objects (if $\Delta R_{ij} < R$, merged). In the case of anti-k$_T$ algorithm, p_T of the object presents as a weight for the distance.

The anti-k$_T$ algorithm has better energy calibration, while the CA algorithm was found to be more efficient at finding hard subjets within the jets in simulation-based studies [41]. Therefore the CA jets are used in this analysis to identify events containing hadronically-decaying Higgs bosons.

Identification

As disscussed in the previous section, this analysis use the Cambridge/Aachen algorithm with a cone radius of R = 0.8 for the identification of jets (CA8 jets). Furthermore, jets are required to pass the following loose identification criteria [42, 43]:

Loose JetID ($>$ 99% efficiency)

- Muon energy fraction smaller than 0.99
- Photon energy fraction smaller than 0.99
- Charged electromagnetic energy fraction smaller than 0.99
- Neutral hadron energy fraction smaller than 0.99
- Charged hadron energy fraction larger than 0
- Number of constituent particles larger than 1

For all jet candidates, $|\eta| < 2.4$ requirement are also applied.

3.3.4 Jet Grooming Algorithms

The jet mass is the main observable in distinguishing a boson-jet from a QCD jet. Jet grooming aims to suppress underlying events and pile-up radiation from the target jet, and improves the discrimination by pushing the jet mass for QCD jets towards lower values while mantaining the jet mass for boson-jets around the boson-mass [44, 45].

Typically three different grooming algorithms are considered, trimming [46], filtering [47] or pruning [48, 49]. Studies of these different grooming methods in CMS are presented in [50].

- **Trimming algorithm:** Trimming is a technique that ignores subjets (smaller jets formed by the breakup of a larger one) below a minimum p_T fraction threshold within the original jet. Trimming reclusters the jet's constituents with a smaller radius R_{sub}. Then accepts only the subjets that have $p_{T,frac} > f_{cut}$, where the $p_{T,frac}$ is defined as subjet p_T devided by the original jet p_T, and f_{cut} is typically taken proportional to H_T, the scalar sum of the p_T of all jet reconstructed in the event.

- **Filtering algorithm:** Filtering reclusters jet constituents with smaller radius R_{filt}, and sorts the subjets in order of their p_T. The final jet is thus redefined as the sum of the four-momentum of the n hardest subjets. By default the factor is n=3, but it's not a fixed number, n depands on the analysis.

- **Pruning algorithm:** The idea is to take a jet of interest and then to recluster it using a sequential clustering algorithm for vetoing soft and large-angle recombinations between pseudojets i and j.
  Clustering proceeds as explained in the previous section, but it is vetoed if the candidates are too far away in ΔR.

$$\text{veto if } \Delta R_{ij} > r_{cut} \times 2m/p_T \tag{3.2}$$

And the energy sharing is too asymmetric.

$$\text{veto if } z_{ij} = \frac{min(p_{T,i}, p_{T,j})}{p_{T,i+j}} < z_{cut} \tag{3.3}$$

Where z_{cut} and r_{cut} are parameters of the algorithm (default value: $z_{cut} = 0.1$, $r_{cut} = 0.5$). If both these conditions are satisfied the softer one of the two candidate is not considered.

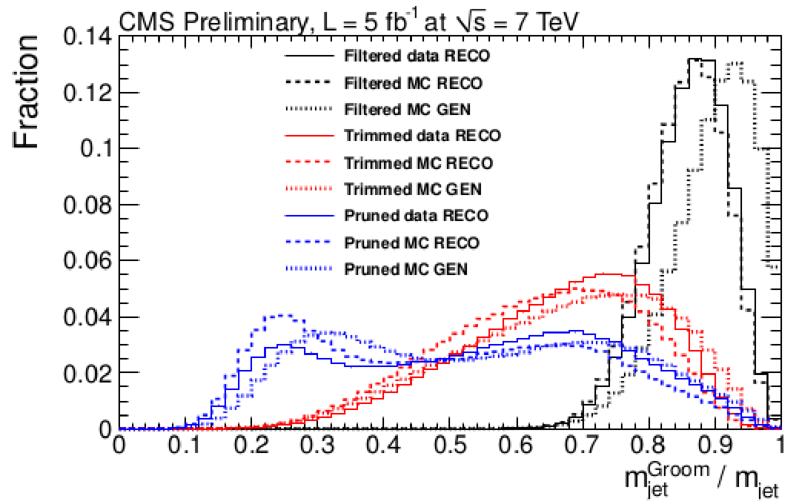


Figure 3-3: Comparison of the jet mass in generic QCD events from the groomed jets divided by the jet mass of matched ungroomed jets for the three grooming techniques, for both data and the PYTHIA 6 Monte Carlo. [1] Events are collected with a single jet trigger.

As shown in Fig. 3-3, the filtering algorithm is the least aggressive grooming technique, with groomed jet masses close to the original case. The trimming algorithm is moderately aggressive which produces a much wider final mass distribution. Pruning is the most aggressive tecnique and a bimodal distribution begins to appear.

In cases where the pruned jet mass is small, jets usually have most of their energy configured in core components with little gluon radiation, which leads to narrow jets. Instead, when the pruned jet mass is large, the jets are split more symmetrically and more similar with the boson-jet structure. In this analysis we use the pruned jet mass because of its capability to improve the jet mass resolution and background rejection.

3.3.5 b-tagging

b-tagging [2, 51] is an algorithm to identify jets originating from b-quarks. It is an important component in analyses searching for new physics. Particularly challenging are those involving top quark or Higgs boson production with decays in the boosted regime.

The methods for b-tagging are based on the unique features of b-jets: B-hadrons have sufficient lifetime that they can travel some distance before they decaying. On the other hand, their lifetimes are not so high as those of light quark hadrons, so they decay inside the detector rather than escape. Another characteristic is, the bottom quark is much more massive than anything it decays into. Thus its decay products tend to have higher transverse momentum. This causes b-jets to be wider, have higher multiplicities and invariant masses, and also to contain low-energy leptons with momentum perpendicular to the jet.

There have different approaches of b-tagging algorithm described in [2]. In this analysis, a complex approach we used will be introduced in the next section.

Identification using Combined Secondary Vertex

Combined Secondary Vertex (CSV) algorithm is a complex approach involves the information of secondary vertices and track-based lifetime. In order to enchance b purity, a secondary vertex must meet the following requirements:

- secondary vertices must share less than 65% of their associated tracks with the primary vertex and the significance of the radial distance between the two vertices has to exceed 3σ;
- secondary vertex candidates with a radial distance of more than 2.5 cm with respect to the primary vertex, with masses compatible with the mass of K^0 or exceeding 6.5 GeV/c^2 are rejected, reducing the contamination by vertices corresponding to the interactions of particles with the detector material and by decays of long-lived mesons;

- the flight direction of each candidate has to be within a cone of $\Delta R < 0.5$ around the jet direction.

There are also requirements to the tracks that forms the primary vertex, the detail reconstruction method and selections are in [2]. After all selections, the following set of variables with high discriminating power and low correlations is used to build likelihood functions:

- the vertex category (real, "pseudo" or "no vertex");
- flight distance significance in the transverse plane ("2D");
- the vertex mass;
- number of tracks at the vertex;
- ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
- pseudorapidities of the tracks at the vertex with respect to the jet axis;
- the 2D IP (impact parameter) significance of the first track that raises the invariant mass above the charm threshold of 1.5 GeV/c^2 (tracks are ordered by decreasing IP significance and the mass of the system is recalculated after adding each track);
- the number of tracks in the jet;
- the 3D IP significances for each track in the jet.

Two likelihood ratios are built from these variables. They are used to discriminate between b and c jets and between b and light-parton jets. They are combined with prior weights of 0.25 and 0.75, respectively. The combined value is the CSV discriminator. By using these additional variables, the CSV algorithm provides discrimination also in cases when no secondary vertices are found, increasing the maximum efficiency with respect to the other algorithms (in the "no vertex" category only the last two variables in the set are available). The distributions of the vertex multiplicity and of the CSV discriminator are shown in Fig. 3-4.

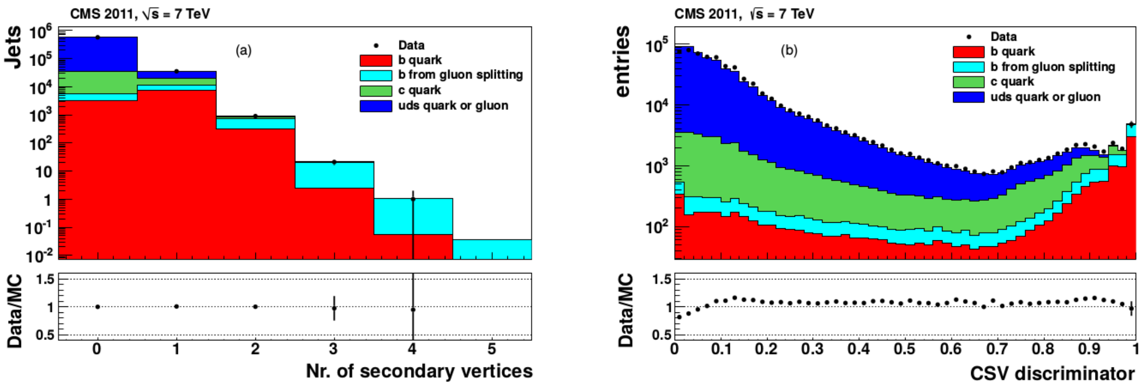
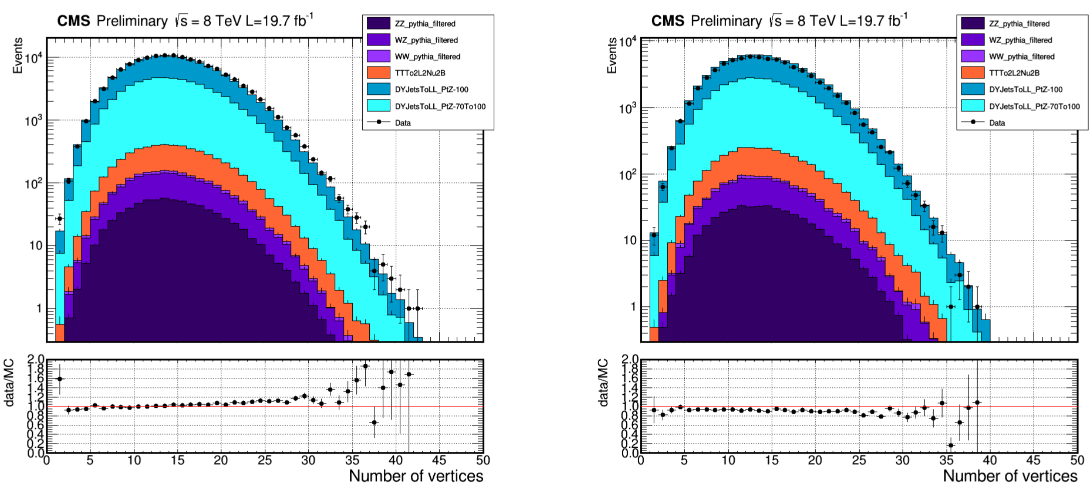


Figure 3-4: [2] Distributions of (a) the secondary vertex multiplicity and (b) the CSV discriminator.

3.4 Pile-up reweighting

At the typical luminosity provided by the LHC, it is common to reconstruct more than one vertex per event. The main event vertex is defined as the one with the highest sum of the p_T^2 of the associated tracks. The presence of additional interactions, known as pile-up (PU).

The simulation generates the pile-up roughly to match the condition in data, however there are still difference between the pile-up numbers in data and MC. It is neccesary to reweight pile-up distribtuions of MC samples to match the data more precisely. By applying a proper weight to each MC event according to the pile-up distribution from data, the MC samples can describe the data better. Fig. 3-5 shows the number of vertices after pile-up reweighting in both two lepton channel.



(a) Number of vertices in muon channel

(b) Number of vertices in electron channel

Figure 3-5: Number of vertices distributions after pile-up reweighting. Data are compared to the combination of all MC background samples. After reweighting, the distributions are almost identical to the data in both channels.

3.5 Event and Object selection

3.5.1 Lepton Requirements

Muon Selection

Besides the muon ID criteria disscussed in section 3.3.1 (Table 3.4), we also require kinematic cuts on the muon candidates. We require that the transverse momentum of the leading muon candidate must be greater than 40 GeV, while the second leading muon transverse momentum minimum threshold is 20 GeV. All muon candidates must be in the psuedo-rapidity region of $|\eta| < 2.4$.

Electron Selection

Kinematic cuts on the electron candidates are also applied. Although the electron ID selection (Table 3.5) already required the pseudo-rapidity of electron supercluster, we cut on the $|\eta| < 2.5$ for electron candidates and all electrons must be outside of [1.4442,1.566] in the η region to avoid the ECAL gap. The p_T requirement is a bit different from the muon case. Since the HLT trigger already selects electron p_T greater than 33 GeV, we require both leading and sub-leading electrons p_T greater than 40 GeV in advance.

3.5.2 Jet Requirement

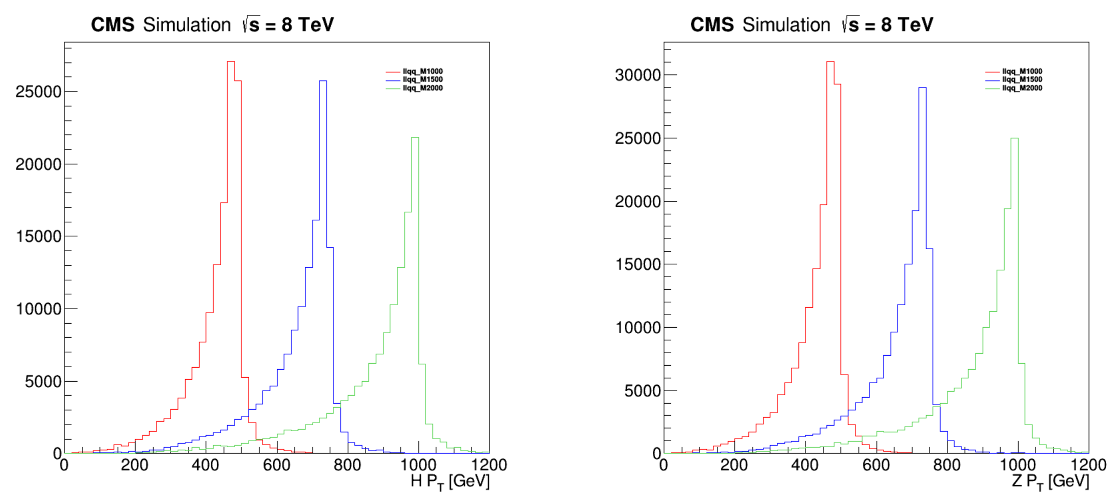
CA8jets in our signal process originate from Higgs decay. If the Z' mass is large enough, the Higgs will be boosted. Therefore we require higher kinematic thresholds to the CA8jets. In every event, there must be at least one CA8jet with $p_T > 80$ GeV, $|\eta| < 2.4$, passing loose jet ID and the pruned-jet mass must be greater than 40 GeV to remove jets from backgrounds.

Futhermore, in order to veto leptons that are mis-identified as jets, leptons overlap with jets are removed by the ΔR cut, i.e. if there's a lepton passing all lepton selections and the spatial distance to a CA8jet smaller than 0.1 ($\Delta R_{jet,lepton} < 0.1$), then the jet will be removed.

3.5.3 Z boson Requirement

The Z boson candidate is reconstructed by adding four-momentum of the selected lepton pair. The Z boson mass is about 91 GeV, therefore we require the reconstructed invariant mass of the Z boson in the mass region [70 GeV, 110 GeV] where is ± 20 GeV to its theoretical mass.

For the CA8jet from Higgs, we require a minimum p_T threshold of 80 GeV. Since the transverse momentum of Z' is zero and the mass diffence between Z and Higgs is negligible (1 TeV >> 125 GeV $\sim$ 91 GeV), the Z and the Higgs boson are back to back at the transverse plane and their p_T's are identical. Therefore we require the same minimum p_T theshold to the Z boson. Fig. 3-6 shows the transverse momentum distributions from the signal samples.



(a) generator level Higgs p_T distribustion

(b) generator level Z boson p_T distribustion

Figure 3-6: Z and Higgs p_T distribution are almost identical. We pick three samples with different mass points of Z', 1000 GeV (red), 1500 GeV (blue) and 2000 GeV (green). These plots are made from the generator level signal samples without any proper selections.

Finally, all selection requirements are summarized in Table 3.6.

Selection	Value	Comments
Trigger	HLT_Mu22_TkMu8	DoubleMu dataset
	HLT_DoubleEle33	DoublePhoton dataset
Leading muon p_T	$p_T > 40$ GeV	
Sub-leading muon p_T	$p_T > 20$ GeV	
Muon η	$\lvert\eta\rvert < 2.4$	
Muon ID	High p_T tracker based	
Muon isolation I_{trk}^{mod}	< 0.1	
Leading electron p_T	$p_T > 40$ GeV	
Sub-leading electron p_T	$p_T > 40$ GeV	
Electron η	$\lvert\eta\rvert < 2.5$	
	out of [1.4442,1.566]	To avoid ECAL gap.
Electron ID	HEEP modified	
Electron isolation		
I_{trk}^{mod}	< 5 GeV	
$I_{ECAL,HCAL}^{mod}$	< 2 GeV $+ 0.03E_T$	Barrel
	< 2.5 GeV	for $E_T < 50$ GeV candidates in Endcap
	< 2.5 GeV $+ 0.03E_T$	for $E_T > 50$ GeV candidates in Endcap
Jet ID	Loose working point	
Jet p_T	$p_T > 80$ GeV	
Jet η	$\lvert\eta\rvert < 2.4$	
Prunedjet mass	> 40 GeV	
Veto jet-lepton overlap	$\Delta R_{jet,lepton} < 0.1$	Remove the jet satisfies this requirement.
Z p_T	$p_T > 80$ GeV	
Z mass window cut	70 GeV $< m_Z <$ 110 GeV	

Table 3.6: Event and object selection requirements used in the analysis.

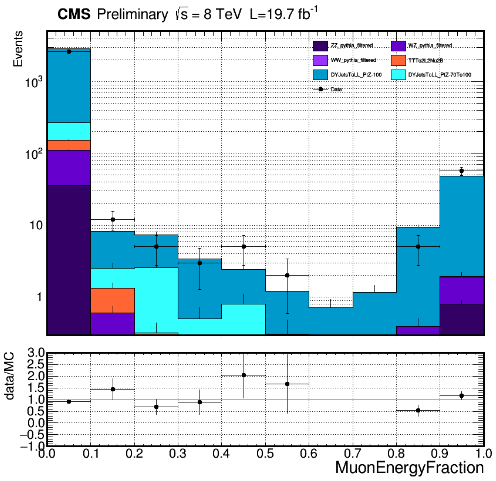
3.6 Data-MC comparison

In this section, a comparison between data and simulation is reported for various kinematic observables. It can be seen that the dominant background contribution comes from the Z+jets production, while sub-leading contributions are from $t\bar{t}$ and dibosons can be negligible.

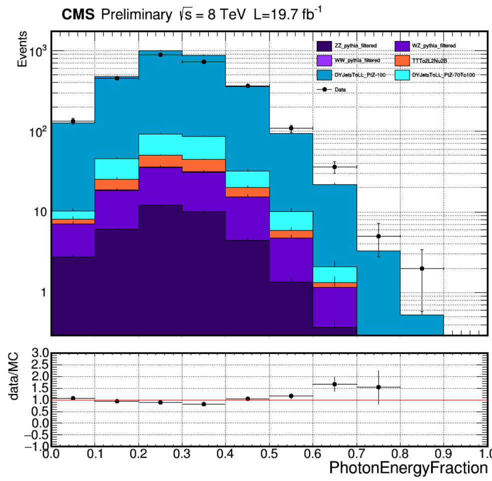
On top of the selections described in previous section, additional regions are defined as following:

- **Signal region (SR)**: Represents the phase space where the signal is expected, defined by the prunedjet mass in 110 GeV $< m_{prunedjet} <$ 140 GeV region. The range is chosen by $\pm$15 GeV to the mass of Higgs.
- **Sidebands (SB)**: Defined by the interval between 70 GeV $< m_{prunedjet} <$ 110 GeV. This region is signal-depleted. In our case, we don't consider prunedjet mass higher than 140 GeV, because of the poor statistics and the excessive contribution of $t\bar{t}$ events.

In the following plots, the data-MC comparison is performed in SB region and all background samples are weighted to the same luminosity as data. The results combined both muon and electron channels. Because of the signal region in data is considered **blind** in this analysis stage, so they are not shown.

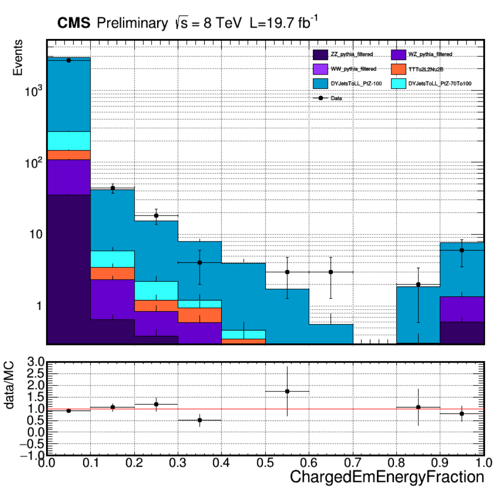


(a) Muon energy fraction

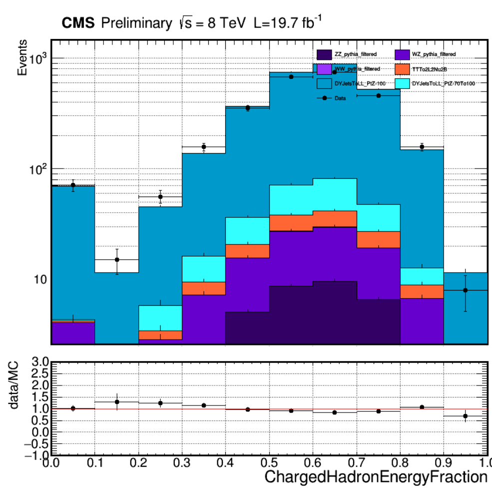


(b) Photon energy fraction

Figure 3-7: Comparison between data and all background samples for two jet variables. The definition of muon/photon energy fraction is muon/photon energy divided by jet energy.

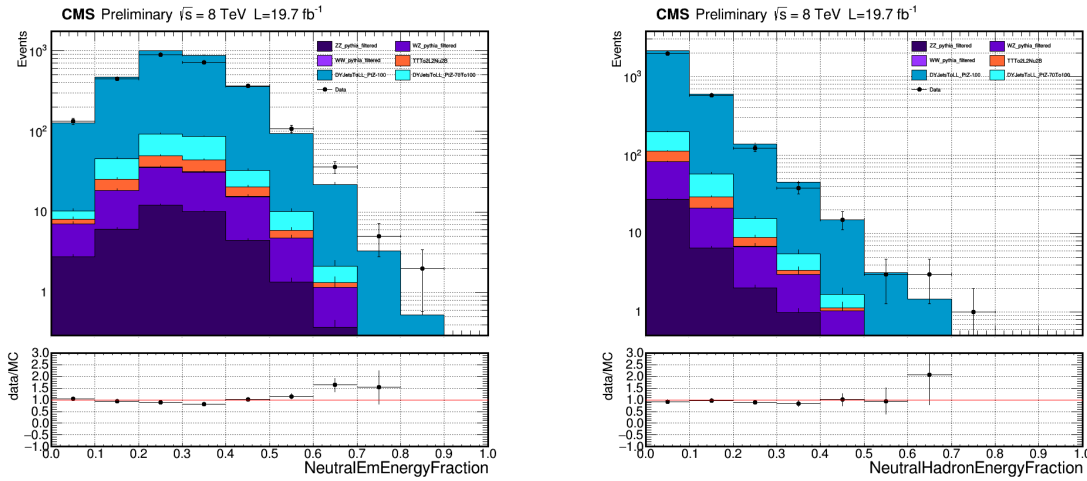


(a) Charged electromagnetic energy fraction



(b) Charged hadron energy fraction

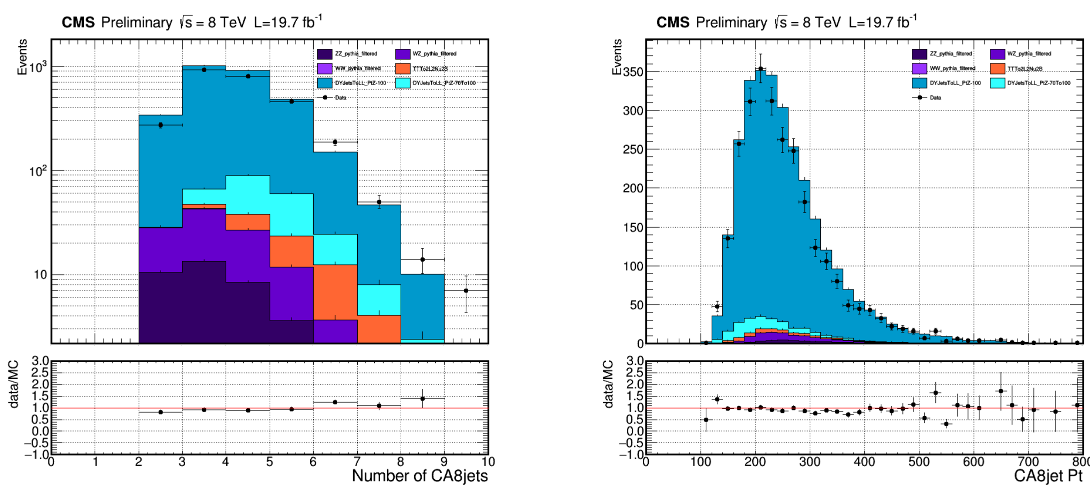
Figure 3-8: Charged electromagnetic/hadron energy fraction is defined by the ratio of the energy of charged particles in ECAL/HCAL to the jet energy.



(a) Neutral electromagnetic energy fraction

(b) Neutral hadron energy fraction

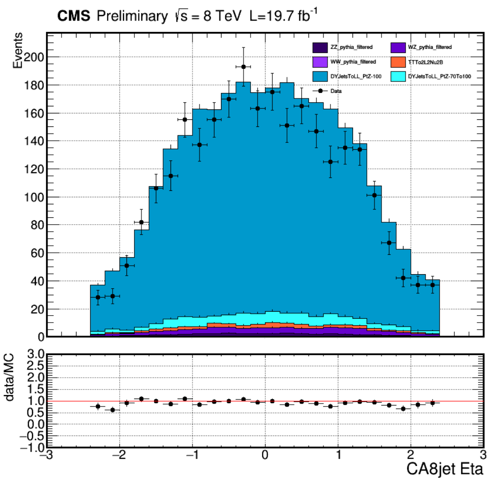
Figure 3-9: Neutral electromagnetic/hadron energy fraction is defined by the ratio of the energy of neutral particles in ECAL/HCAL to the jet energy.



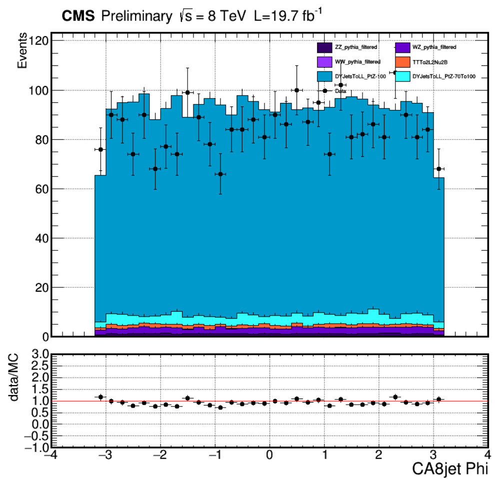
(a) Jet multiplicity

(b) CA8jet p_T

Figure 3-10: Comparison between data and MC in SB region using jet multiplicity (number of jets) and CA8jet transverse momentum.

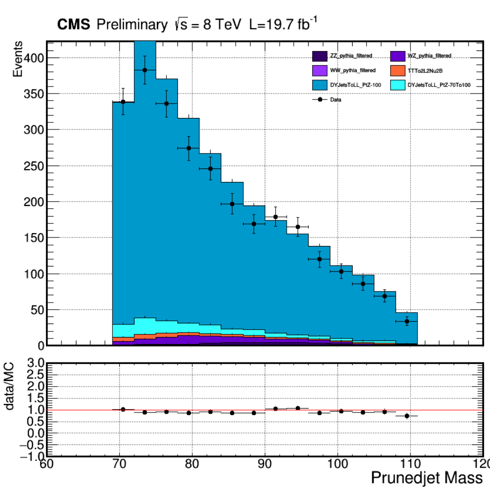


(a) CA8jet η

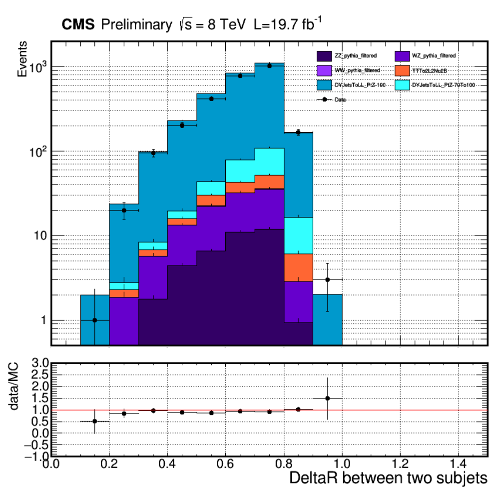


(b) CA8jet ϕ

Figure 3-11: Comparison between data and MC in SB region using CA8jet η and ϕ.

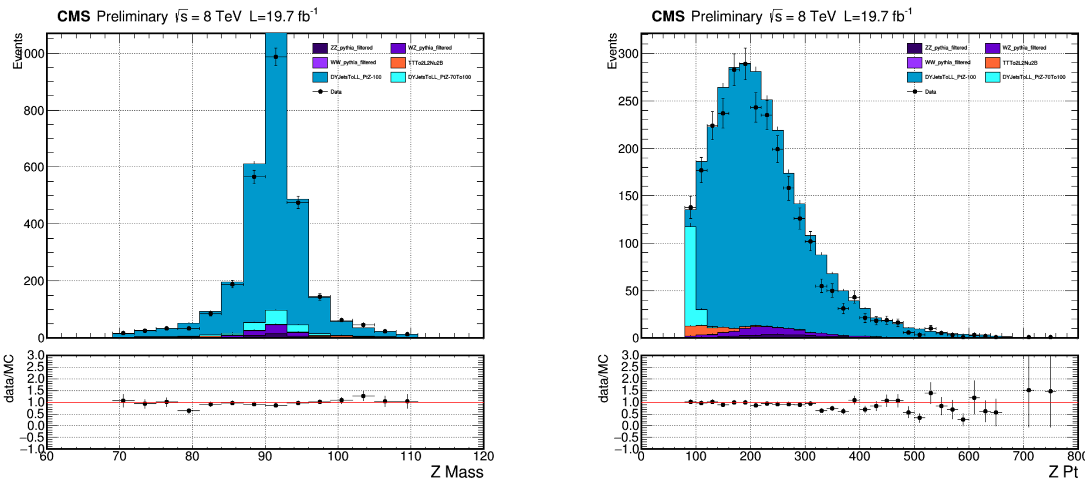


(a) Prunedjet mass



(b) ΔR between two subjets

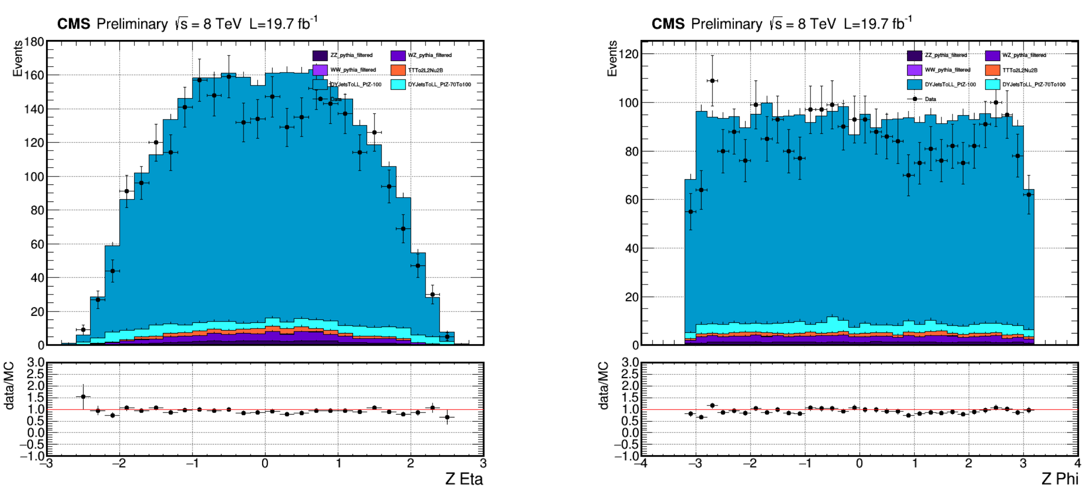
Figure 3-12: Left: the prunedjet mass in the SB region. Right: the spatial distance between two subjets within the CA8jet.



(a) Z mass

(b) Z p_T

Figure 3-13: Comparison between data and MC in SB region using mass and transverse momentum of reconstructed Z boson.



(a) Z η

(b) Z ϕ

Figure 3-14: Comparison between data and MC in SB region using η and ϕ of reconstructed Z boson.

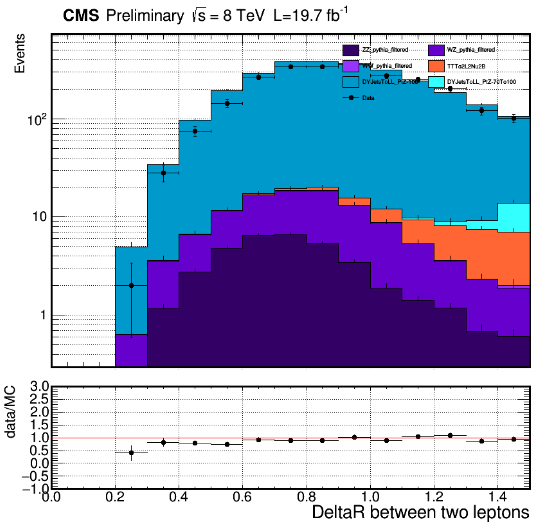


Figure 3-15: ΔR between the two selected leptons.

3.7 Background Estimation

The final aim of this analysis is to test the Z' hypothesis with the observed data, it is important to elaborate a trustworthy strategy for the SM background estimation. Despite the good description of the event kinematics provided by the MC simulation, it is more advisable to minimize the dependence on the MC and develop a data driven strategy.

3.7.1 α ratio method

α ratio is a data driven method in order to estimate the final background in data signal region. We consider the m_{Zh} MC mass spectrum in the SR and SB, then a ratio $\alpha(m_{Zh})$ of the two is created. This factor α allows prediction of the mass spectrum in the signal region starting from the observed distribution in the sidebands. Under the assumption that this estrapolation from the sidebands to the signal region works in the same way both for data and MC, we can estimate the final background distribution in data signal region by multipling the m_{Zh} mass spectrum observed in the sidebands by this α ratio:

$$N_{SR}^{data}(m_{Zh}) = N_{SB}^{data}(m_{Zh}) \times \frac{N_{SR}^{MC}(m_{Zh})}{N_{SB}^{MC}(m_{Zh})} \equiv N_{SB}^{data}(m_{Zh}) \times \alpha(m_{Zh}) \tag{3.4}$$

We divided the spectrum in 14 non-uniform width bins, as shown in Table 3.7, accordingly to the decreasing statistics in the high mass tail.

Bin	**GeV**
1	[680, 720]
2	[720, 760]
3	[760, 800]
4	[800, 840]
5	[840, 920]
6	[920, 1000]
7	[1000, 1100]
8	[1100, 1250]
9	[1250, 1400]
10	[1400, 1600]
11	[1600, 1800]
12	[1800, 2000]
13	[2000, 2200]
14	[2200, 2400]

Table 3.7: Binning of the Zh invariant mass range.

Finally, we multiplied the α ratio to the sidebands data m_{Zh} distribution and obtained the prediction number of backgrounds in data signal region.

3.8 Signal Yields

Since all selections have been settled, we can look at the data in the signal region now. In this section, signal efficiency and distributions of variables in SR will be reported.

3.8.1 signal efficiency

The signal efficiency is defined by the fraction of the number of events passing final selection and the number of generated events in the signal MC samples (Table. 3.1).

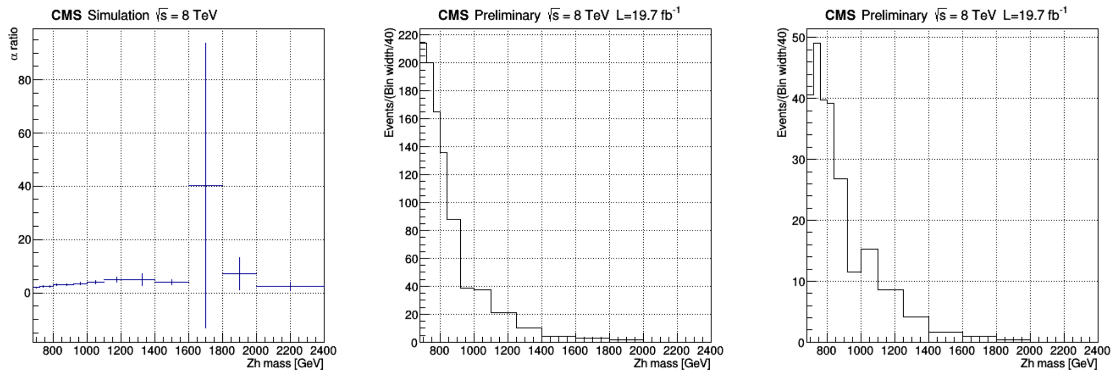


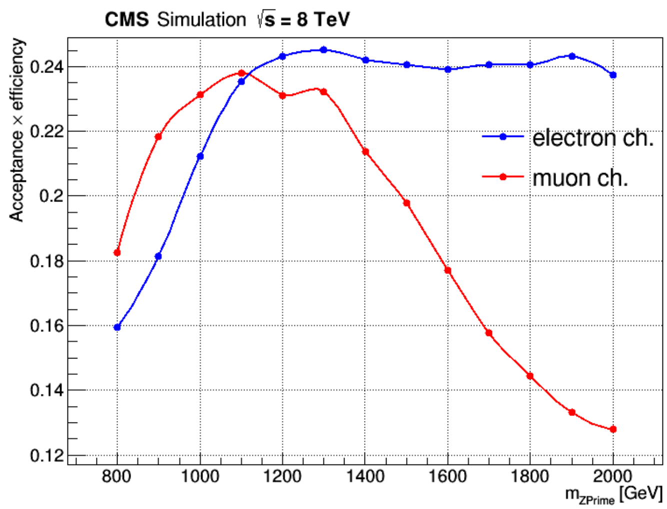
Figure 3-16: Left figure: the α ratio from MC simulation. Central figure: the m_{Zh} distribution observed from data SB. Right figure: the predicted m_{Zh} distribution in data signal region. The m_{Zh} histograms are normalized to the first bin width.

$$\epsilon_{SIG} \equiv \frac{\text{Number of events passing the final selections}}{\text{Number of generated events}} \tag{3.5}$$

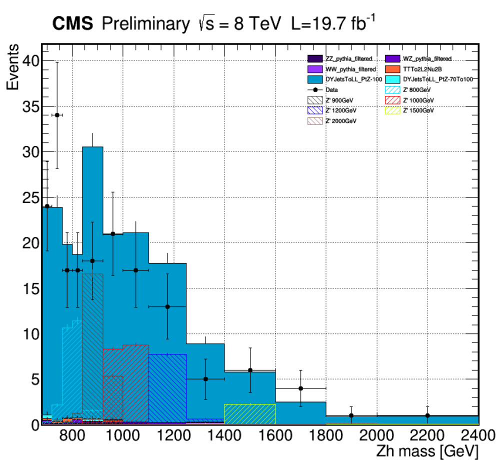
The result of efficiencies are shown in Fig. 3-17 (a). The figure shows a drop of efficiency in muon channel after signal mass of 1100 GeV, which comes from the muon isolation selection. It suggests the defect of the isolation definition. The isolation definition and its efficiency will be improved in the 13 TeV analysis.

3.8.2 The m_{Zh} spectrum

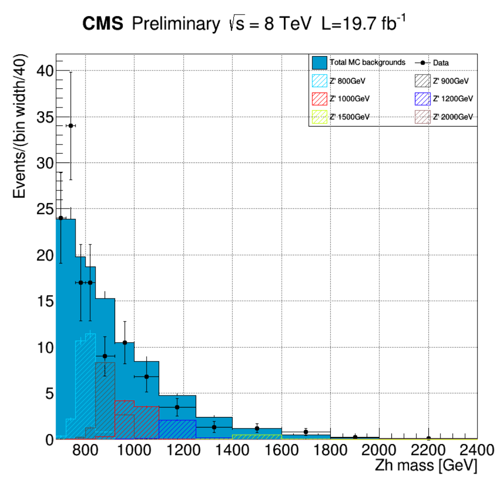
The Z' is reconstructed by combining the four-momentum of selected Z boson and leading jet. A comparison between data and MC of the reconstructed Zh invariant mass distribution in the signal region is shown in Fig. 3-17 (b)(c). Inspecting the SR mass distributions, one can see that data do not present any significant excess from the MC expectation, so we decide to put limits on the production cross section times the branching ratio for the $Z' \to Zh$ process. This Zh invariant mass spectrum will be used for computing the limit as shape information. Furthermore, in the last chapter, the number of events in the signal yields will be used for cut and count method to set the limits.



(a) Signal effciencies for each Z' samples



(b) m_{Zh} spectrum



(c) m_{Zh} spectrum (normalized bin width)

Figure 3-17: (a) The signal efficiencies plot of each Z' mass point. (b) Invariant mass distribution of the MC background simulation compared to the observed data and MC signal shape in the SR. (c) Normalized bin width m_{Zh} distribution (bin width/bin width of the first bin).

3.8.3 The CSV distribution

Despite the reconstructed Zh mass as shape information, we introduced another powerful variable to discriminate signal from background. The overall branching ratio for a SM Higgs decaying into $b\bar{b}$ is about 56% and for Higgs decaying into two quarks only, the BR($h \rightarrow b\bar{b}$) is about 95% [52]. Therefore, the b-tagging CSV working point is a power tool for searching this channel.

Although the recommend CSV selection (loose working point) from BTV group [51] for the boosted Higgs decay result in very low effiency for both background and signal MC samples. Instead of taking selection on the CSV, we use the overall distribution as another source of shape information to set the limits. The method how we retrive the CSV distribution is shown in Table. 3.8.

Category	BTV recommend (CSVL)	Modified
If $\Delta R_{subjets} > 0.3$	subjet CSV > 0.244	use the subjets CSV shape directly
If $\Delta R_{sibjets} < 0.3$	CA8jet CSV > 0.244	use the CA8jet CSV shape directly

Table 3.8: The recommend selection from BTV group for boosted Higgs decay. If the ΔR between the two subjets within CA8jet larger than 0.3, applying CSVL selection on both subjets. If $\Delta R_{subjets} < 0.3$, appliny CSVL on the CA8jet. In this analysis, we use the overall CSV distributions instead of selecting events by this variable (Modified). Note that, only leading CA8jet (the Higgs candidate) and subjets within it are considered in this strategy.

The comparison of the CSV distributions between simulation background, signal and data is shown in Fig. 3-18. Note that, the area of each distribution is set to one in order to compare the shape difference. Inspecting Fig. 3-18, signal shapes tend to distribute on the right side while the backgrounds tend to be on the left, which shows the discriminating power of CSV variable. We only report and use the distribution from the $\Delta R > 0.3$ category to set the final limits because of lack of statics in the CA8jet CSV case. The higher CSV score means the subjet acts more like a b-jet.

Finally, we combine the CSV and the Zh mass distribution, making 2D histograms. Results of MC backgrounds, signal and data are shown in next pages.

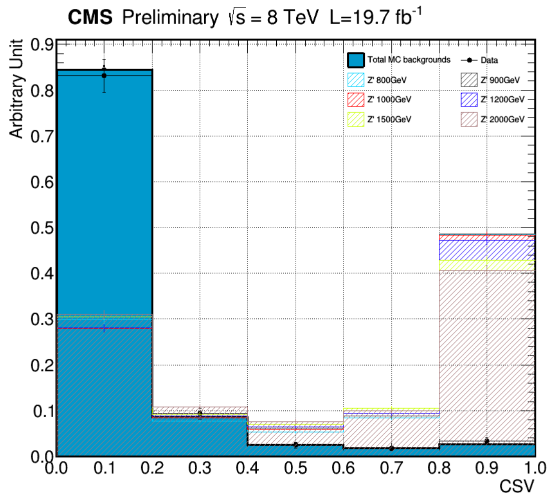
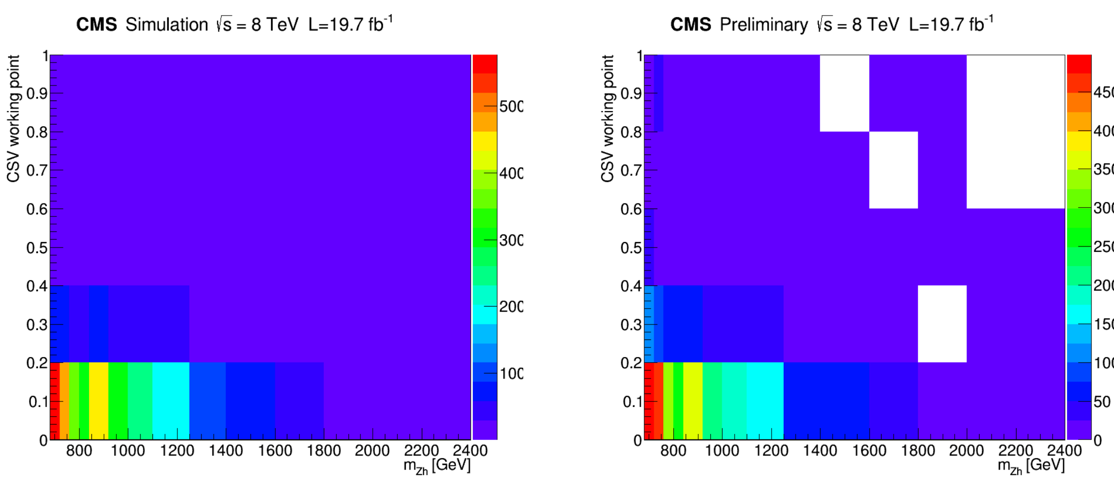


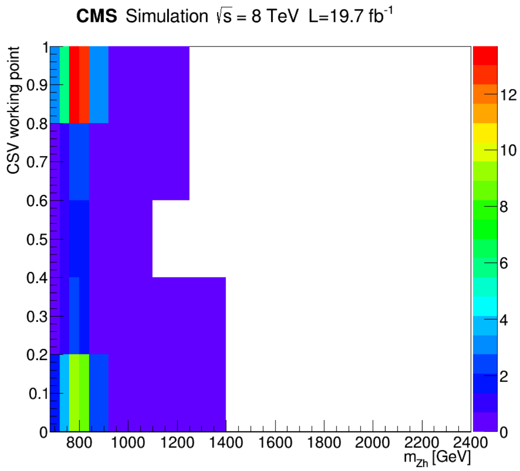
Figure 3-18: The CSV distribution comparison between data and MC samples in SR.



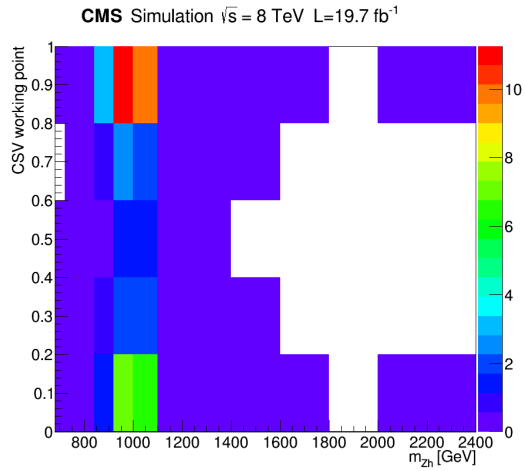
(a) CSV-m_{Zh} distribution of total MC background

(b) CSV-m_{Zh} distribution of total data

Figure 3-19: The combined 2D shape result of data and MC background in SR.

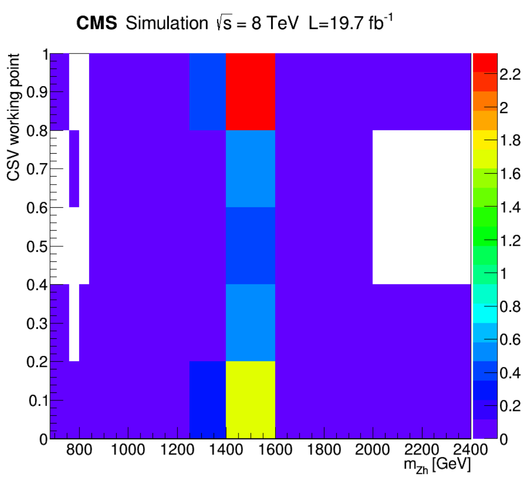


(a) CSV-m_{Zh} distribution of 800 GeV Z' mass signal sample

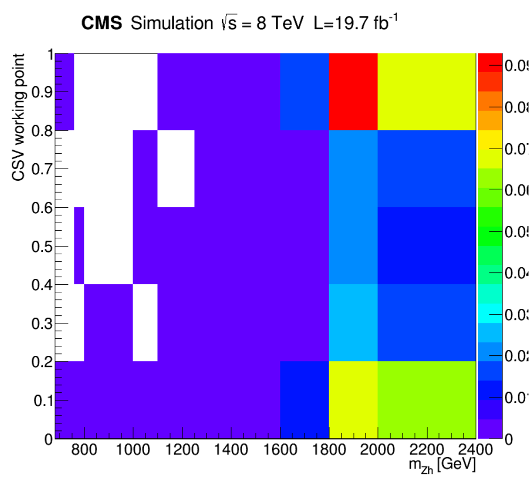


(b) CSV-m_{Zh} distribution of 1000 GeV Z' mass signal sample

Figure 3-20: The combined 2D shape result in SR of 800 GeV and 1000 GeV signal MC samples.



(a) CSV-m_{Zh} distribution of 1500 GeV Z' mass signal sample



(b) CSV-m_{Zh} distribution of 2000 GeV Z' mass signal sample

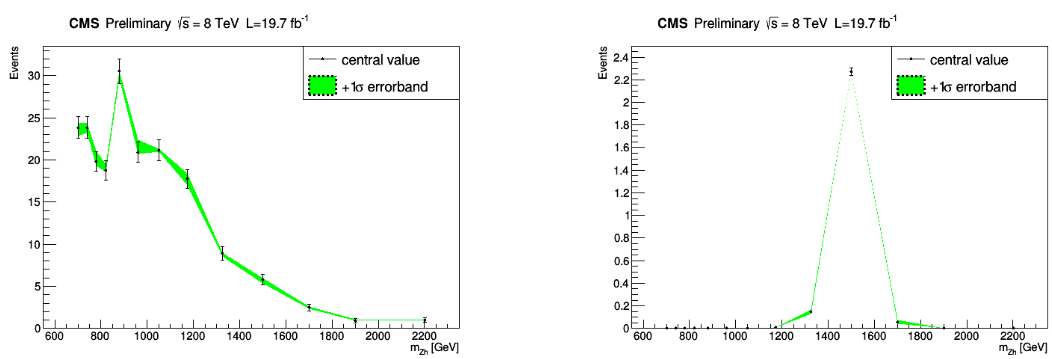
Figure 3-21: The combined 2D shape result in SR of 1500 GeV and 2000 GeV signal MC samples.

3.9 Systematic uncertainties

The background and signal predictions are affected by systematic uncertainties that have to be estimated and taken into account for limit setting. This section includes a list of the relevant systematic uncertainties for this analysis and how they are estimated.

- **Luminosity:** The overall uncertainty of the LHC luminosity delivered to CMS in the 2012 Run-I is measured to be 2.6% [53]

- **Jet energy scale:** The systematic on the jet energy scale was studied by scaling up and down the jet mass and p_T according to the uncertainty associated to the jet energy corrections used. The Zh mass distributions reconstructed by jet four-momentum is affected and therefore we consider the shape uncertainty as well, shown in Fig. 3-22. The overall uncertainty is about 8%.



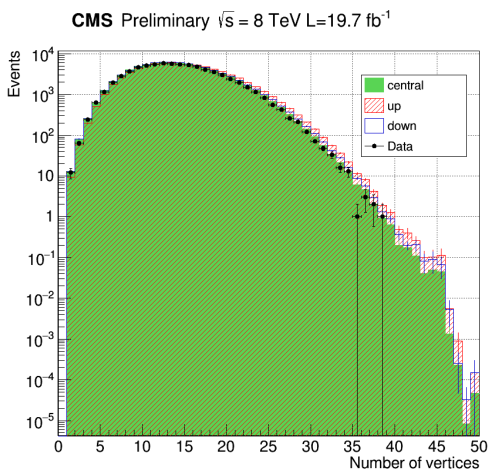
(a) The systematic uncertainty of jet energy scale on background MC m_{Zh} spectrum.

(b) The systematic uncertainty of jet energy scale on signal MC m_{Zh} spectrum.

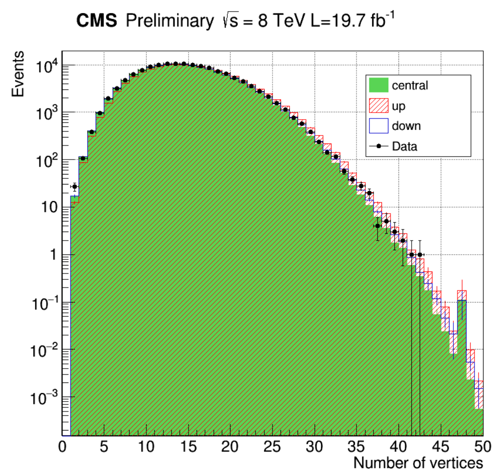
Figure 3-22: SR m_{Zh} distributions for both signal (1500 GeV) and background MC samples. The uncertainty of jet energy scale is shown as the green error band ($\pm 1\sigma$), while the error bar presents the statistic error.

- **CSV distribution normalization:** The uncertainty comes when normalizing the CSV distributions of MC background prediction in order to match the data distribution, which is estimated about 10%.

- **Pile-up reweighting:** As described previously in section 3.4, we reweight the pile-up interactions in MC predictions for better modeling. To calculate the uncertainties on the pile-up simulation, we produce two pile-up distributions where the minimum bias cross section is shifted by $\pm 5\%$ [54]. The impact on the event yields is about 2%. Despite the effects are small, we still consider it as the shape uncertainty. Fig. 3-23 shows the variation distributions of number of vertices for $\pm 1\sigma$.



(a) Number of vertices in electron channel.



(b) Number of vertices in muon channel.

Figure 3-23: (a) shows the distributions of number of vertices in electron channel, comparing central value of the total background prediction and the $\pm 1\sigma$ variation, and data as well. (b) shows the result in muon channel.

- **Lepton ID scale factor:** The muon/electron ID scale factor depends on the kinematics of muon/electron (Table. 3.9). This systematic was studied by applying $\pm 1\sigma$ to the scale factor, the estimated uncertainty is 0.08% for signal yields, 0.1% for the Z+jets background, 0.05% for $t\bar{t}$, and 0.06% for the SM diboson backgrounds. No shape uncertainties are considered.

- **b-jet Ratio:** Since the Drell-Yan process also generates b-jets into our background, the uncertainty of production cross section of Z+b-jets events indeed affects the number of background yield. The cross section ratio defined as σ_{Z+bjet} divided by σ_{Z+jet} and its uncertainties have been studied in [55]. By scaling 1σ on the ratio, we etimated there's 0.2% uncertainty for our Z+jets background yield.

- **SHERPA:** The major background sample in this analysis is Z+jets, which is generated by MADGRAPH and showered by PYTHIA. In order to study the dependence on the MC generator, we use the Z+jets sample generated by SHERPA [56] to estimate the background yield again. The relative uncertainty between SHERPA and MADGRAPH×PYTHIA from the estimated background yield is taken as the systematic by 12%.

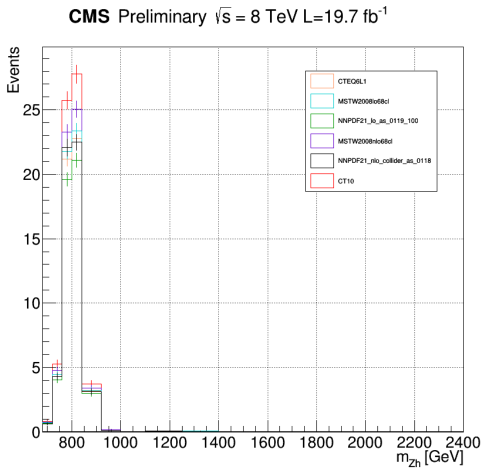
- **Diboson cross section:** Uncertainties of the SM WW/WZ/ZZ production cross section affect their estimated background yield about 5.4%/6.7%/5.5%.

- **PDF uncertainty:** Systematic uncertainties coming from different choice of PDF sets have been considered for this analysis. This study is performed by varing the PDF set when producing the signal samples. The default PDF set we used is CTEQ6L1, replaced by the following PDF set: MSTW2008lo, MSTW2008nlo, NNPDF21_lo, NNPDF21_nlo and CT10. Comparison of distributions from different PDF sets are shown in Fig. 3-24. The estimated overall uncertainty for signal yields is 12% (maximum), shape uncertainty is also taken into account.

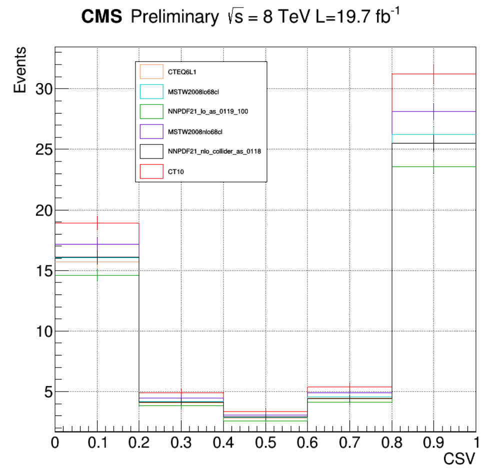
electron p_T [GeV]	$0.0 < \|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.442$	$1.556 < \|\eta\| < 2.0$	$2.0 < \|\eta\| < 2.5$
20-30	1.005 ± 0.003	0.981 ± 0.003	0.980 ± 0.005	1.017 ± 0.006
30-40	1.004 ± 0.001	0.991 ± 0.001	0.992 ± 0.002	1.019 ± 0.003
40-50	1.008 ± 0.001	0.994 ± 0.001	1.004 ± 0.002	1.005 ± 0.001
50-200	1.008 ± 0.001	0.999 ± 0.001	1.006 ± 0.003	1.009 ± 0.002

muon p_T [GeV]	$0.0 < \|\eta\| < 0.8$	$0.8 < \|\eta\| < 2.1$	$2.1 < \|\eta\| < 2.4$
20-40	1.0043 ± 0.0004	1.0074 ± 0.0005	1.022 ± 0.001
40-100	1.0012 ± 0.0004	1.0043 ± 0.0004	1.014 ± 0.001

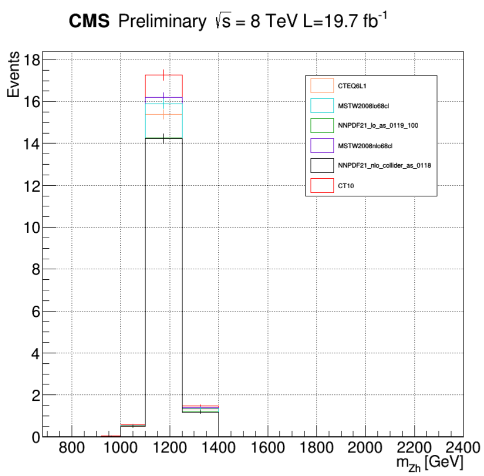
Table 3.9: Data to simulation scale factors for muon and electron identification requirements in various p_T and η ranges.



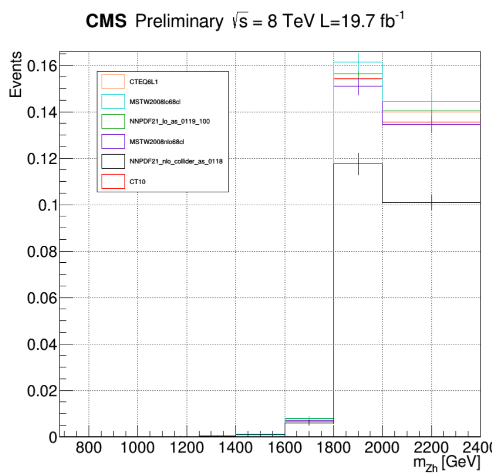
(a) m_{Zh} spectrum of 800 GeV Z' sample



(b) CSV shape of 800 GeV Z' sample



(c) m_{Zh} spectrum of 1200 GeV Z' sample



(d) m_{Zh} spectrum of 2000 GeV Z' sample

Figure 3-24: The comparison between different PDF sets using m_{Zh} spectrum and CSV variable.

Chapter 4

Results and conclusion

In this chapter, the signal, background and data yields after applying full selections are used to determine the upper limits of the production cross section of Z'. The 95% confidence-level limit on the signal contribution in the data is computed by the CL_s method [57, 58] with the RooStats [59] package. In order to extract the limit on the production cross section times the branching ratios, the CMS standard combination tool [60] has been used. Moreover, the Asymptotic method is used to calculate preliminary 95% C.L. upper limits with 1σ and 2σ bands using the CL_s frequentist calculation currently recommended by the LHC Higgs Combination Group.

4.1 Exclusion limit results

4.1.1 Counting result

The expected exclusion limits with the ± 1 and ± 2 σ band, obtained by the provided event yield information, are reported in Fig. 4-1 (a) in terms of upper limits on the signal cross section.

4.1.2 1D shape result

In addition to the event yield information, the m_{Zh} distributions of predicted signal, background and data are provided. All the statistical and systematics uncertainties (with shape uncertainties) are included. The result is shown in Fig. 4-1 (b).

4.1.3 2D shape result

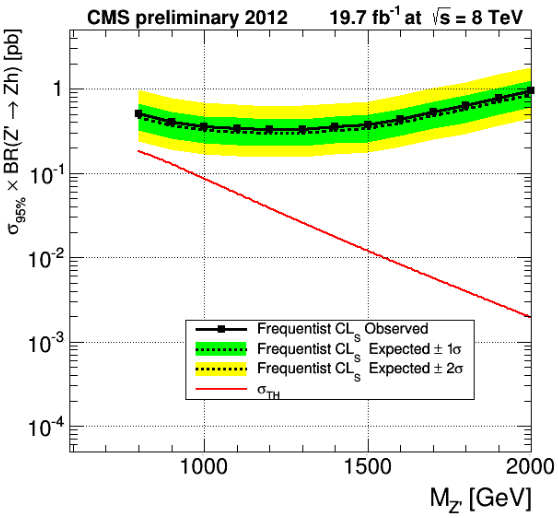
Both the CSV shape and the m_{Zh} spectrum are considered in this strategy. Therefore the systematics related to the CSV shape are also introduced. The result is shown in Fig. 4-1 (c).

4.2 Conclusion

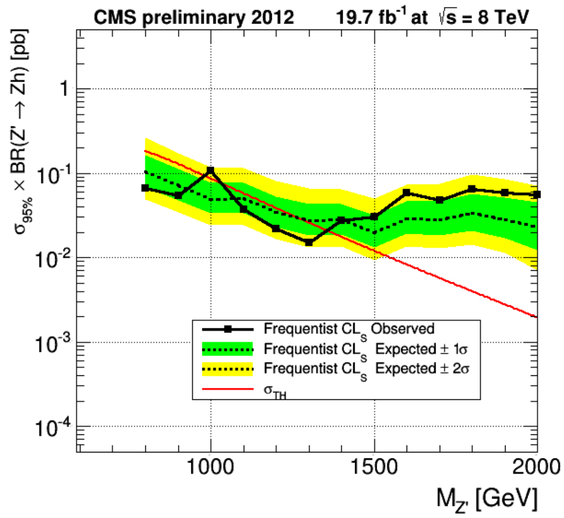
Since no excess above the expected SM background was found, the result is interpreted as an exclusion limit on the production cross section times the branching ratio in the Zh channel as a function of the resonance mass.

Inspecting the limit result, the counting method shows no sensitivity on the cross section limit plot. The expected (observed) lower bound on the Z' mass obtained in the 1D limit result (top figure) is 1257 (1350) GeV, while the 2D result gives 1528 (1477) GeV. The 2D method has improved the limits significantly.

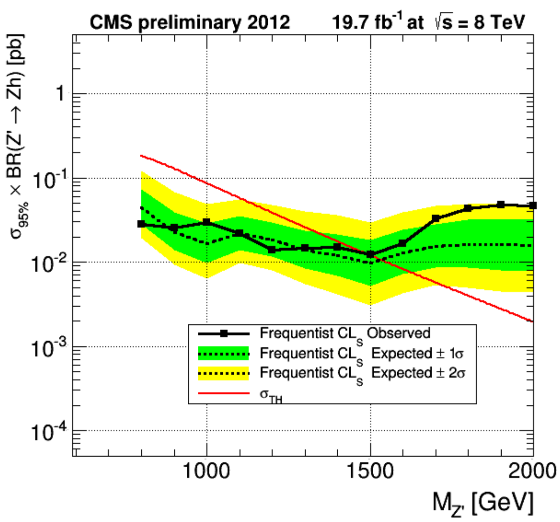
Finally, this analysis puts an upper limit at 95% confidence level on the cross section of $pp \to Z' \to Zh$ at $\sqrt{s} = 8$ TeV. The maximal expected (observed) limit result of 1528 (1477) GeV is determined.



(a) Counting result



(b) 1D shape result



(c) 2D shape result

Figure 4-1: Observed and expected (with $\pm 1(2)\sigma$ band) 95% C.L. upper limit on $\sigma \times \mathrm{BR}(Z' \to Zh)$ including all statistical and systematics uncertainties.

Bibliography

[1] CMS Collaboration. Jet mass in di-jet and vector boson plus jet events. CMS Physics Analysis Summary CMS-PAS-SMP-12-019, CMS, 2012.

[2] CMS Collaboration. Identification of b-quarks jets with the CMS experiment. *JINST*, 8:P04013, 2013.

[3] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phy. Lett. B*, 716:1, 2012.

[4] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phy. Lett. B*, 716:30, 2012.

[5] CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 06:081, 2013.

[6] 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. *Eur. Phys. J. C*, 75:212, 2015.

[7] ATLAS Collaboration. Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ channels in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector. *Phys. Rev. D*, 90:052004, 2014.

[8] ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett. B*, 726:120, 2013.

[9] CMS and ATLAS Collaboration. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[10] Mads T. Frandsen. Minimal Walking Technicolor. arXiv:0710.4333 [hep-ph], http://arxiv.org/abs/0710.4333, 2007.

[11] T. Han, H. E. Logan, B. McElrath, and L.-T. Wang. Phenomenology of the little Higgs model. *Phys. Rev. D*, 67:095004, 2003.

[12] M. Schmaltz and D. Tucker-Smith. LITTLE HIGGS THEORIES. *Annual Review of Nuclear and Particle Science*, 55:no. 1, 229–270, 2005.

[13] M. Perelstein. Little Higgs models and their phenomenolog. *Progress in Particle and Neclear Physics*, 58:no. 1, 247–291, 2007.

[14] R. Contino, D. Pappadopulo, D. Marzocca, and R. Rattazzi. On the effect of resonances in composite Higgs phenomenology. *Journal of High Energy Physic*, 2011:no. 10, 1–50, 2011.

[15] D. Marzocca, M. Serone, and J. Shu. General composite Higgs model. *Journal of High Energy Physics*, 2012:no. 8, 1–52, 2012.

[16] B. Bellazzini, C. Cski, and J. Serra. Composite Higgses. *Eur. Phys. J.*, C74:no. 5, 2766, 2014.

[17] D. Pappadopulo, R. Torre A. Thamm, and A. Wulzer. Heavy vector triplets: bridging theory and data. *Journal of High Energy Physics*, 2014:no. 9, 1–50, 2014.

[18] Andrea Wulzer. An Equivalent Gauge and the Equivalence Theorem. arXiv:1309.6055 [hep-ph], http://arxiv.org/abs/1309.6055, 2013.

[19] V. Barger, W.-Y. Keung, and E. Ma. Gauge model with light W and Z bosons. *Phys. Rev. D*, 22:727, 1980.

[20] Lisa Benato, Yu-Hsiang Chang, Ching-Wei Chen, Michele de Gruttola, Raman Khurana Ji-Kong Huang, Stefano Lacaprara, Yun-Ju Lu, Jacopo Pazzini, Maurizio Pierini, Henry Yee-Shian Tong, Jun-Yi Wu, Shin-Shan Eiko Yu, Marco Zanetti, and Alberto Zucchetta. Search for heavy resonances decaying into a vector boson and a Higgs boson in the $(ll, l\nu, \nu\nu)b\bar{b}$ final state. CMS Note 2015/186, CMS, 2015.

[21] A. Benaglia. The CMS ECAL performance with examples. *JINST*, 09:C02008, 2014.

[22] CMS HCAL Collaboration. Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges. CMS Note 2006/138, CMS, 2006.

[23] Victor Daniel Elvira. Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment. CMS Note 2004/020, CMS, 2004.

[24] J. Brooke on behalf of the CMS Collaboration. Performance of the CMS Level-1 Trigger. arXiv:1302.2469 [hep-ex], http://arxiv.org/abs/1302.2469, 2013.

[25] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shaoand T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[26] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[27] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. The POWHEG-hvq manual version 1.0. arXiv:0707.3081 [hep-ph], http://arxiv.org/abs/0707.3081, 2007.

[28] S. Chatrchyan et al. Performance of CMS muon reconstruction in pp collision events at $\sqrt{7}$ TeV. *JINST*, 7:P10002, 2012.

[29] Wolfgang Adam, Boris Mangano, Thomas Speer, and Teddy Todorow. Track Reconstruction in the CMS tracker. CMS Note 2006/041, CMS, 2006.

[30] CMS Collaboration. Baseline muon selections for Run-I. https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId.

[31] CMS Collaboration. Physics TDR volume I, The CMS Physics Technical Design R I. CERN-LHCC-2006-001, page 521pp, 2006.

[32] W. Adam, R. Frhwirth, A. Strandlie, and T. Todorov. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005.

[33] CMS Collaboration. E/gamma Commissioning and physics deliverables. https://twiki.cern.ch/twiki/bin/view/CMS/EgCommissioningAndPhysicsDeliverables.

[34] CMS Collaboration. HEEP Electron ID and isolation. https://twiki.cern.ch/twiki/bin/viewauth/CMS/HEEPElectronID.

[35] T. S. Williams, D. Newbold, and C. Shepherd-Themistocleous. Modified isolation cuts for the selection of high p_T electrons from the decay of boosted Z bosons. CMS Note 2012/168, CMS, 2012.

[36] Wikipedia. Color confinement. https://en.wikipedia.org/wiki/Color_confinement.

[37] M. Cacciari, G. P. Salam, and G. Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[38] M. Wobisch and T. Wengler. Hadronization corrections to jet cross section in deep-inelastic scattering. arXiv:9907280 [hep-ph], http://arxiv.org/abs/9907280, 1999.

[39] CMS Collaboration. Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV. CMS Physics Analysis Summary CMS-PAS-PFT-10-002, CMS, 2010.

[40] Y.L. Dokshitzer et al. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[41] CMS Collaboration. A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging. CMS Physics Analysis Summary CMS-PAS-JME-09-001, CMS, 2009.

[42] N.Saoulidou and E.Tziaferi. Performance of the Particle-Flow jet identification criteria using proton-proton collisions at = 8 TeV. CMS Note 2014/227, CMS, 2014.

[43] CMS Collaboration. Jet Identification. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID.

[44] P. Tran et al. Measurement of jet mass in V+Jets events. CMS Note 2012/137, CMS, 2012.

[45] A. Bonato, R. Gerosa, and et al. Identifiying hadronically decaying W bosons merged into a single jet. CMS Note 2013/072, CMS, 2013.

[46] J. Thaler, D. Krohn, and L.T. Wang. Jet Trimming. *JHEP*, 100:242001, 2010.

[47] J.M. Butterworth et al. Jet substructure as a new Higgs search channel at the LHC. *PhysRevLett*, 100:242001, 2008.

[48] S. Ellis, C. Vermilion, and J. Walsh. Techniques for improved heavy particle searches with jet substructure. *Phys. Rev.*, D80:051501, 2009.

[49] S. Ellis, C. Vermilion, and J. Walsh. Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches. *Phys. Rev.*, D81:094023, 2010.

[50] S. Chatrchyan et al. Studies of jet mass in di-jet and W/Z + jet events. *JHEP*, 1305:090, 2013.

[51] CMS Collaboration. Performance of b tagging at $\sqrt{s}$ =8 TeV in multihjet, $t\bar{t}$ and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-13-001, CMS, 2013.

[52] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard Model Higgs-Boson Branching Ratios with Uncertainties. arXiv:1107.5909 [hep-ph], http://arxiv.org/abs/1107.5909, 2011.

[53] CMS Collaboration. CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. CMS Physics Analysis Summary CMS-PAS-LUM-13-001, CMS, 2013.

[54] CMS Collaboration. Estimating Systematic Errors Due to Pileup Modeling. https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors.

[55] CMS Collaboration. Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at $\sqrt{s}$ = 7 TeV. *JHEP*, 06:120, 2014.

[56] T. Gleisberg, S. Hoeche, F. Krauss, M. Schoenherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. arXiv:0801.4622 [hep-ph], http://arxiv.org/abs/0801.4622, 2008.

[57] A. L. Read. Presentation of search results: the CL_s technique. *J.Phys.*, G28:2693, 2002.

[58] T. Junk. Confidence level computation for combining searches with small statistics. *Nucl. Insrum. Meth.*, A434:435, 1999.

[59] L. Moneta. The RooStats Project. arXiv:1009.1003 [hep-ph], http://arxiv.org/abs/1009.1003, 2010.

[60] CMS Collaboration. CMS Higgs Combine Tool. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit.