

# Contents

<b>1</b>	<b>Introduction and Theory Overview</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.2	Theory Overview . . . . .	2
1.2.1	Heavy Vector Triplet Model . . . . .	2
1.2.2	Basic Phenomenology . . . . .	3
1.2.3	Explicit Models . . . . .	8
<b>2</b>	<b>The LHC and the CMS Detector</b>	<b>11</b>
2.1	Large Hadron Collider . . . . .	11
2.2	Compact Muon Solenoid . . . . .	13
2.2.1	Tracker . . . . .	15
2.2.2	ECAL . . . . .	16
2.2.3	HCAL . . . . .	17
2.2.4	Muon Chamber . . . . .	19
2.2.5	Trigger System . . . . .	20
<b>3</b>	<b>Analysis Procedures</b>	<b>22</b>
3.1	Monte Carlo Samples and Data sets . . . . .	22
3.1.1	Signal MC . . . . .	22
3.1.2	Background MC . . . . .	24
3.1.3	Data Samples . . . . .	24
3.2	Trigger . . . . .	25
3.3	Physics Objects . . . . .	26

3.3.1	Muon	26
3.3.2	Electron	29
3.3.3	Jet	33
3.3.4	Jet Grooming Algorithms	35
3.3.5	B-tagging	37
3.4	Pile-up reweighting	40
3.5	Event and Object selection	41
3.5.1	Lepton Requirements	41
3.5.2	Jet Requirement	41
3.5.3	Z boson Requirement	42
3.6	Data-MC comparison	44
3.7	Background Estimation	49
3.7.1	$\alpha$ ratio method	49
3.8	Signal Yields	50
3.8.1	signal efficiency	50
3.8.2	The $m_{Zh}$ spectrum	51
3.8.3	The CSV distribution	53
3.9	Systematic uncertainties	56

# List of Figures

1-1	Branching ratios as a function of the resonance mass for the HVT benchmark model A(left) and model B(right). . . . .	10
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. . . . .	10
2-1	Overview of the LHC and relative location of the detectors. . . . .	12
2-2	CERN accelerator complex. . . . .	13
2-3	Structure overview of the CMS detector. . . . .	14
2-4	Schematic layout of tracker. . . . .	15
2-5	The pixel detector inside tracker. . . . .	15
2-6	Schematic layout of the CMS ECAL. . . . .	17
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) . . . . .	18
2-8	Slice view of one quarter of muon chamber system. . . . .	20
2-9	CMS triggering and data acquisition architecture. . . . .	21
3-1	Feynman diagram for $Z' \rightarrow Zh \rightarrow 2l2q$ . . . . .	23
3-2	Example of the $E_{2\times 5}/E_{5\times 5}$ 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). . . . .	32

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. . . . .	36
3-4	[2] Distributions of (a) the secondary vertex multiplicity and (b) the CSV discriminator. . . . .	39
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. . . . .	40
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. . . . .	42
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. . . . .	45
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. .	45
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. .	46
3-10	Comparison between data and MC in SB region using jet multiplicity (number of jets) and CA8jet transverse momentum. . . . .	46
3-11	Comparison between data and MC in SB region using CA8jet $\eta$ and $\phi$ . .	47
3-12	Left: the prunedjet mass in the SB region. Right: the spatial distance between two subjets within the CA8jet. . . . .	47
3-13	Comparison between data and MC in SB region using mass and transverse momentum of reconstructed Z boson. . . . .	48
3-14	Comparison between data and MC in SB region using $\eta$ and $\phi$ of reconstructed Z boson. . . . .	48

3-15 $\Delta R$ between the two selected leptons. . . . .	49
3-16 Left figure: the $\alpha$ 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. . . . .	51
3-17 (a) The signal efficiencies plot of each $Z'$ samples. (b) Invariant mass distribution of the MC background simulation compared to the observed data and MC signal shape in the SR. . . . .	52
3-18 The CSV distribution comparison between data and MC samples in SR. . . . .	54
3-19 The combined 2D shape result of data and MC background in SR. . . . .	54
3-20 The combined 2D shape result in SR of 800 GeV and 1000 GeV signal MC samples. . . . .	55
3-21 The combined 2D shape result in SR of 1500 GeV and 2000 GeV signal MC samples. . . . .	55
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. . . . .	57
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. . . . .	57

# List of Tables

3.1	Signal samples used in the analysis. . . . .	23
3.2	Background samples used in the analysis. . . . .	24
3.3	Data sets used in this analysis. . . . .	25
3.4	Summary of the muon ID selection criteria. . . . .	28
3.5	Summary of the modified HEEPv4.1 electron ID. . . . .	32
3.6	Event and object selection requirements used in the analysis. . . . .	43
3.7	Binning of the $Zh$ invariant mass range. . . . .	50
3.8	The recommend selection from BTV group for boosted Higgs decay. If the $\Delta R$ between the two subjets within CA8jet larger than 0.3, applying CSVL selection on both subjets. If $\Delta R_{subjets} < 0.3$ , applying 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. . . . .	53

# <sup>1</sup> Chapter 1

## <sup>2</sup> Introduction and Theory Overview

### <sup>3</sup> 1.1 Introduction

<sup>4</sup> This thesis presents the analysis details and the results of the search for heavy reso-  
<sup>5</sup> nances decaying into a  $Z$  boson and a Higgs boson ( $h$ ) at the center-of-mass energy of  
<sup>6</sup> 8 TeV, using  $19.7 \text{ fb}^{-1}$  p-p collision data. In turn, the  $Z$  boson is identified through  
<sup>7</sup> its leptonic decays (leptons often refer to  $e$  and  $\mu$  only in experiments.  $l = e, \mu$ ). The  
<sup>8</sup> Higgs boson  $h$  is expected to hadronically decay into a pair of b-quarks. The investi-  
<sup>9</sup> gated final states consist of two charged leptons which are identified in the detector  
<sup>10</sup> and limit the presence of the background, and two b-quarks from the hadronic Higgs  
<sup>11</sup> decay which collects the largest possible fraction of Higgs events.

<sup>12</sup> This thesis is organised as follows. In the latter part of this chapter, the model that  
<sup>13</sup> predicts heavy resonances is introduced, including the expected cross section and the  
<sup>14</sup> specification of model parameters. In chapter 2, the LHC and the CMS experiment  
<sup>15</sup> are described, including the information of each sub-detector and the trigger system  
<sup>16</sup> of the CMS. The details of the analysis are shown in chapter 3. This chapter reveals  
<sup>17</sup> the way to reconstruct physical objects in CMS. By adding some proper kinematic  
<sup>18</sup> selections on those physics objects, the interested events in data collected by the CMS  
<sup>19</sup> detector can be selected. Moreover, this chapter shows the comparison between data  
<sup>20</sup> and simulation. In the last chapter, the results of the search and the conclusion are  
<sup>21</sup> presented.

## 22 1.2 Theory Overview

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

### 32 1.2.1 Heavy Vector Triplet Model

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

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

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

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

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

<sup>44</sup>

<sup>45</sup> In these models, new heavy vector bosons ( $V^\pm, V^0$ ) that couple to the SM Higgs  
<sup>46</sup> doublet (Eq. 1.4) and SM gauge bosons with the parameters  $c_H$  and  $g_V$  and to the  
<sup>47</sup> fermions via the combination  $(g^2/g_V)c_F$ . The parameter  $g_V$  represents the strength  
<sup>48</sup> of the new vector boson interaction, while  $c_H$  and  $c_F$  represent the couplings to the  
<sup>49</sup> Higgs and the fermions respectively, and are expected to be of the order of unity in  
<sup>50</sup> most models.

### <sup>51</sup> 1.2.2 Basic Phenomenology

#### <sup>52</sup> 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  $\theta_N$

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

The mass matrix is

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

In the above equations  $\hat{v}$  denotes the Vacuum Expectation 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

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

- <sup>53</sup> Note that  $M_0$  represents the real mass eigenvalue of the neutral heavy vector boson.
- <sup>54</sup> Moreover, since we assume  $\hat{m}_V > \hat{m}_Z$ , the only variable controls the sign of the
- <sup>55</sup> tangent function is  $c_H$ , which is model dependent. Once the sign of  $c_H$  is determined,
- <sup>56</sup> 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.

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

where it is diagonalized by

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

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

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

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

$$m_W^2 M_\pm^2 = \cos^2 \theta_W m_Z^2 M_0^2 . \quad (1.11)$$

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 hierarchy in the mass relation of SM  $Z$  and  $W$  bosons versus the new vectors.

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

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  $m_Z$

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

The parameter  $\xi$  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

$$\frac{m_W^2}{m_Z^2} \simeq \cos^2 \theta_W . \quad (1.14)$$

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  $V$ s are practically degenerate

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

<sup>57</sup> In the following, when working at the leading order in the limit Eq. (1.12), we can  
<sup>58</sup> ignore the mass splitting and denote the mass of the charged and the neutral states  
<sup>59</sup> collectively as  $M_V$ . It is easy to check that in that limit  $M_V = \hat{m}_V$ .

## <sup>60</sup> Decay Widths

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

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

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  $F = \{l, q, 3\}$ .

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

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

$$\Gamma_{V_\pm \rightarrow f\bar{f}} \simeq 2\Gamma_{V_0 \rightarrow f\bar{f}} \simeq N_C[f] \left(\frac{g^2 c_F}{g_V}\right)^2 \frac{M_V}{48\pi}, \quad (1.18)$$

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

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

<sup>61</sup> 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

<sup>64</sup> Lagrangian is instead useful. The channels that are not shown in the above equations  
<sup>65</sup> are either forbidden or suppressed like the decays to transverse polarizations.

<sup>66</sup> From this section, a very simple picture emerges. At small  $\xi$ , all the decay widths  
<sup>67</sup> are fixed with a given resonance mass  $M_V$  and the couplings  $\{g^2 c_F/g_V, g_V c_H\}$  which  
<sup>68</sup> control the BRs in all relevant channels. Parameters  $c_{VVV}$ ,  $c_{VHH}$  and  $c_{VW}$  are  
<sup>69</sup> basically irrelevant. Thus, the basic phenomenology of this model is well described  
<sup>70</sup> by a good approximation.

### <sup>71</sup> 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

$$c_H \simeq -g^2/g_V^2, c_F \simeq 1, \\ g_V^2 c_H^2 \simeq g^4 c_F^2/g_V^2 \simeq g^2/g_V. \quad (1.20)$$

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

$$\Gamma_{V_0 \rightarrow f\bar{f}'} \simeq N_c[f] \frac{g^4 M_V}{96\pi} \\ \Gamma_{V_0 \rightarrow W^+W^-} \simeq \Gamma_{V_0 \rightarrow Zh} \simeq \frac{g^4 M_V}{192\pi}. \quad (1.21)$$

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 suppression ( $g^2/g_V$ ) in Eq. (1.20).

On the contrary, in model B the  $c_H$  term is unsuppressed

$$c_H \simeq c_F \simeq 1 , \\ g_V^2 c_H^2 \simeq g_V^2 , g^2 c_{c_F} / g_{g_V} \simeq g^2 / g_V . \quad (1.22)$$

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

$$\Gamma_{V_0 \rightarrow f\bar{f}'} \simeq N_c[f] \frac{g^4 M_V}{342\pi} \\ \Gamma_{V_0 \rightarrow W^+W^-} \simeq \Gamma_{V_0 \rightarrow Zh} \simeq \frac{3M_V}{64\pi} . \quad (1.23)$$

- <sup>72</sup> For model B <sub>$g_V=3$</sub>  the dominant BRs are into di-bosons and the fermionic decays are
- <sup>73</sup> extremely suppressed. Moreover, the total width increases with increasing  $g_V$  since it
- <sup>74</sup> is dominated by the di-boson width which grows with  $g_V$  as expected from Eq. (1.22).
- <sup>75</sup> This model B is particularly interesting for the present search, since it predicts signal
- <sup>76</sup> cross sections of the order of fb [17] [20] [Fig. 1-2], branching ratios to vector bosons
- <sup>77</sup> close to unity, and thus being accessible at the LHC. In the latter chapters, the mass
- <sup>78</sup> eigenstate of the neutral heavy vector boson in model B scenario refers to the  $Z'$
- <sup>79</sup> particle, which is the search target of this thesis.

<sup>80</sup>

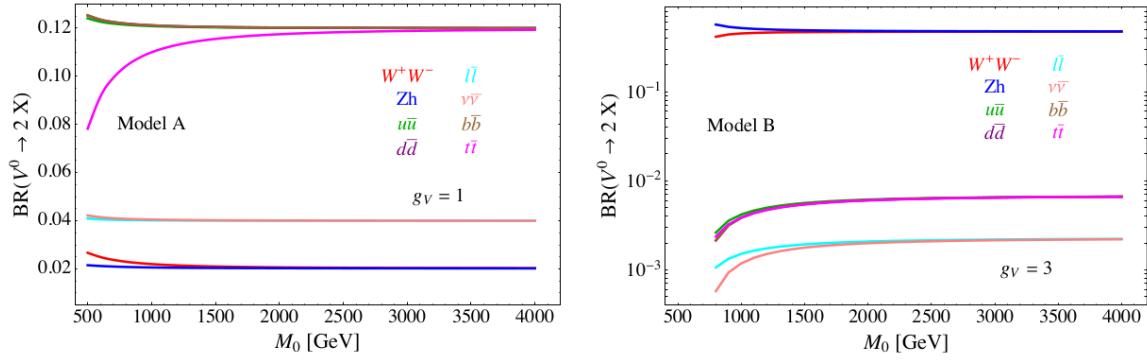


Figure 1-1: Branching ratios as a function of the resonance mass for the HVT benchmark model A(left) and model B(right).

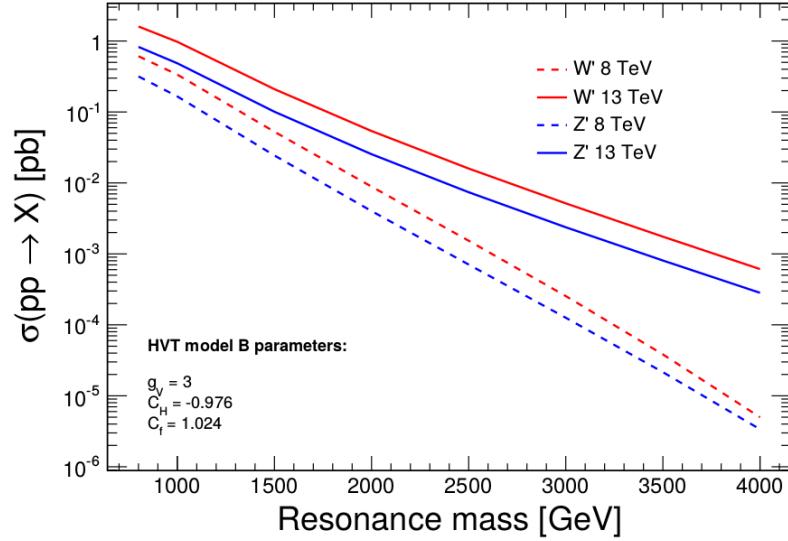


Figure 1-2: Theoretical production cross-section as a function of new resonance particles for HVT model B benchmark. Dash lines are 8 TeV predictions while solid lines are 13 TeV predictions.

81 **Chapter 2**

82 **The LHC and the CMS Detector**

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

86 **2.1 Large Hadron Collider**

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

94 Other accelerators that had been originally built at CERN for previous experi-  
95 ments serve as an injection chain for the LHC now (Fig. 2-2). The proton beam starts  
96 from LINAC, a small linear accelerator, where the energy of protons first reaches at  
97 50 MeV. The proton beam then passes through a booster and goes to the PS, where  
98 it is accelerated up to 25 GeV. After that, it reaches 450 GeV in the SPS. The beam  
99 is finally injected in the LHC ring from the SPS, and it had been accelerated up to  
100 4 TeV in 2012. In early 2015, the proton beam had been accelerated to 6.5 TeV, a

101 value near its designed energy, before undergoing collision.

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



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

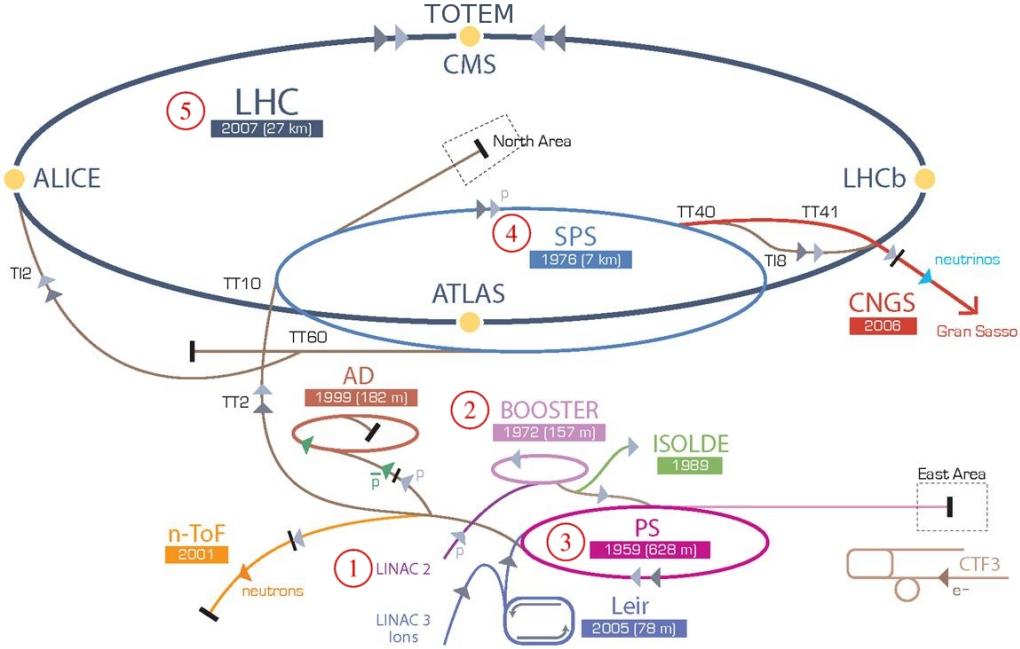


Figure 2-2: CERN accelerator complex.

## 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-detectors. 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.

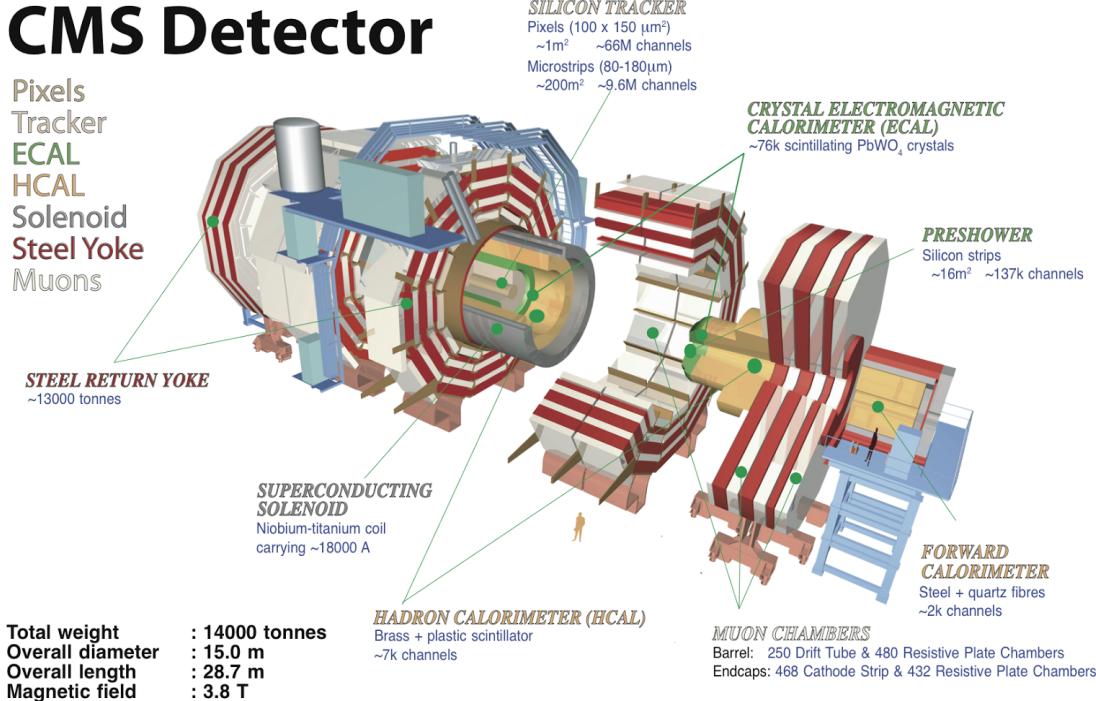


Figure 2-3: Structure overview of the CMS detector.

## <sup>127</sup> 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  $\phi$  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  $\theta$  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  $E_T = E \sin \theta$ .

### <sup>135</sup> 2.2.1 Tracker

<sup>136</sup> Tracker is the most inner part of CMS that records the productions of collisions in the  
<sup>137</sup> first place. It traces the charged particles' trajectories. Physicists can reconstruct the  
<sup>138</sup> vertices of the interaction and the momentum of charged particles by linking tracks  
<sup>139</sup> to the collider's pipe and measuring the curves of particles under magnetic field.

<sup>140</sup> The tracking system is composed of two kinds of detector, the pixel detector and  
<sup>141</sup> 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  $z = \pm 345, \pm 465$  mm.

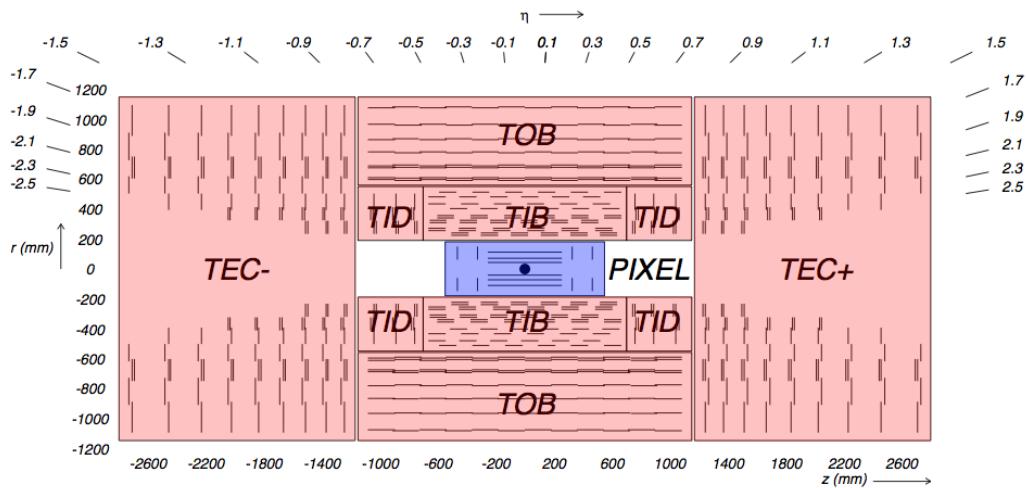


Figure 2-4: Schematic layout of tracker.

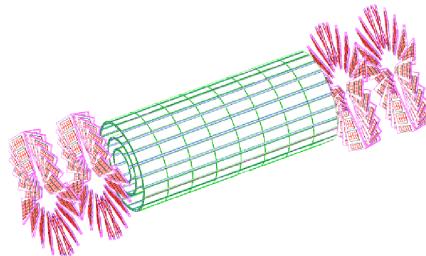


Figure 2-5: The pixel detector inside tracker.

<sup>142</sup>

<sup>143</sup> The pixel detector, though about the size of a shoebox, consists of 1440 segmented  
<sup>144</sup> silicon sensor modules with total 66 million readout channels. Charge carriers are  
<sup>145</sup> distributed over several pixels. The analog pulse height information can be used

146 to calculate the center of certain charge distribution which could improve the hit  
147 information. The spatial resolution is measured to be about  $10 \mu\text{m}$  for the  $r - \phi$   
148 plane or about  $20 \mu\text{m}$  for  $z$  direction measurement.

149 Outside the pixel detector, there comes the silicon strip detector. The barrel  
150 region of silicon strip detector is divided into two parts, the Tracker Inner Barrel  
151 (TIB) and the Tracker Outer Barrel (TOB). The former is composed of four layers  
152 of silicon sensors with a thickness of  $320 \mu\text{m}$  and of strip pitches varying from 80  
153 to  $120 \mu\text{m}$ . The TOB is made of six layers. In this kind of sub-detector, thicker  
154 silicon sensors ( $500 \mu\text{m}$ ) are employed, while the strip pitch varies from 120 to 180  
155  $\mu\text{m}$ . The endcap region ( $|\eta| > 1.6$ ) is covered by the Tracker Inner Disks (TID)  
156 and the Tracker End Cap (TEC). The entire silicon strip detector is comprised of  
157 15200 high-sensitivity modules consisting of detecting unit, supporting structure and  
158 readout electronic system.

### 159 2.2.2 ECAL

160 The Electromagnetic Calorimeter (ECAL) measures the energy of photons, electrons  
161 and positrons. It it is placed just outside the tracker, but still inside the solenoid.  
162 ECAL is made of 74848 lead-tungstate ( $\text{PbWO}_4$ ) crystals. This material is charac-  
163 terized by a high density ( $8.28 \text{ g/cm}^3$ ), which gives the crystals a very compact form  
164 and makes them particularly suitable to be placed inside the magnetic coil. Another  
165 reason, this material has also a fast temporal response ( $\sim 10 \text{ ns}$ ) and its radiation  
166 length ( $X_0$ ) of  $0.89 \text{ cm}$  give ECAL the possibility to fully contain the expansion of  
167 the electromagnetic shower.

168 The arrangement of ECAL is shown in Fig. (2-6). The barrel crystals have a front  
169 face area of  $2.2 \times 2.2 \text{ cm}^2$  and a length of 23 cm. They are positioned at  $r = 1.29$   
170 m in pseudo-rapidity region  $0 < |\eta| < 1.479$ . The crystals in the endcaps have a  
171  $2.47 \times 2.47 \text{ cm}^2$  front face, a 22 cm length and they are positioned at  $z = 3.17 \text{ m}$  in  
172  $1.479 < |\eta| < 3.0$ . A Preshower detector is placed in front of the endcaps crystals.  
173 The active elements of Preshower are two planes of silicon strips with a pitch of 1.9  
174 mm, which lie behind disks of lead absorber at depths of  $2X_0$  and  $3X_0$ . It allows the

175 rejection of photon pairs from  $\pi^0$  decays and improves the estimation of the direction  
176 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),

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \quad (2.1)$$

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

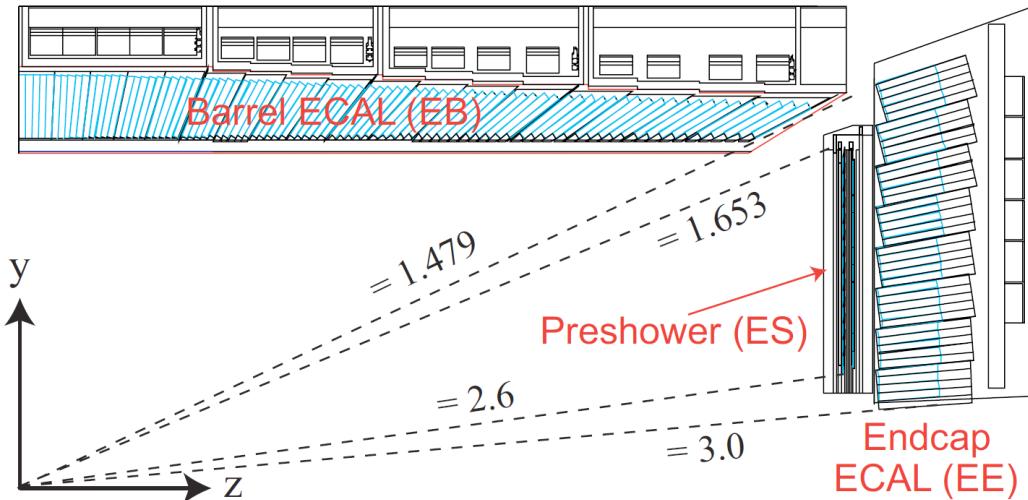


Figure 2-6: Schematic layout of the CMS ECAL.

### 180 2.2.3 HCAL

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

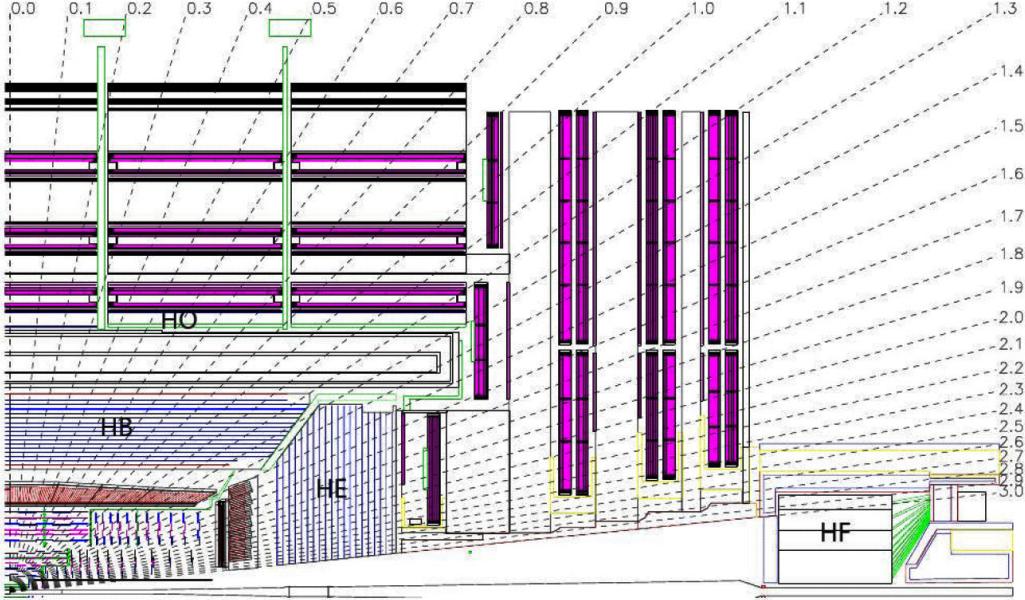


Figure 2-7: Longitudinal view of one quarter of CMS and the location of HB (hadron barrel calorimeter), HE (hadron endcap calorimeter), HF (hadron forward calorimeter) and HO (hadron outer calorimeter)

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

192 HCAL is a sampling calorimeter, whose active elements are plastic scintillators  
 193 interleaved with brass absorber plates and read out by wavelength shifting fibers.  
 194 Brass has been chosen as absorber material for its short interaction length and because  
 195 it is non-magnetic. The thickness of the absorber layers is between 60 mm in the barrel  
 196 and 80 mm in the endcaps. The barrel has 5.46 interaction lengths at  $\eta = 0$  and 10.82  
 197 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

$$\frac{\sigma_E}{E} \simeq \frac{a}{\sqrt{E}} \oplus 5\% \quad (2.2)$$

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

199 **2.2.4 Muon Chamber**

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

205 The muon system fulfills three purposes, muon identification, momentum mea-  
206 surement and triggering. Three different types of gaseous detectors are used for CMS  
207 muon system depending on the requirements.

208 **Drift Tube**

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

215 **Cathod Strip Chamber**

216 In the two endcaps, where the muon flux and the residual inhomogeneous magnetic  
217 field are higher, cathode strip chambers (CSC) are used. CSC is composed of anode  
218 wires and cathod strips in the gas volume. The chambers are arranged in 4 disks  
219 perpendicular to the beam, and in concentric rings (3 rings in the innermost station,  
220 2 in the others) in each of the endcaps.

221 **Resistive Plate Chambers**

222 Resistive plate chambers (RPC) are fast gaseous detectors that provide a muon trigger  
223 system parallel with DTs and CSCs. Each RPC consists of two parallel plates, a

<sup>224</sup> positively charged anode and a negatively charged cathode, both made of a very high  
<sup>225</sup> resistivity plastic material and separated by a gas volume.

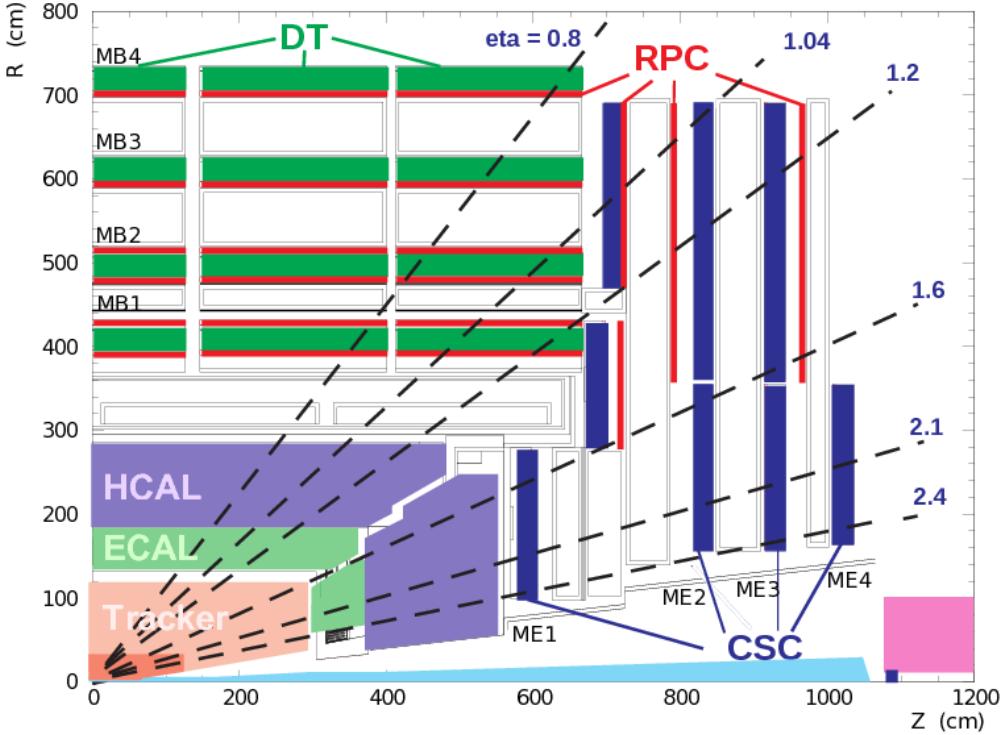


Figure 2-8: Slice view of one quarter of muon chamber system.

### <sup>226</sup> 2.2.5 Trigger System

<sup>227</sup> To have a good chance of producing rare particles, a very large number of collisions  
<sup>228</sup> is required (LHC proton bunches collide every 25 ns). Most collision events in the  
<sup>229</sup> detector are soft and do not produce interesting events. The amount of data from  
<sup>230</sup> each crossing is approximately 1 megabytes, which at the 40 MHz crossing rate would  
<sup>231</sup> result in 40 terabytes of data a second, an amount that the experiment cannot store.  
<sup>232</sup> The task of the trigger system is to reduce the storage rate while keeping a high  
<sup>233</sup> efficiency on the potentially interesting events. In CMS the input rate is reduced in  
<sup>234</sup> two steps, Level-1 Trigger (L1T) and High-Level Trigger (HLT).

235 **Level-1 Trigger**

236 After the Level-1 Trigger selection, the event recording frequency is decreased to  
237 100kHz [24], which is much smaller than the collision rate. The L1T objects are  
238 particles (such as photons, muons and electrons), jet candidates, global transverse  
239 energy and missing transverse energy. Level-1 Trigger just chooses the event with  $E_T$   
240 and  $P_T$  higher than the thresholds.

241 **High Level Trigger**

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

247

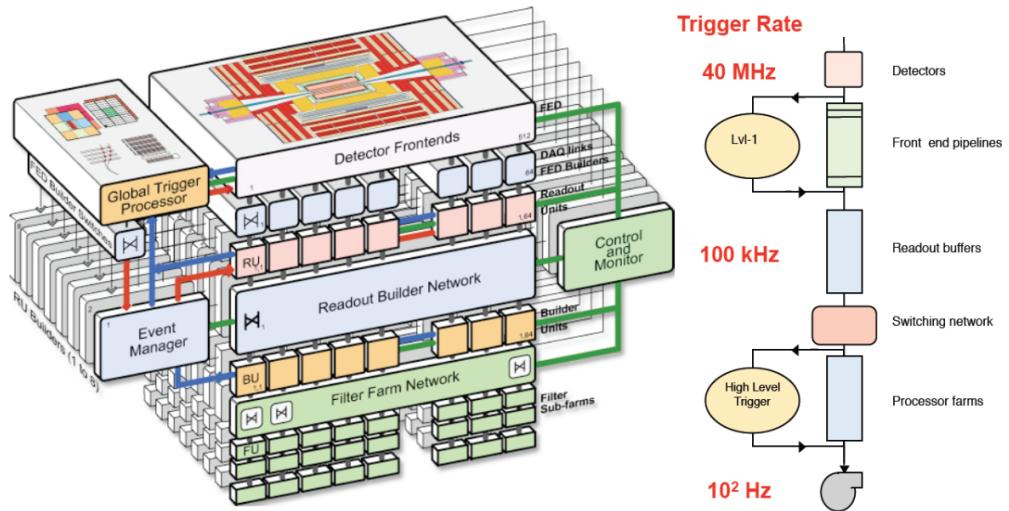


Figure 2-9: CMS triggering and data acquisition architecture.

<sup>248</sup> **Chapter 3**

<sup>249</sup> **Analysis Procedures**

<sup>250</sup> In this chapter, the analysis procedures of the search for  $Z'$  decaying into  $Zh$  in  $llbb$   
<sup>251</sup> final state are reported. The data sets and Monte Carlo (MC) samples we used in this  
<sup>252</sup> analysis will be indicated. Physics objects reconstruction and event selections are also  
<sup>253</sup> introduced. Moreover, background yields and the effects of systematic uncertainties  
<sup>254</sup> will be discussed in the end of this chapter.

<sup>255</sup> **3.1 Monte Carlo Samples and Data sets**

<sup>256</sup> **3.1.1 Signal MC**

<sup>257</sup> As introduced in section 1.2.3, the signal hypothesis is HVT model B benchmark.  
<sup>258</sup> The heavy resonance ( $Z'$ ) is tested using a wide set of masses from 800 GeV to 2000  
<sup>259</sup> GeV, one masspoint every 100 GeV (Table 3.1). The signal is generated by Mad-  
<sup>260</sup> Graph5\_aMC@NLOv5.2.2.1 [25] in LO mode, as a narrow spin-1 neutral resonance  
<sup>261</sup> and is forced to decay in the  $Z' \rightarrow Zh \rightarrow llqq$  channel. Showering and hadronization  
<sup>262</sup> are performed with PYTHIA6 [26].

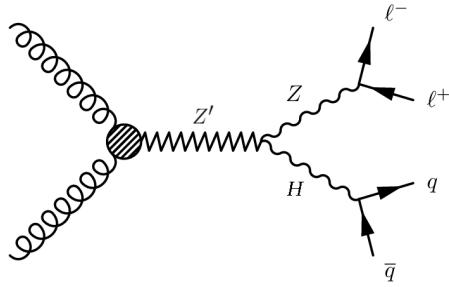


Figure 3-1: Feynman diagram for  $Z' \rightarrow Zh \rightarrow 2l2q$ .

Sample	Number of Processed Events	$\sigma_{LO}(\text{pb})$
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.

263    **3.1.2 Background MC**

264    Since we are looking for new resonances decaying in semi-leptonic final state, the  
265    background samples of this analysis are originated by all SM events with two leptons  
266    and at least one jet as final state. The dominant background contribution is the  
267    produciton of Z boson with jets. This Z+jets sample is produced by MADGRAPH.  
268    In the matrix element level, the Z boson is forced to decay into two leptons, and  
269    further this sample is divided into two samples depending on the Z  $p_T$ , higher than  
270    100 GeV or between 70 and 100 GeV. The contribution of events with Z  $p_T$  less than  
271    70 GeV is negligible due to further cut on the objects  $p_T$  in the selection criteria.

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

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

Sample	Number of Processed Events	$\sigma_{NLO}(\text{pb})$
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.

280    **3.1.3 Data Samples**

281    In this analysis, the full CMS data collected in 2012 are used, corresponding to the  
282    integrated luminosity of  $19.7 \text{ fb}^{-1}$  at the center-of-mass energy  $\sqrt{s} = 8$  TeV. For each  
283    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 ( $\text{pb}^{-1}$ )
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.

302    **3.3 Physics Objects**

303    **3.3.1 Muon**

304    **Reconstruction**

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

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

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

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

328    **Identification**

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

332

333    High- $p_T$  muon ID

334       • Muon identified as a *Global Muon*.

335       • Number of muon hits in the global track  $> 0$ .

336       • Number of matched muon stations  $> 1$ .

337       • Number of pixel hits  $> 0$ .

338       • Number of tracker layers with hits  $> 8$ .

339       • Transverse impact parameter  $d_{xy} < 0.2$  cm.

340       • Longitudinal impact parameter  $d_z < 0.5$  cm.

341       • Relative error on the track transverse momentum  $\sigma_{p_T}/p_T < 0.3$ .

342

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

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

352    In this analysis, a modified isolation criteria is used. The two muons originated  
353    from boosted Z decay are close to each other, and consequently the presence of another

354 muon in the isolation cone could break the function of this variable. In order to solve  
 355 this problem we use a tracker-based isolation relative quantity, explicitely removing  
 356 the momentum flow from any other muon passing our muon selection. Moreover, a  
 357 tracker-based isolation is well motivated also by two additional aspects: it is more  
 358 independent of pile up (pile-up tracks tipically do not match the primary vertex)  
 359 and does not include possible muon radiation. Finally, the modified requirement is  
 360  $I_{rel}^{mod} < 0.1$ .

361

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

Table 3.4: Summary of the muon ID selection criteria.

362 **3.3.2 Electron**

363 **Reconstruction**

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

369 **Clustering**

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

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

380 **Tracking**

381 There are several steps to reconstruct electron tracks in CMS experiment. The hit  
382 on the pixel and the tracker are reconstructed in the first step. The second step is  
383 to find the seed of tracks by matching at least two hits in pixel detector. Lst step  
384 is to fit the trajetory starting from the seed. To perform this fit, the Gaussian-sum  
385 filter (GSF) algorithm is used [32]. In GSF algorithm, the energy loss probability  
386 density function is constructed by multiple Gaussian functions. Finally, the electrons  
387 are built by matching the superclusters to the GSF tracks. Reconstruction efficiencies

388 for electrons  $E_T > 20$  GeV are generally above 95% in EB and 90% in EE [33].

## 389 Identification

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

393

394 Modified HEEPv4.1 electron ID

- 395 • Transversal supercluster energy  $> 35$  GeV.
- 396 • Pseudorapidity of supercluster  $|\eta_{\text{SC}}| < 1.442$  for barrel electrons, or  $1.56 <$   
397  $|\eta_{\text{SC}}| < 2.5$  for endcap electrons.
- 398 • Have either  $E_{2\times 5}/E_{5\times 5} > 0.94$  or  $E_{1\times 5}/E_{5\times 5} > 0.83$ .

399 • Ratio of HCAL deposit energy to ECAL deposit energy (Fig. 3-2) smaller than  
400 0.5.

- 401 • Number of inner layer lost hits smaller than 2.
- 402 • Have  $|d_{xy}| < 0.02$  cm for barrel electrons, or  $|d_{xy}| < 0.05$  cm for endcap electrons.

403

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

- 407 • Modified track isolation is required be to lower than 5 GeV. This variable is  
408 defined as the scalar  $p_T$  sum of the tracks within a  $\Delta R = 0.3$  cone around  
409 the electron, and excluding the  $p_T$  of another electron which passes the above  
410 selections and its track is inside the cone.

- 411 • The electromagnetic calorimeter isolation  $I_{\text{ECAL}}$  is defined as the scalar sum  
412 of  $E_T$  of the crystals in a  $\Delta R = 0.3$  cone around the particular electron (an

413 inner area of full-width 3 crystal), excluding a 4 crystals width around any other  
414 electron. The dimension of the ECAL crystals corresponds roughly at  $\Delta R \sim$   
415 0.01 to 0.02. The threshold of  $I_{ECAL}$  is varying with the electron transverse  
416 energy.

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

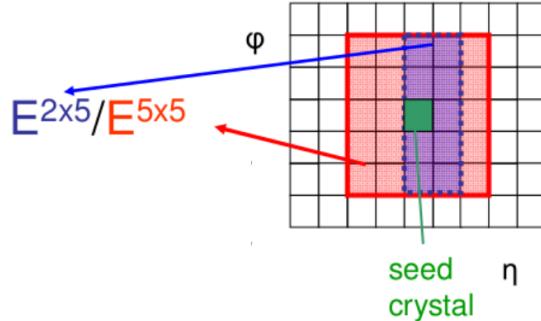


Figure 3-2: Example of the  $E_{2\times 5}/E_{5\times 5}$  computation in the ECAL crystals.  $E_{i\times j}$  is the energy contained in a  $i \times j$  block around the seed crystal (defined as the highest deposit of the energy of the cluster).

Variable	Barrel	Endcap
$E_T$	$> 35 \text{ GeV}$	$> 35 \text{ GeV}$
$ \eta_{SC} $	$ \eta_{SC}  < 1.442$	$1.56 <  \eta_{SC}  < 2.5$
H/E	$< 0.05$	$< 0.05$
$E_{2\times 5}/E_{5\times 5}$	$> 0.94$ or $E_{1\times 5}/E_{5\times 5} > 0.83$	-
Inner Layer Lost Hits	$<= 1$	$<= 1$
$ d_{xy} $	$< 0.02 \text{ cm}$	$< 0.05 \text{ cm}$
$Iso_{Trk}^{mod}$	$< 5 \text{ GeV}$	$< 5 \text{ GeV}$
$Iso_{Ecal,Hcal}^{mod}$	$< 2 \text{ GeV} + 0.03E_T$	$< 2.5 \text{ GeV for } E_T < 50 \text{ GeV}$ $< 2.5 \text{ GeV} + 0.03E_T \text{ for } E_T > 50 \text{ GeV}$

Table 3.5: Summary of the modified HEEPV4.1 electron ID.

422    **3.3.3 Jet**

423    Since gluons and quarks cannot exist in free state due to the color confinement  
424    [36], they fragment into hadrons. The result of this fragmentation (or called the  
425    "hadronization") is a jet of particles depositing energy in the detectors.

426    **Reconstruction**

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

437    The jet clustering algorithms are implemented as follows:

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

438    In Eq. (3.1), the pair of objects ( i , j ) denote two input candidate particles to  
439    the algorithm.  $\Delta R_{i,j}$  is the spatial distance between the two objects, and  $R$  is the  
440    typical cone size of the jet. The parameter  $p$  corresponds to different catagories of  
441    algorithms and will be introduced in the later paragraph.

442    The clustering proceeds by comparing the value of the two quantities,  $d_{iB}$  and  $d_{ij}$ .  
443    If  $d_{ij} > d_{iB}$ , the algorithm will look for another possible combination of ( i , j ). If  
444     $d_{ij} < d_{iB}$ , the 4-momentum of object i will merge into j, and then the object j forms a

445 pseudo-jet, but the algorithm will not stop looking for a new object until all particles  
446 are clustered into jet.

447 The parameter  $p$  can be chosen as follows:

448 •  $p = 0$  : Cambridge/Aachen algorithm;

449 •  $p = -1$  : Anti- $k_T$  algorithm;

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

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

## 457 Identification

458 As discussed in the previous section, this analysis use the Cambridge/Aachen algo-  
459 rithm with a cone radius of  $R = 0.8$  for the identification of jets (CA8 jets). Further-  
460 more, jets are required to pass the following loose identification criteria [42, 43]:

461

462 Loose JetID ( $> 99\%$  efficiency)

463 • Muon energy fraction smaller than 0.99

464 • Photon energy fraction smaller than 0.99

465 • Charged electromagnetic energy fraction smaller than 0.99

466 • Neutral hadron energy fraction smaller than 0.99

467 • Charged hadron energy fraction larger than 0

468 • Number of constituent particles larger than 1

469        For the "charged" variables,  $\eta < 2.4$  requirement are also applied, since there is  
470        no track coverage outside of this region, whereas the "neutral" variables extend to  
471        the whole  $\eta$  region.

472        **3.3.4 Jet Grooming Algorithms**

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

477        Typically three different grooming algorithms are considered, trimming [46], fil-  
478        tering [47] or pruning [48, 49]. Studies of these different grooming methods in CMS  
479        are presented in [50].

480        • **Trimming algorithm:** Trimming is a technique that ignores subjets below a  
481        minimum  $p_T$  fraction threshold within the original jet. Trimming reclusters the  
482        jet's constituents with a smaller radius  $R_{sub}$ . Then accepts only the subjets  
483        that have  $p_{T,frac} > f_{cut}$ , where the  $p_{T,frac}$  is defined as subjet  $p_T$  divided by the  
484        original jet  $p_T$ , and  $f_{cut}$  is typically taken proportional to  $H_T$ , the scalar sum of  
485        the  $p_T$  of all jet reconstructed in the event.

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

490        • **Pruning algorithm:** The idea is to take a jet of interest and then to reclus-  
491        ter it using a sequential clustering algorithm for vetoing soft and large-angle  
492        recombinations between pseudojets i and j.

493        Clustering proceeds as explained in the previous section, but it is vetoed if the  
494        candidates are too far away in  $\Delta R$ .

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

And the energy sharing is too asymmetric.

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

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.

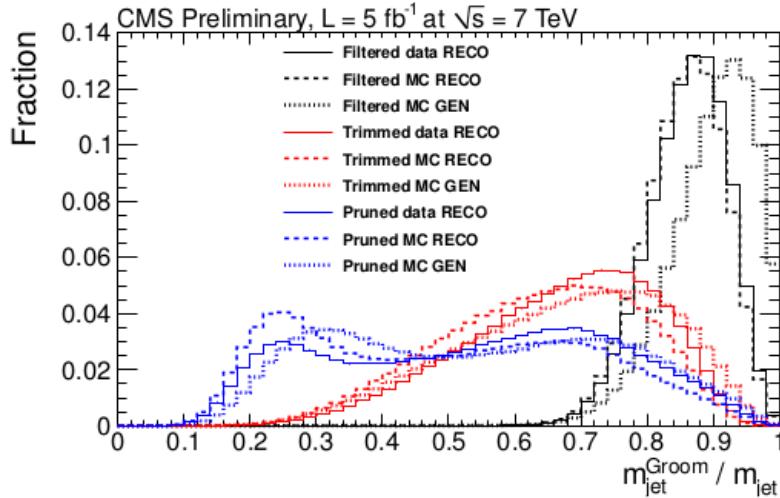


Figure 3-3: Comparison of the jet mass in generic QCD events from the groomed jets divided by the jet mass of matched ungroomed jets for the three grooming techniques, for both data and the PYTHIA 6 Monte Carlo. [1] Events are collected with a single jet trigger.

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 tecnicue 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.

499 Instead, when the pruned jet mass is large, the jets are split more symmetrically and  
500 more similar with the boson-jet structure. In this analysis we use the pruned jet mass  
501 because of its capability to improve the jet mass resolution and background rejection.

### 502 **3.3.5 B-tagging**

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

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

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

## 517 **Identification using Combined Secondary Vertex**

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

- 521 • secondary vertices must share less than 65% of their associated tracks with  
522 the primary vertex and the significance of the radial distance between the two  
523 vertices has to exceed  $3\sigma$ ;
- 524 • 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  $K^0$

526 or exceeding  $6.5 \text{ GeV}/c^2$  are rejected, reducing the contamination by vertices  
527 corresponding to the interactions of particles with the detector material and by  
528 decays of long-lived mesons;

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

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

- 535 • the vertex category (real, "pseudo" or "no vertex");  
536 • flight distance significance in the transverse plane ("2D");  
537 • the vertex mass;  
538 • number of tracks at the vertex;  
539 • ratio of the energy carried by tracks at the vertex with respect to all tracks in  
540 the jet;  
541 • pseudorapidities of the tracks at the vertex with respect to the jet axis;  
542 • the 2D IP (impact parameter) significance of the first track that raises the  
543 invariant mass above the charm threshold of  $1.5 \text{ GeV}/c^2$  (tracks are ordered  
544 by decreasing IP significance and the mass of the system is recalculated after  
545 adding each track);  
546 • the number of tracks in the jet;  
547 • the 3D IP significances for each track in the jet.

548 Two likelihood ratios are built from these variables. They are used to discriminate  
549 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.

556

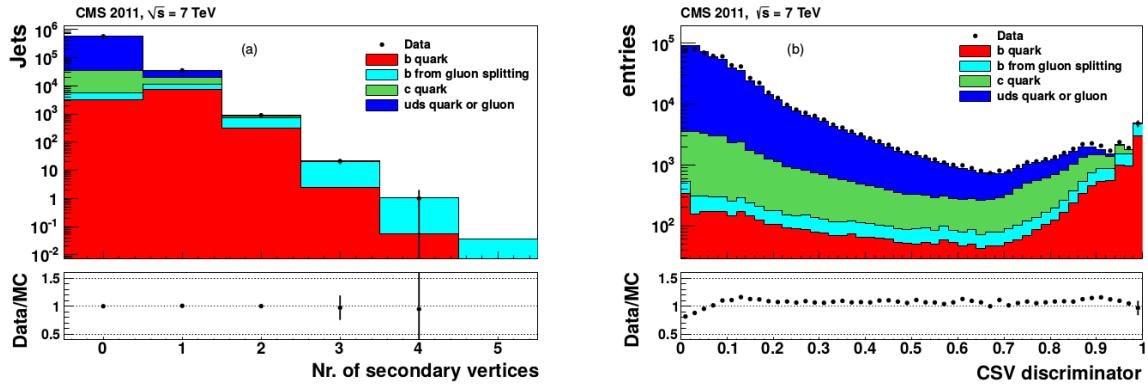


Figure 3-4: [2] Distributions of (a) the secondary vertex multiplicity and (b) the CSV discriminator.

### 557 3.4 Pile-up reweighting

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

562 The simulation generates the pile-up roughly to match the condition in data,  
 563 however there are still differences between the pile-up numbers in data and MC. It is  
 564 necessary to reweight pile-up distributions of MC samples to match the data more  
 565 precisely. By applying a proper weight to each MC event according to the pile-up  
 566 distribution from data, the MC samples can describe the data better. Fig. 3-5 shows  
 567 the number of vertices after pile-up reweighting in both two lepton channels.

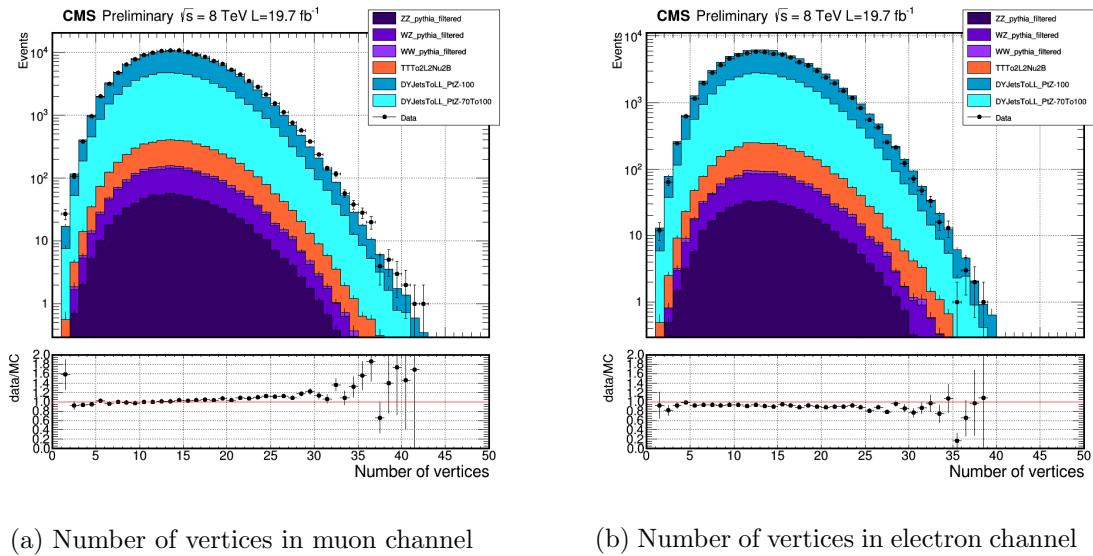


Figure 3-5: Number of vertices distributions after pile-up reweighting. Data are compared to the combination of all MC background samples. After reweighting, the distributions are almost identical to the data in both channels.

568    **3.5 Event and Object selection**

569    **3.5.1 Lepton Requirements**

570    **Muon Selection**

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

576    **Electron Selection**

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

584    **3.5.2 Jet Requirement**

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

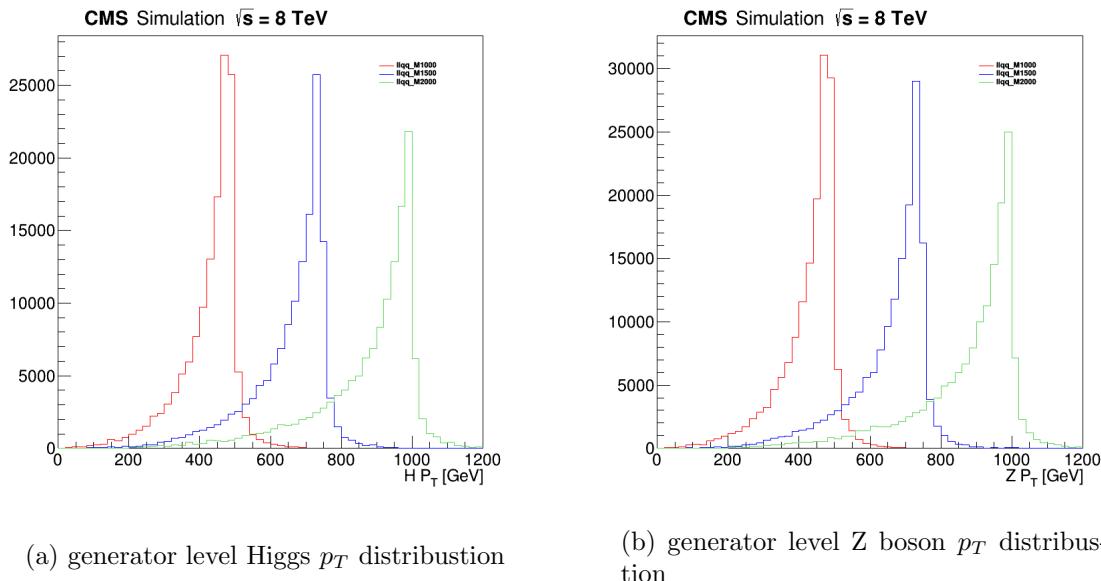
590    Furthermore, in order to veto leptons that are mis-identified as jets, leptons overlap  
591    with jets are removed by the  $\Delta R$  cut, i.e. if there's a lepton passing all lepton  
592    selections and the spatial distance to a CA8jet smaller than 0.1 ( $\Delta R_{jet,lepton} < 0.1$ ),

593 then the jet will be removed.

### 594 3.5.3 Z boson Requirement

595 The Z boson candidate is reconstructed by adding four-momentum of the selected  
 596 lepton pair. Since the Z boson mass is about 91 GeV, we require the reconstructed  
 597 invariant mass of the Z boson in the mass region [70 GeV, 110 GeV] where is  $\pm 20$   
 598 GeV to its theoretical mass.

599 For the CA8jet from Higgs, we require  $p_T$  threshold as 80 GeV. The kinematics of  
 600 reconstructed Z boson and Higgs should be symmetric, because they are both decayed  
 601 from the heavy  $Z'$ , comparing their mass to  $Z'$ , the difference between 125 GeV and  
 602 91 GeV is negligible ( $1 \text{ TeV} \gg 125 \text{ GeV} \sim 91 \text{ GeV}$ ). Therefore we require the same  
 603  $p_T$  threshold to the Z boson. Fig. 3-6 shows the transverse momentum distributions  
 604 from the signal samples.



(a) generator level Higgs  $p_T$  distribution

(b) generator level Z boson  $p_T$  distribution

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.

605 Finally, all selection requirements are summarized in Table 3.6.

Selection	Value	Comments
Trigger	HLT_Mu22_TkMu8 HLT_DoubleEle33	DoubleMu dataset DoublePhoton dataset
Leading muon $p_T$	$p_T > 40$ GeV	
Sub-leading muon $p_T$	$p_T > 20$ GeV	
Muon $\eta$	$ \eta  < 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 $\eta$	$ \eta  < 2.5$ out of [1.4442,1.566]	To avoid ECAL gap.
Electron ID	HEEP modified	
Electron isolation		
$I_{trk}^{mod}$	$< 5$ GeV	
$I_{ECAL,HCAL}^{mod}$	$< 2 \text{ GeV} + 0.03E_T$ $< 2.5 \text{ GeV}$ $< 2.5 \text{ GeV} + 0.03E_T$	Barrel for $E_T < 50$ GeV candidates in Endcap for $E_T > 50$ GeV candidates in Endcap
Jet ID	Loose working point	
Jet $p_T$	$p_T > 80$ GeV	
Jet $\eta$	$ \eta  < 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	$70 \text{ GeV} < m_Z < 110 \text{ GeV}$	

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

## 606 3.6 Data-MC comparison

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

611 On top of the selections described in previous section, additional regions are de-  
612 fined as following:

- 613 • **Signal region (SR):** Represents the phase space where the signal is expected,  
614 defined by the prunedjet mass in  $110 \text{ GeV} < m_{\text{prunedjet}} < 140 \text{ GeV}$  region. The  
615 range is chosen by  $\pm 15 \text{ GeV}$  to the mass of Higgs.
- 616 • **Sidebands (SB):** Defined by the interval between  $70 \text{ GeV} < m_{\text{prunedjet}} <$   
617  $110 \text{ GeV}$ . This region is signal-depleted. In our case, we don't consider pruned-  
618 jet mass higher than  $140 \text{ GeV}$ , because of the poor statistics and the excessive  
619 contribution of  $t\bar{t}$  events.

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

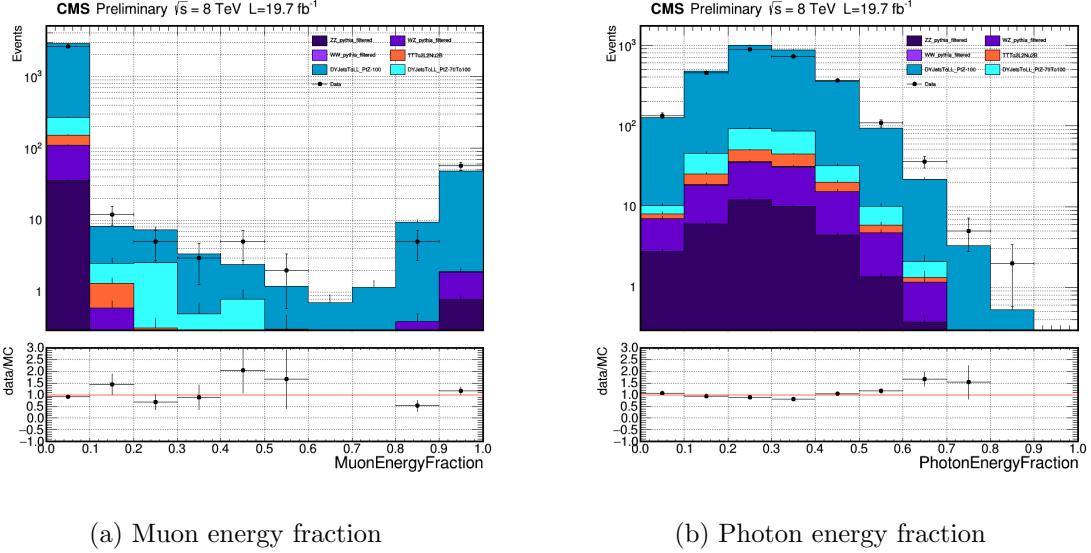


Figure 3-7: Comparison between data and all background samples for two jet variables. The definition of muon/photon energy fraction is muon/photon energy divided by jet energy.

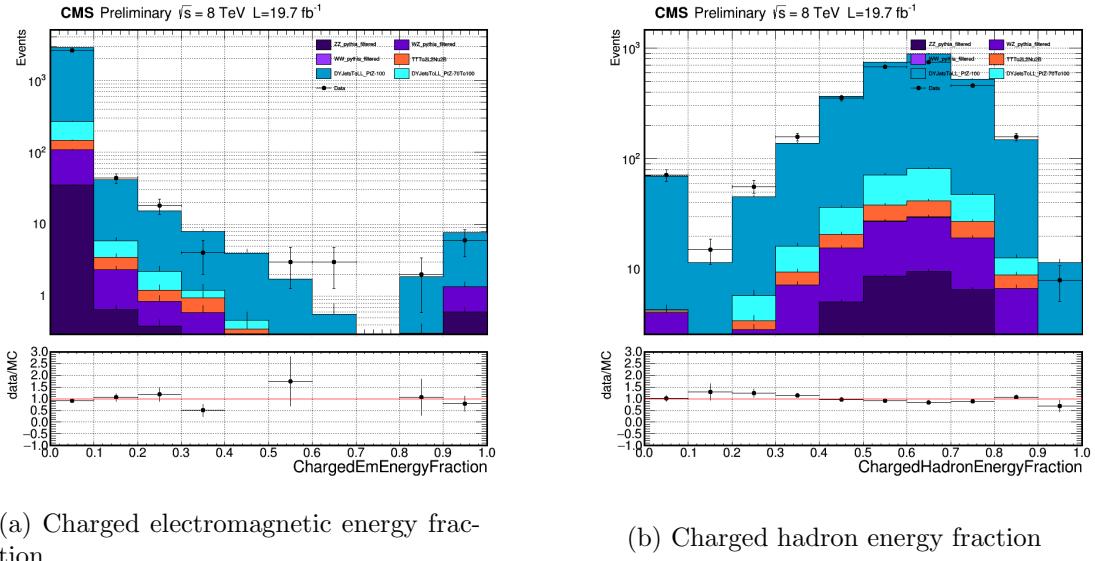


Figure 3-8: Charged electromagnetic/hadron energy fraction is defined by the ratio of the energy of charged particles in ECAL/HCAL to the jet energy.

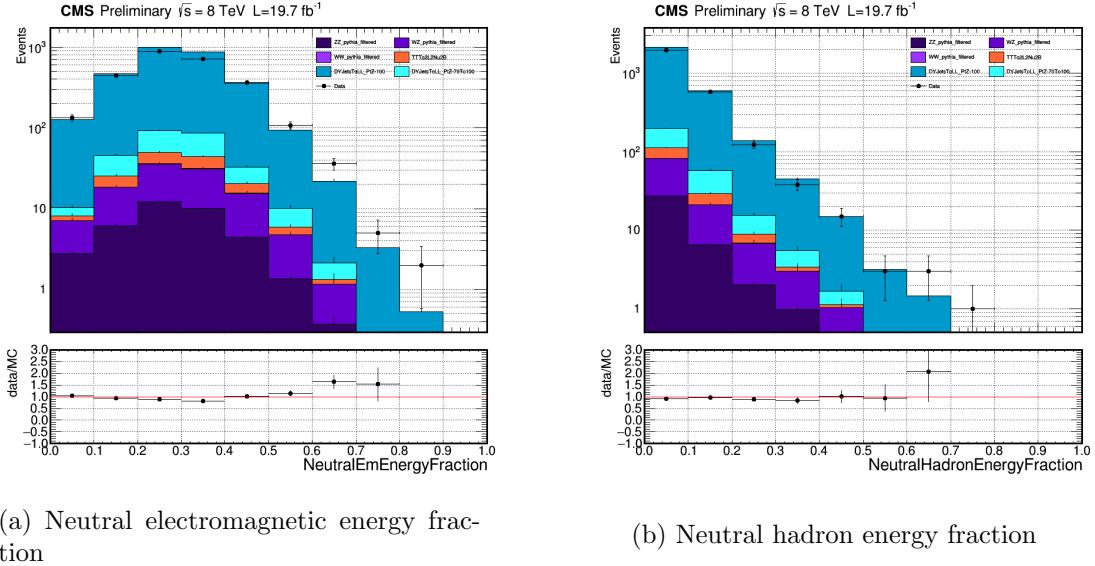


Figure 3-9: Neutral electromagnetic/hadron energy fraction is defined by the ratio of the energy of neutral particles in ECAL/HCAL to the jet energy.

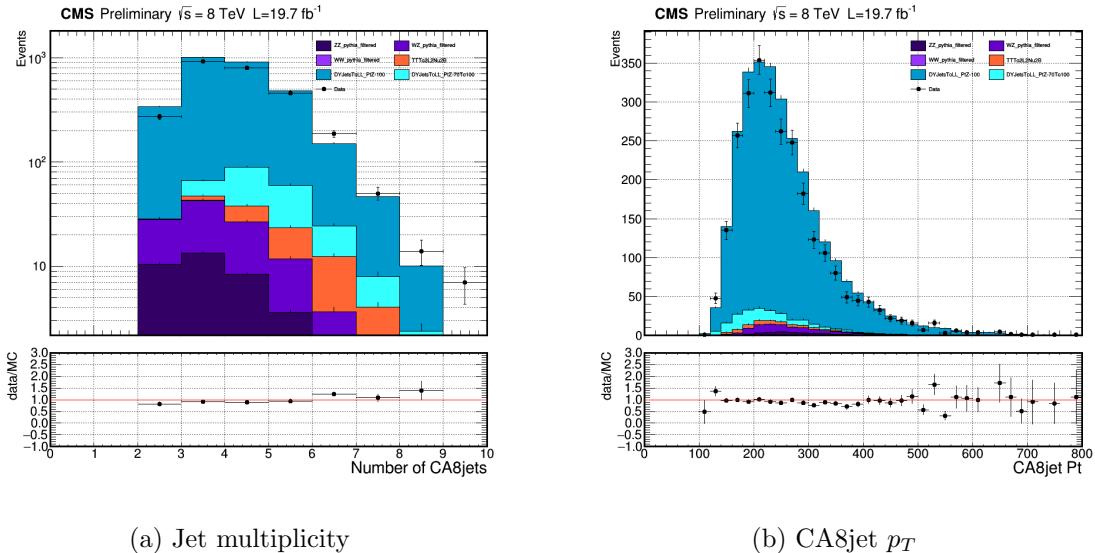


Figure 3-10: Comparison between data and MC in SB region using jet multiplicity (number of jets) and CA8jet transverse momentum.

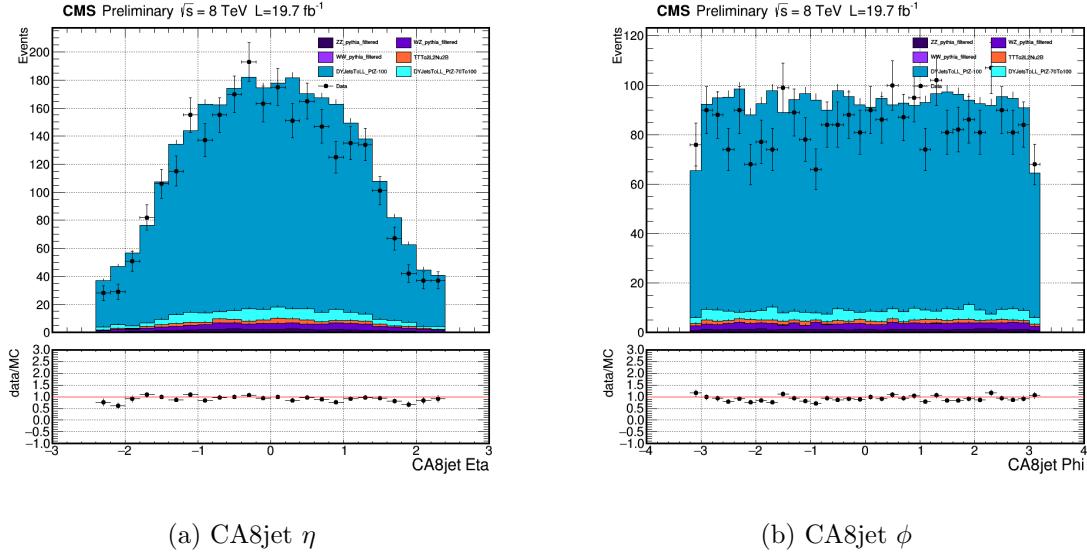


Figure 3-11: Comparison between data and MC in SB region using CA8jet  $\eta$  and  $\phi$ .

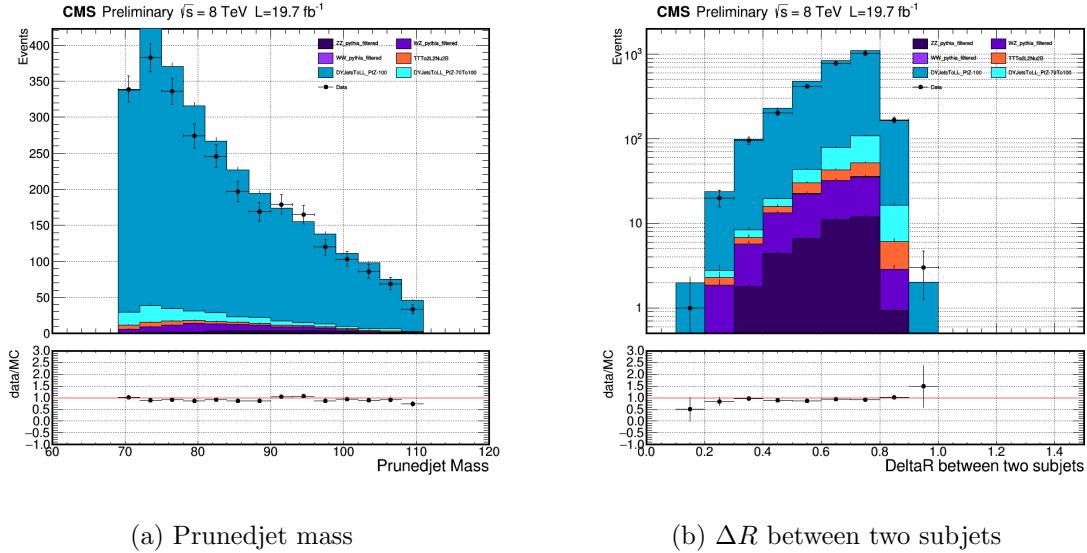


Figure 3-12: Left: the prunedjet mass in the SB region. Right: the spatial distance between two subjets within the CA8jet.

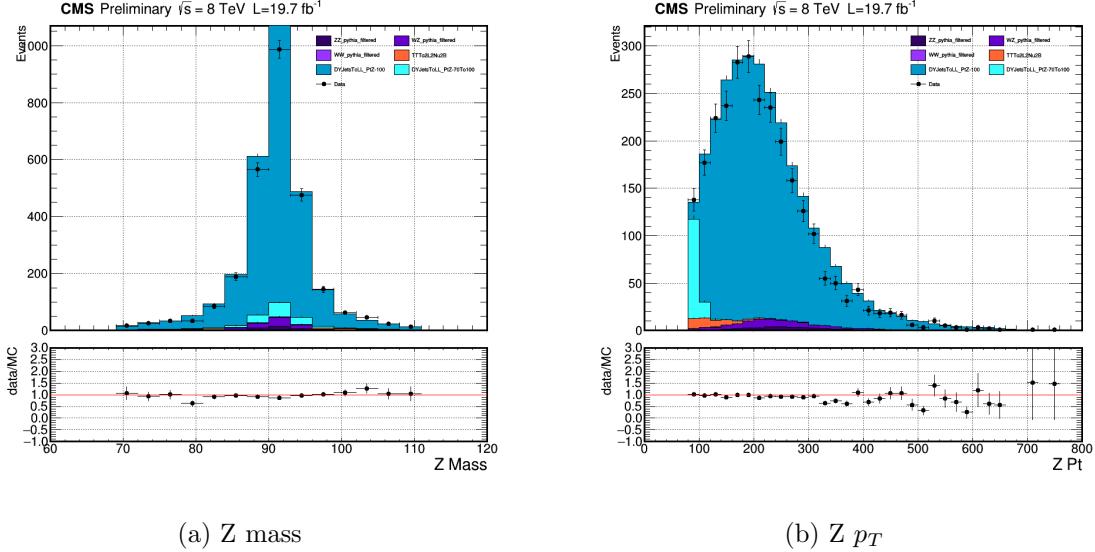


Figure 3-13: Comparison between data and MC in SB region using mass and transverse momentum of reconstructed Z boson.

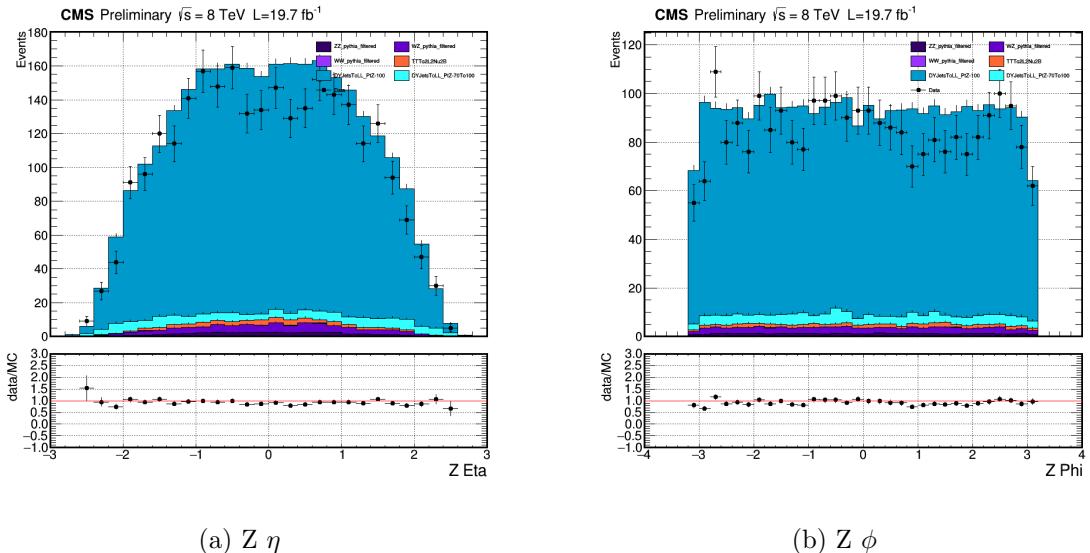


Figure 3-14: Comparison between data and MC in SB region using  $\eta$  and  $\phi$  of reconstructed Z boson.

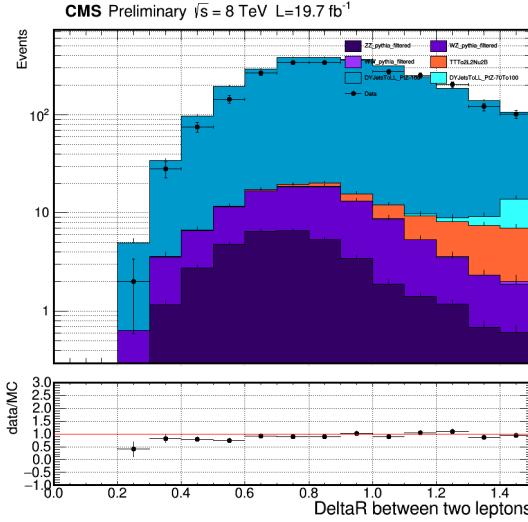


Figure 3-15:  $\Delta R$  between the two selected leptons.

## 624 3.7 Background Estimation

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

### 630 3.7.1 $\alpha$ ratio method

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

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

639 We divided the spectrum in 14 non-uniform width bins, as shown in Table 3.7,  
640 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.

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

## 643 3.8 Signal Yields

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

### 646 3.8.1 signal efficiency

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

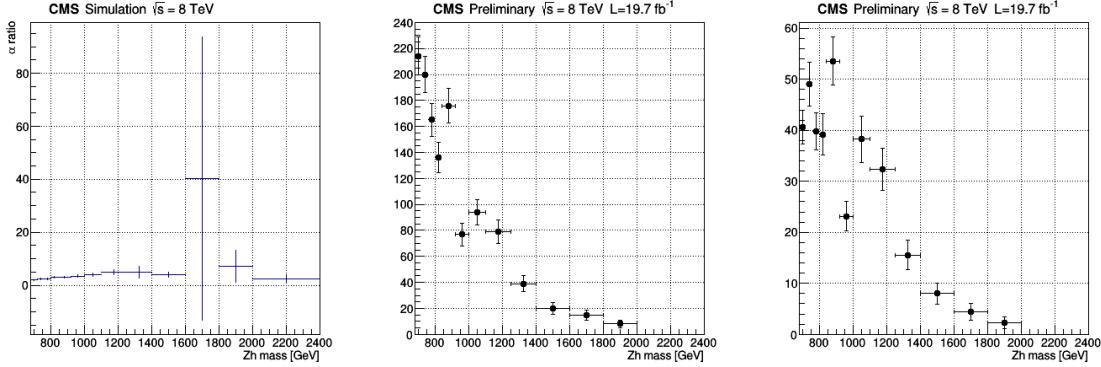


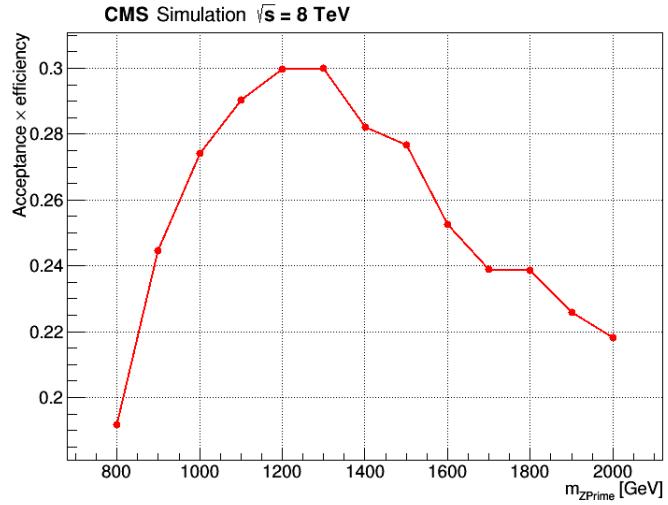
Figure 3-16: Left figure: the  $\alpha$  ratio from MC simulation. Central figure: the  $m_{Zh}$  distribution observed from data SB. Right figure: the predicted  $m_{Zh}$  distribution in data signal region.

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

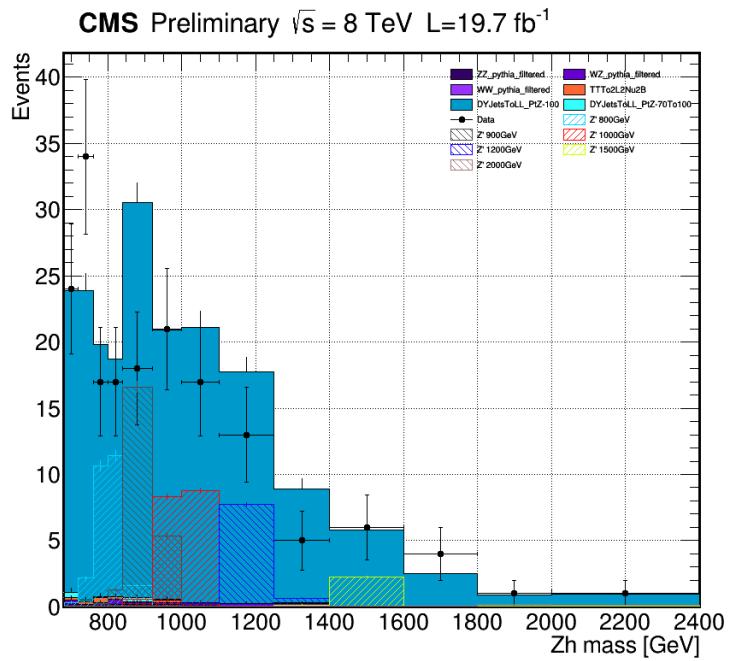
650 In Eq. (3.5), those number of events include both muon and electron channels.  
 651 There are totally thirteen masspoints of  $Z'$  in our signal MC samples. For each  
 652 masspoint, the result of efficiencies are shown in Fig. 3-17 (a).

### 653 3.8.2 The $m_{Zh}$ spectrum

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



(a) Signal efficiencies for each  $Z'$  samples



(b)  $m_{Zh}$  spectrum

Figure 3-17: (a) The signal efficiencies plot of each  $Z'$  samples. (b) Invariant mass distribution of the MC background simulation compared to the observed data and MC signal shape in the SR.

### 663 3.8.3 The CSV distribution

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

669 Although the recommend CSV selection (loose working point) from BTV group  
 670 [51] for the boosted Higgs decay result in very low effiency for both background  
 671 and signal MC samples. Instead of taking selection on the CSV, we use the overall  
 672 distribution as another source of shape information to set the limits. The method  
 673 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 $\Delta R_{subjets} < 0.3$	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  $\Delta R$  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.

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

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

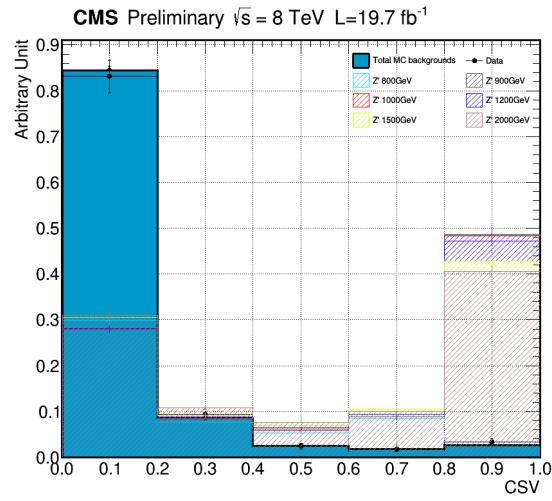


Figure 3-18: The CSV distribution comparison between data and MC samples in SR.

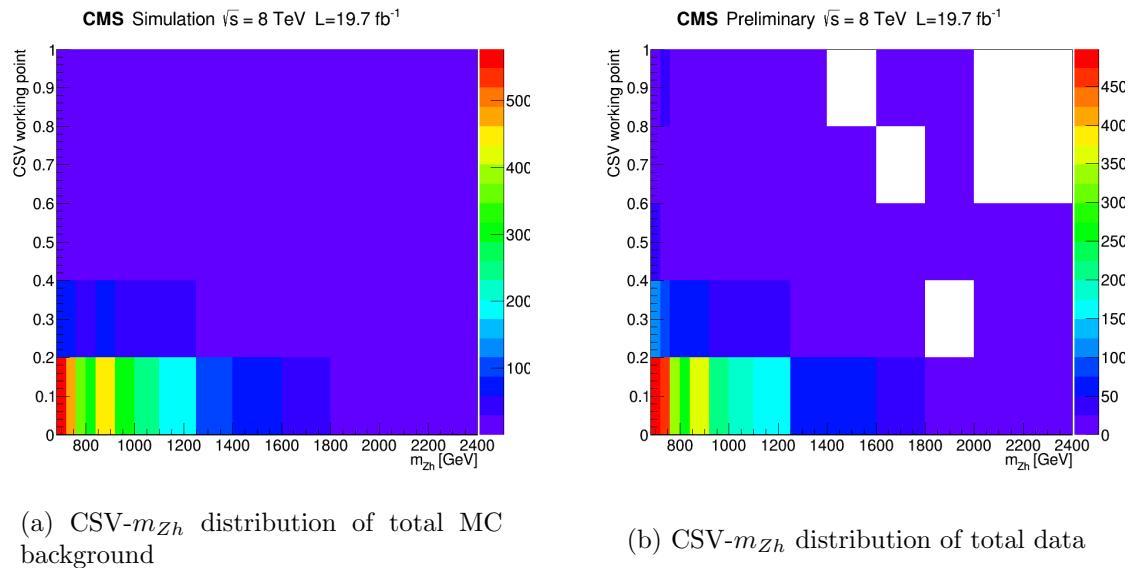


Figure 3-19: The combined 2D shape result of data and MC background in SR.

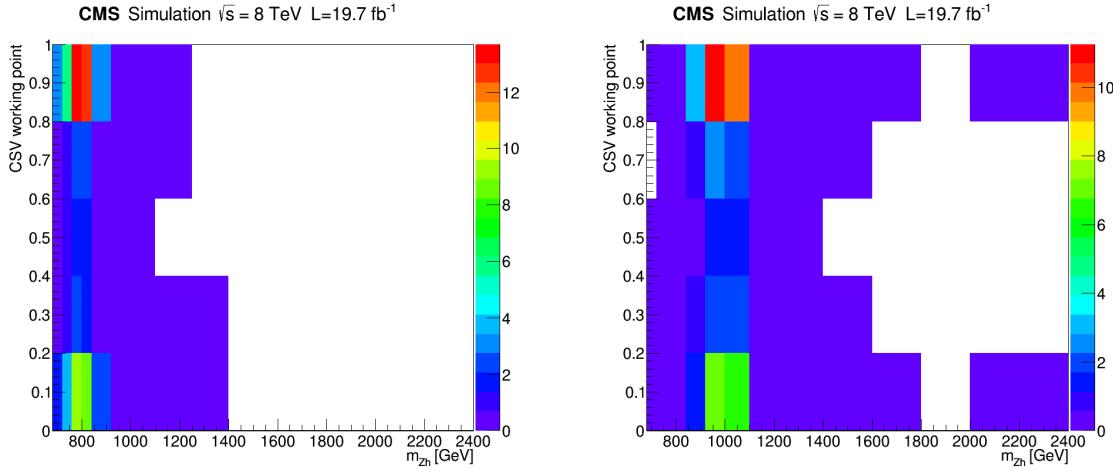


Figure 3-20: The combined 2D shape result in SR of 800 GeV and 1000 GeV signal MC samples.

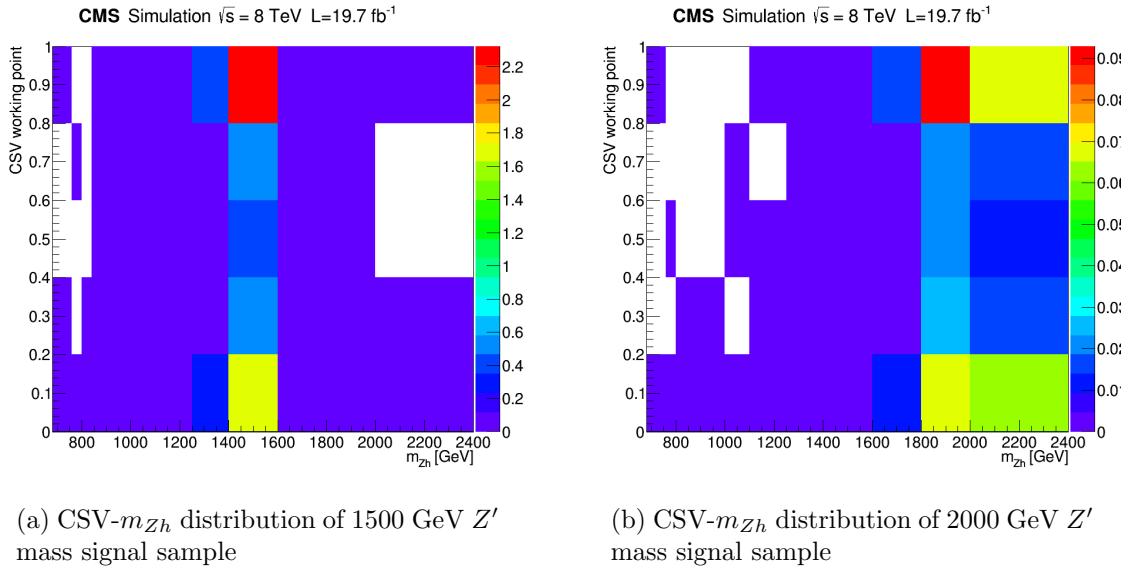


Figure 3-21: The combined 2D shape result in SR of 1500 GeV and 2000 GeV signal MC samples.

## 683 3.9 Systematic uncertainties

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

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

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

695 • **CSV distribution normalization:** This systematic is from normalizing the  
696 CSV distributions of MC background prediction in order to fit the data, it is  
697 estimated about 10% for each background.

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

705 • **Lepton ID scale factor:** Text.

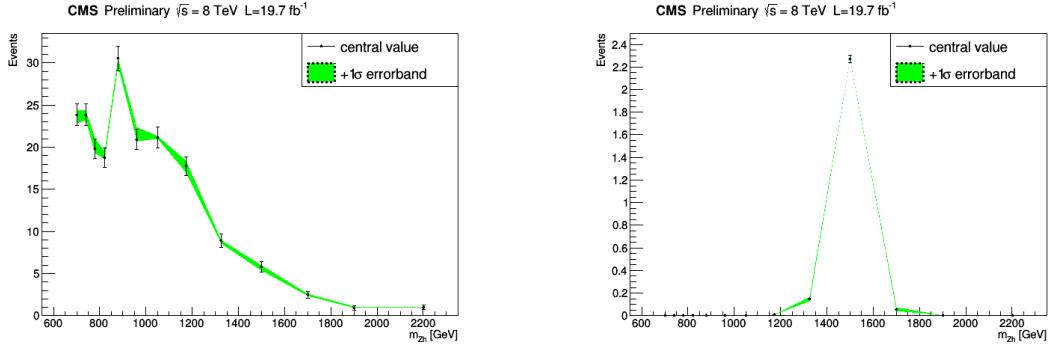


Figure 3-22: SR  $m_{Zh}$  distributions for both signal (1500 GeV) and background MC samples. The uncertainty of jet energy scale is shown as the green error band ( $\pm 1\sigma$ ), while the error bar presents the statistic error.

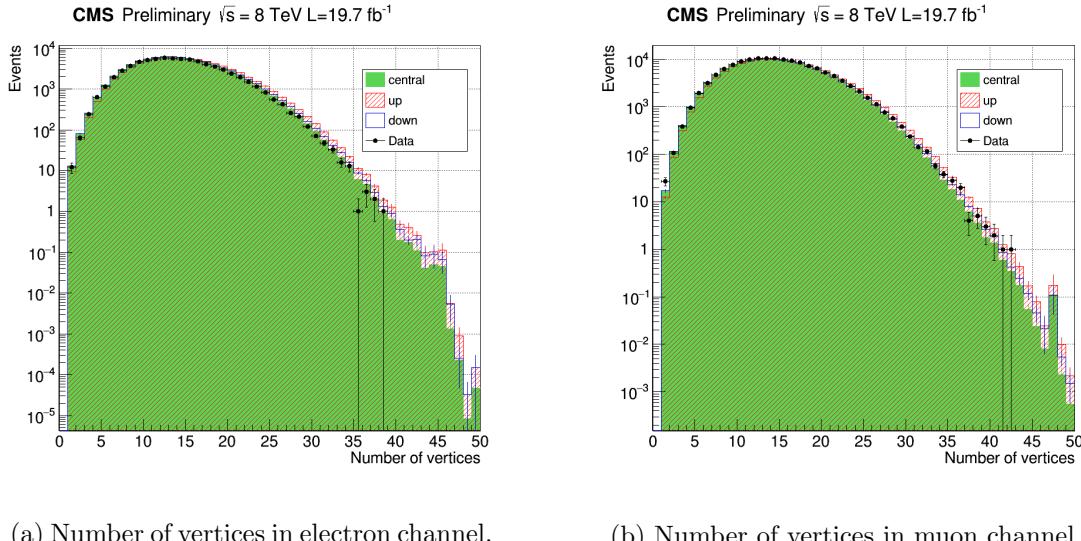


Figure 3-23: (a) shows the distributions of number of vertices in electron channel, comparing central value of the total background prediction and the  $\pm 1\sigma$  variation, and data as well. (b) shows the result in muon channel.

# 706 Bibliography

- 707 [1] CMS Collaboration. Jet mass in di-jet and vector boson plus jet events. CMS  
708 Physics Analysis Summary CMS-PAS-SMP-12-019, CMS, 2012.
- 709 [2] CMS Collaboration. Identification of b-quarks jets with the CMS experiment.  
710 *JINST*, 8:P04013, 2013.
- 711 [3] ATLAS Collaboration. Observation of a new particle in the search for the Stan-  
712 dard Model Higgs boson with the ATLAS detector at the LHC. *Phy. Lett. B*,  
713 716:1, 2012.
- 714 [4] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the  
715 CMS experiment at the LHC. *Phy. Lett. B*, 716:30, 2012.
- 716 [5] CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp  
717 collisions at  $\sqrt{s} = 7$  and 8 TeV. *JHEP*, 06:081, 2013.
- 718 [6] CMS Collaboration. Precise determination of the mass of the Higgs boson and  
719 tests of compatibility of its couplings with the standard model predictions using  
720 proton collisions at 7 and 8 TeV. *Eur. Phys. J. C*, 75:212, 2015.
- 721 [7] ATLAS Collaboration. Measurement of the Higgs boson mass from the  $H \rightarrow \gamma\gamma$   
722 and  $H \rightarrow ZZ^* \rightarrow 4l$  channels in pp collisions at center-of-mass energies of 7 and  
723 8 TeV with the ATLAS detector. *Phys. Rev. D*, 90:052004, 2014.
- 724 [8] ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using  
725 ATLAS data. *Phys. Lett. B*, 726:120, 2013.
- 726 [9] CMS and ATLAS Collaboration. Combined Measurement of the Higgs Boson  
727 Mass in pp Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experi-  
728 ments. *Phys. Rev. Lett.*, 114:191803, 2015.
- 729 [10] Mads T. Frandsen. Minimal Walking Technicolor. arXiv:0710.4333 [hep-ph],  
730 <http://arxiv.org/abs/0710.4333>, 2007.
- 731 [11] T. Han, H. E. Logan, B. McElrath, and L.-T. Wang. Phenomenology of the little  
732 Higgs model. *Phys. Rev. D*, 67:095004, 2003.
- 733 [12] M. Schmaltz and D. Tucker-Smith. LITTLE HIGGS THEORIES. *Annual Review*  
734 *of Nuclear and Particle Science*, 55:no. 1, 229–270, 2005.

- 735 [13] M. Perelstein. Little Higgs models and their phenomenology. *Progress in Particle  
736 and Nuclear Physics*, 58:no. 1, 247–291, 2007.
- 737 [14] R. Contino, D. Pappadopulo, D. Marzocca, and R. Rattazzi. On the effect of  
738 resonances in composite Higgs phenomenology. *Journal of High Energy Physics*,  
739 2011:no. 10, 1–50, 2011.
- 740 [15] D. Marzocca, M. Serone, and J. Shu. General composite Higgs model. *Journal  
741 of High Energy Physics*, 2012:no. 8, 1–52, 2012.
- 742 [16] B. Bellazzini, C. Csaki, and J. Serra. Composite Higgses. *Eur. Phys. J.*, C74:no.  
743 5, 2766, 2014.
- 744 [17] D. Pappadopulo, R. Torre A. Thamm, and A. Wulzer. Heavy vector triplets:  
745 bridging theory and data. *Journal of High Energy Physics*, 2014:no. 9, 1–50,  
746 2014.
- 747 [18] Andrea Wulzer. An Equivalent Gauge and the Equivalence Theorem.  
748 arXiv:1309.6055 [hep-ph], <http://arxiv.org/abs/1309.6055>, 2013.
- 749 [19] V. Barger, W.-Y. Keung, and E. Ma. Gauge model with light W and Z bosons.  
750 *Phys. Rev. D*, 22:727, 1980.
- 751 [20] Lisa Benato, Yu-Hsiang Chang, Ching-Wei Chen, Michele de Gruttola, Ra-  
752 man Khurana Ji-Kong Huang, Stefano Lacaprara, Yun-Ju Lu, Jacopo Pazzini,  
753 Maurizio Pierini, Henry Yee-Shian Tong, Jun-Yi Wu, Shin-Shan Eiko Yu, Marco  
754 Zanetti, and Alberto Zucchetta. Search for heavy resonances decaying into a vec-  
755 tor boson and a Higgs boson in the  $(ll, l\nu, \nu\nu)b\bar{b}$  final state. CMS Note 2015/186,  
756 CMS, 2015.
- 757 [21] A. Benaglia. The CMS ECAL performance with examples. *JINST*, 09:C02008,  
758 2014.
- 759 [22] CMS HCAL Collaboration. Design, Performance, and Calibration of CMS  
760 Hadron-Barrel Calorimeter Wedges. CMS Note 2006/138, CMS, 2006.
- 761 [23] Victor Daniel Elvira. Measurement of the Pion Energy Response and Resolution  
762 in the CMS HCAL Test Beam 2002 Experiment. CMS Note 2004/020, CMS,  
763 2004.
- 764 [24] J. Brooke on behalf of the CMS Collaboration. Performance of the CMS Level-1  
765 Trigger. arXiv:1302.2469 [hep-ex], <http://arxiv.org/abs/1302.2469>, 2013.
- 766 [25] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S.  
767 Shaoand T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of  
768 tree-level and next-to-leading order differential cross sections, and their matching  
769 to parton shower simulations. *JHEP*, 07:079, 2014.

- 770 [26] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics  
 771 and Manual. *JHEP*, 0605:026, 2006.
- 772 [27] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. The POWHEG-hvq manual  
 773 version 1.0. arXiv:0707.3081 [hep-ph], <http://arxiv.org/abs/0707.3081>, 2007.
- 774 [28] S. Chatrchyan et al. Performance of CMS muon reconstruction in pp collision  
 775 events at  $\sqrt{s}$  TeV. *JINST*, 7:P10002, 2012.
- 776 [29] Wolfgang Adam, Boris Mangano, Thomas Speer, and Teddy Todorow. Track  
 777 Reconstruction in the CMS tracker. CMS Note 2006/041, CMS, 2006.
- 778 [30] CMS Collaboration. Baseline muon selections for Run-I.  
 779 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId>.
- 780 [31] CMS Collaboration. Physics TDR volume I, The CMS Physics Technical Design  
 781 R I. CERN-LHCC-2006-001, page 521pp, 2006.
- 782 [32] W. Adam, R. Frhwirth, A. Strandlie, and T. Todorov. Reconstruction of elec-  
 783 trons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of  
 784 Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005.
- 785 [33] CMS Collaboration. E/gamma Commissioning and physics deliverables.  
 786 <https://twiki.cern.ch/twiki/bin/view/CMS/EgCommissioningAndPhysicsDeliverables>.
- 787 [34] CMS Collaboration. HEEP Electron ID and isolation.  
 788 <https://twiki.cern.ch/twiki/bin/viewauth/CMS/HEEPElectronID>.
- 789 [35] T. S. Williams, D. Newbold, and C. Shepherd-Themistocleous. Modified isolation  
 790 cuts for the selection of high  $p_T$  electrons from the decay of boosted Z bosons.  
 791 CMS Note 2012/168, CMS, 2012.
- 792 [36] Wikipedia. Color confinement. [https://en.wikipedia.org/wiki/Color\\_confinement](https://en.wikipedia.org/wiki/Color_confinement).
- 793 [37] M. Cacciari, G. P. Salam, and G. Soyez. The anti- $k_t$  jet clustering algorithm.  
 794 *JHEP*, 04:063, 2008.
- 795 [38] M. Wobisch and T. Wengler. Hadronization corrections to jet cross section in  
 796 deep-inelastic scattering. arXiv:9907280 [hep-ph], <http://arxiv.org/abs/9907280>,  
 797 1999.
- 798 [39] CMS Collaboration. Commissioning of the Particle-Flow reconstruction in  
 799 Minimum-Bias and Jet Events from pp Collisions at 7 TeV. CMS Physics Anal-  
 800ysis Summary CMS-PAS-PFT-10-002, CMS, 2010.
- 801 [40] Y.L. Dokshitzer et al. Better jet clustering algorithms. *JHEP*, 08:001, 1997.
- 802 [41] CMS Collaboration. A Cambridge-Aachen (C-A) based Jet Algorithm for  
 803 boosted top-jet tagging. CMS Physics Analysis Summary CMS-PAS-JME-09-  
 804 001, CMS, 2009.

- 805 [42] N.Saoulidou and E.Tziaferi. Performance of the Particle-Flow jet identification  
 806 criteria using proton-proton collisions at = 8 TeV. CMS Note 2014/227, CMS,  
 807 2014.
- 808 [43] CMS Collaboration. Jet Identification. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID>.
- 809 [44] P. Tran et al. Measurement of jet mass in V+Jets events. CMS Note 2012/137,  
 810 CMS, 2012.
- 811 [45] A. Bonato, R. Gerosa, and et al. Identifying hadronically decaying W bosons  
 812 merged into a single jet. CMS Note 2013/072, CMS, 2013.
- 813 [46] J. Thaler, D. Krohn, and L.T. Wang. Jet Trimming. *JHEP*, 100:242001, 2010.
- 814 [47] J.M. Butterworth et al. Jet substructure as a new Higgs search channel at the  
 815 LHC. *PhysRevLett*, 100:242001, 2008.
- 816 [48] S. Ellis, C. Vermilion, and J. Walsh. Techniques for improved heavy particle  
 817 searches with jet substructure. *Phys. Rev.*, D80:051501, 2009.
- 818 [49] S. Ellis, C. Vermilion, and J. Walsh. Recombination algorithms and jet sub-  
 819 structure: Pruning as a tool for heavy particle searches. *Phys. Rev.*, D81:094023,  
 820 2010.
- 821 [50] S. Chatrchyan et al. Studies of jet mass in di-jet and W/Z + jet events. *JHEP*,  
 822 1305:090, 2013.
- 823 [51] CMS Collaboration. Performance of b tagging at  $\sqrt{s} = 8$  TeV in multihjet,  $t\bar{t}$   
 824 and boosted topology events. CMS Physics Analysis Summary CMS-PAS-BTV-  
 825 13-001, CMS, 2013.
- 826 [52] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard Model  
 827 Higgs-Boson Branching Ratios with Uncertainties. arXiv:1107.5909 [hep-ph],  
 828 <http://arxiv.org/abs/1107.5909>, 2011.
- 829 [53] CMS Collaboration. CMS Luminosity Based on Pixel Cluster Counting - Summer  
 830 2013 Update. CMS Physics Analysis Summary CMS-PAS-LUM-13-001, CMS,  
 831 2013.
- 832 [54] CMS Collaboration. Estimating Systematic Errors Due to Pileup Modeling.  
 833 <https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors>.