

**Search for heavy resonances decaying to a pair of  
Higgs bosons in four b quark final state in  $pp$   
collisions at  $\sqrt{s} = 13$  TeV using the CMS detector at  
the LHC**

by

Gregorio de Leon III

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

Master of Science

at the

NATIONAL CENTRAL UNIVERSITY

© National Central University 2017. All rights reserved.

Author .....

Department of Physics  
September 22, 2017

Certified by .....

Shin-Shan Eiko Yu  
Associate Professor  
Thesis Supervisor

Accepted by .....

Professor  
Chairman, Thesis Committee



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Theoretical Overview . . . . .	2
1.1.1	Self-Coupling Higgs . . . . .	2
1.1.2	Warped Extra Dimensions . . . . .	2
	Kaluza-Klein Graviton . . . . .	2
	Radion . . . . .	2
1.1.3	Lorentz-Boosted Higgs Boson . . . . .	2
1.2	Overview of Analysis Strategy . . . . .	2
1.3	Review of Related Literature . . . . .	4
<b>2</b>	<b>Experimental Apparatus: the LHC and the CMS Detector</b>	<b>9</b>
2.1	Large Hadron Collider . . . . .	9
2.2	Compact Muon Solenoid . . . . .	11
2.2.1	Coordinate System . . . . .	13
2.2.2	Magnet System . . . . .	15
2.2.3	Inner Tracking System . . . . .	16
	Pixel Detector . . . . .	17
	Strip Tracker . . . . .	18
2.2.4	Electromagnetic Calorimeter . . . . .	19
2.2.5	Hadron Calorimeter . . . . .	21
2.2.6	Muon Detector System . . . . .	23
2.3	The Trigger System . . . . .	25
2.3.1	Level 1 Trigger System . . . . .	26
2.3.2	High Level Trigger System . . . . .	27

<b>3 Analysis Strategy</b>	<b>29</b>
3.1 Physical Object Reconstruction . . . . .	29
3.1.1 Jet Reconstruction . . . . .	30
Jet Clustering . . . . .	31
Jet Grooming . . . . .	33
3.2 b-Tagging . . . . .	34
3.2.1 “Double-b Tagger” Algorithm . . . . .	35
3.3 Data and Simulated Samples . . . . .	38
3.3.1 Data Samples . . . . .	38
3.3.2 Trigger . . . . .	38
3.3.3 Simulated Samples . . . . .	39
3.4 Event Selection . . . . .	43
3.4.1 Jet Kinematics Selection . . . . .	43
3.4.2 Higgs Mass Selection . . . . .	43
3.4.3 Dijet Selection . . . . .	44
3.4.4 N-Subjettiness Selection . . . . .	45
3.4.5 Tagging Selection . . . . .	45
3.4.6 Selection Distribution . . . . .	46
3.5 Data and MC Comparison . . . . .	47
<b>4 Background Estimation</b>	<b>55</b>
4.1 “ALPHABET” Method . . . . .	55
4.1.1 Double-b Tagger as the Tagging Variable . . . . .	58
4.1.2 Result . . . . .	58
<b>5 Systematic Uncertainties</b>	<b>61</b>
5.1 Systematics on Signal Efficiency . . . . .	61
5.2 Summary of the Systematic Uncertainties . . . . .	67

<b>6 Results and Discussion</b>	<b>69</b>
6.1 Asymptotic CL <sub>S</sub> Method . . . . .	69
6.2 Limit Plot . . . . .	69
<b>7 Summary, Conclusion, and Outlook</b>	<b>71</b>
7.1 Summary . . . . .	71
7.2 Conclusion . . . . .	71
7.3 Outlook . . . . .	71
<b>Bibliography</b>	<b>73</b>
<b>A Comparison of Different Heavy Resonance Models</b>	<b>83</b>
<b>B Study of Neutrinos inside AK8 b-jets</b>	<b>87</b>



# List of Figures

1.1	Standard Model . . . . .	2
1.2	Feynman diagram of $X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ . . . . .	3
1.3	Higgs branching ratio . . . . .	4
2.1	Overview of the LHC . . . . .	10
2.2	The CERN accelerator complex . . . . .	11
2.3	2015 proton run performance . . . . .	12
2.4	The CMS detector . . . . .	13
2.5	CMS coordinate system . . . . .	14
2.6	CMS tracker system . . . . .	17
2.7	CMS ECAL system . . . . .	20
2.8	CMS HCAL system . . . . .	22
2.9	CMS muon detector system . . . . .	24
2.10	Flow chart of CMS trigger system . . . . .	26
2.11	Architecture of CMS L1 trigger system . . . . .	27
3.1	Jet image from pp collision . . . . .	31
3.2	Jet clustering algorithms . . . . .	32
3.3	Schematic of $b$ -tagging algorithms . . . . .	36
3.4	ROC curve of $b$ -tagging algorithms . . . . .	38
3.5	Trigger efficiency of HLT_PFHT800 as a function of dijet mass . .	40
3.6	Signal and background for $p_T$ , corrected pruned mass, and $\eta$ variables . . . . .	49

3.7	Signal and background for $\tau_{21}$ , double-b tagger, reduced mass, and $\Delta\eta$ variables . . . . .	50
3.8	Acceptance $\times$ Efficiency of signal samples . . . . .	51
3.9	Data and MC comparison for $p_T$ , corrected pruned mass, and $\eta$ variables . . . . .	52
3.10	Data and MC comparison for $\tau_{21}$ , double-b tagger, reduced mass, and $\Delta\eta$ variables . . . . .	53
4.1	Schematic of Alphabet method . . . . .	56
4.2	Dependence of the $\tau_{21}$ on the jet pruned mass . . . . .	57
4.3	Dependence of the double-b tagger on the jet pruned mass . . . . .	57
4.4	Conversion rate and predicted background . . . . .	59
5.1	Trigger efficiency of HLT_PFHT800 as a function of reduced dijet mass . . . . .	65
5.2	Systematic uncertainties on background estimation . . . . .	66
A.1	$p_Z$ , rapidity, and $\cos(\theta^*)$ of heavy resonance models . . . . .	85
A.2	$p_T(H)$ , rapidity of H, and $\Delta\eta_{HH}$ of heavy resonance models . . . . .	86
B.1	Pruned and Soft drop mass . . . . .	88
B.2	Average vector-summed neutrino $p_T$ as a function of pruned mass	89
B.3	Average vector-summed neutrino $p_T$ as a function of jet $p_T$ . . . . .	90

# List of Tables

2.1	Parameters of the CMS magnet. . . . .	16
3.1	2015 dataset . . . . .	39
3.2	Bulk Graviton Monte Carlo samples . . . . .	41
3.3	Radion Monte Carlo samples . . . . .	41
3.4	QCD background Monte Carlo samples . . . . .	42
3.5	Cross sections $\times$ branching ratio of the signal MC samples . . . . .	42
3.6	Tight PF jet identification criteria . . . . .	44
3.7	Summary of event selection criteria used . . . . .	46
3.8	Bulk Graviton event selection cutflow efficiency . . . . .	47
3.9	Radion event selection cutflow efficiency . . . . .	47
5.1	Medium double-b tagger selection data/MC efficiency ratio . . .	65
5.2	Summary of systematic uncertainties . . . . .	67
A.1	Different heavy resonance models . . . . .	83
A.2	Acceptance of fiducial selections for heavy resonance models . . .	84



# Chapter 1

## Introduction

The **standard model** (SM), or colloquially known as the “periodic table of elements for particle physics”, is perhaps the most accurate illustration of numerous theoretical and experimental accomplishments of particle physicists around the world over the past decades. It consists of **quarks** and **leptons**, the building blocks of matter, and its interactions are defined by the exchange of force carriers: the **photons** for electromagnetic interactions, the **W and Z bosons** for weak interactions, and the **gluons** for strong interactions (Figure 1.1). Moreover, the recent discovery of the **Higgs boson** (H) [1–3] at the Large Hadron Collider (LHC) (Section 2.1) has been a great addition to the SM by generating the masses of all elementary particles and breaking the electroweak symmetry [4].

In spite of its achievements, the SM is not yet a perfect model. It still does not yet include one of the four fundamental forces, the gravitational force, and does not provide sufficient explanations for phenomena such as the expansion of the universe, matter-antimatter asymmetry, and the dark matter among others [5]. Also, with the discovery of the H boson, the search for new physics **beyond the SM** (BSM) has become more vivid.

This chapter briefly discusses the theory and strategy of the analysis. Previous studies that are related to the analysis is also shortly reviewed.

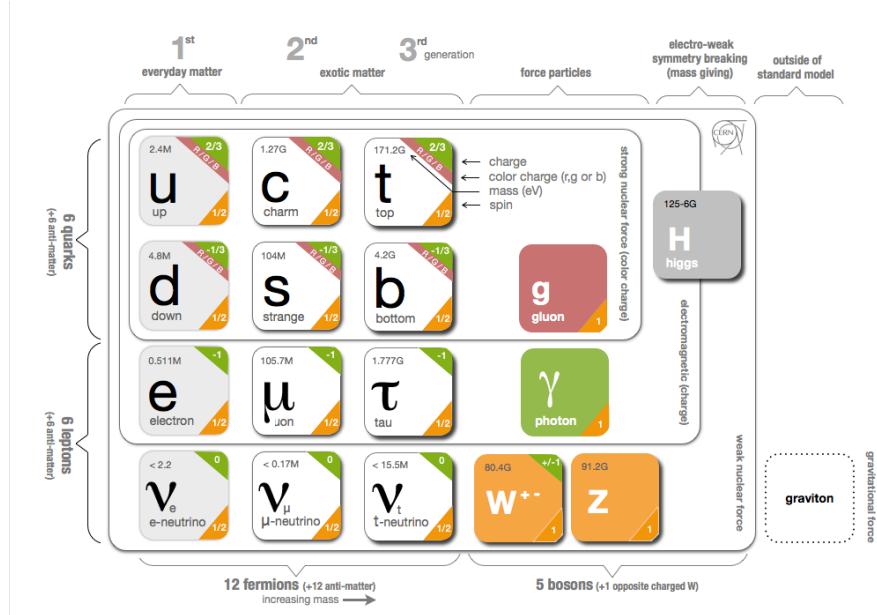


FIGURE 1.1: Standard Model. Figure from Reference [6].

## 1.1 Theoretical Overview

### 1.1.1 Self-Coupling Higgs

### 1.1.2 Warped Extra Dimensions

#### Kaluza-Klein Graviton

#### Radion

### 1.1.3 Lorentz-Boosted Higgs Boson

## 1.2 Overview of Analysis Strategy

The main objective of the analysis is to look for heavy resonances (i.e. Kaluza-Klein Graviton and Radion) that decay into two H bosons with large momentum. Each H boson decays further into a b quark-antiquark pair ( $b\bar{b}$ ) as its final state (See Figure 1.2). Since the resonances are massive ( $\geq 1$  TeV), the H bosons

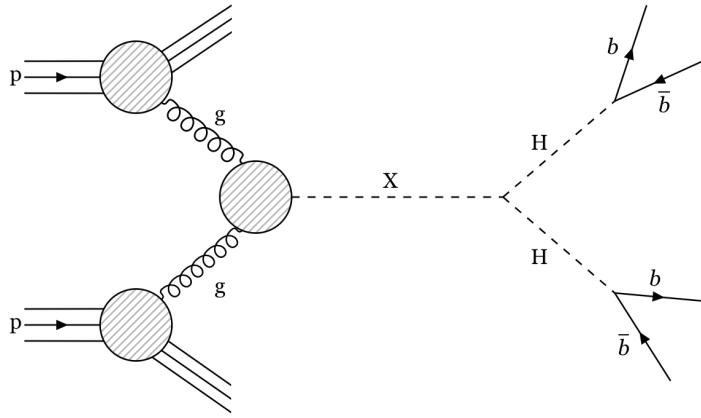


FIGURE 1.2: Feynman diagram of the production of heavy resonance  $X$  that decays into a pair of  $H$  bosons. Each  $H$  boson decays further into  $b\bar{b}$  pair final state.

will be highly Lorentz-boosted. The  $b\bar{b}$  pair will then have a small angular separation that can be reconstructed as one massive jet which corresponds to the  $H$  mass (125 GeV).

The  $H \rightarrow b\bar{b}$  decay is an interesting channel to study because of its large **branching ratio\*** (BR) with a value of 0.5824 at  $M_H = 125$  GeV [7] (See Figure 1.3). However, the signal is also buried within a large quantum chromodynamics (QCD) background ( $\sigma_{bb}(\text{QCD}) \sim 10^7$ ).

A new tagging algorithm, named **double-b tagger** (Section 3.2.1), is used in this analysis that aims to reconstruct the  $b\bar{b}$  pairs inside  $H$  jet. **N-subjettiness** ( $\tau_N$ ) (Section 3.4.4) was also used to measure the degree of jet constituents on how it can be organized as  $N$  subjets. Furthermore, the **Alphabet** method (Section 4), an extension of the  $ABCD$  method, is used as a technique to estimate the background.

The data used in the analysis are collected in proton-proton (pp) collisions at the LHC at center-of-mass energy  $\sqrt{s} = 13$  TeV that corresponds to a  $2.7 \text{ fb}^{-1}$  **integrated luminosity†** during the Run-II in 2015.

\*Branching ratio, or branching fraction, is the probability of a particle decays into one particular final state.

†Integrated luminosity is the instantaneous luminosity integrated over the given time interval,  $L_{int} = \int L dt$ . It directly relates to the number of observed events. The instantaneous

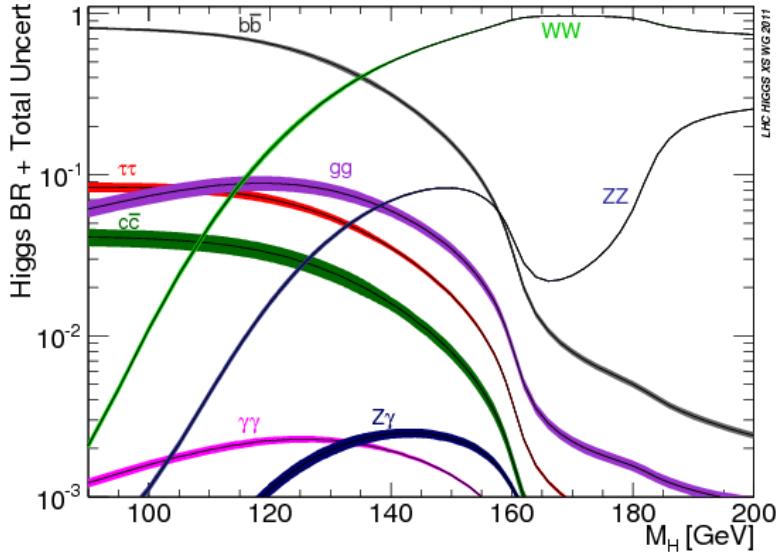


FIGURE 1.3: Branching ratio with uncertainties of Higgs boson for low mass range. Figure from Reference [8].

### 1.3 Review of Related Literature

During the Run-I of the LHC, similar analyses for the search of resonant particle X and non-resonant HH decay channel were done by the ATLAS and CMS collaborations in the pp collisions at  $\sqrt{s} = 8$  TeV.

#### ATLAS Collaboration

In the ATLAS collaboration, a  $\gamma\gamma b\bar{b}$  [9] final state was studied using  $20 \text{ fb}^{-1}$  of pp collisions for both resonant and non-resonant HH production. Jets from the hadronization of  $b$ -quarks ( $b$  jets) are selected using a multivariate tagging algorithm (MV1) [10] that uses a neural network to combine the information gathered in the event reconstruction, and is built to have an efficiency of 70% for tagging  $b$  jets with light-quark or gluon jet and charm-jet rejection, as determined in a MC sample  $t\bar{t}$  events. A 95% confidence level (CL) observed limit on the cross section was set at 2.2 pb for non-resonant production. For production via narrow resonances, the observed exclusion was set from 0.7 to 3.5 pb.

---

luminosity is the proportionality factor between the number of events per second  $dR/dt$  and the cross section  $\sigma_p$ ,  $\frac{dR}{dt} = L \cdot \sigma_p$

A  $b\bar{b}b\bar{b}$  [11] final state was also studied with  $19.5 \text{ fb}^{-1}$  integrated luminosity but no evidence neither for resonant nor non-resonant HH production was observed. It also uses MV1 tagging algorithm to identify the  $b$  jets. An upper limit on the cross section of 3.2 and 2.3 fb was set for Kaluza-Klein Graviton ( $G_{KK}^*$ ) with a mass of 1.0 and 1.5 TeV, respectively, at 95% CL. While an upper limit on the cross section of 202 fb was set for non-resonant SM HH production at 95% CL as well.

Searches for resonant and non-resonant HH production in  $bb\tau\tau$  and  $\gamma\gamma WW^*$  [12] final states using  $20.3 \text{ fb}^{-1}$  of pp collision data were done as well. No significant data excess was observed above the estimated background for both final states. These results were then combined from the  $\gamma\gamma b\bar{b}$  and  $b\bar{b}b\bar{b}$  final states analyses to obtain a 0.69 pb observed limit for non-resonant HH production at 95% CL. The observed limits ranged from 2.1 pb at 260 GeV to 0.011 pb at 1000 GeV for the narrow resonance search.

Recently, the ATLAS collaboration searched for HH production in the  $b\bar{b}b\bar{b}$  final state with  $3.2 \text{ fb}^{-1}$  of pp collision data at  $\sqrt{s} = 13 \text{ TeV}$  [13]. The analysis was done in two ways: Resolved analysis for low mass range (400-1500 GeV) that aims at event topologies with 4  $b$ -tagged jets, and Boosted analysis with mass above 1500 GeV that targets topologies with Lorentz-boosted H bosons resulting in merged jets. The results showed no significant excess was observed in the data beyond the background expectation. The background used in this analysis consists mainly of multi-jets and  $t\bar{t}$  events. 95% CL upper limits in the 24 to 113 fb range were observed for Kaluza-Klein gravitons within Randall-Sundrum model with a mass range of 600 to 3000 GeV. 1.22 pb 95% CL upper limit was observed for non-resonant production with mass above 1350 GeV.

## CMS Collaboration

In the CMS collaboration, a heavy scalar  $\mathcal{H}$  that decays into a pair of SM-like H bosons [14] was studied from a dataset corresponding to an integrated luminosity of  $19.5 \text{ fb}^{-1}$ . Multileptons ( $WW^*\tau\tau$ ,  $ZZ^*\tau\tau$ , and  $\tau\tau\tau\tau$  among other combinations) and lepton+diphoton ( $\gamma\gamma WW^*$ ,  $\gamma\gamma ZZ^*$ , and  $\gamma\gamma\tau\tau$ ) were considered as the final states of this analysis. No significant deviation from the SM expectation was found. The 95% CL upper limit was placed approximately on  $2 \text{ pb}$ .

In addition, a heavy scalar  $\mathcal{H}$  that decays into a pair of SM-like H bosons was studied but the final state consists of two  $\tau$  leptons and two  $b$  jets [15] from a dataset corresponding to an integrated luminosity of  $19.7 \text{ fb}^{-1}$ . Combined secondary vertex (CSV)  $b$  tagging algorithm [16] are used to identify the  $b$  jets. It exploits the information on the decay vertices of long-lived mesons and the transverse impact parameter measurements of charged particles. In the end, no significant excess is observed over the background expectation. The upper limit was set on the  $\mathcal{H}$  mass in the range of  $260\text{-}350 \text{ GeV}$ .

Resonant and non-resonant searches for particles that decay into HH pair with a  $\gamma\gamma b\bar{b}$  [17] final state were done in data corresponding to an integrated luminosity of  $19.7 \text{ fb}^{-1}$ . This analysis also uses the CSV  $b$  tagging algorithm for  $b$  jets identification. To improve the signal sensitivity, the events were classified into two categories: medium purity where each event has exactly 1  $b$ -tagged jet, and high purity where each event has more than 1  $b$ -tagged jet. The results showed no evidence for a signal had been found above the background expectation for both resonant and non-resonant search. The mass range between  $325\text{-}450 \text{ GeV}$  was excluded for the first Kaluza-Klein excitation mode of the graviton in the RS1 Randall-Sundrum model. For the non-resonant search, an upper limit of  $0.71 \text{ pb}$  was set in its cross section.

The resonant HH production in the  $b\bar{b}b\bar{b}$  final state was done in two ways:

one analysis was done in low mass region (270-1100 GeV) [18] with data corresponding to an integrated luminosity of  $17.9 \text{ fb}^{-1}$  and the other was done in high mass region (1150-3000 GeV) [19] with data corresponding to an integrated luminosity of  $19.7 \text{ fb}^{-1}$ . For the low mass region analysis, it uses a  $b$  tagging algorithm called combine multivariate (CMVA) algorithm that combines the results of three tagging algorithm: the jet probability (JP) algorithm, two soft lepton tagging algorithm, and the CSV algorithm. No evidence for a signal was observed. An exclusion was done in the 300-1100 GeV mass range of Radion and 380-830 GeV mass range of Kaluza-Klein Graviton at a 95% CL. For the high mass region analysis, the massive resonance causes the HH pair to be highly Lorentz-boosted. By this means, the  $b\bar{b}$  pair can be reconstructed as one massive jet. The CSV algorithm and  $\tau_N$  were used in this analysis. Again, no evidence for a signal was observed. Radion scalar masses below 1550 GeV were excluded for with scale  $\Lambda_R = 1 \text{ TeV}$ .



# Chapter 2

## Experimental Apparatus: the LHC and the CMS Detector

The data used in this analysis are collected by the **Compact Muon Solenoid** (CMS) detector at the **Large Hadron Collider** (LHC). The CMS is one of the general-purpose detectors at the LHC which aims to search for extra dimensions, dark matter, and etc [20]. In this chapter, a brief overview of the LHC and the CMS detector is presented.

### 2.1 Large Hadron Collider

Built by the **European Organization for Nuclear Research** (also known as “CERN”; derived from *Conseil Européen pour la Recherche Nucléaire*) from 1998 to 2008, the LHC is a particle collider, 27 km in circumference, and lies between 50 m to 175 m beneath the ground of the France-Switzerland border (Figure 2.1). It is the largest and most powerful proton-proton collider in the world. It is also a collaboration of more than 10,000 scientists and engineers coming from different universities and research centers all over the globe [21, 22].

The experiment in the LHC starts with hydrogen gas that is injected into the source chamber of the **Linear accelerator 2** (LINAC2), in order to remove its electrons. The bunch of protons left is accelerated with an electric field in the LINAC2 to reach its energy of 50 MeV. The protons are then fed to the **Proton**

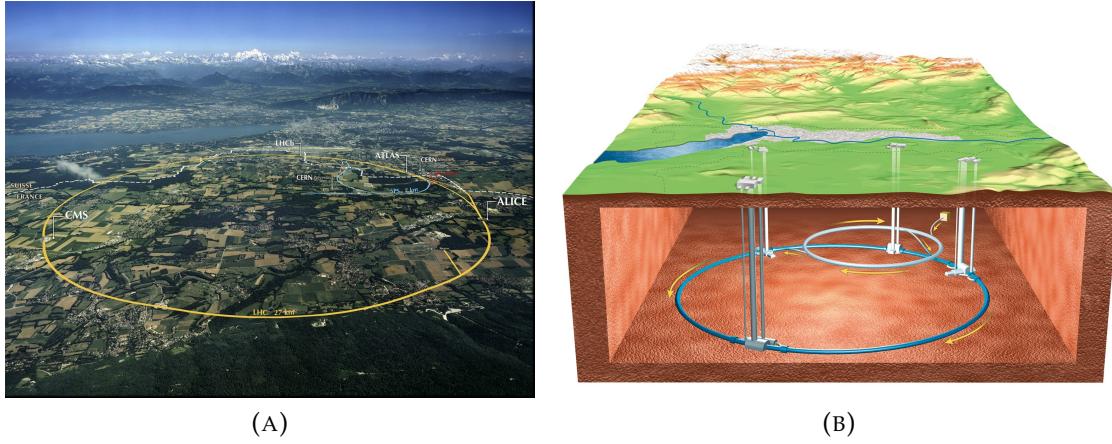


FIGURE 2.1: Overview of the LHC (Top view (A) and underground sketch (B)) at the Swiss-French Border from Reference [23].

**Synchrotron Booster** (PSB) to accelerate up to 1.4 GeV. After the PSB, it is flung on to the **Proton Synchrotron** (PS) where it is accelerated to 25 GeV. Packets of protons are then channeled to the **Super Proton Synchrotron** (SPS) which is specifically designed to receive the protons at this energy and further increase its energy up to 450 GeV. Finally, the SPS launches the protons into the **LHC**. There are two vacuum pipes in the LHC: one circulates the protons clockwise and the other counter-clockwise. A total of 2808 packets of protons is launched by the SPS in the LHC. The LHC boosts the energy of each proton to reach 6.5 TeV [24].

There are four **intersection points** in the LHC, which correspond to the four main experiments; the “**A Large Ion Collider Experiment**” (ALICE), the “**Large Hadron Collider beauty**” (LHCb), the “**A Toriodial LHC ApparatuS**” (ATLAS), and the **CMS**. ALICE is a heavy-ion (Pb-Pb nuclei) detector that studies the physics of strongly interacting matter at extreme energy densities specifically the quark-gluon plasma, a phase of matter thought to have formed just after the big bang. LHCb uses the particle “beauty quark” or “b quark” to study the differences between matter and antimatter. ATLAS and CMS are the two general-purpose detectors at the LHC with same specific goals, to search for extra dimensions, dark matter, and etc., but both detectors have different scientific solutions and system design [20].

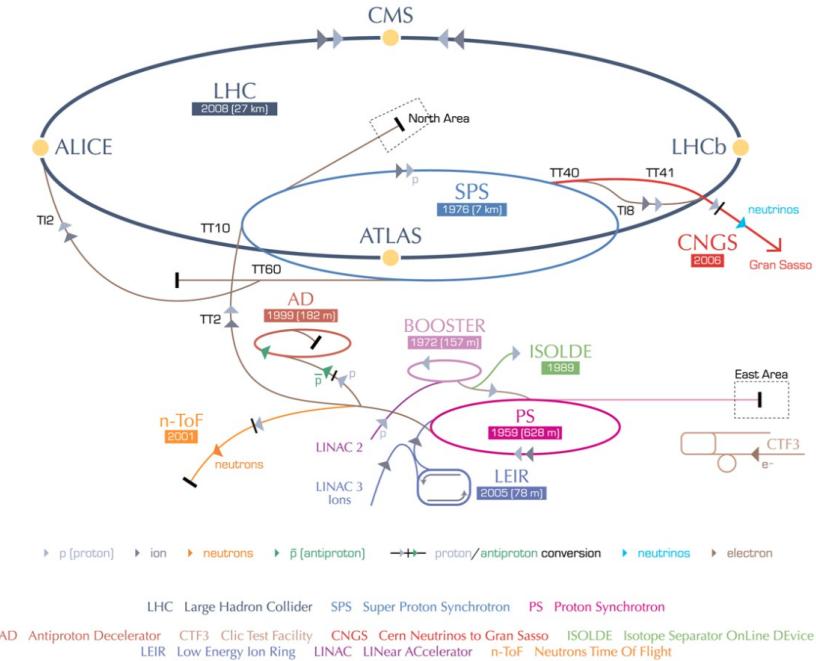


FIGURE 2.2: Schematic view of the CERN accelerator complex. The dark blue circle represents the LHC ring. The smaller machines are used in a chain to help boost the particles to their final energies and provide beams to a whole set of smaller experiments, which also aim to uncover the mysteries of the Universe.

Figure from Reference [25].

A center-of-mass energy of 13 TeV with 25 ns beam bunch spacing was attained during the Run-II of the LHC in 2015. It was reported that 2244 proton bunches per beam were providing an approximate of  $5.0 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$  peak luminosities and  $4 \text{ fb}^{-1}$  total delivered integrated luminosity in both CMS and ATLAS (see Figure 2.3) [26].

## 2.2 Compact Muon Solenoid

Located approximately 100 m beneath the ground of Cessy, France, the CMS is one of the multi-purpose particle detectors operating at the LHC and has multi-objectives such as studying SM, including the recently discovered Higgs Boson, and searching for extra dimensions and dark matter particles. Despite having the same objectives with the ATLAS detector, it has different scientific solutions and distinct designs such as a high-field solenoid, a full-silicon-based

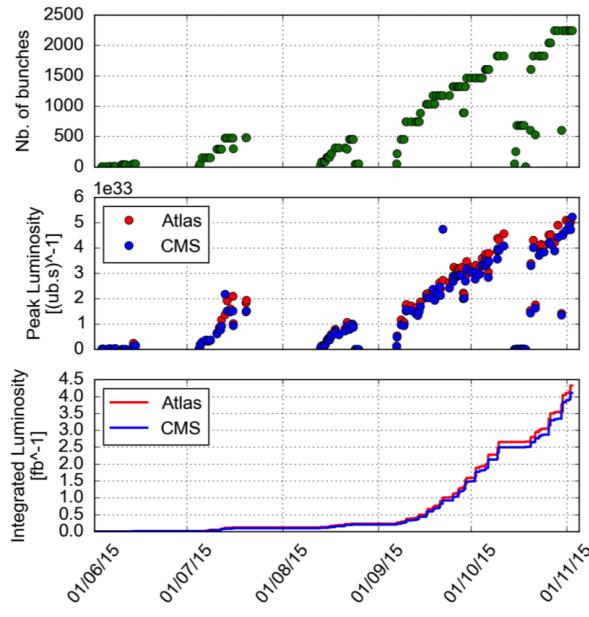


FIGURE 2.3: Progression summary of 2015 proton run performance for the total number of bunches, peak, and integrated luminosities. Figure from Reference [26].

inner tracking system, and a homogenous scintillating-crystals-based electromagnetic calorimeter [20, 27].

CMS satisfies these following detector requirements in order to achieve its objectives [27]:

- Good muon identification and momentum resolution over a wide range of momenta and angles, good dimuon mass resolution ( $\approx 1\%$  at 100 GeV), and the ability to determine unambiguously the charge of muons with  $p$  up to 1 TeV;
- Good charged-particle momentum resolution and reconstruction efficiency in the inner tracker. Efficient triggering and offline tagging of  $\tau$ 's and  $b$ -jets requires pixel detectors close to the interaction region;
- Good electromagnetic energy resolution, good diphoton and dielectron mass resolution ( $\approx 1\%$  at 100 GeV), wide geometric coverage,  $\pi^0$  rejection, and efficient photon and lepton isolation at high luminosities;

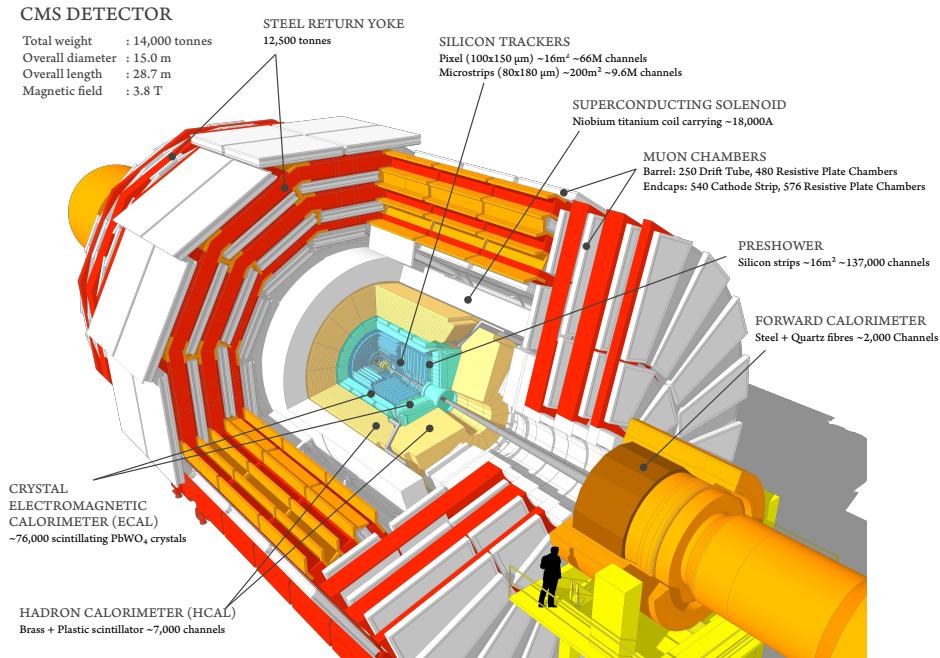


FIGURE 2.4: Cutaway view of the CMS detector from Reference [28].

- Good missing transverse momentum and dijet-mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

The most distinct feature of the CMS is the 13 m long, 6 m inner-diameter, 3.8 T **superconducting solenoid** (Section 2.2.2) located at the heart of the detector. From the inner to the outer layer of the magnet’s bore are the **silicon-based tracker** (Section 2.2.3), the **electromagnetic calorimeter** (ECAL) (Section 2.2.4), and the **hadron calorimeter** (HCAL) (Section 2.2.5). Outside of the magnet are the **muon chambers** (Section 2.2.6) [27].

The overall dimension of the CMS detector is 28.7 m long, 15 m diameter, and the weight is about 14,000 t [28].

### 2.2.1 Coordinate System

The origin of the CMS detector’s coordinate system is the nominal collision point in the experiment. Its  $x$ -axis points towards the center of the LHC ring while its

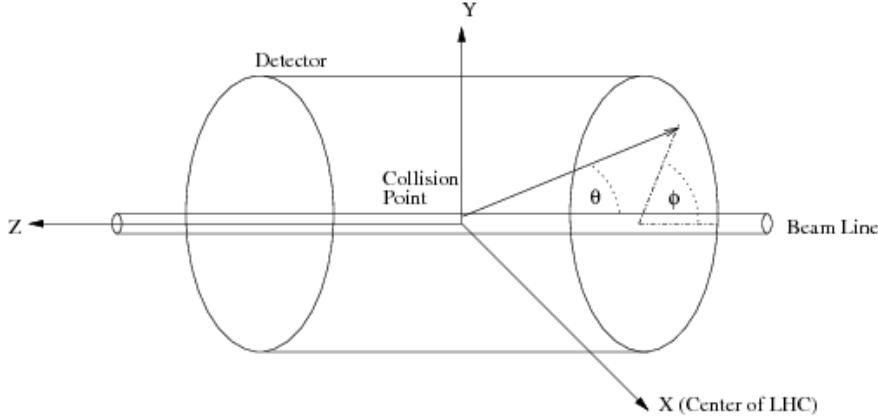


FIGURE 2.5: CMS coordinate system. The  $x$ -axis points towards the center of the LHC ring,  $y$ -axis points vertically upward, and  $z$ -axis points along the beam direction counter-clockwise. Figure from Reference [29].

$y$ -axis points vertically upward. Hence, its  $z$ -axis points along the beam direction counter-clockwise. In the  $x$ - $y$  plane, the radial coordinate is denoted by  $r$  and the azimuthal angle ( $\phi$ ) is measured from the  $x$ -axis (Figure 2.5) [27]. Conventionally, the spherical coordinate system is used due to the detector's axis of symmetry along the  $z$  axis.

Polar angle ( $\theta$ ) is the angle measured relative to the  $z$ -axis (beam axis). However, considering that the rest frame of a hard collision is likely to be boosted relative to the lab frame along the beam direction [30], its difference ( $\Delta\theta$ ) is not a good quantity to use since it is not Lorentz-invariant. Thus, **rapidity** ( $y$ ) [31], a Lorentz-invariant quantity for boosts along the beam direction, is used and defined as:

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) = \tanh^{-1} \left( \frac{p_z}{E} \right) \quad (2.1)$$

For massless particles (particle's mass is negligible when it is traveling near the speed of light which is comparably similar to the events happening inside the collider), rapidity is transformed to **pseudorapidity** ( $\eta$ ) [31] which is defined as:

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right] \quad (2.2)$$

With this, the **angular separation** between objects is defined as  $\Delta R \equiv \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$ , or  $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  if particles involved are massless. In addition, the **transverse momentum** ( $p_T$ ) is defined as the momentum computed from the  $x$  and  $y$  components, which is transverse to the beam axis, and the **transverse energy** ( $E_T$ ) is defined as  $E_T = E \sin\theta$  [27, 29].

### 2.2.2 Magnet System

The 13 m long, 6 m inner-diameter, 3.8 T superconducting solenoid of the CMS weighs 220 t cold mass and can provide an enormous bending power\* of 12 Tm before the muon system measures the muon bending angle. It has a 2.6 GJ stored energy at full current. To support its generation of 3.8 T field (by providing 41.7 MA-turn), the winding is composed of 4 layers, instead of the usual 1 or 2 layers, that is made from an augmented Niobium Titanium (NbTi) conductor [27]. A detailed description of the CMS magnet's parameter is in Table 2.1.

Its main objective is to bend the particles' direction coming from the high-energy collision inside the LHC. The particle's direction is less bent by the magnetic field when it has higher momentum. This anticorrelation between the particle's momentum and bending expresses that its momentum can be measured by tracking its trajectory. Thus, CMS built the most powerful magnet possible to have a highly accurate measurement of particles' momentum, as well as particles in high energies [32].

The magnetic flux is returned within the detector by the 10,000 t steel **yoke**. It is composed of 2 endcaps, having three disks each, and 5 barrel wheels that weigh from 400 t for the lightest to 1920 t for the central wheel, including the coil and its cryostat. It also acts as a filter that only allows muons and weakly interacting particles (e.g. neutrinos) to pass through [27, 32].

---

\*Bending power (or “magnetic rigidity”) is characterized by the field integral  $\int B \times dl$ , where  $B$  is the azimuthal field component, and the integral is taken over the particle's straight-line trajectory without the field, from the inner to the outer radius of the toroid. Its unit is in Tesla-meter (Tm).

TABLE 2.1: Parameters of the CMS magnet.

General Parameters	
Magnetic Length	12.5 m
Cold bore diameter	6.3 m
Central magnetic induction	4 T
Total Ampere-turns	41.7 MA-turns
Nominal current	19.14 kA
Inductance	14.2 H
Stored Energy	2.6 GJ
Cold mass	
Layout	Five modules mechanically and electrically coupled
Radial thickness of cold mass	312 mm
Radiation thickness of cold mass	$3.9 X_0$
Weight of cold mass	220 t
Maximum induction on conductor	4.6 T
Temperature margin wrt operating temperature	1.8 K
Stored energy/unit cold mass	11.6 kJ/kg
Iron yoke	
Outer diameter of the iron flats	14 m
Length of barrel	13 m
Thickness of the iron layers in barrel	300, 630, and 630 mm
Mass of iron in barrel	6000 t
Thickness of iron disks in endcaps	250, 600, and 600 mm
Mass of iron in each endcap	2000 t
Total mass of iron in return yoke	10,000 t

### 2.2.3 Inner Tracking System

Following the requirements of the LHC physics program, the inner tracking system of the CMS is capable of providing an accurate and efficient reconstruction of the paths of charged particles with  $p_T > 1$  GeV within  $|\eta| < 2.5$  range. It can also give an accurate measurement of secondary vertices and impact parameters that are essential for effective heavy flavors identification. Its dimension is 5.8 m long and 2.5 m in diameter enclosing the intersection point. The silicon-based design is due to the requirement that the system must have high granularity and quick response in a way that the paths are identified consistently and the bunch

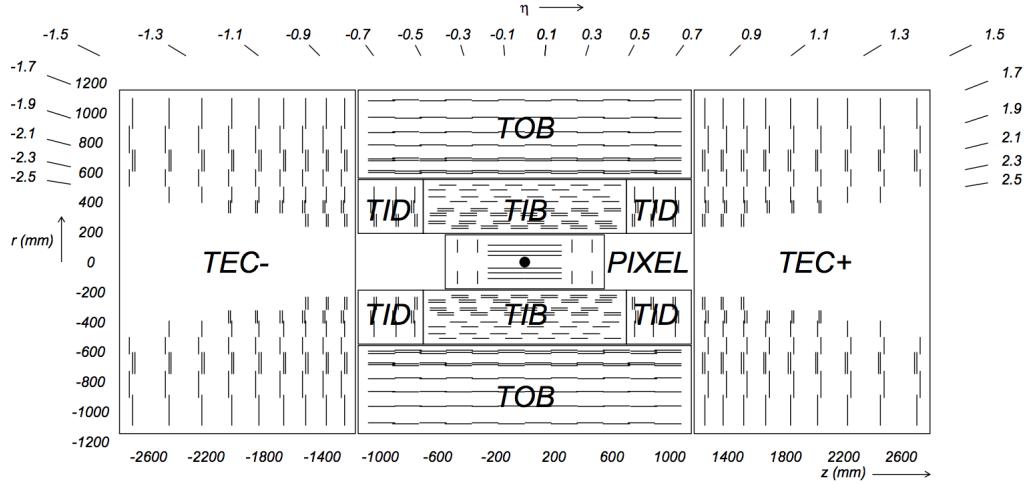


FIGURE 2.6: Schematic cross sectional view of the CMS tracker system. Each line represents a detector module, double lines means back-to-back module. Figure from Reference [27].

crossing are attributed precisely without overheating the system and causing severe radiation damage by the intense particle flux [27].

There are two kinds of tracker subsystem in the CMS tracker; a **pixel detector** and a **strip tracker**. The pixel detector has three barrel layers occupying a radial distance from 4.4 to 10.2 cm while the strip tracker has ten barrel detection layers with a radial coverage from 25.5 to 108 cm. Both subsystems are finished off by an endcap; two disks for pixel detector and 3 plus 9 disks for strip tracker on each barrel's side. This extends the tracker's acceptance up to  $|\eta| < 2.5$ . Overall, the CMS tracker is composed of 1440 pixel and 15,418 strip detector modules. The active silicon area of the CMS tracker is about  $200 \text{ m}^2$  [27].

### Pixel Detector

Being the closest to the interaction region, the pixel subsystem provides precise tracking points in the transverse ( $r - \phi$ ) and longitudinal ( $z$ ) plane; thus it is responsible for the small impact parameter resolution for accurate reconstruction of secondary vertices from  $b$  and  $\tau$  decays, and formation of seed tracks for the outer track reconstruction and high-level triggering. It covers a range of  $|\eta| < 2.5$  to match the acceptance of the central tracker [27].

The dimension of the pixel cell is  $100 \times 150 \mu\text{m}^2$ . It uses a concept called “n-in-n” where the pixels contain n-implants placed in the n-substrate while the back side is p-doped. This concept was chosen, instead of the usual single sided “p-in-n”, because: (1) a high signal charge is guaranteed at moderated bias voltages ( $< 600 \text{ V}$ ) after high hadron fluences, (2) the electrons have high mobility in which the Lorentz angle increases thus satisfying the spatial resolution requirement, and (3) the necessity of constructing the back leads to the implementation of the guard ring system which keeps all sensor edges at ground potential [33].

The 53 cm long three barrel layers (BPix) contain 48 million pixels with a total area of  $0.78 \text{ m}^2$  while the two 6-15 cm radius endcap disks (FPix) contain 18 million pixels covering a total area of  $0.28 \text{ m}^2$  [27].

### Strip Tracker

The strip tracker subsystem, which encloses the pixel detector subsystem, is built with 24,244 single-sided “p-on-n” type microstrip sensors. A total of 9.3 million strip trackers that covers a total active area of  $198 \text{ m}^2$  is used in this subsystem. The sensors are manufactured on 6-inch wafers in a standard planar process having an n-doped float zone silicon with  $\langle 100 \rangle$  crystal orientation. There are two sets of sensors made: (1) the “thin” one with a  $320 \mu\text{m}$  wafer thickness and a  $1.5 - 3.25 \text{ K}\Omega\text{cm}$  substrate resistivity, and (2) the “thick” one with a  $500 \mu\text{m}$  wafer thickness and a  $4 - 8 \text{ K}\Omega\text{cm}$  substrate resistivity [34].

The subsystem can be divided into quarters; the four layers **Tracker Inner Barrel** (TIB), the six layers **Tracker Outer Barrel**, the **Tracker Inner Endcap** (TID) that is made of three small disks of the inner barrel on both side, and the **Tracker Outer Endcap** (TEC) which is composed of nine big disks on both sides of the tracker. The TIB and TID have a combined radius of 55 cm and delivers up to  $4 r - \phi$  measurements on a trajectory using “thin” sensors, parallel to the beam axis for TIB and radial for TID. The TOB has a radius of 116 cm and uses “thick” sensors. It provides  $6 r - \phi$  measurements per trajectory. Lastly, the TEC covers

the region  $124 \text{ cm} < |z| < 282 \text{ cm}$  and  $22.5 \text{ cm} < |r| < 113.5 \text{ cm}$ . It carries up to 7 rings of microstrip sensors (the inner 4 rings use “thin” sensors and the outer 3 rings use “thick” sensors) and delivers  $9 r - \phi$  measurements per trajectory [27, 34].

### 2.2.4 Electromagnetic Calorimeter

The ECAL system is situated right outside the inner tracking system. Its objective is to measure the EM particles’ energy such as electrons and photons. The main material of the ECAL is the lead tungstate ( $\text{PbWO}_4$ ) crystals; 61,200 mounted in the barrel and 7,324 in both of the endcaps.  $\text{PbWO}_4$  crystal is the chosen material for it is highly dense ( $8.28 \text{ g/cm}^3$ ), has short radiation length<sup>†</sup> ( $0.89 \text{ cm}$ ), and small Molière radius<sup>‡</sup> ( $2.2 \text{ cm}$ ). The ECAL’s high-density crystals design makes it radiation resistive, have a quick temporal response ( $\sim 10 \text{ ns}$ ), and provide fine granularity. Being a homogeneous crystal calorimeter, it also provides a good energy resolution which enhances the search for  $H \rightarrow \gamma\gamma$  [27, 35].

The photodetectors used are avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps. The **ECAL barrel** (EB) covers a range of  $|\eta| < 1.479$  with a granularity of 360-fold in  $\phi$  and  $(2 \times 85)$ -fold in  $\eta$ . Its crystals have a front face cross section of  $22 \times 22 \text{ mm}^2$  ( $0.0174 \times 0.0174 \text{ in } \eta - \phi$ ), and  $26 \times 26 \text{ mm}^2$  at the rear face. The centers of the crystals’ front face are at a radius  $1.29 \text{ m}$ . The crystal has a length of  $230 \text{ mm}$  ( $25.8 X_0$ ). Overall, the barrel crystal’s dimension is  $8.14 \text{ m}^3$  in volume and  $67.4 \text{ t}$  in weight. The **ECAL endcap** (EE) covers a range of  $1.479 < |\eta| < 3.0$ . It is  $315.4 \text{ cm}$  away from the interaction point longitudinally, considering the estimated  $1.6 \text{ cm}$  shift towards the interaction point when the magnetic field is turned on. Both endcaps are divided into 2 halves, or *Dees*, with each holding 3662 crystals. The crystals in the EE are

---

<sup>†</sup>Radiation length  $X_0$  is the mean distance traveled by the relativistic charged particle in a material over which its energy decreases to  $1/e$  of its initial value by bremsstrahlung.

<sup>‡</sup>Molière radius  $R_M$  is the distance in the transverse plane of an incoming particle that contains  $\geq 90\%$  of the shower’s energy deposition.

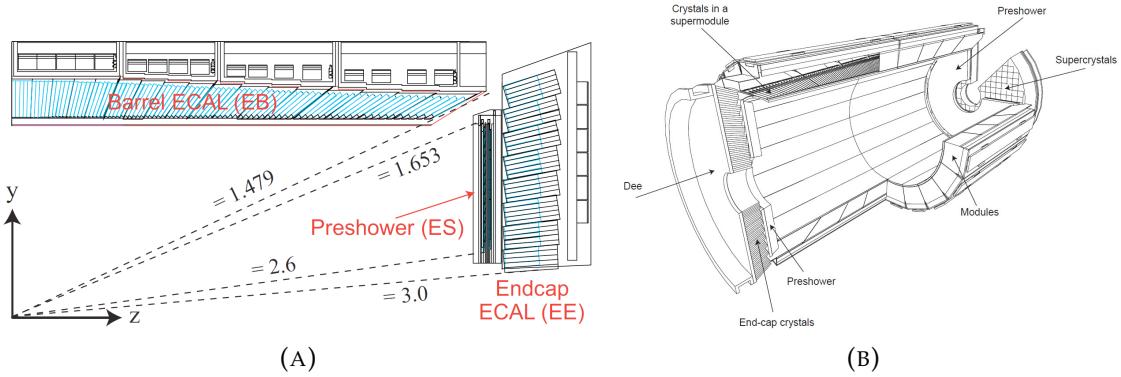


FIGURE 2.7: CMS electromagnetic calorimeter; (A) One-quarter schematic cross sectional view (dashed lines represent fixed  $\eta$  values) from Reference [36], and (B) Layout of the crystal modules, supermodules, and endcaps, with the preshower in front, in the CMS ECAL from Reference [27].

identically shaped and grouped in  $5 \times 5$  crystals (also known as “supercrystals”, or SCs) with a length of 220 mm ( $24.7 X_0$ ). It has a front face cross section of  $28.62 \times 28.62$  mm $^2$ , and  $30 \times 30$  mm $^2$  at the rear face. The dimension of the EE crystal is 2.90 m $^3$  in volume and 24.0 t in weight [27, 35].

In front of the EE crystals, there is a **preshower** detector installed within the range of  $1.653 < |\eta| < 2.6$ . Its objectives are to filter neutral pions ( $\pi^0$ ), to identify electrons from minimum ionizing particles, and to enhance the electron’s and photon’s position determination with high granularity. It is made up of silicon strip sensor with a lead absorber, which initiates the electromagnetic showers, in front. Its total thickness is 20 cm [27].

The energy resolution [37] of the calorimeter can be parameterized as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (2.3)$$

where  $S$  is the stochastic term, which depends on fluctuation in showering, photo-statistics and photodetector gain;  $N$  the noise term, which depends on electronic, digitization, and pile-up noise; and  $C$  is the constant term that depends on leakage of energy, calibration, and non-uniformity of the longitudinal light collection.

### 2.2.5 Hadron Calorimeter

The HCAL system is placed mainly at the outer extent of the ECAL system ( $R = 1.77$  m) and the inner extent of the superconducting solenoid ( $R = 2.95$  m). It measures the hadron jets and neutrinos or exotic particles that result to the apparent missing transverse momentum<sup>§</sup> ( $p_T^{miss}$ ). It is divided into four distinct subsystems; the **HCAL barrel** (HB), the **HCAL endcap** (HE), the **HCAL outer** (HO), and the **HCAL forward** (HF) [27, 38].

The HB is placed inside the superconducting solenoid that covers a range of  $|\eta| < 1.3$ . It consists of 36 azimuthal wedges that divide the barrel into HB+ and HB-. These wedges are made out of flat brass absorber plates which are aligned parallel to the beam axis. The main material used for the absorber is brass; eight layers of 50.5 mm thick plates and six layers of 56.5 mm thick plates. This brass plates are in the middle of two steel plate; 40 mm thick in front and 75 mm thick at the back. The total absorber thickness is 5.82 interaction length<sup>¶</sup> ( $\lambda_l$ ) at  $90^\circ$ . A 3.7 mm thick Kuraray SCSN81 plastic scintillator is used as the active material for the HB [27, 38].

The HE is within the range of  $1.3 < |\eta| < 3$ , which is also inside the superconducting solenoid. Same with HB, it also uses brass as the main material for the absorber and scintillator as the active material; 79 mm thick brass disks interleaved with scintillator wedges, 9 mm thick Bicron BC408 for the first layer and 3.7 mm thick SCSN81 for the other layers. Combining HB and HE, the total length of the HCAL system is  $\sim 10 \lambda_l$ , including the EM crystals [27, 38].

The HO is placed outside the superconducting solenoid that complements the HB. This sequence is due to the constraint in the amount of material that can be put in to absorb the hadronic shower. The HO catches the energy leakage

---

<sup>§</sup>Missing transverse momentum  $p_T^{miss}$  is the imbalance in the momentum measured in the transverse plane that mainly comes from neutrinos (Neutrinos cannot be detected by the detectors).

<sup>¶</sup>Interaction length  $\lambda_l$  is the mean distance traveled by the hadron particles after which its incident number is reduced by a factor of  $1/e$  due to the nuclear/strong interactions.

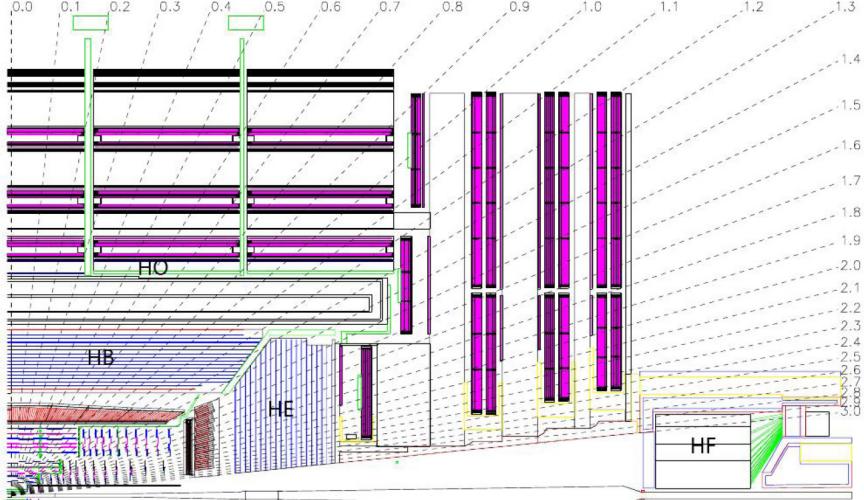


FIGURE 2.8: One-quarter longitudinal view of the CMS hadron calorimeter system showing the layout of its parts: the hadron barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeter. The dashed lines represent fixed  $\eta$  values.

Figure from Reference [27].

from the HB where  $\sim 5\%$  of all hadron's energy above 100 GeV are deposited. In the  $\phi$  plane, it has a 12-fold structure, with  $30^\circ$  component each that are divided into six  $5^\circ$  sectors. In the  $\eta$  plane, it consists of five 19.5 cm thick iron "rings"; Ring 0 covers a range of  $|\eta| < 0.35$ , Rings  $\pm 1$  cover  $0.35 < |\eta| < 0.87$  range, and Rings  $\pm 2$  cover  $0.87 < |\eta| < 1.2$  range. Ring 0 has two layers of HO scintillators on either side at radial distance of 3.82 m and 4.07 m, while the others have single HO scintillator layer at radial distance of 4.07 m. The HO extends the depth of the HCAL system to a minimum of  $11.8 \lambda_l$  excluding the HB-HE boundary region [27, 38].

Lastly, the HF is situated 11.2 m away (along the  $z$ -direction) from the interaction point that is beyond  $|\eta| = 3.0$ . The HF extends the pseudorapidity range of the HCAL system up to  $|\eta| = 5.2$  using a Cherenkov-based, radiation-hard technology. It is made of quartz fiber and steel, with the fibers parallel to the beam axis. It is built in  $20^\circ$  wedges with each containing two  $10^\circ$   $\phi$  sectors [27, 38].

The energy resolution for the HCAL can also be parameterized using the same canonical formula (Equation 2.3), though, the parameters are larger with

usual values of 35 – 100% for  $S$ , 1 – 3 GeV for  $N$ , and 3 – 5% for  $C$  [39].

### 2.2.6 Muon Detector System

Muon is one of the most important particles in a particle detection experiment as it provides clean signatures of interesting physics processes within a large background rate. Hence, a robust and accurate muon measurement is one of the main goals of the CMS, as it is stated in the name of the detector. The muon detector system is placed at the outermost layer of the CMS. It has three primary functions: identification, momentum measurement, and triggering of the muon. Its design is capable of reconstructing the muon's momentum and charge within a large kinematic range of the LHC. It utilized three types of gas-ionization particle subdetector for muon identification: the **drift tube chambers** (DT) in the barrel, the **cathode strip chambers** (CSC) in the endcap, and the **resistive plate chambers** (RPC) in both barrel and endcap [27, 40].

The DT covers a range of  $|\eta| < 1.2$ . It is divided into 12  $\phi$ -segments per wheel creating 4 sections that are arranged in an alternating sequence with the plates of the magnet flux-return yoke. Each section has 8 layers of tubes that measure the coordinates of the muon in the transverse plane, and 4 layers in the longitudinal plane except for the last outer section. A drift cell is the basic element of the DT subsystem which has a transverse area of  $42 \times 13 \text{ mm}^2$  with a  $50 \mu\text{m}$  diameter gold-plated stainless-steel anode wire at the center. This corresponds to a drift time of 380 ns in a gas mixture of 85% Ar + 15% CO<sub>2</sub>. Four layers of cells in a staggered parallel order create a so called *superlayer* (SL). A DT chamber is made of 2 parallel SLs that measures the  $r - \phi$  coordinates, and an orthogonal SL that measures the  $r - z$  coordinates except for the last outer section [27, 40].

Higher muon rates and background levels, and strong non-uniformed magnetic field are the usual cases happening in the endcap regions of the CMS. The CSC is preferred in this kind of situation since it has a fast temporal response,

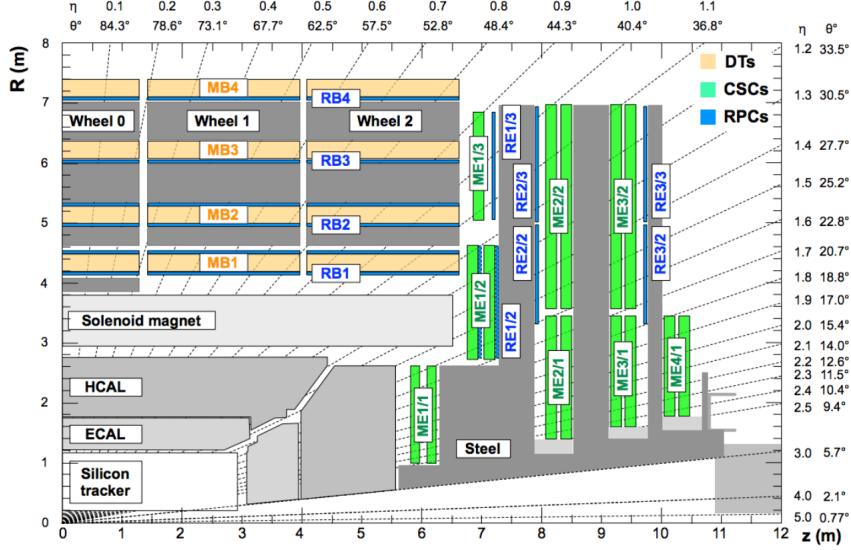


FIGURE 2.9: One-quarter cross sectional view of the CMS muon detector system. Dark grey areas are the various muon stations and steel disks, light orange areas are the drift tubes (DT) that are labeled muon barrel (MB), green areas are the cathode strip chambers (CSC) that are labeled muon endcap (ME), and blue areas are the resistive plate chambers (RPC; RB for barrel and RE for endcap). The dashed lines represent fixed  $\eta$  and  $\theta$  values. Figure from Reference [40].

fine segmentation, and high tolerance on a non-uniform magnetic field. It covers a range of  $0.9 < |\eta| < 2.4$ . It is divided into 4 sections attached on the front of the endcap steel disks, transverse to the beam. A chamber consists of 6 layers that can measure muon coordinates in both transverse and longitudinal plane. Cathode strips on each chamber provide the precise measurements in the  $r - \phi$  bending plane and run radially outward. Additionally, anode wires provide measurements of  $\eta$  and muon's beam-crossing time and run perpendicular to the strips. All chambers have 50% CO<sub>2</sub>, 40% Ar, and 10% CF<sub>4</sub> gas concentration [27, 40].

To have a precise measurement of the beam crossing time at the highest LHC luminosities, the RPCs are added to the CMS muon detector system in both the barrel and endcap region. It has an excellent time resolution, and a fast, independent triggering system with a sharp  $p_T$  threshold over a wide range ( $|\eta| < 1.6$ ). It is a double-gap chamber, operated in avalanche mode to ensure reliable performance at high rates. In the barrel region, it is installed adjoining the 4

sections of the DT; double layer in the 2 inner sections while single layer on the 2 outer sections. The double layering is for allowing the trigger algorithm to function even for low  $p_T$  that may stop before reaching the 2 outer sections. In the endcap region, a plane of RPCs are installed adjoining only the first 3 sections of the CSC for the trigger can use the coincidences between sections to reduce backgrounds, attain excellent  $p_T$  resolution, and improve time resolution for brunch crossing identification. RPCs consists of 95.2% Freon ( $\text{C}_2\text{H}_2\text{F}_4$ , also known as R134a), 4.5% isobutane (i- $\text{C}_4\text{H}_{10}$ ), and 0.5%  $\text{SF}_6$  gas mixture [27, 40].

## 2.3 The Trigger System

The proton beams in the LHC cross every 25 ns, which corresponds to a crossing frequency of 40 MHz. At a nominal design luminosity of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , each crossing results to an average of 20 inelastic pp events producing about 1 MB of zero-suppressed data. This means that for every second, there are approximately 40 TB of data produced which is excessively large for the current technology to process and store. In this case, a **trigger system** is used to reduce the rate of the data [27, 41].

The trigger is the start of selecting events according to interesting physics processes. The rejection or acceptance of an event is evaluated by its suitability on whether it can be included in one of the various data sets (e.g. multi-leptons, leptons+jets, electrons, and other samples) for data analysis. This selection must be done by the trigger in real time [42].

The CMS trigger system is capable of reducing the data from 32 MHz to  $O(100)$  Hz using a different sequential trigger level design. There are two levels or steps; the first step is called the **Level-1 Trigger** (L1) (Section 2.3.1) which is composed of custom-made, highly programmable hardware, and the second is the **High-Level Trigger** (HLT) (Section 2.3.2) which is a software system implemented in a commercial processor farm. Together, the L1 Trigger and the HLT

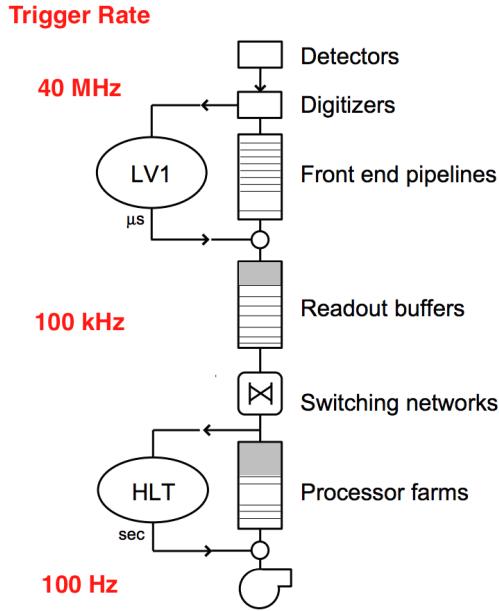


FIGURE 2.10: The data flow of the CMS trigger system. L1 Trigger reduces the rate of accepted events to less than 100 kHz and the HLT further reduces it to a final output of 100 Hz. Figure from Reference [41].

can reduce the rate to be at least a factor of  $10^6$  [27].

### 2.3.1 Level 1 Trigger System

All the data coming from the calorimeters and the muon system are first filtered by the L1 Trigger system. The system starts with the first component called **Trigger Primitive Generators** (TPG) that look for energy deposits, and track segments or hit patterns in the calorimeters and muon system, respectively. The **Regional Triggers** then combine all the information gathered and rank the trigger objects such as electrons, muons, and jets in a limited spatial region. The ranking is based on the object's energy or momentum and quality (L1 parameter measurements). Lastly, the Global Calorimeter Trigger and Global Muon Trigger determine the highest-ranking calorimeter and muon objects, respectively, and transfer it to the **Global Trigger**, the last component of the L1 Trigger System. The Global Trigger makes the last decision (based on the algorithm calculations and the status of the subdetectors and data-acquisition (DAQ) system, which is regulated by the **Trigger Control System**) whether to reject or accept the objects.

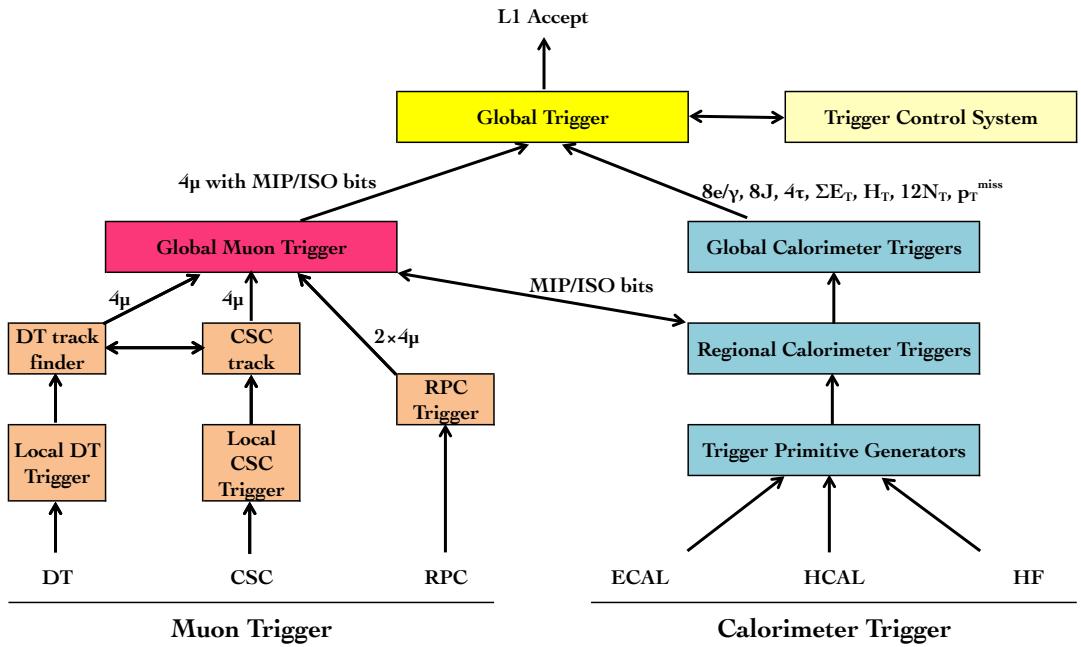


FIGURE 2.11: Architecture of CMS L1 trigger system. Figure from Reference [27].

before transferring it to the HLT for further evaluation [27]. The accepted events are ordinarily called **L1 accept**. The architecture of the CMS L1 Trigger system is depicted in Figure 2.11.

L1 Trigger has to analyze every bunch crossing thus, it only has  $3.2\ \mu\text{s}$  latency in between a given bunch crossing and the distribution of the trigger decision to the detector's front-end electronics [27].

### 2.3.2 High Level Trigger System

The reason of using the HLT system is mainly to avoid the issues when using the “traditional” three-level trigger system. In a three-level trigger system, the rate reduction of  $O(1000)$  can be attained in two stages: the “Level-2” trigger, a custom-made hardware or software-based system, provides the quick large rejection factor and then the “Level-3” trigger, a processor farm, does the selection based on more sophisticated algorithms. On the other hand, the speed and rejection power of the Level-2 trigger, and the flexibility and sophistication of

the Level-3 trigger are already incorporated in the HLT system. Additionally, the CMS HLT can attain a full-granularity data from the whole detector including the tracker, unlike the three-level trigger system which has to use “regional” information from specific parts of the detector [43].

The HLT performs the readout of the electronics after the run in L1 Trigger system (L1 accept). It then executes physics selection algorithm such as applying reconstruction and identification criteria in order to identify interesting physics events. The accepted events, together with a small sample of rejected ones, are then forwarded to the online services that monitor the detector’s performance and provides the means of archiving the events in mass storage. The event selection of the HLT is comparable to the offline processing but is more efficient even with lower resolutions [41].

The processing farm design is for achieving a reduction rate by a factor of  $10^3$  [42]. From the L1 Trigger, the HLT further decreases the rate of 100 kHz to a final output of 100 Hz.

# Chapter 3

## Analysis Strategy

After all the information of the physics object is collected from the different part of the detector, it can now be combined and reconstructed to the analysis of interest. This process is done by the **particle-flow algorithm** [44, 45] (Section 3.1). It is the one responsible for the identification and reconstruction of all the particles inside the collider. The reconstructed particles are then reconstructed again to **jets** (Section 3.1.1) and now can be studied according to the strategy of the analysis.

This chapter breaks down the methodology of the analysis. Techniques used in this analysis are expounded as well.

### 3.1 Physical Object Reconstruction

The performance of the particle-flow (PF) algorithm depends on the following factors [44]:

- Purity and efficiency of track reconstruction;
- Ability of the clustering algorithm to untwine overlapping showers;
- Capability of the linking method to connect together the deposits of every particle in the subdetectors.

The PF algorithm starts with extrapolating the tracks through the calorimeters. When a track falls within the boundaries of one or more clusters, the clusters will be now linked to that certain track. The set of track and cluster/s constitute a **charged hadron** (or electrons if the energy is all from the ECAL) and the building bricks are no longer considered in the rest of the algorithm. Identification of **muons** is done beforehand so that its track will not be misidentified to a charged hadron. Due to the frequent Bremsstrahlung photon emission, *Gaussian-Sum Filter* [46], a special track reconstruction, is used for **electrons** to attach the photon clusters to the electrons and avoid energy double counting. After all the tracks are treated, the residual clusters now result into neutral particles; **photons** in the ECAL and into **neutral hadrons** in the HCAL [44].

After the association of all particle's deposit is the evaluation of its nature and determination of its optimal 4-momentum from the combined information in the subdetectors. If there is an excess in the calibrated calorimeter energy of the clusters (linear combination of the ECAL and HCAL energy deposits) associated to a track with respect to the track momentum that is higher than  $1\sigma$ , this excess will be attributed to an overlapping neutral particle (photon or hadron) that carries an energy corresponding to the difference between the two measurements [44].

Lastly, the resulting particles are then used for jet reconstruction,  $\tau$  reconstruction and identification from its decay products, and isolated particles measurement [44].

### 3.1.1 Jet Reconstruction

A jet is a group of hadrons produced by fragmentation and hadronization of partons (quarks and gluons) that have small relative distances in momentum space. It is collimated as a cone-shape with a base of radius  $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ . These

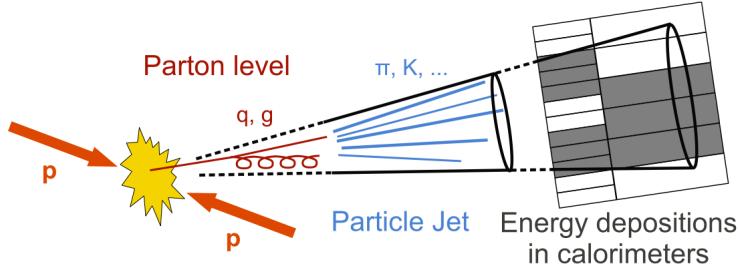


FIGURE 3.1: Sketch of a jet which is a collimated spray of stable particles from the  $pp$  collision. Figure from Reference [51].

pieces of information that are collected from the detector are then combined to reconstruct a jet by **jet reconstruction algorithms**.

There are two classes of algorithms to reconstruct jets. The first one is the **cone algorithm** where it assumes that particles within the jets are already in conical form and thus clustered base on  $\eta - \phi$  space which results in jets that have rigid circular boundaries. However, the cone algorithm has many issues such as overlapping cones and so called *dark towers*\* [47]. Also, it is generally infrared and collinear (IRC) [48] unsafe. The second is the **sequential clustering algorithm** where it assumes that particles inside the jets have small differences in its  $p_T$ 's and thus clustered based on momentum space which results in jets with different areas in  $(\eta - \phi)$  space. This analysis used the sequential clustering algorithm because it is IRC safe and with the help of FASTJET program [49], it is much faster than the cone algorithm [50].

### Jet Clustering

All of the sequential clustering algorithms use the same method:

$$d_{ij} = \min(p_{ti}^{2a}, p_{tj}^{2a}) \frac{\Delta R_{ij}^2}{R^2} \quad (3.1a)$$

$$d_{iB} = p_{ti}^{2a} \quad (3.1b)$$

---

\*Dark towers are energy clusters that are large enough to be considered as a separate jet or to be included in a nearby jet but are not designated into either.

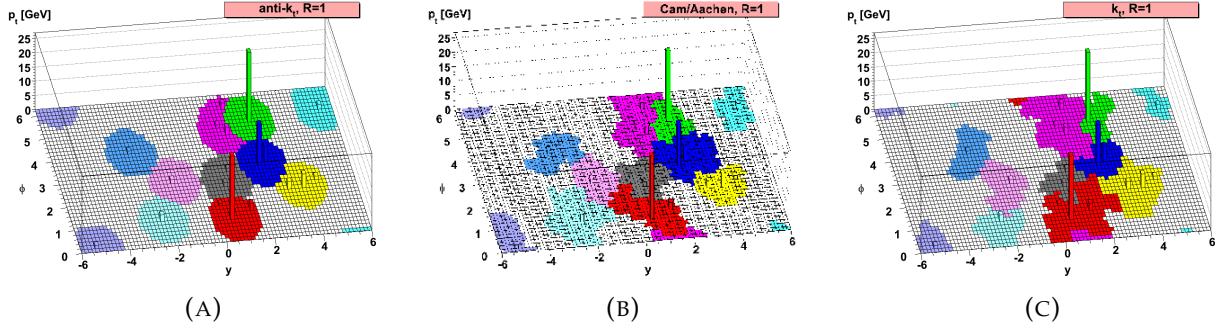


FIGURE 3.2: A sample parton-level objects with randomly injected “ghost” particles clustered using (A) anti- $k_t$  algorithm, (B) Cambridge/Aachen (CA) algorithm, and (C)  $k_t$  algorithm, illustrating the “active” attachment areas of the resulting hard jets. Figure from Reference [52].

Equation 3.1a is the distance variable between  $i^{th}$  and  $j^{th}$  particle, where  $a$  is an exponent corresponding to the clustering algorithm used,  $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$  is the distance in  $\eta - \phi$  space between the two particles, and  $R$  is the radius parameter which determines the final size of the jet. The second equation (Equation 3.1b) is the distance in momentum space between the beam axis and the detected particle [50].

The parameter  $a$  changes its value depending on the clustering algorithm used:

- $a = -1$ : anti- $k_t$  algorithm
- $a = 0$ : Cambridge/Aachen (CA) algorithm
- $a = 1$ :  $k_t$  algorithm

The sequential clustering algorithm aims to look for the minimum of the whole set  $\{d_{ij}, d_{iB}\}$ . If it finds the minimum  $d_{ij}$ , the particles  $i$  and  $j$  will then be combined into one  $(ij)$  using the summation of 4-vectors and the two particles will then be removed in the list of particles. For the minimum  $d_{iB}$ , the  $i^{th}$  particle will be labeled as the final jet and removed in the list of particles. The process of finding the minimum  $d_{ij}$  and  $d_{iB}$  is repeated until all the particles are part of a

jet that has a distance between the jet axes  $\Delta R_{ij} < R$  (inclusive clustering) or a chosen amount of jets have been found (exclusive clustering) [50].

The distance variables of the **Cambridge/Aachen algorithm** (CA) do not depend on the momentum and therefore its area sort of fluctuates and is to some extent susceptible to the underlying event (UE) and pileup (PU). The CA algorithm has the best declustering because of the purely spatial character of the distance variables. Hence, it is best suited for studying jet substructure. The  **$k_T$  algorithm** is dominated by low  $p_T$  and so prefers to cluster soft particles first which results in a considerable fluctuation on its area and also susceptible to the UE and PU. It is used to reconstruct the energy density due to the PU and correct for jet energy and isolation. It is also good on resolving subjets due to its method of clustering and less complicated in declustering than the CA algorithm. Lastly, the **anti- $k_T$  algorithm** is dominated by high  $p_T$  and so prefers to cluster hard particles first. Thus, the area only fluctuates slightly and less sensitive to the UE and PU. Although, its declustering is not good so it is not recommended for studying jet substructure.

Since this analysis studies particles with high  $p_T$ , the anti- $k_T$  algorithm is used for jet clustering with a distance parameter  $R = 0.8$  (**AK8 jets**).  $R = 0.8$  is used to cover the  $p_T$  range of  $b$  quarks that are getting close to each other ( $b\bar{b}$  pair is reconstructed as one massive jet). Clustered jets with a distance parameter  $R = 0.4$  (**AK4 jets**) are used for supporting studies.

### Jet Grooming

It is crucial to remove the soft and wide-angle radiation components of the jet to improve its mass resolution for jet reconstruction. **Jet grooming algorithm** does this process while preserving the hard parton remnants that originates from the heavy particles.

One technique of jet grooming algorithm is called **pruning** [53, 54]. Pruning is based from other jet substructure (*filtering* [55] and *top tagging* [56]) that is

modified to improve heavy particle identification. The pruning procedure starts with a clustered jet by running it again in the jet clustering algorithm and then checked for following conditions in each recombination  $i, j \rightarrow p$ :

$$z_{ij} = \frac{\min(p_{Ti}, P_{Tj})}{p_{Tp}} < z_{cut} \quad (3.2a)$$

$$\Delta R_{ij} > D_{cut} \quad (3.2b)$$

If both of these conditions are met, the softer protojet is discarded and then proceed with the algorithm. The resulting jet is the **pruned jet** [53].

The  $z_{ij}$  is the parameter of the algorithm with a default value of 0.10. The parameter  $D_{cut}$  chooses how much of the jet substructure can be pruned, a smaller value means more pruning. However, a very small value can lead to overpruning. Therefore, the  $D_{cut}$  is scaled with  $2m_j/p_{Tj} \cdot D_{cut} = m_j/p_{Tj}$  for a reasonable starting value [53].

This analysis used the pruning grooming algorithm since it has a better identification of heavy particle that decay into a jet and at the same time, suppresses QCD background.

## 3.2 b-Tagging

Identification of jets coming from  $b$  hadrons is essential since it is extensively used in different measurements such as the bottom or top pair production, H boson searches, and various supersymmetric processes. The **b-tagging algorithms** are used to find those jets which depend on  $b$  hadrons' long lifetime, relative large mass, hard fragmentation fractions, and along with its semileptonic decays [57].

In the CMS experiment, many  $b$ -tagging algorithms have been developed to have a precise and efficient  $b$  jets identification with good background rejection.

One of the many *b*-tagging algorithms is called **double-*b* tagger** [58]. It can identify jets containing two *b* quarks which is suitable for this analysis.

### 3.2.1 “Double-*b* Tagger” Algorithm

The double-*b* tagger is a newly developed *b*-tagging algorithm that has adopted the definition of the CSVv2 [59] algorithm to handle the  $b\bar{b}$  topology but instead of using the jet axis, defined by the primary vertex and jet momentum’s direction, the axes of the two N-subjettiness<sup>†</sup> are used to decipher the two *b* hadrons decay chains for  $H \rightarrow b\bar{b}$  signal (see Figure 3.3) [58].

Discriminating variables that depend on reconstructed tracks, secondary vertices (SVs) including two SV system are used as input to the MVA algorithm to separate the signal ( $H \rightarrow b\bar{b}$  jets) from the background (inclusive QCD jets). These input variables should not depend on the  $p_T$  and mass of the jet to maintain the generality of the algorithm [58].

Each track with  $p_T > 1$  GeV that are associated with a jet with  $\Delta R < 0.8$  is linked to the closest  $\tau$ -axis. The distance of a track to the  $\tau$ -axis is defined as the closest one to the axis and set to be  $< 700 \mu\text{m}$  to reject tracks from pileup. The point on the track that is closest to the  $\tau$ -axis must be within 5 cm of the primary vertex. Pairs of tracks that are compatible with the kaon mass within 30 MeV are rejected to reduce the contamination of the decay products of long-lived particles (e.g. neutral kaons). The impact parameter (IP) of a track with respect to the primary vertex sets apart the *b* hadron’s decay products from prompt tracks. It is computed in three dimensions and the impact parameter significance (SIP) is the IP to its estimated uncertainty ratio. Input variables that are related to the SVs coming from *b* hadron decay have been studied. SVs are distinguished through the Inclusive Vertex Finder (IVF) [60, 61] algorithm using tracks with  $p_T > 0.8$  GeV. The IVF algorithm uses as input the collection of reconstructed tracks in the event. Jets with a cone-size  $\Delta R < 0.7$  and the closest  $\tau$ -axis within

---

<sup>†</sup>See Section 3.4.4 for description of N-subjettiness.

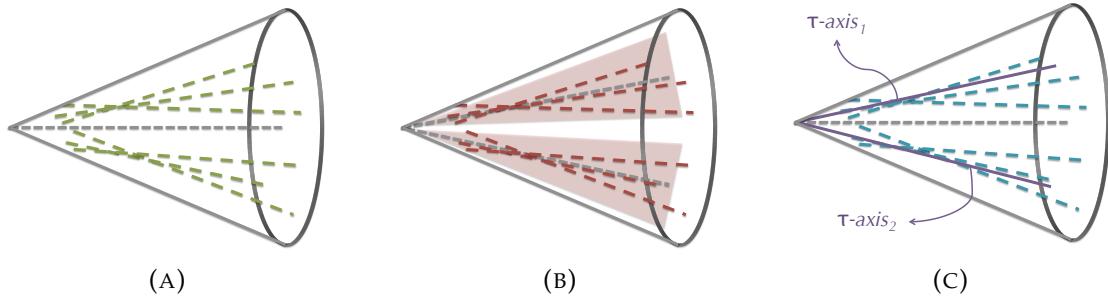


FIGURE 3.3: Schematic of (A) fatjet b-tagging, (B) subjet b-tagging, and (C) double-b tagger approaches.

that jet are to be associated to the reconstructed SVs. The track momenta of the constituent tracks from all the SVs associated to each  $\tau$ -axis are added to compute the SV mass and the SV transverse momentum for that  $\tau$ -axis [58].

The following are the input variables to the double-b tagger MVA discriminant [58]:

- The number of tracks associated with the jet;
- The first four SIP values for selected tracks ordered in decreasing SIP;
- For each  $\tau$ -axis, the first two SIP values are considered for their respective associated tracks ordered in decreasing SIP to further discriminate against single  $b$  quark and light flavor jets from QCD when one or both SV are not reconstructed due to IVF inefficiencies;
- The measured IP significance in the plane transverse to the beam axis, 2D SIP, of the first two tracks (first track) that raises the SV invariant mass above the bottom (charm) threshold of 5.2 (1.5) GeV;
- The number of SV associated with the jet;
- The significance of the 2D distance between the primary vertex and the secondary vertex, flight distance, for the SV with the smallest 3D flight distance uncertainty, for each of the two  $\tau$ -axes;

- The  $\Delta R$  between the SVs with the smallest 3D flight distance uncertainty and its  $\tau$ -axis, for each of the two  $\tau$ -axes;
- The relative pseudorapidity ( $\eta_{rel}$ ) of the tracks from all SVs with respect to their  $\tau$ -axis for the three leading tracks ordered in increasing  $\eta_{rel}$ , for each of the two  $\tau$ -axes;
- The total SV mass, defined as the total mass of all SVs associated to a given  $\tau$ -axis, for each of the two  $\tau$ -axes;
- The ratio of the total SV energy, defined as the total energy of all SVs associated to a given  $\tau$ -axis, and the total energy of all the tracks associated with the fat jet that are consistent with the primary vertex, for each of the two  $\tau$ -axes;
- The information related to the two-SV system, the  $z$  variable, defined as:

$$z = \Delta R(SV_0, SV_1) \cdot \frac{p_{T,SV_1}}{m(SV_0, SV_1)} \quad (3.3)$$

where  $SV_0$  and  $SV_1$  are the SVs with the smallest 3D flight distance uncertainty. The  $z$  variable helps to reject the  $b\bar{b}$  background from gluon splitting relying on the different kinematic properties compared to the  $b\bar{b}$  pair from the decay of a massive resonance.

Variables with enough classifier segregation that show small correlation with the other inputs and better QCD background discrimination (at least 5%) are selected as the discriminating variables. A total of 27 variables are used as an input to the double-*b* tagger MVA discriminant [58].

The performance (ROC curve) of the double-*b* tagger compared to other *b*-tagging algorithms is shown in Figure 3.4.

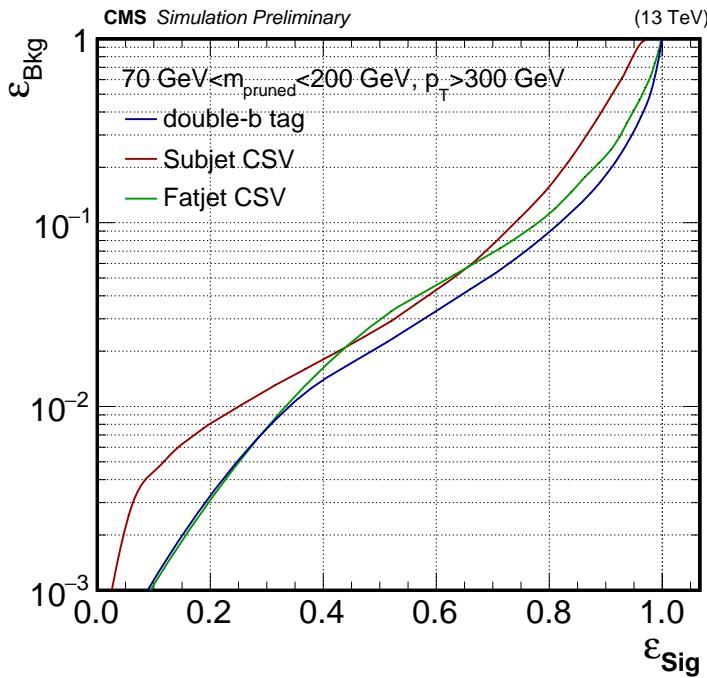


FIGURE 3.4: ROC curve of double-b tagger (blue), subjet CSV (red), and fatjet CSV (green) algorithms. This shows that the double-b tagger has better performance on discriminating the signal from the background than the other techniques.

### 3.3 Data and Simulated Samples

The data and Monte Carlo (MC) samples are processed using the CMS Software (CMSSW) version 7\_6\_X during the Run-II of the LHC in 2015.

#### 3.3.1 Data Samples

The data are collected in the pp collisions at the LHC with the CMS detector at  $\sqrt{s} = 13 \text{ TeV}$  in 2015. It corresponds to an integrated luminosity of  $2.7 \text{ fb}^{-1}$  and measured using the golden JSON file `Cert_13TeV_16Dec2015ReReco_Collisions15_25ns_JSON_Silver_v2.txt`. The datasets used are provided in Table 3.1.

#### 3.3.2 Trigger

Since the final state includes two H jets, the trigger places a requirement on the scalar sum of the transverse momentum of all outgoing quarks and gluons,

TABLE 3.1: List of 2015 dataset with its corresponding integrated luminosity used in this analysis.

Dataset	Integrated Luminosity ( $\text{fb}^{-1}$ )
/JetHT/Run2015C_25ns-16Dec2015-v1/MINIAOD	0.0177
/JetHT/Run2015D-16Dec2015-v1/MINIAOD	2.666

$H_T$ . The trigger path used is the `HLT_PFHT800`. This criterion gives at least 99% trigger efficiency for events above 1200 GeV dijet invariant mass (See Figure 3.5). The trigger efficiency is measured w.r.t. the prescaled trigger `HLT_PFHT350`.

### 3.3.3 Simulated Samples

#### Signal Samples

The signal MC samples are generated with the `MADGRAPH5_AMC@NLO` [62, 63]. The `PDF4LHC` next-to-leading order (NLO) parton distribution function (PDF) set [64–68], taken from `LHAPDF6` [69], with 4-flavor scheme, is used. The showering and hadronization processes are performed by the `PYTHIA8` [70] using the `CUETP8M1-NNPDF2.3LO` [71] tune. Simulation and event reconstruction of particles passing through the detector are then done by the `GEANT4` [72]. Additionally, all signal samples belong to the `RunIIFall15MiniaODv2_76X_mcRun2` campaign with the 25 ns asymptotic conditions and the width of signal is set to 1 MeV, consistent with the narrow-width approximation (NWA).

The list of all signal MC samples used is in Table 3.2 for Bulk Graviton and Table 3.3 for Radion.

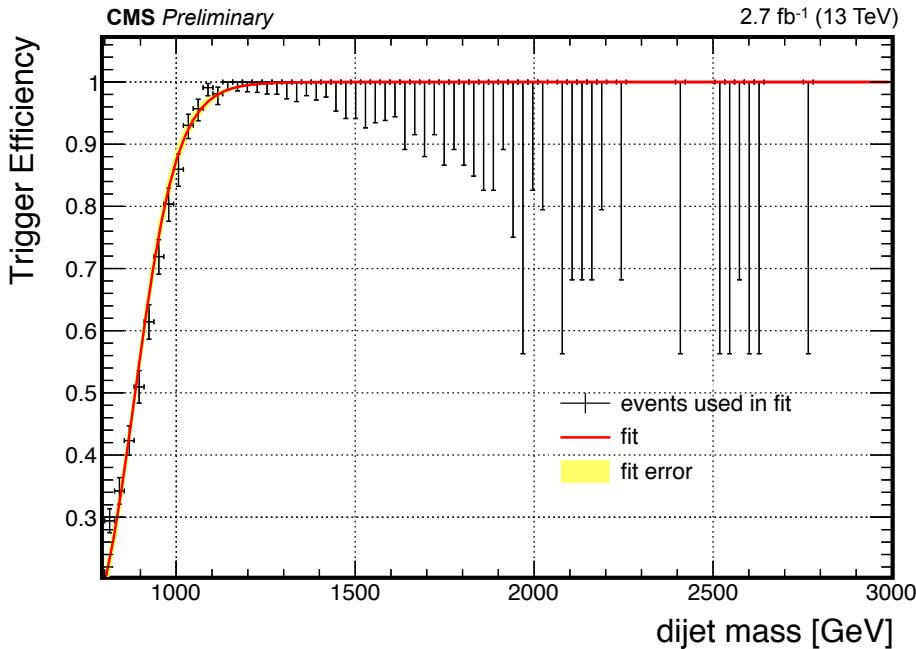


FIGURE 3.5: Trigger efficiency of HLT\_PFT800 as a function of dijet mass measured w.r.t the prescaled trigger HLT\_PFT350.

### Background Samples

The background used in this analysis is derived entirely from data. However, simulated multijet events samples are used to develop the background estimation technique before applying to data. Same with the signal MC samples, the background MC samples are generated with the `MADGRAPH5_AMC@NLO` and interfaced to `PYTHIA8` for parton showering and hadronization.

Other backgrounds such as  $t\bar{t}$ +jets, W+jets and SM diboson process are expected to have a very small contribution to the background (< 1%). Consequently, those backgrounds are not included in the analysis. The background MC samples used are listed in Table 3.4.

TABLE 3.2: List of Bulk Graviton ( $X_G \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ ) Monte Carlo simulation samples used in this analysis.

Process	Number of events
BulkGravTohhTohbbhbb_narrow_M-1000_13TeV-madgraph	50000
BulkGravTohhTohbbhbb_narrow_M-1200_13TeV-madgraph	50000
BulkGravTohhTohbbhbb_narrow_M-1400_13TeV-madgraph	50000
BulkGravTohhTohbbhbb_narrow_M-1600_13TeV-madgraph	50000
BulkGravTohhTohbbhbb_narrow_M-1800_13TeV-madgraph	50000
BulkGravTohhTohbbhbb_narrow_M-2000_13TeV-madgraph	50000
BulkGravTohhTohbbhbb_narrow_M-2500_13TeV-madgraph	49200
BulkGravTohhTohbbhbb_narrow_M-3000_13TeV-madgraph	49800
BulkGravTohhTohbbhbb_narrow_M-4500_13TeV-madgraph	48200

TABLE 3.3: List of Radion ( $X_R \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ ) Monte Carlo simulation samples used in this analysis.

Process	Number of events
RadionTohhTohbbhbb_narrow_M-1000_13TeV-madgraph	50000
RadionTohhTohbbhbb_narrow_M-1200_13TeV-madgraph	50000
RadionTohhTohbbhbb_narrow_M-1400_13TeV-madgraph	50000
RadionTohhTohbbhbb_narrow_M-1600_13TeV-madgraph	49200
RadionTohhTohbbhbb_narrow_M-1800_13TeV-madgraph	50000
RadionTohhTohbbhbb_narrow_M-2000_13TeV-madgraph	49200
RadionTohhTohbbhbb_narrow_M-2500_13TeV-madgraph	49200
RadionTohhTohbbhbb_narrow_M-3000_13TeV-madgraph	49200
RadionTohhTohbbhbb_narrow_M-3500_13TeV-madgraph	48400
RadionTohhTohbbhbb_narrow_M-4500_13TeV-madgraph	50000

## Theory Prediction

This analysis used the NLO cross sections<sup>†</sup> for Bulk Graviton [73] and Radion [74] MC samples. The Bulk Graviton production cross sections are evaluated for  $\kappa/\overline{M_{Pl}} = 0.1$ , while the Radion production cross sections are evaluated for  $\Lambda_R = 3 \text{ TeV}$

<sup>†</sup>Cross section ( $\sigma$ ) is a quantity in which its magnitude is proportional to the probability of a particle yields into a particular outcome. It has the unit of an area.

TABLE 3.4: List of QCD background Monte Carlo simulation samples used in this analysis with its cross sections at leading order (LO) taken from McM.

Process	$\sigma$ (pb)	Number of events
QCD_HT-100to200	2.785E+07 (LO)	81,906,377
QCD_HT-200to300	1.717E+06 (LO)	18,752,566
QCD_HT-300to500	3.513E+06 (LO)	20,312,907
QCD_HT-500to700	3.163E+04 (LO)	19,755,616
QCD_HT-700to1000	6.831E+03 (LO)	15,595,234
QCD_HT-1000to1500	1.207E+03 (LO)	4,966,123
QCD_HT-1500to2000	1.199E+02 (LO)	3,964,488
QCD_HT-2000toInf	2.524E+01 (LO)	1,984,407

and  $kl = 35$ . The cross sections times the decay rate for both samples are provided in Table 3.5.

 TABLE 3.5: Cross sections  $\times$  branching ratio of Bulk Graviton ( $X_G$ ) and Radion ( $X_R$ ) at next-to-leading order (NLO).

$M_X$ (GeV)	$\sigma(pp \rightarrow X_G \rightarrow HH \rightarrow b\bar{b}b\bar{b})$ $\kappa/\overline{M_{Pl}} = 0.1$	$\sigma(pp \rightarrow X_R \rightarrow HH \rightarrow b\bar{b}b\bar{b})$ $\Lambda_R = 3$ TeV and $kl = 35$
	(fb)	(fb)
1000	1.90E-01	2.11E+01
1200	7.60E-02	1.12E+01
1400	3.05E-02	5.99E+00
1600	1.32E-02	3.33E+00
1800	6.20E-03	1.92E+00
2000	3.06E-03	1.15E+00
2500	6.32E-04	3.46E-01
3000	1.49E-04	1.11E-01
3500	3.87E-05	3.78E-02
4500	3.02E-06	4.29E-03

## 3.4 Event Selection

Events with at least one reconstructed pp collision vertex, within  $|z| < 24$  cm of the detector’s center with respect to the longitudinal beam direction are required. Additional vertices that match to other overlapping pp collisions (pileup) are reconstructed using charged particles tracks. Primary interaction vertex is assumed to correspond with the one that maximizes the sum in  $p_T^2$  of these associated tracks. PF algorithm is used to reconstruct the particles as previously described in Section 3.1.

### 3.4.1 Jet Kinematics Selection

As mentioned in Section 3.1.1, this analysis used AK8 jets implemented with FASTJET. The Fall15\_25nsv2 jet energy corrections [75] were applied. Corrections on the jets’ 4-momenta [76, 77] and base on jet’s area in an even-by-event scenario [78–80] are also used. Every jet is further required to pass tight jet identification criteria [81] recommended by the JetMET POG (see Table 3.6).

The jets are arranged according to its  $p_T$  and picked the two with the highest  $p_T$ , namely **leading jet** and **sub-leading jet**, respectively. Together with tight jet identification criteria, the fiducial selections of the jets include  $p_T > 200$  GeV and a pseudorapidity range of  $|\eta| < 2.4$ .

### 3.4.2 Higgs Mass Selection

The pruning algorithm is used after the clustering of the jets as mentioned in Section 3.1.1. The mass of the two leading pruned jets is used in this analysis. Yet, a correction must be done since the **pruned mass** has a large shift with respect to the nominal value  $M_H$  hence a jet energy calibration is applied to the pruned masses. Only the L2-L3 (with L2-L3 residuals in the data) corrections are applied since the L1 corrections are correlated with the pruning procedure.

TABLE 3.6: Requirements of a tight PF jet identification provided by the JetMET POG.

Variable	Cut
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
Muon Fraction	< 0.80
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

Lastly, the corrected pruned mass must be within the 105 GeV to 135 GeV region to avoid contamination from the  $Z^0 H \rightarrow b\bar{b}b\bar{b}$  process. This region will be considered as the **signal region**.

### 3.4.3 Dijet Selection

**Dijet invariant mass** ( $M_{jj}$ ) is the sum of the 4-momentum of the two leading jets and this corresponds to the heavy resonance invariant mass. Nevertheless, to improve the dijet mass resolution and sensitivity, this analysis used **reduced dijet mass** ( $M_{jj}^{red}$ ) as the main observable which is defined as:

$$M_{jj}^{red} \equiv M_{jj} - (M_{jet1} - M_H) - (M_{jet2} - M_H) \quad (3.4)$$

where  $M_{jet1}$  and  $M_{jet2}$  are the pruned jet masses and  $M_H$  is the SM H boson (125 GeV). The reduced dijet mass improves the resolution while maintaining the peak position at  $\approx M_X$ . A selection of  $M_{jj}^{red} > 800$  GeV is applied.

In addition, the two leading jets have to be relatively close in  $\eta$  to reduce the contribution from the QCD background thus a requirement of  $|\Delta\eta_{jj}| < 1.3$  is applied.

### 3.4.4 N-Subjettiness Selection

**N-subjettiness** ( $\tau_N$ ) [82–84] is an inclusive jet shape giving an absolute measurement of the alignment of the jet energy to the subjets and is perfect for boosted object identification. It is defined as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}) \quad (3.5)$$

where  $k$  runs over the constituent particles in a given jet,  $p_{T,k}$  is the transverse momentum of  $k$ ,  $\Delta R_{J,k}$  is the distance between a candidate subjet axis  $J$  and a constituent particle  $k$ . The normalization factor  $d_0$  is defined as:

$$d_0 = \sum_K p_{T,k} R_0 \quad (3.6)$$

where  $R_0$  is the characteristic jet radius used in the original jet clustering algorithm.

The QCD jets that have diffuse shower of wide-angle radiation can typically result to a large  $\tau_1$  but it can have a large value of  $\tau_2$  as well. In this manner, the ratio  $\tau_N/\tau_{N-1}$  is used to have a manageable offline parameter to modify the signal efficiency/background rejection while avoiding to do the particle clustering again. Moreover, the ratio  $\tau_{21} = \tau_2/\tau_1$  is found to be an effective discriminating variable to isolate boosted H bosons from the QCD background with large invariant mass [82]. The two leading jets have to satisfy  $\tau_{21} < 0.6$  since this cut shows better discriminating power.

### 3.4.5 Tagging Selection

The double-b tagger discriminant is divided into three working points: loose cut (discriminant value  $> 0.3$ ), medium cut ( $> 0.6$ ), and the tight cut ( $> 0.9$ ). The signal region is found to be at events where the medium cut is applied to the two

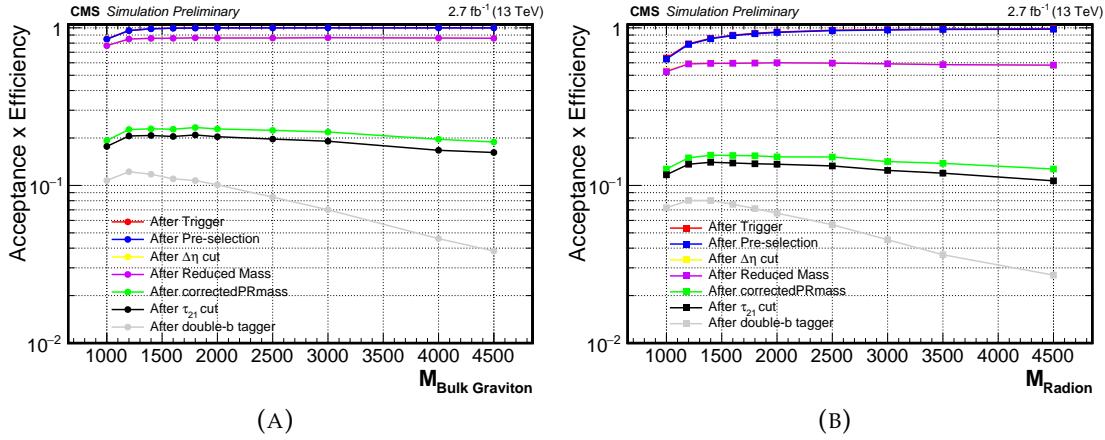


FIGURE 3.6: Acceptance  $\times$  Efficiency of (A) Spin-2 Bulk Graviton and (B) Spin-0 Radion for every mass points.

leading jets thus the medium cut is used in this analysis. The signal efficiency of the medium cut is  $\sim 66\%$  with a mistag rate of 4.4%.

### 3.4.6 Selection Distribution

The summary of all the event selection criteria applied is in Table 3.7. Event selection cutflow efficiency for both Bulk Graviton and Radion are shown in Figure 3.8 and the values are in Tables 3.8 and 3.9, respectively.

Different kinematic distributions for signal and background simulated samples are plotted in Figures 3.6 and 3.7. The event selections are not yet applied on this comparison except for the trigger and fiducial selections. Additionally, a 70 - 200 GeV H mass window [58] is applied on the corrected pruned mass. The signal and background distribution are normalized to  $2.7 \text{ fb}^{-1}$  data luminosity. The cross section of the signal is set to 200 fb to have a better comparison with the background.

## 3.5 Data and MC Comparison

A comparison between data and simulation for different kinematic observables is presented in Figures 3.9 and 3.10. The QCD background is scaled to match the

TABLE 3.7: Summary of all event selection criteria used in this analysis.

Trigger	HLT_PFHT800
	$p_T > 200 \text{ GeV}$
Higgs Jet Selection	$ \eta  < 2.4$ Jet Identification with tight working point
Dijet Selection	$ \Delta\eta  < 1.3$ $M_{jj}^{red} > 800 \text{ GeV}$
Signal Region	$105 \text{ GeV} < \text{L2-L3 corrected pruned mass jet} < 135 \text{ GeV}$
N-subjettiness	$\tau_{21} < 0.6$ for both AK8 jet
Tagging Selection	double-b tagger $> 0.6$

number of events available in data. This comparison is done before applying the  $b$ -tagging selection (double- $b$  tagger discriminant  $> 0.6$ ). A mass window of 70 - 200 GeV [58] on the corrected pruned mass is used for this comparison.

TABLE 3.8: Bulk Graviton event selection cutflow efficiency. The efficiency of each column is taken w.r.t. the previous cut. The last column represents the full event selection efficiency.

Mass (GeV)	Trigger	Fiducial selection	$\Delta\eta$	Reduced mass	Signal region	$\tau_{21}$	double-b tag	Total Efficiency
1000	0.85	1.0	0.92	0.99	0.25	0.92	0.61	0.11
1200	0.96	1.0	0.89	1.0	0.27	0.91	0.59	0.12
1400	0.99	1.0	0.87	1.0	0.27	0.91	0.57	0.12
1600	1.0	1.0	0.86	1.0	0.26	0.90	0.54	0.11
1800	1.0	1.0	0.87	1.0	0.27	0.90	0.51	0.11
2000	1.0	1.0	0.86	1.0	0.26	0.89	0.50	0.10
2500	1.0	1.0	0.86	1.0	0.26	0.88	0.42	0.08
3000	1.0	1.0	0.87	1.0	0.25	0.87	0.37	0.07
4500	1.0	1.0	0.86	1.0	0.22	0.86	0.24	0.04

TABLE 3.9: Radion event selection cutflow efficiency. The efficiency of each column is taken w.r.t. the previous cut. The last column represents the full event selection efficiency.

Mass (GeV)	Trigger	Fiducial selection	$\Delta\eta$	Reduced mass	Signal region	$\tau_{21}$	double-b tag	Total Efficiency
1000	0.64	0.99	0.84	0.99	0.24	0.92	0.61	0.07
1200	0.79	0.99	0.76	0.99	0.25	0.91	0.59	0.08
1400	0.86	0.99	0.70	1.0	0.26	0.90	0.57	0.08
1600	0.90	0.99	0.67	1.0	0.26	0.90	0.55	0.08
1800	0.92	0.99	0.65	1.0	0.26	0.89	0.52	0.07
2000	0.94	0.99	0.64	1.0	0.25	0.90	0.49	0.07
2500	0.97	0.99	0.62	1.0	0.25	0.88	0.42	0.06
3000	0.97	0.99	0.61	1.0	0.24	0.88	0.36	0.05
3500	0.98	0.99	0.60	1.0	0.24	0.87	0.30	0.04
4500	0.99	0.99	0.59	1.0	0.22	0.84	0.25	0.03

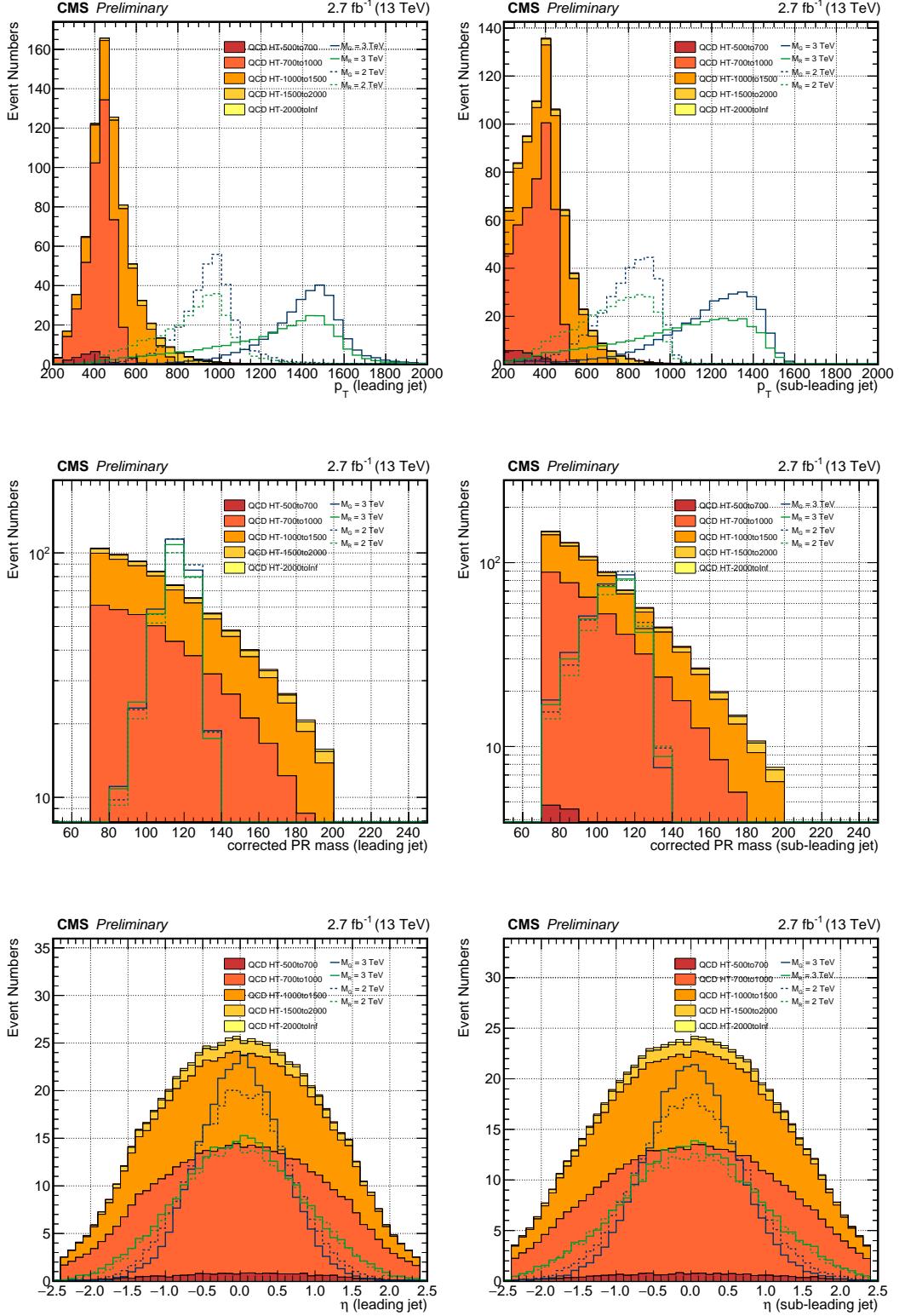


FIGURE 3.7: The comparison of signal and background for  $p_T$  (top), corrected pruned mass (middle), and  $\eta$  (bottom) of leading (left) and sub-leading (right) AK8 Jets. The signal samples of  $M_X = 3 \text{ TeV}$  (solid line) and  $2 \text{ TeV}$  (dashed line) for Bulk Graviton (blue) and Radion (green) are shown. The signal cross section is set to 200 fb for this comparison only.

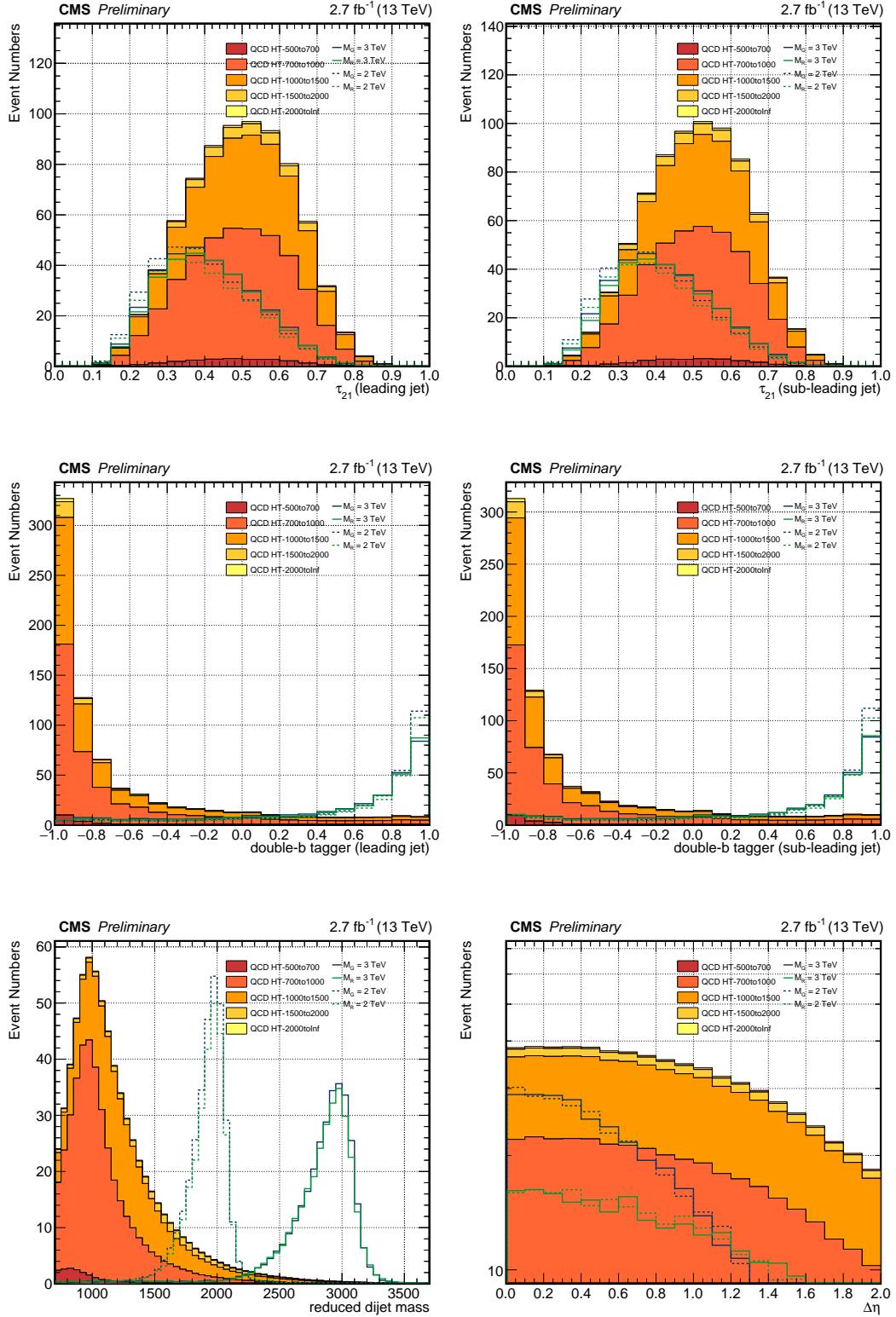


FIGURE 3.8: The comparison of signal and background for  $\tau_{21}$  (top) and double-b tagger (middle) of leading (left) and sub-leading (right) AK8 Jets, reduced mass (bottom-left), and  $\Delta\eta$  (bottom-right). The signal samples of  $M_X = 3 \text{ TeV}$  (solid line) and  $2 \text{ TeV}$  (dashed line) for Bulk Graviton (blue) and Radion (green) are shown. The signal cross section is set to  $200 \text{ fb}$  for this comparison only.

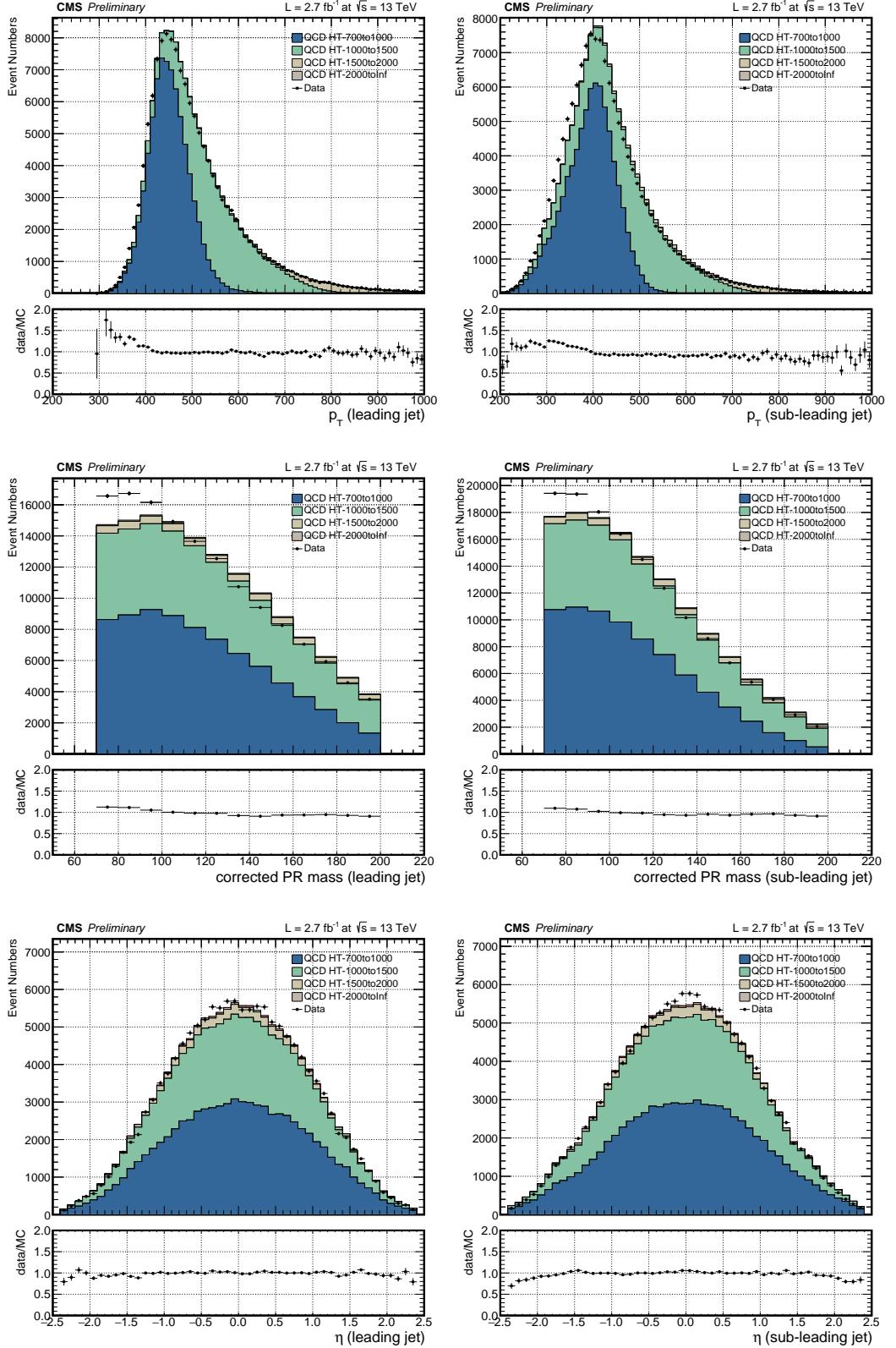


FIGURE 3.9: Data and Monte Carlo comparison for  $p_T$  (top), corrected pruned mass (middle), and  $\eta$  (bottom) of leading (left) and sub-leading (right) AK8 Jets. QCD background MC sample has been normalized to match the number of data events.

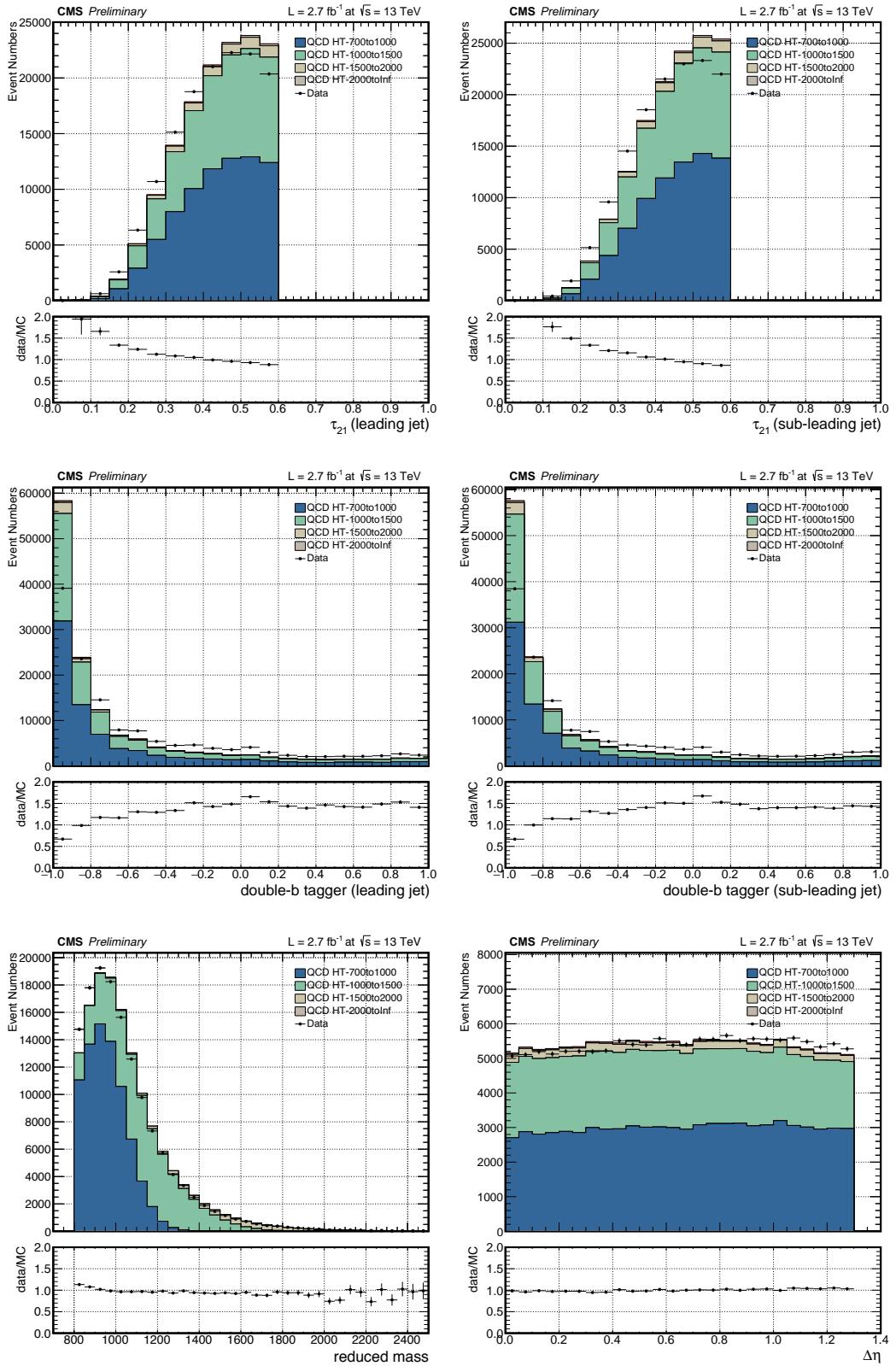


FIGURE 3.10: Data and Monte Carlo comparison for  $\tau_{21}$  (top) and double-b tagger (middle) of leading (left) and sub-leading (right) AK8 Jets, reduced mass (bottom-left), and  $\Delta\eta$  (bottom-right). QCD background MC sample has been normalized to match the number of data events.

# Chapter 4

## Background Estimation

For predicting a data-driven background, the **ABCD method** is one of the background estimation method used which exploits a number of sidebands of the signal region. It uses two uncorrelated variables that divide the parameter space into 4 regions where one region is considered as the signal region and the other regions are the control region. In any case that the two variables are correlated, the *ABCD* method is extended to **Alphabet method**.

### 4.1 “ALPHABET” Method

In this analysis, the Alphabet method uses variables that are defined with respect to the leading jet based on its mass (pruned mass) and tagging variable (double-b tagger). These variables are now used to define several regions, as outlined in Figure 4.1:

- **Pre-tag** is the region where it passes all the selection in Table 3.7 except for the two variables (mass and tagging requirements) for leading jet. This is the superset of the events from the following regions;
- **Sideband (SB)** is the region where it fails the mass requirement;
- **Anti-tag region (AR)** passes the mass requirement but fails the tagging requirement;
- **Signal region (SR)** passes both mass and tagging requirement.

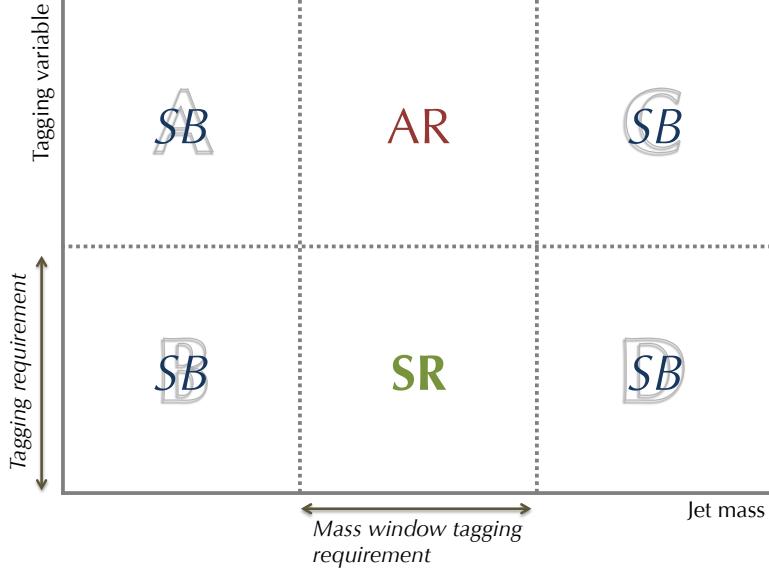


FIGURE 4.1: Schematic representation of the Alphabet method showing all the regions used for predicting the background. All six regions comprise of pre-tag region. The sidebands (SBs) are the regions that fail the H-jet mass requirement. The anti-tag region (AR) passes the H-jet mass requirement but fails the double-b tagger requirement. The signal region (SR) both pass the H-jet mass and double-b tagger requirement.

The shape of the reduced mass distribution, but not the normalization, in the AR and SR is expected to be similar. Since the double-b tagger discriminant does depend on pruned mass (thus Alphabet method), the **conversion rate** ( $R_{p/f}$ ), or **pass-fail ratio**, must be measured as a function of pruned mass which is defined as the ratio of the number of events passing and failing the tagging selection,  $R_{p/f} \equiv N_{pass}/N_{fail}$ . The pass-fail ratio is measured in the SBs blinding the AR and SR. It is then fitted with a quadratic function:  $f(a, b, c; x) = a + bx + cx^2$ , where the  $a$ ,  $b$ , and  $c$  are coefficients obtained from the fit. In this manner, the fit interpolates the pass-fail ratio through the SR and thus prediction can be made by weighing every event in AR with the right pass-fail ratio that matches AR's pruned mass.

$$\text{Background in SR}(m_{\text{pruned}} - M_H) = R_{p/f}(m_{\text{pruned}} - M_H) \times AR(m_{\text{pruned}} - M_H) \quad (4.1)$$

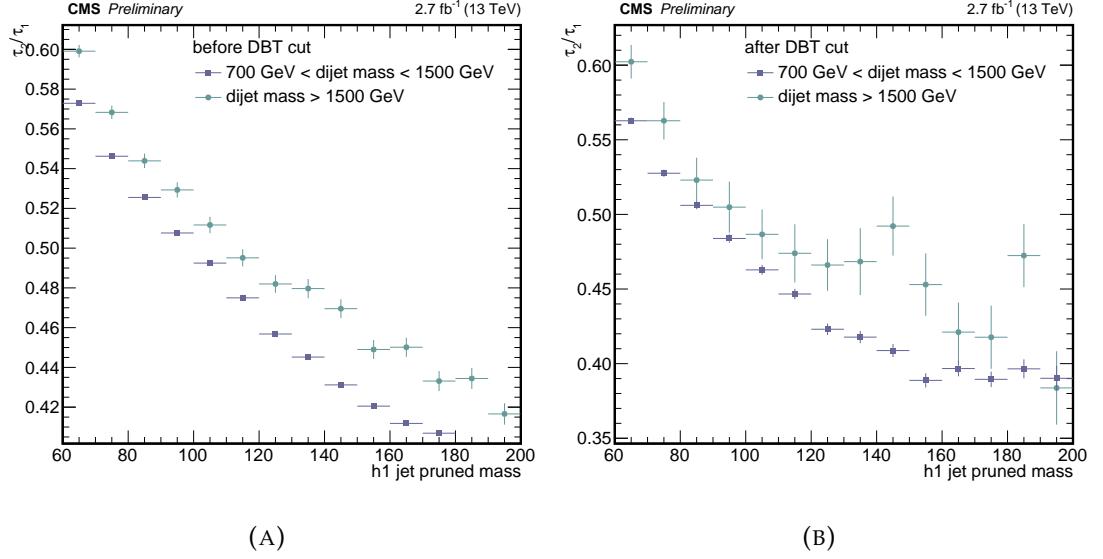


FIGURE 4.2: Dependence of the  $\tau_{21}$  on the jet pruned mass, before (A) and after (B) the double-b tagger selection. The high dijet mass ( $\geq 1500$  GeV), marked in circle markers, is separated from the low dijet mass (700 - 1500 GeV) marked with square markers. This shows that the  $\tau_{21}$  is dependent on the dijet mass event.

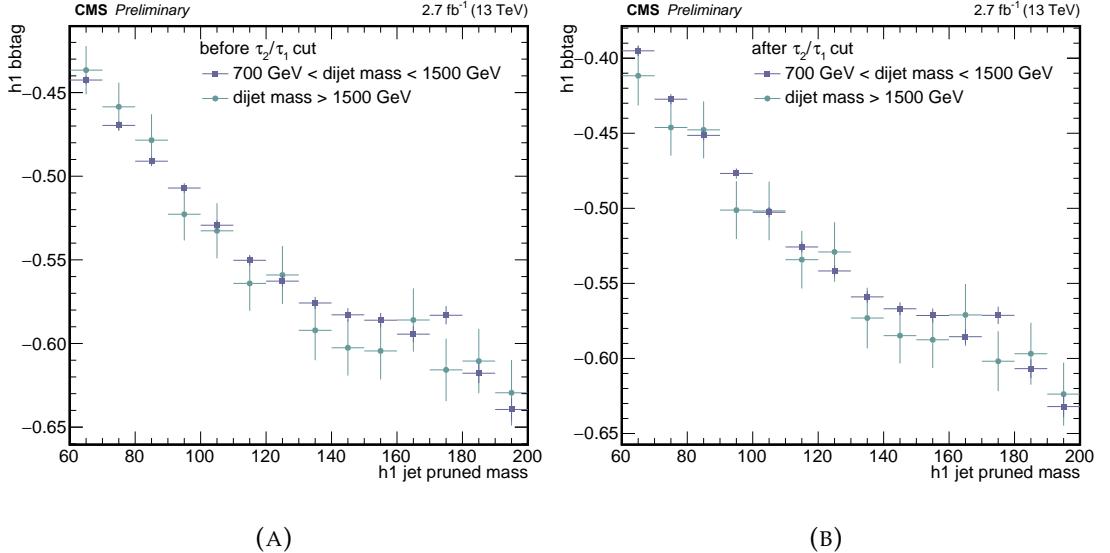


FIGURE 4.3: Dependence of the double-b tagger on the jet pruned mass, before (A) and after (B) the  $\tau_{21}$  selection. The high dijet mass ( $\geq 1500$  GeV), marked in circle markers, is separated from the low dijet mass (700 - 1500 GeV) marked with square markers. The double-b tagger shows no dependence on dijet mass event.

### 4.1.1 Double-b Tagger as the Tagging Variable

The double-b tagger is chosen as the tagging variable because its conversion rate is around 10%, this guarantees enough statistics in the SB to do the estimation. The double-b tagger also does not depend on the dijet mass or jet  $p_T$  unlike the  $\tau_{21}$  (see Figures 4.2 and 4.3). If a sample is dominated by the lower dijet mass, using a tagging variable that has dijet mass or jet  $p_T$  dependence will not be practical since its pass-fail ratio will obviously be different for higher dijet mass. These conditions make the double-b tagger a good candidate to use for the tagging variable.

### 4.1.2 Result

Figure 4.4a shows the pass-fail ratio as a function of pruned mass with the quadratic fit function. As mentioned above, the SR and AR are blinded for this fit. The predicted background is shown in Figure 4.4b along with the observed number events the SR. The shape and normalization of the predicted background show good agreement with the data spectrum. Even so, the data has no significant excess over the predicted background.

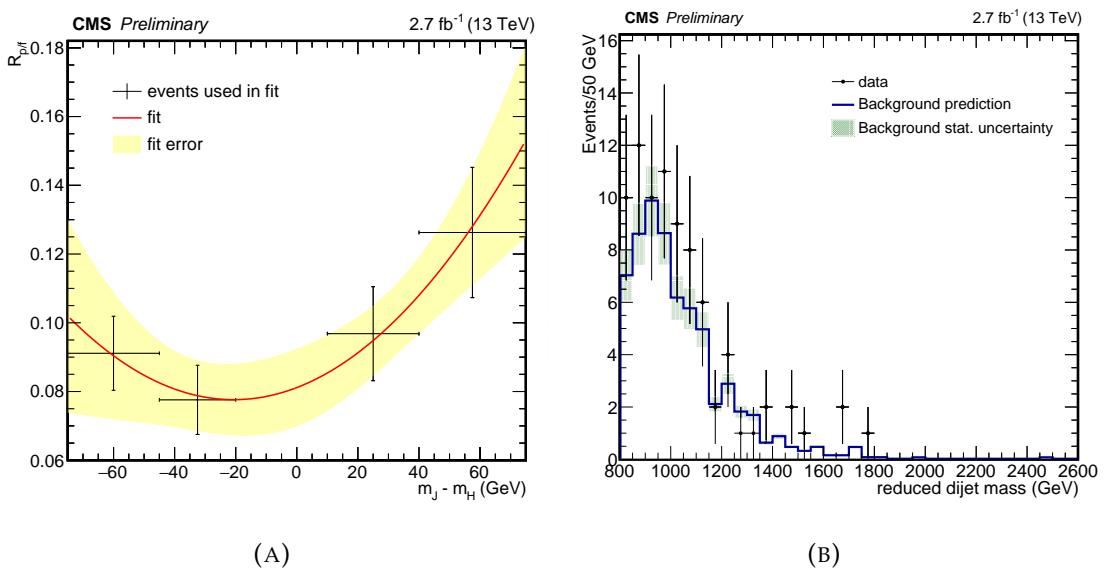


FIGURE 4.4: (A) Pass-fail ratio  $R_{p/f}$  with the quadratic fit and fitting error, and (B) predicted reduced dijet mass distribution in the signal region after applying the "Alphabet" method with uncertainty.



# Chapter 5

## Systematic Uncertainties

Since the background is computed entirely from the data, the following sources of systematic uncertainties only affect the expected signal efficiencies and not the background.

### 5.1 Systematics on Signal Efficiency

#### Luminosity

The Lumi POG assigned a 2.7% total uncertainty on the CMS luminosity in 2015 [85].

#### Pileup

The uncertainty from the pileup is obtained by varying the measured minimum bias cross section of pp collisions at 13 TeV (= 69 mb) by  $\pm 5\%$ . The relative change of the pileup efficiency is then used as the uncertainty of the pileup. The uncertainty obtained ranges from 0.2% to 1.2%, depending on the mass point of the signal MC sample.

## PDF and renormalization/factorization scale

### PDF

The **parton distribution function** (PDF) is defined as a probability density for finding a particle in a resolution scale using a certain longitudinal momentum fraction. The PDF is computed by fitting the observables in the experimental data. There are different groups that define the PDF but the NNPDF3.0 is used as the default PDF set for the Run-II of the LHC [68].

The uncertainty of the PDF is derived from the relative standard deviation\* of the 100 PDF weights efficiency distribution. An uncertainty of 0.8 - 2.7% is obtained depending on the mass point of the signal MC sample.

### Renormalization/Factorization scale

The **renormalization scale** ( $\mu_R$ ) describes the scale of the physical quantity, derived with QCD computations, with the energy of the interaction energy scale while the **factorization scale** ( $\mu_F$ ) characterize the evolution of the PDF of hadron showers with the changes in the interaction energy scale. The uncertainty of the scales are determined by varying the corresponding scale by factors of 0.5 and 2. An uncertainty of approximately  $10^{-5}$  is obtained.

### $\tau_{21}$ scale factor

The signal yield is corrected by applying the efficiency ratio in data to the simulation of the  $\tau_{21}$  selection that has been computed in Reference [86]. The **scale factor** (SF) for the  $\tau_{21}$  has been analyzed using a  $W$  jets in an augmented  $t\bar{t}$  data sample. The SF for the  $\tau_{21} < 0.6$  cut is  $0.979 \pm 0.028$  for each jet. A 5.6% uncertainty is obtained assuming that the two leading jets are correlated.

---

\*Relative standard deviation, or coefficient of variation, is a standard measurement of a dispersion in a probability or frequency distribution. It is the ratio of the standard deviation to the mean,  $\sigma/\mu$ .

### H-tagging uncertainty

An uncertainty of 10% has been added to adapt the SF of the  $\tau_{21}$  to H jets and to take into account the jet mass dependence of the  $\tau_{21}$ . This uncertainty is measured by comparing the PYTHIA 8 and HERWIG++ simulated signal samples on how the two showering algorithms can differentiate the hadronically decaying W and H jets [87].

### Jet energy-momentum scale

The uncertainty in the jet energy-momentum scale is obtained by applying the jet energy corrections [88, 89] on the 4-momentum of each jet and varied within  $\pm 1\sigma$  based on the  $p_T$  and  $\eta$  of the jet. This uncertainty is treated as a shape based uncertainty.

### Jet energy resolution

The jet energy resolution (JER) is characterized by the width of the reconstructed jet energy distribution that was caused by jet energy corrections. In the data, the measured JER is usually worse than in the simulation. Ergo, to take into account this difference with the data, the jets in MC are smeared using the recommended procedure of the JetMET POG [90].

There are two smearing methods that can be utilized to match the jet  $p_T$  resolution in simulation to data; the **scaling method** requires the presence of a matching particle-level jet and the **stochastic smearing** that uses random number from a normal distribution.

The scaling method rescales the corrected 4-momentum of the jet with a factor:

$$c_{JER} = 1 + (s_{JER} - 1) \frac{p_T - p_T^{ptcl}}{p_T} \quad (5.1)$$

where  $p_T$  is the transverse momentum of the jet being rescaled,  $p_T^{ptcl}$  is the transverse momentum of the corresponding jet clustered from the generator-level particles, and  $s_{JER}$  is the data-to-simulation core resolution scale factor. The factor  $c_{JER}$  is set to zero if its value is negative. This method only works if there is a well-matched particle-level jet and can result in a large shift of the response otherwise. The following requirements are to be satisfied to know that there is a well-matched particle-level jet:

$$\Delta R = \frac{R_{cone}}{2}, |p_T - p_T^{ptcl}| < 3\sigma_{JER}p_T \quad (5.2)$$

where  $R_{cone}$  is the jet cone size parameter and  $\sigma_{JER}$  is the relative  $p_T$  resolution as measured in the simulation.

On the other hand, the stochastic smearing does not require a matching particle-level jet and rescales the corrected 4-momentum of the jet with a factor:

$$c_{JER} = 1 + N(0, \sigma_{JER}) \sqrt{\max(s_{JER}^2 - 1, 0)} \quad (5.3)$$

where  $N(0, \sigma_{JER})$  denotes the random number sampled from a normal distribution with a zero mean and  $\sigma_{JER}$  as the standard deviation. Again, the factor  $c_{JER}$  is set to zero if its value is negative.

Yet, to have a better smearing result, the JetMET POG recommends the **hybrid method** which combines the two smearing method. If there is a well-matched particle, the scaling method is applied; otherwise the stochastic smearing will be applied. This uncertainty is treated as a shape based uncertainty.

## Trigger efficiency

Systematic uncertainty on trigger efficiency is studied using events in data that has been collected with a looser trigger selection with a prescaled path requiring  $H_T > 350$  GeV (HLT\_PFHT350). The measured efficiency is then fitted with an

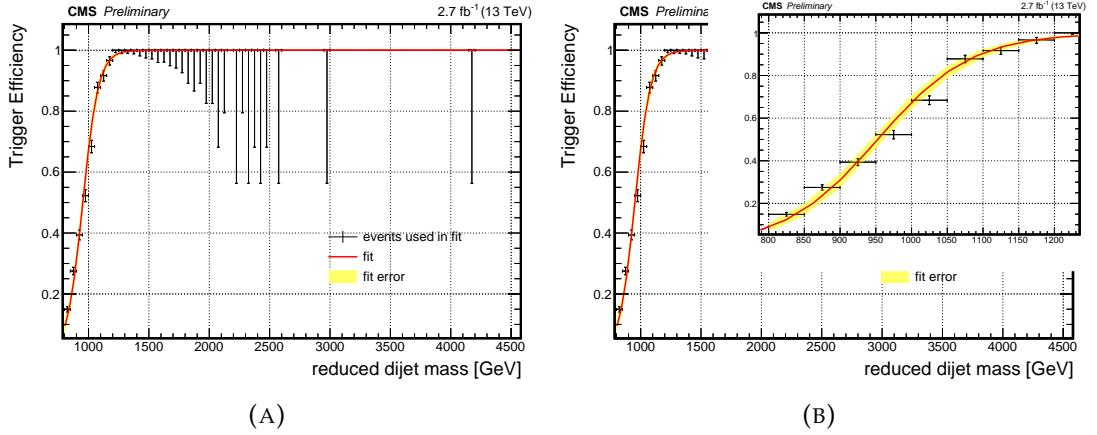


FIGURE 5.1: Trigger efficiency of HLT\_PFHT800 as a function of reduced dijet mass measured w.r.t. the prescaled trigger HLT\_PFHT350 fitted with an error function and varied within  $\pm 1\sigma$  of its uncertainty. Right-hand plot shows a zoom-in image in 800 - 1200 GeV range (turn-on region).

error function and varied within  $\pm 1\sigma$  of its uncertainty. The measured trigger efficiency is shown in Figure 5.1.

At 1 TeV, the uncertainty is  $\sim 6\%$  but as the mass goes higher, the uncertainty decreases until it is  $< 1\%$ . This uncertainty is treated as a shape based uncertainty.

## Double-b tagging

The SFs for the double-b tagger have been analyzed in an augmented gluon splitting to  $b\bar{b}$  data sample. The SFs of medium double-b tagger selection (double-b tagger discriminant  $> 0.6$ ) shown in Table 5.1 are obtained from Reference [58]. The SFs are applied in both leading and sub-leading jet. This uncertainty is treated as a shape based uncertainty.

TABLE 5.1: Medium double-b tagger selection data/MC efficiency ratio or scale factor. The value of the scale factors are provided in Reference [58].

$p_T$ (GeV)	< 400	400 - 500	500 - 600	> 600
SFs	$0.92 \pm 0.09$	$0.99 \pm 0.12$	$0.94 \pm 0.19$	$1.05 \pm 0.21$

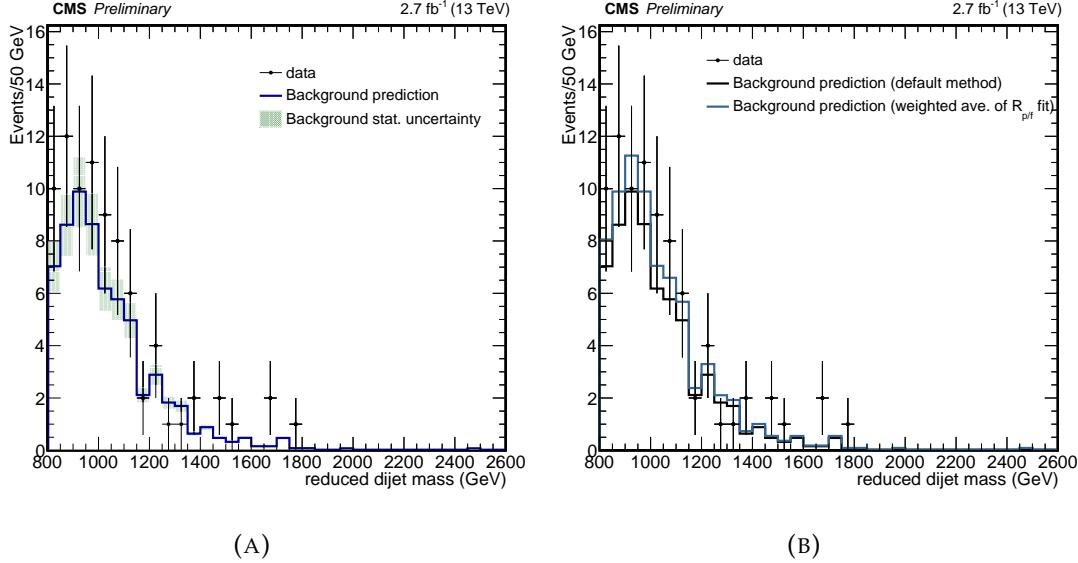


FIGURE 5.2: (A) Background estimation with uncertainty in the fit to the SB region, and (B) background prediction using the default method (black line) and the weighted average of  $R_{p/f}$  fit (blue line).

## Multijet background

There are two sources of uncertainties studied for the multijet background. The first and dominant uncertainty comes from the fitting of the SB region in the pass-fail ratio (Figure 5.2a). The second one is done by using the weighted average<sup>†</sup> of the  $R_{p/f}$  fit to determine the uncertainty on the dependence of the pass-fail ratio in the pruned mass (Figure 5.2b). These uncertainties are treated as shape based uncertainties.

## Jet mass scale

The uncertainty in the jet mass scale is obtained by applying the jet energy corrections [88, 89] on the pruned mass of each jet and varied within  $\pm 1\sigma$  based on the  $p_T$  and  $\eta$  of the jet. This uncertainty is treated as a shape based uncertainty.

<sup>†</sup>Weighted average:  $\bar{x} = \sum_i^n (x_i \sigma_i^{-2}) / \sum_i^n \sigma_i^{-2}$

## 5.2 Summary of the Systematic Uncertainties

All the sources of systematic uncertainties that affects this analysis are summarized in Table 5.2.

TABLE 5.2: Summary of all systematic uncertainties that affects this analysis.

Uncertainties	Variation (%)
Luminosity	2.7
Pileup	0.2 - 1.2
$\tau_{21}$ Scale Factor + H-tagging	5.6 + 10
PDF and Renormalization/Factorization Scale	0.8 - 2.7 / $\sim 10^{-5}$
Jet Energy-Momentum Scale	<i>shape</i>
Jet Energy Resolution	<i>shape</i>
Trigger Efficiency	<i>shape</i>
Double-b Tagging	<i>shape</i>
QCD Multi-jet Background	<i>shape</i>
Jet Mass Scale	<i>shape</i>



# Chapter 6

## Results and Discussion

Although the background prediction shows good agreement with the data, there is still no significant data excess above the background. Even so, it does not mean that there are no  $X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$  events, limits can be set on the reduced dijet mass for the theoretical model of Bulk Graviton and Radion. Setting limits is done to quantify the degree of agreement between

By this means, limit setting is used to interpret the results in this analysis. The limits are obtained by **asymptotic CL<sub>S</sub> method** that is run in the HIGGS COMBINE TOOL [91] with accordance to the recommendation of the LHC Higgs Combination Group [92].

### 6.1 Asymptotic CL<sub>S</sub> Method

### 6.2 Limit Plot



# **Chapter 7**

## **Summary, Conclusion, and Outlook**

### **7.1 Summary**

### **7.2 Conclusion**

### **7.3 Outlook**





# Bibliography

- [1] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30 –61. ISSN: 0370-2693. DOI: <http://dx.doi.org/10.1016/j.physletb.2012.08.021>. URL: <http://www.sciencedirect.com/science/article/pii/S0370269312008581>.
- [2] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1 –29. ISSN: 0370-2693. DOI: <http://dx.doi.org/10.1016/j.physletb.2012.08.020>. URL: <http://www.sciencedirect.com/science/article/pii/S037026931200857X>.
- [3] CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at  $\sqrt{s} = 7$  and 8 TeV". In: *Journal of High Energy Physics* 2013.6 (2013), p. 81. ISSN: 1029-8479. DOI: [10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081). URL: [https://doi.org/10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081).
- [4] A. Nisati and V. Sharma. *Discovery of the Higgs Boson*. World Scientific Publishing Company, 2016. ISBN: 9789814425469. URL: <https://books.google.com.tw/books?id=3Y74DAAAQBAJ>.
- [5] Gordon Kane. "The dawn of physics beyond the standard model". In: *Scientific American* 288.6 (2003), pp. 68–75.
- [6] David Galbraith. *UX: Standard Model of the Standard Model*. 2017. URL: <http://davidgalbraith.org/portfolio/ux-standard-model-of-the-standard-model/>.
- [7] Marco Pieri. *SM Higgs Branching Ratios and Total Decay Widths (update in CERN Report4 2016)*. 2016. URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR>.
- [8] A. Denner et al. "Standard Model Higgs-Boson Branching Ratios with Uncertainties". In: *Eur. Phys. J. C* 71 (2011), p. 1753. DOI: [10.1140/epjc/s10052-011-1753-8](https://doi.org/10.1140/epjc/s10052-011-1753-8). arXiv: [1107.5909 \[hep-ph\]](https://arxiv.org/abs/1107.5909).
- [9] ATLAS Collaboration. "Search for Higgs Boson Pair Production in the  $\gamma\gamma b\bar{b}$  Final State Using  $pp$  Collision Data at  $\sqrt{s} = 8$  TeV from the ATLAS Detector". In: *Phys. Rev. Lett.* 114 (8 2015), p. 081802. DOI: [10.1103](https://doi.org/10.1103)/

## BIBLIOGRAPHY

---

- PhysRevLett.* 114 . 081802. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.114.081802>.
- [10] ATLAS Collaboration. *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*. Tech. rep. ATLAS-CONF-2014-004. Geneva: CERN, 2014. URL: <http://cds.cern.ch/record/1664335>.
- [11] ATLAS Collaboration. “Search for Higgs boson pair production in the  $b\bar{b}b\bar{b}$  final state from pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”. In: *Eur. Phys. J.* C75.9 (2015), p. 412. DOI: [10.1140/epjc/s10052-015-3628-x](https://doi.org/10.1140/epjc/s10052-015-3628-x). arXiv: [1506.00285 \[hep-ex\]](https://arxiv.org/abs/1506.00285).
- [12] ATLAS Collaboration. “Searches for Higgs boson pair production in the  $hh \rightarrow b\bar{b}\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$  channels with the ATLAS detector”. In: *Phys. Rev. D* 92 (9 2015), p. 092004. DOI: [10.1103/PhysRevD.92.092004](https://doi.org/10.1103/PhysRevD.92.092004). URL: <https://link.aps.org/doi/10.1103/PhysRevD.92.092004>.
- [13] ATLAS Collaboration. “Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *Phys. Rev.* D94.5 (2016), p. 052002. DOI: [10.1103/PhysRevD.94.052002](https://doi.org/10.1103/PhysRevD.94.052002). arXiv: [1606.04782 \[hep-ex\]](https://arxiv.org/abs/1606.04782).
- [14] CMS Collaboration. “Searches for heavy Higgs bosons in two-Higgs-doublet models and for  $t \rightarrow ch$  decay using multilepton and diphoton final states in pp collisions at 8 TeV”. In: *Phys. Rev.* D90 (2014), p. 112013. DOI: [10.1103/PhysRevD.90.112013](https://doi.org/10.1103/PhysRevD.90.112013). arXiv: [1410.2751 \[hep-ex\]](https://arxiv.org/abs/1410.2751).
- [15] CMS Collaboration. “Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with  $h \rightarrow \tau\tau$ ”. In: 755 (2016), pp. 217–244. ISSN: 0370-2693. DOI: [http://dx.doi.org/10.1016/j.physletb.2016.01.056](https://dx.doi.org/10.1016/j.physletb.2016.01.056). URL: <http://www.sciencedirect.com/science/article/pii/S0370269316000757>.
- [16] The CMS collaboration. “Identification of b-quark jets with the CMS experiment”. In: *Journal of Instrumentation* 8.04 (2013), P04013. URL: <http://stacks.iop.org/1748-0221/8/i=04/a=P04013>.
- [17] CMS Collaboration. “Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV”. In: *Phys. Rev. D* 94 (5 2016), p. 052012. DOI: [10.1103/PhysRevD.94.052012](https://doi.org/10.1103/PhysRevD.94.052012). URL: <https://link.aps.org/doi/10.1103/PhysRevD.94.052012>.

- [18] CMS Collaboration. "Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 8 TeV". In: *Phys. Lett.* B749 (2015), pp. 560–582. DOI: [10.1016/j.physletb.2015.08.047](https://doi.org/10.1016/j.physletb.2015.08.047). arXiv: [1503.04114 \[hep-ex\]](https://arxiv.org/abs/1503.04114).
- [19] CMS Collaboration. "Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks". In: *Eur. Phys. J. C* 76.CMS-EXO-12-053. CERN-EP-2016-041. CMS-EXO-12-053 (2016), 371. 38 p. URL: <https://cds.cern.ch/record/2134962>.
- [20] Experiments. 2012. URL: [http://cds.cern.ch/record/1997374](https://cds.cern.ch/record/1997374).
- [21] *The Large Hadron Collider*. 2014. URL: [http://cds.cern.ch/record/1998498](https://cds.cern.ch/record/1998498).
- [22] "LHC Guide". 2017. URL: <https://cds.cern.ch/record/2255762>.
- [23] *Building the HL-LHC with the Industry*. URL: <https://project-hl-lhc-industry.web.cern.ch/>.
- [24] *The accelerator complex*. 2012. URL: [http://cds.cern.ch/record/1997193](https://cds.cern.ch/record/1997193).
- [25] Cinzia De Melis. "The CERN accelerator complex. Complexe des accélérateurs du CERN". In: (2016). General Photo. URL: [http://cds.cern.ch/record/2197559](https://cds.cern.ch/record/2197559).
- [26] Mike Lamont. "LHC Performance in Run 2 and Beyond". In: *PoS Lepton-Photon2015* (2016), 001. 8 p. URL: [http://cds.cern.ch/record/2257503](https://cds.cern.ch/record/2257503).
- [27] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (2008), S08004. URL: [http://stacks.iop.org/1748-0221/3/i=08/a=S08004](https://stacks.iop.org/1748-0221/3/i=08/a=S08004).
- [28] Tai Sakuma and Thomas McCauley. "Detector and Event Visualization with SketchUp at the CMS Experiment". In: *Journal of Physics: Conference Series* 513.2 (2014), p. 022032. URL: [http://stacks.iop.org/1742-6596/513/i=2/a=022032](https://stacks.iop.org/1742-6596/513/i=2/a=022032).
- [29] Matthias Schott and Monica Dunford. "Review of single vector boson production in pp collisions at  $\sqrt{s} = 7$  TeV". In: *Eur. Phys. J. C* 74 (2014), p. 2916. DOI: [10.1140/epjc/s10052-014-2916-1](https://doi.org/10.1140/epjc/s10052-014-2916-1). arXiv: [1405.1160 \[hep-ex\]](https://arxiv.org/abs/1405.1160).
- [30] A. Astbury, F.C. Khanna, and R.W. Moore. *Fundamental Interactions: Proceedings of the 22nd Lake Louise Winter Institute, Lake Louise, Alberta, Canada, 19-24 February, 2007*. World Scientific Publishing Company Incorporated, 2008. ISBN: 9789812776099. URL: <https://books.google.com.tw/books?id=h1pDQAAQBAJ>.

## BIBLIOGRAPHY

---

- [31] Heidi Schellman. *Learn Hardon Collider Physics in 3 Days*. 2005. URL: <http://particle.physics.ucdavis.edu/workshops/TASI04/schellman.pdf>.
- [32] *Bending Particles*. URL: <http://cms.cern/detector/bending-particles>.
- [33] Y. Allkofer et al. “Design and performance of the silicon sensors for the CMS barrel pixel detector”. In: *Nucl. Instrum. Methods Phys. Res., A* 584.physics/0702092 (2008). Comments: 22 pages, 21 figures, pp. 25–41. URL: <http://cds.cern.ch/record/1142546>.
- [34] Laura Borrello et al. *Sensor Design for the CMS Silicon Strip Tracker*. Tech. rep. CMS-NOTE-2003-020. Geneva: CERN, 2003. URL: <http://cds.cern.ch/record/687861>.
- [35] CMS Collaboration. “Performance and operation of the CMS electromagnetic calorimeter”. In: *Journal of Instrumentation* 5.03 (2010), T03010. URL: <http://stacks.iop.org/1748-0221/5/i=03/a=T03010>.
- [36] Bora Isildak. “Measurement of the differential dijet production cross section in proton-proton collisions at  $\sqrt{s} = 7$  tev”. PhD thesis. Bogazici U., 2011. arXiv: [1308.6064 \[hep-ex\]](https://arxiv.org/abs/1308.6064). URL: <https://inspirehep.net/record/1251416/files/arXiv:1308.6064.pdf>.
- [37] E. Cheu et al. *Calorimetry in High Energy Physics: Proceedings of the Seventh International Conference*. 1998. ISBN: 9789814545099. URL: <https://books.google.com.tw/books?id=aALQDgAAQBAJ>.
- [38] CMS Collaboration. “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data”. In: *Journal of Instrumentation* 5.03 (2010), T03012. URL: <http://stacks.iop.org/1748-0221/5/i=03/a=T03012>.
- [39] Karl Jakobs. *8. Energy measurements in calorimeters*. 2017. URL: <http://portal.uni-freiburg.de/jakobs/dateien/vorlesungsdateien/particledetectors/kap8>.
- [40] The CMS collaboration. “The performance of the CMS muon detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV at the LHC”. In: *Journal of Instrumentation* 8.11 (2013), P11002. URL: <http://stacks.iop.org/1748-0221/8/i=11/a=P11002>.
- [41] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002. URL: <http://cds.cern.ch/record/578006>.

- [42] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical Design Report CMS. URL: <http://cds.cern.ch/record/706847>.
- [43] CMS Collaboration. “Commissioning of the CMS High-Level Trigger with Cosmic Rays”. In: *JINST* 5 (2010), T03005. DOI: [10.1088/1748-0221/5/03/T03005](https://doi.org/10.1088/1748-0221/5/03/T03005). arXiv: [0911.4889 \[physics.ins-det\]](https://arxiv.org/abs/0911.4889).
- [44] Florian Beaudette. “The CMS Particle Flow Algorithm”. In: *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. 2013, pp. 295–304. arXiv: [1401.8155 \[hep-ex\]](https://arxiv.org/abs/1401.8155). URL: <https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf>.
- [45] Albert M Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector”. In: (2017). arXiv: [1706.04965 \[physics.ins-det\]](https://arxiv.org/abs/1706.04965).
- [46] Wolfgang Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. CMS-NOTE-2005-001. Geneva: CERN, 2005. URL: <https://cds.cern.ch/record/815410>.
- [47] Michael G. Albrow et al. “Tevatron-for-LHC Report of the QCD Working Group”. In: 2006. arXiv: [hep-ph/0610012 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0610012). URL: [http://lss.fnal.gov/cgi-bin/find\\_paper.pl?conf=06-359](https://lss.fnal.gov/cgi-bin/find_paper.pl?conf=06-359).
- [48] Gavin P. Salam and Grègory Soyez. “A practical seedless infrared-safe cone jet algorithm”. In: *Journal of High Energy Physics* 2007.05 (2007), p. 086. URL: [http://stacks.iop.org/1126-6708/2007/i=05/a=086](https://stacks.iop.org/1126-6708/2007/i=05/a=086).
- [49] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “FastJet user manual”. In: *The European Physical Journal C* 72.3 (2012), p. 1896. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2). URL: <https://doi.org/10.1140/epjc/s10052-012-1896-2>.
- [50] Ryan Atkin. “Review of jet reconstruction algorithms”. In: *Journal of Physics: Conference Series* 645.1 (2015), p. 012008. URL: [http://stacks.iop.org/1742-6596/645/i=1/a=012008](https://stacks.iop.org/1742-6596/645/i=1/a=012008).
- [51] Henning Kirschenmann. *Jets at CMS and the determination of their energy scale*. 2012. URL: <http://cms.web.cern.ch/news/jets-cms-and-determination-their-energy-scale>.
- [52] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- $k_t$  jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. URL: [http://stacks.iop.org/1126-6708/2008/i=04/a=063](https://stacks.iop.org/1126-6708/2008/i=04/a=063).

## BIBLIOGRAPHY

---

- [53] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. "Techniques for improved heavy particle searches with jet substructure". In: *Phys. Rev. D* 80 (5 2009), p. 051501. DOI: [10.1103/PhysRevD.80.051501](https://doi.org/10.1103/PhysRevD.80.051501). URL: <https://link.aps.org/doi/10.1103/PhysRevD.80.051501>.
- [54] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. "Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches". In: *Phys. Rev. D* 81 (9 2010), p. 094023. DOI: [10.1103/PhysRevD.81.094023](https://doi.org/10.1103/PhysRevD.81.094023). URL: <https://link.aps.org/doi/10.1103/PhysRevD.81.094023>.
- [55] Jonathan M. Butterworth et al. "Jet substructure as a new Higgs search channel at the LHC". In: *Proceedings, 34th International Conference on High Energy Physics (ICHEP 2008): Philadelphia, Pennsylvania, July 30-August 5, 2008*. 2008. arXiv: [0810.0409 \[hep-ph\]](https://arxiv.org/abs/0810.0409). URL: <https://inspirehep.net/record/798300/files/arXiv:0810.0409.pdf>.
- [56] David E. Kaplan et al. "Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks". In: *Phys. Rev. Lett.* 101 (14 2008), p. 142001. DOI: [10.1103/PhysRevLett.101.142001](https://doi.org/10.1103/PhysRevLett.101.142001). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.101.142001>.
- [57] CMS Collaboration. *Performance of the b-jet identification in CMS*. Tech. rep. CMS-PAS-BTV-11-001. Geneva: CERN, 2011. URL: <https://cds.cern.ch/record/1366061>.
- [58] CMS Collaboration. *Identification of double-b quark jets in boosted event topologies*. Tech. rep. CMS-PAS-BTV-15-002. Geneva: CERN, 2016. URL: <https://cds.cern.ch/record/2195743>.
- [59] CMS Collaboration. *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. Tech. rep. CMS-PAS-BTV-15-001. Geneva: CERN, 2016. URL: <https://cds.cern.ch/record/2138504>.
- [60] CMS Collaboration. *Performance of b tagging at sqrt(s)=8 TeV in multijet, ttbar and boosted topology events*. Tech. rep. CMS-PAS-BTV-13-001. Geneva: CERN, 2013. URL: <https://cds.cern.ch/record/1581306>.
- [61] CMS Collaboration. "Measurement of  $b\bar{b}$  angular correlations based on secondary vertex reconstruction at  $\sqrt{s} = 7$  TeV". In: *Journal of High Energy Physics* 2011.3 (2011), p. 136. ISSN: 1029-8479. DOI: [10.1007/JHEP03\(2011\)136](https://doi.org/10.1007/JHEP03(2011)136). URL: [https://doi.org/10.1007/JHEP03\(2011\)136](https://doi.org/10.1007/JHEP03(2011)136).

- [62] Johan Alwall et al. “MadGraph 5: going beyond”. In: *Journal of High Energy Physics* 2011.6 (2011), p. 128. ISSN: 1029-8479. DOI: [10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128). URL: [https://doi.org/10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128).
- [63] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (2014), p. 79. ISSN: 1029-8479. DOI: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079). URL: [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [64] Stefano Carrazza et al. “A compression algorithm for the combination of PDF sets”. In: *Eur. Phys. J.* C75 (2015), p. 474. DOI: [10.1140/epjc/s10052-015-3703-3](https://doi.org/10.1140/epjc/s10052-015-3703-3). arXiv: [1504.06469 \[hep-ph\]](https://arxiv.org/abs/1504.06469).
- [65] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (2016), p. 023001. URL: <http://stacks.iop.org/0954-3899/43/i=2/a=023001>.
- [66] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. In: *Phys. Rev. D* 93 (3 2016), p. 033006. DOI: [10.1103/PhysRevD.93.033006](https://doi.org/10.1103/PhysRevD.93.033006). URL: <https://link.aps.org/doi/10.1103/PhysRevD.93.033006>.
- [67] L. A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *The European Physical Journal C* 75.5 (2015), p. 204. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-015-3397-6](https://doi.org/10.1140/epjc/s10052-015-3397-6). URL: <https://doi.org/10.1140/epjc/s10052-015-3397-6>.
- [68] Richard D. Ball et al. “Parton distributions for the LHC run II”. In: *Journal of High Energy Physics* 2015.4 (2015), p. 40. ISSN: 1029-8479. DOI: [10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040). URL: [https://doi.org/10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040).
- [69] Andy Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *The European Physical Journal C* 75.3 (2015), p. 132. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-015-3318-8](https://doi.org/10.1140/epjc/s10052-015-3318-8). URL: <https://doi.org/10.1140/epjc/s10052-015-3318-8>.
- [70] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159 –177. ISSN: 0010-4655. DOI: <https://dx.doi.org/10.1016/j.cpc.2015.01.024>. URL: <http://www.sciencedirect.com/science/article/pii/S0010465515000442>.

## BIBLIOGRAPHY

---

- [71] V. Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *The European Physical Journal C* 76.3 (2016), p. 155. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-016-3988-x](https://doi.org/10.1140/epjc/s10052-016-3988-x). URL: <https://doi.org/10.1140/epjc/s10052-016-3988-x>.
- [72] S. Agostinelli et al. "Geant4 - a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250 –303. ISSN: 0168-9002. DOI: [http://dx.doi.org/10.1016/S0168-9002\(03\)01368-8](http://dx.doi.org/10.1016/S0168-9002(03)01368-8). URL: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [73] Maxime Gouzevitch. *CrossSectionsLHC/WED*. 2016. URL: [https://github.com/CrossSectionsLHC/WED/blob/master/KKGraviton\\_Bulk/AnalysesTables/BulkGraviton\\_yr4\\_13TeV\\_kmp10p5\\_hh4b.dat](https://github.com/CrossSectionsLHC/WED/blob/master/KKGraviton_Bulk/AnalysesTables/BulkGraviton_yr4_13TeV_kmp10p5_hh4b.dat).
- [74] Maxime Gouzevitch. *CrossSectionsLHC/WED*. 2016. URL: [https://github.com/CrossSectionsLHC/WED/blob/master/Radion\\_Bulk/AnalysesTables/BulkRadion\\_yr4\\_13TeV\\_NLO\\_LR\\_3TeV\\_hh4b.dat](https://github.com/CrossSectionsLHC/WED/blob/master/Radion_Bulk/AnalysesTables/BulkRadion_yr4_13TeV_NLO_LR_3TeV_hh4b.dat).
- [75] Recommended Jet Energy Corrections and Uncertainties for Data and MC. 2015. URL: [https://twiki.cern.ch/twiki/bin/view/CMS/JECDatamc#Recommended\\_for\\_Data](https://twiki.cern.ch/twiki/bin/view/CMS/JECDatamc#Recommended_for_Data).
- [76] The CMS collaboration. "Determination of jet energy calibration and transverse momentum resolution in CMS". In: *Journal of Instrumentation* 6.11 (2011), P11002. URL: <http://stacks.iop.org/1748-0221/6/i=11/a=P11002>.
- [77] "Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11  $\text{fb}^{-1}$  of data in CMS". In: (2013). URL: <https://cds.cern.ch/record/1545350>.
- [78] *Pileup Removal Algorithms*. Tech. rep. CMS-PAS-JME-14-001. Geneva: CERN, 2014. URL: <https://cds.cern.ch/record/1751454>.
- [79] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The catchment area of jets". In: *Journal of High Energy Physics* 2008.04 (2008), p. 005. URL: <http://stacks.iop.org/1126-6708/2008/i=04/a=005>.
- [80] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Physics Letters B* 659.1 (2008), pp. 119 –126. ISSN: 0370-2693. DOI: <http://dx.doi.org/10.1016/j.physletb.2007.09.077>. URL: <http://www.sciencedirect.com/science/article/pii/S0370269307011094>.

- [81] Andreas Hinzmann. *Jet Identification*. 2016. URL: [https://twiki.cern.ch/twiki/bin/view/CMS/JetID#Recommendations\\_for\\_13\\_TeV\\_2015](https://twiki.cern.ch/twiki/bin/view/CMS/JetID#Recommendations_for_13_TeV_2015).
- [82] Jesse Thaler and Ken Van Tilburg. “Identifying boosted objects with N-subjettiness”. In: *Journal of High Energy Physics* 2011.3 (2011), p. 15. ISSN: 1029-8479. DOI: [10.1007/JHEP03\(2011\)015](https://doi.org/10.1007/JHEP03(2011)015). URL: [https://doi.org/10.1007/JHEP03\(2011\)015](https://doi.org/10.1007/JHEP03(2011)015).
- [83] Jesse Thaler and Ken Van Tilburg. “Maximizing boosted top identification by minimizing N-subjettiness”. In: *Journal of High Energy Physics* 2012.2 (2012), p. 93. ISSN: 1029-8479. DOI: [10.1007/JHEP02\(2012\)093](https://doi.org/10.1007/JHEP02(2012)093). URL: [https://doi.org/10.1007/JHEP02\(2012\)093](https://doi.org/10.1007/JHEP02(2012)093).
- [84] Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn. “N jettness: an inclusive event shape to veto jets”. In: *Phys. Rev. Lett.* 105 (9 2010), p. 092002. DOI: [10.1103/PhysRevLett.105.092002](https://doi.org/10.1103/PhysRevLett.105.092002). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.105.092002>.
- [85] CMS Collaboration. *CMS luminosity measurement for the 2015 data-taking period*. Tech. rep. CMS-PAS-LUM-15-001. Geneva: CERN, 2017. URL: <https://cds.cern.ch/record/2138682>.
- [86] CMS Collaboration. *Search for massive resonances decaying into pairs of boosted W and Z bosons at  $\sqrt{s} = 13$  TeV*. Tech. rep. CMS-PAS-EXO-15-002. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2117062>.
- [87] S. Ahuja et. al. “Search for heavy resonances decaying to a pair of Higgs bosons in four b quark final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV”. CMS Analysis Note 2015/287\_v9, CERN. 2015.
- [88] CMS Collaboration. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. In: *Journal of Instrumentation* 12.02 (2017), P02014. URL: <http://stacks.iop.org/1748-0221/12/i=02/a=P02014>.
- [89] CMS Collaboration. “Jet energy scale and resolution performances with 13TeV data”. In: (2016). URL: <https://cds.cern.ch/record/2160347>.
- [90] Andrey Popov. *Jet Energy Resolution*. 2017. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution>.
- [91] Nicholas Wardle. *Documentation of the RooStats -based statistics tools for Higgs PAG*. 2017. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit>.

## BIBLIOGRAPHY

---

- [92] CMS Collaboration. "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV". In: *Eur. Phys. J.* C75.5 (2015), p. 212. DOI: [10.1140/epjc/s10052-015-3351-7](https://doi.org/10.1140/epjc/s10052-015-3351-7). arXiv: [1412.8662](https://arxiv.org/abs/1412.8662) [hep-ex].
- [93] S-S. Yu et. al. "Comparison of radion/2HDM heavy higgs/Zprime/RS/bulk gravitons  $X \rightarrow hh \rightarrow bbbb$  and  $X \rightarrow WW \rightarrow lvqq$ ". Diboson Meeting, CERN. 2015.
- [94] Andrew J. Larkoski et al. "Soft drop". In: *Journal of High Energy Physics* 2014.5 (2014), p. 146. ISSN: 1029-8479. DOI: [10.1007/JHEP05\(2014\)146](https://doi.org/10.1007/JHEP05(2014)146). URL: [https://doi.org/10.1007/JHEP05\(2014\)146](https://doi.org/10.1007/JHEP05(2014)146).



# Appendix A

## Comparison of Different Heavy Resonance Models

Kinematic distributions and acceptance of different heavy resonance models X were studied. The models studied are listed on Table A.1.

Acceptance is defined as the ratio of the number of events after and before the fiducial selections:

$$\text{Acceptance} = \frac{N_{\text{events}}^{\text{after}}}{N_{\text{events}}^{\text{before}}} \quad (\text{A.1})$$

Fiducial selections used are as follows:

- $p_T > 30 \text{ GeV}$
- $|\eta_H| < 2.5$
- $|\Delta\eta_{HH}| < 1.3$

TABLE A.1: Different heavy resonance models with its spin and process.

Particle Name	Spin	Model	Process
Radion	0	heft_radion	Gluon fusion (GF)
2HDM H	0	2HDM in default MADGRAPH	Drell-Yan (DY)
Bulk Graviton	2	RS_bulk_ktilda	GF
RS Graviton	2	RS_exovv	GF + DY

## Appendix A. Comparison of Different Heavy Resonance Models

---

The Table A.2 shows the results for the acceptance of fiducial selections for resonance models. Models with spin 0 have acceptances of approximately 57%. The Bulk Graviton model has the highest acceptance of 86%. Finally, the RS Graviton is the only model that has different acceptance for different masses.

TABLE A.2: Acceptance of the fiducial selections for heavy resonance models.

Particle Name	Spin	Acceptance (%)	
		$M_X = 2 \text{ TeV}$	$M_X = 4.5 \text{ TeV}$
Radion	0	57.6	57.1
2HDM H	0	57.2	57.4
Bulk Graviton	2	86.3	86.2
RS Graviton	2	65.1	44.3

Figures in A.1 and A.2 shows different kinematic distribution for heavy resonance models [93]. Spin 0 models have the same trends for  $M_X = 2000 \text{ GeV}$ . Spin 2 models that are produced via glusion fusion (GF) have the same trends on its kinematic distributions.

In conclusion, the heavy resonance models can be categorized according to its spin rather than each model.

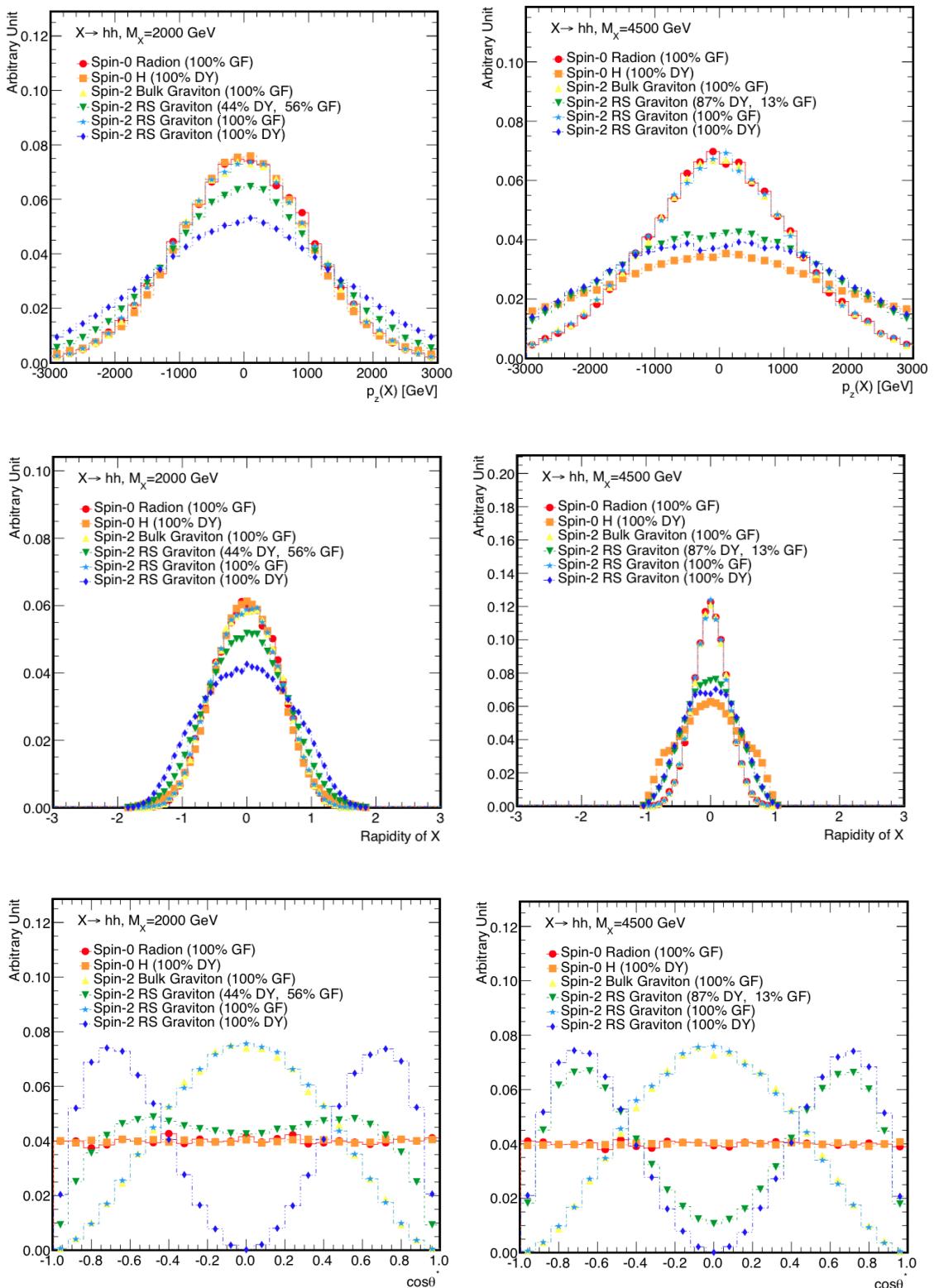


FIGURE A.1:  $p_Z$  (top), rapidity (middle), and  $\cos(\theta^*)$  (bottom) of heavy resonance models for  $M_X = 2000 \text{ GeV}$  (left) and  $M_X = 4500 \text{ GeV}$  (right). Figures from Reference [93].

## Appendix A. Comparison of Different Heavy Resonance Models

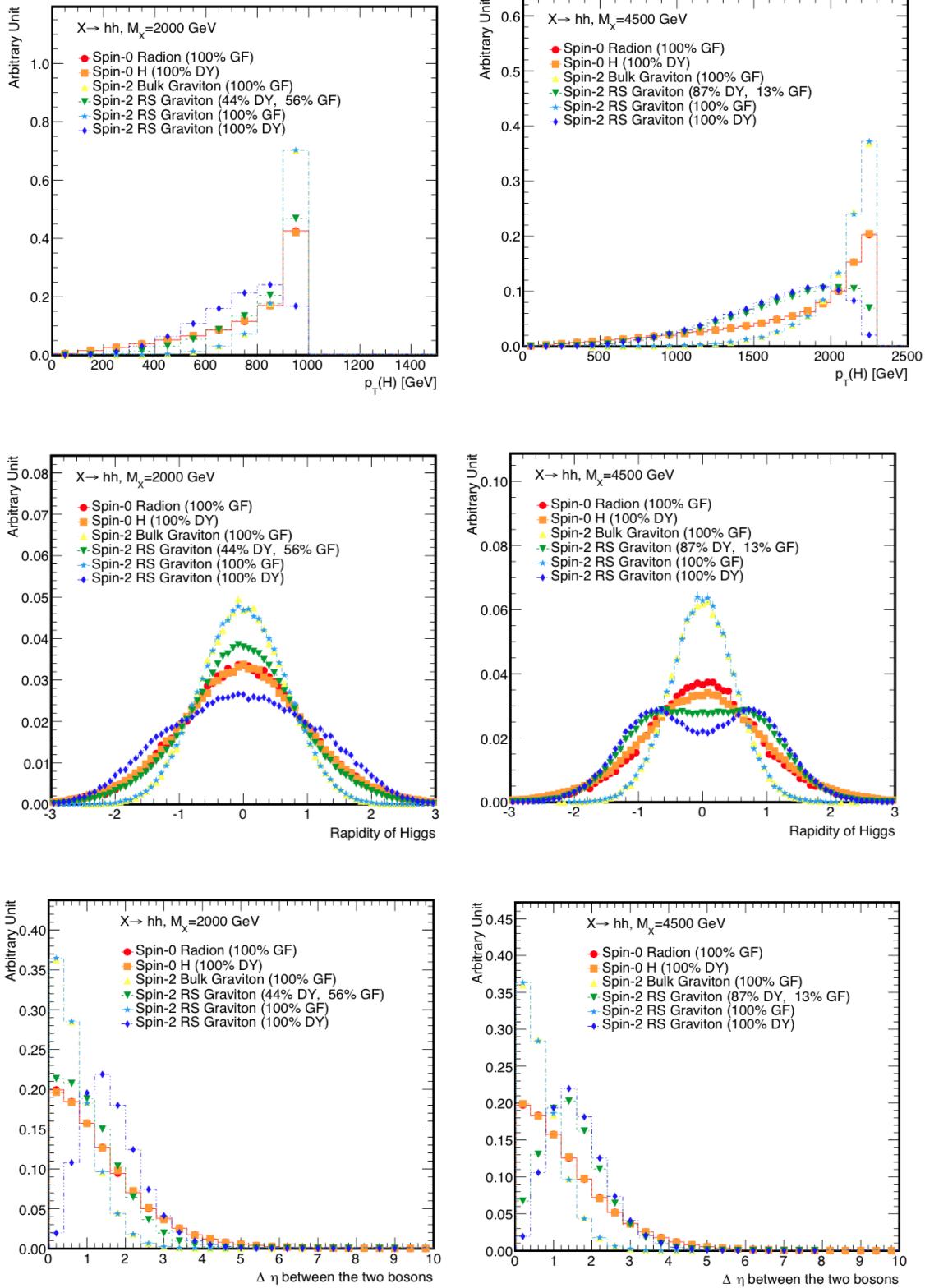


FIGURE A.2:  $p_T(H)$  (top), rapidity of H (middle), and  $\Delta\eta_{HH}$  (bottom) of heavy resonance models for  $M_X = 2000 \text{ GeV}$  (left) and  $M_X = 4500 \text{ GeV}$  (right). Figures from Reference [93].

## Appendix B

# Study of Neutrinos inside AK8 b-jets

In Figure B.1, it was observed that there was a shift towards left, around 5 GeV, on the pruned and soft drop<sup>\*†</sup> mass of the sub-leading jet compare to the leading jet. The difference is suspected to be caused by the contribution of neutrinos on the energy response of a heavy flavor jet since the b-hadrons decays semileptonically, 20-25% of the time, to a charged lepton, a neutrino, and a charm hadron. The charm hadron can further semileptonically decay into a charged lepton, a neutrino, and a strange hadron. However, the Jet Energy Calibration (JEC) does not take into account the difference between the jet flavors so there is a drop on the mass and  $p_T$  of the sub-leading jet.

The study is done by obtaining two leading AK8 jets with the following selections:

- $p_T > 300 \text{ GeV}$
- $|\eta| < 2.5$
- Loose Jet ID requirement
- $0.605 < \text{CISVV2}_{AK8\text{Jet}} < 1.0$

Afterwards, the two leading jets are matched with generator-level neutrinos, originating from a charm or a b hadron, within  $\Delta R < 0.8$ .

---

<sup>\*</sup>Soft drop is a jet grooming technique that depends on two parameters; a soft threshold  $z_{cut}$  and angular exponent  $\beta$  [94].

<sup>†</sup>In this study, the pruned and soft drop are uncorrected though the 4-momentum of the AK8 jets are corrected with CMS official jet energy calibration (JEC).

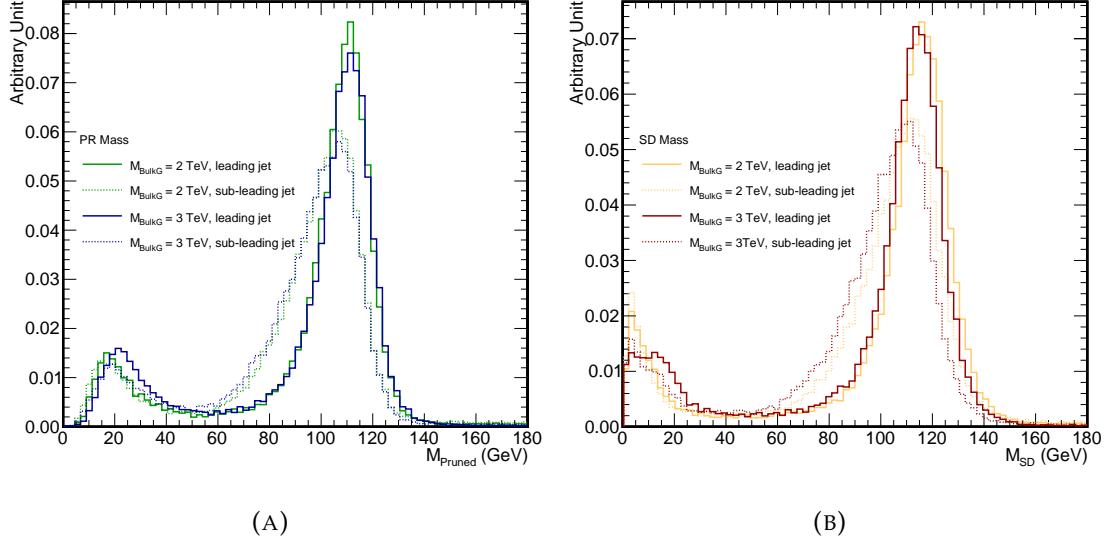


FIGURE B.1: (A) Pruned (blue for  $M_X = 2 \text{ TeV}$  and green for  $M_X = 3 \text{ TeV}$ ) and (B) Soft drop (orange for  $M_X = 2 \text{ TeV}$  and red for  $M_X = 3 \text{ TeV}$ ) mass distribution for leading (solid line) and sub-leading jet (dotted line). The sub-leading jet is approximately 5 GeV lower than the leading jet.

The Figures in B.2 show the average vector-summed neutrino  $p_T$  as a function of the pruned mass for 1, 2, 3.5, and 4 TeV Bulk Graviton signal samples. It can be observed that the neutrino has higher contribution on the sub-leading jet than on the leading jet. Moreover, the contribution of neutrinos on the sub-leading tends to go higher as the mass points goes higher. Same observations can be seen on the Figures in B.3 where instead of pruned mass, the figures are now as a function of the jet  $p_T$ .

In conclusion, the shift on the pruned and soft drop mass distribution of the subleading jet in Figure B.1 are indeed caused by the neutrinos inside the  $b$ -jets.

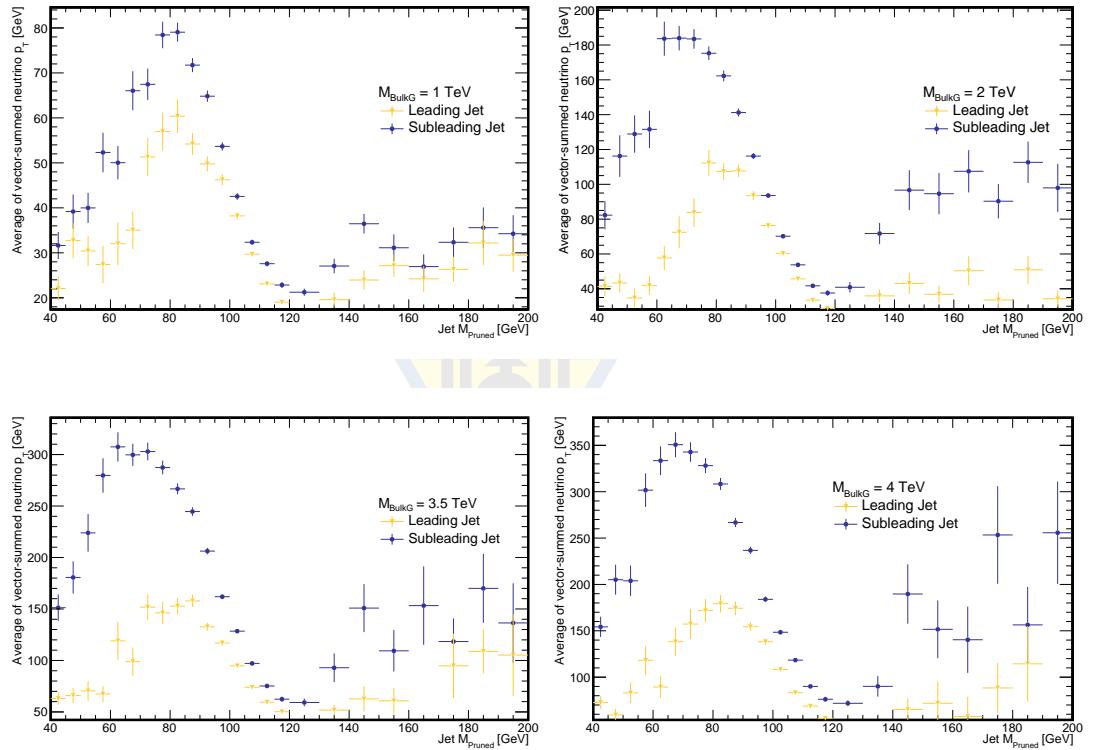


FIGURE B.2: The average vector-summed neutrino  $p_T$  as a function of pruned mass of leading (yellow triangle markers) and sub-leading (blue circle markers) for 1 (top-left), 2 (top-right), 3.5 (bottom-left), and 4 (bottom-right) TeV for Bulk Graviton signal samples.

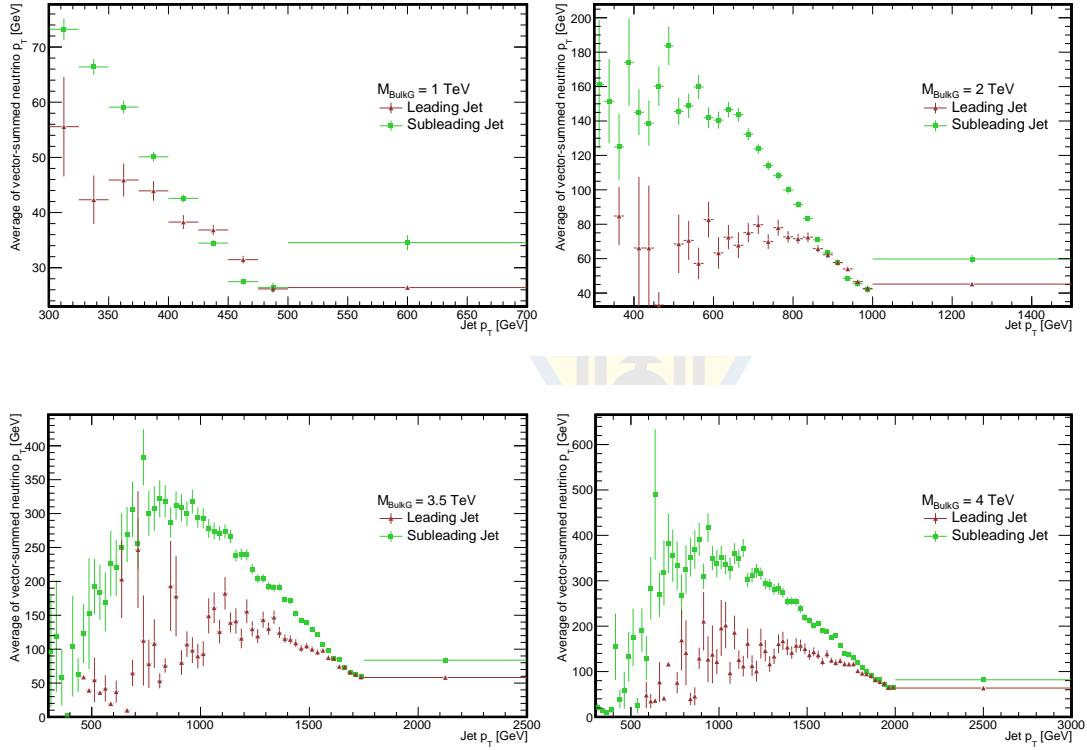


FIGURE B.3: The average vector-summed neutrino  $p_T$  as a function of jet  $p_T$  of leading (red triangle markers) and sub-leading (green square markers) for 1 (top-left), 2 (top-right), 3.5 (bottom-left), and 4 (bottom-right) TeV for Bulk Graviton signal samples.