# ABSTRACT

## STUDIES OF VECTOR BOSON SCATTERING IN THE SEMILEPTONIC $ZV$ CHANNEL WITH THE CMS DETECTOR AT $\sqrt{s} = 13$ TEV

Ramanpreet Singh, Ph.D.
Department of Physics
Northern Illinois University, 2021
Vishnu Zutshi, Director

About Vector boson scattering (VBS)

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

SEPTEMBER 2021

# STUDIES OF VECTOR BOSON SCATTERING IN THE SEMILEPTONIC $ZV$ CHANNEL WITH THE CMS DETECTOR AT $\sqrt{s} = 13$ TEV

BY

RAMANPREET SINGH


A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

Dissertation Director:
Vishnu Zutshi

# ACKNOWLEDGEMENTS

Acknowledge people here.

# DEDICATION

To my family.

# TABLE OF CONTENTS

Chapter Page

# LIST OF TABLES

# LIST OF FIGURES

Figure Page

# LIST OF APPENDICES

# CHAPTER 1

# INTRODUCTION

# CHAPTER 2

# THE STANDARD MODEL

Introduction to Standard Model (SM), Effective Field Theory (EFT), Vector boson scattering (VBS), etc.

## 2.1 Standard model

Introduction to SM [1] goes here.

### 2.1.1 Fundamental Particles

About fundamental particles

#### 2.1.1.1 Fermions

About fermions

# CHAPTER 3

# THE LHC AND CMS EXPERIMENT

The physics analysis is carried out using Compact Muon Solenoid (CMS) experiment at European Council for Nuclear Research (CERN) Large Hadron Collider (LHC) accelerator. This chapter provides overview of LHC and detail of CMS experiment and its sub-detectors for particle tracking and calorimetry.

## 3.1 The Large Hadron Collider

The LHC is the largest accelerator located at CERN in Geneva, Switzerland. The main LHC ring is 27 km in circumference and around 50 to 175 m underground. The LHC is built to collide protons at 14 TeV center-of-mass energy, LHC delivered proton-proton collisions at 7 and 8 TeV during run-1 (2010–2012), and at 13 TeV center-of-mass energy during run-2 (2015–2018) [2].

The Figure 3.1 describes CERN accelerator complex. The protons are sourced by ionizing hydrogen atoms and then fed into linear accelerator (LINAC). The LINAC accelerates the protons to 50 MeV and sent to the booster. Then the booster increases energy of protons to 1.4 GeV and feeds it to the proton syncrotron (PS) which further increases energy to 25 GeV and starts bunching them together with bunches 25 ns apart. Then the proton bunches are passed through super proton synchrotron (SPS) which increases energy to 450 GeV and finally sent to main LHC clockwise and counterclockwise rings where they are accelerated to final energy required which is 6.5 TeV for both bunches going clockwise and counterclockwise to obtain collisions at 13 TeV center-of-mass energy.

The proton-proton collisions occurs at four different location where two general purpose detectors CMS and A Toroidal LHC Apparatus (ATLAS), and two specific purpose detector A Large Ion Collider Experiment (ALICE) and LHC-beauty (LHCb) are located.

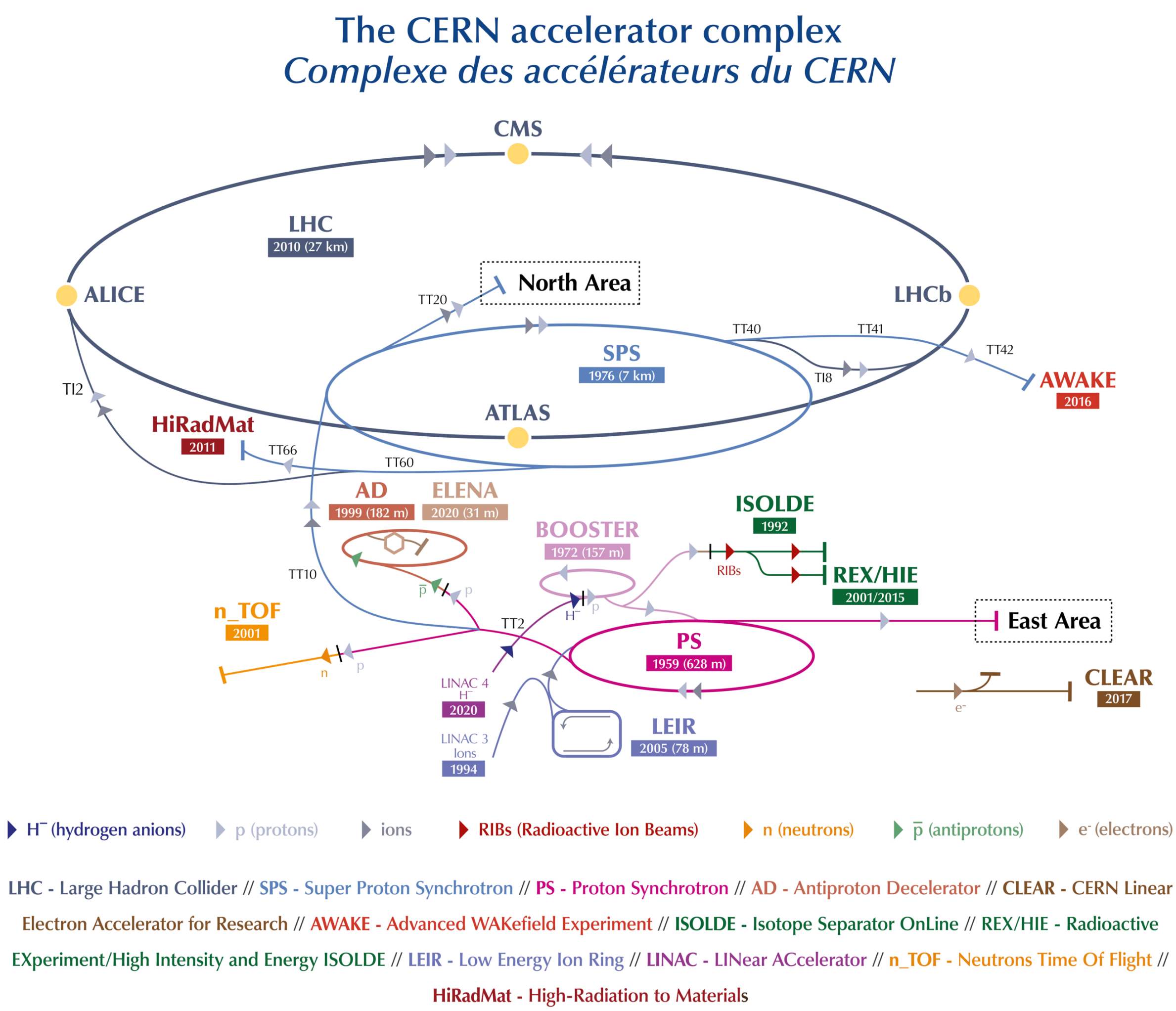


Figure 3.1: A schematic of the CERN accelerator complex [3]

### 3.1.1 Integrated Luminosity

The number of events generated in a collisions for a given process is

$$N = L\sigma \tag{3.1}$$

where $\sigma$ is cross-section of the process and $L$ is the luminosity of the LHC.

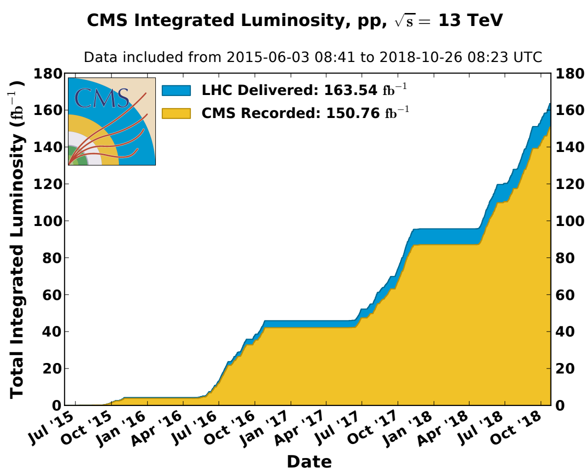


Figure 3.2: Cumulative delivered and recorded luminosity versus time for 2015–2018 proton-proton collisions [4]

Cumulative luminosity delivered and recorded by CMS during run-2 operation in shown in Figure 3.2. For run-2 standard physics analysis luminosity recorded during 2016–2018 is considered, and only runs certified as "golden" by CMS Luminosity physics object group (POG) for analysis are used. The total luminosity for run-2 standard physics is $137.19\,\mathrm{fb}^{-1}$ and separately for years in Table 3.1 [5–7].

Table 3.1: Standard physics luminosity for run-2

| 2016 | 2017 | 2018 | run-2 |
|---|---|---|---|
| $35.92\,\mathrm{fb}^{-1}$ | $41.53\,\mathrm{fb}^{-1}$ | $59.74\,\mathrm{fb}^{-1}$ | $137.19\,\mathrm{fb}^{-1}$ |

## 3.2 The CMS Detector

The CMS detector is a general purpose detector. A cutaway view of the detector is shown in Figure 3.3. The detector is cylindrical with dimensions 21 meters long, and 15 meters in diameter, and the whole detector weighs about 14000 tonnes. The detector is built in slices with central region called "barrel", and two closing end sides called "endcap". A superconducting solenoid generates magnetic field of 3.8 T inside and 2 T outside, and to contain the magnetic field outside of solenoid and support structure of the detector massive steel yokes are used.

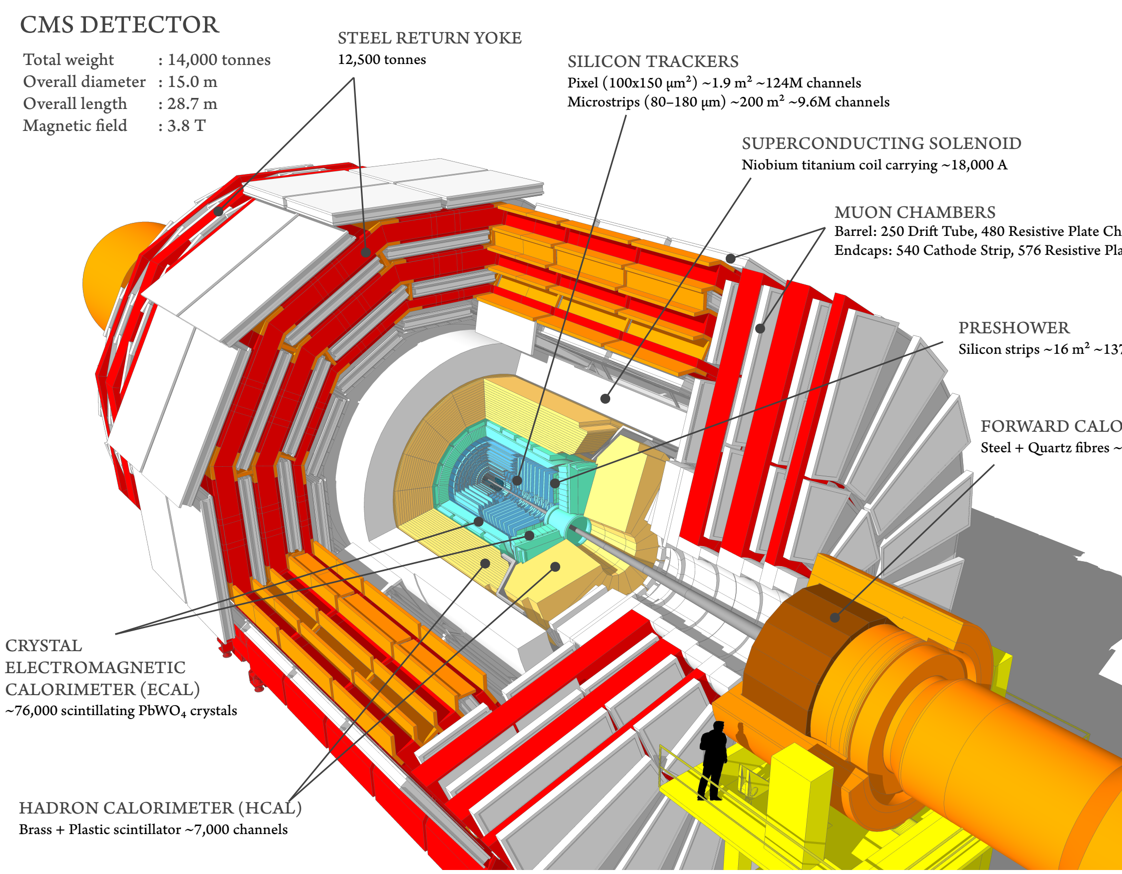


Figure 3.3: The CMS detector cutaway view [8]

The slice view of CMS in Figure 3.4 shows how different particles leave signature in CMS detector. Neutral particles such photons, neutrinos, and hadrons will leave no track in Silicon Tracker (ST), and are identified by only energy deposited or missing energy. Electrons are identified from the track in ST and energy deposit in Electromagnetic Calorimeter (ECAL), hadrons are heavier and they pass through ECAL and deposit their energy completely in Hadronic Calorimeter (HCAL), leaving only small fraction of energy in ECAL. Since muons are minimum ionizing particle (MIP), they pass through whole detector with very small fraction of energy deposit in ECAL and HCAL.

This section describes the subsystems of CMS detector. For detailed technical description refer to [9].

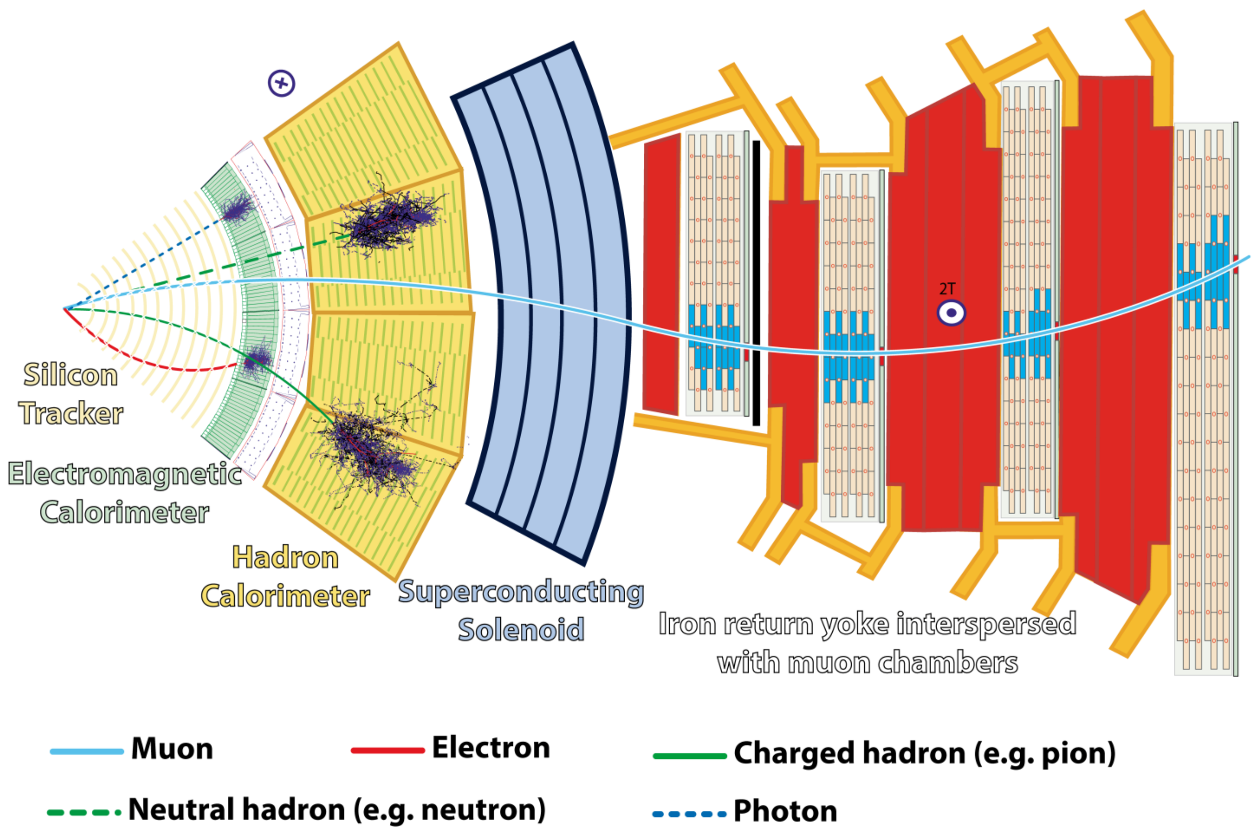


Figure 3.4: The CMS detector slice view [10]

### 3.2.1 The CMS Coordinate System

CMS uses interaction point (IP) of collisions as origin to define right-handed coordinate system. The $z$-axis is along the beamline, the $x$-axis points toward the center of the LHC, and the $y$-axis points upwards, toward Earth's surface. The transverse plane $x-y$ is used as to calculate most commonly used quantities like transverse momentum $p_T$ and energy $E_T$.

To describe the direction of particles leaving the IP, azimuthal $\phi$ and polar $\theta$ angles are used. $\phi$ is measured around the beam axis, and $\theta$ is measured from the beam axis. In collider physics, pseudorapidity $\eta$ (Lorentz invariant) is used to describe direction from beam pipe instead of $\theta$ as,

$$\eta = -\ln[\tan\theta/2] \tag{3.2}$$

and sometimes in terms of rapidity $y$ as,

$$y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z} \tag{3.3}$$

Particles kinematics can be completely described in terms of $p_T$, $\eta$, $\phi$, and $E_T$ or mass. The distance between the two particles $\Delta R$ in $\eta-\phi$ plane is described as,

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.4}$$

### 3.2.2 The Superconducting Magnet

The superconducting magnet is the main part of the CMS detector, it is 12.5 meters long and 6.3 meters in diameter. The magnet is cooled to 4.5 K and 20 kA current flows through it to generate 3.8 T of magnetic field with stored energy of 2.6 GJ.

The Figure 3.5 shows visible superconducting magnet and iron yoke when part of CMS detector was lowered in the underground cavern during installation in 2007.

The key purpose of magnet is to determine the momentum and the sign of charged particles by bending them. The momentum resolution of the particles will decrease with increase in $p_T$, with constant 3.8 T magnetic field inside and it has momentum resolution of $\Delta p/p \approx 10\%$, which is enough to determine unambiguously the sign of muons with momentum of $\approx 1\,\mathrm{TeV}/c$.

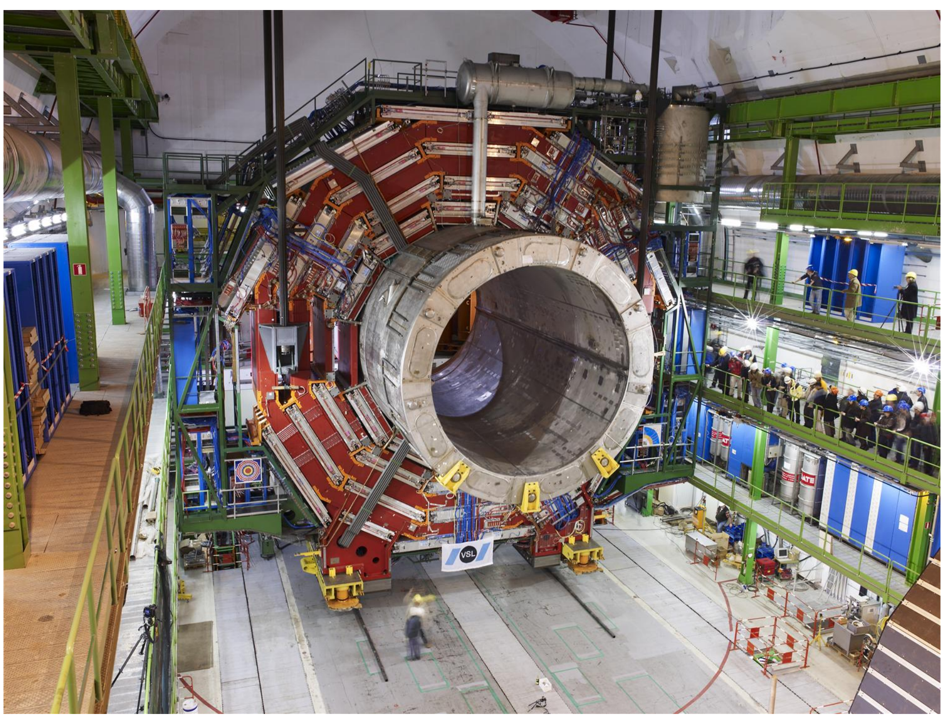

Figure 3.5: The picture of the CMS detector central part when lowered in underground cavern with superconducting magnet and iron yoke visible [11].

### 3.2.3 The Tracking System

The CMS tracking system ST is the innermost part of the detector, it is made up of pixel and strip detectors. The main goal of ST is to reconstruct the tracks of the charged particles with high precision in high pileup environment.

Silicon is most commonly used material for making tracking systems because of it's semiconductor properties, and high radiation hardness which is essential for the innermost detector. When a p-n junction is built on silicon substrate it creates a depletion zone with no charge carriers at the junction, and whenever a charged particle pass through the depletion zone it creates a electron-hole pair, and under reverse bias this electron-hole generates electrical signal. The CMS tracking consists of about 124 million channels of such junctions in pixel detector and 10 million in strip detector.

The pixel detector was upgraded in 2017 and the comparison of layers before and after the upgrade is shown in Figure 3.6. It is made up of four barrel layers and three endcaps, with nearest barrel layer being 3 cm away from beamline for precise measurement of IP. Because of the large number of pixel channels, the readout is done by Application-specific integrated circuits (ASICs).

The outermost part of ST detector is made of silicon strips. It allows large coverage by reducing number of readout channels. It has 10 layers in barrel region and 12 discs in endcap region. For better signal-to-noise ratio and radiation tolerance both pixel and strip operates at -20 °C.

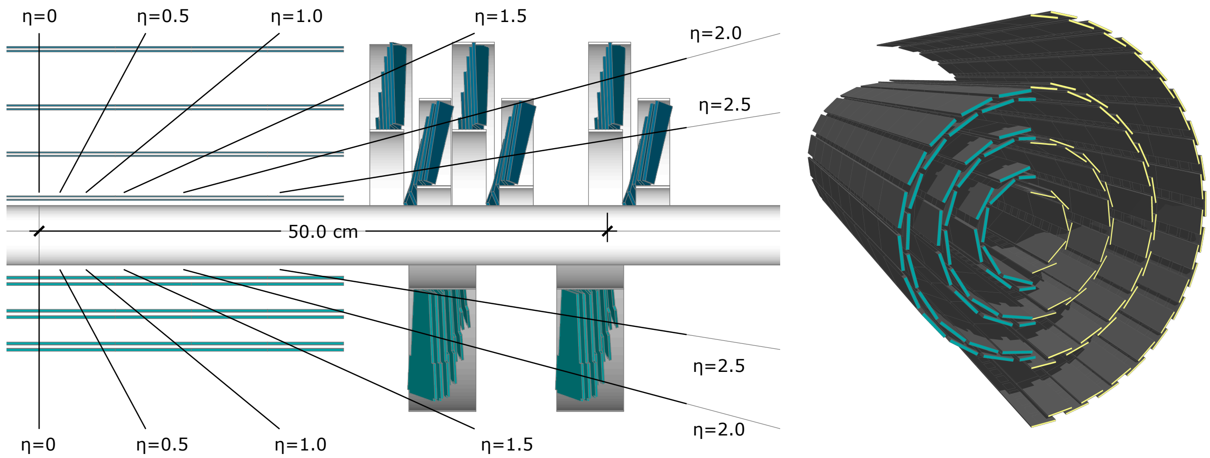


Figure 3.6: The CMS pixel upgrade. The left is cross sectional view of pixel detector layers before upgrade (bottom) and after Phase 1 upgrade (top). The right is view pixel barrel before upgrade (left) and after upgrade (right) [12].

### 3.2.4 The Electromagnetic Calorimeter

The ECAL active material is made of lead tungstate ($PbWO_4$) scintillating crystals and two layers silicon strip for preshower in front of the endcaps. The crystals in central barrel section are mounted in quasi-projective geometry pointing towards IP and covers $|\eta| < 1.48$, and two endcaps extends the coverage to $|\eta| = 3.0$. The schematic layout of ECAL is shown in Figure 3.7 and the picture of endcap quadrant when assembled in Figure 3.8

The main purpose of ECAL is to determine energy and positions of electromagnetically interacting particles. To determine particle need to completely deposit their energy, except electron and photons all other particles pass through ECAL crystals with only small fraction of energy signature in crystals. When electron and photon interacts with $PbWO_4$ it starts the process of electromagnetic shower and continues until the energy the energy of the incident particle is below threshold, which is about 1 MeV.

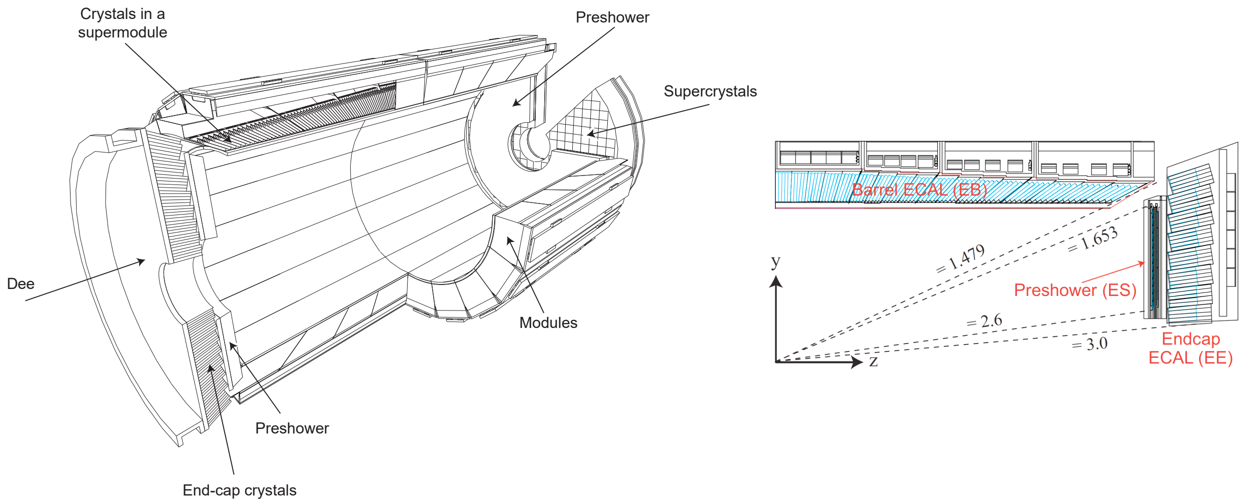


Figure 3.7: The CMS ECAL schematic layout. The left is schematic showing arrangement of superclusters in barrel and endcap (with preshower layers). The right is $y - z$ plane quarter view of ECAL layout [13].

The resolution of the ECAL energy measurements can be described as,

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.5}$$

where $S$ is the stochastic term, $N$ is related to the noise, and C is a constant offset.

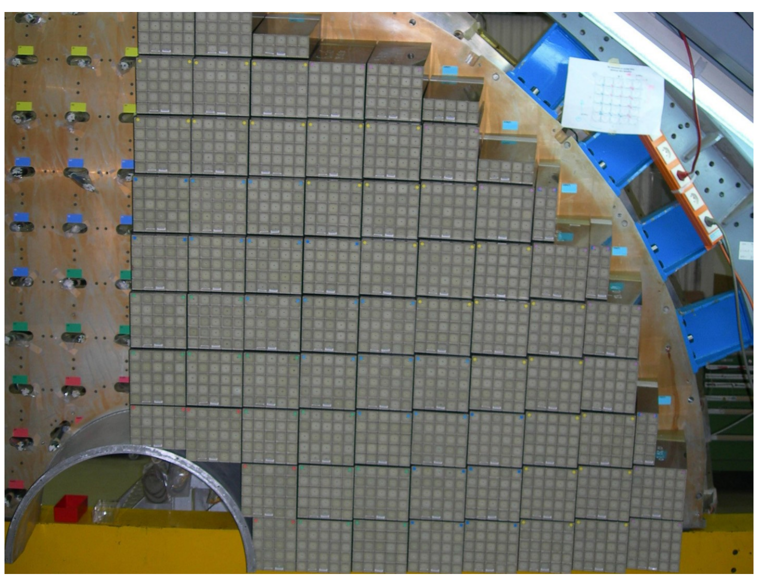

Figure 3.8: The ECAL endcap quadrant assembled view [14].

### 3.2.5 The Hadronic Calorimeter

HCAL is last subdetector inside solenoid after ECAL and the first half of barrel HCAL inserted is shown in Figure 3.9. Similar to ECAL the purpose of HCAL is to shower hadrons, and measure their energy and position. HCAL is made up of towers pointing towards IP and each tower is made up of sampling layers with alternating layers of plastic scintillator and brass. Brass acts as absorber in HCAL and causes hadrons to shower, then from the light output of scintillator receiving secondary shower particles gives the amount of energy deposit in each layer. In phase 1 upgrade the HCAL was upgraded to give energy deposit as function of depths and the depth segmentation schematic is show in Figure 3.10 and the details of upgrade are in technical design report [15].

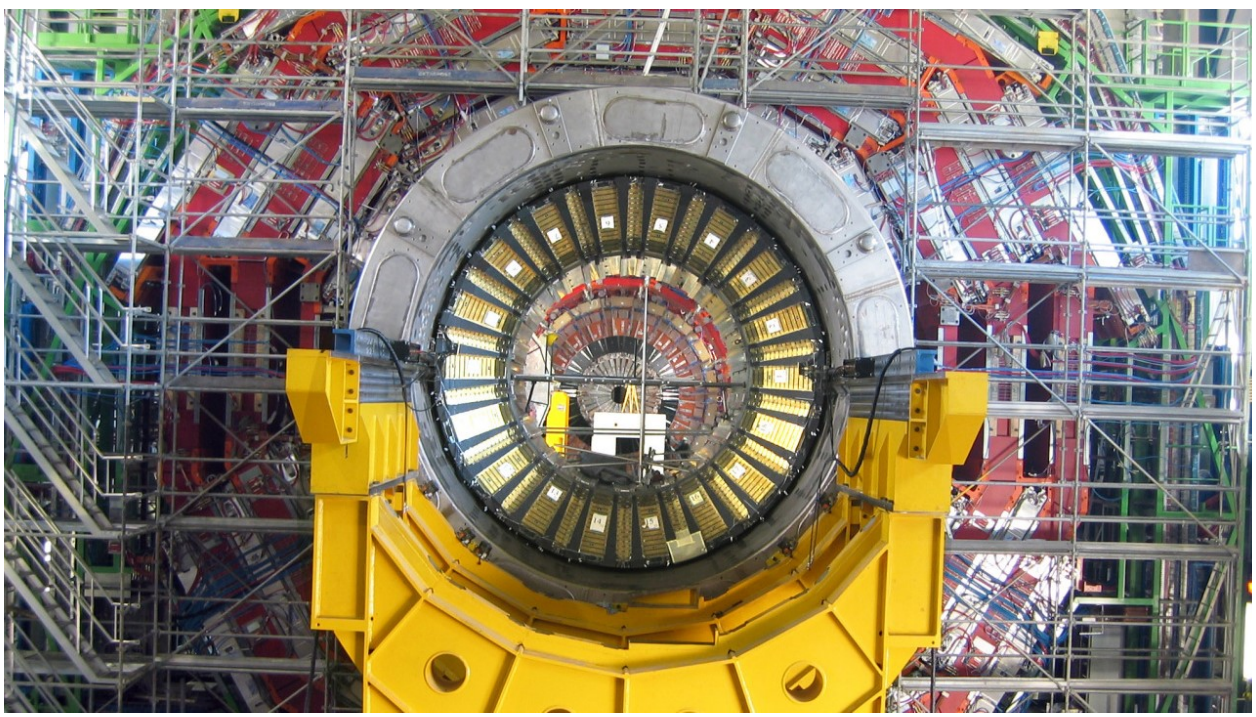

Figure 3.9: The first half of the barrel HCAL inserted into the superconducting solenoid (April 2006) [16].

HCAL consists barrel (HB) and two endcaps (HE) located inside solenoid. These two subsystems combined cover region $|\eta| = 3.0$, which is most of physics analysis done in CMS. There are two another subsystems of HCAL outside solenoid, a forward HCAL (HF) and outer barrel HCAL (HO). HO was added to ensure there is no leak from the particles that make past the solenoid. HF extends the coverage to $|\eta| = 5.0$ and is based Cherenkov radiation principle unlike other subsystems of HCAL, and it uses quartz fiber as active material with steel absorbers. HF is used most commonly used by heavy ion analysis.

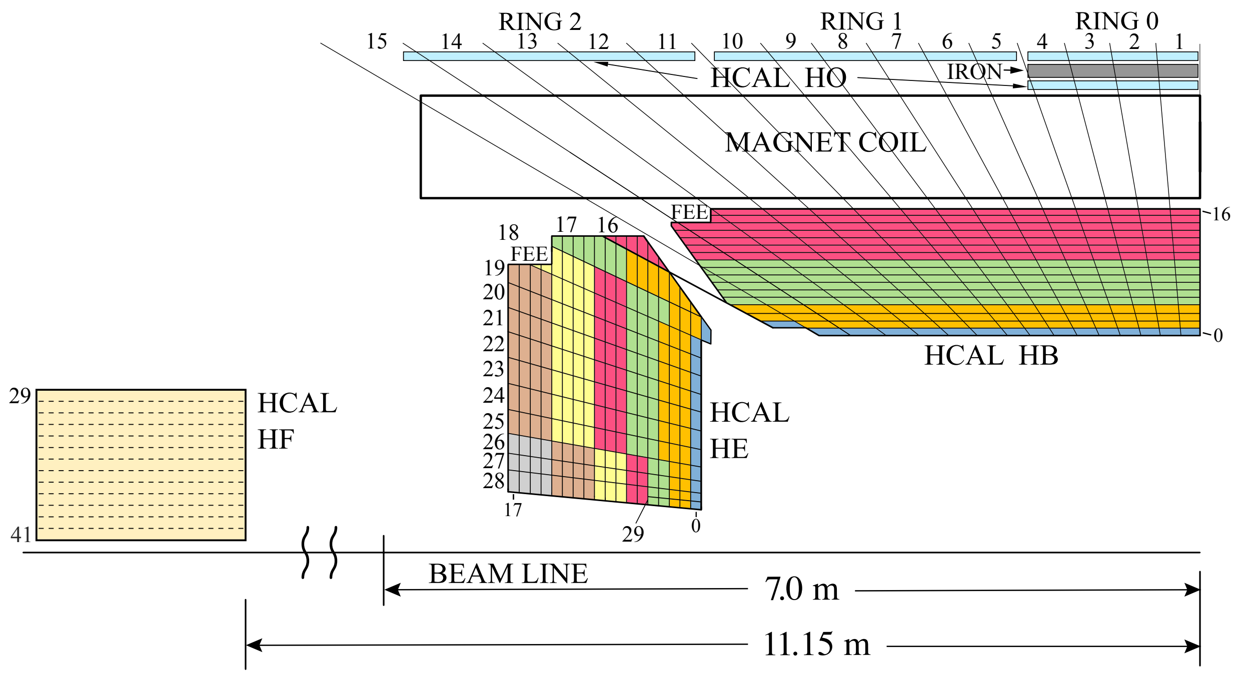


Figure 3.10: The HCAL depth segmentation after phase 1 upgrade [17].

### 3.2.6 Muon Detector

The outermost subsystem in the CMS detector is the muon detector. Unlike electrons, muons are MIPs, they do not lose much of their energy while passing through tracker, calorimeter and solenoid. Muon detector is build to identify, measure momentum and trigger the events with muons. Like other subsystems, muon detector consists of barrel and endcap detector and schematic layout is highlighted in Figure 3.11.

The muon detector consists of three subsystems drift tubes (DTs), cathode strip chambers (CSCs) and resistive plate chambers (RPCs).

The DTs are wire gas detectors filled with Argon and composed of many tube cells of about 4 cm. Muon passing through these tubes ionizes Argon and free electron is detected with wires as cathode. Each DT is about 2 meters by 2.5 meters in size, and there are four layers of the DTs interleaved with iron yoke parallel to the beam pipe in barrel region. The drift time is the order of about 380 ns.

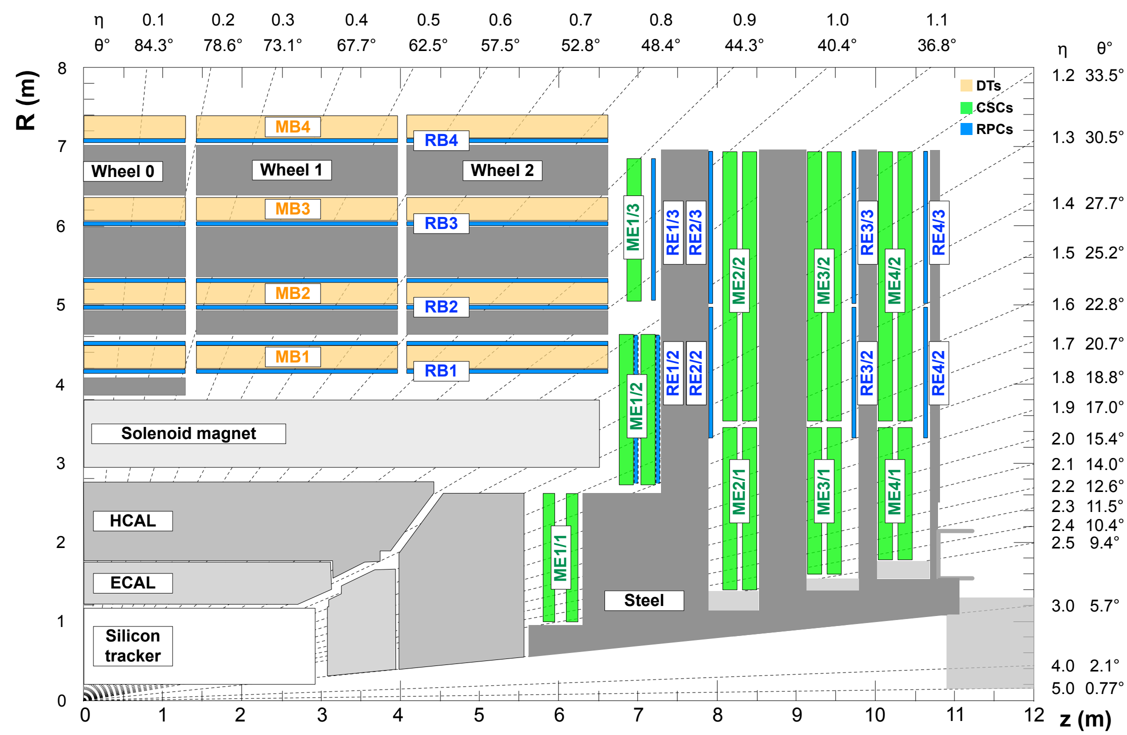


Figure 3.11: The quadrant view of CMS subdetectors layout, and the coverage of the muon detector DTs, CSCs, and RPCs highlighted [18].

The CSCs are based on same principle as DTs, and are made of multi-wire proportional chambers consisting of 6 anode planes interleaved with 7 cathode planes. They have time resolution smaller than 5 ns. The CSCs are used in endcap region, where radiation hardness is required, and non uniform magnetic field does neutrinos effects the the measurement.

The RPCs are made up of two high resistive parallel plates, with oppositely charged plates and gas volume between them. When a charged particles passes through it and ionizes the gas, it creates an avalanche and charge is collected by metallic readout strips. RPCs have poor position resolution but fast readout of the order of 1 ns, which is fast compared to DTs, this is the reason there are 1 or 2 RPCs attached to both DTs, and CSCs.

### 3.2.7 Level 1 Trigger

Since proton-proton collisions happens every bunch crossing which is 25 ns apart, which is equivalent to 40 MHz collisions rate. At this collisions rate, the data storage required will be enormous and CMS can only record up to 1000 events per second. Since most of events does not contain interesting physics events, they can be thrown away. To do this CMS has two tier trigger system Level-1 Trigger (L1T), and High Level Trigger (HLT).

The L1T is the foremost electronic processing system through which event information is processed before it is passed to second trigger system HLT. The L1T is designed to make fast decisions in about 3.8 $\mu$s, and only uses ECAL, HCAL and muon system to make decision. L1T cut downs the data rate from 40 MHz to 100 kHz. The L1T electronics is placed next to the detector in underground cavern for fast transfer of data.

The HLT further reduces the data rate from 100 kHz to about 1 kHz using a computer farm with nearly 26000 cores. HLT uses all the available information from the event to make decision in about 300m s. HLT is modular by design to allow the use of information from different systems to construct multiple paths called HLT paths, for example the single muon

HLT path will save event with at least one muon passing the selection criteria set in HLT path. Events passing at least on HLT path are save for offline physics analysis.

## 3.3 High Granularity Calorimeter Upgrade

About High Granularity Calorimeter (HGCAL) Upgrade

### 3.3.1 Technical Design

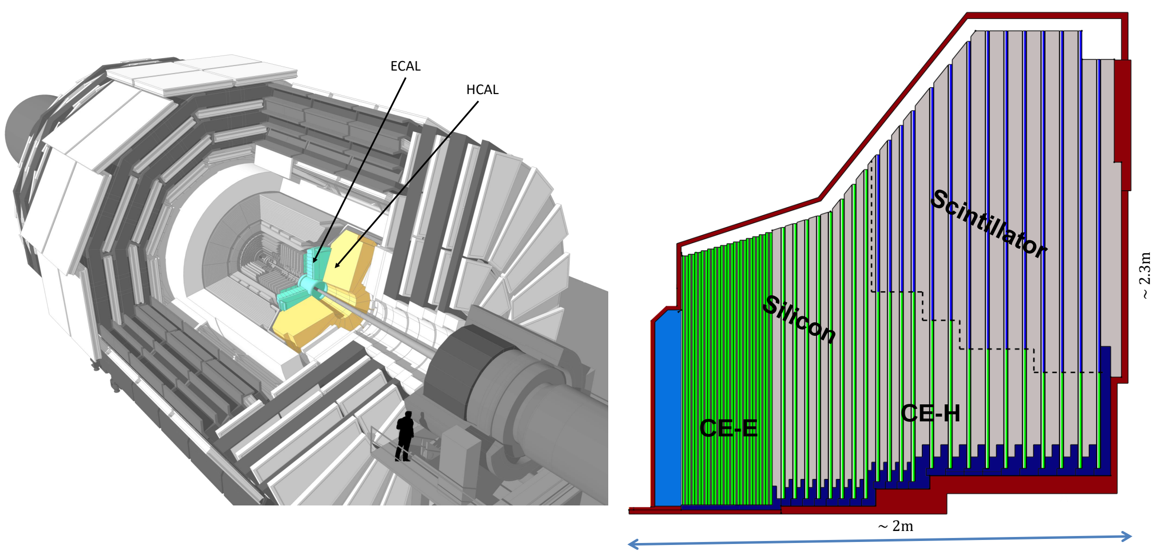


Figure 3.12: Overview of CE [19, 20]

### 3.3.2 Scintillator Tiles

### 3.3.3 End of Life Scenario

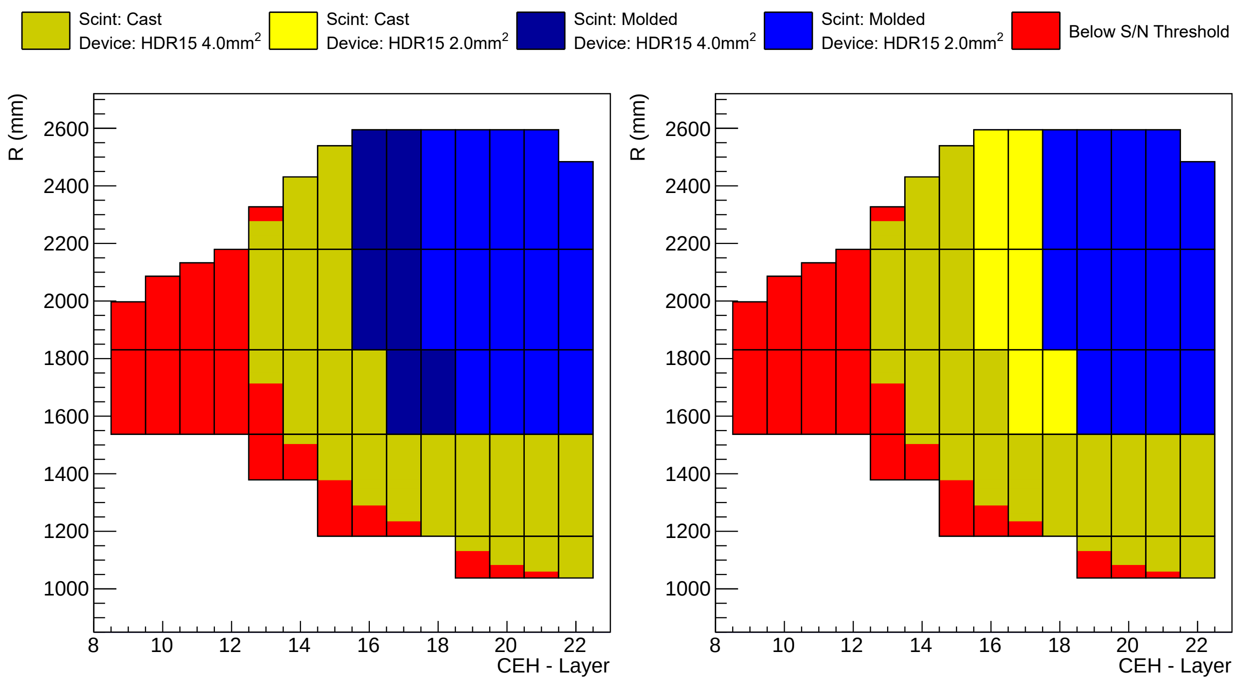


Figure 3.13: HGCAL scenarios

Table 3.2: HGCAL scenarios comparison

| | | Scene A | Scene B |
|---|---|---|---|
| Cast Scintillator | Cell Count | 148, 608 | 176, 256 |
| | Total Area | 185.13 m$^2$ | 239.45 m$^2$ |
| | Percentage | 50.6 % | 65.5 % |
| Injection Molded Scintillator | Cell Count | 91, 008 | 63, 360 |
| | Total Area | 180.5 m$^2$ | 126.18 m$^2$ |
| | Percentage | 49.4 % | 34.5 % |
| SiPMs Count | 2 mm$^2$ | 63, 360 | 91, 008 |
| | 4 mm$^2$ | 100, 224 | 148, 608 |

# CHAPTER 4

# EVENT SIMULATION AND RECONSTRUCTION

The proton-proton collision at LHC produces shower of particles, before the event information can be easily used in an analysis, the data collected goes through iterative process of reconstruct particles produced in collision. CMS uses particle flow (PF) algorithm to reconstruct 4-vectors of muons, electrons, photons, hadrons, jets and missing transverse momentum [21].

To analyze the data collected and compare it with theoretical model, events are simulated using Monte Carlo (MC) event generators and are passed through detector simulation and PF so that MC events can be treated same as real events.

This chapter describes the basic ingredients for object reconstruction, PF candidates and MC event generators used in this analysis.

## 4.1 Track Reconstruction and Calorimeter Clustering

For complete particle reconstruction two main ingredients are tracks left by particle in the detector, and energy deposit in calorimeter. This section describes track reconstruction from the hits in Tracker and Muon Detector, and energy deposit measurement from calorimeter clustering.

Track reconstruction requires reconstructed hits, and seed generation which are described in [22], then the track reconstruction is done using pattern recognition which is based on combinatorial Kalman Filtering (KF) method [23]. It is an iterative process starting from seed layer the track is estimated and then proceeds to next layer one by one, at each successive layer the track trajectory is better known. There can be multiple hits in each new layer, for this multiple trajectory candidates are created. All the trajectory candidates are grown in parallel to avoid bias, and truncated at each layer to prevent exponential increase in number of candidates. Then finally the track is fitted to compute momentum and vertex information.

The main purpose of calorimeter clustering is to determine position and energy deposit of the particle. A cluster in a calorimeter is a local group of energy deposits that are spatially consistent with a electromagnetic or hadronic shower. First the topological clusters are identified, a topological cluster is a contiguous region of energy deposit, then a seed is identified in topological cluster with certain energy threshold, and highest among the 8 neighbors for ECAL and 4 neighbors for HCAL. Now starting with seed energy and position, the neighbors energy are added, and new position is calculated. For the case when we have just one seed, there is only iteration until all the neighbors are added, in case when we have more than one seeds in a cluster, the energy from neighbors is shared, the fraction of energy shared depends on the energy and position of the cluster, after the first iteration of calculation of the energy and position the process is repeated with new values of cluster's energy and position until either the maximum iteration is reached or cluster's energy and position values are converged.

## 4.2 Reconstructed Particles

After tracks and calorimeter clusters are formed, PF links this information from the detectors together to form objects as broadly discussed in Section 3.2 and shown in Figure 3.4. This section describes the properties of those reconstructed particle candidates.

### 4.2.1 Muons

Reconstructing muon with best precision is the key ingredient for many physics searches. Muons reconstruction and identification uses all the information from tracker, calorimeters and muon detector. There are two types of reconstruction performed “Global” and “Tracker” for muon candidates. Global muons are formed combining and refitting muon hits in the muon detector with compatible track from ST, and the tracker muons are formed by extrapolating tracks from ST to segments in muon detector.

Once the muon candidates are found, the kinematics properties ($p_T, \eta, \phi$) are calculated from track fitting, and other properties such as distance form primary vertex (PV) $dxy$, $dz$, number of hits in the tracker and muon system, tracker based relative isolation (4.1) in a cone of $\Delta R = 0.3$, and PF relative based isolation (4.2) in a cone of $\Delta R = 0.4$ are stored for cleaning and isolating muons for physics analysis.

The tracker and PF based relative isolation are defined as,

$$\text{TkIso03} = \left(\sum p_T^{\text{Tracks (PV)}}\right) / \left(p_T^{\mu}\right) \tag{4.1}$$

$$\text{PFRelIso04} = \left(\sum p_T^{\text{CH (PV)}} + \min\left[0, \sum E_T^{\text{NH}} + \sum E_T^{\gamma} - 0.5 \sum p_T^{\text{CH (PU)}}\right]\right) / \left(p_T^{\mu}\right) \tag{4.2}$$

where “Tracks (PV)” refers to all the tracks in tracker and coming from PV, “CH (PV)” and “CH (PU)” refers to charged hadrons coming from PV and pile up (PU) respectively, “NH” refers to neutral hadrons, $\mu$ refers to muon, and $\gamma$ refers to photon.

There are multiple source of muons whenever collision event happens, they can be real muons or hadrons which are misidentified as muons, these hadrons are able “punch” through

HCAL and leaves hit in muon detector. The real muons of interest are called "prompt" muons and others are either usually referred as "fake" or "non-prompt". Fake muons can originate from decay of pions and kaons in flight usually identified with a "kink" in track or from heavy flavor decay of $b$ or $c$-quarks which are identified with tracks not originating form PV. The prompt muons are the ones coming from decay of $H$, $W$, $Z$ bosons and $\tau$ leptons, and have small impact parameter from PV, have hits in both tracker and muon detector, and are typically well isolated.

In addition to muons from collision events, there can be cosmic muons from pion decay in upper atmosphere. Cosmic muons are generally not in-time with collision and far from interaction points.

### 4.2.2 Electrons and Photons

Since there is large amount of material in tracker, electrons often emit bremsstrahlung photons when passing through tracker volume, and photons can further decay to $e^-e^+$ pair which complicates the tracking algorithm. The energy deposit of such electrons emitting bremsstrahlung will have large spread in $\phi$ direction because the magnetic field will bend electrons in $\phi$ whereas photons are unaffected. For this reason electron and photon reconstruction are done together, and the Gaussian-sum filter (GSF) algorithm is used for electron track reconstruction which takes care of kinks in electrons track because of hard emission [24].

An electron is reconstructed when an ECAL cluster matches a GSF track, and a photon is reconstructed when an ECAL cluster with $E_T$ more than 10 GeV is found and have no matching GSF track. To prevent electron and photon from being misidentified as jets certain conditions are applied, for electron the number of GSF track matching with ECAL cluster is limited to maximum of two, and energy deposit in a cone of $\Delta R = 0.15$ in HCAL around the position of electrons and photons is required to be less than 10%.

Similar to muons, after electron and photon reconstruction is done, their kinematics properties are calculated and various other properties required for cut based and Multivariate analysis (MVA) based identification are stored. The detailed description of electrons and photons identification technique and properties used in this dissertation can be found in Reference [25].

### 4.2.3 Hadrons and Jets

Quarks and gluons produced in a collision event are not detected directly, because of color confinement, they go through fragmentation and hadronization making a collimated spray of particles mostly made of hadrons and are called "jets". Charged hadrons are reconstructed when a HCAL cluster can be associated with one or more tracks, if the track association fails the cluster is reconstructed as neutral hadron.

Jet in CMS are reconstructed using FASTJET package [26], which takes input of all PF candidates and associated tracks. The clustering basically combines 4-vectors of particles iteratively and stop when distance between two particles ($d_{ij}$) is higher than stopping distance ($d_{iB}$).

$d_{ij}$ and $d_{iB}$ are defined as,

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.3}$$

$$d_{iB} = p_{Ti}^{2p} \tag{4.4}$$

where $p$ is the parameter for different clustering algorithms, $R$ is the cone size, and $\Delta R_{ij}$ is distance between two particles in iteration.

Anti-$k_T$ (AK) is the most use jet algorithm in physics analysis, this corresponds to $p = -1$, this means the hard particles will be clustered first in this clustering algorithm. The cone size used for standard jets (AK4) is $R = 0.4$, and for large jets (AK8) often called "fatjet" is $R = 0.8$.

To mitigate the effect of PU contamination in jets two most commonly used techniques are charged hadron subtraction (CHS) and Pileup Per Particle Identification (PUPPI) [27]. CHS as the name suggests removes all the PF in the jet clustering which are originating from PU vertices, and it is a standard technique for AK4 jets. PUPPI works by identifying PU in an event from charged PU information, then it assigns a weight to all the other particles inside jet, such as neutral particles, the weight is then used to rescale momentum of those particles. The main limitation of CHS is that it only removes charged PU contribution, for larger jets it can be issue, since it is clustering larger number of particles and can have significant contribution from neutral hadrons, for this reason PUPPI technique is used for AK8 jets.

To improve the jet selection and reject jets originating purely from PU two methods are used in this dissertation, jet identification based on multiplicities and energy fraction of particles contained in the jet, and MVA based PU identification which uses jets shape variables to discriminate prompt jet from pileup jets. The details of PU mitigation and identification used in CMS are described in Reference [28].

After jets reconstruction is complete the in addition to calculating kinematics properties ($p_T$, $\eta$, $\phi$ and mass), various other properties such as $b$-quark tagging and quark-gluon likelihood are also calculated and stored.

#### 4.2.3.1 N-Subjetiness and Deep Taggers

The origin of fatjets are usually when heavy energetic particle often referred to as "boosted" decays hadronically, for example boosted $W$ or $Z$ bosons decaying to a pair of quarks. To find and discriminate the fatjet of interest based on in it's substructure the two technique studied and used are N-Subjetiness [29] and "deep tagger" [30].

N-Subjetiness is defined as,

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\} \tag{4.5}$$

where $k$ runs over constituent particles in a jet, $\Delta R_{J,k}$ is the distance between subjet $J$ and $k$ constituent, and $d_0$ is the normalization constant defined as,

$$d_0 = \sum_k p_{T,k} R_0 \tag{4.6}$$

$\tau_N$ quantifies to what degree a jet can be regarded as made of $N$ jets. The small values of the $\tau_N$ means a jet is more likely to have $N$ or less subjets, and higher value means it will at least have $N+1$ subjets. Rather than using $\tau_N$ alone, ratio of different $\tau_N$ variables is used, which more discriminating for cases like $W$ vs QCD jets. Figure 4.1 shows distribution of $\tau_{21}$ and $\tau_{32}$ shapes in signal and background. $\tau_{21}$ is used to discriminate fatjets with 2-prongs ($W/Z/H$) and $\tau_{32}$ with 3-prongs ($t$ quark) substructure against QCD jets.

Deep Tagger for AK8 are Machine learning (ML) based tagger developed to determine origin of a fatjet. These taggers are trained on particle level information from PF and provide multi class tagging probabilities. In addition to there are version of these taggers is also which is de-correlated from the mass of jet, this is important for analysis including this

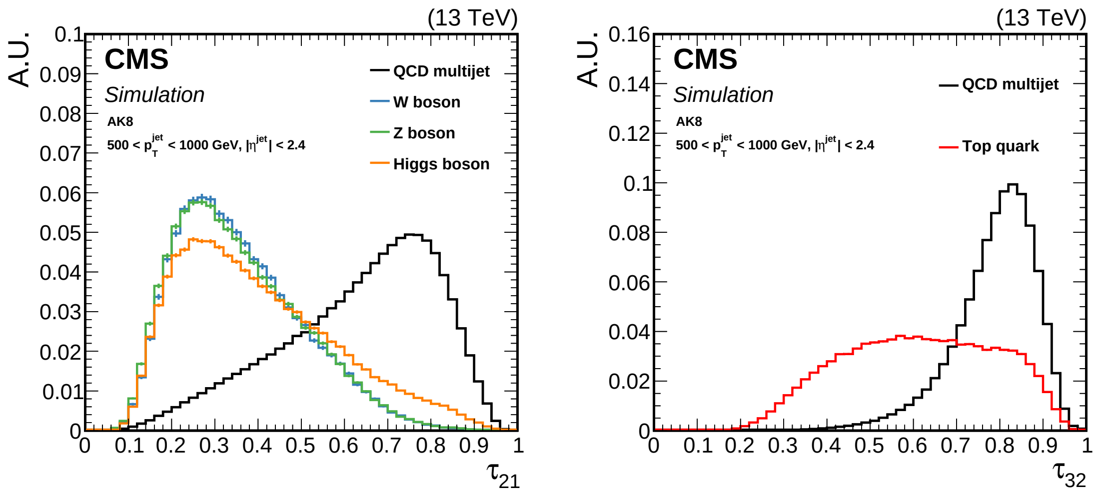


Figure 4.1: Comparison of $\tau_{21}$ and $\tau_{32}$ shapes for signal and background in AK8 jets. The left is $\tau_{21}$ distribution showing discrimination $W/Z/H$ jets vs QCD jet, and the right is $\tau_{32}$ distribution for $t$ quark vs QCD jets [30].

dissertation where we utilize mass regions of fatjet to normalize background contribution. Figure 4.2 describes the architecture of "DeepAK8" tagging.

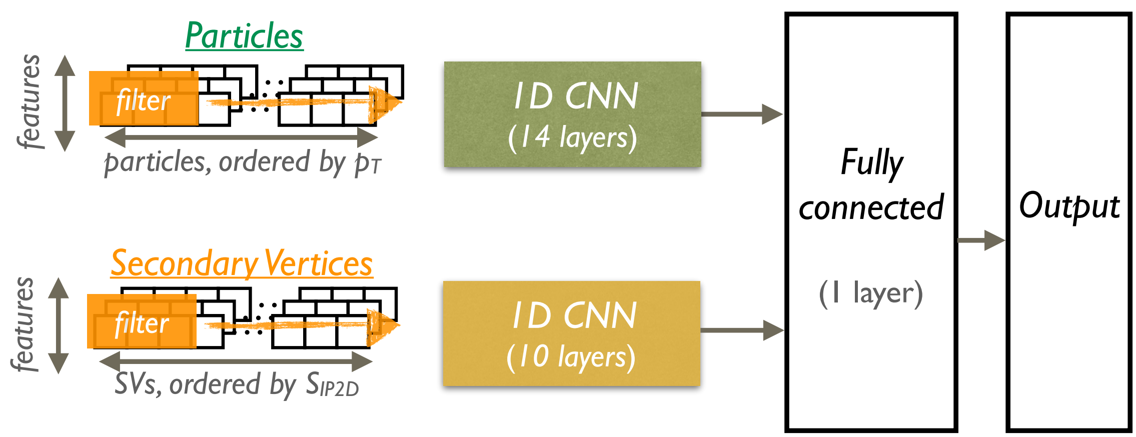


Figure 4.2: The network architecture of DeepAK8 [30]

#### 4.2.3.2 Softdrop Mass

Fatjets can also have contamination coming from wide angle soft initial state radiation (ISR) and multiple hadron scattering, which affects the mass calculation of the jet, to remove such contamination and have better mass reconstruction, the “softdrop” mass algorithm [31] is used.

Softdrop is a declusttering algorithm which removes the particle from the jet with radius $R_0$, when the following condition between two particles is satisfied,

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut} \left( \frac{\Delta R_{12}}{R_0} \right)^{\beta} \tag{4.7}$$

where $\Delta R_{12}$ is the distance between the two particles, $z_{cut}$ and $\beta$ are the parameters for tuning softdrop declusttering. For fatjets used in CMS, they have softdrop applied with $\beta = 0$ and $z_{cut} = 0.1$ which vetoes both soft and soft-collinear emissions in a jet.

### 4.2.4 Missing transverse momentum

Invisible particles like neutrinos cannot be detected at CMS directly. Kinematics of such particles can be determined using laws of conservation of total momentum. In case of proton-proton collision, the actual collision happens between quarks contained in proton and quarks carry fraction of proton momentum and can not be determined exactly, for this reasons kinematic determination of invisible particles is limited to transverse plane only.

After all the particles are reconstructed in an event, their $p_T$'s can be used to determine missing transverse momentum as,

$$\vec{p}_T^{miss} = -\sum \vec{p}_T \tag{4.8}$$

It's usually neutrinos which contributes to missing transverse momentum, and they have very small mass, the missing transverse momentum is then equivalent to Missing Transverse Energy (MET), which is most often used term in physics analysis.

## 4.3 Monte Carlo Simulation

### 4.3.1 Generators

### 4.3.2 Hadronization

# CHAPTER 5

# VBS MEASUREMENT IN $ZVJJ$ FINAL STATE

Analysis of VBS

## 5.1 Dataset and Simulation

### 5.1.1 Data

### 5.1.2 MC Simulations

## 5.2 Event Selection

### 5.2.1 HLT Trigger

### 5.2.2 Lepton Selection

#### 5.2.2.1 Muon

#### 5.2.2.2 Electron

### 5.2.3 VBS Tagged Jets

### 5.2.4 $V$ Jet Candidate

# CHAPTER 6

# RESULTS

# REFERENCES

[1] C. N. Yang and R. L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". *Phys. Rev.* 96 (1 Oct. 1954), pp. 191–195. DOI: 10.1103/PhysRev.96.191. URL: https://link.aps.org/doi/10.1103/PhysRev.96.191.

[2] L. Evans and P. Bryant. "LHC Machine". *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[3] E. Mobs. "The CERN accelerator complex - 2019" (July 2019). Accessed August 16, 2021. URL: http://cds.cern.ch/record/2684277/files/.

[4] Public CMS Luminosity Information. "Cumulative delivered and recorded luminosity versus time for run-2 pp data" (2018). Accessed August 17, 2021. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[5] The CMS Collaboration. "CMS Luminosity Measurements for the 2016 Data Taking Period" (2017). URL: https://cds.cern.ch/record/2257069.

[6] The CMS Collaboration. "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV" (2018). URL: https://cds.cern.ch/record/2621960.

[7] The CMS Collaboration. "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV" (2019). URL: https://cds.cern.ch/record/2676164.

[8] T. Sakuma. "Cutaway diagrams of CMS detector" (May 2019). Accessed August 17, 2021. URL: https://cds.cern.ch/record/2665537.

[9] The CMS Collaboration. "The CMS experiment at the CERN LHC". *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088/1748-0221/3/08/s08004.

[10] The CMS Collaboration. “CMS Detector Slice” (Jan. 2016). Accessed August 18, 2021. URL: https://cds.cern.ch/record/2120661.

[11] M. Brice and C. Marcelloni. “The central part of CMS is lowered” (Feb. 2007). Accessed August 22, 2021. URL: https://cds.cern.ch/record/1020311.

[12] T. Sakuma. “SketchUp Drawings of the Phase 1 pixel detector alongside the current pixel detector” (June 2012). Accessed August 17, 2021. URL: https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=6473.

[13] The CMS Collaboration. “The CMS ECAL performance with examples”. *Journal of Instrumentation* 9.02 (Feb. 2014), pp. C02008–C02008. DOI: 10.1088/1748-0221/9/02/c02008.

[14] The CMS Collaboration. “Images of CMS ECAL Endcap (EE)” (Nov. 2008). Accessed August 22, 2021. URL: https://cds.cern.ch/record/1431479.

[15] J. Mans et al. “CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter” (Sept. 2012). URL: https://cds.cern.ch/record/1481837.

[16] The CMS Collaboration. “Images of the CMS HCAL Barrel (HB)” (2008). Accessed August 22, 2021. URL: https://cds.cern.ch/record/1431485.

[17] The CMS Collaboration. “HCAL Depth Segmentation Phase 1 Upgrade” (2012). Accessed August 22, 2021. URL: https://home.fnal.gov/~chlebana/CMS/Phase1.

[18] The CMS Collaboration. “Performance of the CMS Muon Detectors in early 2016 collision runs” (2016). Accessed August 22, 2021. URL: https://twiki.cern.ch/twiki/pub/CMSPublic/MuonDPGPublic160729.

[19] D. Barney. “Overview slide of CE with main parameters” (Oct. 2019). Accessed August 24, 2021. URL: https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=13251.

[20] D. Barney and T. Sakuma. “Sketchup images highlighting the sub-detectors” (Sept. 2017). Accessed August 24, 2021. URL: https://cds.cern.ch/record/2628519.

[21] The CMS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. DOI: 10.1088/1748-0221/12/10/p10003.

[22] S. Cucciarelli et al. “Track reconstruction, primary vertex finding and seed generation with the Pixel Detector”. CMS-NOTE-2006-026 (Jan. 2006). URL: https://cds.cern.ch/record/927384.

[23] W. Adam et al. “Track Reconstruction in the CMS tracker”. CMS-NOTE-2006-041 (Dec. 2006). URL: https://cds.cern.ch/record/934067.

[24] W. Adam et al. “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”. *Journal of Physics G: Nuclear and Particle Physics* 31.9 (July 2005), N9–N20. DOI: 10.1088/0954-3899/31/9/n01.

[25] The CMS Collaboration. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. *JINST 16 (2021) P05014* (Dec. 12, 2020). DOI: 10.1088/1748-0221/16/05/P05014.

[26] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet user manual” (Nov. 25, 2011). DOI: 10.1140/epjc/s10052-012-1896-2.

[27] D. Bertolini et al. “Pileup Per Particle Identification”. *JHEP 1410 (2014) 59* (July 22, 2014). DOI: 10.1007/JHEP10(2014)059.

[28] The CMS Collaboration. “Pileup mitigation at CMS in 13 TeV data”. *Journal of Instrumentation* 15.09 (Sept. 2020), P09018–P09018. DOI: 10.1088/1748-0221/15/09/p09018.

[29] J. Thaler and K. V. Tilburg. "Identifying Boosted Objects with N-subjettiness". *JHEP 1103:015,2011* (Nov. 10, 2010). DOI: 10.1007/JHEP03(2011)015.

[30] The CMS Collaboration. "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques". *JINST 15 (2020) P06005* (Apr. 17, 2020). DOI: 10.1088/1748-0221/15/06/P06005.

[31] A. J. Larkoski et al. "Soft Drop". *JHEP 1405 (2014) 146* (Feb. 11, 2014). DOI: 10.1007/JHEP05(2014)146.

# APPENDIX A

# MACHINE LEARNING

Write about ML, MVA, Boosted decision tree (BDT), etc.

# APPENDIX B

# DATA ANALYSIS

Data Analysis Code and Stuff