

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, 2022
Vishnu Zutshi, Director

This dissertation reports on the search for the vector boson scattering (VBS) in semileptonic channel ZV , where the Z boson decays to a pair of leptons, and the other vector boson V (either W or Z) decays hadronically to a pair of jets, in association of two jets i.e. $ZVjj \rightarrow l^+l^-jjjj$, $l = e$ or μ . The search used 137 fb^{-1} of proton-proton collision data collected by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) with center-of-mass energy (\sqrt{s}) of 13 TeV from 2016 to 2018. VBS is a key process in understanding the nature of electroweak symmetry breaking (EWSB) in the framework of Standard Model (SM).

Studies and instrumentation done towards the future detector upgrade to the CMS experiment which will replace the current endcap Hadronic Calorimeter (HCAL) and Electromagnetic Calorimeter (ECAL) with High Granularity Calorimeter (HGCAL) during Long Shutdown 3 (LS3) are also reported. Optimal configuration of scintillator tiles coupled to silicon photomultipliers (SiPMs) in the HGCAL is measured and suggested using testbeam measurement of scintillator tiles and noise measurement of SiPMs. Automated wrapping of scintillator tiles in Enhanced Specular Reflector (ESR) film are also discussed.

NORTHERN ILLINOIS UNIVERSITY
DE KALB, ILLINOIS

SUMMER 2022

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

BY

RAMANPREET SINGH
© 2022 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

²³

ACKNOWLEDGEMENTS

²⁴

Acknowledge.

DEDICATION

25

To my family.

TABLE OF CONTENTS

	Page
²⁷ LIST OF TABLES	viii
²⁸ LIST OF FIGURES	ix
²⁹ LIST OF APPENDICES	xiii
 Chapter	
³⁰ 1 INTRODUCTION	1
³¹ 1.1 Standard model	2
³² 1.1.1 Quantum Electrodynamics (QED)	4
³³ 1.1.2 Quantum Chromodynamics (QCD)	5
³⁴ 1.1.3 Electroweak Theory	6
³⁵ 1.1.4 Electroweak Symmetry Breaking and Higgs Mechanism	8
³⁶ 1.2 Vector Boson Scattering	11
³⁷ 1.2.1 Motivation	11
³⁸ 1.2.2 Topology of VBS in Semileptonic ZV Final State	13
³⁹ 2 THE LHC AND CMS EXPERIMENT	16
⁴⁰ 2.1 The Large Hadron Collider	16
⁴¹ 2.1.1 Integrated Luminosity	18
⁴² 2.2 The CMS Detector	19
⁴³ 2.2.1 The CMS Coordinate System	22
⁴⁴ 2.2.2 The Superconducting Magnet	23
⁴⁵ 2.2.3 The Tracking System	24

	Page
46 Chapter	Page
47 2.2.4 The Electromagnetic Calorimeter	25
48 2.2.5 The Hadronic Calorimeter	27
49 2.2.6 Muon Detector	28
50 2.2.7 Trigger	30
51 3 EVENT SIMULATION AND RECONSTRUCTION	32
52 3.1 Track Reconstruction and Calorimeter Clustering	32
53 3.2 Reconstructed Particles	34
54 3.2.1 Muons	34
55 3.2.2 Electrons and Photons	35
56 3.2.3 Hadrons and Jets	36
57 3.2.3.1 N-Subjetiness and Deep Taggers	38
58 3.2.3.2 Softdrop Mass	40
59 3.2.4 Missing transverse momentum	41
60 3.3 Monte Carlo (MC) Simulation	42
61 3.3.1 Generators	42
62 3.3.2 Parton Shower and Hadronization	43
63 3.3.3 Tunes	44
64 4 VBS MEASUREMENT IN ZVJJ FINAL STATE	45
65 4.1 Dataset and Simulation	46
66 4.1.1 Data	46
67 4.1.2 MC Simulations	47
68 4.2 Event Selection	52
69 4.2.1 VBS Tagged Jets and V Jet Candidate	53

		Page
71	Chapter	
70	4.3 Control and Signal Region	54
72	4.3.1 MC Scale Factors	54
73	4.3.2 DY+Jets Normalization	55
74	4.4 Kinematics Distributions	57
75	4.4.1 Boosted ZV DY+Jets Control Region	58
76	4.4.2 Resolved ZV DY+Jets Control Region	65
77	4.5 Machine Learning Modeling	73
78	4.5.1 Algorithm: Gradient Boosted Decision Tree	73
79	4.5.2 Training and Results	74
80	4.5.3 MVA Score Inference	82
81	4.6 Measurement	83
82	4.6.1 Statistical and Systematic Uncertainties	84
83	4.6.2 Significance	85
84	4.6.3 Impact Plots	87
85	4.6.4 Postfit Plots	89
86	5 HIGH GRANULARITY CALORIMETER UPGRADE	92
87	5.1 Technical Design and Requirements	92
88	5.2 Scintillator Tiles and SiPMs	94
89	5.2.1 Scintillator Materials	95
90	5.2.2 SiPMs	96
91	5.2.3 Scintillator Tiles	96
92	5.3 Automated Wrapping of Scintillator Tiles	99
93	5.4 Signal-to-Noise Ratio	103
94	5.4.1 Formulation	103

	Page
95 Chapter	Page
96 5.4.2 Testbeam and SiPM Noise Inputs	104
97 5.4.3 Scenarios	105
98 5.4.4 Results and Conclusion	106
99 6 CONCLUSIONS	109
100 REFERENCES	111
101 APPENDICES	119

LIST OF TABLES

	Table	Page
103	4.1 Trigger paths used to select events in CMS collision data	47
104	4.2 List of MC samples for Signal and Background modeling	49
105	4.3 List of MC samples for Signal and Background modeling	50
106	4.4 List of MC samples for Signal and Background modeling	51
107	4.5 Normalization factors from 1D bin of p_T of Z	56
108	4.6 Combined dataset Training and Testing Statistics in Signal Region	75
109	4.7 Training Input Variable Ranking	75
110	4.8 TMVA configuration of trained models	76
111	4.9 Expected Signal Significance	86
112	4.10 Prefit and postfit yields of MC processes	90
113	4.11 Prefit and postfit yields of MC processes	91
114	5.1 HGCAL scenarios comparison	108

LIST OF FIGURES

	Figure	Page
116	1.1 Standard model list of matter and interaction particles	3
117	1.2 3D representation of Higgs potential	10
118	1.3 The cross-sections for longitudinal polarized VBS involving three and four- 119 boson coupling with and without Higgs coupling included.	12
120	1.4 Tree level Feynman diagram of VV VBS.	13
121	1.5 Tree level processes that can happen in scattering represented by blob in VV 122 VBS.	13
123	1.6 Tree level Feynman diagram of ZV VBS process at LHC	14
124	2.1 A schematic of the CERN accelerator complex.	17
125	2.2 Integrated delivered and recorded luminosity for 2015–2018 proton-proton 126 collisions	18
127	2.3 The CMS detector cutaway view	20
128	2.4 Schematic view of the CMS detector.	21
129	2.5 The picture of the CMS detector central part when lowered in underground 130 cavern with superconducting magnet and iron yoke	23
131	2.6 The CMS pixel upgrade.	25
132	2.7 The CMS ECAL schematic layout	26
133	2.8 The ECAL endcap quadrant assembled view	26
134	2.9 The first half of the barrel HCAL inserted into the superconducting solenoid (April 2006)	28
135	2.10 The HCAL depth segmentation after phase 1 upgrade	29

Figure	Page
2.11 The quadrant view of CMS subdetectors layout, and the coverage of the muon detector DTs, CSCs, and RPCs highlighted	30
3.1 Comparison of τ_{21} and τ_{32} shapes for signal and background in AK8 jets	39
3.2 The network architecture of DeepAK8	40
4.1 Normalization factors from 1D bin of p_T of Z and VBS trailing jet	56
4.2 DY+Jets Control Region: Leading electron kinematics in Boosted ZV Channel	58
4.3 DY+Jets Control Region: Trailing electron kinematics in Boosted ZV Channel	59
4.4 DY+Jets Control Region: Leading muon kinematics in Boosted ZV Channel	60
4.5 DY+Jets Control Region: Trailing muon kinematics in Boosted ZV Channel	61
4.6 DY+Jets Control Region: Hadronic boson kinematics in Boosted ZV Channel	62
4.7 DY+Jets Control Region: Leading VBS tagged jet kinematics in Boosted ZV Channel	63
4.8 DY+Jets Control Region: Trailing VBS tagged jet kinematics in Boosted ZV Channel	64
4.9 DY+Jets Control Region: Leading electron kinematics in Resolved ZV Channel	65
4.10 DY+Jets Control Region: Trailing electron kinematics in Resolved ZV Channel	66
4.11 DY+Jets Control Region: Leading muon kinematics in Resolved ZV Channel	67
4.12 DY+Jets Control Region: Trailing muon kinematics in Resolved ZV Channel	68
4.13 DY+Jets Control Region: Hadronic boson leading jet kinematics in Resolved ZV Channel	69
4.14 DY+Jets Control Region: Hadronic boson trailing jet kinematics in Resolved ZV Channel	70

165 Figure		Page
166 4.15 DY+Jets Control Region: Leading VBS tagged jet kinematics in Resolved 167 ZV Channel	71	71
168 4.16 DY+Jets Control Region: Trailing VBS tagged jet kinematics in Resolved 169 ZV Channel	72	72
170 4.17 Decision Tree	74	74
171 4.18 Inputs Variables for training Boosted ZV BDT Classifier	77	77
172 4.19 Inputs Variables for training Resolved ZV BDT Classifier	78	78
173 4.20 Inputs Variables for training Resolved ZV BDT Classifier	79	79
174 4.21 Correlation Matrix for Signal and Background	80	80
175 4.22 MVA Score ROC Curve	81	81
176 4.23 MVA Score in Control Region for Boosted ZV Channel	82	82
177 4.24 MVA Score in Control Region for Resolved ZV Channel	82	82
178 4.25 MVA Score in Signal Region for Boosted ZV Channel	83	83
179 4.26 MVA Score in Signal Region for Resolved ZV Channel	83	83
180 4.27 Impact Plots	87	87
181 4.28 Impact Plots	88	88
182 4.29 MVA Score postfit in Signal Region for Boosted ZV Channel	89	89
183 4.30 MVA Score postfit in Signal Region for Resolved ZV Channel	89	89
184 5.1 Overview of HGCAL position and quadrant view of CE	93	93
185 5.2 CE-H mixed layer 22	94	94
186 5.3 SiPM	97	97
187 5.4 Scintillator tile with dimple	97	97
188 5.5 ESR wrapper cut for tile size R18–19	98	98
189 5.6 R18–19 scintillator tile wrapped in ESR	98	98

	Figure	Page
190	5.7 Automated scintillator tile wrapper overview	99
191	5.8 Tile folding station and ESR magazine	100
192	5.9 Sticker and tile dispenser.	102
193	5.10 Tile folding station	102
194	5.11 Scintillator performance with various active area size of SiPM.	107
195	5.12 HGCAL scenarios	108
196	B.1 Impact Plots.	122
197	B.2 Impact Plots.	123
198		

LIST OF APPENDICES

Appendix	Page
200 A ANALYSIS CODE	119
201 B ADDITIONAL FIGURES	121
202 B.1 Impact Plots	122

203

CHAPTER 1

204

INTRODUCTION

It doesn't matter how beautiful your theory is, it doesn't matter how smart you are.

If it doesn't agree with experiment, it's wrong. In that simple statement is the key to science.

— Richard P. Feynman

205 “Particle Physics” is the branch of physics which deals with the fundamental particles
206 and the interactions between them. Fundamental particles are the subatomic particles which
207 at our present level of understanding are not made of other particles. There are two types
208 of fundamental particles “matter” and “interaction” particles. As the name suggests matter
209 particles are the fundamental constituent of matter and the interactions among them is
210 governed by how they exchange interaction particles. Standard Model (SM) of particle
211 physics is the theory that classifies these fundamental particles and describes three out of
212 four fundamental interaction forces; electromagnetic, weak, and strong.

213 This chapter briefly introduces the theory of SM, Higgs mechanism, spontaneous elec-
214 troweak symmetry breaking (EWSB), vector boson scattering (VBS), and the motivation for
215 the search of VBS in semileptonic decay channel ZV with leptonic decay of Z , and hadronic
216 decay of V (W/Z) to pair of quarks.

217

1.1 Standard model

218 In SM, the matter particles are fermions, and the interaction particles are bosons. SM
 219 also includes anti-fermions, which are fermions with equal mass but opposite sign of charge.
 220 Figure 1.1 lists mass, electric charge, and spin of fermions and bosons in the SM.

221 Fermions obey Fermi-Dirac statistics and have half integer spin. They can be further
 222 divided into leptons which have integral electric charge, and quarks which have fractional
 223 electric charge. There are three generations of quarks and leptons discovered to date, each
 224 generation only differing in mass. In addition to the electric charge, quarks also have three
 225 types of “color” charge (red, green and blue). Quarks cannot be isolated because of “color
 226 confinement”, which requires net color charge to be zero for a free particle, for this reason we
 227 can only have certain composition of quarks. Baryons (proton, neutrons, etc.) are made up
 228 of three quarks each with different color charge, and mesons (pions, kaons, etc.) are made
 229 of two quarks with color and anti-color charge.

230 Bosons obey Bose-Einstein statistics and have integral spin. They are described by a
 231 local gauge theory and are also called gauge boson. Photons are the interaction particle of
 232 electromagnetic force, they are massless and only interact with charged particles. Gluons are
 233 the mediator of strong force between the quarks, they are massless and carry color charge.
 234 W^\pm and Z are the vector bosons and mediator of weak force, unlike photons and gluons
 235 they are massive. W^+ and W^- are antiparticles of each other, and Z is its own antiparticle.
 236 The last gauge boson, Higgs is a massive scalar boson with zero spin, zero electric, and no
 237 color charge. Higgs boson is not a force carrier, but rather explains why only some particles
 238 have mass.

239 The SM is built in the framework of quantum field theory (QFT), in which particles
 240 are excitation of the fields and interactions arise from local gauge invariance. The SM is a

²⁴¹ $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge theory, $U(1)_Y$, $SU(2)_L$, $SU(3)_C$ are the gauge symmetries
²⁴² of quantum electrodynamics (QED), weak interaction and quantum chromodynamics (QCD)
²⁴³ respectively, where the indices stands for “hypercharge” (Y), “left-handed” (L) and “color”
²⁴⁴ (C).

Standard Model of Elementary Particles

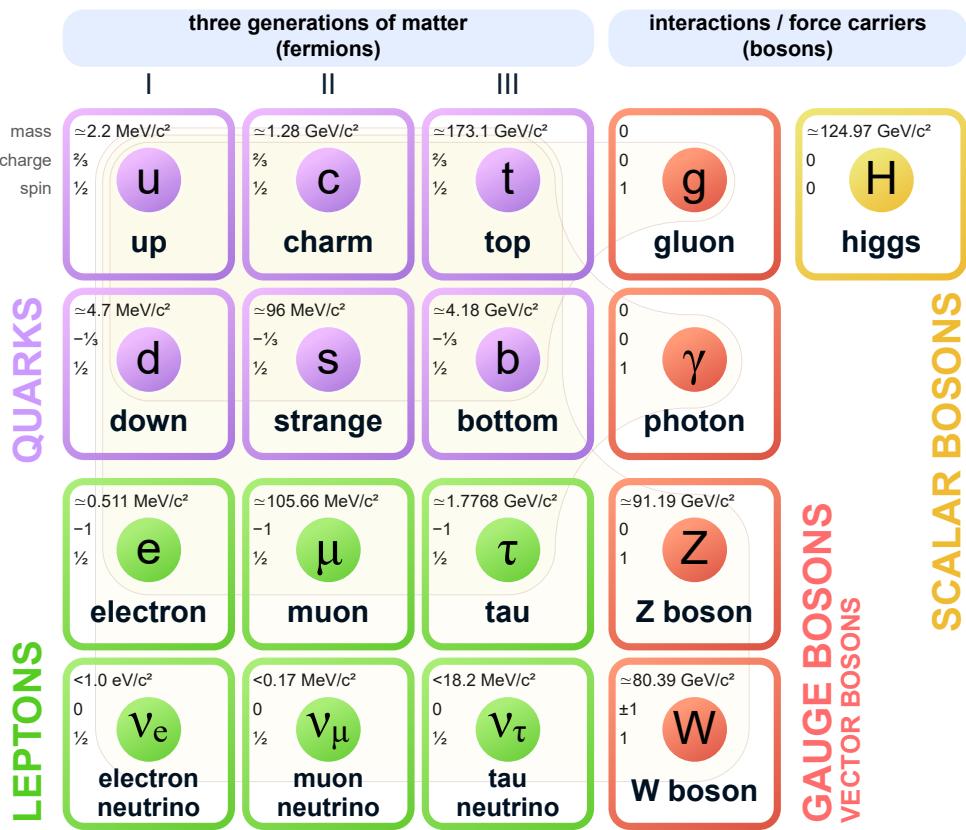


Figure 1.1: Standard model list of matter and interaction particles [1]

245

1.1.1 Quantum Electrodynamics (QED)

QED is a quantum field theory of electrodynamics, it describes the interaction of photons to the charged fermions. The QED is local gauge invariant and symmetric with $U(1)_Q$ symmetry group, defined as,

$$U(1)_Q = \exp(iQ\theta(x)) \quad (1.1)$$

246 where $\theta(x)$ is any spacetime function also called gauge parameter, and Q is the coupling
247 constant of photon field to the fermions which is equivalent to the charge of fermion.

Under this transformation, fermion spinor $\psi(x)$ and four-potential A_μ electromagnetic tensor will transform as,

$$\psi(x) \rightarrow U(1)_Q \psi(x) \quad (1.2)$$

$$A_\mu \rightarrow A_\mu - \frac{1}{e} \partial_\mu \theta \quad (1.3)$$

The general Lagrangian of QED for fermions and their interaction with photon field is given by,

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.4)$$

where m is the mass of fermion, D_μ is the covariant derivative, and $F_{\mu\nu}$ is the electromagnetic field tensor defined as,

$$D_\mu = \partial_\mu + iQA_\mu \quad (1.5)$$

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (1.6)$$

248

1.1.2 Quantum Chromodynamics (QCD)

249 The strong interactions are represented by $SU(3)_C$ gauge group, invariant under trans-
 250 formations of color charge degree of freedom, based on Yang-Mills theory [2]. Since “elec-
 251 trodynamics” is the theory of electric charge, this theory of color (*chromo* in Greek) charge
 252 is called “chromodynamics”, hence the name quantum chromodynamics (QCD).

A quark spinor in initial state can be represented as,

$$\psi = \begin{pmatrix} \psi_{red} \\ \psi_{blue} \\ \psi_{green} \end{pmatrix} = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} \quad (1.7)$$

$SU(3)_C$ is an exact symmetry, it means the difference between colors cannot be measured experimentally, thus the color labels in quark spinor are arbitrary. $SU(3)_C$ transformation is defined as,

$$SU(3)_C = \exp \left(i\theta^a(x) \frac{\lambda^a}{2} \right) \quad (1.8)$$

253 where λ^a for $a = 1, \dots, 8$, are the Gell-Mann matrices, and $\theta^a(x)$ are any gauge parameters.
 254 These eight generators of symmetry corresponds to eight gauge vector boson gluons.

Similar to QED, the covariant derivative for QCD can be formed as,

$$D_\mu = \partial_\mu + ig_s \frac{\lambda^a}{2} G_\mu^a \quad (1.9)$$

255 where g_s is the coupling constant of gluon to the quarks, and G_μ^a are the eight gauge
 256 fields corresponding to gluons.

The corresponding field strength tensor in QCD can be formed as,

$$F_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f^{abc} G_\mu^b G_\nu^c \quad (1.10)$$

where f^{abc} are the structure constants of $SU(3)_C$ which satisfy $[\lambda^a, \lambda^b] = i f^{abc} \lambda^c$ relation.

The full Lagrangian for QCD can now be constructed as,

$$\mathcal{L}_{QCD} = \bar{\psi}^i (i\gamma^\mu D_\mu{}^{ij} - m\delta^{ij}) \psi^j - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu} \quad (1.11)$$

for mass m , indices i and j runs from 1 to 3.

The main difference of gluon field with respect to photon field is the presence of third term in field strength tensor which allows triplet and quartic self coupling of gluons.

1.1.3 Electroweak Theory

Weak interaction occurs for all the fermions via the exchange of massive vector bosons W^\pm , Z . Since the unification of electromagnetic and weak interaction into electroweak (EW) interaction by Glashow, Weinberg, and Salam [3–5], the weak interaction is better understood in terms of EW theory.

Weak interaction only couples to left-handed fermions and it is same whether the fermion is charged or not. The underlying gauge group of EW interaction is $SU(2)_L \otimes U(1)_Y$ and has

two transformations one for the left-handed doublet L and the right handed singlet fermions ψ_R which are defined as,

$$SU(2)_L \otimes U(1)_Y = \exp \left(i\theta^a(x) \frac{\sigma^a}{2} + i\theta(x) \frac{Y}{2} \right), \quad (\text{doublet}) \quad (1.12)$$

$$= \exp \left(i\theta(x) \frac{Y}{2} \right), \quad (\text{singlet}) \quad (1.13)$$

where Y is the hypercharge (linear combination of electric charge and weak isospin component), and σ^a for $a = 1, 2, 3$ are the Pauli spin matrices generator of $SU(2)$ symmetry. Left-handed fermion L doublets are,

$$L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau_L \end{pmatrix}, \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \quad (1.14)$$

and right-handed singlets are,

$$\psi_R = e_R, \mu_R, \tau_R, u_R, d_R, c_R, s_R, t_R, b_R \quad (1.15)$$

The covariant derivative of EW is then,

$$D_\mu L = \left(\partial_\mu + ig_w \frac{\sigma^a}{2} W_\mu^a + ig \frac{Y}{2} B_\mu \right) L \quad (1.16)$$

$$D_\mu \psi_R = \left(\partial_\mu + ig \frac{Y}{2} B_\mu \right) \psi_R \quad (1.17)$$

where W_μ^a and B_μ are the gauge fields. The EW Lagrangian can now written as,

$$\mathcal{L}_{EW} = i\bar{L}\gamma^\mu D_\mu L + i\bar{\psi}_R \gamma^\mu D_\mu \psi_R - \frac{1}{4} W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \quad (1.18)$$

where $B_{\mu\nu}$ and $W_{\mu\nu}^a$ are fields strength, defined as,

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.19)$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g_w \epsilon^{abc} W_\mu^b W_\nu^c \quad (1.20)$$

the linear combination of B_μ and W_μ gauge field, with a weak mixing angle θ_w gives 4 vectors boson W^+ , W^- , Z , and γ of SM,

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp W_\mu^2) \quad (1.21)$$

$$Z_\mu = \cos \theta_w W_\mu^3 - \sin \theta_w B_\mu \quad (1.22)$$

$$A_\mu = \sin \theta_w W_\mu^3 + \cos \theta_w B_\mu \quad (1.23)$$

$$\tan \theta_w = g/g_w \quad (1.24)$$

266 Similar to QCD, the presence of third term in field strength tensor allows the self triple
267 (WWZ , $WW\gamma$) and quartic ($WWWW$, $WWZZ$, $WWZ\gamma$, $WW\gamma\gamma$) couplings.

268 1.1.4 Electroweak Symmetry Breaking and Higgs Mechanism

269 *Spontaneous symmetry breaking* is the phenomena which explains why the ground state
270 is not invariant under the symmetry of the Lagrangian. “Spontaneous” means that the
271 symmetry breaking is not done by external agent but rather by Lagrangian itself in ground
272 state.

273 The EW theory unifies weak interaction and QED but the gauge boson in EW theory
274 are all massless, if we were to add mass terms like $-m^2 W_\mu W^\mu$ by hand, it will no longer
275 be gauge invariant. The solution to this, without breaking gauge invariance, is spontaneous

²⁷⁶ symmetry breaking, but this requires addition of new scalar field called Higgs field via
²⁷⁷ Brout–Englert–Higgs mechanism (BEH) [6, 7], and this symmetry breaking is known as
²⁷⁸ electroweak symmetry breaking (EWSB).

BEH introduces a complex scalar field as $SU(2)_L$ doublet with non-zero vacuum expectation value (VEV),

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^3 \end{pmatrix} \quad (1.25)$$

and BEH Lagrangian is,

$$\mathcal{L}_{BEH} = |D_\mu \phi|^2 - V(\phi) \quad (1.26)$$

where D_μ is same as EW covariant derivate in Equation 1.16, and $V(\phi)$ is,

$$V(\phi) = \mu^2 |\phi|^2 + \lambda(|\phi|^2)^2 \quad (1.27)$$

the parameter λ is required to be positive, for $\mu^2 > 0$ the minima is at 0, which is not an interesting case, but for $\mu^2 < 0$ vacuum state energy is given by,

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \quad (1.28)$$

by the choice of non-zero VEV v , scalar field can be parameterized as,

$$v = \sqrt{\frac{-\mu^2}{\lambda}} \quad (1.29)$$

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h(x) + v \end{pmatrix} \quad (1.30)$$

where $h(x)$ is the Higgs field and BEH spontaneously breaks electroweak symmetry,

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{EM} \quad (1.31)$$

279 Visually the Higgs potential is shown in Figure 1.2. The ball position at the center
280 represents unbroken symmetry, and at the minima represents spontaneous broken symmetry
in the ground state of potential.

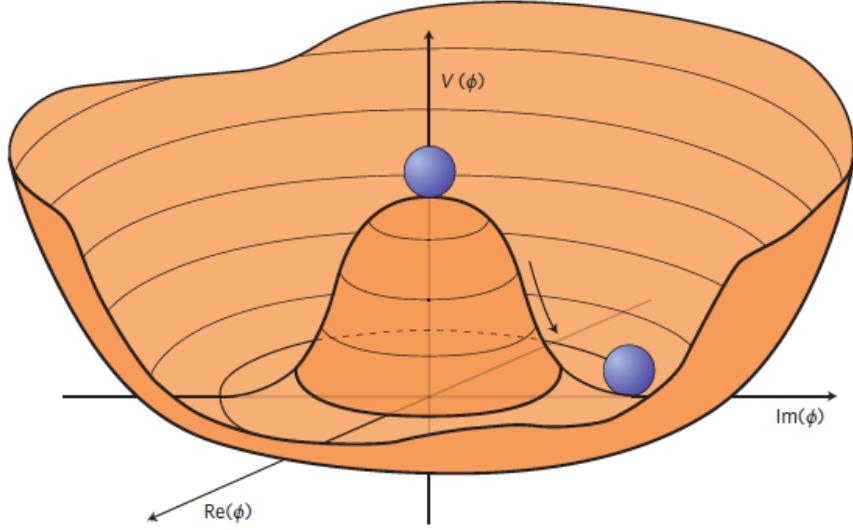


Figure 1.2: 3D representation Higgs potential [8].

281

After EWSB, the BEH Lagrangian contains the following mass terms,

$$m_W^2 W_\mu^+ W^{-\mu}, \quad m^2 Z Z_\mu Z^\mu, \quad m^2 h^2 \quad (1.32)$$

and gauge fields,

$$-\frac{1}{4} A_{\mu\nu} A^{\mu\nu}, \quad -\frac{1}{4} W_{\mu\nu}^+ W^{-\mu\nu}, \quad -\frac{1}{4} Z_{\mu\nu} Z^{\mu\nu}, \quad -\frac{1}{4} (\partial_\mu h)(\partial^\mu h) \quad (1.33)$$

²⁸² thus explaining existence of three massive vector boson (W^\pm , Z), one massless vector
²⁸³ boson (γ), and one massive scalar boson Higgs (H).

With experimentally measured value of VEV v approximately 246 GeV, The masses of bosons can be written in terms of v as,

$$m_A = 0, \quad m_W = \frac{g_W v}{2}, \quad m_Z = \frac{\sqrt{g_W^2 + g^2} v}{2}, \quad m_H = \sqrt{2\lambda} v \quad (1.34)$$

²⁸⁴ 1.2 Vector Boson Scattering

²⁸⁵ The vector boson scattering (VBS) is a key process to experimentally study the nature of
²⁸⁶ EWSB mechanism, because massive vector boson at high energy scattering can probe EW
²⁸⁷ theory at TeV scale. This section describes the motivation behind studying VBS, and the
²⁸⁸ topology of scattering of ZV channel studied in this dissertation.

²⁸⁹ 1.2.1 Motivation

A massless spin-1 boson can exists in two transverse polarization as,

$$\varepsilon_\pm^\mu = \mp \frac{1}{\sqrt{2}} (0, 1, \pm i, 0) \quad (1.35)$$

and massive vector bosons can also exists in one longitudinal polarization,

$$\varepsilon_L^\mu = \frac{1}{m} (p_z, 0, 0, E) \quad (1.36)$$

²⁹⁰ This means the longitudinal polarized VBS cross-section will scale as E/m , whereas the
²⁹¹ scattering cross-section of transverse polarized boson remains constant. Figure 1.3 shows

the cross-section of longitudinal polarized VBS $V_L V_L \rightarrow V_L V_L$ from low to high energies.
 Perturbatively, the cross-section of longitudinal polarized VBS will scale with center of mass
 energy \sqrt{s} and eventually unitarity is violated at ≈ 1.2 TeV scale [9, 10]. Figure 1.3 also
 shows how the existence of light Higgs boson and inclusion of Higgs to vector boson coupling
 diagrams in longitudinal polarized VBS can restore unitarity, and any deviation from SM
 Higgs to vector bosons coupling will be visible in such a di-boson spectrum.

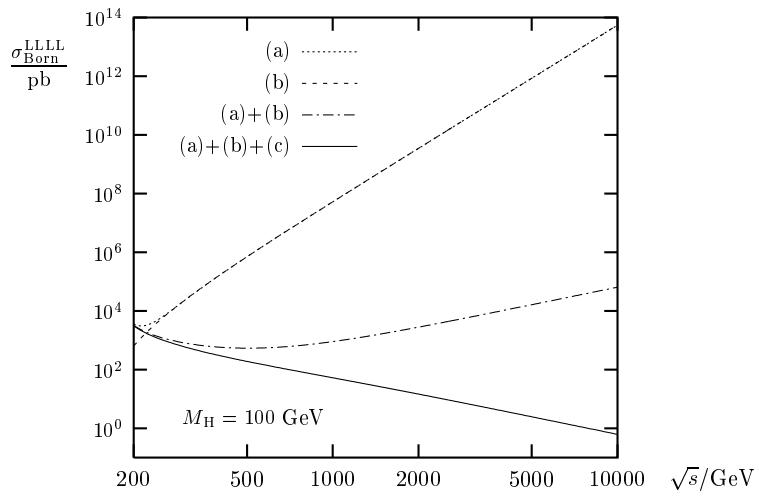


Figure 1.3: The cross-sections for longitudinal polarized VBS involving three and four-boson coupling with and without Higgs coupling diagram included. (a) is for the diagrams with three-boson coupling, (b) is for diagrams with four-boson coupling, and (c) is for diagrams with Higgs to vector boson coupling. [11]

Thus VBS measurement serves the two-fold purpose of doing precise measurements of
 the Higgs coupling to vector bosons at higher energy scales, and providing a framework to
 test for anomalous quartic gauge couplings in the SM.

301 1.2.2 Topology of VBS in Semileptonic ZV Final State

302 As discussed in previous section, VBS in SM can proceed via exchange of vector boson,
 303 Higgs boson, or quartic gauge coupling of four vector bosons. Figure 1.4, 1.5 shows tree level
 304 Feynman diagrams for VBS process.

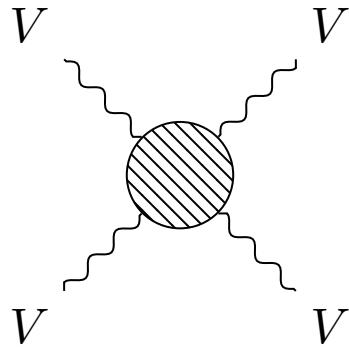


Figure 1.4: Tree level Feynman diagram of VV VBS, where V can be W^\pm , Z , or γ

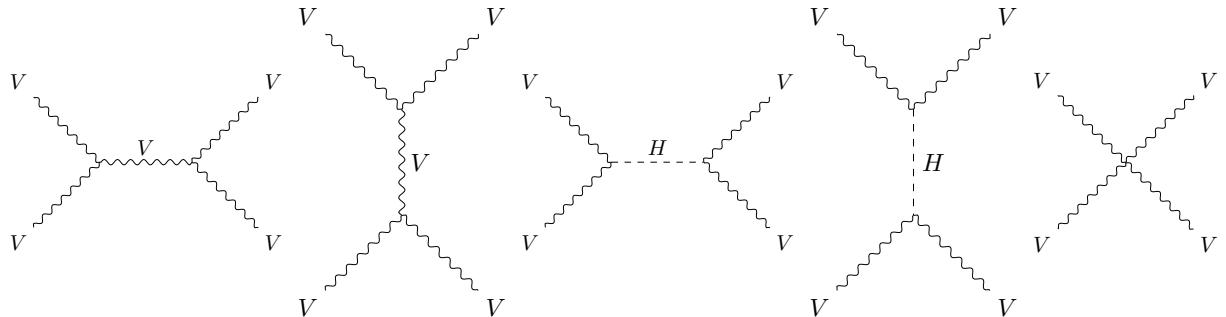


Figure 1.5: Tree level processes that can happen in scattering represented by blob in VV VBS, starting from s and t-channel exchange of vector boson, Higgs boson, and the last one is quartic coupling of vector bosons.

305 In proton-proton collisions, the actual interaction happens with the constituent quarks.
 306 For VBS to occur, the incoming (colliding) quarks have to radiate vector boson, then the
 307 scattering process between those vector boson can proceed via exchange of vector boson,

308 Higgs boson, or quartic coupling. The tree level Feynman diagram of a ZV VBS process in
309 proton-proton collision is shown in Figure 1.6.

310 The outgoing quarks are a key signature of VBS in hadron collider experiments because
311 they will have large pseudorapidity¹ difference between them, corresponding to a large in-
312 variant mass of outgoing quark pair. These outgoing quarks are first identified and tagged
313 to filter out most of the QCD background.

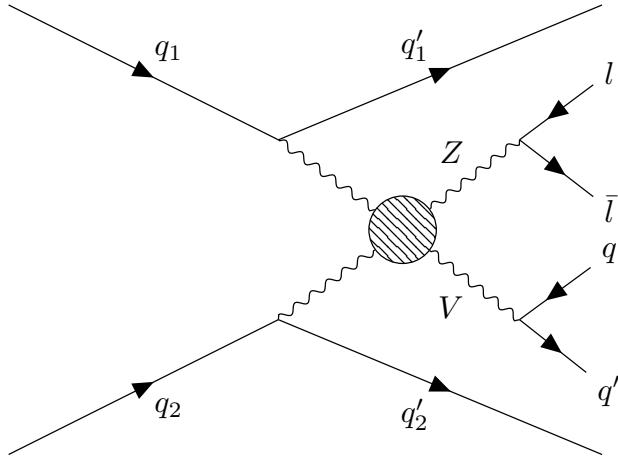


Figure 1.6: Tree level Feynman diagram of ZV VBS process at LHC. The diagram shows the production of two vector bosons being radiated from incoming quarks, in final state after scattering (blob), Z decays to pair of leptons, V (W/Z) decays to pair of quarks, and plus two outgoing quarks.

314 The type of leptonically decaying vector boson can be determined from the number of
315 leptons in final state i.e. whether it was W (1 lepton + 1 neutrino) or Z (2 leptons). But
316 given the hadronic activity in the proton-proton collisions, it is challenging to determine the
317 type of hadronically decaying vector boson and is generally denoted by V .

318 Depending upon the decay channel of vector bosons, VBS can be categorized into fully
319 leptonic or hadronic, or semileptonic channels. The fully leptonic channel has the cleanest
320 signal, but the signal yield is also lower. The fully hadronic channel has the highest signal

¹defined as $\eta = -\ln[\tan \theta/2]$, where θ is angle between the z -axis and the beam-axis.

321 yield but also have large background because of hadronic activity in the final state. The
322 semileptonic final state sits in-between these two channels, with cleaner boson reconstruction
323 from leptons, and larger yields from hadronically decaying boson.

324 Out of two channels in the semileptonic, WV and ZV , WV has been studied and pub-
325 lished by the Compact Muon Solenoid (CMS) [12]. Anomalous quartic gauge coupling limits
326 derived using semileptonic channels with 2016 data of the CMS have also been published [13].

327 This analysis looks for the VBS signature with ZV in final state with the Z decaying
328 to two opposite sign same flavor (OSSF) leptons, and V decaying to a pair of quarks in
329 association with two outgoing quarks. Chapter 2 describes the proton-proton collision data
330 used in this analysis and how various subdetectors of the CMS experiment work to capture
331 the event information. Then in Chapter 3 the particle reconstruction and event simulations
332 are discussed. In the Chapter 4 the object and events selection used are listed, the sig-
333 nal vs background classifier developed is described, and finally the measurement of signal
334 significance is discussed.

335

CHAPTER 2

336

THE LHC AND CMS EXPERIMENT

337

The physics analysis is carried out using data collected by the CMS experiment at European Center for Nuclear Research (CERN) Large Hadron Collider (LHC) accelerator. This chapter provides overview of the LHC and details of the CMS experiment and its sub-detectors for particle tracking and calorimetry.

341

2.1 The Large Hadron Collider

342

The LHC is the largest accelerator located at CERN in Geneva, Switzerland. The main LHC ring is 27 km in circumference and 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 run1 (2010–2012), and at 13 TeV center-of-mass energy during run2 (2015–2018) [14].

347

Figure 2.1 describes CERN accelerator complex. The protons are sourced by ionizing hydrogen atoms and then fed into a linear accelerator (LINAC). The LINAC accelerates the protons to 50 MeV and delivers them to the booster. Then the booster increases the energy of the protons to 1.4 GeV and feeds them to the proton syncrotron (PS) which further increases their energy to 25 GeV and starts bunching them together with bunches 25 ns apart. The proton bunches are passed through super proton synchrotron (SPS) which brings their energy to 450 GeV and bunches are then finally sent to main LHC clockwise and counterclockwise

354 rings where they are accelerated to final energy required which is 6.5 TeV for both bunches
 355 going clockwise and counterclockwise to obtain collisions at 13 TeV center-of-mass energy.

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

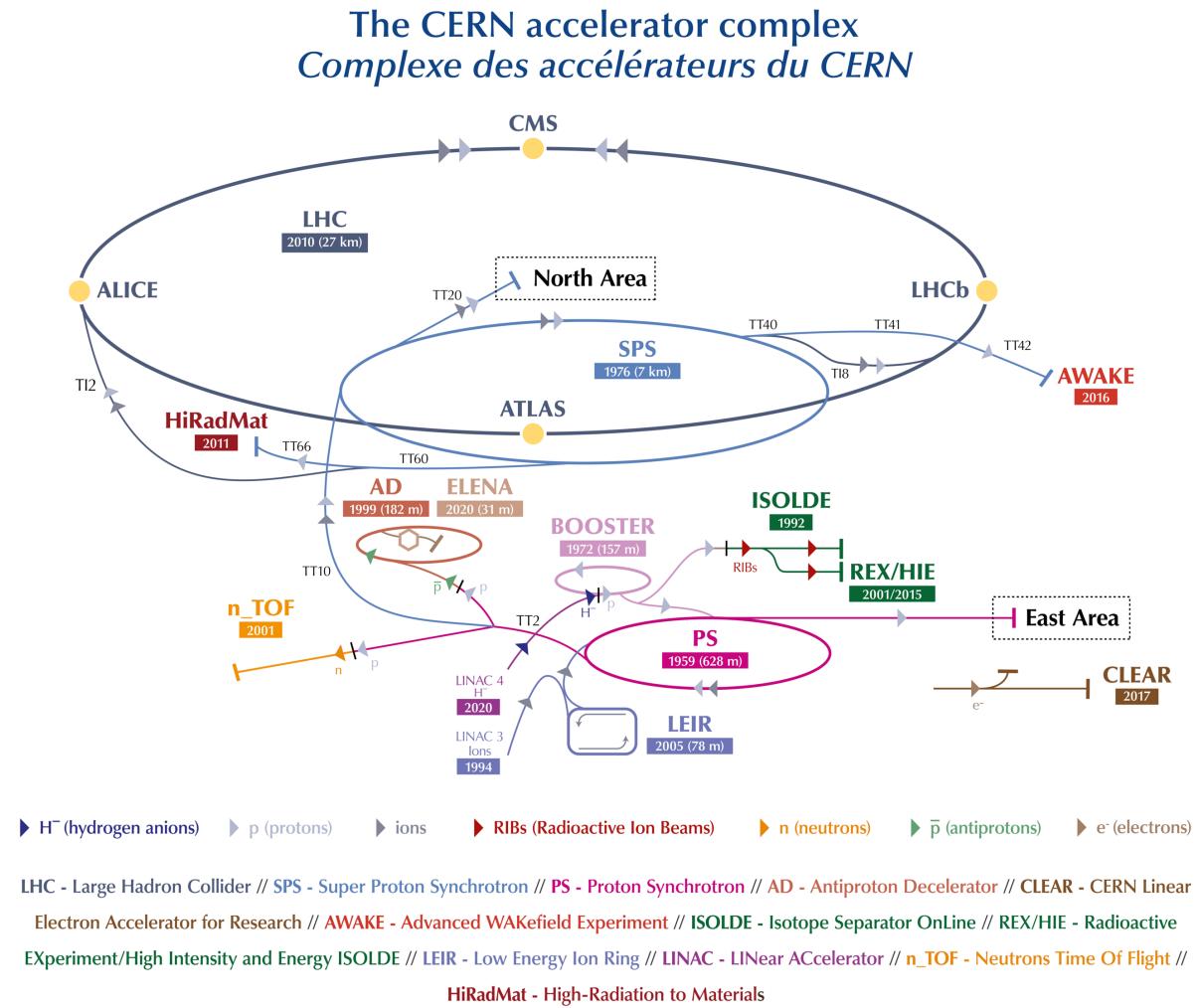


Figure 2.1: A schematic of the CERN accelerator complex [15]

359

2.1.1 Integrated Luminosity

Luminosity is a measure of how many events can happen per unit area and per unit time for a collision or scattering process, and it is defined as,

$$L = \frac{1}{\sigma} \frac{dN}{dt}, \quad L_{\text{int}} = \int L dt \quad (2.1)$$

360 where σ is cross-section of the process and L is the instantaneous luminosity ($\text{cm}^{-2} \text{s}^{-1}$),
 361 and L_{int} is the integrated luminosity usually expressed in inverse units of cross-section (pb^{-1} ,
 362 fb^{-1} , etc.).

363 For proton-proton collision at LHC the nominal instantaneous luminosity is of the order
 364 of $10^{34} \text{ cm}^{-2} \text{s}^{-1}$, and integrated luminosity delivered and recorded by CMS during run2¹
 365 operation is shown in Figure 2.2.

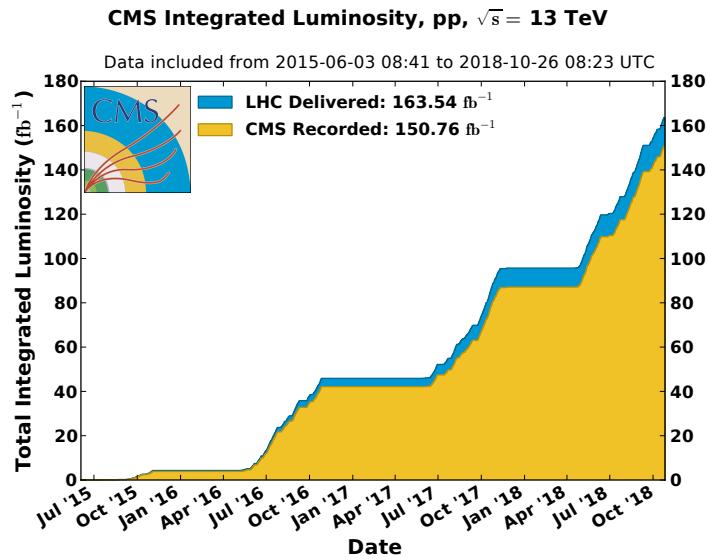


Figure 2.2: Integrated delivered and recorded luminosity for 2015–2018 proton-proton collisions [16]

¹2nd run period (2015–2018) of proton-proton collisions at LHC

366 For run2 standard physics analysis, luminosity recorded during 2016–2018 years is con-
 367 sidered, and only runs certified as “golden” by CMS Luminosity physics object group (POG)
 368 are used for the analysis. The runs are certified as “golden” if all the subdetector, triggers
 369 and physics objects perform as expected. The total integrated luminosity for run2 standard
 370 physics is 137.19 fb^{-1} (35.92 fb^{-1} during 2016, 41.53 fb^{-1} during 2017, and 59.74 fb^{-1} during
 371 2018 period) [17–19].

372 2.2 The CMS Detector

373 The CMS detector is a general purpose high energy particle physics detector. A cutaway
 374 view of the detector is shown in Figure 2.3. The detector is cylindrical with dimensions:
 375 21 meters long, and 15 meters in diameter, with the whole detector weighing about 14000
 376 tonnes. The detector is built in slices with central region called the “barrel”, and the two
 377 enclosing sides called “endcaps”. A superconducting solenoid generates magnetic field of
 378 3.8 T inside and 2 T outside. To contain the magnetic field outside of solenoid and support
 379 structure of the detector, massive steel yokes are used.

380 Figure 2.4 shows a schematic view of the CMS detector, and how different particles leave
 381 signature in various sub-detectors. Neutral particles such as photons, neutrinos, and neutral
 382 hadrons will leave no track in the Silicon Tracker (ST), and are identified only by the energy
 383 deposited in the calorimeters or the missing energy in an event. Electrons are identified from
 384 the track in the ST and energy deposit in the Electromagnetic Calorimeter (ECAL). Since
 385 hadrons are heavier and they pass through the ECAL and deposit their most of energy in
 386 the Hadronic Calorimeter (HCAL), leaving only small fraction of energy in the ECAL. Since
 387 muons are minimum ionizing particles (MIPs), they pass through whole detector with very
 388 small fraction of the energy deposited in the ECAL and HCAL.

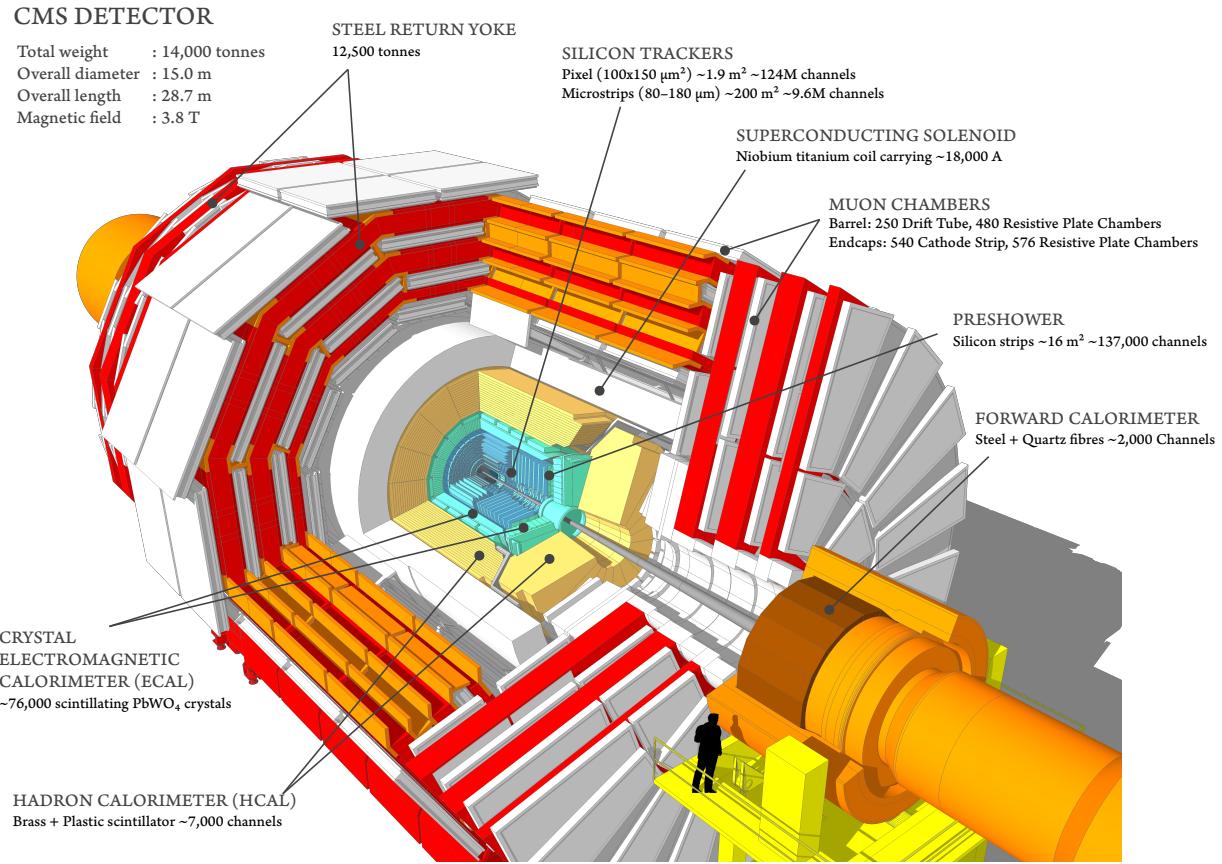


Figure 2.3: The CMS detector cutaway view [20]

389 This following section gives an overview of the subsystems of CMS detector. For detailed
 390 technical description refer to [21].

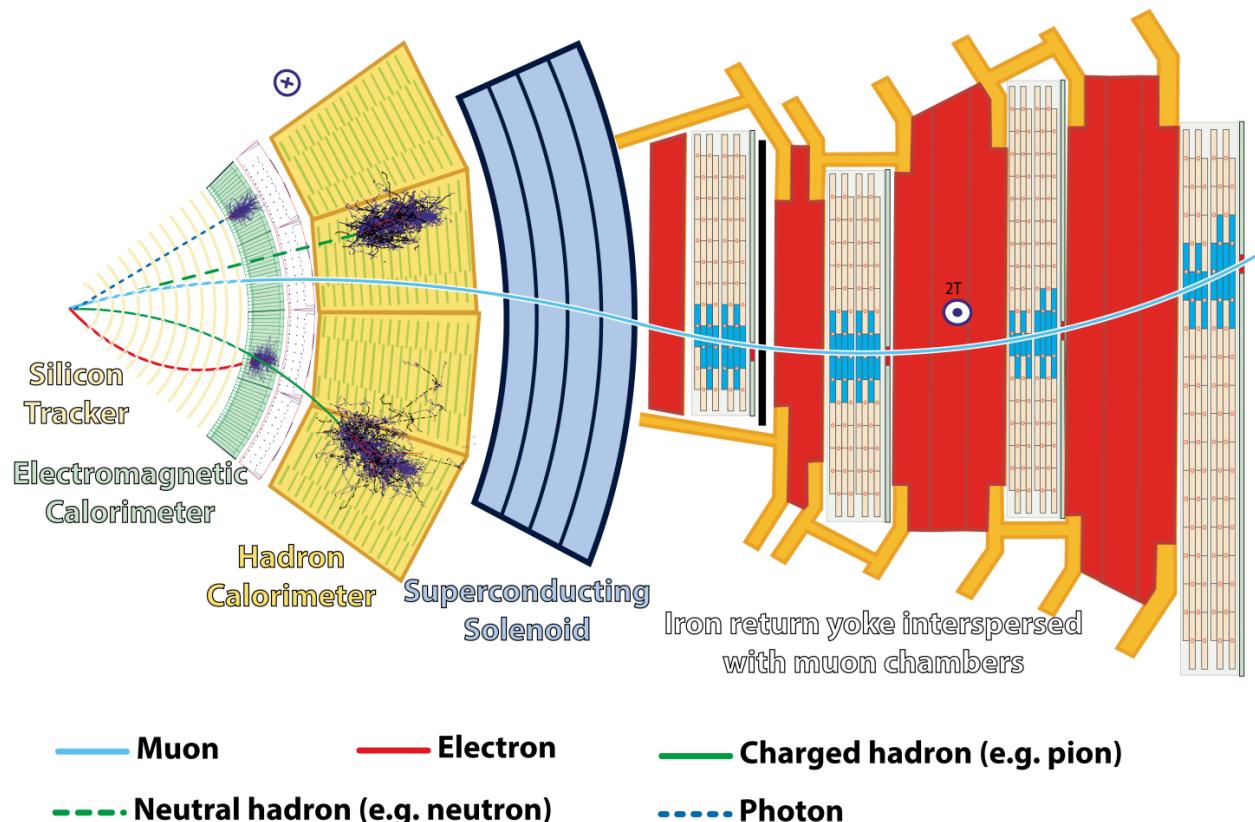


Figure 2.4: Schematic view of the CMS detector [22]

391

2.2.1 The CMS Coordinate System

392 CMS uses interaction point (IP) of the collisions as origin to define the right-handed
 393 coordinate system. The z -axis is along the beamline, the x -axis points toward the center of
 394 the LHC, and the y -axis points upwards toward the Earth's surface. The transverse plane
 395 $x - y$ is used to calculate most commonly used quantities like transverse momentum p_T and
 396 energy E_T .

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

$$\eta = -\ln[\tan \theta/2] \quad (2.2)$$

and sometimes in terms of rapidity y as,

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \quad (2.3)$$

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

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (2.4)$$

397 **2.2.2 The Superconducting Magnet**

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

401 Figure 2.5 shows the installation (in 2007) of the superconducting magnet with iron yoke
 402 in the underground cavern.

403 The key purpose of the magnet is to help determine the momentum and the sign of
 404 charged particles by bending them. The momentum resolution of the particles decreases
 405 with increase in their p_T . Constant magnetic field of 3.8 T gives momentum resolution of
 406 $\Delta p/p \approx 10\%$. This resolution is enough to determine unambiguously the sign of muons with
 407 momentum of $\approx 1 \text{ TeV}/c$.

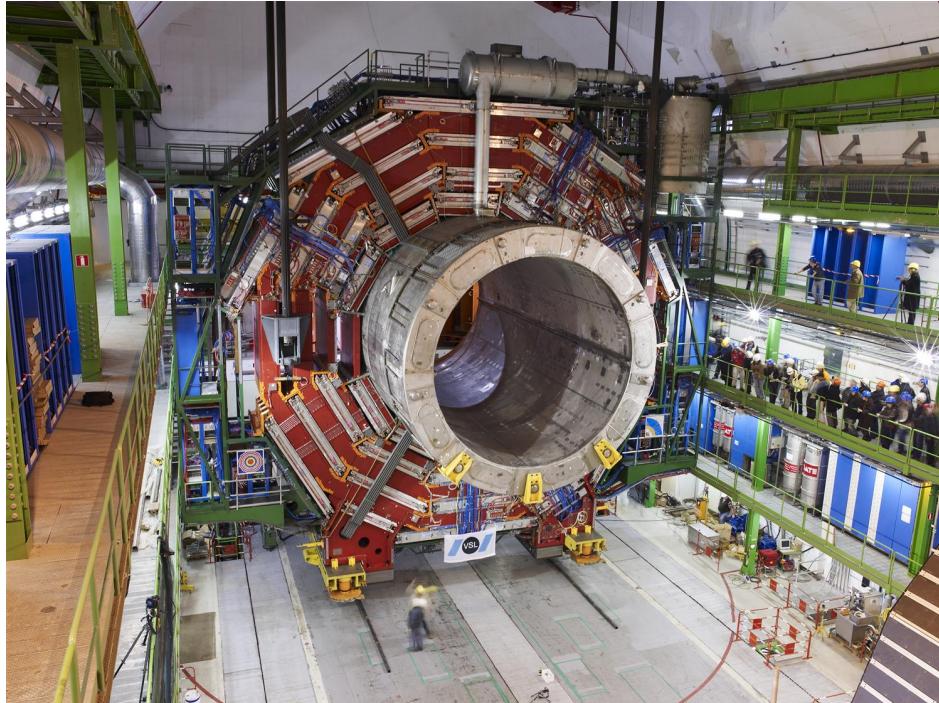


Figure 2.5: The picture of the CMS detector central part when lowered in underground cavern with superconducting magnet and iron yoke [23].

408

2.2.3 The Tracking System

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

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

419 The pixel detector was upgraded in 2017 and the comparison of layers before and after
 420 the upgrade is shown in Figure 2.6. It is made up of four barrel layers and three endcaps,
 421 with nearest barrel layer being 3 cm away from the beamline for precise measurement of the
 422 IP. Because of the large number of pixel channels, the readout is done by Application-specific
 423 integrated circuits (ASICs).

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

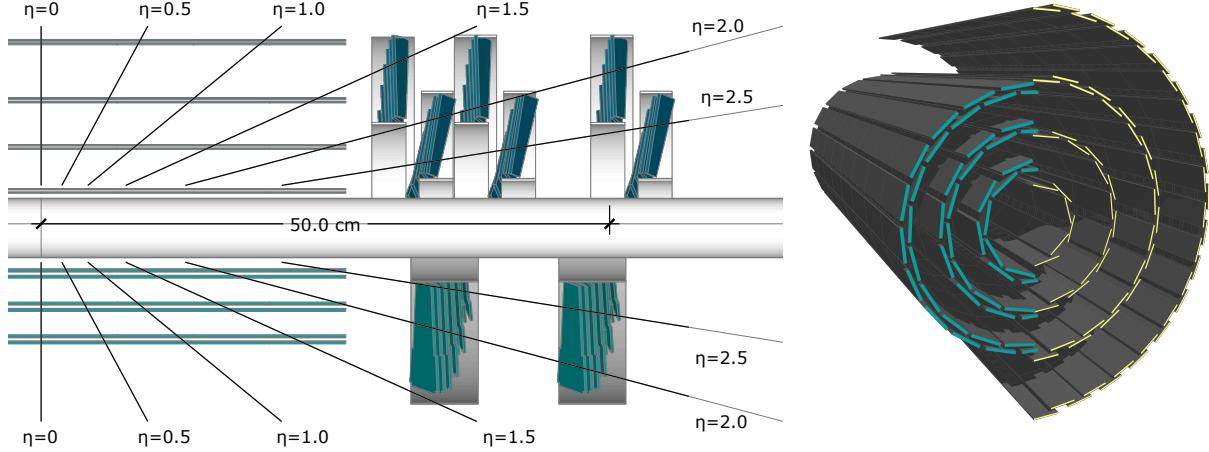


Figure 2.6: The CMS pixel upgrade. The left is cross sectional view of pixel detector layers before upgrade (bottom) and after Phase 1 upgrade (top). The right is view pixel barrel before upgrade (left) and after upgrade (right) [24].

2.2.4 The Electromagnetic Calorimeter

428 The ECAL active material is made of lead tungstate (PbWO_4) scintillating crystals and
 429 two layers silicon strip for preshower in front of the endcaps. The crystals in central barrel
 430 section are mounted in quasi-projective geometry pointing towards IP and cover $|\eta| < 1.48$.
 431 The two endcaps extends the coverage to $|\eta| = 3.0$. The schematic layout of ECAL is shown
 432 in Figure 2.7 and the picture of endcap quadrant when assembled is shown in Figure 2.8.
 433

434 The main purpose of the ECAL is to determine energy and positions of electromagneti-
 435 cally interacting particles. Except electron and photons all other particles pass through
 436 ECAL crystals with only small fraction of energy signature in the crystals. When electron
 437 and photon interacts with PbWO_4 it starts an process of electromagnetic shower which con-
 438 tinues until the energy the energy of the incident particle is below threshold, which is about
 439 1 MeV.

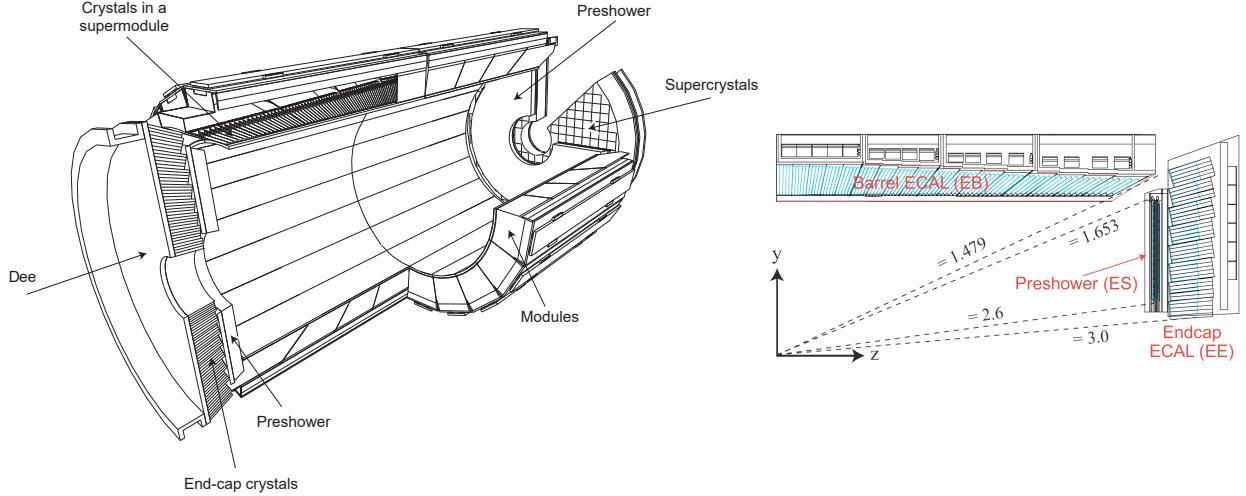


Figure 2.7: The CMS ECAL schematic layout. The left schematic shows the arrangement of superclusters in barrel and endcap (with preshower layers). On the right is the $y - z$ plane quarter view of ECAL layout [25].

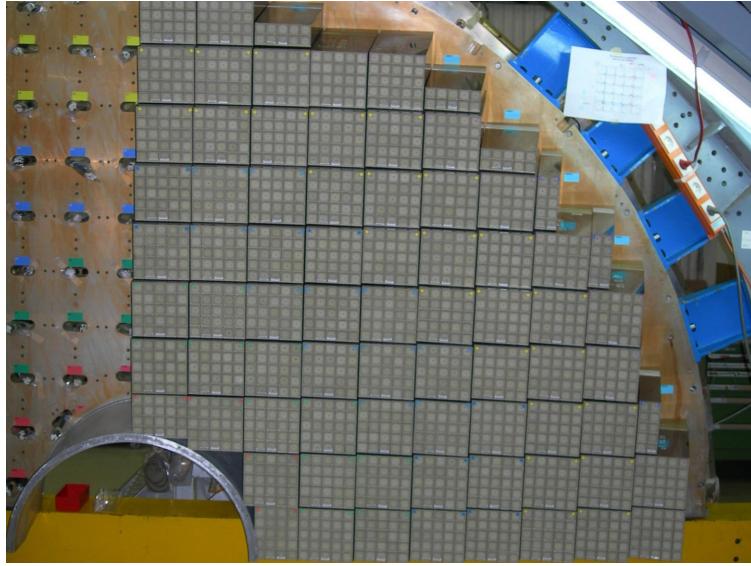


Figure 2.8: The ECAL endcap quadrant assembled view [26].

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

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (2.5)$$

⁴⁴⁰ where S is the stochastic term, N is related to the noise, and C is a constant offset.
⁴⁴¹ The energy resolution in the central region of ECAL barrel ($|\eta| < 0.8$) for electrons from Z
⁴⁴² decays was 2% and 2–5% elsewhere. The energy resolution for photons from Higgs decay
⁴⁴³ varied from 1.1% to 2.6% in the barrel region and from 2.2% to 5% in the endcaps [27].

⁴⁴⁴ **2.2.5 The Hadronic Calorimeter**

⁴⁴⁵ Similar to ECAL, the purpose of HCAL is to shower hadrons, and measure their energy
⁴⁴⁶ and position. The first half of barrel HCAL inserted into solenoid during installation is
⁴⁴⁷ shown in Figure 2.9.

⁴⁴⁸ HCAL is made up of towers pointing towards IP where each tower is made up of sampling
⁴⁴⁹ layers; with alternating layers of plastic scintillator and brass. Brass acts as absorber in the
⁴⁵⁰ HCAL and causes hadrons to shower. Scintillators emits light when secondary particles from
⁴⁵¹ hadronic shower pass through them. The light output collected from the scintillator layers,
⁴⁵² combined with total absorption length of Brass layers gives the total energy deposited in the
⁴⁵³ HCAL. In phase 1 upgrade the HCAL was upgraded to give energy deposit as a function of
⁴⁵⁴ depth. The depth segmentation schematic is show in Figure 2.10 and the details of upgrade
⁴⁵⁵ are in technical design report [28].

⁴⁵⁶ HCAL consists barrel (HB) and two endcaps (HE) located inside solenoid. HCAL covers
⁴⁵⁷ region with $|\eta| = 3.0$. There are two other subsystems of HCAL outside solenoid, a forward
⁴⁵⁸ HCAL (HF) and outer barrel HCAL (HO). HO was added to ensure there is no leakage from
⁴⁵⁹ the particles that make it past the solenoid. HF extends the coverage to $|\eta| = 5.0$. HF is
⁴⁶⁰ based on Cherenkov radiation principle and uses quartz fiber as active material with steel
⁴⁶¹ absorbers.



Figure 2.9: The first half of the barrel HCAL inserted into the superconducting solenoid (April 2006) [29].

462

2.2.6 Muon Detector

463

The outermost subsystem in the CMS detector is the muon detector. Unlike electrons, muons are MIPs i.e. they do not lose much of their energy while passing through tracker, calorimeter and solenoid. Muon detector is built to identify, measure momentum and trigger the events with muons. Like other subsystems, the muon detector consists of a barrel and endcap detector. The schematic layout is highlighted in the Figure 2.11. As shown in the schematic, the muon detector consists of three subsystems drift tubes (DTs), cathode strip chambers (CSCs) and resistive plate chambers (RPCs).

470

The DTs are wire gas detectors filled with Argon and composed of many tube cells of about 4 cm. Muon passing through these tubes ionize Argon and the free electron is detected on the wire cathode. Each DT is about 2 meters by 2.5 meters in size, and there are four

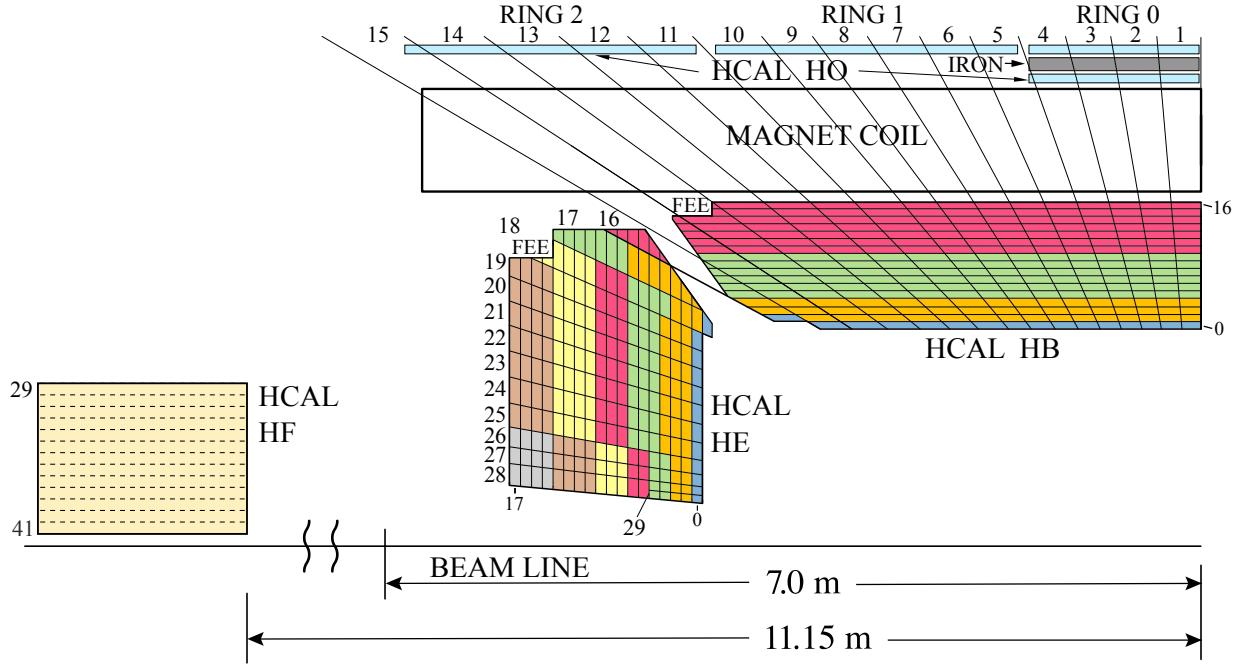


Figure 2.10: The HCAL depth segmentation after phase 1 upgrade [30].

473 layers of the DTs interleaved with the iron yoke, parallel to the beam pipe in barrel region.

474 The drift time is the order of about 380 ns.

475 The CSCs are based on same principle as DTs, and are made of multi-wire proportional
476 chambers consisting of 6 anode planes interleaved with 7 cathode planes. They have time
477 resolution better than 5 ns. The CSCs are used in endcap region, where radiation hardness
478 is required, and non uniform magnetic field does not effects the measurement.

479 The RPCs are made up of two oppositely charged high resistive parallel plates with a
480 gas² volume between them. When a charged particles passes through, it ionizes the gas, and
481 it creates an avalanche. The charge is collected by metallic readout strips. RPCs have poor
482 position resolution but fast readout of the order of 1 ns, which is fast compared to DTs, this
483 is the reason there are 1 or 2 RPCs attached to both DTs, and CSCs.

²mixture of 96.2% $C_2H_2F_4$, 3.5% Iso- C_4H_{10} , 0.3% SF_6

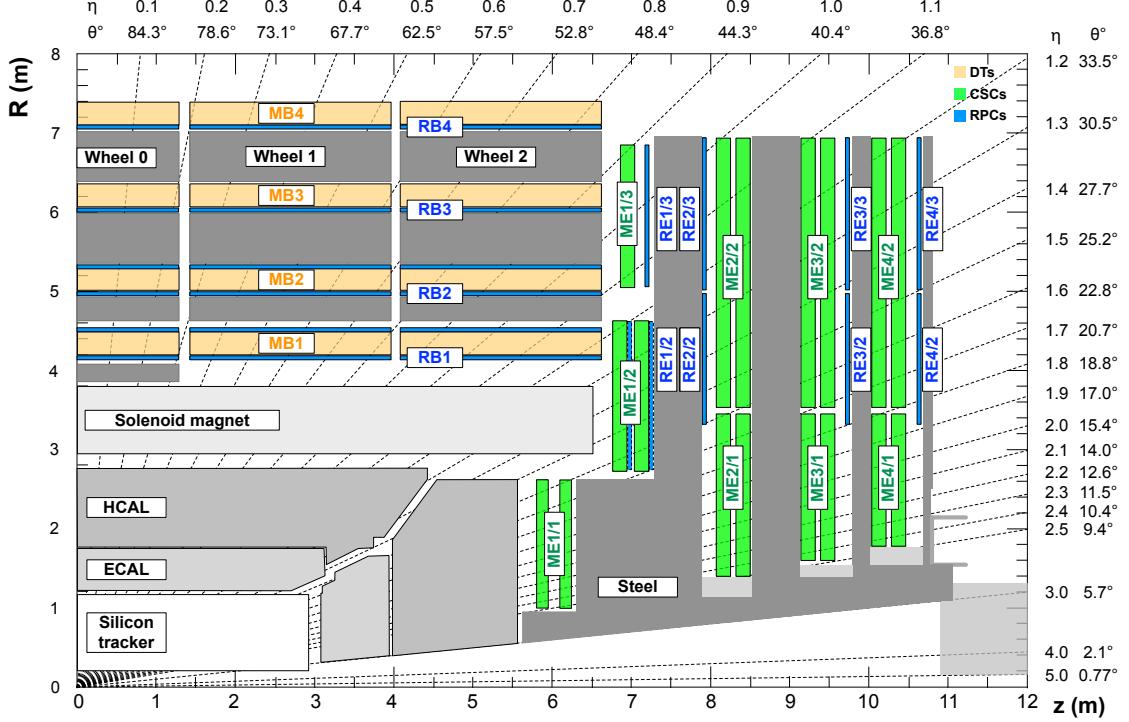


Figure 2.11: The quadrant view of CMS subdetectors layout, and the coverage of the muon detector DTs, CSCs, and RPCs highlighted [31].

484

2.2.7 Trigger

485 Since proton-proton collisions occurs every bunch crossing which are 25 ns apart; equivalent to a 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 do 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).

490 The L1T is the first electronic processing system through which event information is processed before it is passed to second trigger system: HLT. The L1T is designed to make a fast decisions in about $3.8\ \mu s$, and only uses ECAL, HCAL and muon system to make a decision regarding throwing or keeping an event. L1T cut downs the data rate from 40 MHz

494 to 100 kHz. The L1T electronics is placed next to the detector in underground cavern for
495 fast transfer of data.

496 Events accepted by L1T are passed to the HLT. The HLT further reduces the data rate
497 from 100 kHz to about 1 kHz using a computer farm with nearly 26000 cores. HLT uses all
498 the available information from the event to make decision in about 300 ms. HLT is modular
499 by design to allow the use of information from different systems to construct multiple paths
500 called HLT paths, for example the single muon HLT path will save event with at least one
501 muon passing the selection criteria set in HLT path. Events passing at least on HLT path
502 are saved for offline physics analysis.

CHAPTER 3

EVENT SIMULATION AND RECONSTRUCTION

505 The proton-proton collision at the LHC produces a shower of particles. Before the event
 506 information can be easily used in an analysis, the data collected goes through an iterative
 507 process of reconstruction of particles produced in the collision. CMS uses a particle flow (PF)
 508 algorithm to reconstruct 4-vectors of muons, electrons, photons, hadrons, jets and missing
 509 transverse momentum [32].

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

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

3.1 Track Reconstruction and Calorimeter Clustering

516 For complete particle reconstruction the two main ingredients are the track left by par-
 517 ticle in the detector and the energy deposit in calorimeter. This section describes track
 518 reconstruction from the hits in the tracker and muon detector, and measurement of energy
 519 deposit from calorimeter clustering.

520 Track reconstruction requires reconstructed hits, and seed generation which are described
 521 in [33]. The track reconstruction is done using pattern recognition which is an iterative
 522 process and based on a combinatorial Kalman Filtering (KF) method [34]. Starting from

523 first hits pair from initial layers are used to estimate a track. Next hits information of
 524 subsequent layers are included to grow the tracks into possible trajectories. There can be
 525 multiple hits in each new layer, for this multiple trajectory candidates are created. All the
 526 trajectory candidates are grown in parallel to avoid bias, and truncated at each layer to
 527 prevent exponential increase in number of candidates. Then finally the track is fitted to
 528 compute momentum and vertex information.

529 The main purpose of the calorimeter clustering is to determine the position and energy of
 530 the particle. A cluster in a calorimeter is a local group of energy deposits that are spatially
 531 consistent with a electromagnetic or hadronic shower. First the topological clusters are
 532 identified, a topological cluster is a contiguous region of energy deposits. Next seeds are
 533 identified in topological cluster with certain energy threshold and are local maxima within
 534 topological cluster. In ECAL seed is a crystals with energy threshold of 0.23 GeV and
 535 have highest energy deposit among the 8 neighboring crystals in a topological cluster. In
 536 HCAL seed is a tower with energy threshold of 0.08 GeV highest energy deposit among the
 537 4 neighboring towers in a topological cluster.

538 Each seed is grown into energy cluster by gathering neighboring crystals (or towers) en-
 539 ergy. For the case when there is one seed, there is only one iteration until all the neighboring
 540 crystals (or towers) are added. In case when there are more than one seeds, the energy from
 541 neighboring crystal (or towers) is shared. The fraction of energy shared depends on the
 542 energy and position of the clusters being reconstructed. Each iteration uses re-calculated
 543 energy and position of the clusters. Iteration continues until either the maximum iteration
 544 is reached or clusters energy and positions values have converged.

545

3.2 Reconstructed Particles

546 After tracks and calorimeter clusters are formed, PF links this information from the
 547 detectors together to form objects as broadly discussed in Section 2.2 and shown in Figure 2.4.
 548 This section describes the properties of those reconstructed particle candidates.

549

3.2.1 Muons

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

556 Once the muon candidates are found, the kinematic properties (p_T, η, ϕ) are calculated
 557 from track fitting. Other properties such as distance from primary vertex (PV) dxy, dz ,
 558 number of hits in the tracker and muon system, tracker based relative isolation (3.1) in a
 559 cone of $\Delta R = 0.3$, and PF based relative isolation (3.2) in a cone of $\Delta R = 0.4$ are stored for
 560 cleaning and isolating muons for physics analysis.

The tracker and PF based relative isolation are defined as,

$$\text{TkIso03} = \left(\sum p_T^{\text{Tracks (PV)}} \right) / (p_T^\mu) \quad (3.1)$$

$$\text{PFRellIso04} = \left(\sum p_T^{\text{CH (PV)}} + \min \left[0, \sum E_T^{\text{NH}} + \sum E_T^\gamma - 0.5 \sum p_T^{\text{CH (PU)}} \right] \right) / (p_T^\mu) \quad (3.2)$$

561 where “Tracks (PV)” refers to all the tracks in tracker and coming from PV, “CH (PV)”
 562 and “CH (PU)” refers to charged hadrons coming from PV and pile up (PU) respectively,
 563 “NH” refers to neutral hadrons, μ refers to muon, and γ refers to photon.

564 There are multiple source of muons whenever collision event happens, they can be real
 565 muons or hadrons which are misidentified as muons, these hadrons are able “punch” through
 566 HCAL and leaves hit in muon detector. The real muons of interest are called “prompt” muons
 567 and others are either usually referred as “fake” or “non-prompt”. Fake muons can originate
 568 from decay of pions and kaons in flight usually identified with a “kink” in track or from
 569 heavy flavor decay of b or c-quarks which are identified with tracks not originating form PV.
 570 The prompt muons are the ones coming from decay of H, W, Z bosons and τ leptons, and
 571 have small impact parameter from PV, have hits in both tracker and muon detector, and
 572 are typically well isolated.

573 Muon identification for loose and tight muons follows CMS Muon Physics Object Group
 574 recommendation and are defined in Reference [35].

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

578 3.2.2 Electrons and Photons

579 Since there is large amount of material in the tracker, electrons often emit bremsstrahlung
 580 photons when passing through tracker volume, and energetic photons can decay to e^-e^+
 581 pair which complicates the tracking algorithm. The energy deposit of such electrons emit-
 582 ting bremsstrahlung will have large spread in ϕ direction because the magnetic field will
 583 bend electrons in ϕ whereas photons are unaffected. For this reason electron and photon

584 reconstruction are done together, and the Gaussian-sum filter (GSF) algorithm is used for
 585 electron track reconstruction which takes care of kinks in electrons track because of hard
 586 emission [36].

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

594 Similar to muons, after electron and photon reconstruction is done, their kinematics
 595 properties are calculated and various other properties required for cut based and Multivariate
 596 analysis (MVA) based identification are stored. The detailed description of electrons and
 597 photons identification technique and properties used in this dissertation can be found in
 598 Reference [37].

599 3.2.3 Hadrons and Jets

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

605 Jet in CMS are reconstructed using the FASTJET package [38], which takes as input all of
 606 the PF candidates and the associated tracks. The clustering start with combining 4-vectors

607 of the two particles i and j , if the distance between them (d_{ij}) is lower than stopping distance
 608 (d_{iB}). Then particles i and j are removed from the input collection, and the clustering is
 609 continued with combined ij and the next candidate in the collection. The clustering stops
 610 when, $d_{ij} > d_{iB}$ and i is called a jet.

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

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (3.3)$$

$$d_{iB} = p_{Ti}^{2p} \quad (3.4)$$

611 where p is the parameter for different clustering algorithms, R is the cone size, and ΔR_{ij}
 612 is distance between two particles in iteration.

613 Anti- k_T (AK) is the most used jet algorithm in CMS physics analyses. This corresponds
 614 to $p = -1$, which means the hard particles will be clustered first in this clustering algorithm.
 615 The cone size used for standard jets (AK4) is $R = 0.4$, and for large jets (AK8) often called
 616 “fatjet” is $R = 0.8$.

617 To mitigate the effect of PU contamination in jets two most commonly used techniques
 618 are charged hadron subtraction (CHS) and Pileup Per Particle Identification (PUPPI) [39].
 619 CHS as the name suggests removes all the PF particles in the jet clustering which are
 620 originating from PU vertices, and it is a standard technique for AK4 jets. PUPPI rescales
 621 4-vectors of all the particles in a jet with a weight. The PUPPI weight is derived from a
 622 local metric which is built using information of charged hadrons whether they are originating
 623 from leading vertex or PU. The main limitation of CHS is that it only removes charged PU
 624 contribution, for larger jets it can be issue, since it is clustering larger number of particles
 625 and can have significant contribution from neutral hadrons, for this reason PUPPI technique
 626 is used for AK8 jets.

627 Jet energy corrections are applied to the reconstructed jets to correct for jet energy scale
 628 (JES) differences in MC and data events. Three levels of corrections are applied. The first
 629 correction that is basically for the removal of pileup and electronic noise and is applied to
 630 both MC and data events. The second correction applied only to MC events corrects for jet
 631 response as a function of p_T and η to the particle level jet response. Third correction applied
 632 only to collision data are derived from the differences in p_T and η response of jets in data
 633 and MC samples, using $\gamma + \text{jets}$, $Z + \text{jets}$, and dijet MC events [40].

634 To improve the jet selection and reject jets originating purely from PU two methods
 635 are used in this dissertation, jet identification based on multiplicities and energy fraction
 636 of particles contained in the jet, and MVA based PU identification which uses jets shape
 637 variables to discriminate prompt jet from pileup jets. The details of PU mitigation and
 638 identification used in CMS are described in Reference [41].

639 After jet reconstruction is complete, in addition to calculating kinematic properties (p_T ,
 640 η , ϕ and mass), various other properties such as b-quark tagging and quark-gluon likelihood
 641 are also calculated and stored.

642 **3.2.3.1 N-Subjetiness and Deep Taggers**

643 The origin of fatjets are usually when heavy energetic particle such as W or Z bosons
 644 decays to a pair of quarks and the resulting jets overlap each other. This usually happens
 645 when decaying the particle is boosted in lab frame, and the fatjet is usually referred as
 646 boosted jet. To find and discriminate the fatjet of interest based on its substructure the
 647 two common technique used in CMS are N-Subjetiness [42] and “deep tagger” [43].

N-Subjetiness is defined as,

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

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,

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

648 where R_0 is the jet radius ($= 0.8$, for fatjet).

649 τ_N quantifies to what degree a jet can be regarded as made of N jets. The small values

650 of the τ_N means a jet is more likely to have N or less subjets, and higher value means it will

651 at least have $N + 1$ subjets. Rather than using τ_N alone, ratio of different τ_N variables is

652 preferred for rejecting QCD jets. Figure 3.1 shows distribution of $\tau_{21} = \tau_2/\tau_1$ and $\tau_{32} = \tau_3/\tau_2$

653 shapes for different MC samples. τ_{21} is used to discriminate fatjets with 2-prongs (W/Z/H)

654 and τ_{32} with 3-prongs (t-quark) substructure against QCD jets.

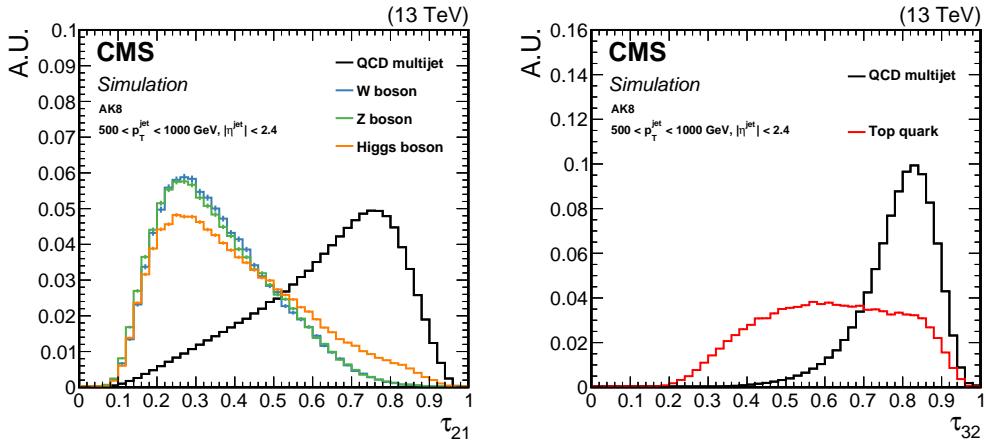


Figure 3.1: Comparison of τ_{21} and τ_{32} shapes for signal and background in AK8 jets. The left is τ_{21} distribution showing discrimination W/Z/H jets vs QCD jet, and the right is τ_{32} distribution for t-quark vs QCD jets [43].

655 Deep Tagger for AK8 is a Machine learning (ML) based tagger developed to determine
 656 the origin of a fatjet. Figure 3.2 describes the architecture of “DeepAK8” tagging. These
 657 taggers are trained on particle level information from PF, and provide multi class tagging
 658 probabilities such as Z vs. QCD, W vs. QCD, etc. Since deep taggers utilize deep neural
 659 network with particle level information their classification power is higher than N-subjetiness,
 660 and are worth the study for future analysis.

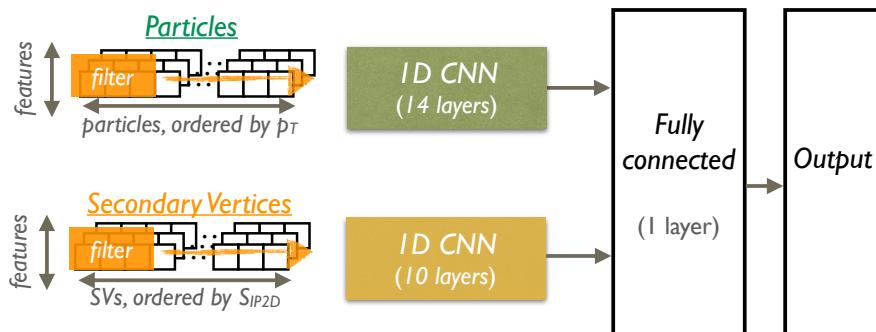


Figure 3.2: The network architecture of DeepAK8 [43]

661 3.2.3.2 Softdrop Mass

662 Fatjets can also have contamination coming from wide angle soft initial state radiation
 663 (ISR) and multiple hadron scattering, which affects the mass calculation of the jet, to remove
 664 such contamination and have better mass reconstruction, the “softdrop” mass algorithm [44]
 665 is used.

Softdrop is a declustering algorithm which removes the particle from the jet with radius R_0 ($= 0.8$, for fatjet), when the following condition between two particles is satisfied,

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0} \right)^\beta \quad (3.7)$$

666 where ΔR_{12} is the distance between the two particles, z_{cut} and β are the parameters for
 667 tuning softdrop declustering. For fatjets used in CMS, the softdrop algorithm is applied with
 668 $\beta = 0$ and $z_{cut} = 0.1$ which vetoes both soft and soft-collinear emissions in a jet.

669 **3.2.4 Missing transverse momentum**

670 Invisible particles like neutrinos cannot be detected at CMS directly. Kinematics of such
 671 particles can be determined using the law of conservation of total momentum. In case of
 672 proton-proton collision, the actual collision happens between quarks contained in proton
 673 which carry an undetermined fraction of proton momentum. For this reasons kinematic
 674 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,

$$\vec{p}_T^{miss} = - \sum \vec{p}_T \quad (3.8)$$

675 It's usually neutrinos which contributes to missing transverse momentum and they have
 676 very small mass, the missing transverse momentum is then equivalent to Missing Transverse
 677 Energy (MET), which is most often used term in physics analyses. Experimentally, noise
 678 in detector or beam halo of colliding beams can cause anomalously large missing transverse
 679 momentum. There are various algorithms available in CMS that can suppress such events.

680

3.3 Monte Carlo (MC) Simulation

681 In the proton-proton collision, determination of exact underlying process from recon-
 682 structed final state is not possible. To do measurements on the process of interest, we need
 683 to quantify how much is the signal (process of interest) and how much is the background
 684 (processes with same final state and similar kinematics). MC simulations are employed to
 685 generate underlying processes that can help in the modeling of signal and background process
 686 and compare them against the collision data.

687 Main tools and framework used at CMS for generating MC events are described in this
 688 section.

689

3.3.1 Generators

690 There are three main components to the event generator, parton distribution function
 691 (PDF), matrix element and parton shower. PDF encodes the probability of finding a patron
 692 with given fraction of energy at given proton momentum. Matrix element encodes the
 693 theoretical cross-section of the hard process.

694 Our signal $ZVjj$ is a pure EW process generated at leading order (LO) with MAD-
 695 GRAPH5_aMC@NLO and MADSPIN [45, 46]. At LO this process has six EW vertices, and
 696 zero QCD vertices. The generation is done in two steps with both vector boson generated on
 697 shell with MADGRAPH and decayed to leptons and quarks via MADSPIN. For example, to
 698 generate ZZ process, the syntax used is,

```
$ madgraph >
generate p p > z z j j QED=4 QCD=0
$ madspin >
decay z > l+ l-
decay z > j j
```

699 where p and j are defined as collection of quarks of gluons, z is the on-shell Z boson, and
 700 l^+ , l^- are the leptons consisting of electrons, muons and taus. The syntax indicates $\text{QED}=4$
 701 $\text{QCD}=0$ implies that there are four EW vertices, since we are first generating Z on-shell, and
 702 zero QCD vertices. The last two EW vertices comes from decay of Z bosons to fermions.

703 **3.3.2 Parton Shower and Hadronization**

704 After events are generated with the hard process of interest, they go through the process of
 705 parton shower and hadronization to simulate QCD bremsstrahlung radiation and make final
 706 state hadrons to preserve overall color charge of an event. This is done because quarks and
 707 gluons coming from hard process scattering cannot stay free and produce quark-antiquark
 708 pair or radiate gluons. The parton shower software simulates this process until partons have
 709 reached low energy of the order of 1 GeV.

710 After all the partons have showered, they go through the process of hadronization to
 711 produce final state hadrons. Hadronization is done using the models that are tuned on
 712 experimental data. Most of the models used for MC in standard model physics at CMS are

713 based on the Lund string model in the PYTHIA framework [47]. Most showers are usually
 714 collimated and they appear as jets in the final state, but sometimes a highly energetic gluon
 715 can radiate a high energy gluon which will eventually make a new jet, it can be either ISR
 716 during the hard process, or final state radiation (FSR) during the parton shower. This
 717 increases jets multiplicity of an event, since we have four jets in our final state this step is
 718 important for accurate modeling of our background processes.

3.3.3 Tunes

720 In hadron-hadron collisions there are two additional interactions, beam-beam remnants
 721 (BBRs) and multiple-parton interactions (MPIs), they are modeled as underlying event
 722 (UE). BBR is what left behind by incoming beam hadron and does not take part in ISR or
 723 hard scattering, and these remnants need to be color connected with the rest of the event.
 724 The MPIs are the additional multiple interaction that a parton can go through in the same
 725 hadron-hadron collision, these interactions can be hard or semi-hard. To model UE properly
 726 in standard MC event generator, a set of parameters (referred as “tune”) are needed to be
 727 adjusted so that it provides better fit to the experiment data. CMS made use of the tune
 728 TuneCUETP8M1 [48] (2016 MCs samples) and TuneCP5 [49] (2017 and 2018 MCs samples) in
 729 the PYTHIA for modeling UE.

730

CHAPTER 4

731

VBS MEASUREMENT IN ZVJJ FINAL STATE

732 As discussed in Section 1.2 this analysis targets VBS of ZV with two jets in final state.

733 The goal of the analysis is to reduce contribution from background processes as much as

734 possible without losing much of signal, and measure signal strength and significance.

735 Since Z decays leptonically and V is decaying hadronically, the phase-space of this analysis

736 can be either l^+l^-jjjj or l^+l^-Jjj , where l are leptons (electrons or muons), j are narrow jets

737 and J is a boosted (wider) jet. The phase-space is divided into two broad regions signal and

738 controlled, signal region is constructed based on theory such that it is mostly signal process

739 and control region is basically orthogonal to the signal region where we expect contributions

740 mostly from background processes. The main purpose of control region in this analysis is to

741 provide normalization factors the dominant background.

742 The analysis is performed “blind” to avoid intrinsic bias i.e. until the analysis procedure

743 is finalized, the collision data is only used in control regions. Once the analysis technique is

744 optimized using MC samples and validated against collision data in control regions. Once

745 the analysis technique is satisfactory and approved by CMS Physics Group then the results

746 are “un-blinded” i.e. measurements are done using collision data in signal region.

747

4.1 Dataset and Simulation

748 As discussed in Section 2.2.7 only events those pass Level-1 Trigger and HLT paths
 749 are saved for further processing, MC simulation also have these identical step during event
 750 generation to mimic Level-1 Trigger and HLT paths.

751 CMS processes the datasets centrally and provides various tiers of datasets such as
 752 “RECO” datasets, which contains reconstructed objects and no skimming. Average size
 753 of an event saved at “RECO” tier is 480 kB per event and on average an analysis will pro-
 754 cess 3 billion events, which makes this tier not practical in terms of computer processing time
 755 and storage if each analysis starts from “RECO”. CMS centrally processes these datasets
 756 further and removes certain objects or features to reduce the average event size but still
 757 covering majority of the analysis to make use of the reduced datasets.

758 This analysis uses NANOAOD tier with version “v7” of datasets which has average event
 759 size of 2–3 kB.

760

4.1.1 Data

761 The collision data events used in this analysis are all certified by CMS Data Quality
 762 Monitoring (DQM) and Data Certification (DC) group. The primary trigger object in HLT
 763 paths are leptons p_T threshold, and since in our final state we are looking for Z boson decaying
 764 into two leptons, we require single and double lepton trigger for our analysis. Depending
 765 on the detector conditions and LHC storage capacity we have slightly different threshold in
 766 triggers across different years. Table 4.1 contains the list of HLT paths used in this analysis.

Table 4.1: Trigger paths used to select events in CMS collision data

Dataset	Year	HLT Path	p_T	Threshold
Single Muon	2016	HLT_IsoMu24		24 GeV
		HLT_IsoTkMu24		24 GeV
	2017	HLT_IsoMu27		27 GeV
	2018	HLT_IsoMu24		24 GeV
Single Electron	2016	HLT_Ele27_WPTight_Gsf		27 GeV
		HLT_Ele25_eta2p1_WPTight_Gsf		25 GeV
	2017	HLT_Ele35_WPTight_Gsf		35 GeV
	2018	HLT_Ele32_WPTight_Gsf		32 GeV
Double Muon	2016	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ		17, 8 GeV
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ		17, 8 GeV
		HLT_TkMu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ		17, 8 GeV
	2017	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8		17, 8 GeV
		HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8		17, 8 GeV
	2018	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8		17, 8 GeV
Double Electron	2016	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ		23, 12 GeV
		HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL		23, 12 GeV
	2017	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL		23, 12 GeV
	2018	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL		23, 12 GeV

767

4.1.2 MC Simulations

768 The EW VBS process which is our signal is generated with MADGRAPH5+PYTHIA8
 769 at LO with α_{EW}^6 order i.e. all vertices in tree level Feynman diagram are EW vertices.
 770 The QCD induced VBS background process which is very similar to our signal is generated
 771 with same configuration but with $\alpha_{EW}^4 \alpha_{QCD}^2$ order i.e. two of the six vertices are QCD.
 772 The dominant background to analysis Drell-Yan (DY) + Jets and, top-quark based pro-
 773 cesses consisting of single top-quark (t and s-channel), single top-quark in association with
 774 W boson (tW) and top-quark pair ($t\bar{t}$) production. The DY + Jets are generated at LO

775 with **MADGRAPH5+PYTHIA8** in bins of HT i.e. scalar sum of all the jets p_T in the event,
776 to have more statistics for higher HT bins. Top-quark background process are generated at
777 next-to-next-to-leading order (NNLO), s-channel is generated with **MADGRAPH5_aMC@NLO**
778 +**PYTHIA8** and others t-channel, tW, t \bar{t} are generated with **POWHEG+PYTHIA8**. The com-
779 plete list of Table 4.2, 4.3, 4.4 contains the list of Signal and Background MC samples used
780 for modeling in this analysis.

Table 4.2: List of MC samples for Signal and Background modeling

Process	Year	Cross Section (pb)	
VBS_EWK (Signal)	2016	WminusTo2JZTo2LJJ_EWK_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.02982
	2016	WplusTo2JZTo2LJJ_EWK_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.05401
	2016	ZTo2LZTo2JJ_EWK_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.01589
	2017, 2018	WminusTo2JZTo2LJJ_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.02982
	2017, 2018	WplusTo2JZTo2LJJ_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.05401
	2017, 2018	ZTo2LZTo2JJ_EWK_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.01589
	2016	WminusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.3488
	2016	WplusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.575
	2016	ZTo2LZTo2JJ_QCD_LO_SM_MJJ100PTJ10_TuneCUETP8M1_13TeV-madgraph-pythia8	0.3449
	2017, 2018	WminusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.3488
VBS_QCD (Background)	2017, 2018	WplusTo2JZTo2LJJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.3488
	2017, 2018	ZTo2LZTo2JJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.575
	2017, 2018	ZTo2LZTo2JJ_QCD_LO_SM_MJJ100PTJ10_TuneCP5_13TeV-madgraph-pythia8	0.3449
	2016	DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	169.9
	2016	DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	147.4
DY + Jets LO (Background)	2016	DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	41.04
	2016	DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5.674
	2016	DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.358
	2016	DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.6229
	2016	DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.1512
	2016	DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	0.003659
	2017	DYJetsToLL_M-50_HT-70to100_TuneCP5_13TeV-madgraphMLM-pythia8	167.33
	2017	DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	161.1
	2017	DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	48.66
	2017	DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	6.968
	2017	DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	1.743
	2017	DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.8052
	2017	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.1933
	2017	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.003468

Table 4.3: List of MC samples for Signal and Background modeling

Process	Year	Cross Section (pb)
DY + Jets LO (Background)	2018	DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8
	2018	0.003468
	2016	ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV_PSweights_powheg-pythia8
Top (Background)	2016	ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
	2016	ST_tW_top_5f_NoFullyHadronicDecays_13TeV_PSweights_powheg-pythia8
	2016	ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
	2016	ST_t-channel1_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
	2016	ST_t-channel1_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8
	2016	ST_s-channel1_4f_leptonDecays_13TeV_PSweights_amcatnlo-pythia8
	2016	ST_s-channel1_4f_hadronicDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8
	2016	ST_s-channel1_4f_InclusiveDecays_13TeV-amcatnlo-pythia8
	2016	TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8
	2016	TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8
	2016	TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8
	2017	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
	2017	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8
	2017	TTToHadronic_TuneCP5_13TeV-powheg-pythia8
	2017	ST_s-channel1_antitop_leptonDecays_13TeV-PSweights_powheg-pythia
	2017	ST_s-channel1_top_leptonDecays_13TeV-PSweights_powheg-pythia
	2017	ST_t-channel1_antitop_5f_TuneCP5_PSweights_13TeV-powheg-pythia8
	2017	ST_t-channel1_top_5f_TuneCP5_13TeV-powheg-pythia8
	2017	ST_tW_antitop_5f_InclusiveDecays_TuneCP5_13TeV-powheg-pythia8
	2017	ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8

Table 4.4: List of MC samples for Signal and Background modeling

Process	Year	Dataset Name	Cross Section (pb)
Top (Background)	2018	TTToSemileptonic_TuneCP5_13TeV-powheg-pythia8	365.34
	2018	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	86.99
	2018	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
	2018	ST_s-channel_antitop_leptonDecays_13TeV-PSweights_powheg-pythia	1.33
	2018	ST_s-channel_top_leptonDecays_13TeV-PSweights_powheg-pythia	2.13
	2018	ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8	27.19
	2018	ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8	45.7
	2018	ST_tW_DS_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	12.04
	2018	ST_tW_DS_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8	12.04

781

4.2 Event Selection

782 In the first stage of selection, events are selected if there are minimum number of objects
 783 for the analysis categories i.e. at least two leptons of same flavor ($p_T > 10 \text{ GeV}$), and either
 784 four narrow jets (AK4) (Resolved ZV category) or two narrow jets (AK4) plus one wider jet
 785 (AK8) (Boosted ZV category).

786 After initial skimming step, first loose leptons (muons and electrons) are vetoed and tight
 787 leptons are selected, if there are exactly two tight leptons of same flavor, and oppositely
 788 charged then the event is selected, and Z candidate is constructed from these two leptons
 789 which are OSSF. And events with Z candidate mass in the range [76, 106] GeV are selected.

790 Following are the loose and tight selections for the muons and electrons:

- 791 • **Loose Muons:** Muons with $p_T < 10 \text{ GeV}$, $|\eta| > 2.4$, failing loose ID [35], or PF
 792 relative isolation in cone $R = 0.4$ (“`pfRelIso04`”) more than 0.40.
- 793 • **Tight Muons:** Muons with $p_T > 20 \text{ GeV}$, passing tight ID [35], “`pfRelIso04`” less than
 794 0.15 and impact parameters $d_{xy} < 0.01, d_z < 0.1$.
- 795 • **Loose Electrons:** Electrons with $p_T < 10 \text{ GeV}$, $|\eta| > 2.5$, or failing loose cut based
 796 ID [37].
- 797 • **Tight Electrons:** Electrons with $p_T > 20 \text{ GeV}$ and MVA based ID depending on year
 798 of the dataset. For 2016 dataset: passing “`mvaSpring16GP_WP90`” (MVA based ID,
 799 90% efficiency) and “`pfRelIso03_all`” less than 0.0571 for $|\eta| > 1.479$ and less than
 800 0.0588 otherwise. For 2017 and 2018 dataset: passing “`mvaFall17V2Iso_WP90`” (MVA
 801 based ID + Isolation, 90% efficiency) and “`pfRelIso03_all`” less than 0.06 is required.

802 **4.2.1 VBS Tagged Jets and V Jet Candidate**

803 After selecting leptons and constructing Z candidate, first we look for V as a “FatJet”
 804 (merged jet, AK8) candidate with p_T of 200 GeV, within $|\eta| < 2.4$ and in softdrop mass
 805 range of [40, 150] GeV. Then all the FatJets are checked for isolation from selected leptons
 806 with $\Delta R = 0.8$ and then N-subjetiness $\tau_{21} < 0.45$ cut is applied. Finally a FatJet with mass
 807 nearest to the average mass of W and Z is selected.

808 • **Boosted:** Now after above described selection, if we have a FatJet as a candidate for
 809 V then the event is categorized as “boosted”.

810 • **VBS Jets:** Now whether the event is a boosted event or not, narrow jets (AK4) are
 811 considered to be tagged as VBS jets, this is done by first cleaning jets by applying
 812 tight “JetID”, tight “pileup JetID” (PUID), $p_T > 30$, and checking isolation against
 813 selected leptons with $\Delta R = 0.4$, selected FatJet (if any) with $\Delta R = 0.8$ and other jets
 814 with $\Delta R = 0.4$. Then all the combinations of jets are checked as a pair, to find the
 815 pair with highest invariant mass with minimum of 500 GeV, and the jets forming that
 816 pair are tagged as VBS jets. If no such pair is found, the event is not selected.

817 • **Resolved:** If there is no boosted event but there are VBS tagged jets, then from the
 818 rest of jet collection, a pair is searched with each jets within $|\eta| < 2.4$, and invariant
 819 mass nearest to the average mass of W and Z and in the range of [40, 150] GeV. Jets
 820 from selected pair are then labelled as V jet candidates and event is categorized as
 821 “resolved”.

822

4.3 Control and Signal Region

823 To model our background as accurate as possible, the selected events are divided into
 824 two regions based usually a observed property of the selected objects.

825 For this analysis, the mass window of V is selected to define the region boundaries and
 826 two regions are defined as:

- 827 • **Signal Region:** $65 < M_V < 105$ GeV, it's an 20 GeV window around the mass of V .

828 This is a narrow mass window, and we expect cleaner signal in this region.

- 829 • **Control Region:** $40 < M_V < 65$ and $105 < M_V < 150$ GeV, in these mass sidebands
 830 around signal region are rich in background. This region is dominant in DY plus jets
 831 events and will be used to model and normalize this background process.

832

4.3.1 MC Scale Factors

833 Whenever we apply some cut on object properties on data and MC, the efficiency of
 834 the cut can be different for the data than it's for MC. To account for this difference in
 835 efficiency of data and MC, scale factors are calculated and applied to MC. The scale factors
 836 are usually function of p_T and η of the object being used for selection. These scale factors
 837 are then applied as event weight i.e. they are multiplied to the cross-section weight factor
 838 applied to the MC events. Following are the scale factors used in this analysis:

- 839 • **Pile-Up Reweighting:** During run2 data taking at CMS the average number of
 840 pileup in an event can varied from 5 to 100, and to have same shape for this variable
 841 in MC, reweighting scale derived from the difference is applied to MC.

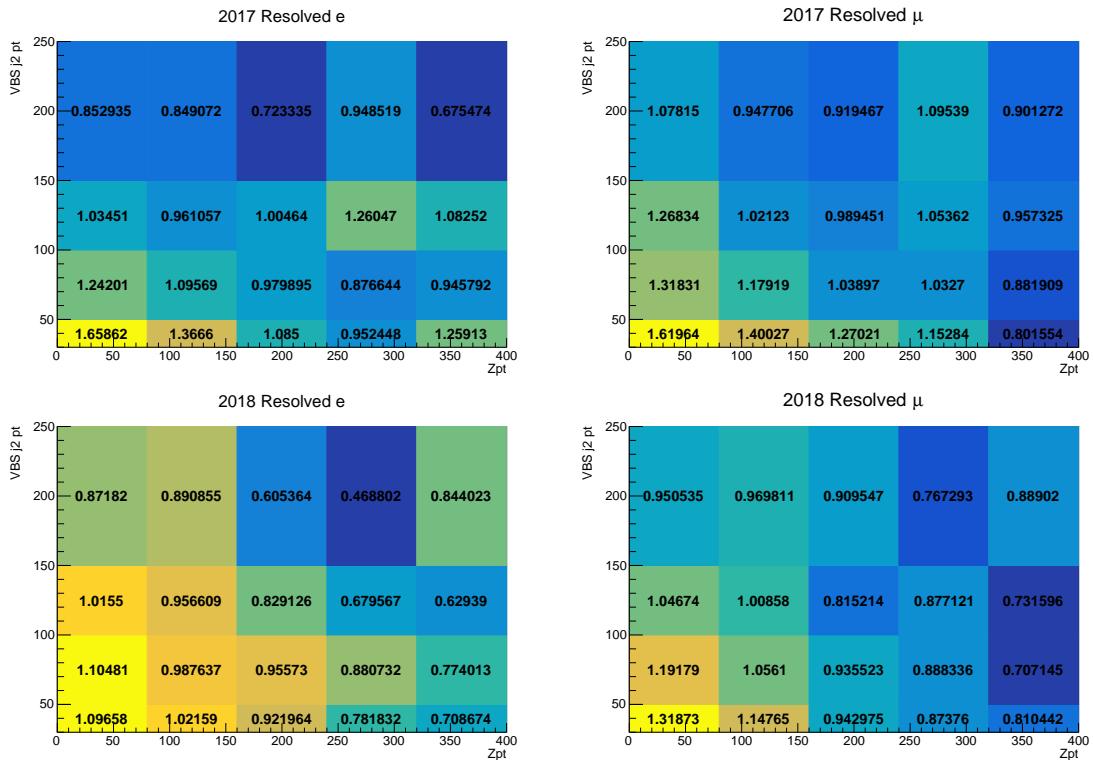
- 842 • **Trigger Efficiency:** Since the HLT trigger paths are defined by the p_T threshold of
843 the object in the trigger, they can have different efficiency in data and MC.
- 844 • **Lepton Identification/Isolation:** Scale factors to account for tight selection of elec-
845 trons and muons are derived from “tag and probe” method and applied to the MC.
- 846 • **Tight Pile-Up Jet Identification:** Since to reduce jets coming from pileup vertex,
847 tight Pile-Up Identification is applied to the VBS tagged jets, a scale factor is calculated
848 for both VBS jets and their product is applied to MC.
- 849 • **L1 Prefire Correction:** L1 Trigger prefiring scale factors recommended by CMS for
850 2016 and 2017 year datasets [50].

851 4.3.2 DY+Jets Normalization

852 To normalize and model our main background MC DY+Jets, we extract scale factors
853 from bins of variables in control region and apply it to both control and signal region. For
854 boosted category of all years 2016, 2017, 2018 and resolved in 2016 1D bins of p_T of Z were
855 used, and for resolved in 2016 and 2017 2D bins of p_T of Z and VBS trailing jet were used.
856 Table 4.5 lists the normalization factors extracted in 1D bin of p_T of Z and Figure 4.1 lists
857 factors extracted in 2D bin of p_T of Z and VBS trailing jet.

Table 4.5: Normalization factors from 1D bin of p_T of Z

Z p_T bin	Resolved				Boosted			
	2016		2016		2017		2018	
	e	μ	e	μ	e	μ	e	μ
[0, 80]	1.11	1.19	1.09	1.03	0.92	1.02	0.86	1.01
[80, 160]	1.07	1.12	1.00	1.03	0.83	0.84	0.86	0.81
[160, 240]	1.14	1.13	1.04	0.88	0.82	0.80	0.75	0.68
[240, 320]	0.93	1.14	0.83	1.15	0.99	0.93	0.73	0.76
[320, 400]	1.07	1.08	0.73	0.82	1.23	0.85	0.70	0.76
[400, 480]	1.02	1.06	0.62	0.70	0.91	0.84	0.67	0.67
[480, inf]	0.84	0.81	0.40	0.56	0.45	0.76	0.64	0.68

Figure 4.1: Normalization factors from 1D bin of p_T of Z and VBS trailing jet. Top to Bottom: 2017 and 2018 resolved. Left to Right: electron and muon channel.

858

4.4 Kinematics Distributions

859 To understand how the kinematics of various final state objects and derived candidate
860 behave, they are studied in the control region by stacking up all the background histograms
861 and overlaying them with the data as points. Signal MC is also overlayed on the histograms,
862 since it's small, for visualization purposes it's scaled with a factor shown in the legends of
863 the plots.

864 The kinematics plots also include a ratio plot of data and total background MC, and
865 uncertainties in histograms of MC are shown as shaded region which includes total statistical
866 bin error, total JES and theoretical renormalization (QCD scale) error on the DY+Jets and
867 VBS_QCD MC samples.

868 Section 4.4.1 and 4.4.2 shows kinematics plots in control region for boosted and resolved
869 ZV category respectively.

4.4.1 Boosted ZV DY+Jets Control Region

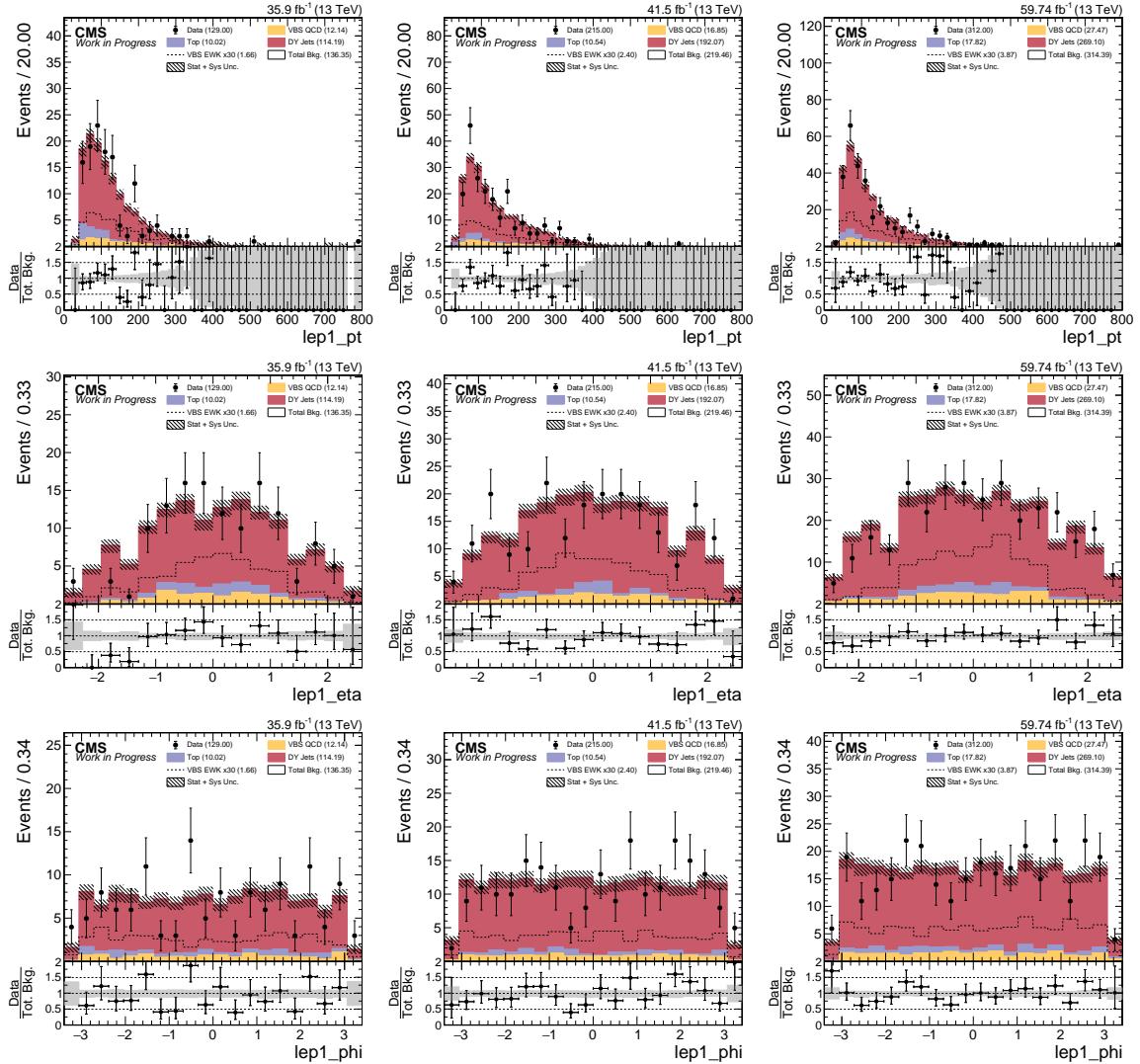


Figure 4.2: DY+Jets Control Region: Leading electron kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

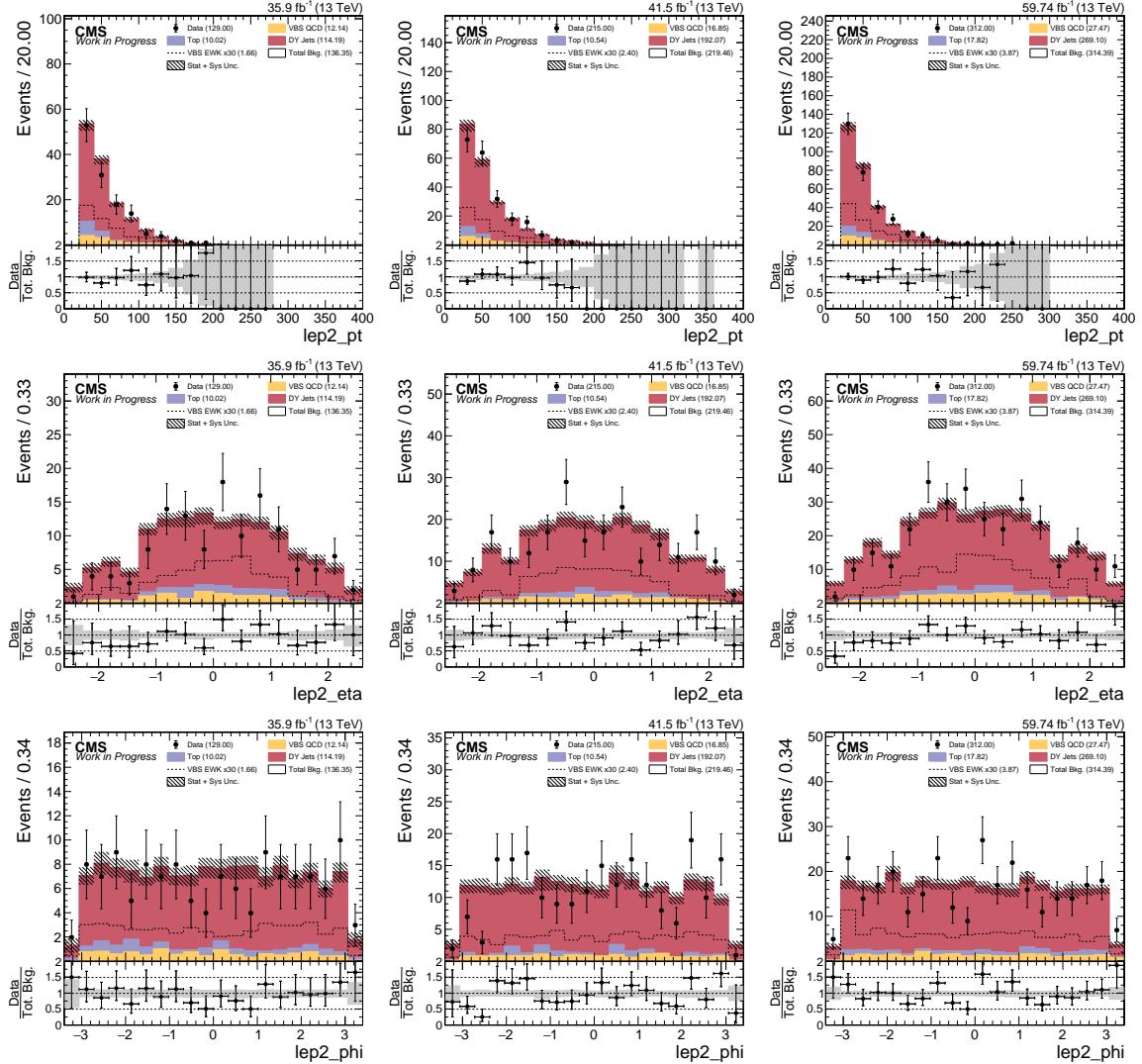


Figure 4.3: DY+Jets Control Region: Trailing electron kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

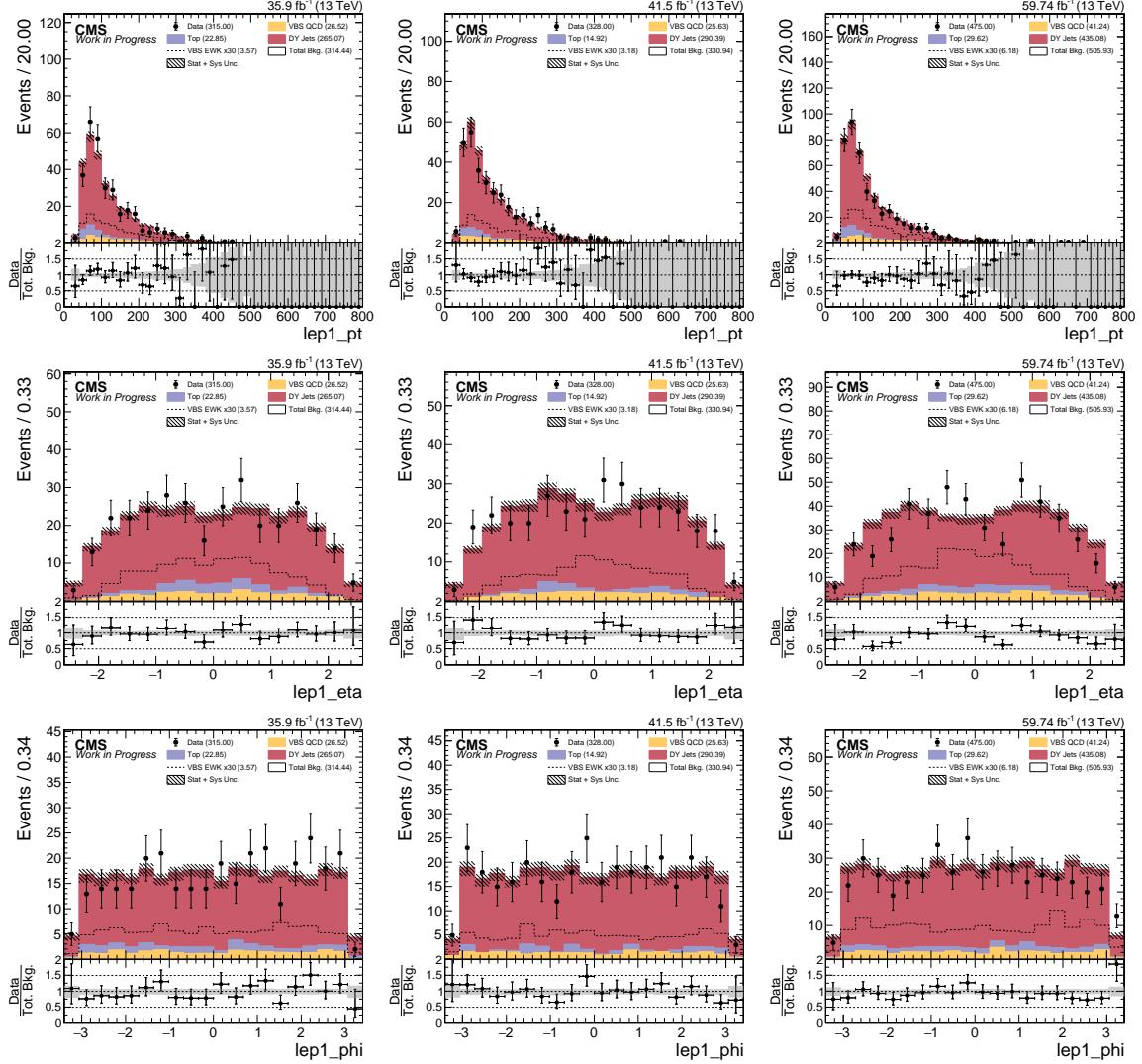


Figure 4.4: DY+Jets Control Region: Leading muon kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

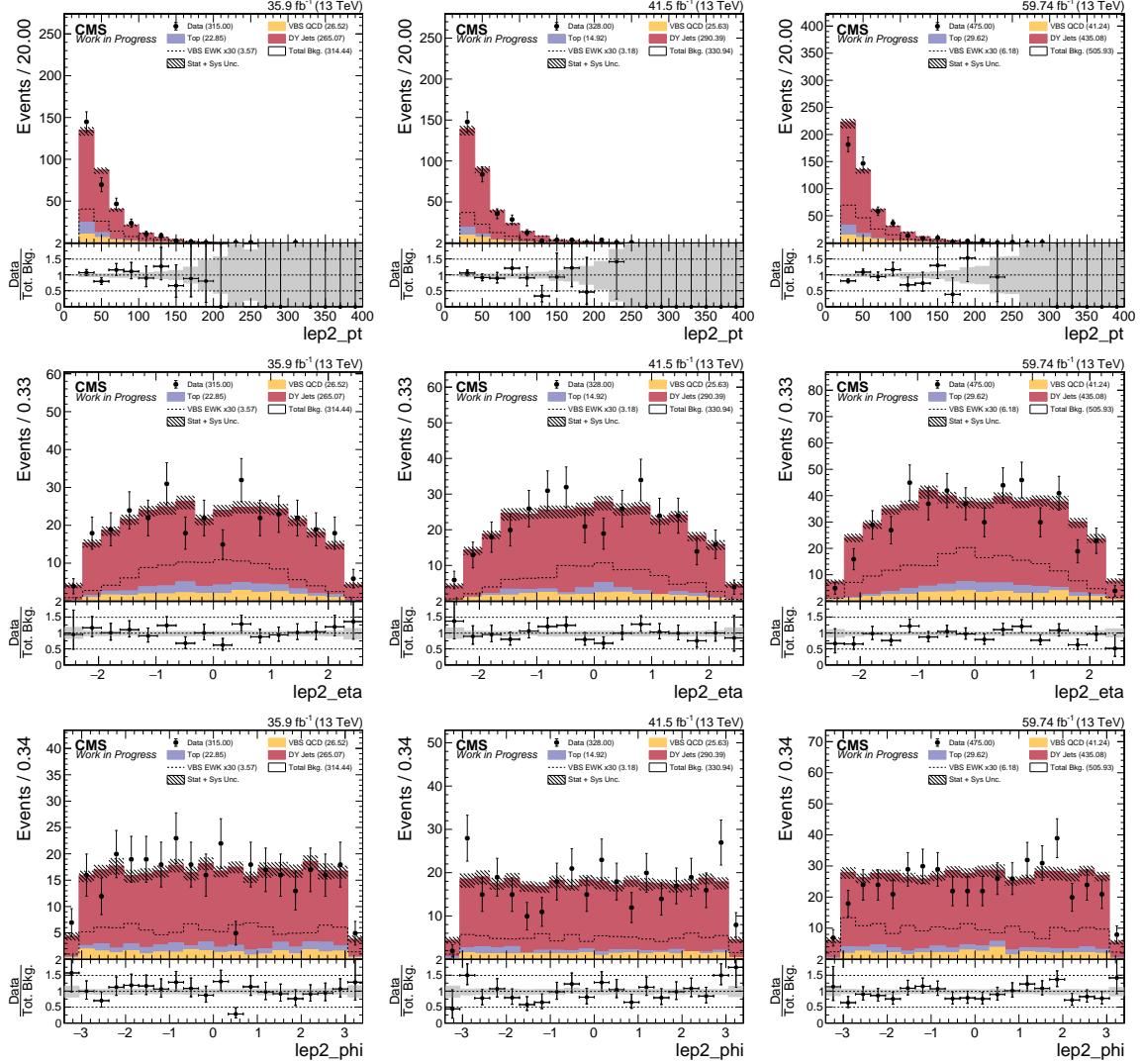


Figure 4.5: DY+Jets Control Region: Trailing muon kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

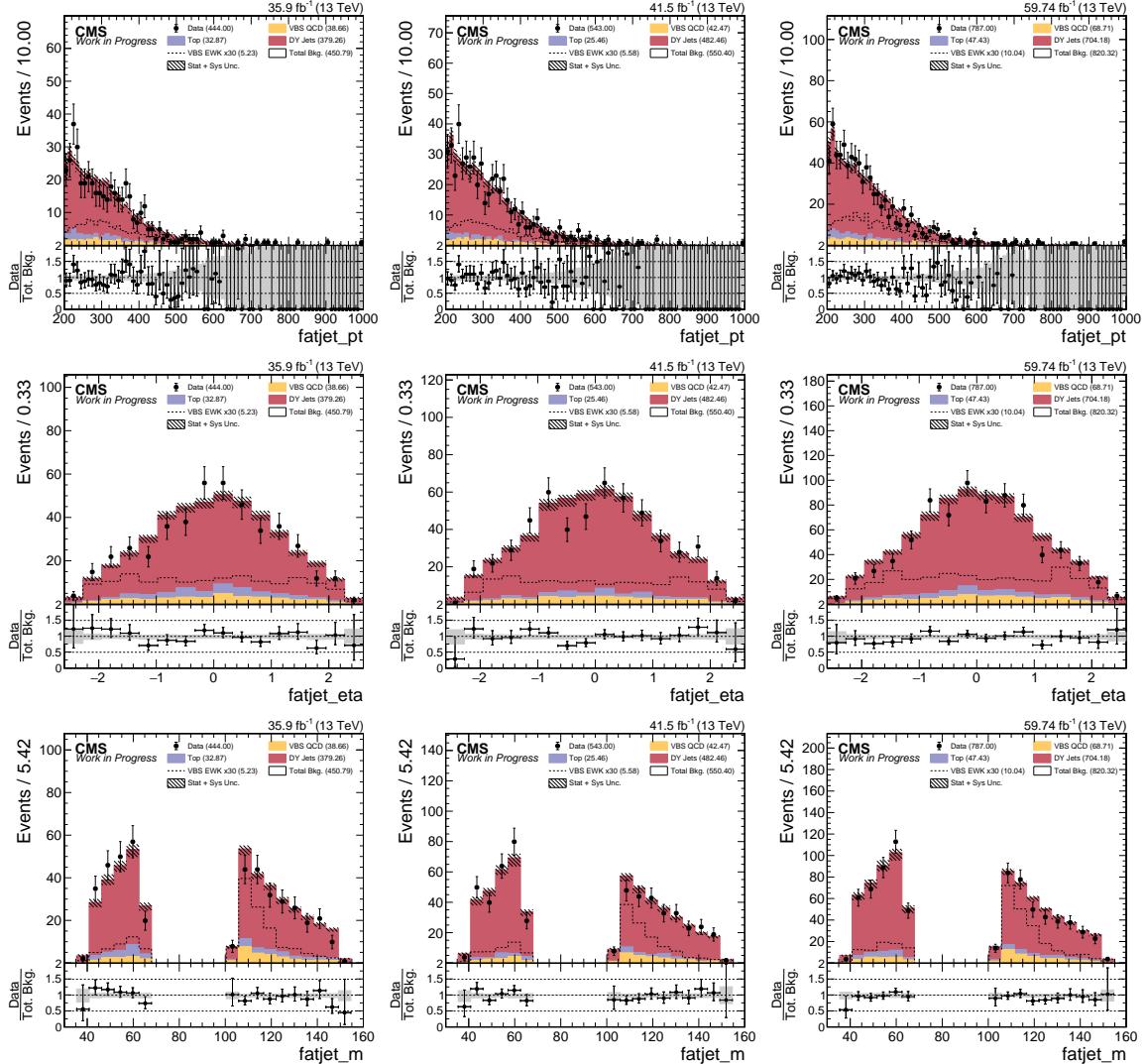


Figure 4.6: DY+Jets Control Region: Hadronic boson kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , mass m .

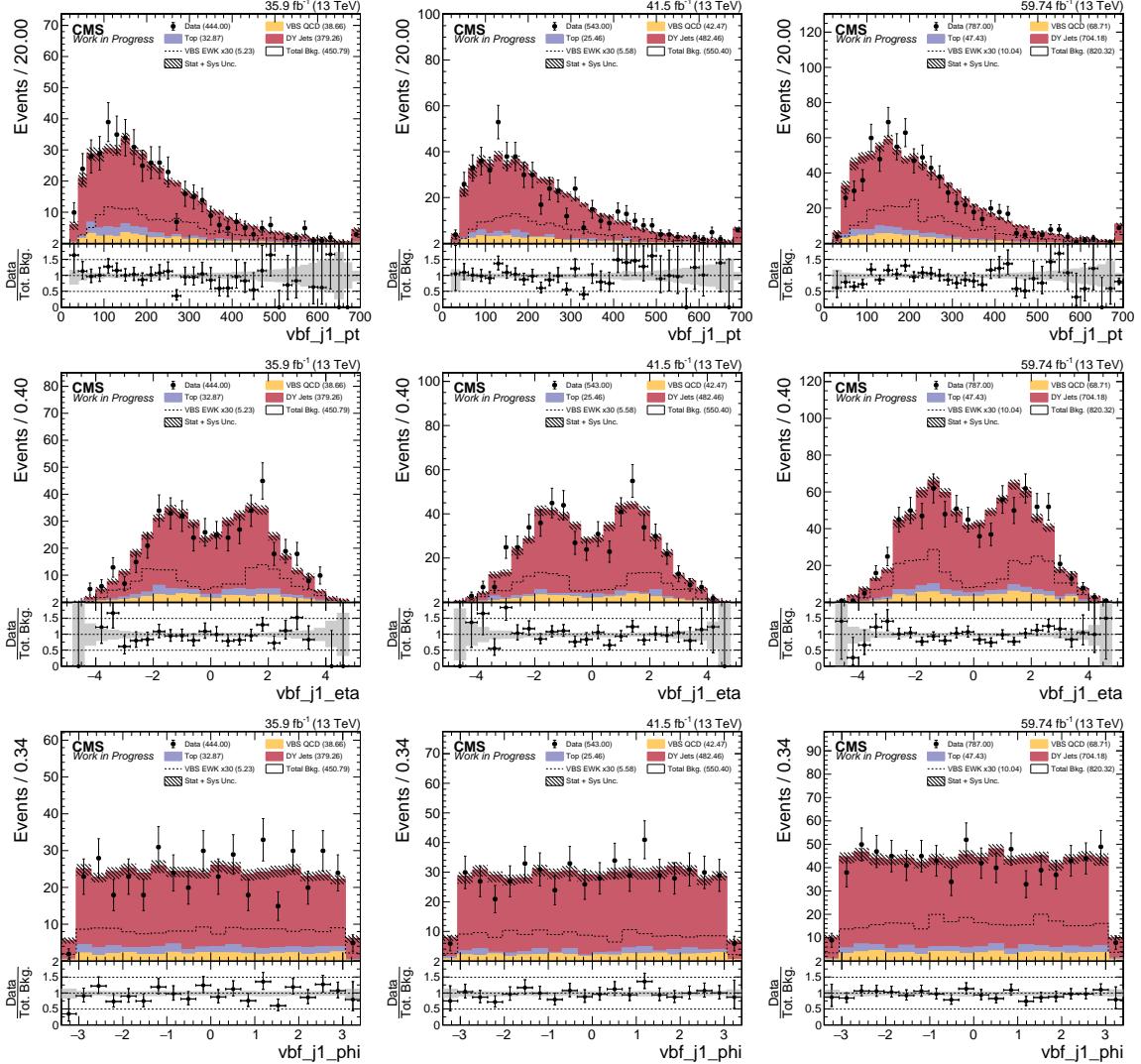


Figure 4.7: DY+Jets Control Region: Leading VBS tagged jet kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

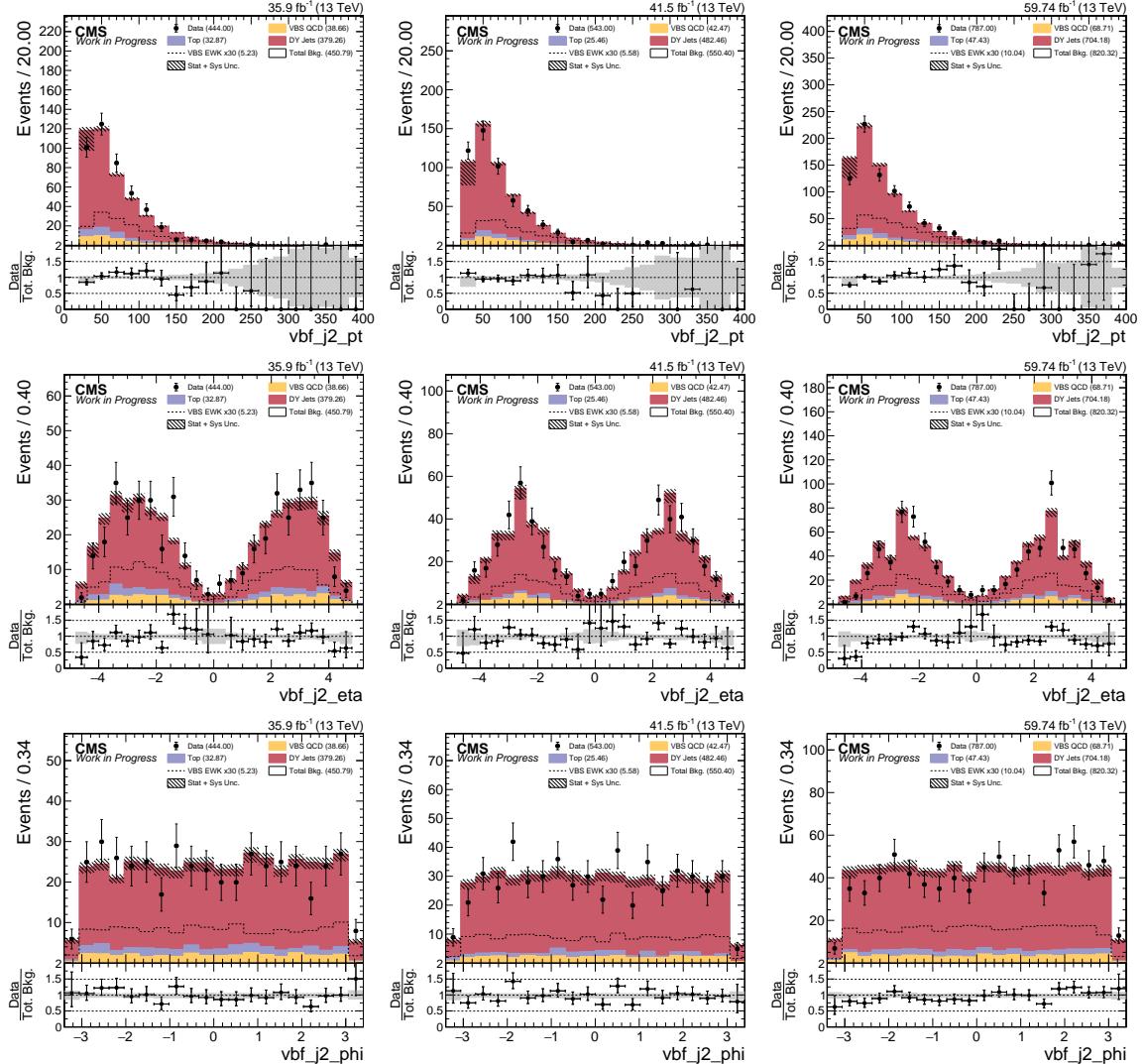


Figure 4.8: DY+Jets Control Region: Trailing VBS tagged jet kinematics in Boosted ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

4.4.2 Resolved ZV DY+Jets Control Region

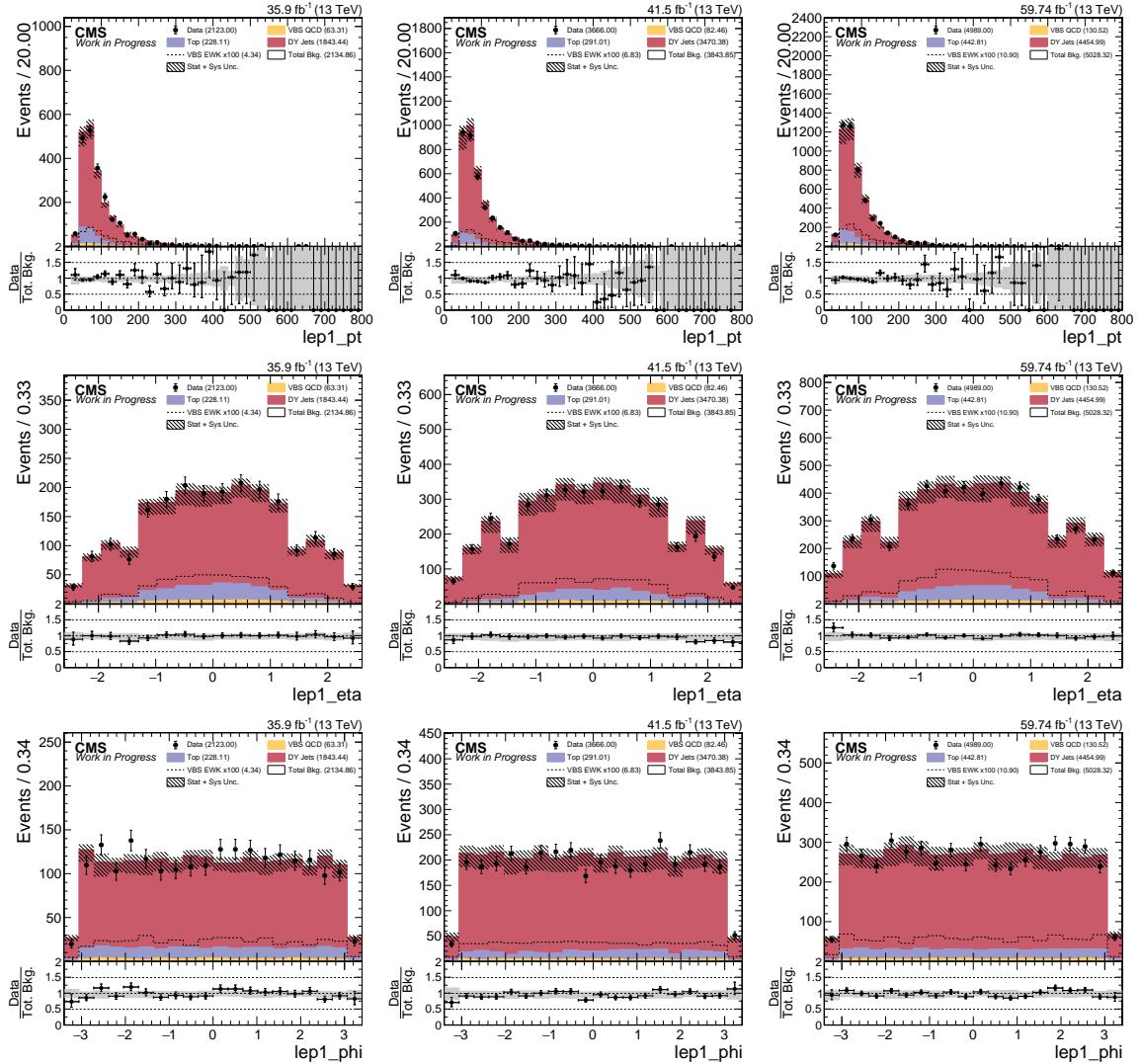


Figure 4.9: DY+Jets Control Region: Leading electron kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

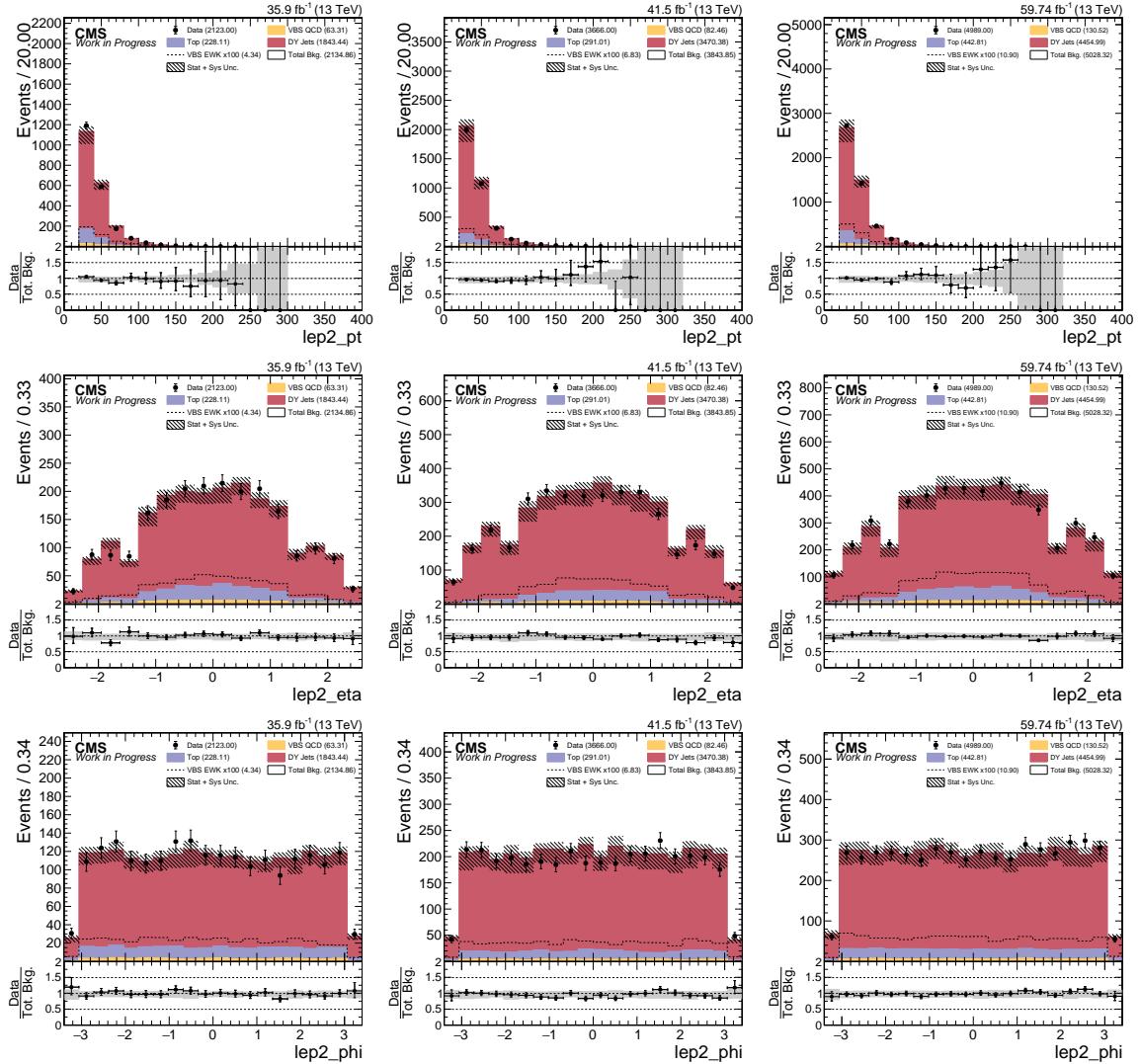


Figure 4.10: DY+Jets Control Region: Trailing electron kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

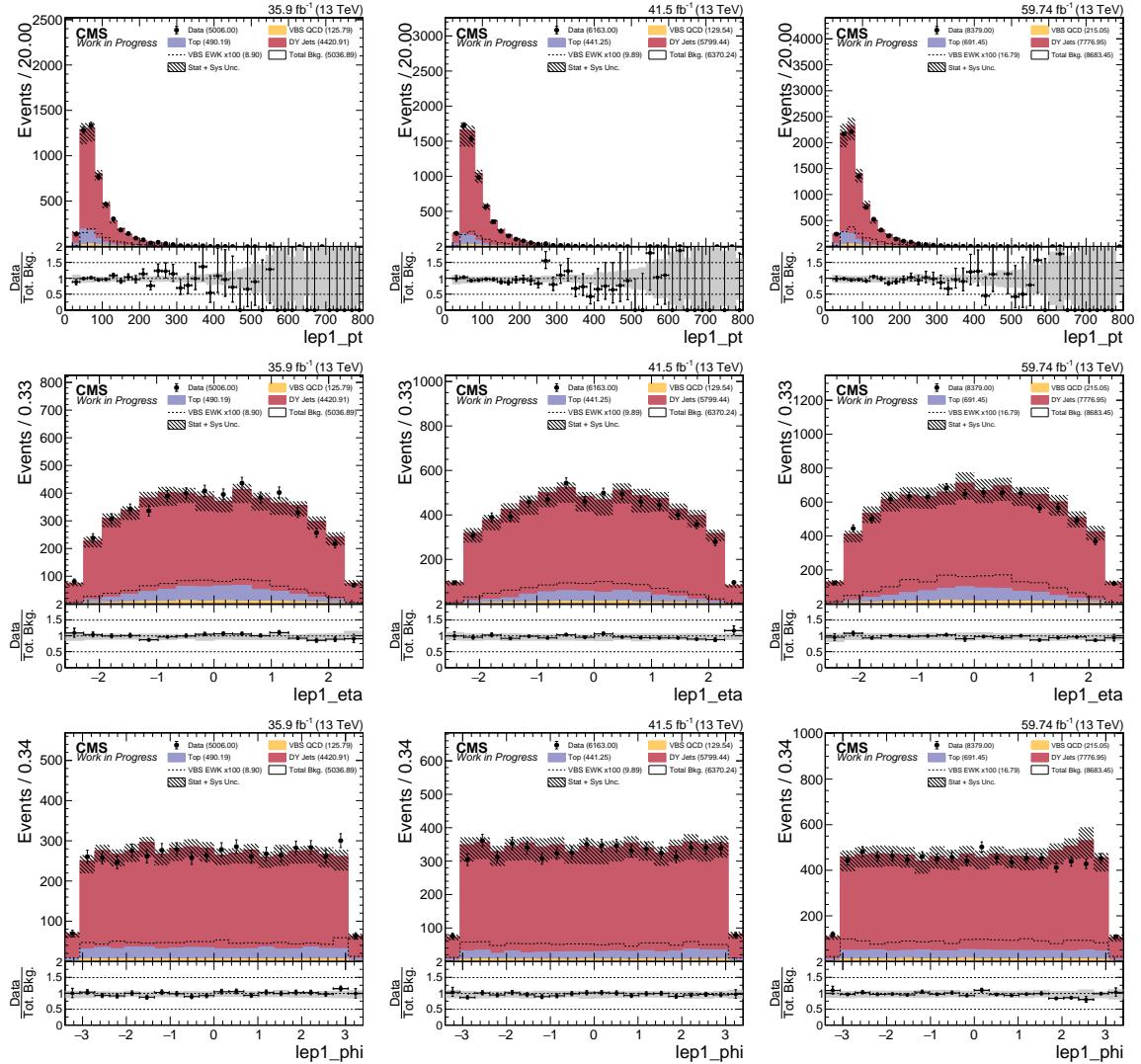


Figure 4.11: DY+Jets Control Region: Leading muon kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

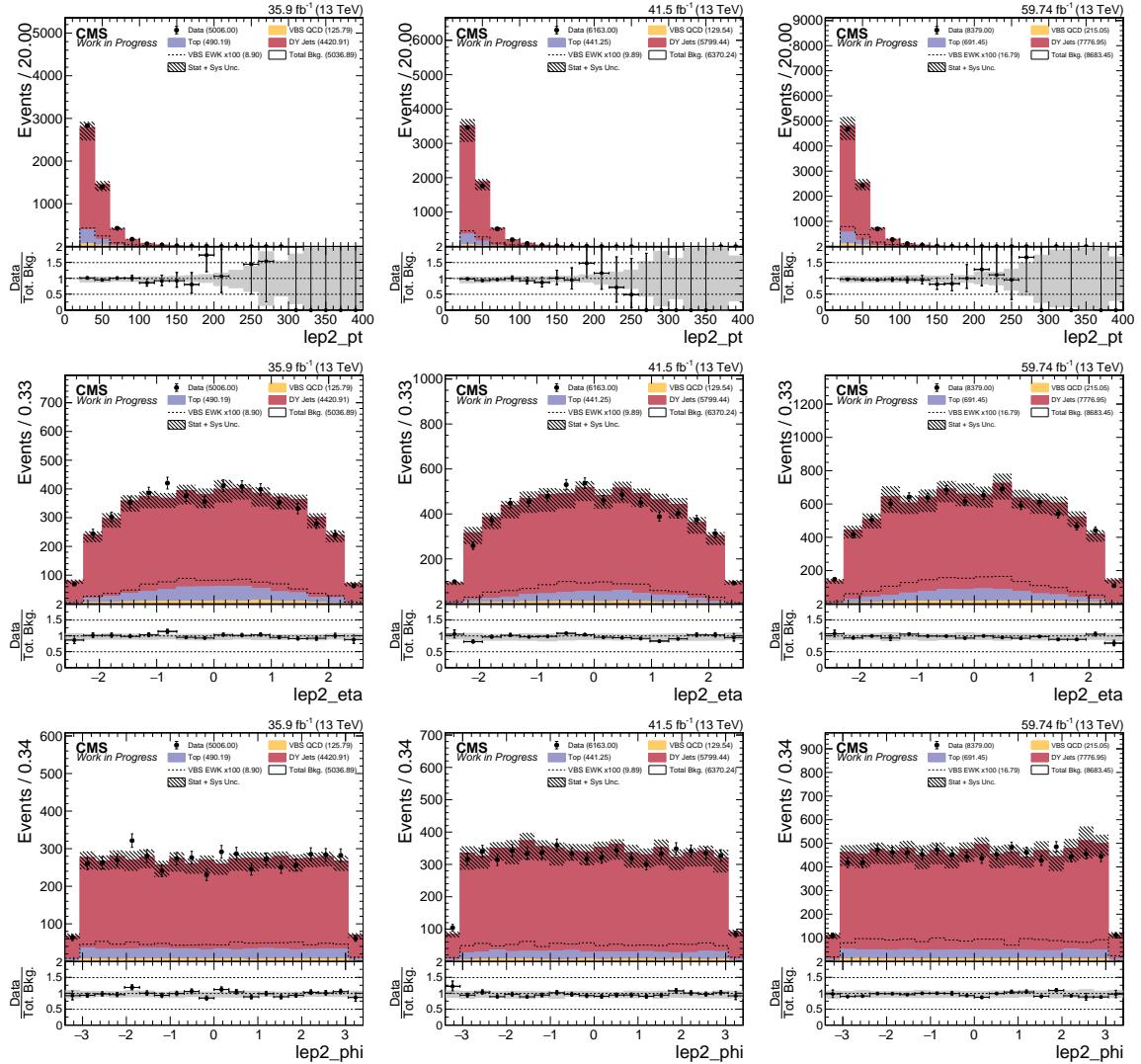


Figure 4.12: DY+Jets Control Region: Trailing muon kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

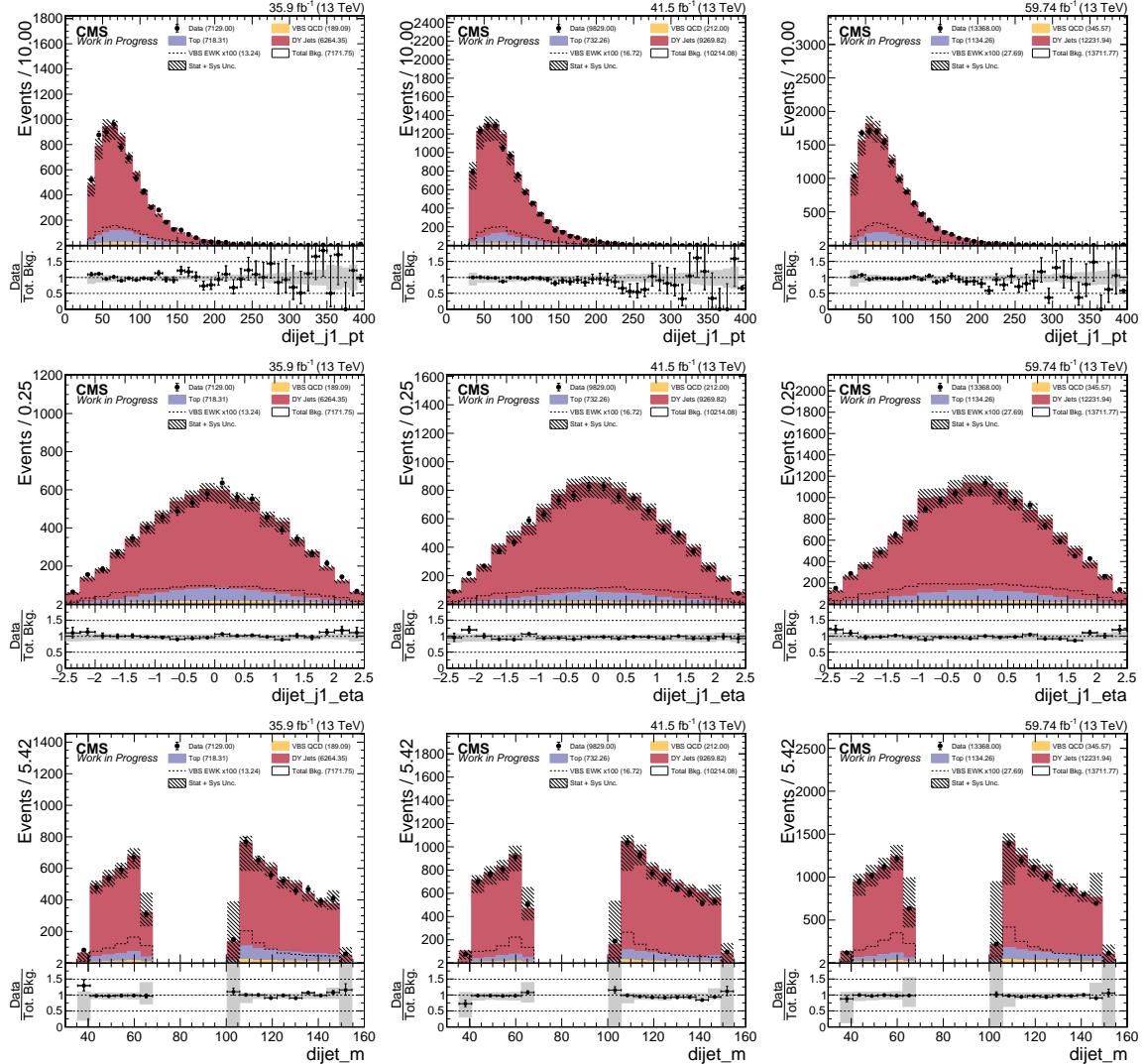


Figure 4.13: DY+Jets Control Region: Hadronic boson leading jet kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , invariant mass m_{jj} .

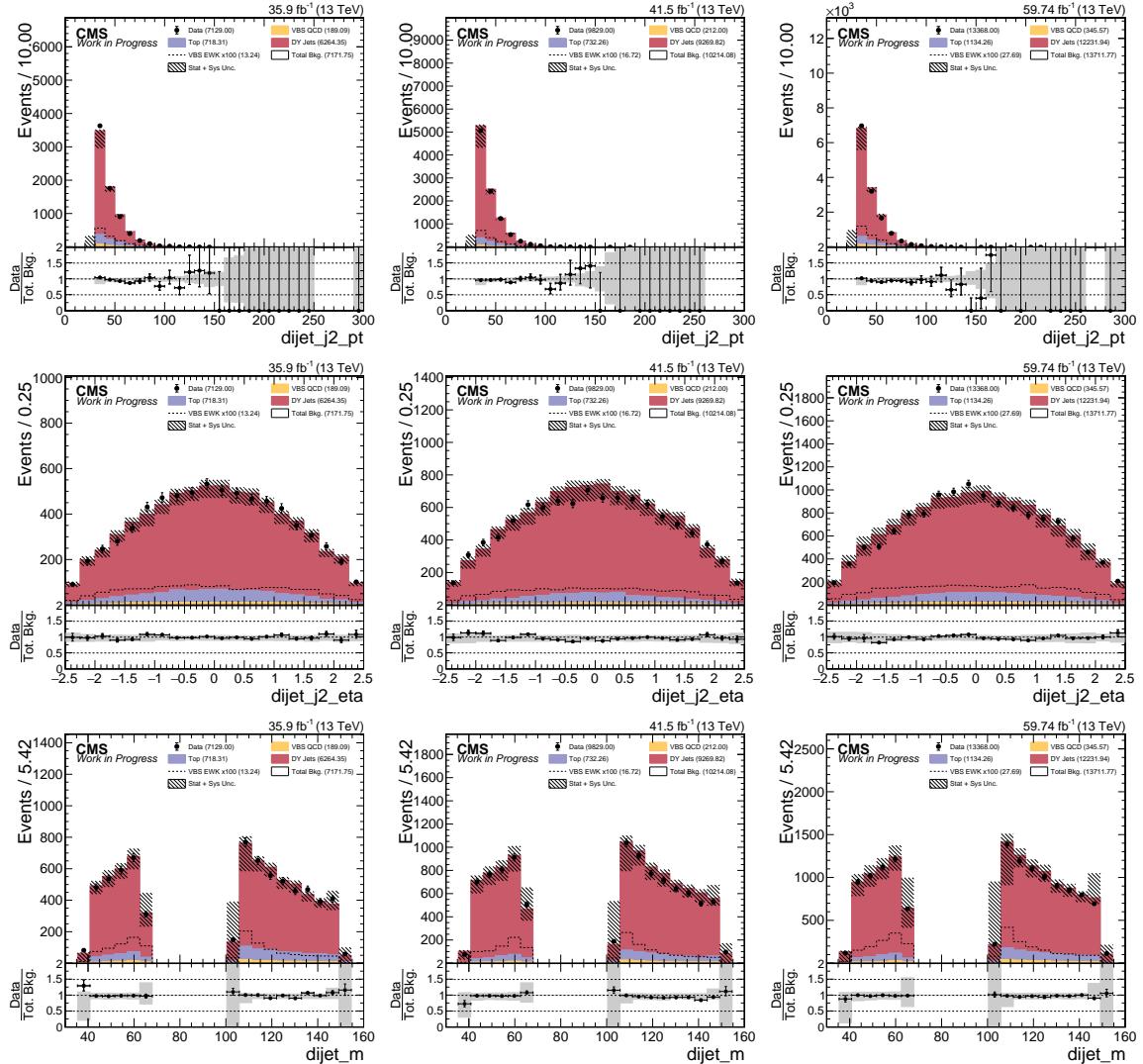


Figure 4.14: DY+Jets Control Region: Hadronic boson trailing jet kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , invariant mass m_{jj} .

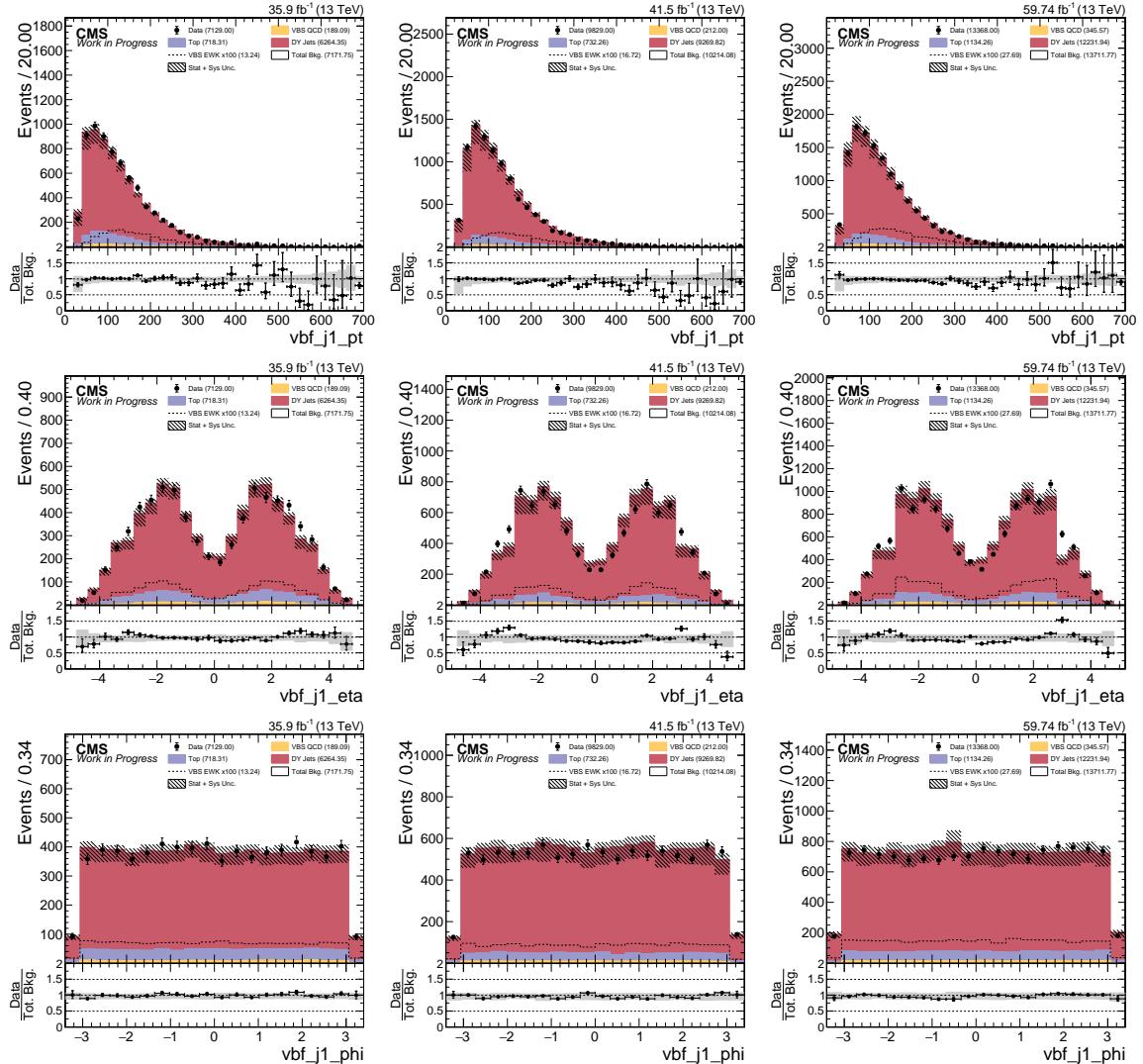


Figure 4.15: DY+Jets Control Region: Leading VBS tagged jet kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

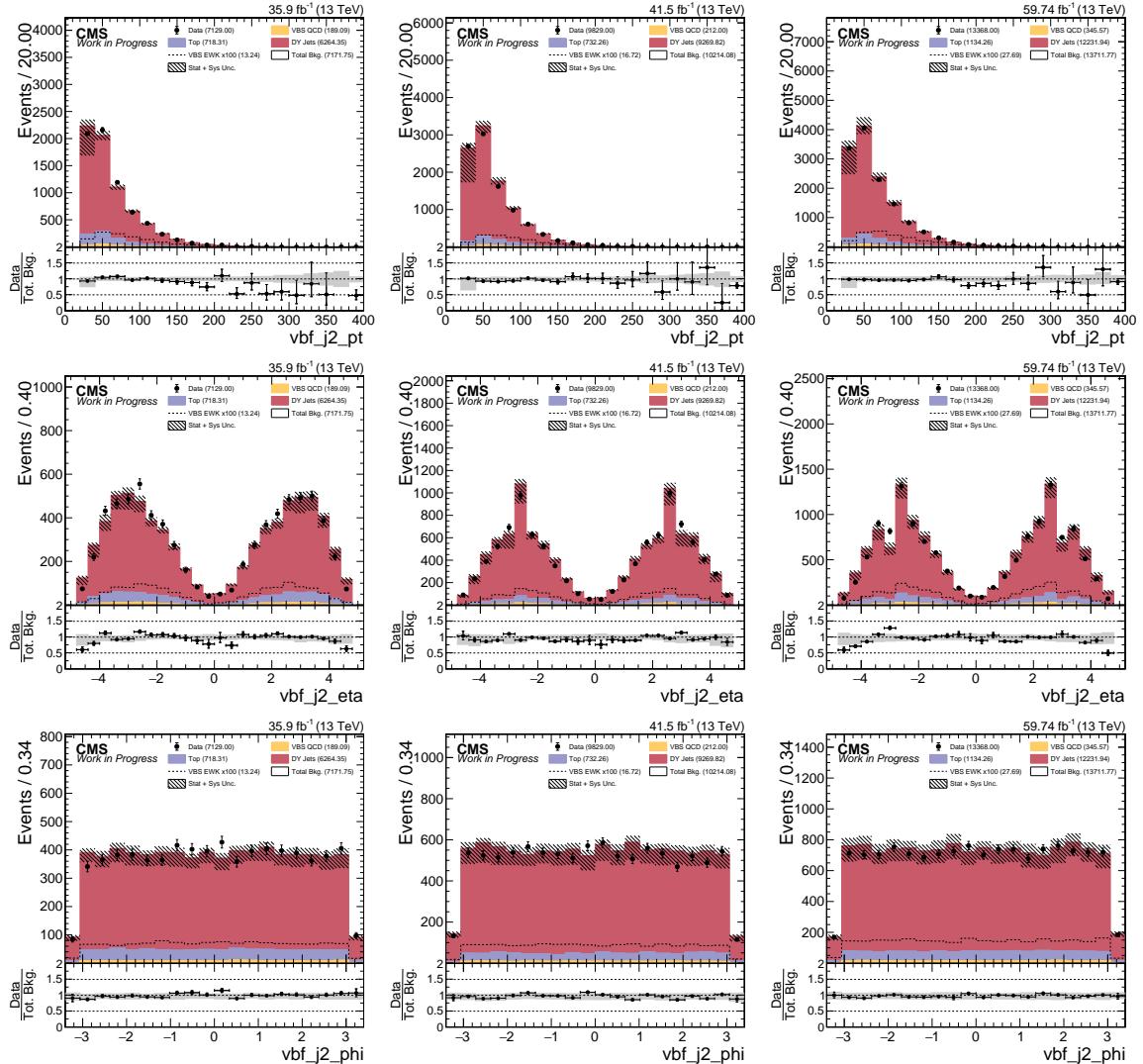


Figure 4.16: DY+Jets Control Region: Trailing VBS tagged jet kinematics in Resolved ZV Channel. Error bars include statistical uncertainty on total background, JES and QCD scale systematic on DY+Jets and VBS_QCD MC. From Left to Right: 2016, 2017, and 2018. From Top to Bottom: p_T , η , ϕ .

872

4.5 Machine Learning Modeling

873 Instead of traditional cut-based analysis, we decided to use MVA a.k.a. ML technique to
 874 build a signal vs background classifier. The main reasoning behind using a MVA technique
 875 is so that we can build a model which can learn our analysis topology from looser selection
 876 regions and still let us keep higher statistics for final measurement.

877

4.5.1 Algorithm: Gradient Boosted Decision Tree

878 Boosted decision tree (BDT) is a MVA algorithm with *boosting* applied sequentially in
 879 series on the MVA algorithm, and then taking weighted average to get the improved result,
 880 in this case the boosted¹ MVA algorithm is *decision tree*. A decision tree is a binary tree,
 881 example shown in Figure 4.17, at each node a decision is made using the input variables
 882 which provides the best separation² at that node, this process is repeated until events at a
 883 node reaches below certain threshold, and then depending on the majority of the events in
 884 the final node, the node is classified as signal or background.

885 There are several boosting algorithm, the one discussed here and used in this analysis is
 886 gradient boost. Gradient boost is an additive model of adding decision trees in series and
 887 with each tree in sequence training on the adjusted dataset to minimize the *loss-function*
 888 with respect to the previous model. This improves the model sequentially and it is faster
 889 since only a subset of training dataset is used in subsequent decision trees. Toolkit for
 890 Multivariate Analysis (TMVA) part of ROOT package used in this analysis for implement-

¹Not to be confused analysis category Boosted ZV, they are unrelated.

²defined as, $\frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$, where \hat{y}_S and \hat{y}_B are probability density functions of signal and background.

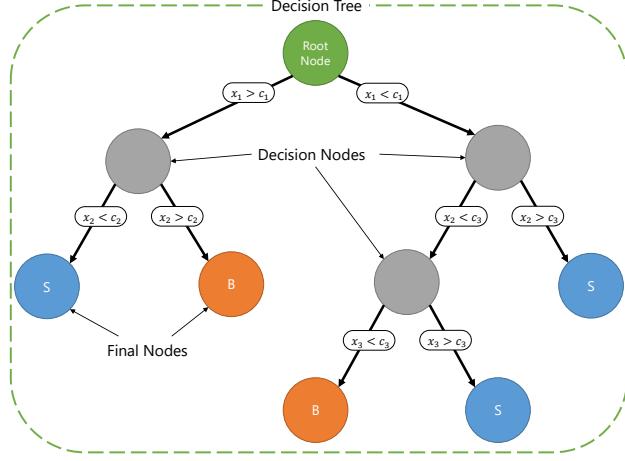


Figure 4.17: Decision Tree.

891 ing gradient boosted BDT has implementation of binomial log-likelihood loss function for
 892 classification [51].

893 4.5.2 Training and Results

894 Two models were trained for boosted and resolved topology and the training was done
 895 using combined MC from all years 2016, 2017, 2018 to benefit from larger statistics see
 896 Table 4.6, signal MC “VBS_EWK” was trained against background “DY + Jets LO” since
 897 that is dominant background in our analysis. Each models BDT hyper-parameters were
 898 tuned to prevent under and over-fitting, this is done by comparing the signal efficiency
 899 training and testing datasets at three different background rates (1%, 10% and 30%), if the
 900 difference is not more than 5%, then the hyper-parameters are acceptable.

901 Input variables used in each model training were pruned i.e. input variables which does
 902 not impact much on the performance of the model in terms of area under curve (AUC) of
 903 the Receiver Operating Characteristic (ROC) were dropped from the final training models.
 904 Distribution of final input variables used in training are shown in Figure 4.18 and Figure 4.19.

Table 4.6: Combined dataset Training and Testing Statistics in Signal Region

		Number of Events		
Channel	Dataset	Training	Testing	Total
Boosted	Signal	7404	7405	14809
Boosted	Background	46991	46991	93982
Resolved	Signal	23425	23425	46850
Resolved	Background	209368	209368	418736

905 After training TMVA BDT evaluates input variables and ranks them in terms of impor-
 906 tance and separation they provide in classification is listed in Table 4.7, and the correlation
 907 matrix of variable is show in Figure 4.21.

Table 4.7: Training Input Variable Ranking

Channel	Variable	Variable Name	Importance	Separation
Boosted	M_{JJ}^{VBS}	vbf_m	0.2496	0.1348
	Zeppenfeld Z_V^*	zeppLep_deta	0.2396	0.1116
	QGL j_1^{VBS}	vbf2_AK4_qgid	0.1889	0.02413
	QGL j_2^{VBS}	vbf1_AK4_qgid	0.1780	0.02330
	M_{VV}	dibos_m	0.1439	0.005308
Resolved	Zeppenfeld Z_V^*	zeppLep_deta	0.1955	0.1219
	M_{JJ}^{VBS}	vbf_m	0.1822	0.07998
	HT^*	ht_resolved	0.1693	0.04201
	QGL j_1^{VBS}	vbf2_AK4_qgid	0.1403	0.02159
	QGL j_2^{VBS}	vbf1_AK4_qgid	0.1341	0.03235
	M_{VV}	dibos_m	0.09098	0.01112
	η_{lep2}	lep2_eta	0.08760	0.01755

908 The under and over-fitting of trained model is checked by Kolmogorov—Smirnov (K-S)
 909 test and ROC curves comparison between training and testing datasets. If the tests are not
 910 acceptable, then the training is redone with adjusted parameters. Table 4.8 lists the final
 911 training configuration in the TMVA framework for for boosted and resolved models. The
 912 Figure 4.22 show MVA score and ROC curves of the BDT models. Boosted trained model
 913 MVA has 78 % and resolved has 79 % of AUC.

Table 4.8: TMVA configuration of trained models

Boosted and Resolved ZV	
BoostType	Grad
NTrees	800
MaxDepth	3
MinNodeSize	3%
nCuts	100
Shrinkage	0.010
UseBaggedBoost	True
BaggedSampleFrac	0.6

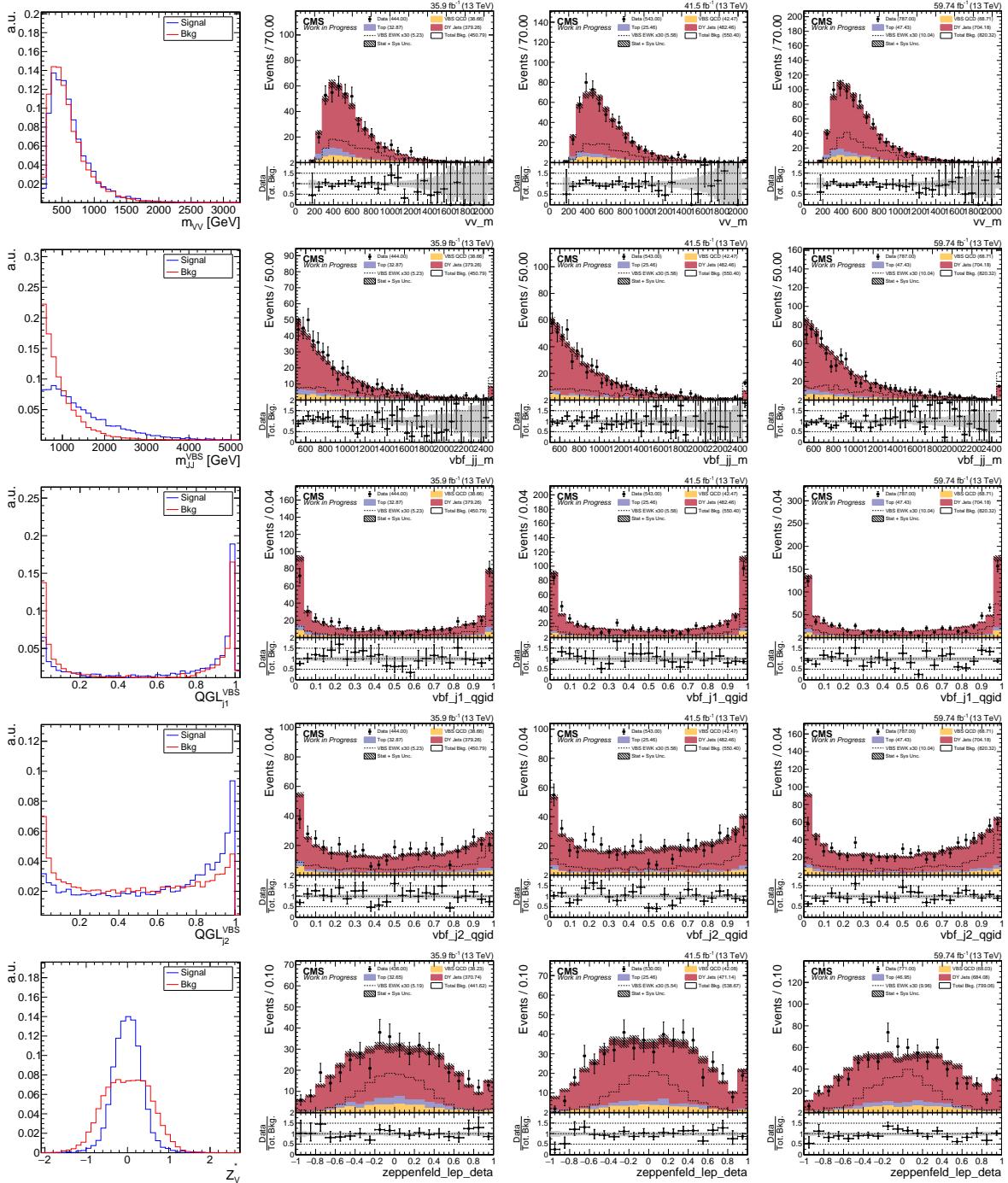


Figure 4.18: Inputs Variables for training Boosted ZV BDT Classifier. Left to Right: Shape comparison of signal and background (combined dataset), control region kinematic distribution in 2016, 2017 and 2018 dataset. Top to Bottom: Diboson invariant mass, VBS tagged jets invariant mass, Quark Gluon Likelihood (QGL) of leading VBS tagged jet, QGL of trailing VBS tagged jet, Zeppenfeld variable of leptonic boson.

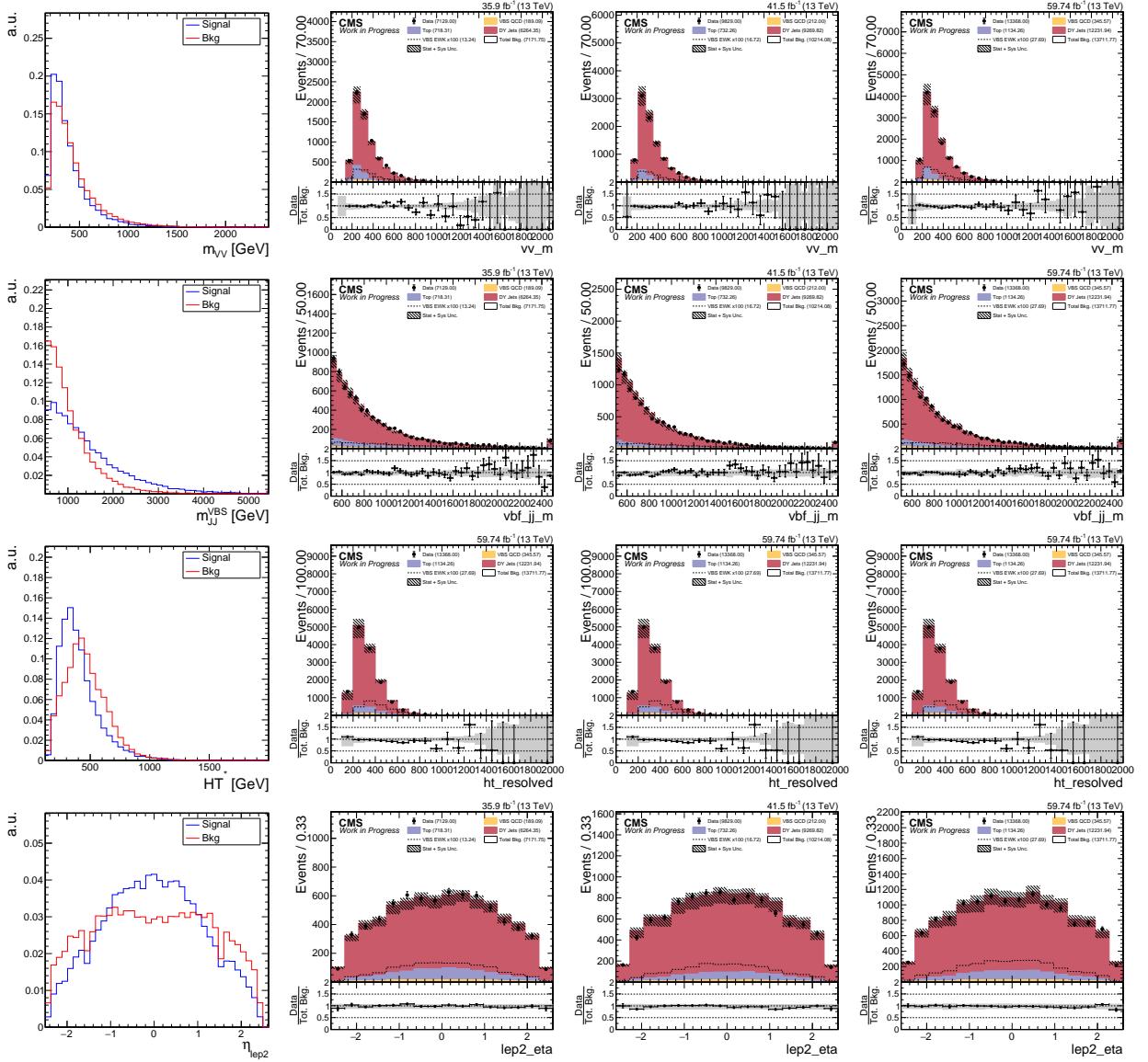


Figure 4.19: Inputs Variables for training Resolved ZV BDT Classifier. Left to Right: Shape comparison of signal and background (combined dataset), control region kinematic distribution in 2016, 2017 and 2018 dataset. Top to Bottom: Diboson invariant mass, VBS tagged jets invariant mass, HT^* (p_T sum of jets), trailing lepton η .

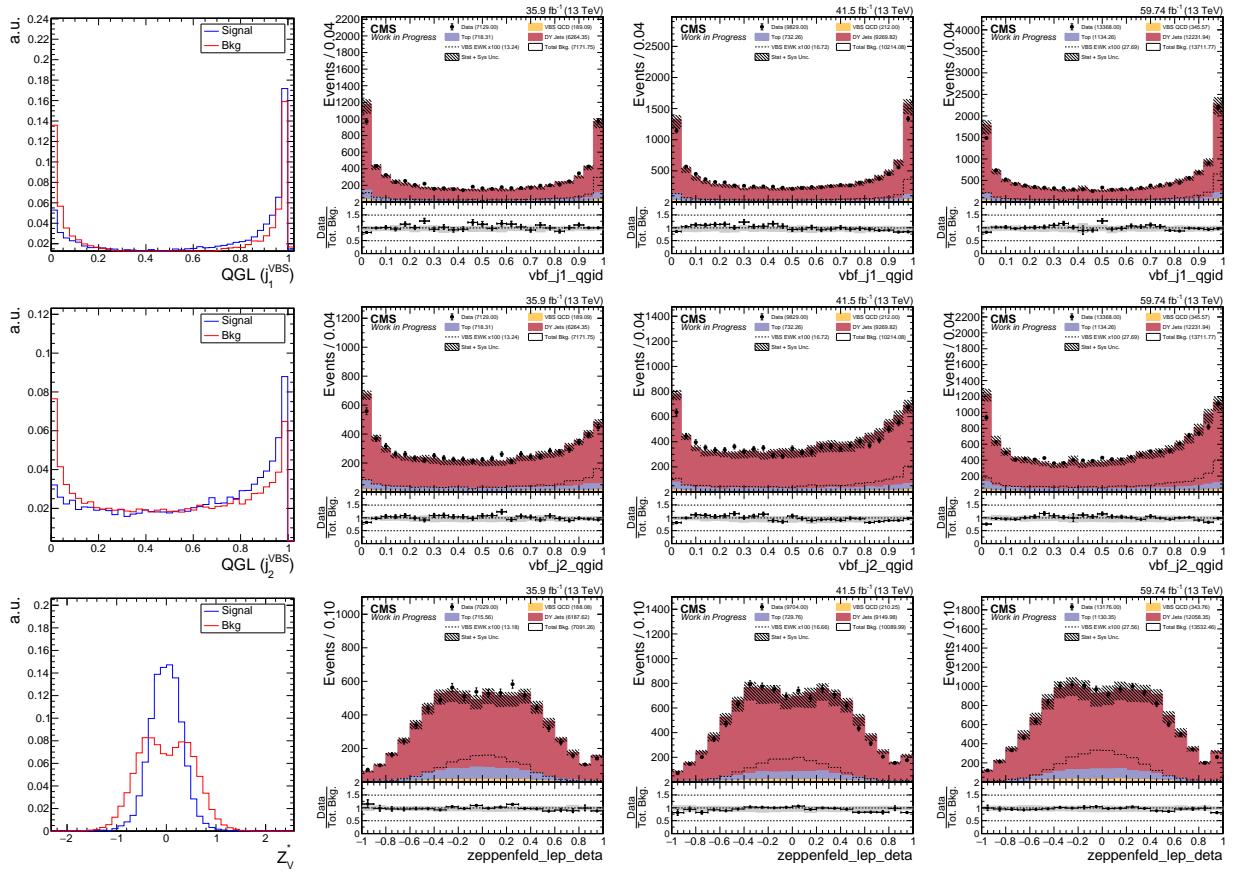


Figure 4.20: Inputs Variables for training Resolved ZV BDT Classifier. Left to Right: Shape comparison of signal and background (combined dataset), control region kinematic distribution in 2016, 2017 and 2018 dataset. Top to Bottom:QGL of leading VBS tagged jet, QGL of trailing VBS tagged jet, Zeppenfeld variable of leptonic boson.

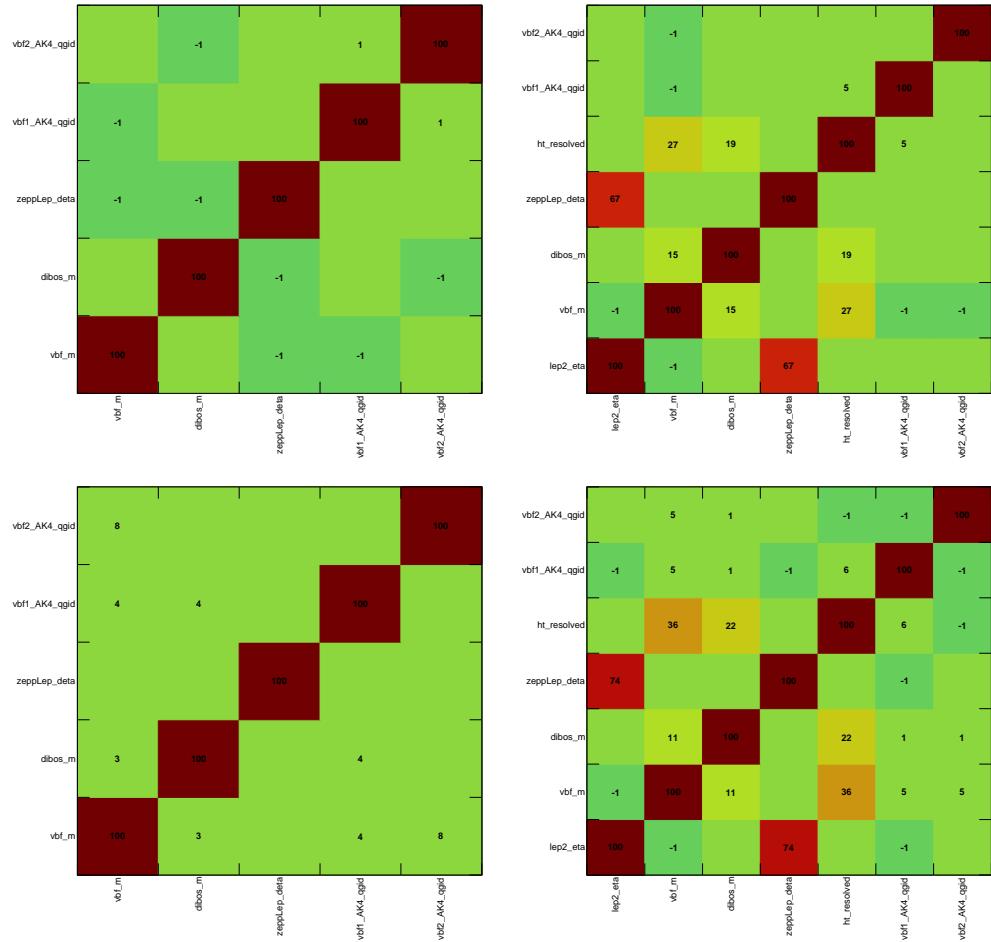


Figure 4.21: Correlation Matrix for Signal and Background. From Left to Right: Boosted, Resolved. From Top to Bottom: Signal, Background

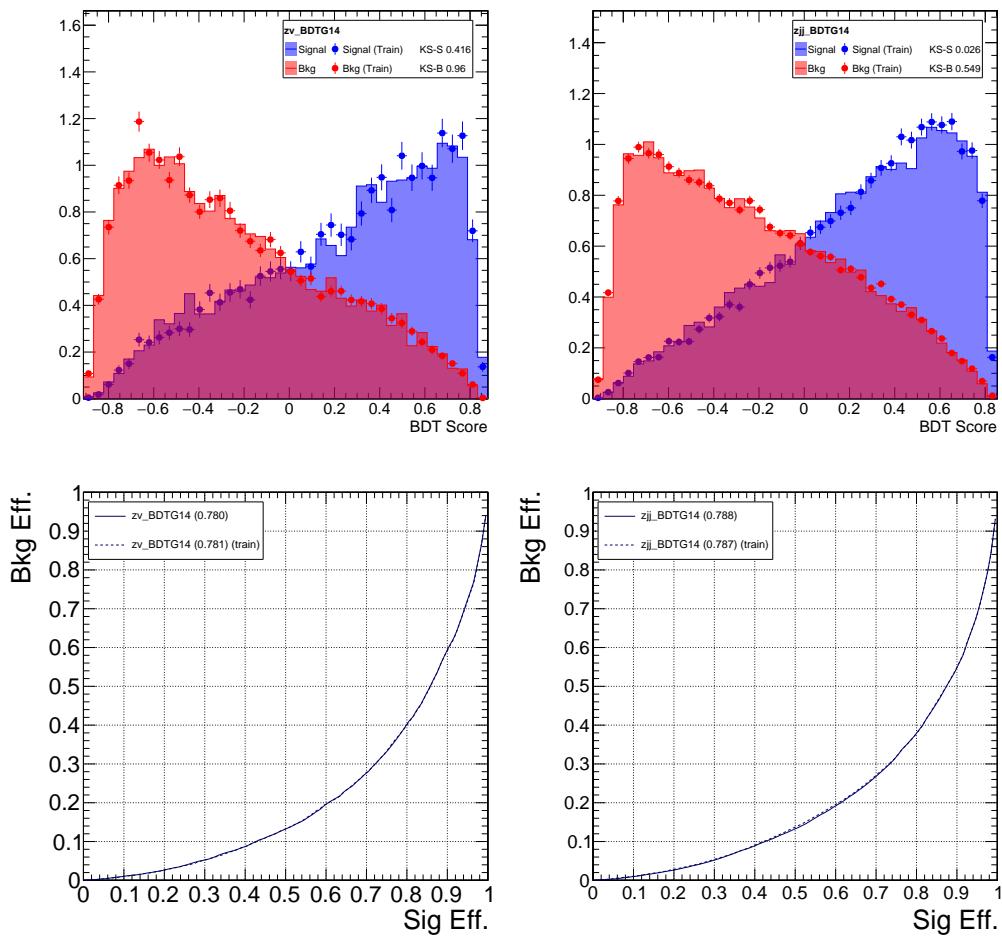


Figure 4.22: From Left to Right: Boosted, Resolved. Top to Bottom: MVA Score of BDT models, ROC Curves.

914

4.5.3 MVA Score Inference

915 Using the training models for boosted and resolved ZV, the MVA score is inferred for
 916 the complete dataset of data and MC. Then the binning of MVA score is derived such that
 917 the signal yield is almost equal in all the bins in signal region. MVA score distribution for
 918 control region with data is shown in Figure 4.23 and 4.24, and blinded MVA score for signal
 919 region is shown in Figure 4.25 and Figure 4.24.

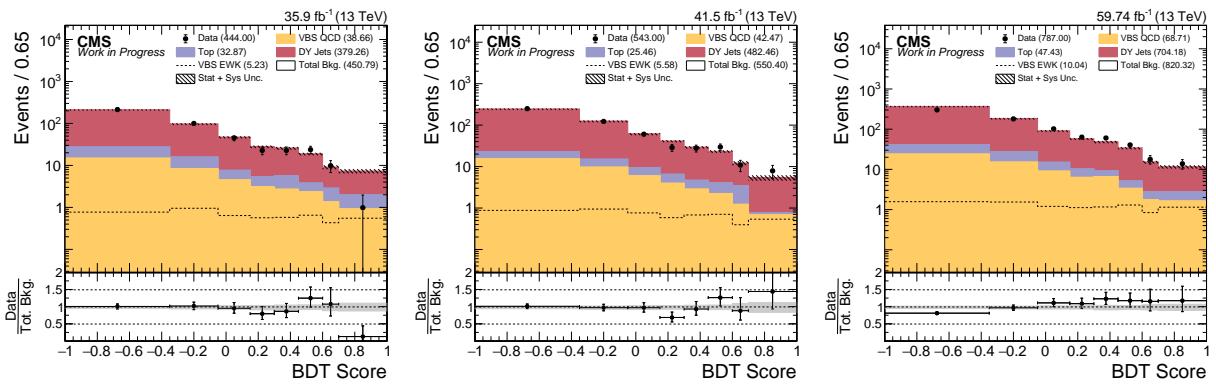


Figure 4.23: MVA Score in Control Region for Boosted ZV Channel.

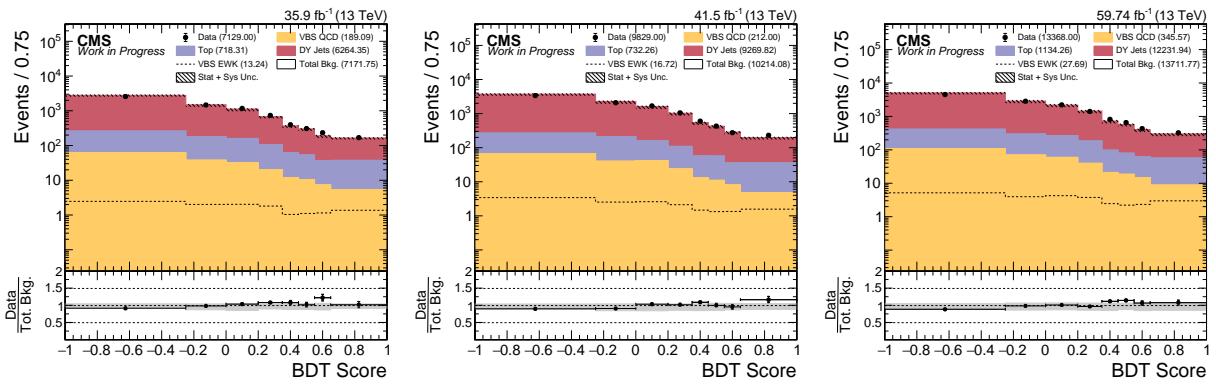


Figure 4.24: MVA Score in Control Region for Resolved ZV Channel.

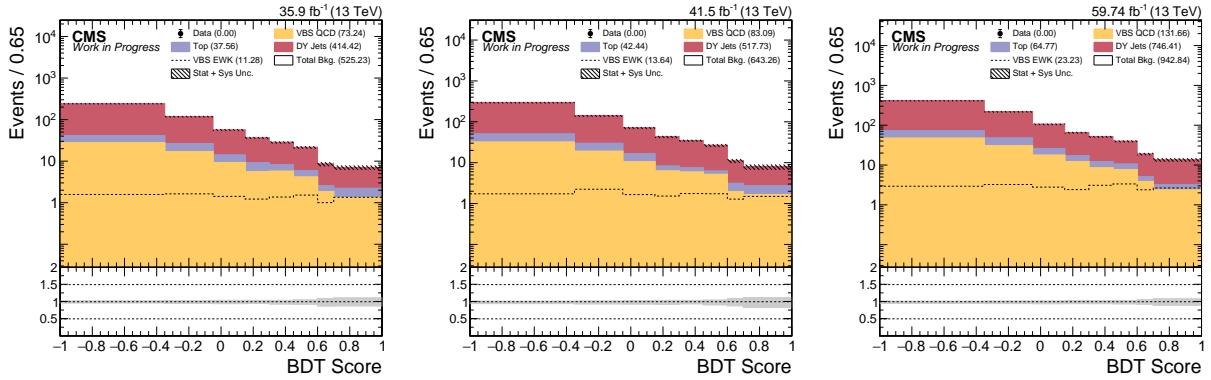


Figure 4.25: MVA Score in Signal Region for Boosted ZV Channel.

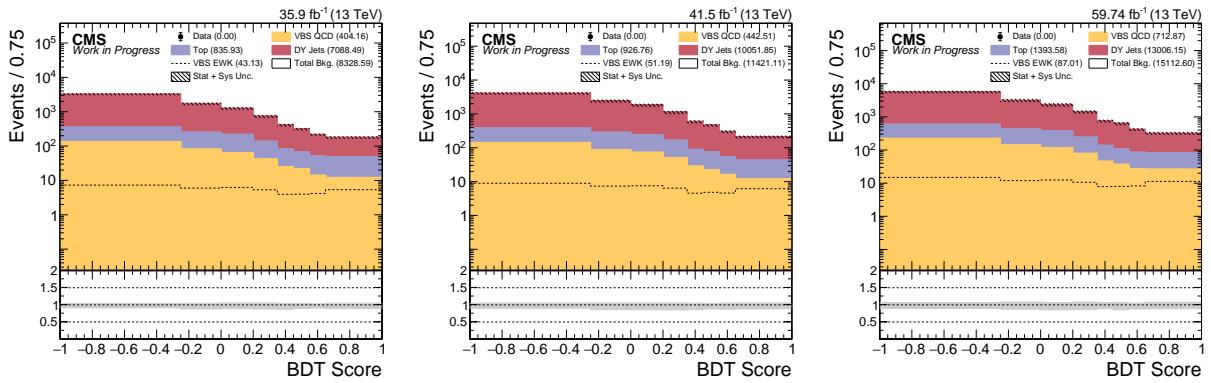


Figure 4.26: MVA Score in Signal Region for Resolved ZV Channel.

920

4.6 Measurement

921 To measure our signal, the MVA score in signal region is used as the observable to
922 calculate expected significance and study impact of systematics using impact plots.

923 **CombineLimit** is command line tool which takes input configuration as plain text file
924 and can perform standard high energy physics statistical analysis. **CombineLimit** is built on
925 **RooFit** and **RooStats** packages which are also part of ROOT package [52, 53].

926

4.6.1 Statistical and Systematic Uncertainties

927 MC statistical uncertainties for the bins with more than 10 events are taken into account
 928 using “Barlow—Beeston lite” method, and using Poisson probability density function (p.d.f.)
 929 otherwise [54].

930 Systematic uncertainties for correction of multiplicative type such as integrated luminos-
 931 ity, efficiency, etc., are treated with most common log-normal distribution model, i.e. mul-
 932 tiplying with factor κ is equivalent to $+1\sigma$ variation, and multiplying with $1/\kappa$ to get -1σ
 933 variation. For relative uncertainty $\Delta x/x$, the κ can be set as $1 + \Delta x/x$. For example, 2.5%
 934 uncertainty with log-normal model on integrated luminosity, κ will be 1.025. Uncertainties
 935 set this way are referred to as “flat” systematic uncertainties.

936 For systematic uncertainties such as JES, observable used in the measurement is also
 937 calculated with JES $+1\sigma$ and -1σ scale factor separately, and the variation up and down
 938 value are directly taken from the histogram bin, these type of uncertainties are referred to
 939 as “shape” systematic uncertainties.

940 Following are the systematic uncertainties used in this analysis,

- 941 • **Integrated Luminosity:** 2.5% uncertainty on all 2016, 2017, and 2018 year datasets.
- 942 • **L1 Prefire Correction:** Shape systematics for 2016 and 2017 year.
- 943 • **Lepton Reconstruction and Identification Efficiency:** 1% uncertainty for elec-
 944 tron and muons.
- 945 • **Lepton momentum scale:** Shape systematics for electron and muon momentum
 946 scale, shapes are obtained by varying p_T 3% up and down in selection.
- 947 • **JES :** Shape systematics obtained by varying jet p_T with JES up and down, for all the
 948 years dataset. The uncertainties are treated separately depending on the source [40].

- **PU Reweighting:** Shape systematics obtained by varying normalization up and down for all the years dataset.
- **Renormalization and factorization scale:** Renormalization and factorization scale uncertainties are calculated for VBS_EWK and VBS_QCD MC datasets, as prescribed by CMS the renormalization and factorization scale factor are varied up and down by the factor of 2, and the envelope of the all the variation is taken as 1σ shape systematics.
- **PDF systematics:** PDF uncertainties are obtained as uniform shape systematic by reweighting MC events cross-section with PDF set.

4.6.2 Significance

Signal significance (Z) is the quantity used to report if there is excess of signal events in presence of background events (b) by testing alternate ($n = \mu s + b$) hypothesis against null ($n = b$) hypothesis, where μ is the signal strength.

Signal significance (Z) in the asymptotic limit is calculated using profile likelihood ratio as the test statistics [55],

$$Z = \sqrt{q_0} \quad (4.1)$$

$$q_\mu = -2 \log \frac{L(data|\mu, \hat{\theta}_\mu)}{L(data|\hat{\mu}, \hat{\theta})}, \quad 0 \leq \mu \leq \hat{\mu} \quad (4.2)$$

where $\hat{\mu}$ is the maximum likelihood estimate of μ , and $\hat{\theta}_\mu$ are the nuisance parameters.

As per LHC statistical recommendation [56], likelihood $L(data|\mu, \hat{\theta}_\mu)$ is constructed as,

$$L(data|\mu, \hat{\theta}_\mu) = \text{Poisson}(data|\mu s(\hat{\theta}) + b(\hat{\theta})) \cdot \pi(\hat{\theta}) \quad (4.3)$$

where $\pi(\hat{\theta})$ is the p.d.f. for nuisance parameter.

For calculating expected significance for $\mu = 1$, the observed data is not used, instead n is set equal to total MC events. The expected significance calculated for each category per year dataset, for combined dataset per category and for combined dataset and category is listed in Table 4.9.

Table 4.9: Expected Signal Significance

Channel	2016	2017	2018	Combined
Boosted	0.64	0.66	0.95	1.33
Resolved	0.40	0.37	0.56	0.78
Combined				1.52

971

4.6.3 Impact Plots

972 Impact plots are way to visualize which nuisance parameter θ have largest effect on
 973 parameter of interest μ and measure correlation from the direction of $+1\sigma$ (correlated) and
 974 -1σ (anti-correlated). Figure 4.27, 4.28 shows impacts for top 60 nuisance parameter, and
 975 Figure B.1, B.2 in the Appendix B.1 for the rest of nuisance parameter.

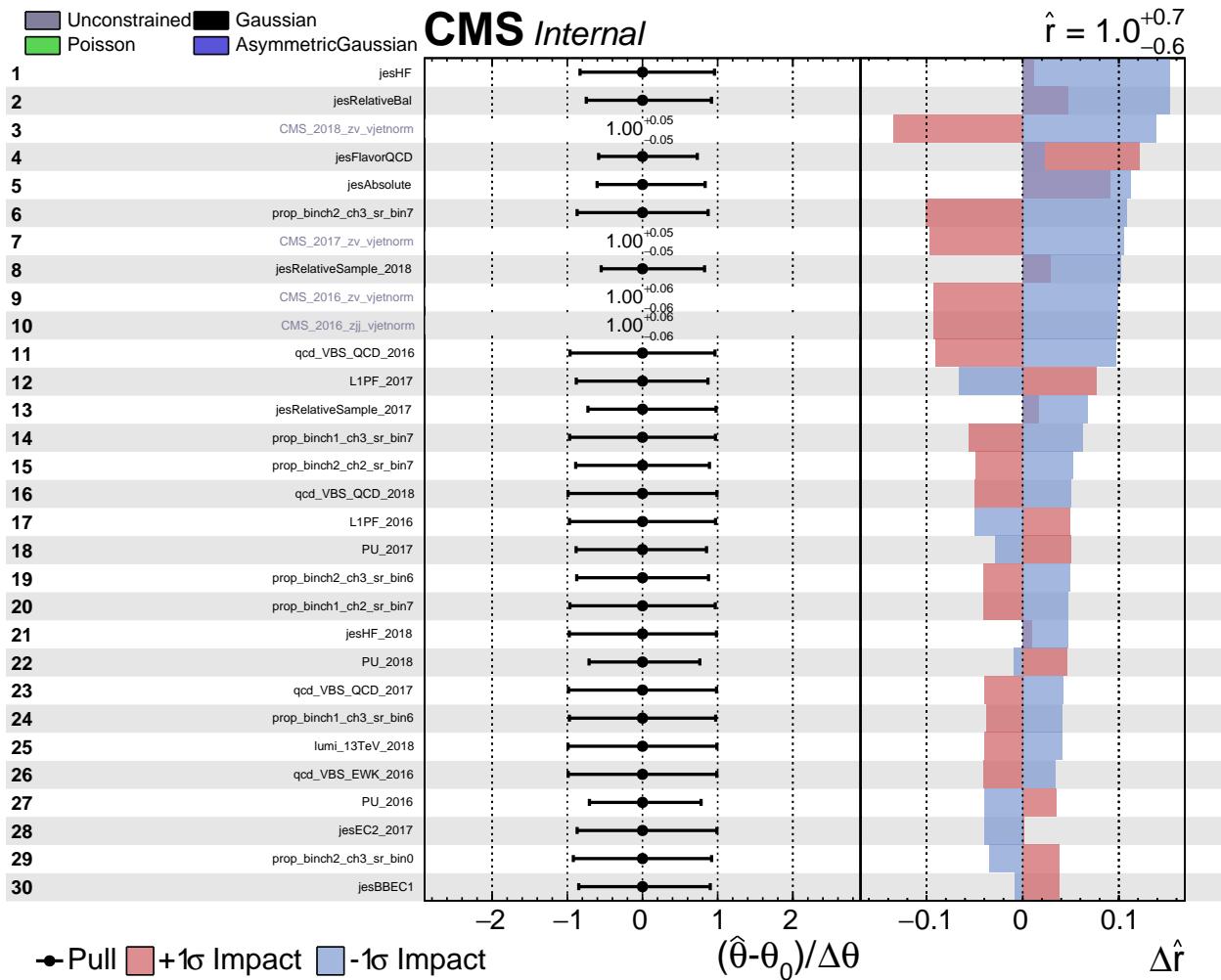


Figure 4.27: Impact Plots

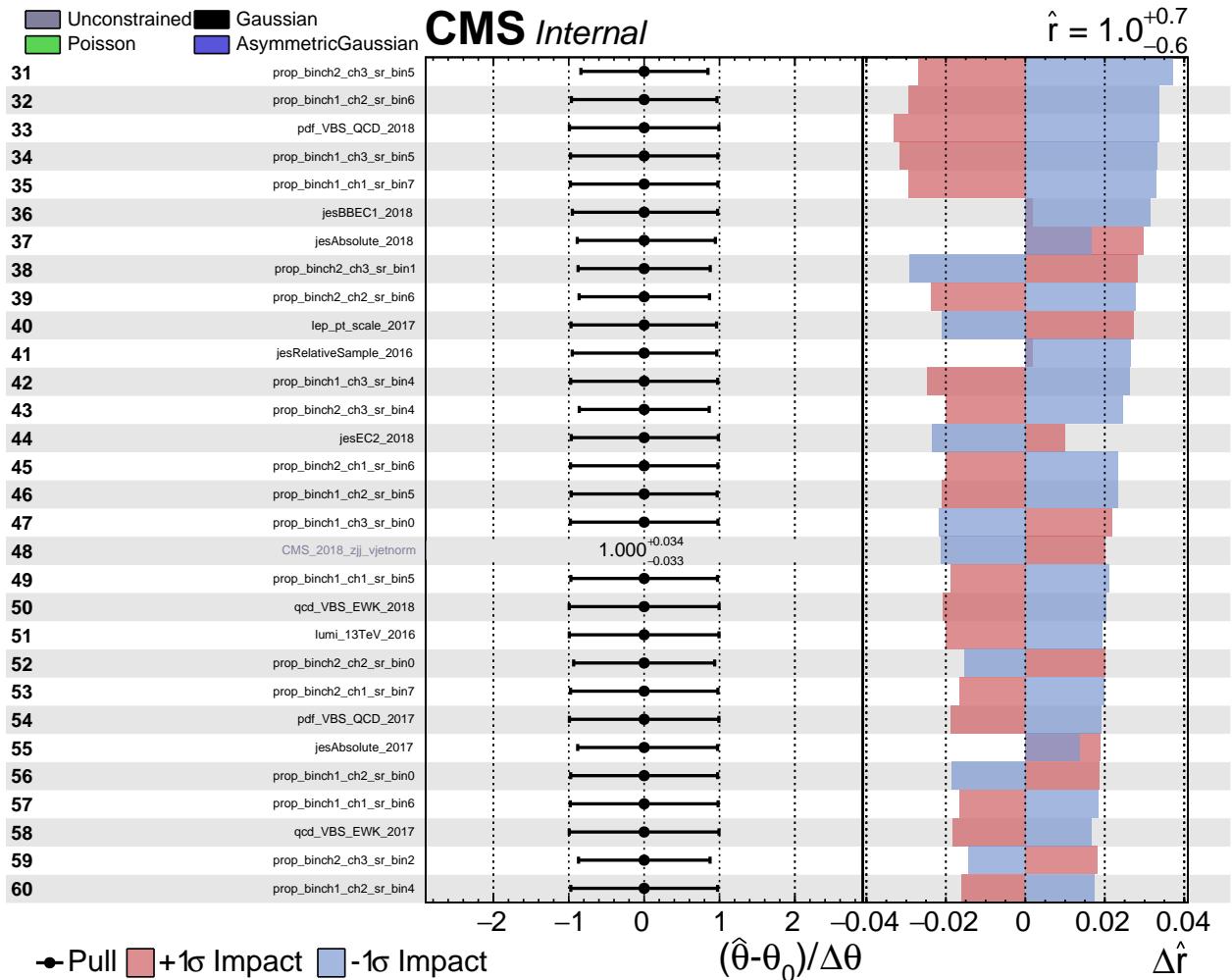


Figure 4.28: Impact Plots

976

4.6.4 Postfit Plots

Postfit distribution in signal region using Asimov dataset for Boosted ZV are shown in Figure 4.23 and for Resolved ZV are shown in Figure 4.30. MC processes overall prefit and postfit (signal + background and background-only) values in control and signal region are listed in Table 4.10 for Boosted ZV and in Table 4.11 for Resolved ZV category.

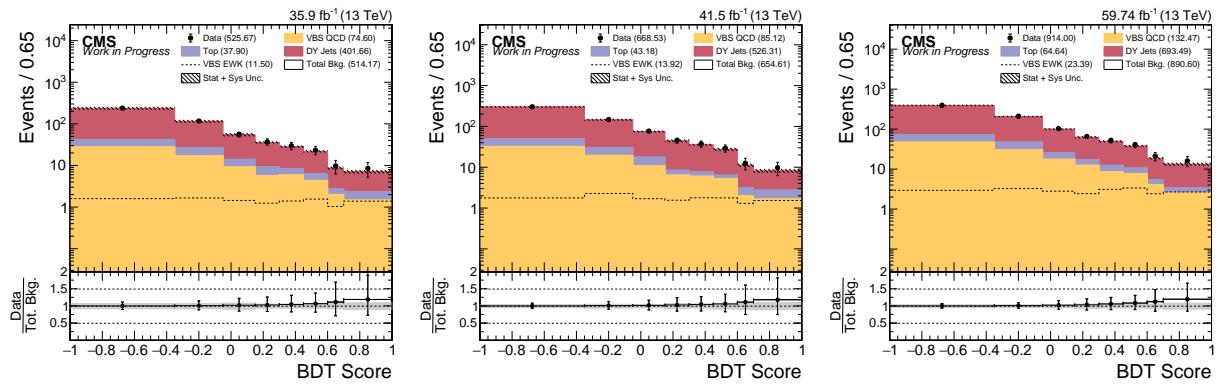


Figure 4.29: (Asimov Data) MVA Score postfit in Signal Region for Boosted ZV Channel.

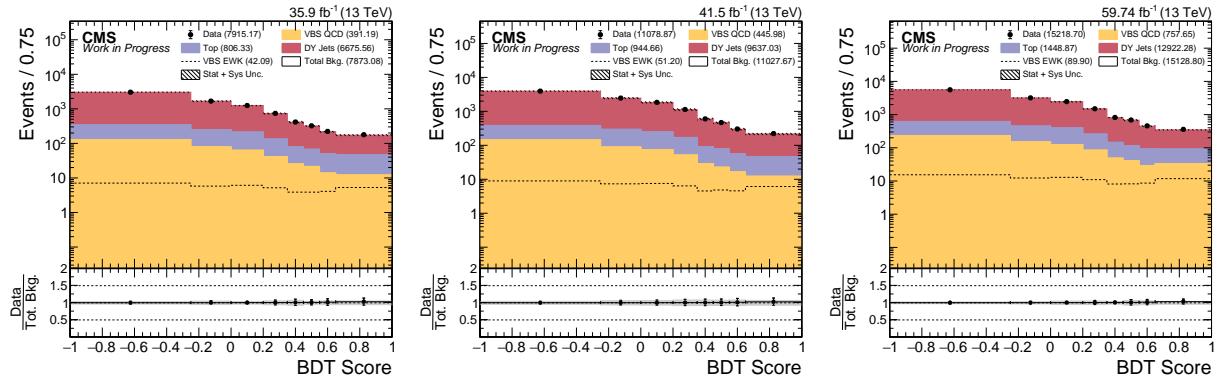


Figure 4.30: (Asimov Data) MVA Score postfit in Signal Region for Resolved ZV Channel. (Asimov Data)

Table 4.10: Prefit and postfit yields of MC processes

Region	Process	Pre-fit	S+B Fit	B-Only Fit
Boosted ZV 2016				
CR	DY+Jets	379.258 ± 11.645	366.292 ± 20.823	372.397 ± 17.543
CR	Top	32.874 ± 1.214	32.936 ± 1.146	32.949 ± 1.255
CR	VBS_EWK	5.230 ± 0.156	5.239 ± 8.585	0.000 ± 0.000
CR	VBS_QCD	38.659 ± 1.250	38.730 ± 1.199	38.748 ± 1.286
SR	DY+Jets	414.424 ± 37.150	401.658 ± 33.875	410.467 ± 33.476
SR	Top	37.557 ± 7.538	37.905 ± 7.583	38.173 ± 7.722
SR	VBS_EWK	11.280 ± 1.097	11.499 ± 19.267	0.000 ± 0.000
SR	VBS_QCD	73.244 ± 5.459	74.605 ± 4.790	75.410 ± 4.942
Boosted ZV 2017				
CR	DY+Jets	482.465 ± 14.680	477.902 ± 22.557	484.890 ± 18.593
CR	Top	25.458 ± 1.980	25.347 ± 1.839	25.335 ± 1.986
CR	VBS_EWK	5.584 ± 0.167	5.593 ± 8.960	0.000 ± 0.000
CR	VBS_QCD	42.475 ± 1.295	42.524 ± 1.262	42.543 ± 1.333
SR	DY+Jets	517.728 ± 28.192	526.307 ± 29.981	537.022 ± 21.508
SR	Top	42.445 ± 5.199	43.179 ± 5.010	43.436 ± 5.155
SR	VBS_EWK	13.635 ± 1.135	13.918 ± 22.190	0.000 ± 0.000
SR	VBS_QCD	83.087 ± 5.566	85.124 ± 5.179	86.000 ± 5.310
Boosted ZV 2018				
CR	DY+Jets	704.184 ± 21.318	658.283 ± 26.685	670.205 ± 23.886
CR	Top	47.431 ± 1.545	47.601 ± 1.435	47.630 ± 1.489
CR	VBS_EWK	10.045 ± 0.296	10.078 ± 10.857	0.000 ± 0.000
CR	VBS_QCD	68.708 ± 2.167	68.950 ± 1.982	68.996 ± 2.065
SR	DY+Jets	746.410 ± 41.307	693.492 ± 29.658	711.877 ± 26.629
SR	Top	64.774 ± 3.933	64.638 ± 2.715	65.198 ± 3.368
SR	VBS_EWK	23.232 ± 1.294	23.391 ± 24.686	0.000 ± 0.000
SR	VBS_QCD	131.658 ± 7.743	132.474 ± 6.005	133.943 ± 6.877

Table 4.11: Prefit and postfit yields of MC processes

Region	Process	Pre-fit	S+B Fit	B-Only Fit
Resolved ZV 2016				
CR	DY+Jets	6264.349 ± 190.352	6149.237 ± 92.197	6162.392 ± 85.695
CR	Top	718.306 ± 21.460	734.229 ± 21.392	735.064 ± 21.753
CR	VBS_EWK	13.240 ± 0.395	13.527 ± 29.811	0.000 ± 0.000
CR	VBS_QCD	189.095 ± 5.645	193.259 ± 5.631	193.479 ± 5.720
SR	DY+Jets	7088.493 ± 633.668	6675.562 ± 330.385	6710.664 ± 266.921
SR	Top	835.933 ± 69.920	806.328 ± 40.673	809.491 ± 39.805
SR	VBS_EWK	43.133 ± 2.811	42.092 ± 91.329	0.000 ± 0.000
SR	VBS_QCD	404.160 ± 34.797	391.192 ± 19.933	392.868 ± 20.162
Resolved ZV 2017				
CR	DY+Jets	9269.820 ± 280.782	8831.423 ± 109.274	8846.941 ± 100.456
CR	Top	732.264 ± 22.326	738.436 ± 21.576	739.105 ± 22.234
CR	VBS_EWK	16.719 ± 0.499	16.839 ± 39.897	0.000 ± 0.000
CR	VBS_QCD	212.000 ± 6.326	213.551 ± 6.143	213.741 ± 6.281
SR	DY+Jets	10051.850 ± 952.129	9637.035 ± 645.399	9681.931 ± 511.040
SR	Top	926.759 ± 64.180	944.658 ± 45.063	948.703 ± 44.739
SR	VBS_EWK	51.194 ± 3.178	51.201 ± 118.771	0.000 ± 0.000
SR	VBS_QCD	442.505 ± 34.337	445.977 ± 23.882	447.682 ± 24.279
Resolved ZV 2018				
CR	DY+Jets	12231.940 ± 370.848	11810.172 ± 129.492	11836.949 ± 122.098
CR	Top	1134.261 ± 33.468	1169.084 ± 28.849	1170.880 ± 31.799
CR	VBS_EWK	27.689 ± 0.816	28.538 ± 48.075	0.000 ± 0.000
CR	VBS_QCD	345.569 ± 10.189	356.174 ± 8.782	356.722 ± 9.687
SR	DY+Jets	13006.154 ± 1365.518	12922.276 ± 541.237	12998.396 ± 527.575
SR	Top	1393.580 ± 115.044	1448.865 ± 68.813	1455.162 ± 65.230
SR	VBS_EWK	87.014 ± 5.266	89.902 ± 149.233	0.000 ± 0.000
SR	VBS_QCD	712.871 ± 58.615	757.654 ± 35.477	762.648 ± 32.311

981

CHAPTER 5

982

HIGH GRANULARITY CALORIMETER UPGRADE

983 By the start of Run4 period of proton-proton collision at LHC, the collision energy
 984 is expected to reach its full design limit of 14 TeV and commissioning of High Luminosity
 985 LHC (HL-LHC) is expected to increase the luminosity by 10 times and integrated luminosity
 986 collected by the end of Run5 will be 3000 fb^{-1} .

987 With increased luminosity the CMS detector will get higher dose of radiation and average
 988 number of pileup interactions will be of the order $O(140)$ and endcap calorimeters ECAL
 989 and HCAL will suffer irreparable damage due to the much higher radiation dose received
 990 from increased luminosity in those regions. The High Granularity Calorimeter (HGCAL)
 991 is an upgrade that will replace current endcap calorimeters ECAL and HCAL. HGCAL is
 992 expected to be completed, and installed during Long Shutdown 3 (LS3) (2026–2028).

993 This chapter will discuss broadly the design of HGCAL, especially the scintillator section
 994 of the HGCAL and studies done at Northern Illinois Center for Accelerator and Detector
 995 Development (NICADD) towards its upgrade, for complete design see the Technical Design
 996 Report (TDR) [57].

997

5.1 Technical Design and Requirements

998 As mentioned the HGCAL will replace current endcap ECAL and HCAL, Figure 5.1
 999 shows exactly where the new detector will be placed. The right image in the Figure 5.1
 1000 shows side view ($z - r$ plane) of the detector, starting from the left i.e. innermost layers is

1001 Calorimeter Endcap Electromagnetic (CE-E) whose active layers are made of all silicon cells
 1002 ($\approx 0.5 - 1 \text{ cm}^2$), then majority of the detector in longitudinal length is Calorimeter Endcap
 1003 Hadronic (CE-H) whose starting few active layers are also all silicon cells, and then its mixed
 1004 with silicon cells in lower rings of the layer and the scintillator tiles ($\approx 5 - 31 \text{ cm}^2$) in the rest.
 1005 Silicon is radiation hard material i.e. the response of silicon cells will not degrade under high
 1006 rate of radiation dose, but silicon is also expensive and for such high cell count in HGCAL
 1007 it significantly increase the budget of the project, for this reason scintillator tiles are used
 1008 wherever radiation doses are low. Some of the main reasons for high cell count, lateral and
 1009 longitudinal granularity are to preserve energy resolution after 3000 fb^{-1} , precision timing
 1010 measurement of the showers to reject energy deposit from pileup, and been able to observe
 1011 narrower jets with $R = 0.2$. To be able to deliver these requirements both silicon cells and
 1012 scintillator tiles need to have good signal-to-noise ratio (S/N) even after 3000 fb^{-1} .

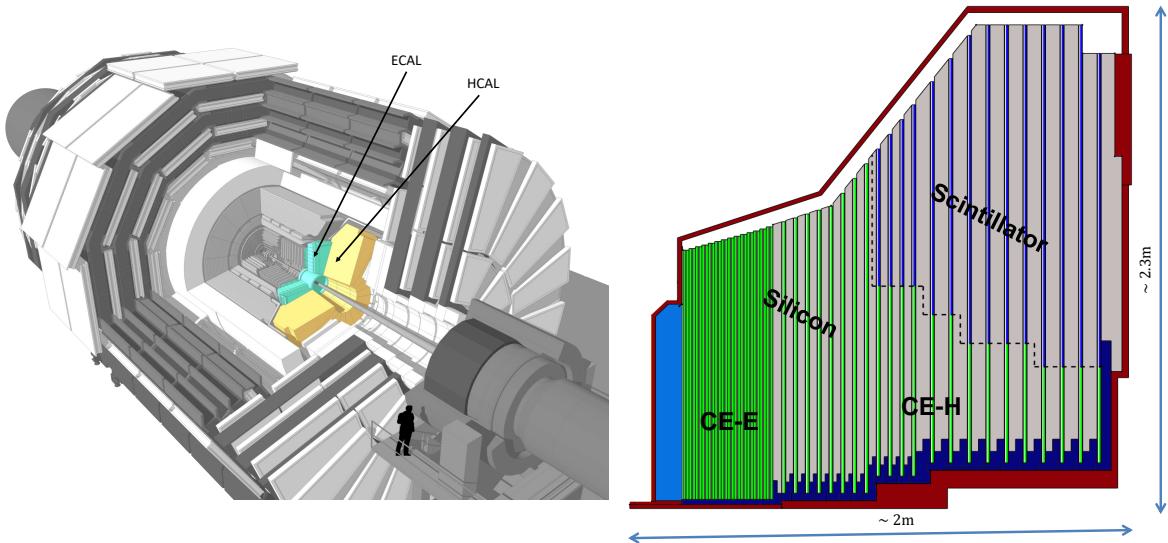


Figure 5.1: Overview of HGCAL position and quadrant view of Calorimeter Endcap (CE). Left image shows current endcap ECAL and HCAL highlighted. Right image shows quadrant view of HGCAL from the side [58, 59].

1013

5.2 Scintillator Tiles and SiPMs

1014 Silicon cells will be directly fabricated 8 inch silicon wafers (432 cells) as shown in Fig-
 1015 ure 5.2 as yellow and green colored hexagons, whereas each scintillator tiles will need to
 1016 prepared separately and assembled in the form of “tileboard” (about 8×8 tiles). Figure 5.2
 1017 shows scintillator tiles boundary with red grid lines, there are 288 scintillator tiles in each
 1018 ring, and each layer has different number of rings of scintillator tiles with maximum of 40
 1019 rings. To reduce the production cost and assembly complexities scintillator size are same for
 1020 every two rings, and ring number is used to identify tile size, for example R18–19 is the tile
 1021 size of tiles in ring number 18 and 19.

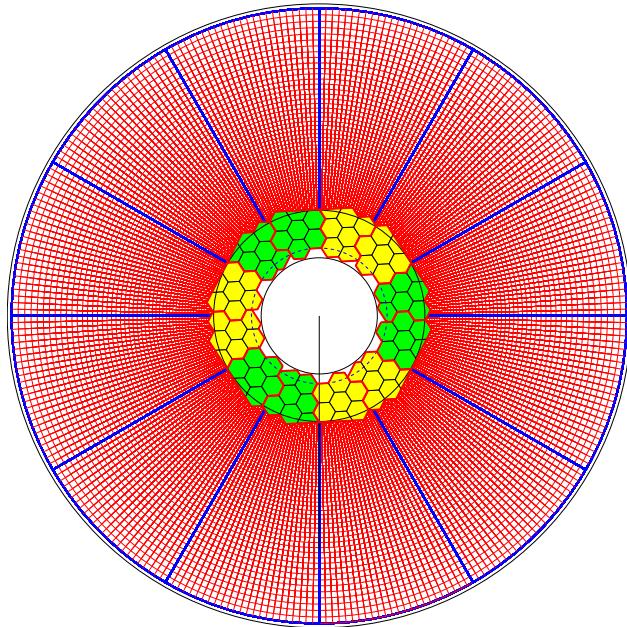


Figure 5.2: CE-H mixed layer 22. Yellow and green hexagons are the 8 inch silicon modules and the red grid lines shows 11,520 (40 rings) scintillator tiles. [57].

1022

5.2.1 Scintillator Materials

1023 Materials which scintillates i.e. emits light whenever a ionizing radiation passes through
 1024 it. Scintillator material can be crystals (current barrel and endcap ECAL), liquid or plastic
 1025 and are broadly divided into two categories organic or inorganic scintillators. HGCAL will
 1026 make use of organic plastic scintillator, and the most commonly used scintillator of this
 1027 type are polyvinyltoluene (PVT) and PS. PVT and PS are base material and by themselves
 1028 cannot serve the purpose of scintillator tiles for couple of reasons the light emitted is at lower
 1029 wavelengths (ultraviolet) and they are barely transparent to this light, for these reasons they
 1030 have small percentage of dopants added which absorbs the scintillation light and re-emits at
 1031 larger wavelengths (visible) and a second dopant is added to further increase the attenuation
 1032 length of the light emitted, so that light can be collected and coupled with light detection
 1033 system.

1034 PVT based scintillator will be referred to as “cast” scintillator, and they are usually
 1035 available in large sheets from vendors and need to be machined to make dimple and reduce
 1036 size. polystyrene (PS) based scintillator are called “injection molded” scintillator because
 1037 they are molded into desired tile shape.

1038 Cast scintillators are generally brighter (in terms of light output) than injection molded
 1039 scintillators, but the cost of cast scintillator per area is generally higher. Choice of scintillator
 1040 in addition to silicon photomultiplier (SiPM), which will discuss next impacts how will
 1041 optimizing cost and still retain good *S/N* for lifetime of HGCAL.

1042

5.2.2 SiPMs

1043 SiPM is a device that convert incident photon to electric current with large gain of 10^5 to
 1044 10^6 electrons. SiPM achieves this by pixels ($10\text{ }\mu\text{m}$ to $100\text{ }\mu\text{m}$ in size) connected in parallel,
 1045 where each pixel is a avalanche photodiode (APD) combined with quenching resistor. SiPM
 1046 of active area 2 mm^2 with $15\text{ }\mu\text{m}$ pixel has approximately 9,000 pixels, commercially SiPM
 1047 are also known as Multi-Pixel Photon Counter (MPPC) [60]. Figure 5.3 shows SiPM next
 1048 to tip of a pen.

1049 SiPM operates in reverse bias (Geiger Mode), with voltage above breakdown over voltage
 1050 (V_0), with linear relationship between V_0 and gain. In addition to low operating voltage
 1051 ($40 - 60\text{ V}$) the power consumption is also low, which increases with the number of pixel.
 1052 For HGCAL Hamamatsu S14160 [61] is being considered with area 2, 4 and 9 mm^2 . Larger
 1053 area means large signal, but also means higher power consumption which can be significant
 1054 given the granularity of the HGCAL.

1055 For HGCAL the SiPMs will be mounted on printed circuit board (PCB), and prepared
 1056 scintillator tiles will be directly glued on over SiPMs with dimple centering on SiPM's active
 1057 area.

1058

5.2.3 Scintillator Tiles

1059 Scintillator tiles coupled directly with SiPM alone can not provide sufficient S/N to reject
 1060 noise with certainty, additionally the response will be non-uniform centered at SiPM. Higher
 1061 signal and uniform response, both can be achieved by coating or wrapping in reflective
 1062 material [62]. Enhanced Specular Reflector (ESR) is a multi-layer highly reflective material
 1063 with $65\text{ }\mu\text{m}$ thickness, and is the material chosen for wrapping scintillator tiles with this.



Figure 5.3: SiPM

1064 As visible from Figure 5.2 the number of individual tiles in each layer is already very
 1065 high, and to wrap 100–200 thousand tiles with ESR is a very challenging task. At NICADD
 1066 we are building an automated wrapping machine (Section 5.3) to wrap scintillator tiles with
 1067 speed and repeatability. Figure 5.4 shows the R18–19 size scintillator tile, Figure 5.5 is the
 1068 ESR cut into shape for wrapping R18–19 tiles.

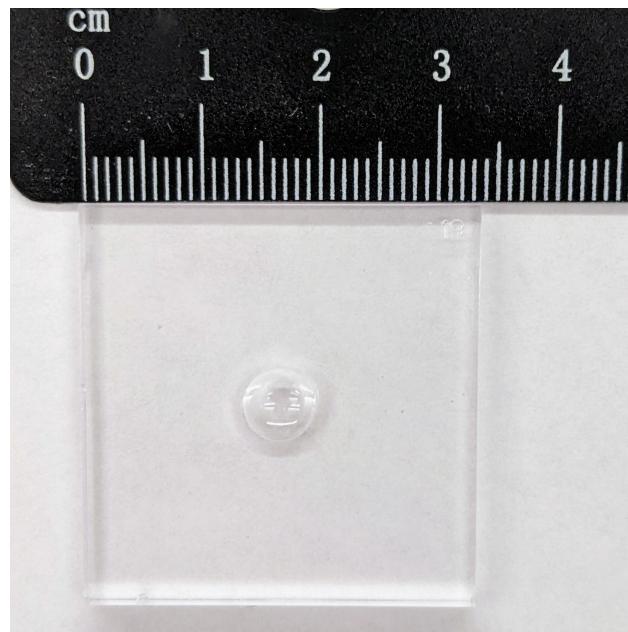


Figure 5.4: Scintillator tile with dimple

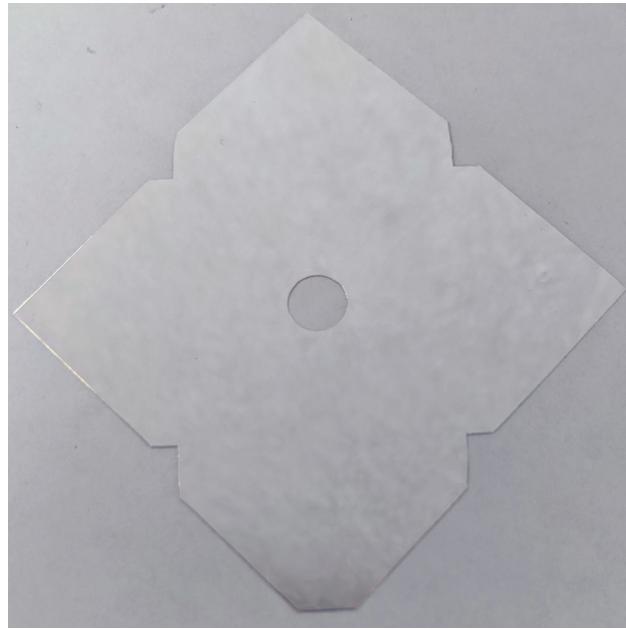


Figure 5.5: ESR wrapper cut for tile size R18–19

₁₀₆₉ The final wrapped tile in complete automated process with wrapping machine being
₁₀₇₀ developed and build at NICADD is shown in Figure 5.6.



Figure 5.6: R18–19 scintillator tile wrapped in ESR. Left: Up side of the tile, with Kapton sticker holding the flaps. Right: Bottom side with center hole over dimple.

1071 **5.3 Automated Wrapping of Scintillator Tiles**

1072 As mentioned in Section 5.2.3, the number of wrapped scintillator tiles in ESR required
1073 for the HGCAL is very large, 100–200 thousand tiles. In addition to large number of tiles,
1074 the tiles will be of different sizes and repeatability in wrapped tile quality is required for
1075 reliable performance of the detector. Figure 5.7 and 5.8 shows the overview of automated
1076 wrapped being developed at NICADD.

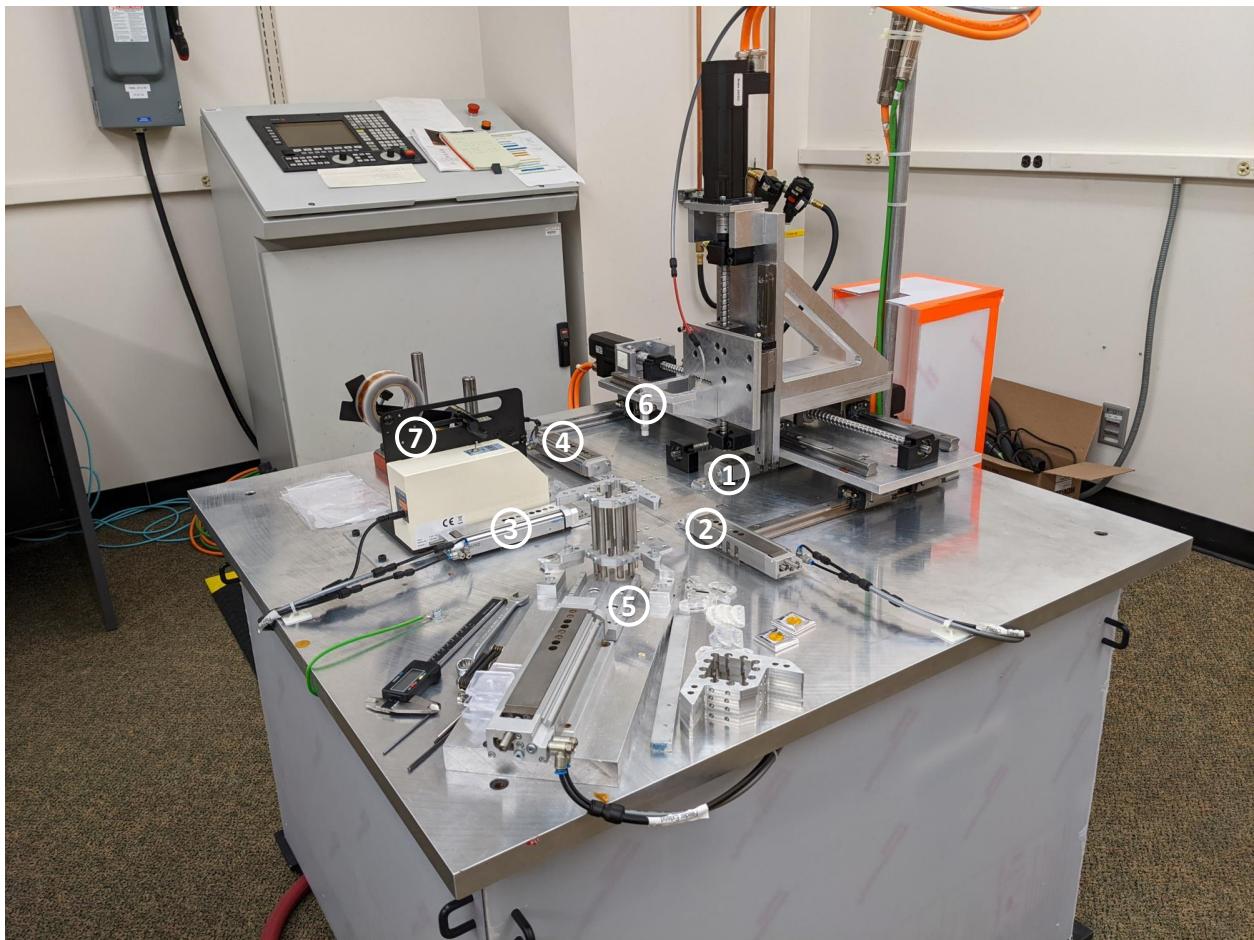


Figure 5.7: Automated scintillator tile wrapper overview. Shown in picture with labels are, 1–4: Actuator arms for folding the cut ESR flaps over the tiles, 5: Tile magazine and dispenser assembly, 6: z -axis end-effector with vacuum suction, 7: Kapton sticker dispenser.

1077 The machine is built to provide motion in *xyz*-axes and controlled by G-code program-
 1078 ming language. The second main components of the wrapper are actuators and vacuum
 1079 suction, actuator have arm that can extend and retract quickly with the pressurized air,
 1080 both pressurized air and vacuum are controlled by solenoid switch which is programmed the
 1081 help of programmable logic controller and can be used in G-code programs.

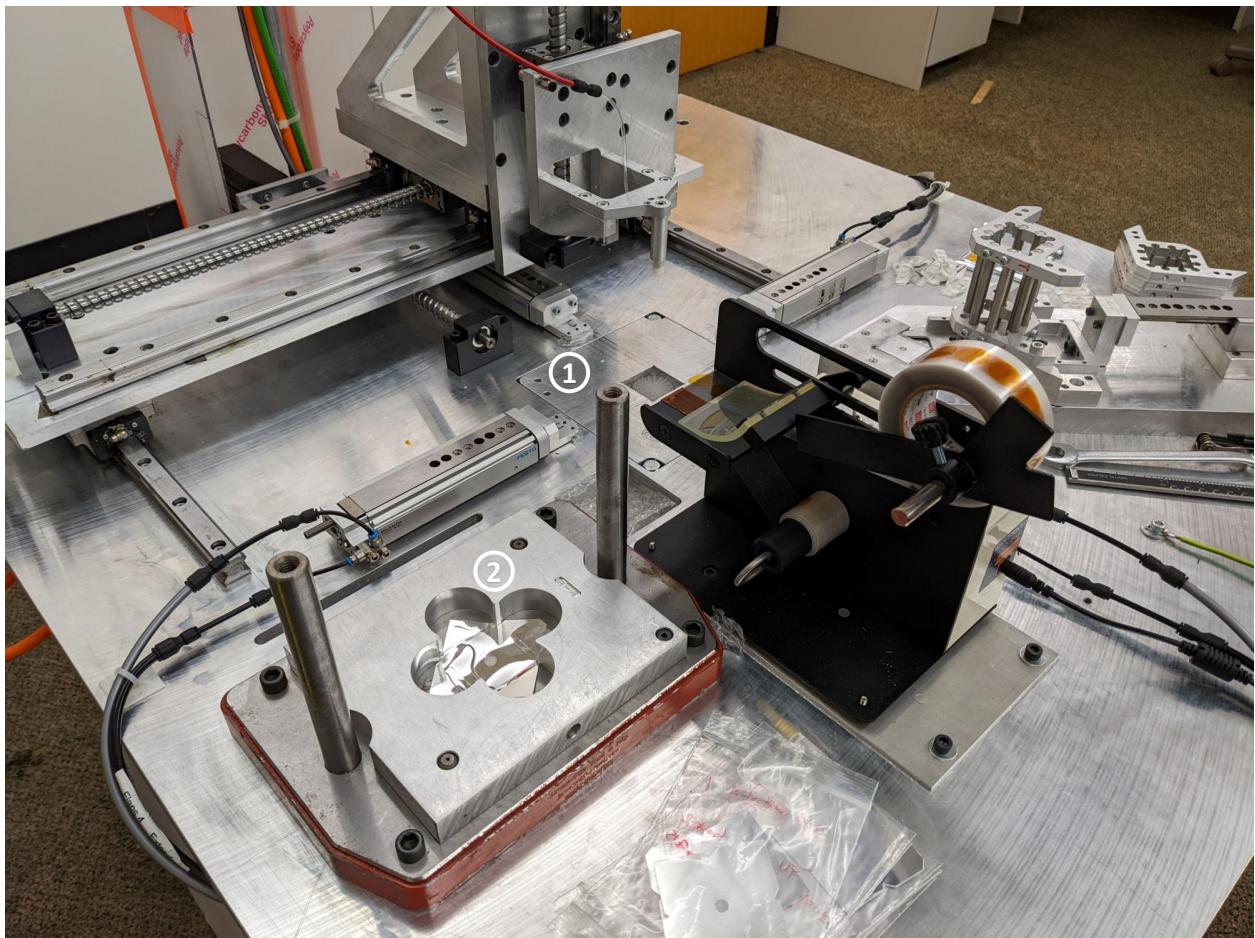


Figure 5.8: Tile folding station and ESR magazine. Shown in picture with labels 1: tile folding station, and 2: ESR magazine.

1082 The final wrapped tile as shown in Figure 5.6 is done by the machine in following steps:

- 1083 • **ESR Pickup, and Placement:** *z*-axis end-effector picks up the cut ESR using vac-
 1084 uum suction from the ESR magazine, and then places it precisely over the tile pocket

1085 in the folding station. Placed ESR over tile pocket is held into its place with the help
1086 of vacuum suction, which is built into tile pocket seat.

1087 • **Tile Dispenser, Pickup, and Placement:** Tiles will be stacked in the magazine
1088 at the tile dispenser location and an actuator arm will push out a tile completely and
1089 held in place (right image in Figure 5.9) until z -axis end-effector makes contact and
1090 vacuum suction is turned on. Now, the actuator arm is retracted and tile is picked up
1091 and placed over ESR at the folding station.

1092 • **ESR Flaps Folding:** After ESR and tile are placed, and since they are held strongly
1093 in place with vacuum, the tile pocket retracts with the help of actuator connected to it
1094 from the bottom. At this stage tile is inside pocket and flushed, and ESR flaps are out
1095 and vertical. Now the four actuator arms as shown in Figure 5.10 extends and closes
1096 ESR flaps.

1097 • **Sticker Application:** Now the folding actuator arms stay in place until sticker is
1098 applied using z -axis end-effector which picks up the sticker from sticker dispenser (left
1099 image Figure 5.9) and applies it on the center of the tile where all flaps corners meet.

1100 • **Wrapped Tile:** Final step in wrapping is retracting folding actuator arms, pushing out
1101 the tile and turning of the vacuum suction of tile pocket. Now z -axis end-effector picks
1102 up the wrapped tile and drops into the collection basket. Wrapped tile in Figure 5.6
1103 is wrapped automatically with this machine.

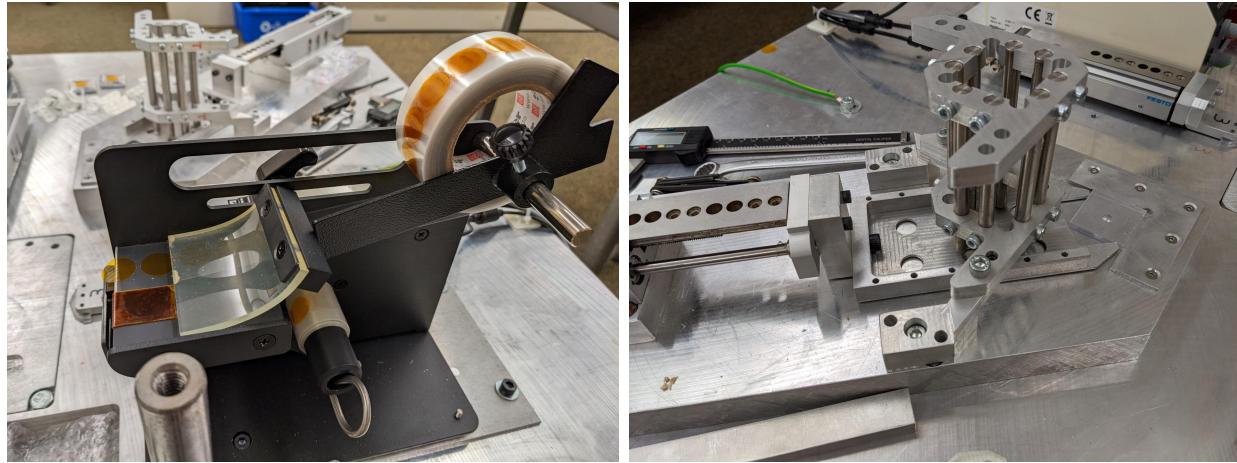


Figure 5.9: Sticker and tile dispenser, Left: Sticker dispenser with kapton tape roll. Right: Tile dispenser with tile magazine.

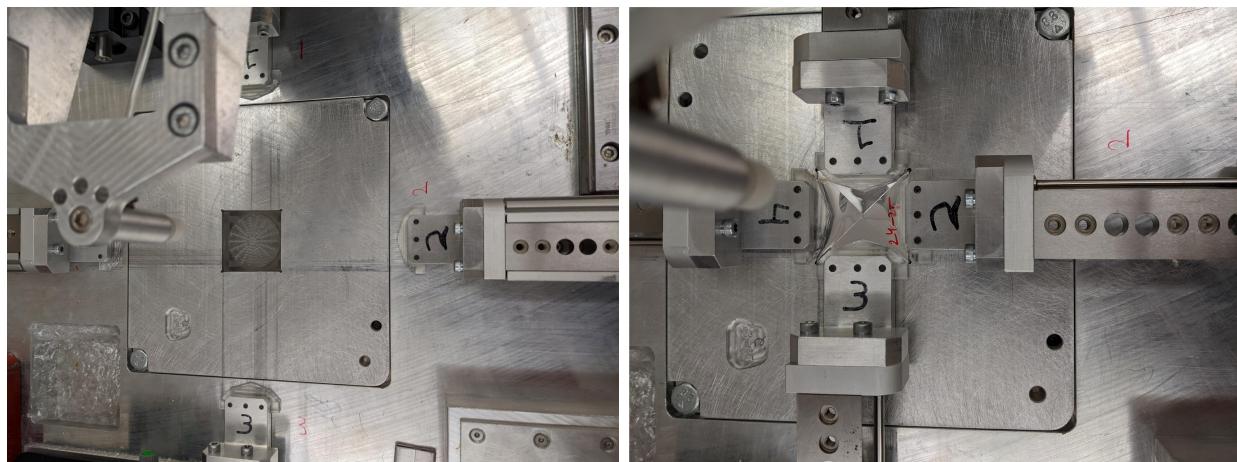


Figure 5.10: Tile folding station. Left: Flushed tile pocket at the center and retracted actuators. Right: After all the arms are extended during wrapping.

1104

5.4 Signal-to-Noise Ratio

1105 As discussed earlier in the Section 5.1, for HGCAL to retain its precision till the end-
 1106 of-life it needs good S/N , which is defined as $S/N > 3$. In this section we will discuss
 1107 formulation and input to the calculation of S/N , and the results of optimal configuration by
 1108 minimizing cost and retaining good S/N .

1109

5.4.1 Formulation

1110 MIP Signal-to-Noise ratio for the scintillator tile coupled directly with SiPM is formulated
 1111 as [63]:

$$\frac{S}{N} = \frac{(\text{MIP}^*) \sqrt{\frac{A_{t,\text{ref}}}{A_t}} \left(\frac{A_s}{A_{s,\text{ref}}} \right) (\text{Radiation Loss})}{(\text{SiPM}_{\text{noise},\text{base}}) \sqrt{\frac{A_s}{A_{s,\text{base}}}} \sqrt{1.88 \frac{T_s - T_{s,\text{base}}}{10^\circ \text{C}}} \sqrt{\frac{f}{f_{\text{base}}}}} \quad (5.1)$$

1112

where,

- 1113 • MIP^* : is the MIP measurement of the scintillator tile with a scale factor to account
 1114 for SiPM device photon detection efficacy (PDE) difference used during testbeam
 1115 measurement and SiPM expected to be used.
- 1116 • A_t : is the area of the tile for which the S/N is being evaluated and subscript **ref**
 1117 means the area of tile corresponding to MIP measurement.
- 1118 • A_s : similar to A_t , it is instead the area of the SiPM device coupled to scintillator tile.

- **Radiation Loss:** is the loss in light output due to the radiation dose received, it is expressed as:

$$e^{-R/D_c} \quad (5.2)$$

$$D_c = (6.0 \text{ Mrad}) \left(\frac{R}{1 \text{ krad/hr}} \right)^{0.35} \quad (5.3)$$

1119 where R is the dose rate in krad/hr, and D_c is the dose constant, and both are obtained
 1120 from FLUKA simulations.

- $\text{SiPM}_{\text{noise},\text{base}}$: is the root means square (RMS) value in photo electrons (p.e.) of signal noise received from SiPM from thermal excitation of electrons in pixels, also called dark current rate (DCR). In addition to thermal effects, irradiation of silicon also increases this noise. base in the subscript refers to DCR measurement conditions such as temperature of the SiPM $T_{s,\text{base}}$, area of SiPM $A_{s,\text{base}}$ and fluence f_{base} .
- T_s : is the temperature of the HGCAL hence the SiPM at which it will be operated, which is -30° C .
- f : is the amount of fluence SiPM will receive over its lifetime of operation in HGCAL i.e. after 3000 fb^{-1} .

1130 5.4.2 Testbeam and SiPM Noise Inputs

1131 Fermi National Accelerator Laboratory (FNAL) conducted testbeam measurement in
 1132 January 2020 on both cast and injection molded scintillator tiles wrapped in ESR with
 1133 SiPM using FNAL 120 GeV testbeam facility. The scintillator tiles used in testbeam were of

₁₁₃₄ dimensions $30 \times 30 \text{ mm}^2$ square tiles, and SiPM device used was Hamamatsu S13360–1350CS
₁₁₃₅ ($1.3 \times 1.3 \text{ mm}^2$) [60, 64].

₁₁₃₆ The MIP measured from testbeam are 35 p.e. for cast scintillator tiles, and 25 p.e.
₁₁₃₇ for injection molded with SiPM operated at voltage of 54.26 V, which is V_0 of 2.5V (I-V
₁₁₃₈ method) (equivalent to 3.0V when measured with gain method). Currently SiPM device
₁₁₃₉ class expected to used in HGCAL is Hamamatsu S14160 with $15 \mu\text{m}$ pixel size (dubbed as
₁₁₄₀ HDR15) and $2, 4, 9 \text{ mm}^2$ in area operated at V_0 of 2V (I-V method), using ratio of PDEs of
₁₁₄₁ these devices we can calculate PDE scale factor as,

$$= \frac{\text{PDE of S14160 at } V_O = 2V}{\text{PDE of S13360 at } V_O = 3V} = \frac{34.9}{40} = 0.8725 \quad (5.4)$$

₁₁₄₂ this gives, MIP* value to be 30.5 p.e. for cast and 21.8 p.e. for injection molded scintillator
₁₁₄₃ tiles.

₁₁₄₄ DCR measurement for HDR15 (2 mm^2) SiPMs irradiated to $5 \times 10^{13} \text{ n/cm}^2$ operated at
₁₁₄₅ $V_0 = 2V$ (I-V method) and at temperature -30° C is equivalent to RMS value of 19 p.e.
₁₁₄₆ with 15 ns integration time period.

₁₁₄₇ Using the testbeam measurement of scintillator tiles, and irradiated SiPM DCR mea-
₁₁₄₈ surement end-of-life scenario estimation of detector performance was done for combinations
₁₁₄₉ of types of scintillator tiles and different area of HDR15 SiPMs.

₁₁₅₀ 5.4.3 Scenarios

₁₁₅₁ Five combinations of scintillator material and SiPM active area were considered in two
₁₁₅₂ different scenarios as:

- ₁₁₅₃ • **Scene A:** In this scene, SiPMs with larger active area are preferred, followed by
₁₁₅₄ injection molded over cast scintillator.

₁₁₅₅ 1. Injection Molded Scintillator Tiles and SiPM of active area 2 mm^2 .

₁₁₅₆ 2. Injection Molded Scintillator Tiles and SiPM of active area 4 mm^2 .

₁₁₅₇ 3. Cast Scintillator Tiles and SiPM of active area 2 mm^2 .

₁₁₅₈ 4. Cast Scintillator Tiles and SiPM of active area 4 mm^2 .

₁₁₅₉ 5. Cast Scintillator Tiles and SiPM of active area 9 mm^2 .

- ₁₁₆₀ • **Scene B:** In this scene, brighter scintillator i.e. cast over injection is preferred, followed by increasing size of SiPMs.

₁₁₆₂ 1. Injection Molded Scintillator Tiles and SiPM of active area 2 mm^2 .

₁₁₆₃ 2. Cast Scintillator Tiles and SiPM of active area 2 mm^2 .

₁₁₆₄ 3. Injection Molded Scintillator Tiles and SiPM of active area 4 mm^2 .

₁₁₆₅ 4. Cast Scintillator Tiles and SiPM of active area 4 mm^2 .

₁₁₆₆ 5. Cast Scintillator Tiles and SiPM of active area 9 mm^2 .

₁₁₆₇ Individual's *S/N* of each combination when used alone is shown in Figure 5.11 after 3000

₁₁₆₈ fb^{-1} . Clearly injection molded scintillator cannot be used in leftmost layers, and even with

₁₁₆₉ cast scintillator it is possible only when using SiPM with large active area.

₁₁₇₀ 5.4.4 Results and Conclusion

₁₁₇₁ Since for assembly of scintillator tiles on tileboard, it is preferred to have single type of
₁₁₇₂ scintillator with SiPM combination. For this reason, each scene is evaluated in the preference
₁₁₇₃ order and tileboard is assigned a combination only if all the rings in it are able to satisfy
₁₁₇₄ $S/N > 3$. Figure 5.12 shows final results of how both scenes fill tileboards in CE-H.

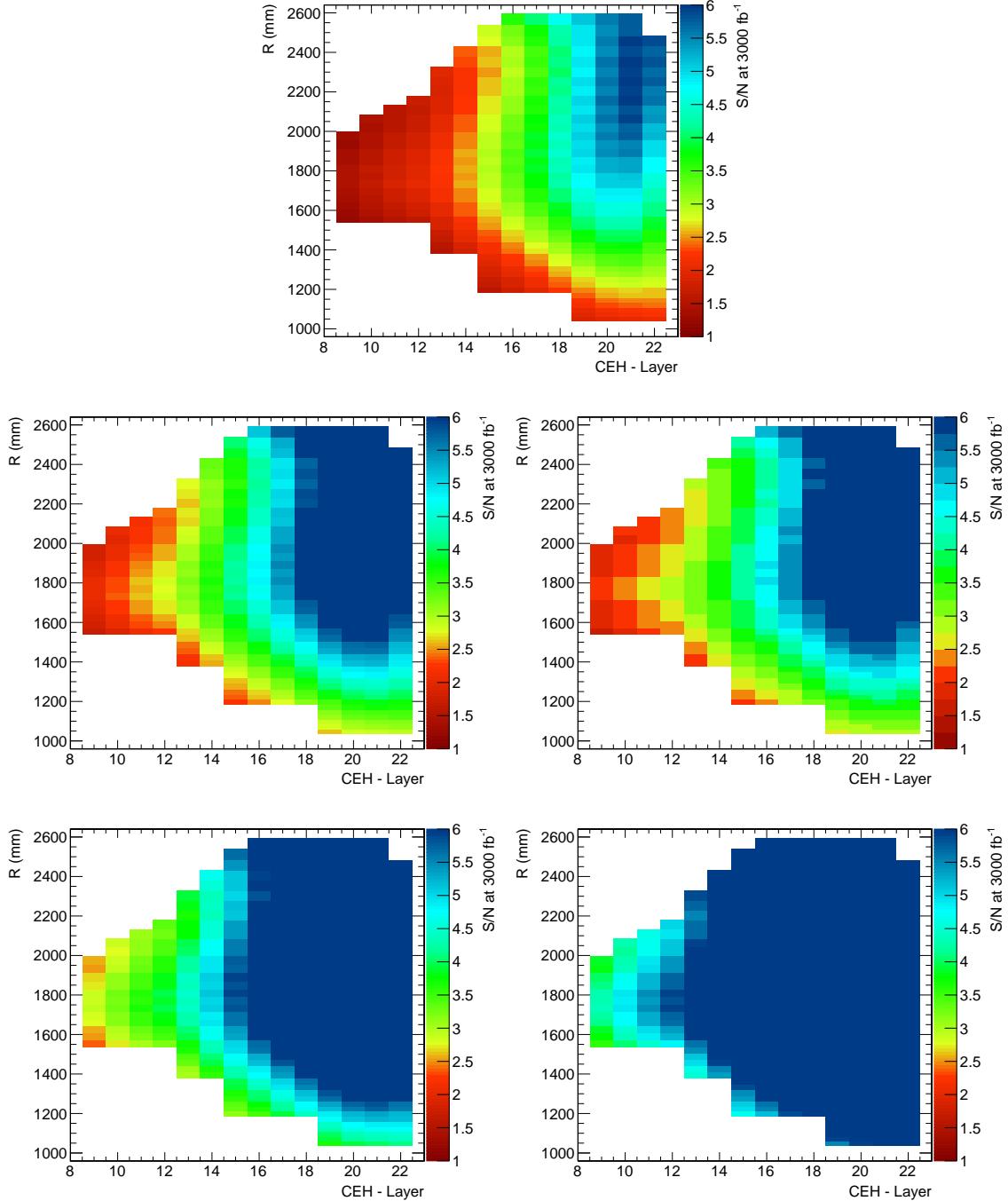


Figure 5.11: Scintillator performance with various active area size of SiPM. Top row from Left to Right: Injection Molded Scintillator with SiPM 2 and 4 mm^2 active area device. Bottom row from Left to Right: Cast Scintillator with SiPM 2, 4 and 9 mm^2 active area device.

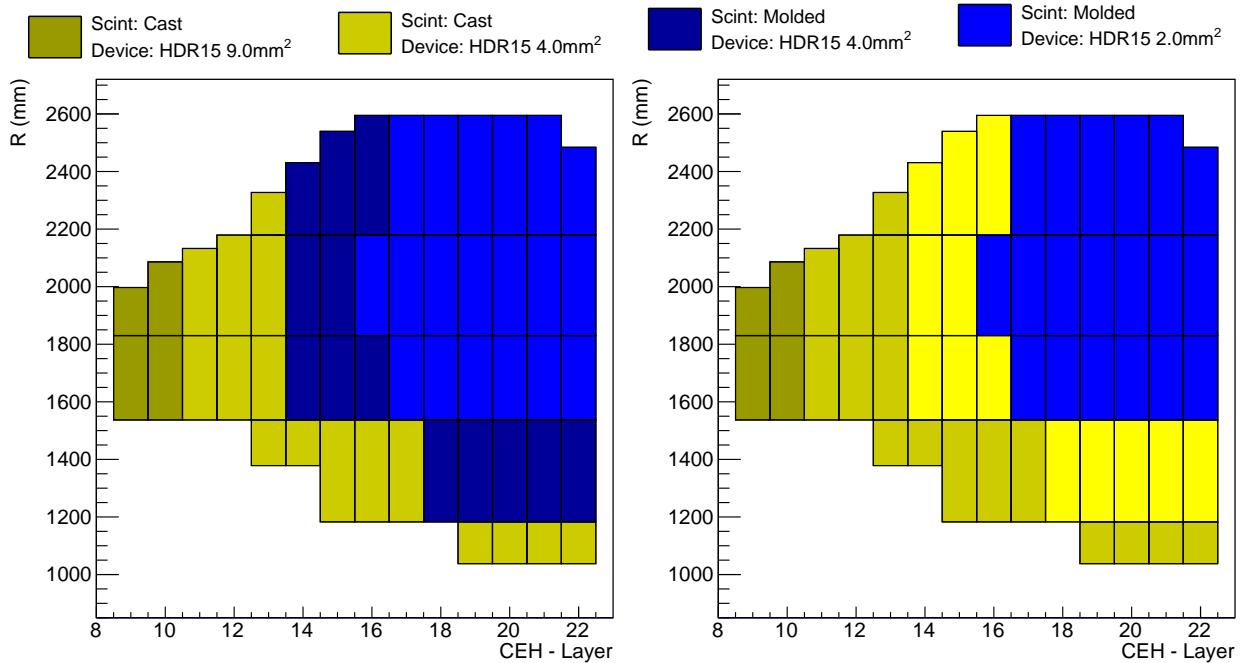


Figure 5.12: HGCAL scenarios. Left: Scene A, Right: Scene B

Table 5.1: HGCAL scenarios comparison

		Scene A	Scene B
Cast Scintillator	Cell Count	84, 096	153,216
	Total Area	101.59 m ²	197.46 m ²
	Percentage	27.8 %	54.0 %
Injection Molded Scintillator	Cell Count	155, 520	63, 360
	Total Area	264.04 m ²	168.17 m ²
	Percentage	72.2 %	46.0 %
SiPMs Count	2 mm ²	86, 400	69, 120
	4 mm ²	138, 240	155, 520
	9 mm ²	14, 976	14, 976

1175

CHAPTER 6

1176

CONCLUSIONS

1177 In this dissertation the contribution towards development of SM VBS analysis in semilep-
1178 tonic $ZVjj$ channel are discussed, and work done towards development and instrumentation
1179 of new the detector HGCAL for the CMS experiment are also discussed.

1180 This analysis is done with the 137 fb^{-1} of integrated luminosity data using 13 TeV proton-
1181 proton collision dataset collected by the CMS experiment during the run period 2016 to 2018.
1182 MVA approach was used to model signal versus background classifier using gradient boosted
1183 BDT in the signal region. To correct and normalize DY plus jets background model, a control
1184 region defined using hadronic boson mass was used. Expected significance of 1.5σ is reported
1185 for EW VBS $ZVjj$. The analysis is currently being developed further and under consideration
1186 for pre-approval by the Standard Model Physics group of the CMS collaboration.

1187 For analysis like this which requires higher jet multiplicity in an event, the sensitivity of
1188 the analysis suffers greatly from the pileup contamination. HGCAL in addition to replacing
1189 dying ECAL and HCAL hardware, it will also help many VBS analysis, since the jets in
1190 endcap suffers the most from pileup contamination, by allowing to construct narrow jets using
1191 lateral and longitudinal granularity of the silicon and scintillator cells, pileup contamination
1192 can be significantly reduced. HGCAL is expected to installed during LS3 which is currently
1193 expected to be from end of year 2025 to the start of year 2029.

1194 In Chapter 5 of this dissertation, optimal configuration of scintillator tiles coupled with
1195 SiPMs were studied and suggested by simulating end-of-life scenarios of the HGCAL. Results
1196 made use of testbeam measurement of the scintillator tiles conducted by FNAL and the cold
1197 noise measurement of SiPMs.

1198 To wrap scintillator tiles with ESR film is challenging task because of some inflexibility
1199 in it. To wrap more than hundred thousand for already difficult task requires automation to
1200 wrap faster with repeatability. Complete automated wrapping of scintillator tiles with ESR
1201 with wrapping machine at NICADD were also discussed in the Chapter 5.

REFERENCES

- [1] Wikipedia contributors. “Standard Model of Elementary Particles” (July 2021). Accessed September 07, 2021. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [2] 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](https://doi.org/10.1103/PhysRev.96.191). URL: <https://link.aps.org/doi/10.1103/PhysRev.96.191>.
- [3] S. L. Glashow. “The renormalizability of vector meson interactions”. *Nuclear Physics* 10 (Feb. 1959), pp. 107–117. DOI: [10.1016/0029-5582\(59\)90196-8](https://doi.org/10.1016/0029-5582(59)90196-8).
- [4] S. Weinberg. “A Model of Leptons”. *Physical Review Letters* 19 (Nov. 1967), pp. 1264–1266. DOI: [10.1103/physrevlett.19.1264](https://doi.org/10.1103/physrevlett.19.1264).
- [5] A. Salam and J. C. Ward. “Weak and electromagnetic interactions”. *Il Nuovo Cimento* 11 (Feb. 1959), pp. 568–577. DOI: [10.1007/bf02726525](https://doi.org/10.1007/bf02726525).
- [6] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. *Physical Review Letters* 13 (Aug. 1964), pp. 321–323. DOI: [10.1103/physrevlett.13.321](https://doi.org/10.1103/physrevlett.13.321).
- [7] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. *Physical Review Letters* 13 (Oct. 1964), pp. 508–509. DOI: [10.1103/physrevlett.13.508](https://doi.org/10.1103/physrevlett.13.508).
- [8] J. Ellis. “Higgs Physics”. en (2015). DOI: [10.5170/CERN-2015-004.117](https://doi.org/10.5170/CERN-2015-004.117).
- [9] B. W. Lee, C. Quigg, and H. B. Thacker. “Weak interactions at very high energies: The role of the Higgs-boson mass”. *Physical Review D* 16 (Sept. 1977), pp. 1519–1531. DOI: [10.1103/physrevd.16.1519](https://doi.org/10.1103/physrevd.16.1519).

- [10] B. W. Lee, C. Quigg, and H. B. Thacker. “Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass”. *Physical Review Letters* 38 (Apr. 1977), pp. 883–885. DOI: 10.1103/physrevlett.38.883.
- [11] A. Denner and T. Hahn. “Radiative Corrections to $W^+W^- \rightarrow W^+W^-$ in the Electroweak Standard Model”. *Nucl.Phys.B* 525:27-50, 1998 (Nov. 12, 1997). DOI: 10.1016/S0550-3213(98)00287-9.
- [12] CMS Collaboration. “Evidence for WW/WZ vector boson scattering in the decay channel $l\nu qq$ produced in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV” (Dec. 10, 2021). arXiv: 2112.05259 [hep-ex].
- [13] CMS Collaboration. “Search for anomalous electroweak production of vector boson pairs in association with two jets in proton-proton collisions at 13 TeV”. *Phys. Lett. B* 798 (2019) 134985 (May 17, 2019). DOI: 10.1016/j.physletb.2019.134985. arXiv: 1905.07445 [hep-ex].
- [14] L. Evans and P. Bryant. “LHC Machine”. *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.
- [15] E. Mobs. “The CERN accelerator complex - 2019” (July 2019). Accessed August 16, 2021. URL: <http://cds.cern.ch/record/2684277/files/>.
- [16] 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>.
- [17] The CMS Collaboration. “CMS Luminosity Measurements for the 2016 Data Taking Period” (2017). URL: <https://cds.cern.ch/record/2257069>.
- [18] 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>.

- 1248 [19] The CMS Collaboration. “CMS luminosity measurement for the 2018 data-taking pe-
1249 riod at $\sqrt{s} = 13$ TeV” (2019). URL: <https://cds.cern.ch/record/2676164>.
- 1250 [20] T. Sakuma. “Cutaway diagrams of CMS detector” (May 2019). Accessed August 17,
1251 2021. URL: <https://cds.cern.ch/record/2665537>.
- 1252 [21] The CMS Collaboration. “The CMS experiment at the CERN LHC”. *Journal of Instru-*
1253 *mentation* 3.08 (Aug. 2008), S08004–S08004. DOI: [10.1088/1748-0221/3/08/s08004](https://doi.org/10.1088/1748-0221/3/08/s08004).
1254 URL: <https://doi.org/10.1088/1748-0221/3/08/s08004>.
- 1255 [22] The CMS Collaboration. “CMS Detector Slice” (Jan. 2016). Accessed August 18, 2021.
1256 URL: <https://cds.cern.ch/record/2120661>.
- 1257 [23] M. Brice and C. Marcelloni. “The central part of CMS is lowered” (Feb. 2007). Accessed
1258 August 22, 2021. URL: <https://cds.cern.ch/record/1020311>.
- 1259 [24] T. Sakuma. “SketchUp Drawings of the Phase 1 pixel detector alongside the current
1260 pixel detector” (June 2012). Accessed August 17, 2021. URL: <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB>ShowDocument?docid=6473>.
- 1261 [25] The CMS Collaboration. “The CMS ECAL performance with examples”. *Journal of
1263 Instrumentation* 9.02 (Feb. 2014), pp. C02008–C02008. DOI: [10.1088/1748-0221/9/02/c02008](https://doi.org/10.1088/1748-0221/9/02/c02008).
- 1262 [26] The CMS Collaboration. “Images of CMS ECAL Endcap (EE)” (Nov. 2008). Accessed
1264 August 22, 2021. URL: <https://cds.cern.ch/record/1431479>.
- 1265 [27] CMS Collaboration. “Energy calibration and resolution of the CMS electromagnetic
1267 calorimeter in pp collisions at $\sqrt{s} = 7$ TeV”. *JINST* 8 (2013) P09009 (June 9,
1268 2013). DOI: [10.1088/1748-0221/8/09/P09009](https://doi.org/10.1088/1748-0221/8/09/P09009). arXiv: [1306.2016 \[hep-ex\]](https://arxiv.org/abs/1306.2016).
- 1269 [28] J. Mans et al. “CMS Technical Design Report for the Phase 1 Upgrade of the Hadron
1271 Calorimeter” (Sept. 2012). URL: <https://cds.cern.ch/record/1481837>.

- 1272 [29] The CMS Collaboration. “Images of the CMS HCAL Barrel (HB)” (2008). Accessed
1273 August 22, 2021. URL: <https://cds.cern.ch/record/1431485>.
- 1274 [30] The CMS Collaboration. “HCAL Depth Segmentation Phase 1 Upgrade” (2012). Ac-
1275 cessed August 22, 2021. URL: <https://home.fnal.gov/~chlebana/CMS/Phase1>.
- 1276 [31] The CMS Collaboration. “Performance of the CMS Muon Detectors in early 2016
1277 collision runs” (2016). Accessed August 22, 2021. URL: <https://twiki.cern.ch/twiki/pub/CMSPublic/MuonDPGPublic160729>.
- 1279 [32] The CMS Collaboration. “Particle-flow reconstruction and global event description
1280 with the CMS detector”. *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–
1281 P10003. DOI: [10.1088/1748-0221/12/10/p10003](https://doi.org/10.1088/1748-0221/12/10/p10003).
- 1282 [33] S. Cucciarelli et al. “Track reconstruction, primary vertex finding and seed generation
1283 with the Pixel Detector”. CMS-NOTE-2006-026 (Jan. 2006). URL: <https://cds.cern.ch/record/927384>.
- 1285 [34] W. Adam et al. “Track Reconstruction in the CMS tracker”. CMS-NOTE-2006-041
1286 (Dec. 2006). URL: <https://cds.cern.ch/record/934067>.
- 1287 [35] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction
1288 with proton-proton collisions at $\sqrt{s} = 13$ TeV”. *JINST* 13 (2018) P06015 (Apr. 12,
1289 2018). DOI: [10.1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015). arXiv: [1804.04528](https://arxiv.org/abs/1804.04528).
- 1290 [36] W. Adam et al. “Reconstruction of electrons with the Gaussian-sum filter in the CMS
1291 tracker at the LHC”. *Journal of Physics G: Nuclear and Particle Physics* 31.9 (July
1292 2005), N9–N20. DOI: [10.1088/0954-3899/31/9/n01](https://doi.org/10.1088/0954-3899/31/9/n01).
- 1293 [37] The CMS Collaboration. “Electron and photon reconstruction and identification with
1294 the CMS experiment at the CERN LHC”. *JINST* 16 (2021) P05014 (Dec. 12, 2020).
1295 DOI: [10.1088/1748-0221/16/05/P05014](https://doi.org/10.1088/1748-0221/16/05/P05014).

- [38] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet user manual” (Nov. 25, 2011). DOI: [10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2).
- [39] D. Bertolini et al. “Pileup Per Particle Identification”. *JHEP* **1410** (2014) 59 (July 22, 2014). DOI: [10.1007/JHEP10\(2014\)059](https://doi.org/10.1007/JHEP10(2014)059).
- [40] CMS Collaboration. “Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018” (Apr. 2020). URL: <https://cds.cern.ch/record/2715872>.
- [41] 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](https://doi.org/10.1088/1748-0221/15/09/p09018).
- [42] 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](https://doi.org/10.1007/JHEP03(2011)015).
- [43] 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](https://doi.org/10.1088/1748-0221/15/06/P06005).
- [44] A. J. Larkoski et al. “Soft Drop”. *JHEP* **1405** (2014) **146** (Feb. 11, 2014). DOI: [10.1007/JHEP05\(2014\)146](https://doi.org/10.1007/JHEP05(2014)146).
- [45] 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” (May 1, 2014). DOI: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079). arXiv: 1405.0301 [hep-ph].
- [46] P. Artoisenet et al. “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”. *Journal of High Energy Physics, Volume 2013, Issue 3* (Dec. 14, 2012). DOI: [10.1007/JHEP03\(2013\)015](https://doi.org/10.1007/JHEP03(2013)015). arXiv: 1212.3460 [hep-ph].

- 1319 [47] B. Andersson et al. “Parton fragmentation and string dynamics”. *Physics Reports*
 1320 97.2-3 (July 1983), pp. 31–145. DOI: [10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7).
- 1321 [48] CMS Collaboration. “Event generator tunes obtained from underlying event and mul-
 1322 tiparton scattering measurements”. *Eur. Phys. J. C* 76 (2016) 155 (Dec. 2, 2015). DOI:
 1323 [10.1140/epjc/s10052-016-3988-x](https://doi.org/10.1140/epjc/s10052-016-3988-x). arXiv: [1512.00815](https://arxiv.org/abs/1512.00815) [hep-ex].
- 1324 [49] CMS Collaboration. “Extraction and validation of a new set of CMS PYTHIA8 tunes
 1325 from underlying-event measurements”. *Eur. Phys. J. C* 80 (2020) 4 (Mar. 28, 2019).
 1326 DOI: [10.1140/epjc/s10052-019-7499-4](https://doi.org/10.1140/epjc/s10052-019-7499-4). arXiv: [1903.12179](https://arxiv.org/abs/1903.12179) [hep-ex].
- 1327 [50] CMS Collaboration. “Performance of the CMS Level-1 trigger in proton-proton colli-
 1328 sions at $\sqrt{s} = 13$ TeV”. *JINST* 15 (2020) P10017 (June 17, 2020). DOI: [10.1088/1748-0221/15/10/P10017](https://doi.org/10.1088/1748-0221/15/10/P10017). arXiv: [2006.10165](https://arxiv.org/abs/2006.10165) [hep-ex].
- 1330 [51] A. Hoecker et al. “TMVA - Toolkit for Multivariate Data Analysis” (Mar. 4, 2007).
 1331 arXiv: [physics/0703039](https://arxiv.org/abs/physics/0703039) [physics.data-an].
- 1332 [52] W. Verkerke and D. Kirkby. “The RooFit toolkit for data modeling” (June 14, 2003).
 1333 arXiv: [physics/0306116](https://arxiv.org/abs/physics/0306116) [physics.data-an].
- 1334 [53] L. Moneta et al. “The RooStats Project” (Sept. 6, 2010). arXiv: [1009.1003](https://arxiv.org/abs/1009.1003).
- 1335 [54] R. Barlow and C. Beeston. “Fitting using finite Monte Carlo samples”. *Computer
 1336 Physics Communications* 77.2 (Oct. 1993), pp. 219–228. DOI: [10.1016/0010-4655\(93\)90005-w](https://doi.org/10.1016/0010-4655(93)

 1337 90005-w).
- 1338 [55] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”.
 1339 *Eur.Phys.J.C*71:1554,2011 (July 10, 2010). DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0).
 1340 arXiv: [1007.1727](https://arxiv.org/abs/1007.1727) [physics.data-an].

- 1341 [56] ATLAS Collaboration, CMS Collaboration, LHC Higgs Combination Group. “Procedure
1342 for the LHC Higgs boson search combination in Summer 2011”. CMS-NOTE-
1343 2011-005, ATL-PHYS-PUB-2011-11 (Aug. 2011). URL: <https://cds.cern.ch/record/1379837>.
- 1344
- 1345 [57] CMS Collaboration. “The Phase-2 Upgrade of the CMS Endcap Calorimeter” (Nov.
1346 2017). DOI: 10.17181/CERN.1V8M.1JY2. URL: <https://cds.cern.ch/record/2293646>.
- 1347
- 1348 [58] D. Barney. “Overview slide of CE with main parameters” (Oct. 2019). Accessed August
1349 24, 2021. URL: <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB>ShowDocument?docid=13251>.
- 1350
- 1351 [59] D. Barney and T. Sakuma. “Sketchup images highlighting the sub-detectors” (Sept.
1352 2017). Accessed August 24, 2021. URL: <https://cds.cern.ch/record/2628519>.
- 1353 [60] Hamamatsu. “S13360 series MPPC” (June 2022). Accessed June 13, 2022. URL: https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/ssd/s13360_series_kapd1052e.pdf.
- 1354
- 1355
- 1356 [61] Hamamatsu. “S14160 series MPPC” (June 2022). Accessed June 13, 2022. URL: https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRARY/ssd/s14160_s14161_series_kapd1064e.pdf.
- 1357
- 1358
- 1359 [62] G. Blazey et al. “Directly coupled tiles as elements of a scintillator calorimeter with
1360 MPPC readout”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 605.3 (July 2009),
1361 pp. 277–281. DOI: 10.1016/j.nima.2009.03.253.
- 1362
- 1363 [63] J. Mans and P. de Barbaro. “The baseline readout of the scintillator section of the
1364 CMS HL-LHC Endcap Calorimeter” (Feb. 2017).

- ¹³⁶⁵ [64] A. Belloni et al. “Test beam study of SiPM-on-tile configurations”. *Journal of Instrumentation* 16.07 (July 2021), P07022. DOI: 10.1088/1748-0221/16/07/p07022.
- ¹³⁶⁶

1367

APPENDIX A

1368

ANALYSIS CODE

1369 Analysis code used for the analysis is hosted on Github (<https://github.com>) platform.

- 1370 • Custom NanoAOD production from MiniAOD to include missing PDF weights.

1371 <https://github.com/singh-ramanpreet/VBS-customNanoAODProduction/>

- 1372 • “NanoSkim”: Intermediate skimming step for the analysis phase space with minimal
1373 selection to save time, when run it again during analysis development.

1374 <https://github.com/singh-ramanpreet/VVjjSemileptonic-NanoSkim>

- 1375 • “Selection”: This repo contains the code for main event selection of this analysis, it
1376 also calculates and embed scale factors for various objects.

1377 <https://github.com/singh-ramanpreet/VVjjSemileptonic-Selection>

- 1378 • “Analysis”: This repo contains code MVA training, MVA inference and embedding,
1379 making Data/MC histograms, making datacards for the statistical analysis with “com-
1380 bine”.

1381 <https://github.com/singh-ramanpreet/VVjjSemileptonic-Analysis>

1382

APPENDIX B

1383

ADDITIONAL FIGURES

1384 Add backup plots.

1385

B.1 Impact Plots

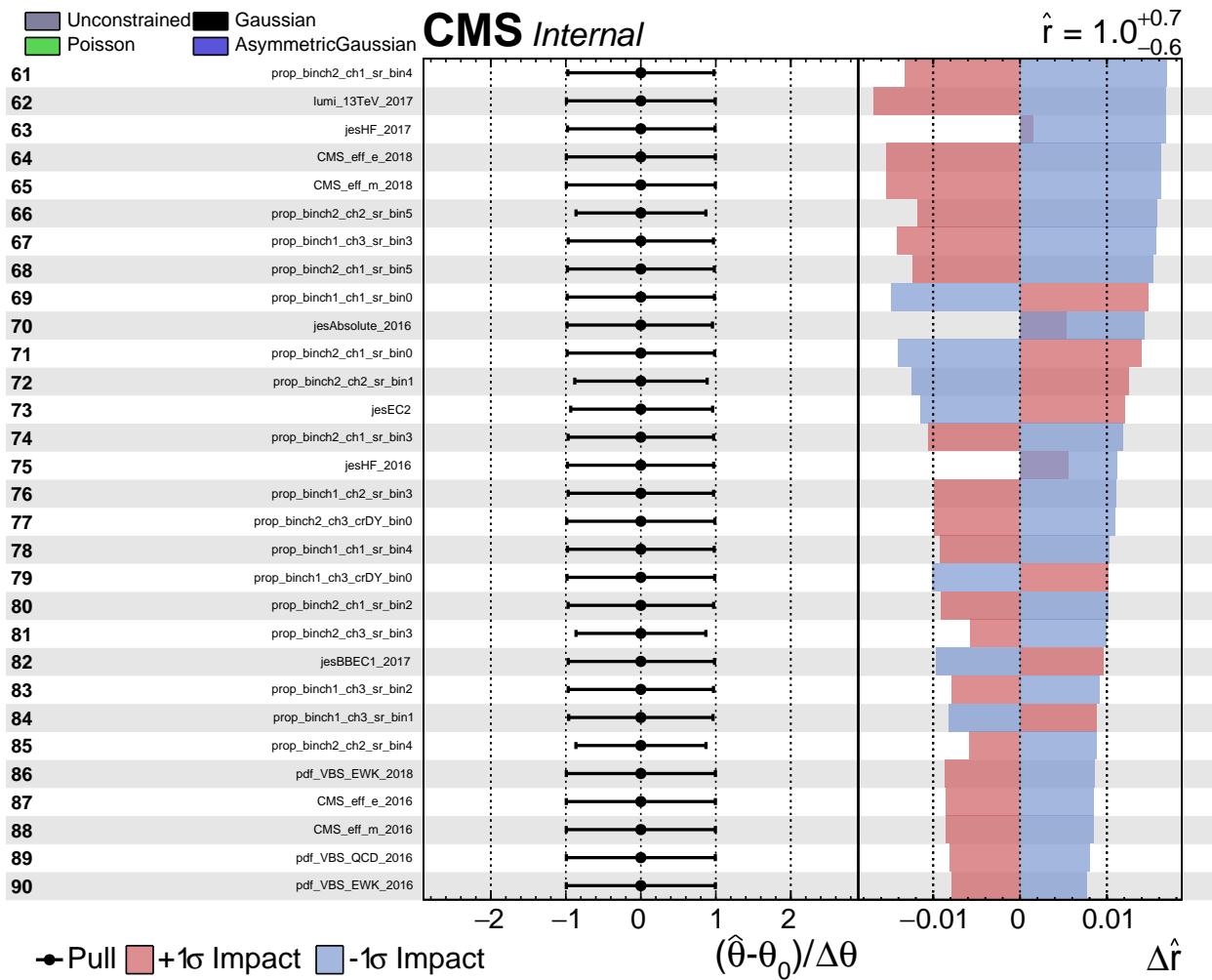


Figure B.1: Impact Plots

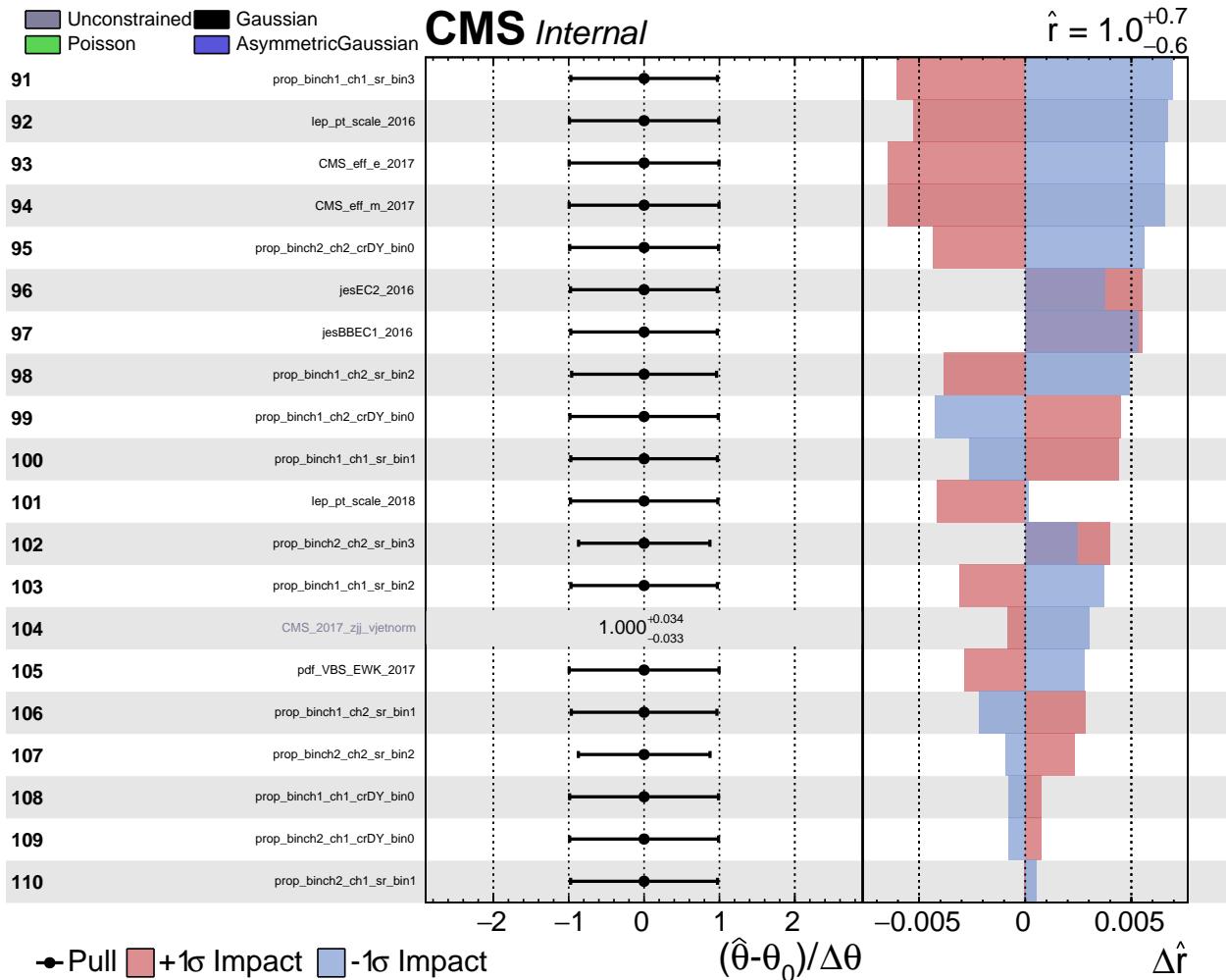


Figure B.2: Impact Plots