

**Search for Non-Resonant Higgs boson pair
production in the $HH \rightarrow b\bar{b}W^+W^-$ decay
channel in $p - p$ collisions using CMS data at
 $\sqrt{s} = 13$ TeV at the LHC**

By
Gourab Saha
PHYS05201604016

Saha Institute of Nuclear Physics, Kolkata

A thesis submitted to the
Board of Studies in Physical Sciences
In partial fulfillment of requirements
For the Degree of
DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE



April, 2023

Homi Bhabha National Institute¹

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Gourab Saha** entitled **Search for Non-Resonant Higgs boson pair production in the $HH \rightarrow b\bar{b}W^+W^-$ decay channel in $p - p$ collisions using CMS data at $\sqrt{s} = 13$ TeV at the LHC** and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor in Philosophy.

Subir Sarkar

Date: 31/03/2023

Chair - Prof. Subir Sarkar

Suchandra Dutta

Date: 31/03/2023

Guide / Convener - Prof. Suchandra Dutta

Subendu Rakshit

Date: 31.03.2023

Examiner - Prof. Subhendu Rakshit (IIT Indore)

Supratik Mukhopadhyay

Date: 31 Mar 23

Member 1 - Prof. Supratik Mukhopadhyay

Satyaki Bhattacharya

Date: 31/03/23

Member 2 - Prof. Satyaki Bhattacharya

Dilip Kumar Ghosh

Date: 31/3/2023

Member 3 - Prof. Dilip Kumar Ghosh (IACS, Kolkata)

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 31/03/2023

Guide: Suchandra Dutta

Place: KOLKATA

Prof. Suchandra Dutta

¹This page is to be included only for final submission after successful completion of viva voce.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.


Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.



Gourab Saha

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me.
The work is original and has not been submitted earlier as a whole or in part for a degree
/ diploma at this or any other Institution / University.


Gourab Saha

List of Publications arising from the thesis

Journal

1. "Searching for exotic Higgs bosons at the LHC", Bhattacharyya, G., Dwivedi, S., Ghosh, D. K., Saha, G., & Sarkar, S., Phys. Rev. D, 106(5), 2022, 055032. doi:10.1103/PhysRevD.106.055032 [Published]
2. "Test beam performance of a CBC3-based mini-module for the Phase-2 CMS Outer Tracker before and after neutron irradiation", The Tracker Group of the CMS Collaboration, (5 2022). ArXiv [Physics.Ins-Det]. Retrieved from <http://arxiv.org/abs/2205.00961> [Accepted to JINST]
3. "Searching for exotic Higgs bosons from top quark decays at the HL-LHC", Bhattacharyya, G., Chakraborty, I., Ghosh, D. K., Jha, T., & Saha, G., (12 2022), ArXiv [Hep-Ph]. Retrieved from <http://arxiv.org/abs/2212.09061> [Communicated to PRD]

Conferences

1. Saha, G. (2021). Performance of 2S Modules of the CMS Phase-2 Tracker in a Test Beam Environment. JPS Conf. Proc., 34, 010022. doi:10.7566/JPSCP.34.010022
2. The performance of 2S mini module of phase-2 outer tracker in a test beam analysis 23rd DAE-BRNS High Energy Physics Symposium 2018, IIT Madras
Oral presentation



Gourab Saha

DEDICATIONS

Dedicated to my Father ...

ACKNOWLEDGMENTS

First of all I would like to express my utmost gratitude to my supervisor Prof. Suchandra Dutta for her mentorship and for introducing me to the exciting world of high energy physics experiments. She has inspired me greatly with her passion, interest and knowledge about particle and detector physics.

I am grateful to Prof. Subir Sarkar for his constant guidance. I am extremely fortunate to have Prof. Sarkar in each and every step of learning programming languages and several other aspects of computation during my PhD tenure. He taught me the beauty of data analysis from the very beginning. The time I spent with Prof. Satyaki Bhattacharya has been a continuous lesson in work ethics, discipline, perseverance and humility. Without his ideas and lectures, statistics would have been merely a tool to me. I sincerely thank him for all the things I could learn from our uncountable discussion sessions. I am very thankful to our director Prof. Gautam Bhattacharyya, who helped me immensely in phenomenological studies of several exotic physics problems. I am also thankful to Prof. Dilip Kumar Ghosh for his constant guidance in my research.

I have had the good fortune to collaborate with all the fantastic folks at SINP, where we have shared ideas and acquired knowledge. I would like to express my gratitude to my friends (batch-mates, seniors and juniors) Atanu da, Sourav da, Suvankar da, Kuntal da, Shamik da, Arnab da, Ashim da, Arghya da, Bishe da, Rajarshi da, Rome da, Rupam da, Sukanya di, Wadut, Jhuma, Aritra, Debabrata, Pritam, Sweta, Anindita, Suman, Sayan, Aranya, Bithika and Promita for the wonderful environment we shared together. I hope some of the good friendships that we formed will last forever.

I am extremely thankful to Florian, Agni from the $b\bar{b}W^+W^-$ analysis team, and Pieter from CP3, Belgium. Their constant supports and hundreds of discussions helped me a lot to develop my Python programming and data analysis skills. I would also like to thank Christian, Alexandra, Karl, Saswati di, Dennis, Peter and Benjamin for their help in the

bbWW analysis. In the phenomenological studies, I was very fortunate to get in touch with some beautiful persons Siddharth, Indrani di, Nabarun da and Tapoja di who helped me in several aspects of my research. Also I would like to thank Mr. Pradip Dutta Sharma for being very helpful and considerate with administrative issues.

I would like to thank my wife Duhita for being a great friend and for supporting me in tough times.

Finally, I want to express my gratitude to my family, especially to my mother, sister, maternal uncles, aunts and my brother Puchai for their unwavering support and trust in me. I am grateful for every bit of their efforts and constant support, which allowed me to make whatever little addition I could contribute to this great body of knowledge.

Contents

Summary	xi
List of Figures	xiii
List of Tables	xvii
1 Introduction	1
2 The Theoretical Perspective	4
2.1 The Standard Model of Particle Physics	4
2.2 The Brout-Englert-Higgs Mechanism	8
2.3 Higgs Self-Interactions	10
2.3.1 HH Production Processes	11
2.3.2 HH production in the BSM scenario	13
3 The Experimental Apparatus	18
3.1 The Large Hadron Collider	18
3.2 The Compact Muon Solenoid Experiment	20
3.2.1 Magnet	21
3.2.2 Tracker	22
3.2.3 Electromagnetic Calorimeter	24
3.2.4 Hadron Calorimeter	26

3.2.5	Muon System	27
3.2.6	The CMS Trigger System and Datasets	29
4	Physics Object Reconstruction in CMS	31
4.1	Track Reconstruction	32
4.2	Beam Spot Reconstruction	35
4.3	Reconstruction of Primary Vertices	35
4.4	Electrons and Photons	37
4.5	Muon	40
4.6	Jets	41
4.6.1	b -jet	43
4.7	Tau	45
4.8	Missing Transverse Energy	46
4.9	Machine Learning	47
5	Search for Non-Resonant HH Production in $b\bar{b}W^+W^-$ Decay Mode	51
5.1	Datasets	52
5.2	Monte Carlo Simulation	53
5.2.1	Signal	55
5.2.2	Background	57
5.3	Event Selection	60
5.3.1	Physics Object Selection	60
5.3.2	Pre-selection Criteria	64
5.3.3	Multivariate Analysis	67
5.4	Background Estimation	74
5.5	Systematic Uncertainties	76
5.6	Statistical Analysis	77

5.6.1	Results	80
6	A Phenomenological Study: Search for Exotic Higgs Bosons at the LHC and beyond	86
6.1	Signal Processes and Choice of Benchmark Points	88
6.2	Backgrounds	91
6.3	Monte Carlo Simulation	91
6.4	A1: Search for exotic scalars produced in association with a Z boson at the LHC and beyond	92
6.5	A2: Searching for exotic Higgs bosons from top quark decays at the HL-LHC	97
7	Performance of 2S Modules of the CMS Phase-2 Tracker in a Test Beam Environment	105
7.1	The CMS Phase-2 Tracker	106
7.2	Concept of p_T Modules	108
7.3	Prototype 2S Modules with CMS Binary Chip (CBC)	109
7.4	Beam Test Setup	111
7.5	Data Acquisition (DAQ) and Track Reconstruction	112
7.6	Data analysis	114
7.7	Results and discussions	116
7.7.1	Signal and noise: Threshold scan	116
7.7.2	Tracking and Module Resolution	118
7.7.3	Uniformity of Efficiency	119
7.7.4	Performance of p_T Discrimination	119
8	Conclusion	122

Summary

The thesis covers a physics analysis with CMS data, two phenomenological studies and a CMS tracker beam test data analysis.

The physics analysis is a search for the production of pairs of Higgs bosons (HH), where one Higgs boson decays to a pair of bottom quarks and the other to two W bosons. The search is based on proton-proton collisions recorded at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1} . The analysis focuses on final states containing at least one leptonically decaying W boson. The expected upper limits on the cross section of HH production is estimated. Specifically, assuming the Standard Model kinematics, the limit on the inclusive cross section for the non-resonant HH production is expected to be 17 times the value predicted by the Standard Model, at a 95% confidence level. The limits are also presented as a function of various Higgs boson coupling modifiers and anomalous Higgs boson couplings through several shape benchmark points motivated by Effective Field Theory.

A class of flavor models based on the group S_3 can contain enlarged scalar sectors with non standard couplings to fermions and gauge bosons. Under the influence of such model, phenomenological analyses are performed to search for exotic scalars through their mass and purely off-diagonal couplings to the fermions. The results of two such analyses, (a) production of χ in association with Z from H and, (b) production of χ from $t\bar{t}$, are discussed in the thesis. A search for a heavy exotic scalar (H) and a comparatively a light pseudo-scalar (χ) with non-standard couplings to leptons and quarks is performed

at the present and upcoming phases of the LHC. For (a), it is observed that even with a 20% uncertainty on the backgrounds, a moderately large region of parameter space containing a pair of off-diagonal Yukawa couplings ($Y_{uc}^\chi, Y_{\mu\tau}^\chi$) can be explored at the HL-LHC. The second one (b) shows a large background contamination with lower signal yield and therefore, several Machine Learning based methods are used. However, with a 10% uncertainty for χ with low mass (~ 40 GeV), it is observed that a feasible range of χ_{ct} can be explored at the HL-LHC.

The thesis also covers the necessity for the installation of a new CMS tracker for the high luminosity phase of the LHC (HL-LHC). The tracking detector consists of two parts, the inner tracker and the outer tracker. The outer tracker is equipped with modules that contain two parallel silicon strip sensors and front-end ASICs. These modules can determine the transverse momentum of charged particles and reconstruct track bends by correlating hits in the two parallel sensors. The track information from high momentum particles will be sent to the Level-1 trigger system of CMS at the bunch crossing rate of the LHC *i.e.* 40 MHz. The front-end ASIC used in the outer tracker modules is the CMS Binary Chip (CBC). This thesis presents the results of a beam test campaign at the DESY beam test facility to study the performance of the modules equipped with non-irradiated and irradiated sensors, and the p_T discrimination logic of the CBCs.

List of Figures

2.1	<i>Particle content of the Standard Model.</i>	5
2.2	<i>The schematic representation of the potential $V(\phi)$ of the Higgs field ϕ for $\mu^2 < 0$.</i>	9
2.3	<i>Higgs (a) tri-linear and, (b) quartic self interactions.</i>	11
2.4	<i>HH production cross section [1] in p-p collisions for several production modes estimated using PDF4LHC15 parton density function. The bands represent the linear combination of the theoretical errors and PDF uncertainties.</i>	12
2.5	<i>Feynman diagrams for GGF HH production via (a) Higgs tri-linear self-interaction and, (b) Higgs-to-top Yukawa interaction.</i>	12
2.6	<i>Feynman diagrams of the HH production via vector boson fusion.</i>	13
2.7	<i>Variation of HH production cross section with κ_λ for different production mechanisms. The dashed and solid lines represent the LO and NLO predictions, respectively. The bands denote the linear addition of PDF and scale uncertainties.</i>	14
2.8	<i>Feynman diagrams at the leading order that contribute to the EFT GGF process. The blobs at the interaction vertices indicate the BSM couplings.</i>	15
2.9	<i>Branching fractions of all possible decay modes of HH assuming $m_H = 125 \text{ GeV}$.</i>	16
2.10	<i>The expected and observed upper limits on the signal strength i.e. the ratio of estimated HH production cross section to the expectation from the SM in different final states and combining all of them.</i>	16
3.1	<i>An aerial view of the LHC with four detectors.</i>	18
3.2	<i>A complete layout of the LHC complex.</i>	19
3.3	<i>A schematic view of the CMS coordinate system.</i>	21

3.4	<i>A schematic diagram of the CMS detector with the major sub-detectors.</i>	21
3.5	<i>The layout of one quarter of the Phase-1 CMS tracker in the $r - z$ plane.</i>	22
3.6	<i>Hit efficiency vs. instantaneous luminosity for CMS Phase-1 pixel detector measured using 2018 data.</i>	23
3.7	<i>A layout of the η coverage of the ECAL.</i>	25
3.8	<i>The energy resolution of the ECAL measured in beam test analysis.</i>	26
3.9	<i>An $r - z$ schematic representation of the four main components of HCAL, the hadron barrel (HB), hadron endcap (HE), hadron outer (HO), and hadron forward (HF) calorimeters, as they were located during the 2016 LHC run. Layers in yellow, green, and magenta are joined as depths 2, 3, and 4 interaction lengths, respectively, while layers in blue are gathered together as depth=1.</i>	27
3.10	<i>A layout of one quarter of the CMS detector with the components of Muon system: four DT chamber (MB1-MB4) in the barrel region, four CSC stations in the endcap (ME1-ME4) and RPC stations.</i>	28
3.11	<i>A schematic diagram of (a) two level trigger system of the CMS and, (b) level-1 trigger system with local, regional and global triggers.</i>	29
4.1	<i>A cross section view of the CMS detector describing different types of particle interaction with the detector material.</i>	32
4.2	<i>Variation of the p_T resolution of the tracks of simulated muons with η. The solid symbols depict the half-width containing 68% of the residual distribution around the mean value and represent real resolution. The open symbols illustrate the resolution if 90% of the residual distribution around the mean value is evaluated, as well as the influence of extreme values.</i>	34
4.3	<i>The impact parameter resolution of muon tracks for (a) transverse and, (b) longitudinal directions. The solid symbols reflect the half-width containing 68% of the residual distribution around the mean value and represent real resolution. The open symbols illustrate the resolution if 90% of the residual distribution around the mean value is evaluated, as well as the influence of extreme values.</i>	34
4.4	<i>The track reconstruction efficiency for muons coming from Z decay as a function of the number of reconstructed primary vertices.</i>	36
4.5	<i>A schematic diagram of the domino construction step of the Hybrid algorithm.</i>	38

4.6	<i>An example of a super-cluster algorithm is to collect all clusters that satisfy a given geometry condition.</i>	38
4.7	<i>Muon reconstruction and identification efficiencies for (a) LooseID and, (b) TightID, as a function of η estimated with 2016 data.</i>	42
4.8	<i>A schematic view of the origin of a b-quark from a secondary vertex.</i>	44
4.9	<i>Performance of various b tagging algorithms.</i>	45
4.10	<i>(a) τ_h identification efficiency, and (b) misidentification probability, for different WPs.</i>	46
4.11	<i>An ensemble of Decision Trees are used for prediction. Here the events of different classes are weighted using Gradient Boosting technique where the residual between the true and predicted output values are minimised to perform optimum classification. All the trees are trained sequentially and finally the combination of all of them is used for final prediction.</i>	48
4.12	<i>Diagram of a Convolutional Neural Network used to classify quark jets from gluon. Here the pictures of HCAL $\eta - \phi$ plane with energy deposition by jets are used as input. The probability of jet to be a quark or gluon like is considered as the output.</i>	49
5.1	<i>A schematic representation of the CMS simulation workflow.</i>	55
5.2	<i>An illustration of a two-stage neural network architecture equipped with the LBN and a subsequent DNN.</i>	70
5.3	<i>Confusion matrices showing the performance of the multi-class classification of the networks. The plots show the composition of each predicted label, under the given assumption that all input classes are of equal event weight sum.</i>	71
5.4	<i>ROCs for binary discrimination performance of $gg \rightarrow HH$ (SM) process against all background processes.</i>	71
5.5	<i>Distributions of the DNN responses for the signal and background categories with respect to the bin number for the SL channel for 2016 era only.. . . .</i>	74
5.6	<i>A schematic representation of the ABCD data-driven method. Three DY control regions are used to estimate the transfer weight factors and then it is applied in the signal region.</i>	75

5.7	<i>The algorithm for application of the fake factor on the non-prompt contribution computed from data and Monte Carlo simulation.</i>	76
5.8	<i>95% CL expected upper limit on the production cross section at $\kappa_\lambda = 1$ (SM case) coupling compared between all three years of DL, SL and the combination of DL and SL. The last row is the combination of all.</i>	81
5.9	<i>For the various benchmark points listed in Table 5.4 (JHEP04) and Table 5.5 (JHEP3), 95% CL expected limits on the cross-section given in fb are presented (right). The SM point gives an upper limit that is equivalent to the actual SM measurement.</i>	82
5.10	<i>The expected upper limit on the HH cross section at 95% confidence level as a function of κ_λ.</i>	82
5.11	<i>The expected upper limit on the VBF HH cross section at 95% confidence level as a function of κ_{2V}.</i>	83
5.12	<i>Likelihood scans as a function of κ_t, κ_λ and κ_{2V} in the two dimensional plane of these coupling modifiers. The best-fit value is presented as cross-bar, and the uncertainty at various levels on both POIs are displayed as contours.</i>	83
6.1	<i>Two major Feynman diagrams for the signal process A1 in (a) t-channel and (b) s-channel. The dashed vertical lines are indications of the on-shell production of H/χ in association with Z, and their subsequent decays into $\mu^\pm\tau^\mp$ and $\mu^+\mu^-/e^+e^-$, respectively.</i>	88
6.2	<i>The Feynman diagrams for signal events produced from $t\bar{t}$ pair, where one top quark decays to $b/\bar{b}W^\pm$ and the other one to χ and a light flavored jet. W can decay leptonically or to jets and χ decays to a pair of μ and τ.</i>	89
6.3	<i>Significance plots for the four mass benchmark configuration points at $\mathcal{L} = 300\text{ fb}^{-1}$ following the cut based analysis. The two plots in the upper panel are for (m_χ, m_H) in GeV : (a) (20, 160), (b) (20, 170). The lower panel contains the other two mass points, in GeV, (c) (60, 160) and (d) (60, 170).</i>	95
6.4	<i>$\Delta R(\mu', \tau_h)$ for m_H (in GeV) = (a) 160, (b) 250 with different choices of m_χ. . . .</i>	96
6.5	<i>Variation of significance with m_H for $m_\chi = 20, 60\text{ GeV}$ and $Y_\ell = Y_q$ (a) 0.003 (b) 0.005 and (c) 0.01 by scanning α in the range (0 – 0.2) for $\mathcal{L} = 1000\text{ fb}^{-1}$. For each colored shade, the outer and inner edges correspond to $\alpha = 0$ and 0.2, respectively.</i>	97

6.6	<i>ROCs of three MVA techniques for $m_\chi = 60$ GeV</i>	102
6.7	<i>Required integrated luminosity to achieve a 2σ exclusion and 5σ discovery. θ is the amount of uncertainties on the backgrounds.</i>	104
7.1	<i>Complete LHC/HL-LHC plan</i>	106
7.2	<i>One quarter of the CMS Phase-2 tracker layout in the $r - z$ view.</i>	107
7.3	<i>An illustration of the stub formation logic used for p_T discrimination in the 2S module of CMS Phase-2 Outer Tracker.</i>	109
7.4	<i>Sketch of a 2S module for the CMS Phase-2 Outer Tracker.</i>	110
7.5	<i>Comparator input, output and pulse shapes of two hit detection modes of the CBC.</i>	111
7.6	<i>A Schematic representation of a typical beam test setup.</i>	112
7.7	<i>Schematic diagram of the data acquisition scheme that consists of a common controller serves the dedicated data producers that ensures the synchronization between the telescope, FEI4 and the DUT.</i>	113
7.8	<i>Variation of reconstructed cluster efficiency with eight TDC phases for the DUT with non-irradiated sensors.</i>	115
7.9	<i>Variation of (a) cluster efficiency, and (b) stub efficiency with V_{CTH} for non-irradiated module.</i>	117
7.10	<i>Reconstructed signal shape for non-irradiated module.</i>	117
7.11	<i>The residual distributions for the two sensors of a non-irradiated 2S module prototype.</i>	118
7.12	<i>Variation of cluster efficiency with track x position, for both non-irradiated sensors of the DUT.</i>	120
7.13	<i>Variation of stub efficiency with (a) angle and (b) emulated p_T.</i>	121
8.1	<i>For the various benchmarks listed in Reference [2], generator-level distributions of the double Higgs invariant mass m_{HH}. The benchmark that was determined to be the most representative within the cluster is shown by the red line, while all the distributions of the identified cluster are represented by the blue lines.</i>	131

List of Tables

4.1	<i>Minimum BDT scores to identify electrons with two WPs.</i>	39
4.2	<i>Loose, Medium and Tight working points of the DeepJet b-tagging algorithm. . .</i>	44
5.1	<i>Branching ratios (BR) of several HH decay modes with DL, SL and FH final states.</i>	51
5.2	<i>The production cross section of the non-resonant GGF HH signal samples at NNLO precision.</i>	56
5.3	<i>The production cross sections of the non-resonant VBF HH signal samples. . . .</i>	56
5.4	<i>Twelve shape benchmarks (BP) [2] and additional one benchmark 8a [3]. . . .</i>	57
5.5	<i>Seven EFT benchmark points (BP) described in Reference [4].</i>	57
5.6	<i>The relevant SM backgrounds used in $HH \rightarrow b\bar{b}W^+W^-$ analysis.</i>	58
5.7	<i>List of the major backgrounds used in the analysis of data recorded in the 2016 data-taking period. The process name, sample name and the corresponding cross sections are tabulated.</i>	59
5.8	<i>Loose, fakeable and tight selection criteria for electrons. The requirement on the output of the BDT trained by the EGamma POG and on the observables $\sigma_{i\eta i\eta}$, H/E, and $1/E - 1/p$ are varied as function of η of the electron candidate. The conditions on the EGamma POG MVA, the Deep Jet discriminant and the relative isolation of the nearest jet to the electron are tightened (relaxed) for fakeable electrons that fail (pass) the requirement $\text{prompt-}e \text{ MVA} > 0.30$, in order to reduce the systematic uncertainty on the fake lepton background estimate on the jet flavor composition [5, 6]. A hyphen (–) indicates selection criteria that are not applied.</i>	62

5.9	<i>Loose, fakeable and tight selection criteria for muons. The conditions on the Deep Jet discriminant and the relative isolation of the nearest jet to the muon are tightened (relaxed) for fakeable muons that fail (pass) the requirement of prompt-μ $MVA > 0.5$. A hyphen (–) indicates selection criteria that are not applied.</i>	63
5.10	<i>Selection of Ak4 jets and Ak4 b-tagged jets.</i>	64
5.11	<i>Selection of Ak8 jets and the b-tagged ones.</i>	65
5.12	<i>Pre-selection conditions applied in the SL and DL channels.</i>	66
5.13	<i>Event categories after pre-selection. The first row describes the selection for the Resolved category with two specific categories: Resolved 1b, Resolved 2b and, the Boosted category is mentioned in the last row.</i>	66
5.14	<i>Events produced at $\mathcal{L} = 138 \text{ fb}^{-1}$ and the yields after pre-selection for the SM benchmark points of the $HH \rightarrow b\bar{b}W^+W^-$ GGF and VBF signal processes and major backgrounds.</i>	67
5.15	<i>Low level variables for SL and DL, and some important high level variables for SL only.</i>	69
5.16	<i>DNN output classes for DL channel.</i>	69
5.17	<i>Sub-categorizations for the merged processes in the SL and DL channels.</i>	72
5.18	<i>Several sources of systematic uncertainties and their impacts on the event yield. The descriptions are written in Appendix-1 6.5.</i>	77
5.19	<i>Summary of HH production measurements from the CMS and ATLAS collaborations in various decay channels with the integrated luminosity. Upper limits are presented as a multiple of the SM prediction for both the inclusive (mainly GGF) and VBF production processes. ATLAS Run-1 results were obtained by combining $bbbb$, $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, and $WW\gamma\gamma$ (at various integrated luminosity values), while the 2016 combination result (with a small amount of data from 2015) of ATLAS was obtained by combining the same channels with the addition of $bbWW$ and $WWWW$. CMS Run-1 results were obtained by merging $bbbb$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$, with the addition of $bbVV$ in 2016. The branching ratios used in each decay channel are taken from the SM alone. The Run-2 combination results are shown in the Run-2 categories combining the processes mentioned above it.</i>	84
6.1	<i>Analysis channels and final state particles.</i>	90

6.2	<i>Benchmark points chosen for A1 and A2.</i>	90
6.3	<i>Summary of acceptance cuts.</i>	92
6.4	<i>The signal and SM background effective cross sections (fb) after each successive baseline cut (C0-C6) and final event yields for $\mathcal{L} = 300 \text{ fb}^{-1}$ at 14 TeV LHC run. Signal event samples are generated for a few representative values of m_χ and m_H (in GeV) and for a range of Y_ℓ and Y_q. Signal cross sections are calculated at LO, $t\bar{t} + \text{jets}$ cross sections at $N^3\text{LO}$, while the other SM backgrounds are estimated at NLO.</i>	94
6.5	<i>Event yields at $\mathcal{L} = 3 \text{ ab}^{-1}$ after baseline selection for DL and SL channel.</i>	100
6.6	<i>Important input variables used for all of three MVA methods</i>	101
6.7	<i>Different BDTD parameters used for three different benchmark points.</i>	101
6.8	<i>Description of XGBoost parameters</i>	101
6.9	<i>Summary of DNN parameters</i>	102
7.1	<i>Some of the main parameters of PS and 2S modules of the CMS Phase-2 tracker.</i>	108
7.2	<i>Summary of the fit parameters shown in Figure 7.11.</i>	119
8.1	<i>List of HLTs used in the analysis. Checkmarks (✓) and hyphens (–) are added after each HLT path to indicate whether the path was enabled or disabled during the 2016, 2017, and 2018 data-taking periods.</i>	125
8.2	<i>List of JSON files used to implement the “good-run” selection in the 2016, 2017, and 2018 data-taking periods.</i>	126
8.3	<i>List of 2016 datasets. The /DoubleEG, /DoubleMuon, and /MuonEG datasets are used for the dilepton channel only.</i>	126
8.4	<i>List of 2017 datasets. The /DoubleEG, /DoubleMuon, and /MuonEG datasets are used for the dilepton channel only.</i>	127
8.5	<i>List of 2018 datasets. The /DoubleMuon and /MuonEG datasets are used for the dilepton channel only. The /EGamma datasets contains events passing the single-electron and double-electron triggers and replace the former /SingleElectron and /DoubleEG datasets.</i>	127

8.6	\cancel{E}_T filters applied to events selected in data and to simulated events. A hyphen (−) indicates that the filter is not applied.	130
-----	--	-----

Chapter 1

Introduction

The aim of particle physics is to study the properties and interaction of the basic building blocks of matter. A renormalizable quantum field theory of fundamental interactions, known as the Standard Model (SM) of Particle Physics, describes phenomena at the electroweak scale. The SM provides a unified understanding of the strong, weak, and electromagnetic forces. The model, which has been put to stringent tests in several collider and non-collider experiments over the past few decades, has shown to be an accurate representation of particle physics up to the TeV scale. The discovery of the Higgs boson by the ATLAS [7] and CMS [8] experiments at the LHC has completed the particle spectrum of the Standard Model and it has opened up a new and intriguing path for studying potential signs of novel physics at the TeV scale and beyond. The study of double Higgs production through Higgs self-interaction provides an unique opportunity to investigate the structure of the Higgs potential which will shed further light on the electroweak symmetry breaking mechanism.

Despite the remarkable agreement of the SM predictions with experimental results, several observations, namely, finite neutrino mass, matter-antimatter asymmetry etc. can not be explained by the SM which necessitated to go beyond the SM.

Various beyond Standard Model (BSM) physics can affect the HH production cross sec-

tion by including either new interactions or particles in the higher order quantum loops that may lead to deviations of the Higgs self-interaction and related coupling strengths from the SM predictions. In addition, the scalar sector of the SM can be extended by introducing a flavor symmetry to study the potential BSM effects.

The LHC machine is scheduled to undergo a substantial upgrade during 2026 – 28, in order to attain instantaneous luminosities that are 5-7 times higher than the current value of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. The LHC is expected to deliver $\sim 300 \text{ fb}^{-1}$ data per year to both ATLAS and CMS experiments during the High Luminosity phase of the LHC (HL-LHC). As a result of such a large increase in the instantaneous luminosity, the additional pile-up (PU) interactions will grow sharply from the existing value of 40 – 50 to ~ 200 which will adversely affect the trigger rates of each experiment. The CMS experiment will upgrade (Phase-2) several sub-detectors in order to operate efficiently in the extremely high radiation environment at the HL-LHC. A new tracker with improved radiation hardness, reduced material budget, higher granularity, longer trigger latency, etc., will replace the present one. To enable CMS to incorporate tracking information in the Level 1 (L1) trigger system, the new tracker will be equipped with detector modules that can distinguish high- p_T tracks from the low p_T ones at the detector front-end itself. The prototypes of such tracker modules are tested in the test beam facilities to study their performance at the high luminosity conditions.

The thesis is organized in eight chapters. An overview of the Standard model including quantum chromodynamics, electroweak theory, spontaneous symmetry breaking and Higgs mechanism, aspects of the production of double Higgs boson and its decay to $b\bar{b}W^+W^-$ are described in Chapter 2. A brief description of LHC, the CMS detector and the associated trigger system follows in Chapter 3. Chapter 4 describes physics object reconstructions in CMS. The $HH \rightarrow b\bar{b}W^+W^-$ analysis strategy and results are discussed in Chapter 5. A phenomenological study to search for exotic scalars through their flavor violating decays at the LHC is described in Chapter 6. The beam test data analysis strat-

egy and the corresponding performance studies for a prototype of strip-strip outer tracker module are discussed in Chapter 7. In Chapter 8, the major results presented in this thesis are summarized.

Chapter 2

The Theoretical Perspective

2.1 The Standard Model of Particle Physics

The Standard Model (SM) [9–11] is a renormalizable quantum field theory that describes strong, weak, and electromagnetic interactions in a single framework. The SM is based on the local gauge invariance of the Lagrangian under the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$, where C stands for color charge, L for left chiral fermions and Y for Hypercharge. The strong force is described by the invariance under the $SU(3)_C$ group which establishes the existence of massless gluons (g) with color charges as the mediator of strong force. The $SU(2)_L \times U(1)_Y$ symmetry unifies the weak and electromagnetic forces, which are mediated by spin-1 W/Z bosons and photon (γ), respectively. Figure 2.1 shows the SM particles in all three generations with their mass, charge and spin quantum number.

The QCD is based on the local gauge invariance under the $SU(3)_C$ group. The Lagrangian density (\mathcal{L}) of a massless spin- $\frac{1}{2}$ fermion is:

$$\mathcal{L} = \bar{\psi}(x) (i\gamma^\mu \partial_\mu) \psi(x) \quad (2.1)$$

where ψ is the fermionic field and the Dirac matrices are denoted by γ^μ . The fermionic



Figure 2.1: Particle content of the Standard Model.

field $\psi(x)$ is transformed under the operation of $SU(3)_C$ group as:

$$\psi(x) \sim e^{ig\frac{\lambda^a}{2}\theta_a(x)}\psi(x) \quad (2.2)$$

where, $\frac{\lambda^a}{2}$ ($a = 1 - 8$) are the 3×3 Gell-Mann matrices [12]. In order to preserve the Lagrangian density invariant under the transformation mentioned in Equation 2.2, the partial derivatives ($\partial_\mu\psi(x)$) must be reinterpreted as covariant derivatives as shown in equation 2.3. The covariant derivative includes the gauge vector fields ($A_\mu^a(x)$) that correspond to the eight gluons *i.e.* the mediators of strong force.

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - igA_\mu^a(x)\frac{\lambda^a}{2} \quad (2.3)$$

The total QCD Lagrangian density with a summation over all the quark fields becomes:

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu\partial_\mu)\psi - g\bar{\psi}\gamma^\mu\frac{\lambda^a}{2}\psi A_\mu^a - \frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a \quad (2.4)$$

The first term describes the free-field quark propagation, whereas the second term appears

from the incorporation of the covariant derivatives that explains quark-gluon interactions. The strength of this interaction is represented by “g” in Equation 2.4, which is usually expressed in terms of the strong coupling constant $\alpha_s = g^2/4\pi$. The third term is the vector field’s kinetic term. The introduction of gauge bosons (gluons) and their interaction with fermion fields (quarks) is a result of requiring Lagrangian invariance under a local gauge transformation.

The electroweak interaction can be explained by the same local gauge invariance mechanism under the $SU(2)_L \times U(1)_Y$ group. Three generators of the $SU(2)_L$ group correspond to three fields (W_μ^i) of the gauge bosons associated to the weak isospin quantum number (I_3). The $U(1)_Y$ group, associated to the weak Hypercharge Y , has a single generator and a single field (B_μ) arises after applying the gauge invariance. The relation between the electric charge and these quantum numbers is given by the Gell-Mann-Nishijima formula:

$$Q = I_3 + \frac{Y}{2} \quad (2.5)$$

The $SU(2)_L$ group provides doublets of left-chiral (Ψ_L) fermions and singlets of right-chiral (ψ_R) fermions. Under the $SU(2)_L \times U(1)_Y$ group, the leptons are represented as,

$$\begin{pmatrix} \ell_L \\ \nu_L \end{pmatrix} \quad (2.6)$$

ℓ_R

The SM excludes right-chiral neutrinos in order to avoid a neutrino mass term in the Lagrangian. The W bosons exclusively interact with the left-chiral fields since they are created from the first two generators of the $SU(2)_L$ group. The photon and Z boson are allowed to interact with both left and right-chiral fields.

In a similar manner to the leptons, the quarks can be represented under the $SU(2)_L \times U(1)_Y$

group as:

$$\begin{pmatrix} u_L \\ d'_L \end{pmatrix} \begin{pmatrix} c_L \\ s'_L \end{pmatrix} \begin{pmatrix} t_L \\ b'_L \end{pmatrix} \quad (2.7)$$

$$\begin{matrix} u_R & c_R & t_R \\ d'_R & s'_R & b'_R \end{matrix}$$

where, the left-chiral fields form doublets and the right-chiral ones form singlets under $SU(2)_L$. The weak interaction eigenstates typically do not overlap with the mass eigenstates of the quarks. The mass and weak interaction eigenstates for the quarks of the up sector can be considered to be identical without losing generality. The weak eigenstate of each *down* type quark is considered as a superposition of the mass eigenstates of all three *down* type quarks. The prime (\prime) designates the down type quarks as weak interaction eigenstates instead of mass eigenstates. Because of this, quarks from two distinct generations can interact through the W boson, a phenomenon known as the quark mixing that was first proposed by N. Cabibbo [13] and this phenomena leads to CP -violation as described by M. Kobayashi and T. Maskawa [14].

Using the notations mentioned above, the complete electroweak Lagrangian can be written as:

$$\mathcal{L} = i\bar{\Psi}_L \not{D} \Psi_L + i\bar{\psi}_R \not{D} \psi_R + i\bar{\psi}'_R \not{D} \psi'_R - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} - \frac{1}{4} W_i^{\mu\nu} W_{\mu\nu}^i \quad (2.8)$$

which includes charged and neutral currents as well as the free fermion Dirac Lagrangian. Any explicit mass term of the gauge fields would break the invariance. Mass terms for fermions are also not permitted because the left and right chiralities of the fields would result in a mass term $m\psi\bar{\psi} = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$, where ψ_L is a component of an $SU(2)_L$ doublet, would incorrectly mix singlets and doublets.

2.2 The Brout-Englert-Higgs Mechanism

Based on the description of the strong and electroweak interactions, any explicit mass term in the SM Lagrangian would violate the gauge invariance. So, the model hypothesizes all the fermions and gauge bosons to be massless which is in contradiction with the experimental observation. The existence of massive vector bosons and fermions require an extension of the SM and it was proposed in 1964 independently by Englert and Brout [15], and Higgs [16] using the concept of Spontaneous Symmetry Breaking (SSB). This is known as the Brout-Englert-Higgs (BEH) mechanism.

When considering a particular energy level in a system whose Lagrangian (\mathcal{L}) has a particular symmetry, two situations can occur: non-degenerate energy level with a unique eigenstate, or degenerate level whose eigenstates are not invariant, but transform linearly amongst themselves under the symmetry transformation of \mathcal{L} . When one of the degenerate states is randomly chosen as the ground state, it no longer shares the symmetries of \mathcal{L} . The asymmetry thus obtained is not due to the addition of a non-invariant asymmetric term to \mathcal{L} , but rather to the arbitrary selection of one out of the continuum of possible ground states.

The simplest approach to break the symmetry in the BEH mechanism is to add a complex scalar doublet that is invariant under translation and Lorentz transformations. An $SU(2)_L$ scalar doublet with complex field can be written as:

$$\Phi = \begin{pmatrix} \phi_a \\ \phi_b \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (2.9)$$

where, ϕ_i denotes the individual scalar field and the field Lagrangian is:

$$\mathcal{L}_{BEH} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \quad (2.10)$$

The second term in the Equation 2.10 represents the potential and it is defined as:

$$V(\Phi^\dagger \Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad \lambda > 0 \quad (2.11)$$

The minimization of the potential depends on the sign of μ^2 . Positive μ^2 makes $\phi_a = \phi_b = 0$ at the minimum of the potential *i.e.* no SSB happens. But when $\mu^2 < 0$, the minima of the potential appears at $|\phi_a|^2 + |\phi_b|^2 = v^2/2$, where $v = \sqrt{\mu^2/\lambda}$ is the vacuum expectation value (*vev*) and the shape of the potential can be demonstrated as shown in Figure 2.2. Any choice of a particular ground state is associated to the other states via a global phase

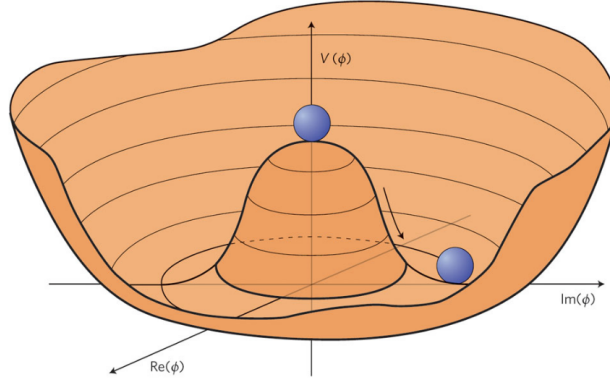


Figure 2.2: The schematic representation of the potential $V(\phi)$ of the Higgs field ϕ for $\mu^2 < 0$.

transformation. In order to break the symmetry, any specific value of the field can be considered without losing any generality. It can be proved that out of the four scalar fields, for a specific choice, only one physical field *i.e.* the Higgs field remains and the Higgs boson is the quantum of that field. The remaining three scalar fields correspond to the three mass-less degrees of freedom or the Goldstone bosons. They can be considered as the longitudinal polarizations of the gauge bosons Z and W , which, in turn, acquire mass after the SSB. The gluon and the photon remain massless because of the unbroken symmetry of $SU(3)_C$ and $U(1)_{em}$.

For simplicity, if $\phi_a = 0$ and $\phi_b = v/\sqrt{2}$ are chosen to achieve the minima of the Higgs potential, the field Φ can be expanded around the *vev* by a perturbation $H(x)$ *i.e.* a physical

Higgs boson.

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (2.12)$$

The BEH Lagrangian [17] shows that the Higgs field is associated to Higgs boson of mass $m_H^2 = 2\lambda v^2 = -2\mu^2$. It also establishes the source of mass of the weakly interacting bosons W and Z , along with their interactions with one or two Higgs *i.e.* HWW , HZZ , $HHWW$ and $HHZZ$. Finally it predicts the tri-linear and quartic self-interactions of the Higgs boson.

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda v H^3 + \frac{1}{4}\lambda H^4 - \frac{\lambda}{4}v^4 \quad (2.13)$$

Equation 2.13 shows the presence of two free parameters, the vev (v) and the mass of Higgs boson (m_H). The precise experimental measurement of the masses of W and Z , and the measurement of Fermi constant G_F extract the vev around 246 GeV. The Higgs boson self coupling, responsible for the mass generation of Higgs, is the final puzzle to completely understand the details of the nature of the Higgs potential.

The fermions get their masses from the Higgs field through Yukawa interaction which couples the left and right chiral fields. The Yukawa term, after the symmetry breaking, can be written as:

$$\mathcal{L} = - \sum m_f (\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L) \left(1 + \frac{H}{v}\right) \quad (2.14)$$

Here, \sum is over both *up* and *down* type fermions. The fermion masses (m_f) and their couplings to the Higgs boson (y_f) are related by $m_f = y_f \frac{v}{\sqrt{2}}$.

2.3 Higgs Self-Interactions

The study of the double Higgs (HH) production, predicted by both the SM and various BSM physics scenarios, is an effective way for a complete understanding of the Higgs sector. HH production processes involve tri-linear Higgs self-interaction, the strength of

which determines the shape of the Higgs potential.

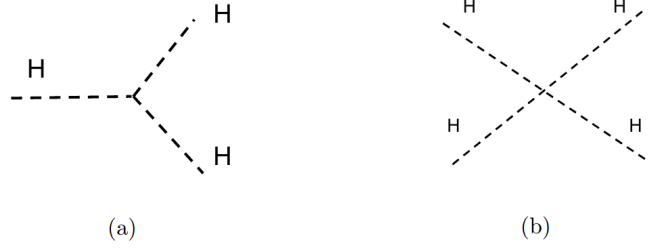


Figure 2.3: Higgs (a) tri-linear and, (b) quartic self interactions.

Figure 2.3 shows the Higgs tri-linear (λ_{HHH}) and quartic (λ_{HHHH}) self-couplings predicted by the SM. The Higgs self-couplings depend on m_H and the vev (v) such as: $\lambda_{HHH} = m_H^2/2v$ and $\lambda_{HHHH} = m_H^2/8v^2$. Both the couplings are predicted to be very small in the SM and therefore, probing the self-couplings at the LHC is extremely challenging. Measurement of λ_{HHH} ¹ is a major focus of the LHC. Any deviation from the self-interaction predicted by the SM would be a signature of new physics.

2.3.1 HH Production Processes

In proton-proton collisions, HH can be produced via five different mechanisms: gluon-gluon fusion (GGF), vector boson fusion (VBF), associated production with vector boson or top quark pair and single top associated production. The corresponding HH production cross sections with center of mass energy is shown in Figure 2.4 [18]. As can be seen from the figure, all the HH processes are quite rare at the LHC. Therefore, only the GGF and VBF production channels are presently being studied.

- The **gluon gluon fusion (GGF)** *i.e.* $gg \rightarrow HH$ proceeds via a loop of heavy quark (*e.g.* top) through Higgs-to-top (y_t) Yukawa interaction (Figure 2.5b) followed by a tri-linear Higgs self-interaction (Figure 2.5a). The two contributing diagrams to the GGF production mode as shown in Figure 2.5, have nearly equal strengths and they

¹For simplicity, the tri-linear coupling λ_{HHH} will be denoted by λ in the further discussions.

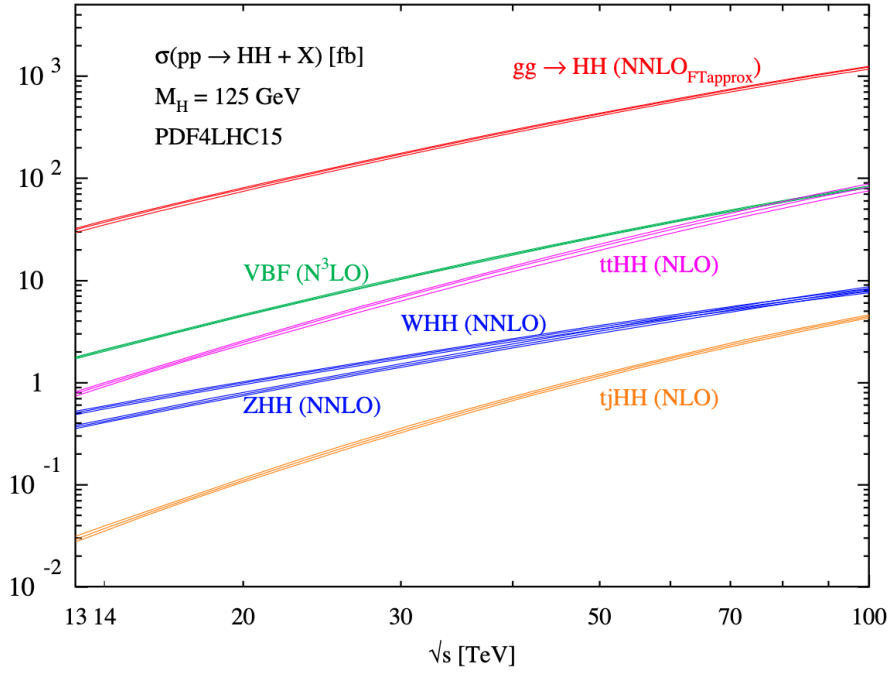


Figure 2.4: HH production cross section [1] in p - p collisions for several production modes estimated using PDF4LHC15 parton density function. The bands represent the linear combination of the theoretical errors and PDF uncertainties.

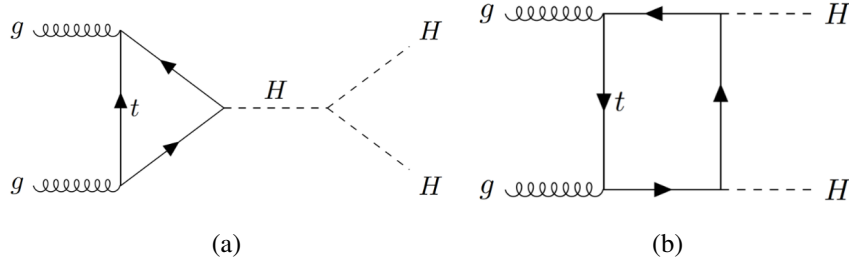


Figure 2.5: Feynman diagrams for GGF HH production via (a) Higgs tri-linear self-interaction and, (b) Higgs-to-top Yukawa interaction.

interfere destructively, making the HH analyses very challenging. The production cross section of HH in the GGF mode is $33.49^{+4.3\%}_{-6.0\%} (scale) \pm 5.9\% (theory) \text{ fb}$ [19] as measured with the LHC data collected during 2015-18 at $\sqrt{s} = 13 \text{ TeV}$.

- The **vector boson fusion (VBF)** i.e. $qq' \rightarrow HHjj$ involves the tri-linear self-coupling (Figure 2.6 left) and the VVH or $VVHH$ interaction (Figure 2.6 middle and right) where $V \equiv W, Z$. Although the cross section is twenty times smaller than the GGF, the VBF provides a distinct signature because of the presence of two

forward jets.

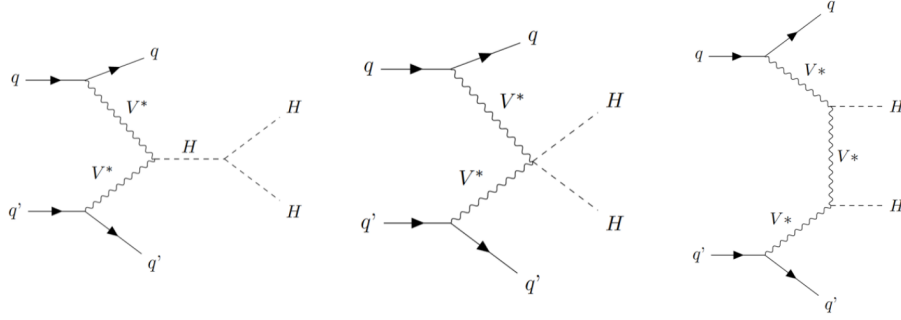


Figure 2.6: *Feynman diagrams of the HH production via vector boson fusion.*

2.3.2 HH production in the BSM scenario

Study of HH production can probe several BSM scenarios. In the SM, the contribution of the two GGF processes described above interfere destructively and make the HH final states sensitive to physics beyond the SM. Any variation of the Higgs couplings or presence of any heavy resonances decaying to HH can modify the amount of destructive interference which may result in an enhancement of the HH production cross section. In the resonant production, new resonances at the TeV scale can be directly observed in ATLAS and CMS. The presence of BSM physics can also introduce several anomalies in the Higgs couplings or new particles in quantum loops that can be inferred by the enhancement in the non-resonant HH production cross section and a significant modification to the event kinematics.

Non-resonant BSM HH production

As predicted by the SM, the Higgs tri-linear self-coupling (λ) can be determined entirely by m_H and v_{ev} . Several BSM approaches predict modification to the parameter λ which is usually quantified by the coupling modifier $\kappa_\lambda = \lambda/\lambda_{SM}$, where λ_{SM} is the SM prediction. In addition to the self-coupling, BSM scenarios can affect the Higgs-to-top Yukawa

interactions (y_t) with $\kappa_t = y_t/y_t^{SM}$, where y_t^{SM} is the SM value. Modification in VVH [20] and $VVHH$ interactions (through κ_{2V}) can be studied with the VBF processes. The variation of HH production cross section with κ_λ for different production mechanisms [18] is shown in Figure 2.7.

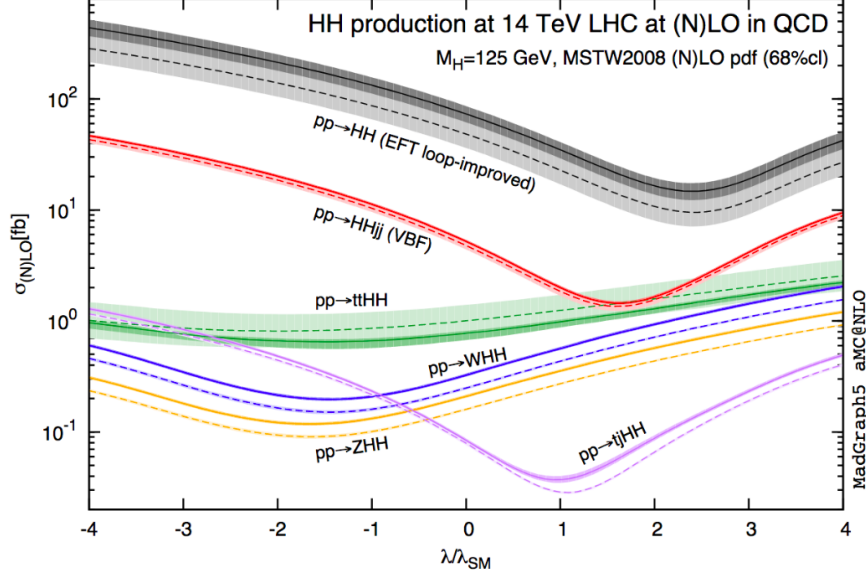


Figure 2.7: Variation of HH production cross section with κ_λ for different production mechanisms. The dashed and solid lines represent the LO and NLO predictions, respectively. The bands denote the linear addition of PDF and scale uncertainties.

The dependence of the HH production on all possible Higgs interactions can be generalized by the effective field theory (EFT) [21] approach. According to the EFT formulation, higher-dimensional operators modify the interactions between the Higgs boson and the other SM fields by introducing new couplings. The EFT Lagrangian can be written in terms of κ_λ and κ_t with three new BSM contact interaction vertices: $t\bar{t}HH$ interaction parameterized with c_2 , $ggHH$ with c_{2g} and ggH with c_g . The corresponding Feynman diagrams involving the five BSM interactions are shown in Figure 2.8.

The five dimensional EFT coupling parameter space creates a lot of combinations among different interactions. To maintain the feasibility of such studies, all the coupling values are clustered in several “shape benchmarks” [3, 4, 22] that cover the entire 5D parameter space. The shape benchmark points are computed with the help of the distributions of m_{HH} and $\cos(\theta^*)$, where θ^* is the polar angle of one of the Higgs of the HH pair at the

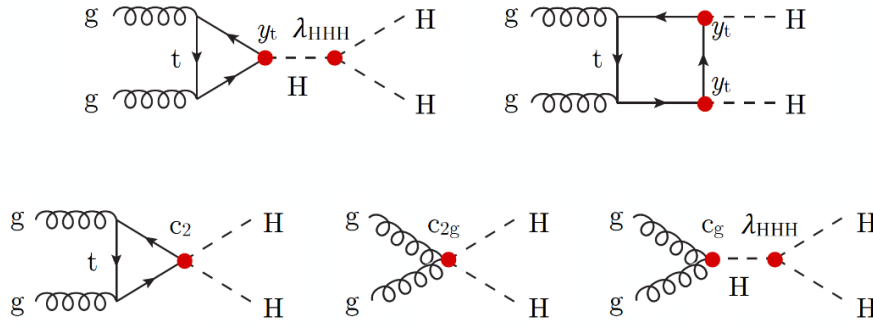


Figure 2.8: Feynman diagrams at the leading order that contribute to the EFT GGF process. The blobs at the interaction vertices indicate the BSM couplings.

center-of-mass frame. The dependence of twelve representative shape benchmarks [22] with m_{HH} is shown in Figure 8.1 in Appendix 6.5. The cancellation of the diagrams of the triangle and box types that contribute to the GGF at LO may be used to explain the variation. The variation of $\cos(\theta^*)$ is relatively flat because of the dominance of the s-wave diagram [23]. The variation of m_{HH} plays a prime role in the clusterization.

Experimental status on the non-resonant HH searches

The HH production cross section is about three orders of magnitude smaller than that of the single Higgs production. It is important to analyse all the decay modes of HH shown in Figure 2.9. However, considering the branching fractions and expected background contamination, $HH \rightarrow b\bar{b}\tau\tau$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$ are the most sensitive channels studied by CMS and ATLAS so far. Figure 2.10 shows the upper limit on the HH production cross section for several decay channels and their combination [24] with the full Run-2 ($\mathcal{L} = 138 \text{ fb}^{-1}$) data.

Analysis of one of the challenging decay modes of HH , namely, $b\bar{b}W^+W^-$ is presented in this thesis. Although $b\bar{b}W^+W^-$ has the second highest branching ratio, the corresponding analysis becomes difficult due to very high contribution from the SM background processes. Only the fully-leptonic and semi-leptonic decays of W boson pair are considered in the analysis. Fully hadronic decay channel of the same is not analyzed because of poor

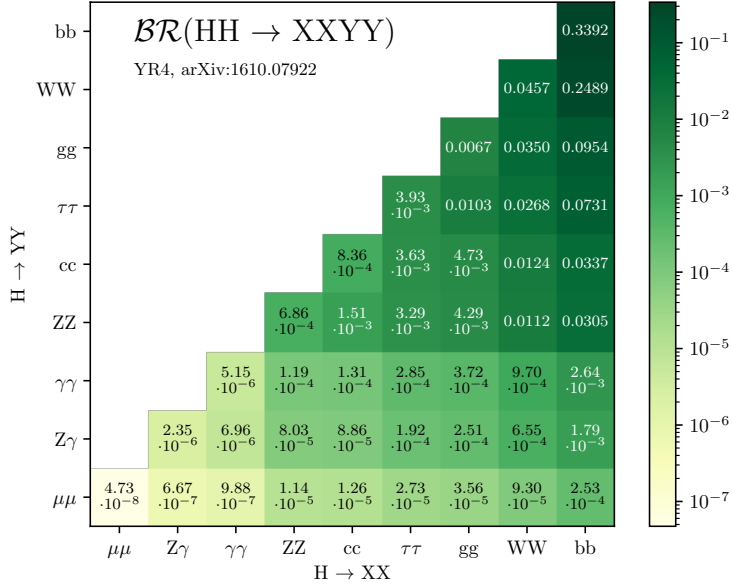


Figure 2.9: Branching fractions of all possible decay modes of HH assuming $m_H = 125 \text{ GeV}$.

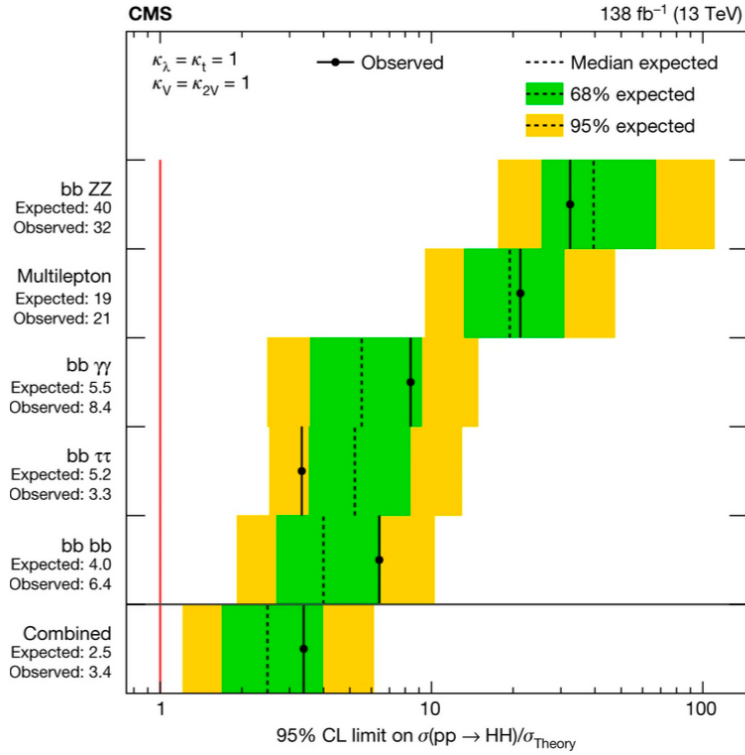


Figure 2.10: The expected and observed upper limits on the signal strength i.e. the ratio of estimated HH production cross section to the expectation from the SM in different final states and combining all of them.

jet energy resolution, low selection efficiency and extremely high background contamination. The analysis strategy and results are discussed in detail in [Chapter 5](#).

Chapter 3

The Experimental Apparatus

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) was built with a goal to find the Higgs boson and to probe new physics at multi TeV energy scale in proton-proton collisions. The collider is placed in a 26.7-kilometer-long underground tunnel in the France-Switzerland border, which was originally constructed for the Large Electron-Positron (LEP) collider. Figure

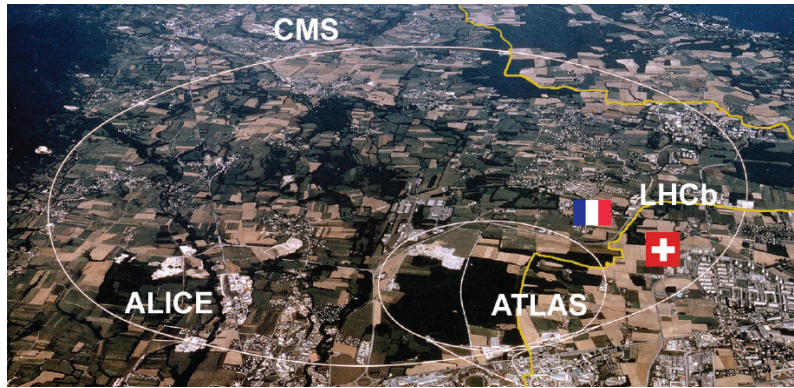


Figure 3.1: An aerial view of the LHC with four detectors.

3.1 shows the LHC and the four detectors namely ALICE [25], ATLAS [26], CMS [27], and LHCb [28] placed at four collision points. ATLAS and CMS are the two general-purpose experiments whereas, ALICE and LHCb are dedicated to heavy-ion physics, and

to explain matter-antimatter asymmetry, respectively. The LHC has been designed to collide proton bunches at an interval of 25 ns with an instantaneous peak luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$, resulting in about a billion interactions every second at CMS and ATLAS. The data used for the studies described in this thesis were collected by the CMS experiment at $\sqrt{s} = 13 \text{ TeV}$ in proton-proton collisions during 2016 – 18.

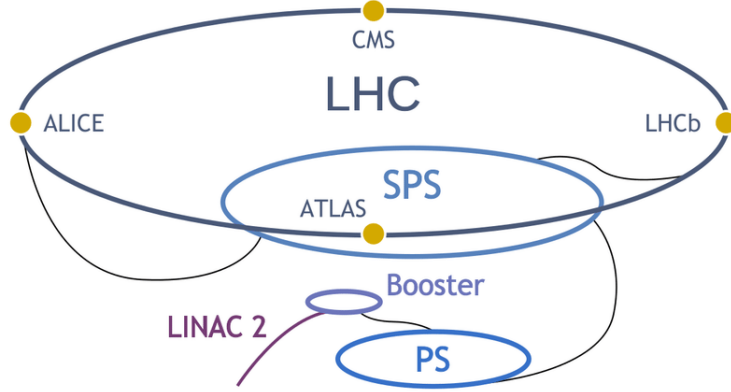


Figure 3.2: A complete layout of the LHC complex.

The protons are obtained by ionizing Hydrogen gas and then passed through a chain of acceleration stages, namely a linear accelerator (LINAC2), a proton synchrotron booster (PSB), a proton synchrotron (PS), and a super proton synchrotron (SPS) as shown in Figure 3.2. The particles are initially accelerated to 750 keV before being focused into a segmented beam by a radio frequency quadrupole. Then the LINAC2 accelerates them to 50 MeV before sending them to the PSB, which in turn accelerates the particles to 1.4 GeV in 530 ns. The protons are then delivered into the PS, where they are accelerated to 25 GeV and separated into bunches with uniform spacing. At design performance, 81 bunch packets are created with a gap of 25 ns or 8 m between them. After being accelerated to 430 GeV in 4.3 seconds at the SPS, the protons are finally transferred to the LHC ring.

3.2 The Compact Muon Solenoid Experiment

One of the two multipurpose detectors at the LHC is the Compact Muon Solenoid (CMS) detector. It is situated at the fifth interaction point of the LHC at 100 meters underground. The CMS detector has the following salient features:

- Excellent Muon identification efficiency and momentum resolution over a wide range of momenta and angles, as well as a very good di-muon mass resolution ($O(1\%)$) at 100 GeV.
- The silicon based tracking system has very good reconstruction efficiency and momentum resolution for charged particles. The inner tracker, made of pixel detectors, powers precise vertex reconstruction and offline tagging of b-jets and τ -jets.
- Electromagnetic calorimeter (ECAL) with an excellent electron and photon energy resolution, di-photon mass resolution $O(1\%)$ at 100 GeV, a wide geometric coverage, efficient electron and photon isolation, and ability to separate a π^0 from a photon.
- Hadron calorimeter (HCAL) with hermetic coverage and precise lateral segmentation for reliable estimation of di-jet mass and missing transverse energy.

The CMS detector distinguishes itself with a solenoid magnet that generates a strong axial magnetic field of 3.8 T.

CMS uses a right-handed co-ordinate system as shown in Figure 3.3. The p-p collision point is considered as the origin of the coordinate system of the CMS detector. The Z axis is along the direction of the proton beam (anti-clockwise), the X-axis is pointing radially inward towards the center of the LHC ring and the Y-axis is pointing vertically upward. The azimuthal angle ϕ is calculated in the X-Y plane, r and θ denote the radial distance in the same plane and the polar angle from the Z-axis, respectively. The polar angle is widely used in terms of pseudo-rapidity defined as $\eta = -\ln(\tan\frac{\theta}{2})$. Figure 3.4 shows a

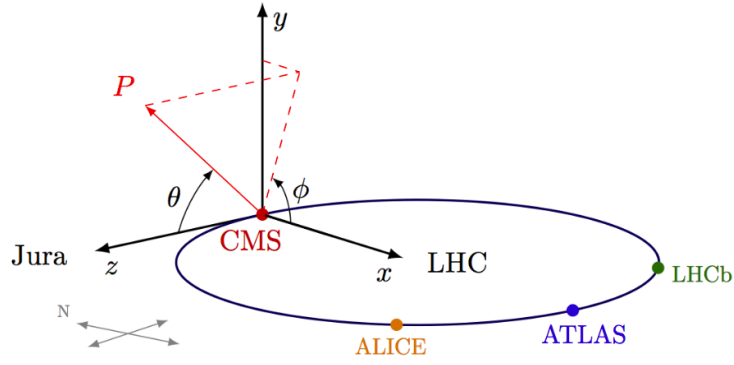


Figure 3.3: A schematic view of the CMS coordinate system.

schematic view of the CMS detector and how the sub-detectors are placed relative to each other. A detailed description of the CMS detector can be found in References [27, 29].

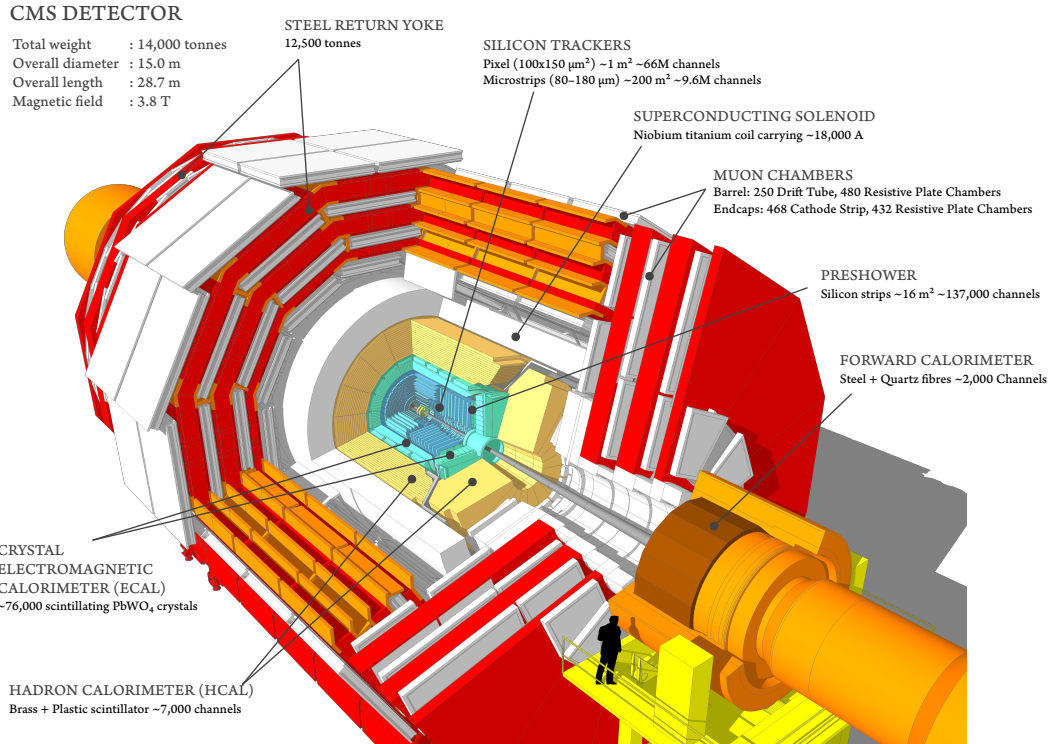


Figure 3.4: A schematic diagram of the CMS detector with the major sub-detectors.

3.2.1 Magnet

A 12.5 m long super-conducting solenoid [27] in the CMS detector generates a strong uniform magnetic field of 3.8 Tesla. The enclosure of the solenoid, which has an interior

diameter of 6 m, serves as a support structure for the sub-detectors. The tracker, electromagnetic calorimeter and hadron calorimeter are installed radially outwards from the beam pipe within the solenoid. A 1.8 m thick saturated iron yoke returns the magnetic field that enables the muon system placed outside the solenoid to reconstruct muons with high efficiency.

3.2.2 Tracker

The CMS tracking system is the closest sub-detector to the collision point. It is designed to provide precise and efficient measurement of the trajectories of the charged particles along-with efficient vertex reconstruction [4.3](#). The tracker consists of several cylindrical layers in the central region and discs in the forward region equipped with silicon pixel and strip detectors covering a region of $|\eta| < 2.5$ as shown in the layout of the $r - z$ plane in [Figure 3.5](#).

When a charge particle passes through the detector layers, it ionizes the detector active medium. The tracker registers hits at different layers and reconstructs charged particle tracks by combining the hits. The charged particles bend in the magnetic field depending on their momentum. The radius of curvature of the trajectory is used to calculate the momentum of the charged particles. Reconstructed tracks are used to determine the primary (collision point) and secondary (decay point of any long-lived particle) event vertices.

The track reconstruction algorithm used in CMS is briefly described in [4.1](#).

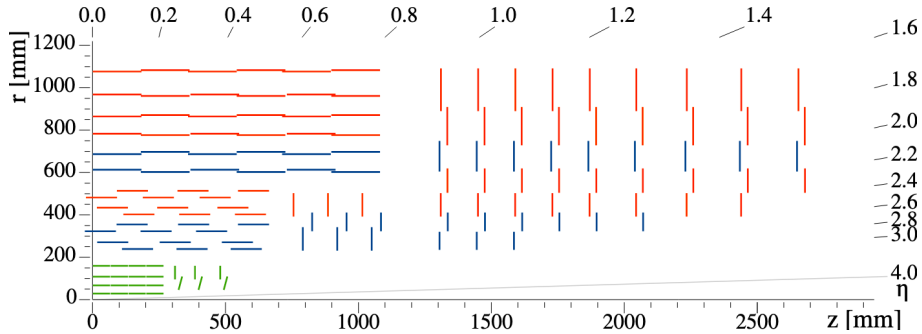


Figure 3.5: The layout of one quarter of the Phase-1 CMS tracker in the $r - z$ plane.

The inner tracker (IT) system (marked as green in Figure 3.5) is composed of 125 million square silicon pixels covering an active area of $\sim 2 \text{ m}^2$. Each pixel has a size of $100 \mu\text{m}$ in the $r - \phi$ and $150 \mu\text{m}$ in the z directions. The silicon sensors have a $285\text{-}300 \mu\text{m}$ thickness. Four layers (L1-L4) are located at radii of 2.9, 6.8, 10.9 and 16 cm in the barrel region (BPIX). On either side of the barrel region, the endcap (FPIX) discs (D1-D3) are placed at a distance of 29.1, 39.6 and 51.6 cm from the collision point. Each IT module is built with a sensor having 160×416 pixels connected to 16 read-out chips (ROC). The highly granular pixel detectors provide three dimensional measurement of hits needed for precise tracking and vertex reconstruction. Figure 3.6 shows the variation of hit efficiency [30] per BPIX layers and FPIX discs with instantaneous luminosity.

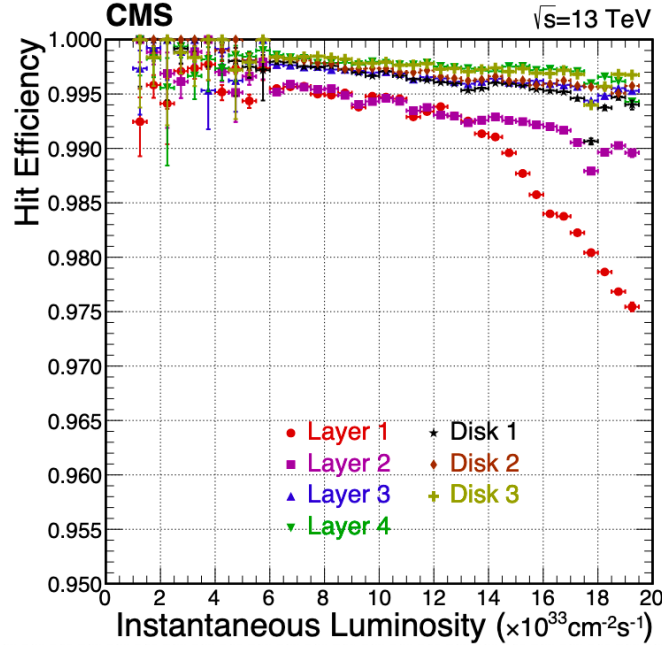


Figure 3.6: Hit efficiency vs. instantaneous luminosity for CMS Phase-1 pixel detector measured using 2018 data.

In order to maintain a good pattern recognition and optimum b-tagging performance, the prime requirement is to achieve a high hit-detection efficiency. Figure 3.6 shows the hit-detection efficiency is higher than 99% for all the layers and discs except L1 for instantaneous luminosity $> 1.4 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

The outer tracker (OT) consists of 9.6 million silicon strips surrounding the inner pixel

detectors. As shown in Figure 3.5, the OT is divided into four regions: four inner barrel layers (TIB), six outer barrel layers (TOB), three inner discs (TID) on either side of the TIB, and nine endcap discs (TEC) on each side of the barrel. The barrel layers provide coverage radially from 25 cm to 110 cm, and the endcap discs have a geometric coverage of 110 cm in radius and 280 cm in Z. The OT sensors for radial distance < 60 cm have a thickness of $320\text{ }\mu\text{m}$ and the ones with radial distance > 60 cm are of thickness $500\text{ }\mu\text{m}$. As shown in Figure 3.5, the blue lines represent the double sided stereo modules at the first two layers of TIB and TOB, TID rings 1 and 2 and TEC rings 1, 2 and 5. The stereo modules provide 3-dimensional position of the hits. These are built with two silicon sensors mounted back-to-back with a relative angle of 100 mrad to reduce the probability of having ghost hits. The pitch of the silicon strips varies from $80\text{-}250\text{ }\mu\text{m}$ depending on the radial position of the OT modules. At higher radial distances, as the particle flux density decreases, higher strip pitch is sufficient.

3.2.3 Electromagnetic Calorimeter

The ECAL is a homogeneous calorimeter made of PbWO_4 crystals with a geometric coverage up to $|\eta| < 3$. The ECAL measures the energy of photons and electrons from the electromagnetic showers formed by the particles in the detector volume. PbWO_4 crystals have high density (8.28 g.cm^3), small radiation length (0.89 cm) and Moliere radius (2.2 cm), which provide the compactness of the ECAL. Very fast response time (*i.e.* emission of 80% scintillation light in 25 ns) of the ECAL crystals allows it to operate efficiently in the high pile-up environment of the LHC. As shown in Figure 3.7, the ECAL consists of the central barrel (EB: $|\eta| < 1.479$) and endcaps (EE: $1.479 < |\eta| < 3.0$). To collect the scintillation light, avalanche photo-diodes (APD) and vacuum photo triodes (VPT) are used in the readout system of barrel and endcap, respectively. The internal radius of EB is 1.29 m . The EB is composed of 61200 crystals and is mechanically divided into 36 super-modules where each of them covers a 10 degree angular region. The crystals have

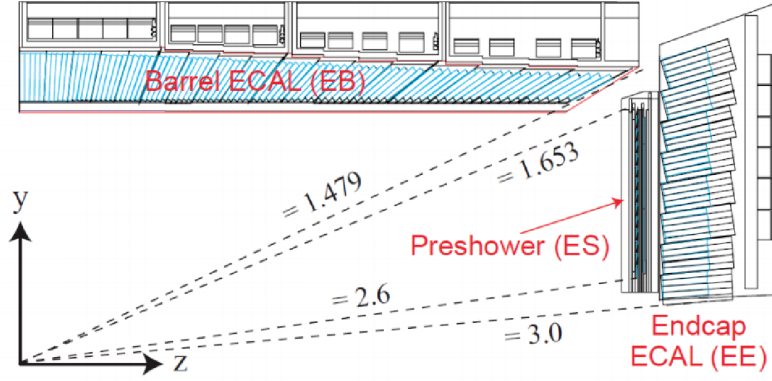


Figure 3.7: A layout of the η coverage of the ECAL.

a $2.2 \times 2.2 \text{ cm}^2$ frontal surface area and a length of 23 cm. On both sides of the EB, the EE sections are placed at 3.17 m distance from the nominal interaction point. The EE is split into four Dees, each covering 180 degrees in azimuth. The EE has 14648 tapered crystals placed in a rectangular $x - y$ grid and clustered in 5×5 mechanical structures known as super-crystals. An ECAL preshower (ES) detector is installed in front of the EE to discriminate between photon and neutral pion showers. The ES is a sampling calorimeter with two lead (Pb) planes followed by silicon sensors and covers $1.653 < |\eta| < 2.61$.

The energy resolution of the ECAL is,

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (3.1)$$

where, S , N and C represent the intrinsic stochastic term, the noise term and the constant term, respectively. The stochastic term represents the variations caused by shower formation in the crystal. The energy scale of physics interest for CMS is about 100 GeV, where the constant term dominates. The constant term is calculated using electron beams of different energies in dedicated beam test environments [31]. Figure 3.8 shows the energy resolution of the ECAL acquired in the beam tests. It demonstrates that the energy resolution $< 1\%$ can be reached at energies greater than 20 GeV while for energies $\mathcal{O}(100)$ GeV, the energy resolution turns out to be ~ 0.4 .

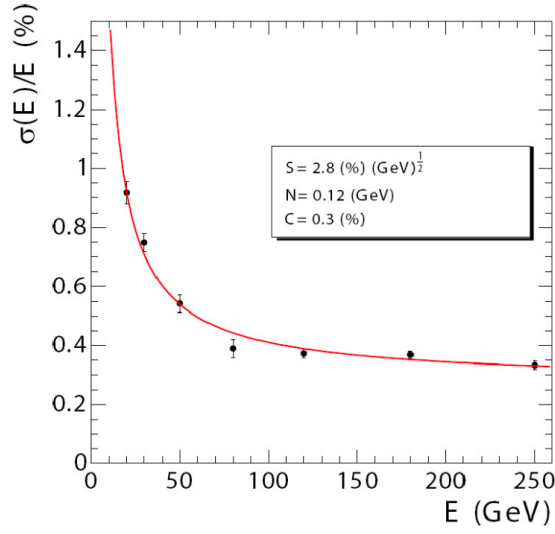


Figure 3.8: *The energy resolution of the ECAL measured in beam test analysis.*

3.2.4 Hadron Calorimeter

The HCAL is a sampling calorimeter made of brass as the absorber and plastic scintillator as the scintillating material. Brass is used because it is structurally robust and non-magnetic, whereas plastic scintillator is chosen for its long-term stability and radiation hardness. Figure 3.9 shows the layout of the HCAL with the individual parts, including the hadron barrel (HB), hadron endcap (HE), hadron forward (HF), and hadron outer (HO).

The HB is divided into 32×72 towers and has an overall coverage of $|\eta| < 1.3$. Each tower has a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. There are 17 layers of plastic scintillator in the HB, which serve as the active material. The region $1.3 < |\eta| < 3.0$ is covered by the hadron calorimeter endcap (HE). Brass plates with 79 mm thickness are used as absorber in the HE.

Outside the solenoid, two more sub-detectors of the HCAL *i.e.* hadron forward (HF) and hadron outer (HO) are placed to accommodate the full containment of the hadronic showers. The pseudo-rapidity region $3.0 < |\eta| < 5.0$ is covered by the HF. The HF is 11.2 meters away from the nominal interaction point of the CMS, and at such a distance,

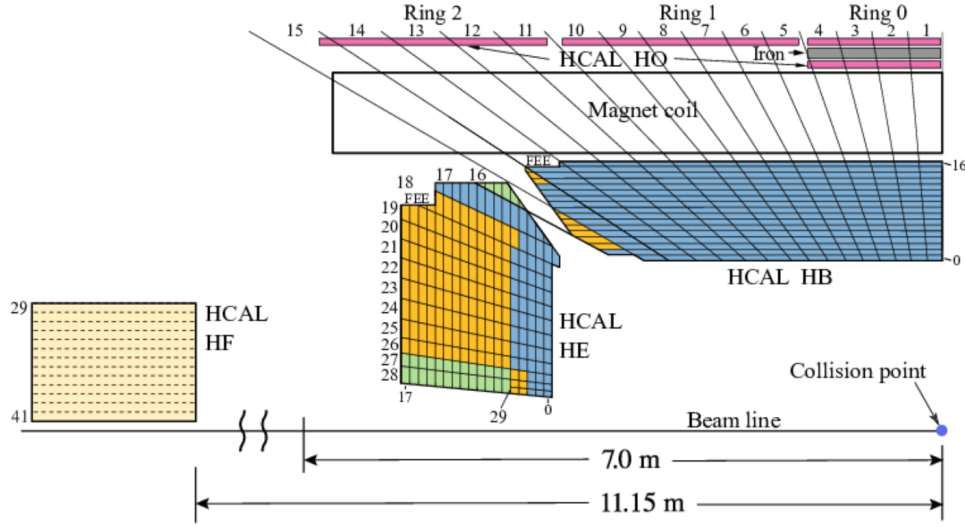


Figure 3.9: An $r - z$ schematic representation of the four main components of HCAL, the hadron barrel (HB), hadron endcap (HE), hadron outer (HO), and hadron forward (HF) calorimeters, as they were located during the 2016 LHC run. Layers in yellow, green, and magenta are joined as depths 2, 3, and 4 interaction lengths, respectively, while layers in blue are gathered together as $\text{depth}=1$.

the absorber depth is considered to be 1.65 m. In order to have sufficient sampling depth in the center of the barrel, the HO is placed between the magnetic coil and muon system. The HO is composed of five rings, with two scintillator layers on either side of an iron plate in the center of the rings. Only one scintillator layer is present in the other rings. The HO works as a tail catcher using the solenoid as an extra layer of absorber.

3.2.5 Muon System

The outer-most sub-detector of CMS is the Muon spectrometer which is placed outside of the magnet, interleaved with iron plates used to generate the return yoke. The iron plates accumulate the magnetic field lines that generate a strong magnetic field of ~ 2 T. A central barrel and two endcap regions make up the Muon system. The Muon system consists of three different types of gaseous detectors: drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC). Figure 3.10 displays a longitudinal view of one quarter of the Muon system.

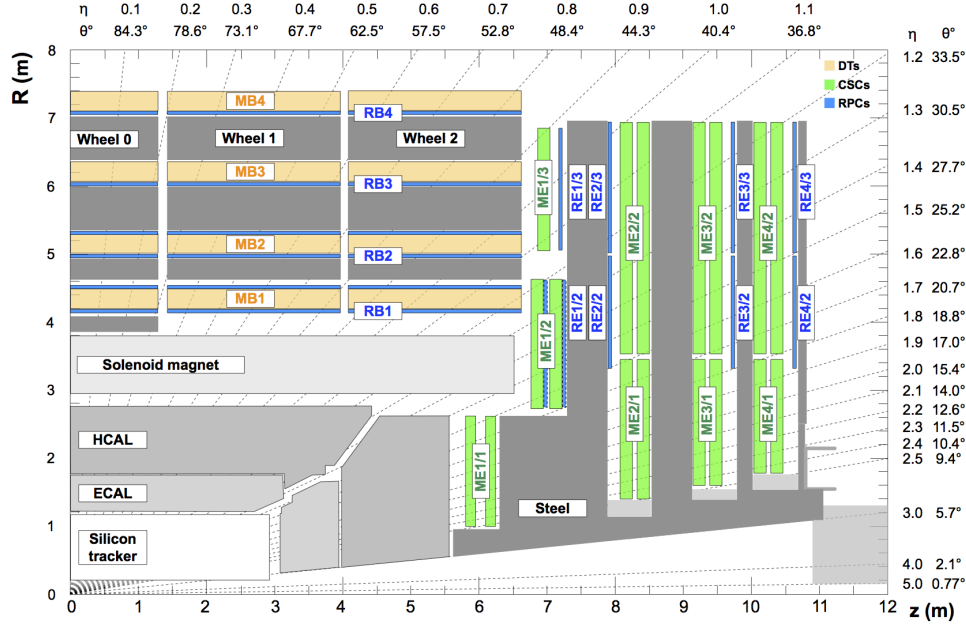


Figure 3.10: A layout of one quarter of the CMS detector with the components of Muon system: four DT chamber (MB1-MB4) in the barrel region, four CSC stations in the endcap (ME1-ME4) and RPC stations.

The central barrel detector is made up of drift tubes (DT) that cover up to $|\eta| < 1.2$. In this region, the expected muon flux and strength of the local magnetic field are minimal. The barrel area is made up of four DT chamber stations (MB1, MB2, MB3, and MB4) in the form of coaxial cylinders around the beam axis. Each station is composed of four DT layers called super-layers (SL). There are three SLs in the first three stations (MB1, MB2, and MB3), two of which measure the $r - \phi$ co-ordinate and the third the $r - z$ co-ordinate. In the last station (MB4), there are two SLs that measure the $r - \phi$ co-ordinate.

The endcap comprises cathode strip chambers (CSC) that span the range $0.9 < |\eta| < 2.4$. CSCs were chosen for the endcap because they perform well in strong magnetic field and with high particle flux. A CSC has fast response time. To ensure excellent spatial resolution, each endcap is made of four finely divided stations. Resistive plate chambers (RPC) with very fast response time are used in both barrel and endcap covering $|\eta| < 2.1$. The RPCs are designed to provide muon track timing information to the muon trigger system. Two RPCs are installed per DT in the barrel, and one RPC is installed per CSC in the endcap. A brief description of muon reconstruction and respective efficiencies are

given in Section 4.5.

3.2.6 The CMS Trigger System and Datasets

The proton beams collide at every 25 ns at the LHC. Majority of these collisions are soft in nature and uninteresting for further analysis. As the cross section of the soft interactions are much higher than the collisions with hard scattering, the goal of a trigger system is to reduce the selection rate of the soft events keeping the rate of the interesting physics events to the maximum extent possible.

The CMS trigger system is divided into two levels: Level-1 (L1) and High Level Trigger (HLT) as shown in Figure 3.11a. The L1 is largely made up of programmable electronics. At the L1 system, the digitized data with low resolution from the calorimeters and the muon system are used. The L1 decision is designed to be completed in $\sim 1 \mu s$ and the front-end electronics retain event information for $3 \mu s$, which corresponds roughly to data from 128 bunch crossings (BX). The L1 reduces the event rate from 40 MHz to 100 KHz. Following an L1 accept, data from all the sub-detectors are read out with full resolution for further processing at the HLT.

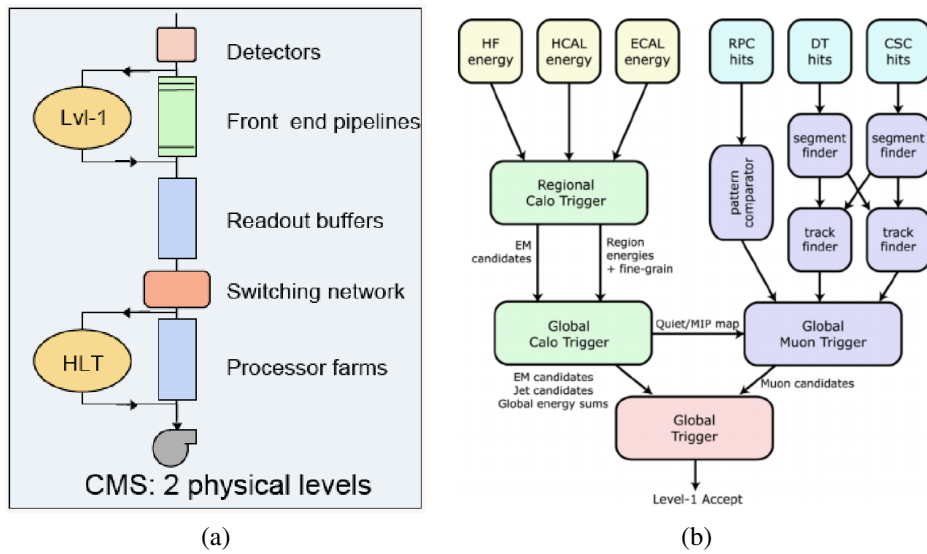


Figure 3.11: A schematic diagram of (a) two level trigger system of the CMS and, (b) level-1 trigger system with local, regional and global triggers.

The HLT is a software system that runs on commercial computer processors and reconstructs physics objects such as electrons, muons, taus, jets, photons and missing transverse energy. The reconstruction algorithms used in the HLT are the same as the ones used for offline event reconstruction. The HLT takes $\sim 10\mu s$ to process an event and reduces the data rate further to 100 Hz.

The collision events selected by the HLT algorithms are used for physics analyses. Physics objects with several p_T threshold and quality cuts act as the trigger objects at the HLT. The collision events are stored in several primary datasets based on the property of the objects available at the HLT. For example, `Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8` is a HLT path that requires the leading muon with $p_T > 17$ GeV and the sub-leading one with $p_T > 8$ GeV. Both of the muons require to pass a very very loose working point of the tracker isolation condition. The last two conditions are on the impact parameter at z direction and a minimum invariant mass cut of 8 GeV of the two muons, respectively. This is one of the numerous double-muon HLT paths that registers an event into the `DoubleMuon` primary dataset. Similarly, collision events are classified into other primary datasets namely, `SingleMuon`, `SingleElectron` and `DoubleEGamma` (EGamma for 2018), `MuonEG` etc.

Chapter 4

Physics Object Reconstruction in CMS

The physics objects are reconstructed by combining information from all the sub-detectors of CMS. Charged particles bend inside the detector volume by the 3.8 T magnetic field and the p_T and sign of the charge of the particles can be measured from the curvature. Electrons and photons deposit most of their energy through electromagnetic interaction inside the ECAL, while the hadrons deposit their energy through nuclear interaction in the HCAL. Muons can be detected from their signatures in the tracking and the muon systems. Neutrinos escape detection and cause missing transverse energy. A schematic of a transverse slice of the CMS detector with the interaction of various particles is shown in Figure 4.1. The Particle Flow algorithm [32] is used in CMS that correlates information from all the sub-detectors and exploits the excellent energy and spatial resolution of the detector to reconstruct high level objects.

In this chapter, the reconstruction of tracks, beam spot, primary vertices and all the physics objects namely, electrons (e), photons (γ), muons (μ), tau (τ) leptons, jets and missing transverse energy (E_T^{miss}) are discussed.

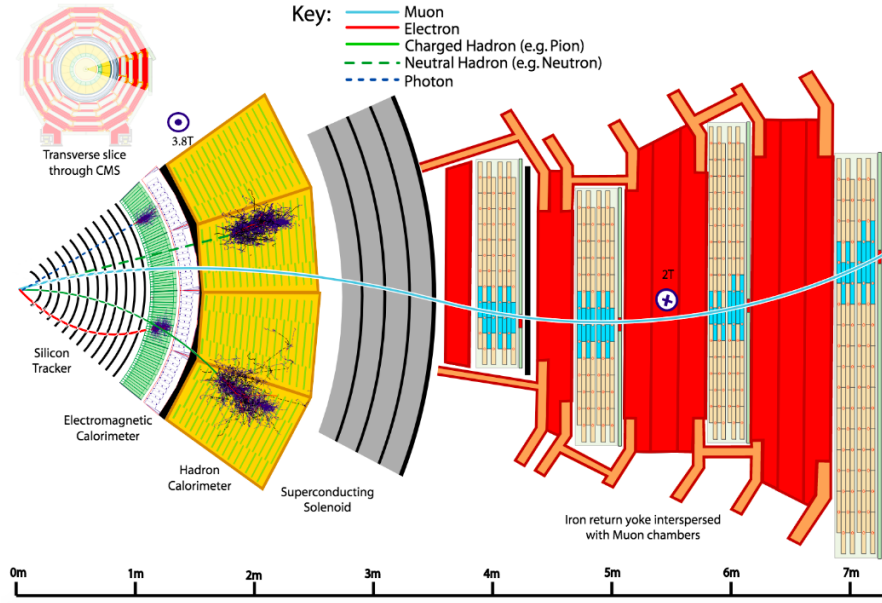


Figure 4.1: A cross section view of the CMS detector describing different types of particle interaction with the detector material.

4.1 Track Reconstruction

When a charged particle passes through the tracker, it produces “hits” in several tracker layers along its trajectory. A “Kalman Filter”[33] based combinatorial track finder (CTF) algorithm is used in an iterative way [34] to reconstruct the tracks through the following stages.

- The **Seed Generation** stage provides track candidates associating a few (2 or 3) hits from the inner pixel detector layers compatible with the trajectory of a charged particle. The seed generation step uses the information of beam spot, described in Section 4.2, and primary vertices (Section 4.3) that are reconstructed with the information from pixel detector only [34]. After seeding, a tracking region is specified in terms of several acceptable track parameters, namely p_T threshold, maximum closest approach to the reconstructed beam spot or pixel vertex, etc.
- The **Track Finding** stage extrapolates the seed trajectories to hits from successive tracker layers along the tracking region. The information required at each layer

includes the hit position with uncertainty and the amount of material crossed by a particle to estimate the effect of multiple Coulomb scattering and energy loss. This stage collects the hits and estimates the track parameters irrespective of the availability of the full information of the entire particle trajectory.

- The **Track Fitting** stage tries to find the best possible estimate of the parameters of each trajectory. Constraints like input of coarse beam spot region at the seeding stage include bias to the extrapolated trajectories. The tracks are re-fitted using a Kalman Filter and smoother. A Runge-Kutta propagator is used to include the effect of inhomogeneous magnetic field and material.
- The **Track Selection** stage uses several quality cuts on the fitted tracks. The track-finding process considers a large number of fake tracks [34] *i.e.* reconstructed tracks not associated with a charged particle, that can be reduced substantially by imposing the quality conditions like minimum number of layers with one hit, good χ^2/dof value, compatibility with the reconstructed primary vertices (PV) (Section 4.3), etc.

All of the above stages are executed in each of the six iteration of the iterative tracking. The basic idea is that the initial iterations find tracks with relatively large p_T and closest to the interaction region. After each iteration, the hits associated with tracks already found are removed. It reduces the combinatorial complexity and simplifies subsequent iterations to search for more tracks. Tracks found in each iteration are merged to build the final collection of reconstructed tracks.

Figure 4.2 shows the dependence of track- p_T resolution as a function of η for single muon Monte Carlo events. Over the entire η range, 1-10 GeV muons show resolution within 1-4%. In Figure 4.3, it can be seen that the track impact parameter resolution in the transverse and longitudinal directions are $\sim 30 - 100 \mu\text{m}$ and $\sim 10 - 20 \mu\text{m}$, respectively for muons with $p_T = 100 \text{ GeV}$.

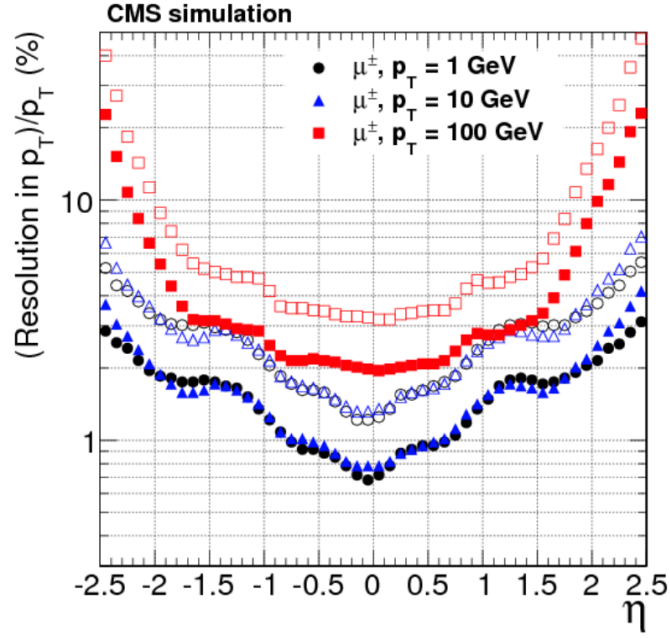


Figure 4.2: Variation of the p_T resolution of the tracks of simulated muons with η . The solid symbols depict the half-width containing 68% of the residual distribution around the mean value and represent real resolution. The open symbols illustrate the resolution if 90% of the residual distribution around the mean value is evaluated, as well as the influence of extreme values.

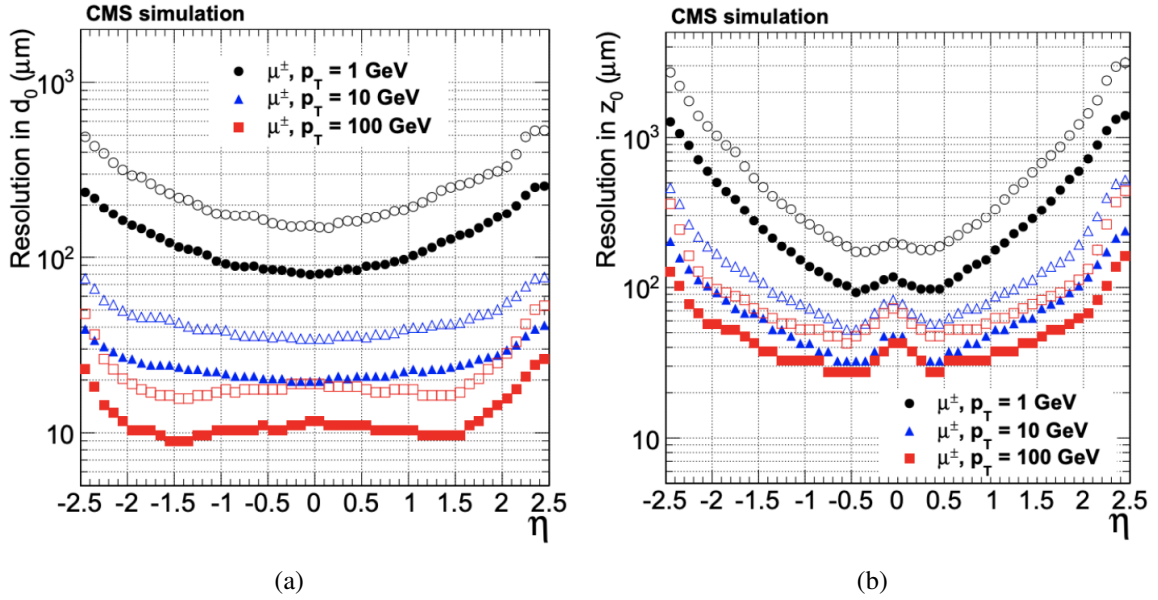


Figure 4.3: The impact parameter resolution of muon tracks for (a) transverse and, (b) longitudinal directions. The solid symbols reflect the half-width containing 68% of the residual distribution around the mean value and represent real resolution. The open symbols illustrate the resolution if 90% of the residual distribution around the mean value is evaluated, as well as the influence of extreme values.

4.2 Beam Spot Reconstruction

The beam spot is a 3-dimensional profile of the proton-proton collision region inside the CMS detector. The width of the proton beams in the transverse direction at the CMS interaction point is expected to be $\sim 16\mu\text{m}$. The position and spread of the luminous region or the beam spot provides a constraint on the interaction region which is crucial for primary vertex reconstruction. The beam spot information is used in the HLT to estimate the position of the interaction point prior to the primary vertex reconstruction.

The algorithm used to estimate the position and size of the beam spot profile in the transverse plane is the $d_0 - \phi$ algorithm [35], where d_0 is the three dimensional impact parameter of the tracks defined as the perpendicular distance between a reconstructed track and reconstructed primary vertex. The beam spot is reconstructed by averaging over all the luminosity sections (LS) *i.e.* the events collected during a period of 23 seconds. It is a robust and fast χ^2 fit that requires a sample of ~ 1000 tracks to provide a precision at the micron level. A detailed description can be found in [34].

4.3 Reconstruction of Primary Vertices

The reconstruction of the position and associated uncertainty of all the proton-proton interaction vertices, known as primary vertices (PV) in each event is crucial for physics performance. PVs are reconstructed using all the reconstructed tracks in an event. The reconstruction consists of three steps: track selection, clustering of tracks that appear to originate from the same vertex and, fitting for the position of vertices.

In the track selection step, the following conditions are satisfied:

- Significance of the transverse impact parameter relative to the center of the beam spot < 5 .

- Number of hits in strip and pixel layers traversed by a track: pixel ≥ 2 , pixel + strip ≥ 5 .
- Normalized χ^2 of the fit of the trajectory of a particle < 20 .

The selected tracks are then grouped according to their z -coordinates at the point where they are closest to the center of the beam spot. The proton-proton interaction vertices in a single LHC bunch crossing are reconstructed by the deterministic annealing (DA) algorithm [36] and the fitting is performed by an adaptive vertex fitter technique [37]. The PV having the highest $\sum p_T^2$ of the associated tracks, is considered as the event vertex. The details of the reconstruction of PV is discussed in Reference [34].

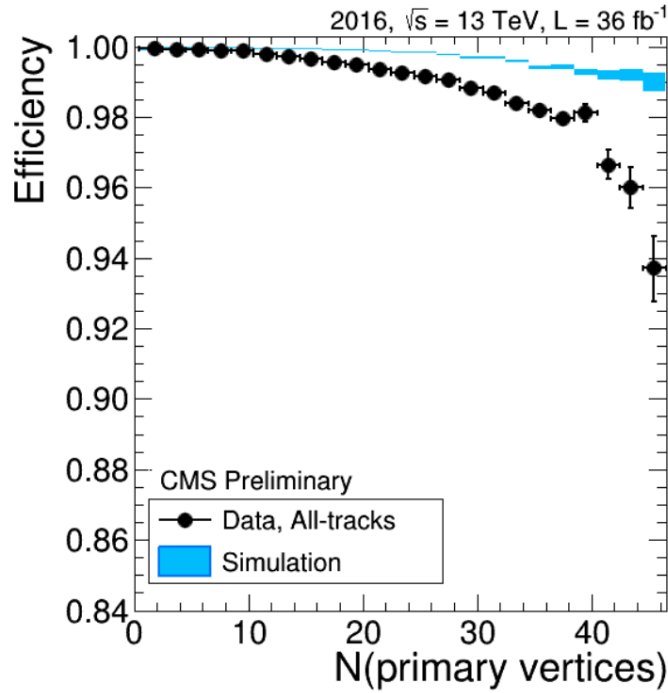


Figure 4.4: The track reconstruction efficiency for muons coming from Z decay as a function of the number of reconstructed primary vertices.

Figure 4.4 shows the track reconstruction efficiency for muons with $p_T > 10$ GeV as a function of PV reconstructed in an event using 2016 data. The tracking efficiency is 99% over the whole detector region, however it decreases as the number of primary vertices increases, falling to 94 – 95% for number of PV > 40 .

4.4 Electrons and Photons

Electrons and photons, collectively referred as the “egamma” objects, are reconstructed using their deposited energies in the ECAL crystals. The deposited energy of an egamma object is collected by forming clusters of ECAL crystals and super clusters (SC). Finally, the tracks reconstructed in the tracker are matched with the SCs to distinguish electrons from photons.

There are two clustering algorithms [38] used in CMS as described below.

- The **Island Algorithm** is designed to search for small energy deposits in individual clusters. The algorithm first selects the seed crystals with the maximum E_T and then starts collecting crystals around the seeds to form the clusters. The algorithm dynamically scans in both ϕ and η directions. The E_T threshold of the seed is the only parameter for this algorithm and it requires to be traded-off between an optimal energy resolution and reduction of noisy hits.
- The **Hybrid Algorithm** is used to reconstruct electrons with relatively higher energies ($E_T > 10$ GeV) in the barrel. The algorithm works with fixed arrays of crystals. It starts from a seed and then forms 1×3 or 1×5 crystal “domino” (see Figure 4.5) constructed with three/five crystals along the η direction. The dominoes are then clustered in the ϕ direction and each distinct cluster of dominoes follows a seed domino with $E_{seed_domino} > E_{seed}$.

An electron or photon deposits energy that may falls outside the primary shower cluster. Therefore, a cluster of clusters, commonly referred as the super-cluster is formed by a fixed area search in the $\eta - \phi$ plane around the seed cluster to recover the total energy of an electromagnetic shower. In the axial magnetic field of CMS, an egamma object is expected to form a shower mostly spread along ϕ but narrower along η . Figure 4.6 illustrates a super-cluster in the $\eta - \phi$ plane.

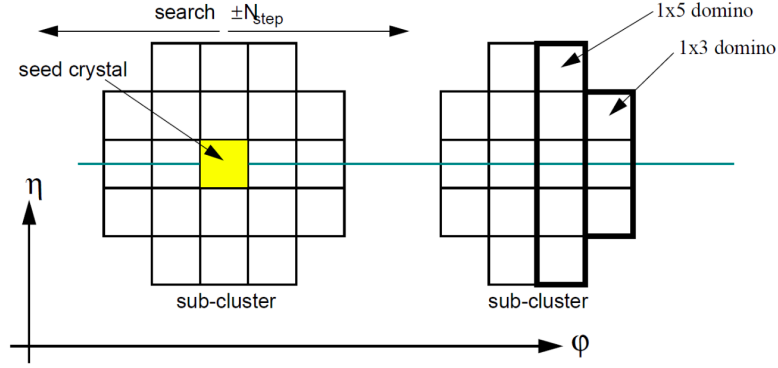


Figure 4.5: A schematic diagram of the domino construction step of the Hybrid algorithm.

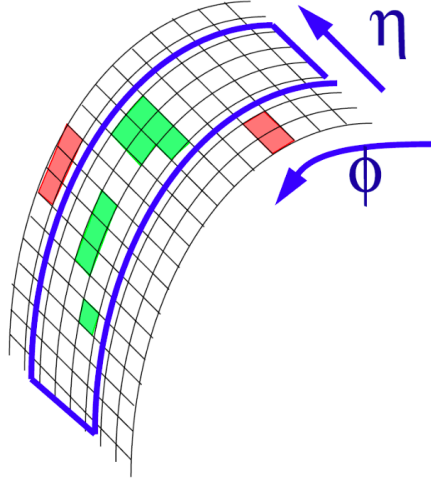


Figure 4.6: An example of a super-cluster algorithm is to collect all clusters that satisfy a given geometry condition.

The position measurement of a shower is performed by the energy weighted mean position of the crystals in a cluster. Depending on the variation of radiation length with the particle type and the lateral shower shape, the weighted mean of the logarithm of crystal energy [38] results in a precise measurement of the incident particle energy.

Any SC with $p_T > 10$ GeV is considered as a photon candidate. The formation of an ECAL SC for electron and photon is similar. Track requirement is not needed for photon reconstruction. The detailed discussion about the photon reconstruction can be found in Reference [39]. In order to reconstruct electrons, the PF algorithm matches the tracks with the super clusters. The full reconstruction has four steps: the track seed selection, track building and fitting using the Gaussian sum filter (GSF) tracking method and finally

track to SC matching.

The multivariate analysis to identify electrons at the analysis level is performed by the boosted decision tree (BDT) technique. The BDT is trained with all the electron candidates having $p_T > 10$ GeV categorized in three η regions, two in barrel and one in endcap. It takes several input features related to cluster shape, tracking and track-cluster matching. The final BDT response defines two working points (WP) for each η region: LooseWP (90% signal efficiency) and TightWP (80% signal efficiency), as mentioned in Table 4.1.

Regions	LooseWP	TightWP
Barrel ($ \eta < 0.8$)	0.837	0.941
Barrel ($0.8 < \eta < 1.479$)	0.715	0.899
Endcap ($ \eta > 1.479$)	0.357	0.758

Table 4.1: *Minimum BDT scores to identify electrons with two WPs.*

Isolation is another important metric to distinguish the prompt electrons from the electrons inside jets. The electrons originating from the event vertex are required to satisfy stringent conditions on the track impact parameters on the longitudinal and transverse directions. The total amount of energy deposited by the charged and neutral particles inside a cone (≤ 0.4) around the electron tracks is referred as electron isolation and the relative isolation can be written as:

$$I_{rel}^\ell = \frac{\sum p_T^{charged} + \max[0, \sum p_T^{neutral-had} + \sum p_T^\gamma - \beta \sum p_T^{PU}]}{p_T^\ell} \quad (4.1)$$

Here, $\sum p_T^{charged}$, $\sum p_T^{neutral-had}$ and $\sum p_T^{PU}$ are the scalar sum p_T s of the charged hadrons originating from the PV, neutral hadrons and photons, respectively. $\sum p_T^{PU}$ represents the sum- p_T of charged hadrons from pile-up vertices, and β denotes the fraction of the charged component. Usually, a cone of radius $0.2 - 0.4$ is used to calculate the isolation. However, the size of the cone is made to shrink inversely proportional to the p_T of the lepton in order to increase the efficiency for leptons reconstructed in events with “boosted” topologies and/or high hadronic activity to pass the isolation criteria. Isolation with narrow cone,

referred to as “mini-isolation”, reduces the effect of pileup. Therefore, instead of having a fixed cone size, the opening angle ($\sim 2M/p_T$, $M \equiv$ mass of the mother particle) between the decay products of a two-body decay decides the expected cone size.

4.5 Muon

The PF algorithm combines information from the tracker and muon systems to reconstruct the muons. There are three types of muon candidates used in CMS analyses.

- **Standalone muons:** The standalone muon reconstruction uses the information from the muon system only. Both tracking detectors (DT and CSC) of the muon chamber and the RPCs are used in the reconstruction. DT and CSC generate seeds that are used to produce probable muon tracks. The standalone muon tracks are then extrapolated to the nominal interaction point. Finally, a vertex-constrained fit is performed to obtain the standalone muon tracks.
- **Global muons:** The standalone muon tracks are extrapolated to the tracker volume and then combined with the hits in the tracker layers. Considering the muon energy loss in the detector material and the effect of multiple scattering, the muons from the innermost layer of muon system are extrapolated to the outermost layer of the tracker and then the entire array of hits is fitted to get the global muon candidates.
- **Tracker muon:** A large fraction of low p_T (6-7 GeV) muons can not provide enough hits in the muon chamber. Therefore, it is not possible to reconstruct the standalone and global muon candidates. Also, some high energetic muons may escape the gap between the wheels. So, a complementary approach is very useful where tracks with $p_T > 0.5$ GeV and momentum $p > 2.5$ GeV are considered as possible muon candidates. The tracks are then propagated through the ECAL, HCAL and the Muon Chamber. An extrapolated track qualifies as a tracker muon

if it matches at least one muon segment from a muon station (a short track reconstructed by one DT or CSC chamber).

The tracking and muon system are able to reconstruct $\sim 99\%$ of the muons produced in $p - p$ collisions either as tracker or global muons, referred as the Loose muons. Further stringent requirements on muon selection enhance the probability of a muon to be a prompt one and the tight conditions are:

- Muons must be reconstructed as both global and tracker muons.
- The global fit must satisfy $\chi^2/dof < 10$
- The following conditions must be satisfied by the associated track in the tracker.
 - matched with muon candidates having hits on at least two muon stations.
 - > 10 tracker hits.
 - at least one pixel hit.
 - transverse impact parameter $\delta_{xy} < 2$ mm.

Figure 4.7 shows the muon reconstruction and identification efficiencies with η for loose and tight identification criteria.

For further selection, $\Delta\beta$ correction is applied on the muon candidates to mitigate the contribution from pile-up as mentioned in Equation 4.1.

4.6 Jets

In proton-proton collisions, the quarks may radiate gluons and each gluon in turn split into a pair of quarks eventually producing a parton shower described by perturbative QCD. This is followed by a non-perturbative regime which represents the transition from partons

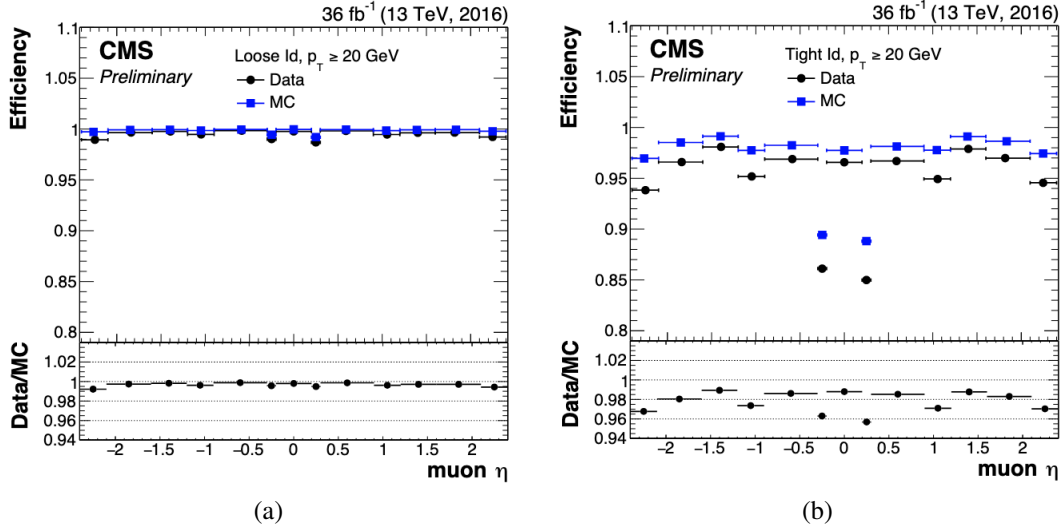


Figure 4.7: Muon reconstruction and identification efficiencies for (a) LooseID and, (b) TightID, as a function of η estimated with 2016 data.

to hadrons. The unstable hadrons *e.g.* π , K etc. decay and finally all the stable particles reach the detector. The particles grouped into cluster having collimated energetic particles which form jets and the direction of the jets are along the direction of primary partons produced in the collisions. The reconstruction of a jet must take into account all the hadronization products of the jet itself to measure the kinematics of the initial partons accurately.

Depending on the sub-detector information used, two types of reconstructed jets are considered: calorimeter or CALO jets and PF jets. The CALO jets are reconstructed using the energy deposits in the calorimeter towers only, whereas the PF jets use the PF candidates for reconstruction.

The effect of pile-up on the jet energy measurement and jet substructure are partially mitigated by the charged hadron subtraction (CHS) algorithm [32]. It uses the tracking information to remove the charged particles associated with a pile-up vertex. Another algorithm, pile-up per particle identification (PUPPI) is also used to mitigate the effect of pile-up. The detail discussion about the pile-up mitigation can be found in [40].

Several jet clustering algorithms [41] are used to form jets. The jet clustering algorithms

depend primarily on two distance parameters: the distance between an object and the beam (d_{iB}), and the distance between two objects (d_{ij}).

$$d_{ij} = \min(p_{T_i}^{2p}, p_{T_j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{T_i}^{2p} \quad (4.2)$$

Here the parameter $p = -1, 0, 1$ defines the type of the clustering algorithm, ΔR_{ij} is the distance between two objects i and j , R is the radius of the jet to be reconstructed. The jet clustering is formulated in a few steps. For all the PF objects, d_{iB} and d_{ij} are estimated using Equation 4.2 and then the minimum distance is found. If $d_{ij} < d_{iB}$, the two PF objects seem to be the constituents of a jet and therefore, they are combined to form a single entity. Otherwise, the i^{th} object is considered as a jet and will not be used in further clustering.

CMS uses the anti- k_T algorithm for jet clustering. Here, p is considered to be -1 which shows that the algorithm clusters jets around its harder constituent. The jet radius usually has two values, jets with $R = 0.4$ are referred as the Ak4 jets and the jets with $R = 0.8$ are called Ak8 jets. If a massive particle is so boosted that its decay products gets very close to each other, it is difficult to distinguish the decay products by Ak4 jet formation. Ak8 jets are fat jets which have information of the massive particle in its substructure.

4.6.1 b -jet

The heavy flavor jets are essential to study many interesting physics problems. In order to tag b -jets, several Machine Learning based algorithms have been developed. The jets originating from the b quarks which are hadronized to B mesons (B^0 or B^\pm) of a large lifetime of ~ 15 ps. Depending on the energy, a B meson can traverse a distance of ~ 1 cm before its decay inside the tracker volume. The precise position resolution of the CMS tracker can measure the displaced decay vertex *i.e.* the Secondary Vertex (SV) of the b quark (Figure 4.8) which is a crucial parameter for b -jet tagging.

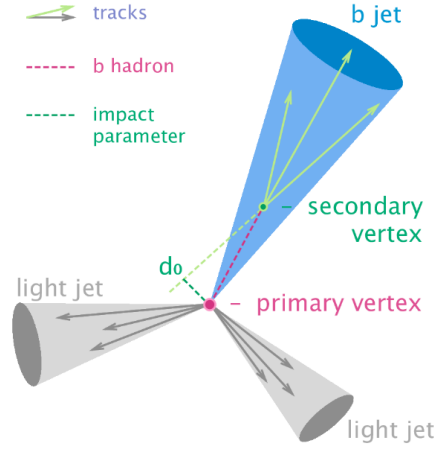


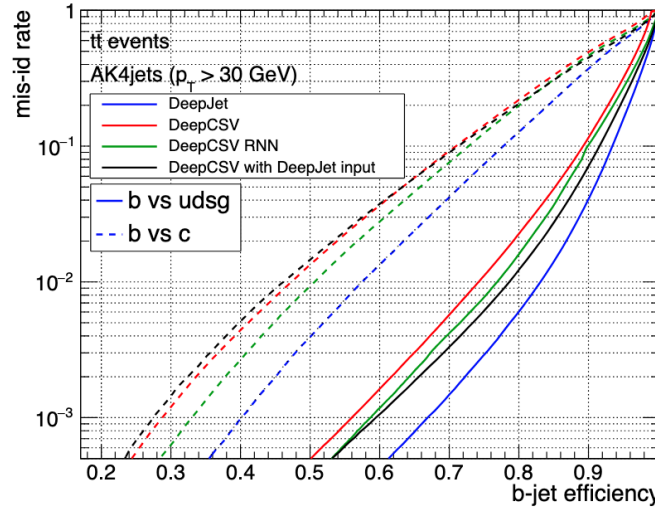
Figure 4.8: A schematic view of the origin of a b -quark from a secondary vertex.

In CMS, the Combined Secondary Vertex (CSV) [42] or the DeepCSV [43] algorithms are used to distinguish the b -jets from light quark jets. Both methods include the information about the SVs and tracks as the input features. These multivariate methods finally produce a response based on these features and several working points are determined to tag the b -jets as per analysis requirements. As a recent development, a DeepJet tagger equipped with convolution and recurrent neural network (RNN) layers, shows better performance than both CSV and DeepCSV. Figure 4.9 demonstrates the performance of several b -tagging algorithms [44] through the b -jet identification efficiencies and the misidentification efficiencies of b selected as c , quark or gluon jets and it is evident that the DeepJet algorithm performs better.

For each Run-2 era (2016-18), three working points of the deepJet algorithm are used for the physics analysis as shown in Table 4.2.

Era	Working points		
	Loose	Medium	Tight
2016	0.722	0.309	0.061
2017	0.749	0.303	0.052
2018	0.726	0.277	0.049

Table 4.2: Loose, Medium and Tight working points of the DeepJet b -tagging algorithm.


 Figure 4.9: Performance of various b tagging algorithms.

4.7 Tau

The τ lepton decays to ν_τ and W , where the latter in turn decays to lepton-neutrino or quark-antiquark pair. The branching fraction of τ decaying to hadrons is $\sim 66\%$. The leptonic decays involve a large amount of missing transverse energy (\cancel{E}_T) which is described in the next section. The hadronic decay of τ generates more collimated jets compared to the QCD jets. The reconstruction of the hadronic τ , denoted as τ_h , is performed by the hadron plus strip (HPS) algorithm [45]. The algorithm is able to reconstruct two τ_h decay modes:

- **One prong** decay contains one charged hadron and ≥ 0 neutral hadrons (π^0).
- **Three prong** decay contains three charged hadrons and ≥ 0 neutral hadrons (π^0).

The HPS algorithm considers the jets, reconstructed by the anti- k_T algorithm, as input. The π^0 s promptly decays to a pair of photons which usually converted to e^+e^- pairs due to high conversion probability. The CMS magnetic field causes a spatial separation of the e^+e^- in the $\eta - \phi$ plane on the ECAL surface. The fraction of energy of τ_h carried by the electrons or photons is reconstructed by forming clusters within a certain region of $\Delta\eta \times \Delta\phi$, referred as “strips”. The HPS algorithm has provision for p_T dependent dynamic

clustering. The jets used in the τ_h reconstruction are required to have $p_T > 0.5$ GeV and their origin must be compatible with the primary vertex. Finally, the reconstructed jets and strips are used to measure the position and momentum of the τ_h . A detailed description of the HPS algorithm can be found in Reference [46].

The isolation of the τ_h candidates is a crucial parameter to distinguish them from quark or gluon jets. A multivariate analysis based discriminant is used for this purpose. Similarly, dedicated boosted decision tree (BDT) based discriminants are used to reduce the fake probability of τ_h to be reconstructed as electron or muon. The BDT is trained using τ_h candidates following the requirements of $p_T^{\tau_h} > 20$ GeV and $|\eta^{\tau_h}| < 2.3$. To model the quark and gluon jets, the QCD multijet, W +jets, and $t\bar{t}$ events are used. Figure 4.10 demonstrates the variation of τ_h identification efficiency (left) and misidentification probability (right) with p_T of the generated τ_h and reconstructed jet, respectively [46].

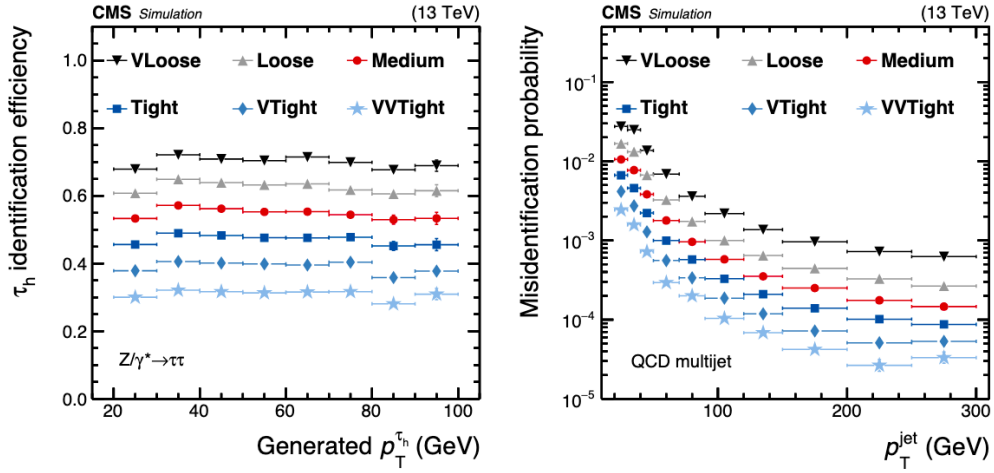


Figure 4.10: (a) τ_h identification efficiency, and (b) misidentification probability, for different WPs.

4.8 Missing Transverse Energy

At the time of the collisions, the transverse momentum of two colliding protons at the CMS interaction point are close to zero. Momentum conservation dictates the total trans-

verse momentum of all the particles produced after the collision should add up to zero. A finite value of the vectorial sum of p_T of all the particles may indicate the presence of missing particles such as neutrinos, other weakly interacting nonstandard particles etc.

The magnitude of the vectorial sum of p_T is known as the missing transverse energy (E_T). The direction of E_T is opposite to the direction of the sum p_T vector.

In CMS, three different algorithms exist for E_T reconstruction: PF E_T that uses the full particle flow algorithm, Calo E_T that uses the calorimeter energy towers and, Track-Correlated E_T where the calo E_T is corrected using the information of tracks of inner tracking system. CMS uses mainly the PF E_T for physics analyses.

The pile-up interactions contribute to the overall E_T estimation. “Type-0” E_T correction is implemented to remove the charged and neutral contribution arising from the pile-up vertices. It might be noted that the raw E_T uses the raw p_T values of the jets which may not be appropriate for physics analyses sensitive to E_T . “Type-1” correction is introduced to propagate the effect of jet energy correction to the E_T estimation [47].

4.9 Machine Learning

From particle identification and analysis to simulation and modeling, Machine Learning (ML) or Artificial Intelligence (AI) is being used extensively in particle physics experiment. Two basic ML based algorithms, BDT and DNN are described below.

BDT stands for Boosted Decision Tree, which is a type of ML algorithms used in data analysis. It is a supervised learning algorithm that can be used for both binary and multi-class classifications, as well as regression problems. In BDT, a decision tree is created using a set of input variables to make a prediction about the target variable. A decision tree is a flowchart-like structure where each internal node represents a test on an input variable, each branch represents the outcome of the test, and each leaf node represents

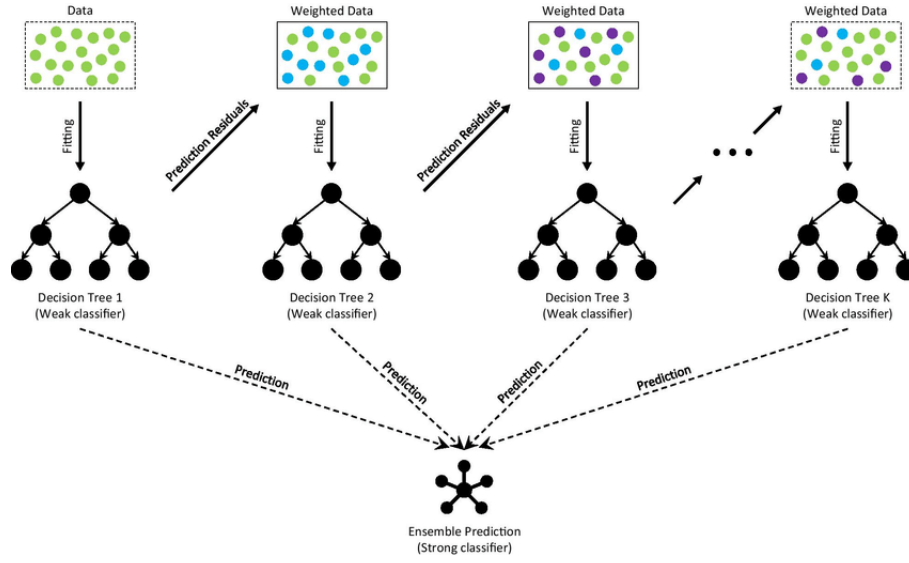


Figure 4.11: An ensemble of Decision Trees are used for prediction. Here the events of different classes are weighted using Gradient Boosting technique where the residual between the true and predicted output values are minimised to perform optimum classification. All the trees are trained sequentially and finally the combination of all of them is used for final prediction.

a class label or a numerical value. The tree is constructed recursively by splitting the dataset into subsets based on the best feature that separates the data most effectively. The term “Boosted” in BDT refers to a process of creating an ensemble of decision trees by iteratively training weak decision trees on the same dataset. The algorithm focuses on the data points that are difficult to classify correctly, and assigns higher weights to these points in each iteration to improve the accuracy of the model. Several types of “Boosting” algorithms are used like, Adaptive boost, Gradient boost, Extreme Gradient boost etc. BDT has been used for a variety of tasks such as identifying particles *e.g.* electrons, muons, jets by developing ID variables, distinguishing signal from background events, and improving the resolution of measured quantities. A schematic diagram [48] of a BDT is shown in Figure 4.11.

Apart from the tree based methods, Neural Networks (NN) are also capable of performing similar tasks. DNN stands for Deep Neural Network, which is a type of artificial neural network with multiple layers of nodes between the input and output layers. A neural network is a ML based model inspired by the structure and function of the human brain, consisting of interconnected nodes (or neurons) that process information. A DNN

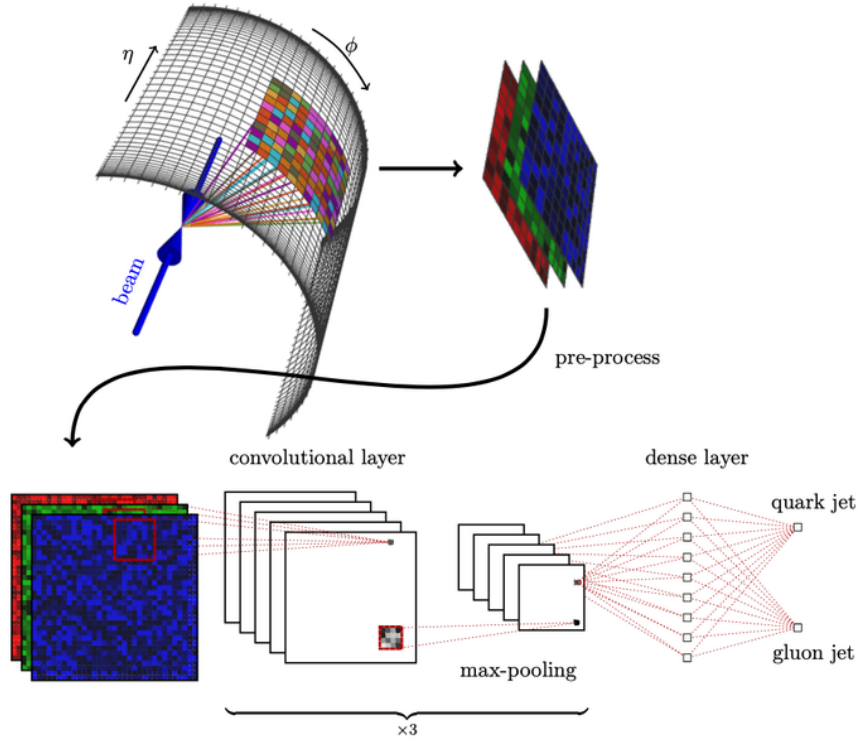


Figure 4.12: Diagram of a Convolutional Neural Network used to classify quark jets from gluon. Here the pictures of HCAL $\eta - \phi$ plane with energy deposition by jets are used as input. The probability of jet to be a quark or gluon like is considered as the output.

consists of multiple hidden layers of nodes, each layer performing a non-linear transformation of the input data. The number of layers in a DNN can vary depending on the complexity of the problem being solved. DNNs are trained using a backpropagation algorithm, which adjusts the weights and biases of the network to minimize the difference between the predicted output and the actual output. DNNs are widely used in a variety of applications, including image recognition, speech recognition, natural language processing, and predictive analytics. They have shown remarkable success in tasks such as image and speech recognition, where they have achieved state-of-the-art performance. Convolutional Neural Network (CNN) is an efficient technique to process input data, such as images. Figure 4.12 [49] shows a schematic diagram of a CNN used to discriminate quark jets from gluon jets. Under the same category, several types of ML techniques *e.g.* Auto-Encoders as anomaly detection techniques are being used for unsupervised or semi-supervised learning, Graph Neural Network are used for almost all the purposes in data

analysis or object reconstruction especially for unstructured data, etc.

Chapter 5

Search for Non-Resonant HH

Production in $b\bar{b}W^+W^-$ Decay Mode

A search for non-resonant HH production in the $b\bar{b}W^+W^-$ final state, performed with the full Run-2 $p-p$ collision data collected by CMS during 2016-18 at $\sqrt{s} = 13$ TeV, which corresponds to an integrated luminosity (\mathcal{L}) of 138 fb^{-1} , is presented in this chapter. The $HH \rightarrow b\bar{b}W^+W^-$ has the second highest branching ratio ($\sim 25\%$) among all the HH decay modes as shown in Figure 2.9. The final state of $b\bar{b}W^+W^-$ depends on the di-leptonic (DL), semi-leptonic (SL) or fully-hadronic (FH) decay of W boson pair. In the DL final state, other decay modes of HH i.e. $b\bar{b}ZZ$ and $b\bar{b}\tau^+\tau^-$ have subdominant contributions. The total branching fractions of HH to DL, SL and FH final states are shown in Table 5.1.

Final State	Process	BR (%)
DL	$HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}\ell^+\nu\ell^-\bar{n}u$	2.64
	$HH \rightarrow b\bar{b}ZZ \rightarrow b\bar{b}\ell^+\ell^-\nu\bar{n}u$	0.12
	$HH \rightarrow b\bar{b}\tau^+\tau^- \rightarrow b\bar{b}\ell^+\nu\ell^-\bar{n}u$ ($\ell \equiv e, \mu$)	0.89
SL	$HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}\ell^+\nu qq$	10.94
FH	$HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}qqqq$	11.31
	$HH \rightarrow b\bar{b}ZZ \rightarrow b\bar{b}q\bar{q}q\bar{q}$	0.4

Table 5.1: Branching ratios (BR) of several HH decay modes with DL, SL and FH final states.

In the full Run-2 data, HH (GGF) events for leptonic final states are expected to be:

$$\begin{aligned} N_{HH(GGF)} &= \sigma_{HH(GGF)} \times BR(HH \rightarrow b\bar{b}\ell^+\nu\ell^-\bar{\nu}, b\bar{b}\ell^+\nu q\bar{q}) \times \mathcal{L} \\ &= 31.05 \text{ fb} \times 14.6\% \times 138 \text{ fb}^{-1} \equiv 625 \end{aligned} \quad (5.1)$$

whereas the contribution from the relevant SM backgrounds is expected to be $O(10^9)$ events. Here ℓ includes e , μ and τ . The fully hadronic decay of WW is not considered because of several factors. As evident from Table 5.1, the number of fully hadronic events expected to be produced in the full Run-2 data is ~ 500 . Although the number is closer to the leptonic event count, poor jet energy resolution and huge background contamination (especially from $t\bar{t}$ +jets and multi-jet backgrounds) will make the final state inefficient for analysis. Therefore, only the DL and SL channels are considered.

To deal with such high background contamination in DL and SL final states, a few basic pre-selection cuts are applied to select events in several signal regions (SR) based on the jet multiplicity for the double and semi-leptonic decay processes only. After the preselections, multivariate analysis approach is used to extract the signal contribution. Finally, a statistical analysis is performed to estimate the upper limit on the HH production cross section at 95% confidence level .

5.1 Datasets

Only those data-taking periods where all the sub-detectors were fully operational are included in the analysis. The event and luminosity information of the “good-runs” are stored in the “Golden-JSON” files (Table 8.2 in Appendix 6.5) for all the three eras i.e. 2016 – 18. The primary datasets are formed on the basis of the HLT configuration as described in Section 3.2.6. The datasets used in this analysis are:

- **SingleMuon:** Dataset containing events triggered by SingleMuon HLT conditions where at least one μ candidate above some p_T threshold is required. Some additional conditions like, tracker muon conditions, muon isolation etc. are also used. The dataset is used for all three eras.
- **DoubleMuon:** Dataset with at least two μ candidates with relatively lower p_T threshold than the requirement of the SingleMuon dataset. Lower p_T requirement with more than one muons maintains the required HLT rate. This dataset is used for all the three eras.
- **SingleElectron:** Dataset with at least one e candidate available for 2016 and 2017. In addition to the p_T threshold on electron, several working points of the electron identification like pixel match, missing hits, $1/E - 1/p$ etc. and/or the isolation conditions are used.
- **DoubleEG:** Dataset with at least two e/γ candidates used for 2016 and 2017.
- **MuonEG:** Dataset with at least one μ and one e/γ candidates used for all three eras..
- **EGamma:** In 2018 era, SingleElectron and DoubleEG are merged in to the EGamma dataset.

The list of Golden JSON files are shown in Table 8.2 (Appendix 6.5). More information about the datasets analyzed for different channels are given in Tables 8.3, 8.4 and 8.5 of Appendix 6.5.

5.2 Monte Carlo Simulation

Monte Carlo simulation is a fundamental component of a physics analysis. A full Monte Carlo simulation is performed in the following stages.

Generation: The hard scattering of proton-proton *i.e.* the collision with maximum momentum transfer between two quarks or gluons are primarily generated by the Event Generators like, MadGraph, Powheg etc. The event generators calculate the amplitude of different Feynman diagrams using Matrix-Element method. The generated events are finally stored in the LHE format for further processing. One of the most important physics inputs used here is the Parton Distribution Function (PDF). The PDFs used in 2016 data-taking period is mostly NNPDF3.0, while in the 2017 and 2018 data-taking eras the NNPDF3.1 [50–52] set was used.

Showering and Hadronization: In this next step of event generation, HERWIG, Pythia etc. are used for showering and hadronization of hard scattered events. At the end, the final state particles with their kinematic description, propagate from the collision point towards the CMS detector. For the present analysis, Pythia is used. To incorporate the effect of multi-parton interactions, several underlying events (UE) tunes like, CP5, CUETP8M1, CUETP8M2 or CUETP8M2T4 [53–55] are used. All the samples are produced assuming $m_t = 172.5$ GeV and $m_H = 125$ GeV.

Detector Simulation: Interaction of the final state particles with the sensitive volumes of the CMS detector is simulated with Geant-4. The full detector simulation is integrated with the CMS software CMSSW. It produces a collection of time-stamped energy deposits for every sensor.

Digitization: In this stage, the effect of the CMS front-end electronics is emulated. It results in digital information structured as raw data acquired by the real CMS experiment.

Reconstruction: The reconstruction step onwards, same sequence of algorithms is used for simulation and real collision events. In this step, physically interpretable objects such as tracks, clustered calorimetric deposits and particle flow candidate are reconstructed.

Mini (/Nano)-AOD: Finally, the reconstruction output is reduced to a minimal set of variables which is sufficient to carry out the majority of the physics analyses. MiniAOD uses the Event Data Model (EDM) structure whereas, NanoAOD further concises the event information and keeps all of it in a flat ROOT tree.

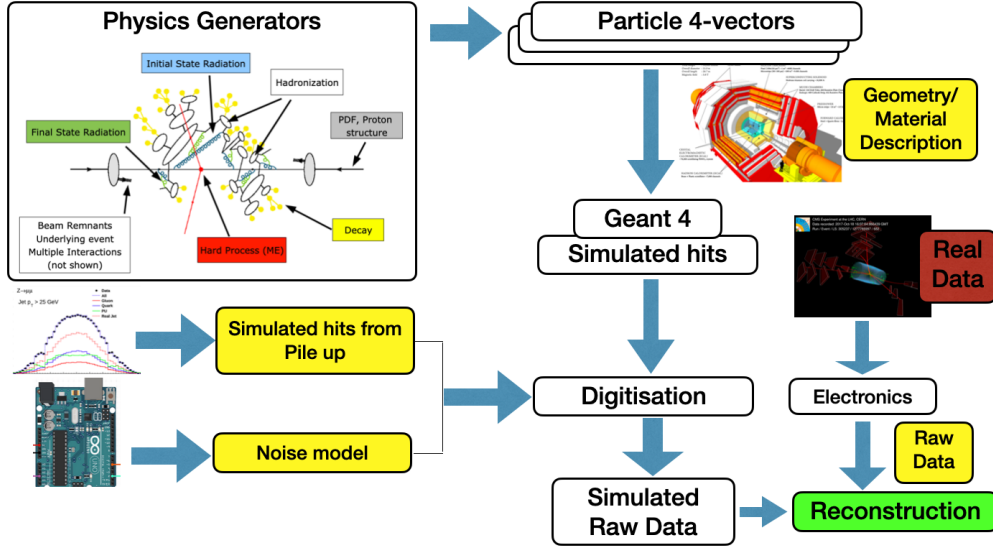


Figure 5.1: A schematic representation of the CMS simulation workflow.

Figure 5.1 describes the full simulation process described above. All the simulated signal and background events used for the purpose of estimating the expected signal and background yields in the analysis and to train machine learning algorithms, are discussed in the next two sections.

5.2.1 Signal

The HH signal samples for non-resonant gluon gluon fusion (GGF) are produced at the next-to-leading (NLO) order implemented with POWHEGv2 [56, 57]. These GGF- HH NLO samples are produced for four different benchmark values of the Higgs tri-linear self-coupling modifier $\kappa_\lambda = 0, 2.45, 5$ & 1 SM. For different values of κ_λ , the strength of destructive interference between the tree and box diagrams shown in Figure 2.5 varies which affects the HH cross sections and the event kinematics. The HH events can also be

produced by the Vector Boson Fusion (VBF) mechanism. The VBF samples are produced at the leading-order with POWHEGv2 using six different combinations of the coupling modifiers κ_λ , κ_V and κ_{2V} . For both GGF and VBF, the parton shower and hadronization are simulated with PYTHIA8 [58] using the parameter set CUETP8M1 for 2016 and TUNECPS [59, 60] for 2017 and 2018.

A statistical inference tool [61] dedicated for the HH non-resonant analyses embeds the HH physics model where one can scan a wide range of coupling modifiers by using a few signal benchmark points. The GGF non-resonant signal coupling benchmark points with respective cross sections are shown in Table 5.2. Table 5.3 shows the corresponding benchmark points for the VBF signal processes.

κ_λ	σ [pb]
1	0.02675
0	0.06007
2.45	0.01133
5	0.07903

Table 5.2: The production cross section of the non-resonant GGF HH signal samples at NNLO precision.

Separate signal samples are produced for $HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}\ell\nu qq'$, $b\bar{b}\ell\nu\ell\nu$, and $HH \rightarrow b\bar{b}\tau^+\tau^-$. The $b\bar{b}\tau^+\tau^-$ decay can contribute to both of the fully-leptonic (DL) and semi-leptonic (SL) channels based on the decay of τ leptons. Fully leptonic decay of the τ pair will contribute to the DL channel. If one τ decays to e/μ and the other τ decays

κ_V	κ_{2V}	κ_λ	σ [pb]
1	1	1	0.001668
1	1	0	0.004454
1	1	2	0.001375
1	2	1	0.01374
0.5	1	1	0.01046
1.5	1	1	0.0638
1	0	1	0.02617

Table 5.3: The production cross sections of the non-resonant VBF HH signal samples.

BP	1	2	3	4	5	6	7	8	9	10	11	12	8a
κ_λ	7.5	1.0	1.0	-3.5	1.0	2.4	5.0	15.0	1.0	10.0	2.4	15.0	1.0
κ_t	1.0	1.0	1.0	1.5	1.0	1.0	1.0	1.0	1.0	1.5	1.0	1.0	1.0
c_2	-1.0	0.5	-1.5	-3.0	0.0	0.0	0.0	0.0	1.0	-1.0	0.0	1.0	0.5
c_g	0.0	-0.8	0.0	0.0	0.8	0.2	0.2	-1.0	-0.6	0.0	1.0	0.0	0.8/3
c_{2g}	0.0	0.6	-0.8	0.0	-1.0	-0.2	-0.2	1.0	0.6	0.0	-1.0	0.0	0.0

Table 5.4: Twelve shape benchmarks (BP) [2] and additional one benchmark 8a [3].

BP	1	2	3	4	5	6	7
κ_λ	3.94	6.84	2.21	2.79	3.95	5.68	-0.10
κ_t	0.94	0.61	1.05	0.61	1.17	0.83	0.94
c_2	-1/3	1/3	-1/3	1/3	-1/3	1/3	1
c_g	0.5×1.5	0.0×1.5	0.5×1.5	-0.5×1.5	$1/6 \times 1.5$	-0.5×1.5	$1/6 \times 1.5$
c_{2g}	$1/3 \times (-3)$	$-1/3 \times (-3)$	$0.5 \times (-3)$	$1/6 \times (-3)$	$-0.5 \times (-3)$	$1/3 \times (-3)$	$-1/6 \times (-3)$

Table 5.5: Seven EFT benchmark points (BP) described in Reference [4].

hadronically (denoted by τ_h) but fails the p_T or η cut or the τ_h identification criteria, it will contribute to the SL channel. Therefore, in the present analysis, contribution from $b\bar{b}\tau^+\tau^-$ channels are taken into consideration to estimate the overall sensitivity.

As discussed in Section 2.3, the EFT shape benchmark points incorporate the clusters of BSM coupling values. Using clustering techniques, two sets of benchmark points were created. A preliminary strategy based on statistics of twelve points were obtained through two-sample tests, and one more point was found later. Seven more shape benchmarks were generated by a second strategy employing unsupervised Machine Learning techniques. The two sets of couplings for the determined benchmarks are shown in Tables 5.4 and 5.5, respectively. The signal samples for all the 21 shape benchmark points including the SM point are used to compute the upper limits.

5.2.2 Background

The relevant SM background processes considered in this analysis given in Table 5.6:

The DY and WJets backgrounds are modeled using the inclusive and exclusive samples. The inclusive samples include events with several jet multiplicities which cover a large

Group	Processes
ttjets	$t\bar{t}$ +jets (fully-leptonic, semi-leptonic, fully-hadronic)
ST	Single top quarks and top quark pairs in association with W
SH	Processes with single Higgs boson <i>e.g.</i> and associated to top pair $H \rightarrow \tau\tau, H \rightarrow ZZ, H \rightarrow WW, H \rightarrow \mu\mu, H \rightarrow b\bar{b}, t\bar{t}H,$ tHq, tHW, VH
ttV(X) [$X \equiv V, H$]	top pair with vector bosons <i>i.e.</i> W, Z or H <i>e.g.</i> $t\bar{t}W, t\bar{t}Z, t\bar{t}WW, t\bar{t}WH, t\bar{t}ZH$
DY	Drell-Yan processes: $Z \rightarrow \ell^+ \ell^-$
WJets	Single W boson production with jets <i>e.g.</i> W +jets, $W(i)$ +jets ($i \equiv 1 - 4$), and h_T binned samples
VV	Di-boson production such as, WW, WZ, ZZ, Wgg , and Zgg
VVV	Tri-boson processes WWW, WWZ, WZZ and ZZZ
Rare	Vector Boson Scattering (VBS) process $WWqq$ and $t\gamma, W\gamma$ $t\bar{t}\gamma, tZq$, and $4t$

Table 5.6: The relevant SM backgrounds used in $HH \rightarrow b\bar{b}W^+W^-$ analysis.

range of h_T *i.e.* the scalar sum- p_T of the jets. The inclusive sample of a physics process tries to mimic the kinematics of the same present in real $p - p$ collision data. Therefore, the events with high jet multiplicity or high h_T is expected to have lower statistics in the inclusive ones. The exclusive samples can populate the low statistics region of an inclusive sample. The exclusive samples are produced with some fixed number of jets or h_T range. Finally, the exclusive DY events are stitched to the inclusive one using the parton information from the generator level. Similarly, the WJets jet binned and h_T binned samples are merged to the inclusive one. The stitching procedure is documented in [62]. The major SM background processes with the respective CMS sample names and cross sections are shown in Table 5.7.

Process	Sample name	cross section [pb]
ttjets	TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	8.84×10^1 [63, 64]
	TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	3.66×10^2 [63, 64]
	TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	3.78×10^2 [63, 64]
ST	ST_s-channel_4f_leptonDecays_13TeV_PSweights-amcatnlo-pythia8	3.36 [65]
	ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	1.36×10^2 [65]
	ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1	8.10×10^1 [65]
	ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	3.58×10^1 [65]
	ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1	3.58×10^1 [65]
DY	DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFFFX-pythia8	6.08×10^3 [66]
	DYToLL_0J_13TeV-amcatnloFFFX-pythia8	4.84×10^3 [66]
	DYToLL_1J_13TeV-amcatnloFFFX-pythia8	8.98×10^2 [66]
	DYToLL_2J_13TeV-amcatnloFFFX-pythia8	3.36×10^2 [66]
Wjets	WJetsToLNU_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6.15×10^4 [67]
	W1JetsToLNU_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.44×10^3 [68, 69]
	W2JetsToLNU_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3.25×10^3 [68, 69]
	W3JetsToLNU_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1.15×10^3 [68, 69]
	W4JetsToLNU_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6.34×10^2 [68, 69]
ttW	TTWJetsToLNU_TuneCP5_13TeV-amcatnloFFFX-madspin-pythia8	1.96×10^{-1} [65, 70]
	TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFFFX-madspin-pythia8	4.05×10^{-1} [65, 70]
ttZ	TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	2.81×10^{-1} [64, 70]
	TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	5.87×10^{-1} [64, 70]
WW	WWTo2L2Nu_13TeV-powheg	1.22×10^1 [68]
	WWToLNUQQ_13TeV-powheg	5.05×10^1 [68]
WZ	WZTo3LNU_TuneCUETP8M1_13TeV-amcatnloFFFX-pythia8	4.43 [68]
	WZTo2L2Q_13TeV-amcatnloFFFX-madspin-pythia8	5.60 [71]

Table 5.7: List of the major backgrounds used in the analysis of data recorded in the 2016 data-taking period. The process name, sample name and the corresponding cross sections are tabulated.

5.3 Event Selection

Events of interest are selected in two distinct final states (channels):

- **Single-lepton channel (SL):** $HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}\ell\nu qq'$
- **Double-lepton channel (DL):** $HH \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}\ell\nu\ell\nu$

In each channel, a set of loose event selection conditions (pre-selection) are applied with an aim to keep a high efficiency for the signal while reducing the background contamination substantially. The final separation between the signal and backgrounds is performed with a deep neural network (DNN) based signal extraction method, described later in Section 5.3.3.

5.3.1 Physics Object Selection

Several selections are applied on the final state particles, namely leptons ($\ell \equiv e, \mu$), τ , jets (including b -jet) and \cancel{E}_T before any event selection.

Electron and Muon Identification

The identification of e and μ is performed in two stages:

- Basic e/μ identification and isolation criteria developed by the EGamma/Muon physics object group (POG) [72, 73] to separate genuine leptons from jets faking leptons.
- MVA techniques to separate the “prompt-leptons” originating from the decays of W , Z or τ from the “non-prompt” ones produced, *e.g.* originating from charm (c) and bottom (b) quarks. The variables used as input to the MVA are p_T , η and isolation of the lepton, properties of the jet nearest to the lepton, transverse and

longitudinal impact parameters of the lepton track with respect to the event vertex, and basic lepton identification criteria. Further details of this approach can be found in Reference [74].

Lepton isolation is computed in the same way for both electrons and muons. Electrons and muons in signal events are expected to be isolated. In the analysis, the “mini-isolation”, described in Section 4.4, is used for the selection of leptons.

Based on all the selection criteria, the leptons are categorized as “Loose”, “Fakeable” and “Tight”. Leptons passing the loose selection are used for the purpose of vetoing events containing lepton pairs of low mass. The fakeable leptons are used to remove the overlap (“cleaning”) between different types of objects, to estimate the fake lepton background from control regions in data, and to compute global kinematic properties of the events. The tight lepton selection is similar to the fakeable one, but has more stringent requirements on the prompt- e and prompt- μ MVA based discriminators. The tight leptons are used to select events in the signal regions (SR) of the SL and DL channels.

Electron

The first step of the electron identification is performed by a boosted decision tree (BDT) method which has been trained to discriminate electrons against jets. The training is performed by the CMS EGamma POG in DY MC samples with the XGBoost [75] algorithm. The discriminant is based on the PF information without any condition on isolation. Three working points (WP) have been defined based on the BDT output score: “WP-loose”, “WP-90” and “WP-80”, corresponding to 98%, 90% and 80% signal efficiency, respectively. The complete electron selection conditions are described in Table 5.8.

Electrons			
Observable	Loose	Fakeable	Tight
p_T	$> 7 \text{ GeV}$	$> 10 \text{ GeV}$	$> 10 \text{ GeV}$
$ \eta $	< 2.5	< 2.5	< 2.5
$ d_{xy} $	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$
$ d_z $	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$
d/σ_d	< 8	< 8	< 8
I_e	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
$\sigma_{i\eta i\eta}$	–	$< \{ 0.011 / 0.030 \}^1$	$< \{ 0.011 / 0.030 \}^1$
H/E	–	< 0.10	< 0.10
$1/E - 1/p$	–	> -0.04	> -0.04
Conversion rejection	–	✓	✓
Missing hits	≤ 1	$= 0$	$= 0$
EGamma POG MVA	$> \text{WP-loose}^2$	$> \text{WP-90} (> \text{WP-loose})^2 \dagger$	$> \text{WP-loose}^2$
Deep Jet of nearby jet	–	$< \text{WP-tight} (< \text{WP-medium})^3$	$< \text{WP-medium}^3$
Jet relative isolation ⁴	–	$< 0.7 (-) \dagger$	–
Prompt- e MVA	–	–	> 0.30

¹ Barrel / endcaps.

² WPs as defined by EGamma POG.

³ WPs as defined by JetMET POG.

⁴ Either the PF-relative isolation computed within a cone of fixed size $\Delta R = 0.4$ or, $\approx p_T^{\text{jet}}/p_T^e - 1$ if the electron is matched to a jet within $\Delta R < 0.4$.

\dagger Fails (passes) the requirement prompt- e MVA > 0.30 .

Table 5.8: *Loose, fakeable and tight selection criteria for electrons. The requirement on the output of the BDT trained by the EGamma POG and on the observables $\sigma_{i\eta i\eta}$, H/E , and $1/E - 1/p$ are varied as function of η of the electron candidate. The conditions on the EGamma POG MVA, the Deep Jet discriminant and the relative isolation of the nearest jet to the electron are tightened (relaxed) for fakeable electrons that fail (pass) the requirement prompt- e MVA > 0.30 , in order to reduce the systematic uncertainty on the fake lepton background estimate on the jet flavor composition [5, 6]. A hyphen (–) indicates selection criteria that are not applied.*

Muon

The muons are initially selected by applying the loose PF muon identification criteria provided by the Muon POG after p_T and a few more acceptance cuts as mentioned in Table 5.9. All the muon selection cuts are described in the same Table.

In the analysis, the muons and electrons are cross-cleaned against each other by removing the overlap using $\Delta R_{e,\mu} > 0.4$.

Muons			
Observable	Loose	Fakeable	Tight
p_T	$> 5 \text{ GeV}$	$> 10 \text{ GeV}$	$> 10 \text{ GeV}$
$ \eta $	< 2.4	< 2.4	< 2.4
$ d_{xy} $	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$
$ d_z $	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$
d/σ_d	< 8	< 8	< 8
I_μ	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
PF muon	$> \text{WP-loose}^1$	$> \text{WP-loose}^1$	$> \text{WP-medium}^1$
Deep Jet of nearby jet	—	$< \text{WP-interp.} (< \text{WP-medium})^{2\dagger}$	$< \text{WP-medium}^2$
Jet relative isolation ³	—	$< 0.8 (-)^\dagger$	—
Prompt- μ MVA	—	—	> 0.5

¹ WPs as defined by Muon POG.

² Upper cut on the Deep Jet score defined with a linear interpolation from Deep Jet WP-medium at $p_T = 20 \text{ GeV}$ to Deep Jet WP-loose at $p_T = 45 \text{ GeV}$, taking the Deep Jet WPs as defined by JetMET POG.

³ Defined as $p_T^{\text{jet}}/p_T^\mu - 1$ if the muon is matched to a jet within $\Delta R < 0.4$ or as the PF-relative isolation computed within a cone of fixed size $\Delta R = 0.4$ otherwise.

\dagger Fails (passes) the requirement prompt- μ MVA > 0.5 .

Table 5.9: *Loose, fakeable and tight selection criteria for muons. The conditions on the Deep Jet discriminant and the relative isolation of the nearest jet to the muon are tightened (relaxed) for fakeable muons that fail (pass) the requirement of prompt- μ MVA > 0.5 . A hyphen (—) indicates selection criteria that are not applied.*

Hadronic τ Selection

Hadronic τ decays (τ_h) are reconstructed by the “hadrons plus strips” (HPS) algorithm and identified by a multivariate approach based on a convolutional deep neural network and referred to as “Deep Tau v2.1”, which has been trained to discriminate τ_h from quark and gluon jets as well as from electrons and muons. The present analysis uses τ_h only for the purpose of applying a τ_h -veto. More specifically, events containing one or more τ_h reconstructed in one of the decay modes h^\pm , $h^\pm + 1\pi^0$, $h^\pm + 2\pi^0$, $h^\pm h^\mp h^\pm$, or $h^\pm h^\mp h^\pm + 1\pi^0$, and passing $p_T > 20 \text{ GeV}$ and $|\eta| < 2.3$, and the byMediumDeepTau2017v2VSjet discriminant, are vetoed in the single lepton channel. This is done in order to avoid overlap with the $e\tau_h$ and $\mu\tau_h$ channels of the $HH \rightarrow b\bar{b}\tau^+\tau^-$ final states [76]. Only those τ_h that pass the fakeable selection and do not overlap with electrons or muons within $\Delta R < 0.3$ are considered for the τ_h -veto.

Ak4-Jets Selection

Jets are reconstructed using the anti- k_t algorithm [77] as described in Section 4.6. Jets arising from calorimeter noise *i.e.* fake jets are rejected by requiring reconstructed jets to pass the Loose WP of the PF jet identification criteria in 2016 and the tight WP in 2017 and 2018, following the recommendations of the JetMET POG [78].

Ak4 jets		
Observable	selection	for b jets
p_T	$> 25 \text{ GeV}$	$> 25 \text{ GeV}$
$ \eta $	< 2.4	< 2.4
jetId ¹	Loose WP	Loose WP
jetPulId ²	Loose WP	Loose WP
clean	$\Delta R_{jet,\ell} > 0.4$	$\Delta R_{jet,\ell} > 0.4$
btagDeepFlavB	–	> 0.3093

¹ PF jet identification criteria recommended by the JetMET POG.

² ID, developed to discriminate jets originating from the PV, from the pileup jets, applicable for $p_T \leq 50 \text{ GeV}$.

Table 5.10: *Selection of Ak4 jets and Ak4 b-tagged jets.*

The Ak4-jets are also required to be cleaned against the fakeable electrons and muons. Ak4-jets originating from the hadronization of b quarks are identified by the “Deep Jet” algorithm, as described in Chapter 4. Three levels of b-tagging selection are provided by the b -tag and vertexing (BTV) POG [79] based on the b -jet selection efficiency and misidentification rates. In this analysis, the medium b -tagging WP is used (Table 4.2).

Ak8 Jet Selection

The Ak8-jets are required to pass the selection conditions as described in Table 5.11.

5.3.2 Pre-selection Criteria

A set of common selection criteria are applied on the events in both SL and DL channels as described in the following:

Ak8 jets		
Observable	selection	for b jets
p_T	$> 200 \text{ GeV}$	$> 200 \text{ GeV}$
$ \eta $	< 2.4	< 2.4
jetId	Loose WP	Loose WP
hasValidSubJets ¹	True	True
softdrop mass ²	$30 - 210$	$30 - 210$
subjettiness $(\tau_{21})^3$	≥ 0.75	≥ 0.75
clean ⁴	$\Delta R_{jet,\ell} > 0.8$	$\Delta R_{jet,\ell} > 0.8$
btagDeepFlavB ⁵	–	> 0.3093

¹ Two subjects must be there both having $p_T > 20 \text{ GeV}$ and $|\eta| < 2.4$.

² Mass of an Ak8-jet using “modified mass drop tagger” algorithm [80, 81]

³ Number of subjects in an Ak8-jet is “N-subjettiness” (τ_N). The discrimination of boosted $H \rightarrow b\bar{b}$ decays from quark and gluon jets is improved by demanding that the ratio τ_2/τ_1 , which quantifies the compatibility of the Ak8-jet with the “two-prong” structure expected for the decay of a W , Z , or H boson into two quarks, satisfies the condition $\tau_2/\tau_1 < 0.75$

⁴ Cleaning against the leading fakeable lepton by a ΔR of 0.8.

⁵ Any one of the two subjects must pass $p_T > 30 \text{ GeV}$ and the medium working point of btagDeepB.

Table 5.11: *Selection of Ak8 jets and the b -tagged ones.*

- **HLT:** Events selected in the SL channel are required to pass either the single-electron or the single-muon trigger, depending on whether the offline reconstructed lepton is an electron or a muon. In the DL channel, the acceptance for the HH signal is increased by using a combination of single-lepton and double-lepton triggers. All the HLT paths used in this analysis are shown in Table 8.1 in Appendix 6.5.
- **\cancel{E}_T Filters:** Events selected in both channels are required to pass the filter algorithms given in Table 8.6 in Appendix 6.5, as recommended by the JetMET POG. These filters perform additional event cleaning against beam halo effects, detector noise, etc. The filters are applied in both data and simulation.
- **Low mass veto:** Events containing a pair of leptons passing the loose selection and having an invariant mass $< 12 \text{ GeV}$ are rejected, because these events are not well modeled by the Monte Carlo simulation.

Following the common selection mentioned above, a few specific conditions are applied on the events to build the signal regions for the DL and SL channels.

CHAPTER 5. SEARCH FOR NON-RESONANT HH PRODUCTION IN $B\bar{B}W^+W^-$ DECAY MODE

Selection	SL	DL
$HH \rightarrow$	$b\bar{b}W^+W^- \rightarrow b\bar{b}\ell\nu qq'$	$b\bar{b}W^+W^- \rightarrow b\bar{b}\ell\nu\ell\nu$
Trigger	Single-lepton trigger	Single- and double-lepton triggers
Lepton p_T	$p_T > 30 \text{ GeV (e) or } p_T > 25 \text{ GeV (}\mu\text{)}$	$p_T > 25 / 15 \text{ GeV}$
Lepton η	$ \eta < 2.5 \text{ (e) or } \eta < 2.4 \text{ (}\mu\text{)}$	
τ_h veto	applied	not applied
Charge requirements	–	$\sum_{\ell} q = 0$
Jet multiplicity	$\geq 1 \text{ AK8-jet and } \geq 1 \text{ AK4-jet}$ or $\geq 3 \text{ AK4-jets}$	$\geq 1 \text{ AK8-jet}$ or $\geq 1 \text{ AK4-jet}$
b tagging requirements	$\geq 1 \text{ AK8-jet with subjet passing medium WP}$ or $\geq 1 \text{ AK4-jet passing medium Deep Jet WP}$	
Dilepton mass	$ m_{\ell\ell} - m_Z > 10 \text{ GeV}^*$	

* Applied on all lepton pairs of same flavor, opposite charge, and passing the loose lepton selection.

Table 5.12: *Pre-selection conditions applied in the SL and DL channels.*

At the end of the pre-selection step, events are divided into three mutually exclusive categories based on the b -tagged Ak4 and Ak8 jet multiplicities as described in Table 5.13.

Event category	SL	DL
Resolved	Ak4 jets ≥ 3 Ak8 b -jets = 0	Ak4 jets ≥ 2 Ak8 b -jets = 0
Resolved 1b	Ak4 b -jets = 1	
Resolved 2b	Ak4 b -jets = 2	
Boosted	Ak8 b -jets ≥ 1	

Table 5.13: *Event categories after pre-selection. The first row describes the selection for the Resolved category with two specific categories: Resolved 1b, Resolved 2b and, the Boosted category is mentioned in the last row.*

All the above pre-selection conditions are applicable for both the GGF and VBF production modes of the signal. For the VBF mode, a few more selections are applied. Any event produced by vector boson fusion should contain two forward Ak4-jets referred as the “VBF jets”. The VBF jets are selected as:

- jet $p_T > 30 \text{ GeV}$ and $|\eta| < 4.7$.
- Jets in the range $2.7 \leq |\eta| \leq 3.0$ should have $p_T \geq 60 \text{ GeV}$. This cut is useful to remove spurious jets that are due to noise in the ECAL endcap.
- Jets should pass the tight working point of the JetID.
- The jet candidates should not overlap with the leading or sub-leading leptons by

$$\Delta R < 0.4.$$

- In the resolved category, the candidates should not overlap with the Ak4-jets originating from H for DL and H, W for SL by $\Delta R < 0.8$.
- In the boosted category, they should not overlap with the leading b -tagged Ak8-jet by $\Delta R < 1.2$.
- Pairs of Ak4-jets ($j1, j2$) passing the above cuts are selected as the VBF-jet pair candidates if the pair satisfies:
 - $|\Delta\eta_{j1,j2}| > 3$.
 - $M_{j1,j2} > 500 \text{ GeV}$.
- In case more than one VBF jet pair candidates are found, the pair with the highest invariant mass is selected as the VBF jet pair.

Process	Expected yield	Yield: SL	Yield: DL
$t\bar{t}$ (fully-leptonic)	1.2×10^7	1.5×10^6	1.5×10^6
$t\bar{t}$ (semi-leptonic)	5.2×10^7	1.1×10^7	1.9×10^2
$t\bar{t}$ (fully-hadronic)	5.0×10^7	2.9×10^2	0
DY	8.4×10^8	2.0×10^4	4.2×10^2
W +jets	8.5×10^9	2.8×10^6	2.1×10^2
ST	3.9×10^7	1.3×10^6	7.9×10^4
HH (GGF)	468.8	96.1	–
	113.1	–	15.0
HH (VBF)	26.1	3.83	–
	6.3	–	0.55

Table 5.14: Events produced at $\mathcal{L} = 138 \text{ fb}^{-1}$ and the yields after pre-selection for the SM benchmark points of the $HH \rightarrow b\bar{b}W^+W^-$ GGF and VBF signal processes and major backgrounds.

The corresponding event yields at $\mathcal{L} = 138 \text{ fb}^{-1}$ are mentioned in Table 5.14.

5.3.3 Multivariate Analysis

The pre-selection is followed by a multivariate analysis (MVA) separately for the SL and DL channels. The MVA is composed of a deep neural network (DNN) with a physics

motivated pre-processor to the DNN. The training is performed separately for each channel adding all the lepton flavor combinations (single electron, single muon for SL and double electron, double muon, electron + muon for DL), events from all eras (2016-18) and different topologies (resolved, boosted). A single network covers both GGF and VBF production modes.

The machine learning method used with the multi-classification approach is a specialized physics motivated architecture known as the Lorentz-Boost Network (LBN) [82], together with a DNN having residual blocks (ResNet) [83]. The LBN builds high-level variables from basic variables like particle four vectors and passes the extracted high-level variables along with user defined ones to the DNN. The DNN with ResNet architecture is expected to perform well with a large number of trainable variables and large statistics.

Input and Output

The input variables (features) used to train the neural networks for SL and DL are different. The input features include several low-level variables like the four momenta of the final state particles, charge, b -tag score etc. fed to the LBN to extract high-level variables and user defined variables are directly fed to the DNN. In case of the low level variables, if the object is not available in an event all the related observables are set to zero. All the low level variables for both the channels, and some of the important¹ high level variables for the SL are shown in Table 5.15. It should be mentioned that six (four) instead of the minimum amount of four (two) jets in SL (DL) are considered for the fact that the jets from the hard process may not always be identified correctly.

The classification task contains seven and nine output classes for SL and DL, respectively. The seven output classes of the DNN for SL are described in Table 5.16.

¹Feature importance is estimated using the permutation invariance [84] method. Using one million events from the non-trained dataset, a single feature is shuffled randomly among all the events keeping the others fixed, and the loss is calculated. The relative difference between the loss with the shuffled feature and the original one determines the importance of the corresponding feature.

Input feature	Description
Low level variables for DL and SL	
One-hot encoded PDGId and Charge of leading (both) lepton(s)	
Six (four) Ak4-jets with two highest DeepJet b -tagging score and the last four (two) sorted in p_T order	
Leading Ak8-jet with subjettiness ($\tau_1, \tau_2, \tau_3, \tau_4$) and the softdrop mass	
\cancel{E}_T where the invariant mass and p_z are set to zero	
High level variables for SL	
h_T	Scalar sum p_T of all the jets in an event
$M_{b\bar{b}}$	Invariant mass of the two b -jets with highest b -tagging score
M_{HH}	
$p_{T_{b\bar{b}}}$	p_T of the $b\bar{b}$ system
$\sqrt{\hat{s}_{min}}$	Minimum parton level center-of-mass energy [85]
$\Delta R_{b,\bar{b}}$	ΔR between the two b -jets selected
p_{T_ℓ}	p_T of the leading lepton
$M_{W \rightarrow jj}$	Invariant mass of the two leading p_T jets (exclusive to the $H \rightarrow b\bar{b}$ system)
$M_{j1,j2}$ (VBF)	Invariant mass of the VBF jets pair
$ \Delta\Phi_{\ell, \cancel{E}_T} $	$\Delta\Phi$ between the leading lepton and \cancel{E}_T
etc.	

Table 5.15: Low level variables for SL and DL, and some important high level variables for SL only.

Class	Description
HH (GGF)	HH GGF NLO samples for four different values of κ_λ as mentioned in Table 5.2
HH (VBF)	HH All the VBF samples mentioned in Table 5.3
ttjets	$t\bar{t}$ + jets, $t\bar{t}V(V) + \text{jets}$ and $t\bar{t}\gamma + \text{jets}$ ($V \equiv W, Z$)
ST	Processes involving single top quark, such as: $t/\bar{t} + \text{jets}$ and $tV + \text{jets}$
WJets	Processes with single W categorized in multiplicity of jets and h_T
H	Processes with single Higgs boson: $gg \rightarrow H$, VBF H , tHV , $t\bar{t}H + \text{jets}$, ZH etc.
Other	DY, VBS $W^+W^+ + 2\text{jets}$, $VV(V)$ etc.

Table 5.16: DNN output classes for DL channel.

For DL, the DY is considered as a separate output class as it provides a large contribution. Also a new class, composed of $ttV(X)$ ($X \equiv V, H$) type of processes *i.e.* $t\bar{t}V$, $t\bar{t}VV$ and $t\bar{t}VH$, is considered in the ttjets and H output classes of the SL.

Architecture and Performance

The neural network architecture connects the LBN with the DNN. The LBN uses the particle 4-momenta and constructs several combinations of intermediate particles and the corresponding rest frames. The number of combination is a hyperparameter and if there are N input particles and M rest frames, there will be $2 \times N \times M$ trainable weights at

this stage. The LBN performs Lorentz Transformations to boost the particles in their dedicated rest frame and the network performs autonomous feature extraction. The output of the LBN contains the 4-vector, p_T , p , invariant mass of each output particles, as well as $\cos(\theta^*)$ *i.e.* the angular difference between the directions of one particle in other's rest frame. A schematic of the LBN+DNN is shown in Figure 5.2. The detailed description of the working principle of the LBN can be found in Reference [82].

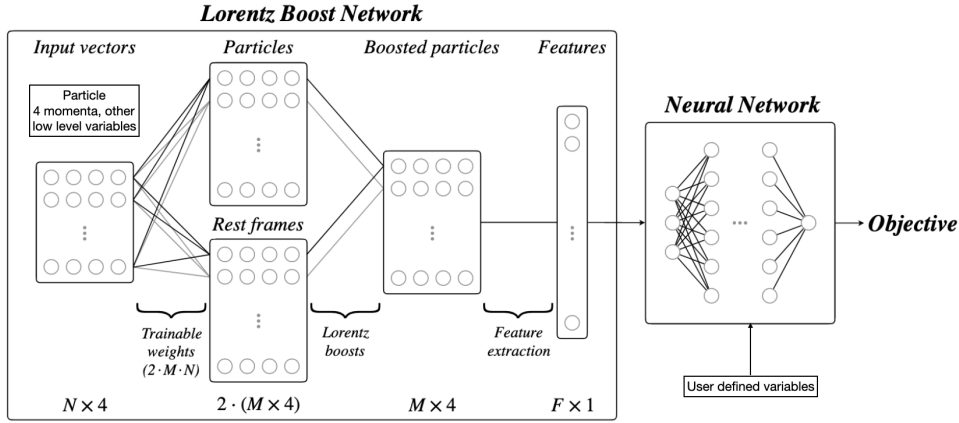


Figure 5.2: An illustration of a two-stage neural network architecture equipped with the LBN and a subsequent DNN.

There are twelve output particles from the LBN, resulting in a total number of inputs for feed forward network of 235 (229) in SL (DL). The feed forward network *i.e.* the DNN consists of three ResNet blocks of two layers of 235 (229) units with ReLU activation in SL (DL). This is followed by the last layer of seven (nine) units in SL (DL) with Softmax activation, which produces the output of the network. Each layer employs a batch normalization which is applied before the activation function. The usual binary cross-entropy is used as the loss function to be optimized and the Adam optimizer is used to determine the weights through back propagation. The learning rate is set to 0.001 (0.01) for SL (DL) and $\sim 10^{-8}$ L2 regularizer is used to control the over-training.

Figure 5.3 summarizes the performance of the networks through the corresponding confusion matrices for the SL and DL channels. The prominent diagonal elements are indicative of a working multi-class classification.

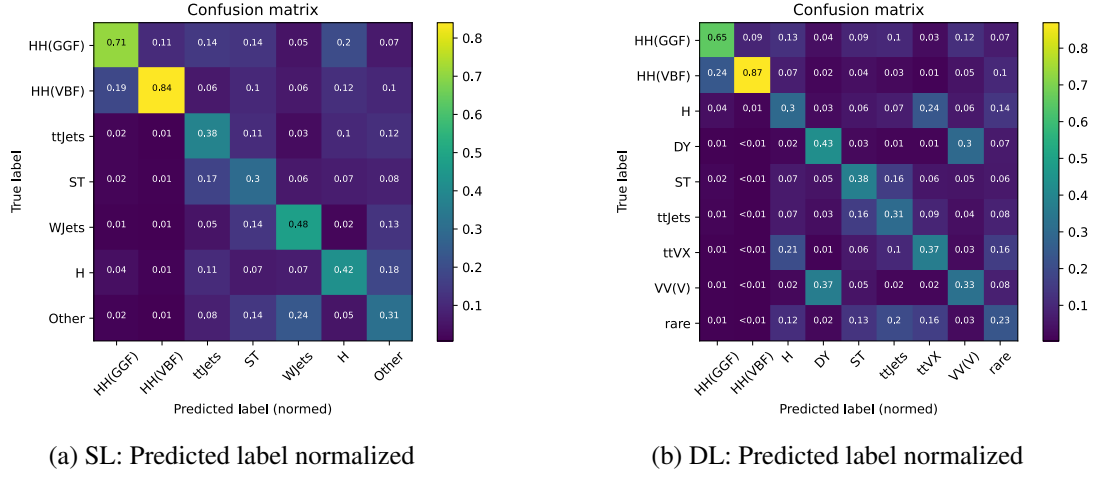


Figure 5.3: Confusion matrices showing the performance of the multi-class classification of the networks. The plots show the composition of each predicted label, under the given assumption that all input classes are of equal event weight sum.

The effective binary classification performance to discriminate $gg \rightarrow HH$ from all the background processes is summarized in the receiver operating characteristic (ROC) curve with an area under the curve *i.e.* $AUC = 0.85$ (0.94), as shown in Figure 6.6 for SL and DL.

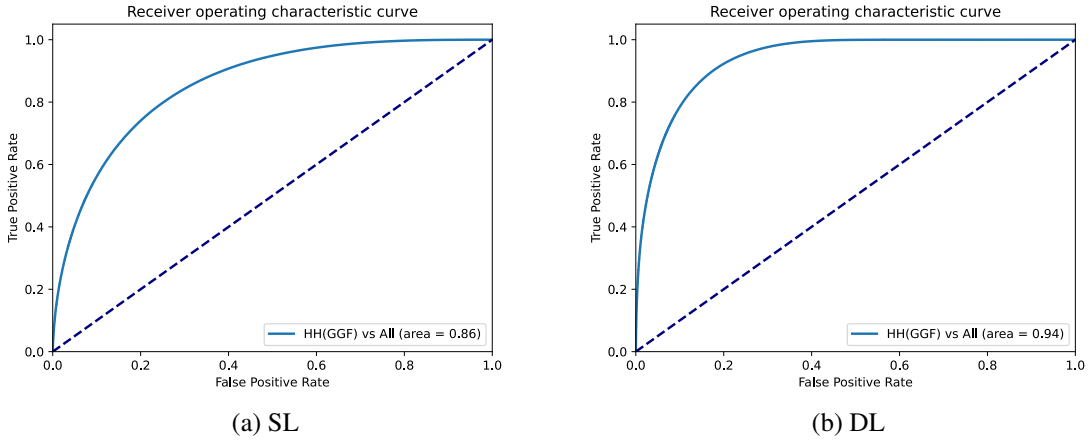


Figure 5.4: ROCs for binary discrimination performance of $gg \rightarrow HH$ (SM) process against all background processes.

Grouping and Sub-Categorization

The multi-class DNN produces one output distribution for each process. These distributions are merged into groups of processes in order to maintain sufficient statistics. All the categories for both SL and DL channels *i.e.* Resolved 1b, Resolved 2b and Boosted, including all the statistical and systematic uncertainties, are used in a simultaneous maximum likelihood fit to the observed/pseudo data. The signal enriched distributions are used for the extraction of the HH signal. The background distributions are useful to constrain the overall background normalization. The categories are summarized in table 5.17 for both the SL and DL channels.

Process Group	Sub-Categories		
SL			
HH (GGF)	Resolved-1b	Resolved-2b	Boosted
HH (VBF)	Resolved-1b	Resolved-2b	Boosted
Top + Higgs	Resolved		Boosted
WJets + Others	Inclusive		
DL			
HH (GGF)	Resolved-1b	Resolved-2b	Boosted
HH (VBF)	Resolved-1b	Resolved-2b	Boosted
Top + Others	Resolved		Boosted
DY + VV(V)	Inclusive		

Table 5.17: Sub-categorizations for the merged processes in the SL and DL channels.

Re-Binning Strategy

The shape of the final discriminant plays the most important role in a shape based analysis, where the test statistic *i.e.* the maximum likelihood ratio includes the bin by bin expected number of signal and background events with their uncertainties. The likelihood fit turns out to be very expensive for a large number of bins of the discriminant. Therefore, it is essential to keep an optimum number of bins maintaining the shape of the respective distribution.

Initially, for all the groups (Table 5.17), the DNN responses are produced with 400 bins. Then, the histograms for the signal and background categories are rebinned using the

following concepts.

- **Signal** categories *i.e.* Resolved 1b, Resolved 2b and Boosted, show that the DNN response value is higher for the signal process than the other processes for that signal category. Therefore, the rebinning strategy for the signal provides a large impact on the sensitivity. The strategy follows aggregating bins starting from the rightmost side of the DNN distribution until the following criterion is satisfied

$$\sum_i \text{Bin Content}(P_i) - \sqrt{\sum_i \sigma_{P_i}^2} > t \quad (5.2)$$

Here the \sum runs over all the processes (each is P_i) of a particular signal category, σ_P is the statistical uncertainty of the bin content and t is the predefined threshold of content for each bin of the rebinned histogram. The signal-to-background ratio grows in the rightmost bins while the threshold values are set with quadratically rising values, resulting in background distributions with a monotonically declining form and sufficient contributions in each bin to assure fit stability. For the resolved signal categories, fifteen bins are considered and for the boosted ones, five bins are considered for rebinning.

- **Background** categories are deprived of the HH signal events. The shape of the DNN distributions of the background categories do not impose much effect on the performance of the analysis sensitivity, but their inclusion in the fit provides constraint on the background normalization. Therefore, the DNN histograms of the background categories are rebinned using quantile binning, where the contributions from all the processes in a background category is summed and then the corresponding cumulative distributions are split into several predefined quantile bins. The resolved and inclusive *i.e.* resolved + boosted Background categories have five quantile bins and the boosted ones have three.

The distributions of the rebinned DNN responses for the SL channel for 2016 era are

shown in Figure 5.5.

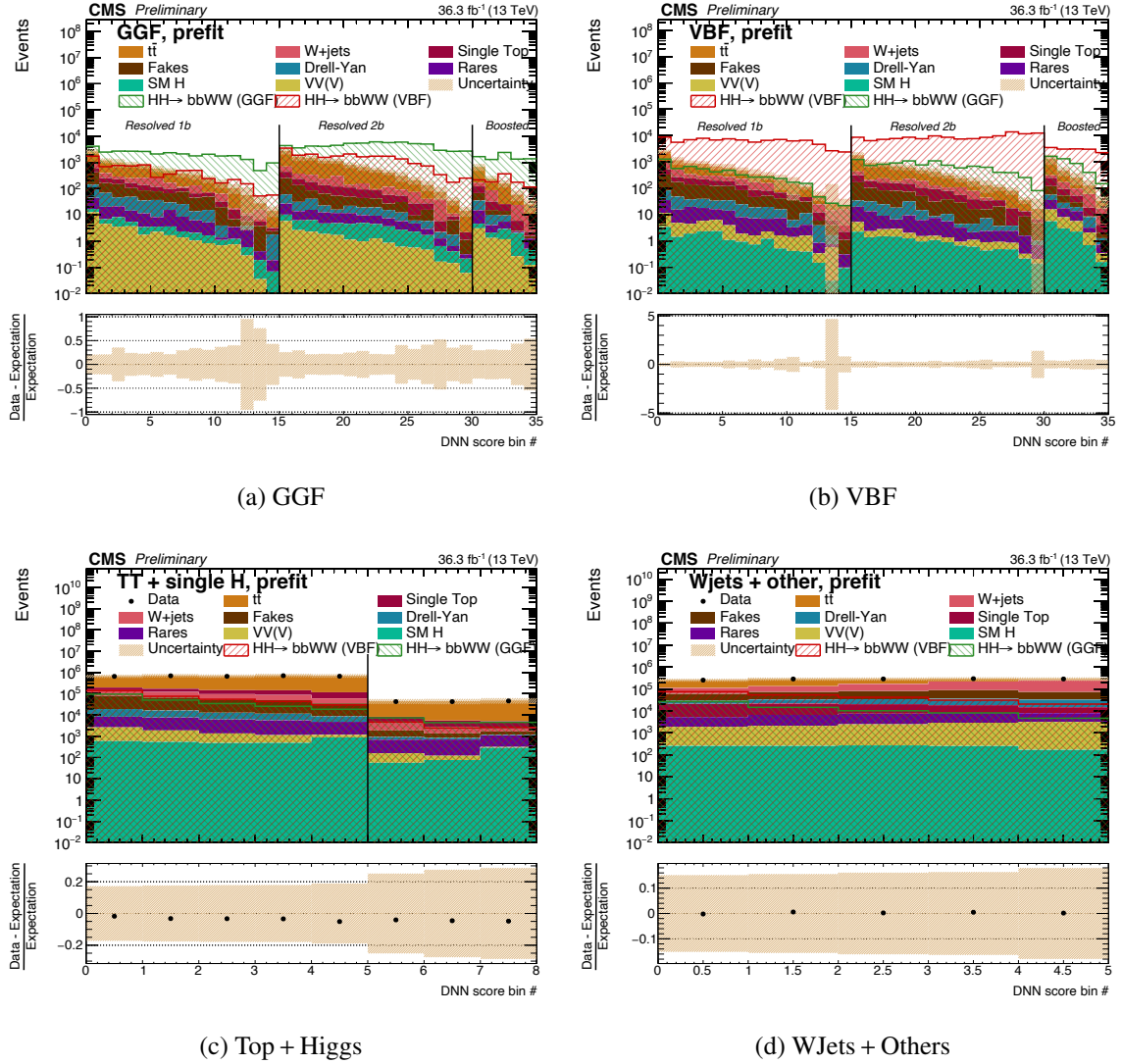


Figure 5.5: Distributions of the DNN responses for the signal and background categories with respect to the bin number for the SL channel for 2016 era only..

5.4 Background Estimation

For the DL channel, the major sources of background in the signal region are $t\bar{t}$, single-top, DY, while $t\bar{t}$, single-top and W +jets backgrounds contribute the most to the SL channel. The contribution of all of these backgrounds are estimated from the Monte Carlo simulation except the DY. For the DL channel, the selection efficiency of the DY events is very

low which affects statistical precision. Therefore, a data-driven method is developed to estimate the DY contribution in the DL channel. In addition, the non-prompt and jet faking leptons are not well modeled in the Monte Carlo simulation and therefore, to extract the contribution of the fake background, a data-driven method is implemented, especially for the SL channel.

The data-driven technique used to estimate the DY contribution is based on the DY control regions (CR). The signal region denoted as “Z-veto” (Green Box) requires $|M_{\ell^+\ell^-} - M_Z| > 10 \text{ GeV}$ and, the CR region, referred as “Z-peak” (Pink box) that requires an inversion of this condition of SR. Additionally, a Z-peak region without any b -jet requirement populates another DY CR (Blue box). The ABCD method (described in Figure 5.6) is used to estimate the weight factors in the DY CRs which are then transferred to the SR. The

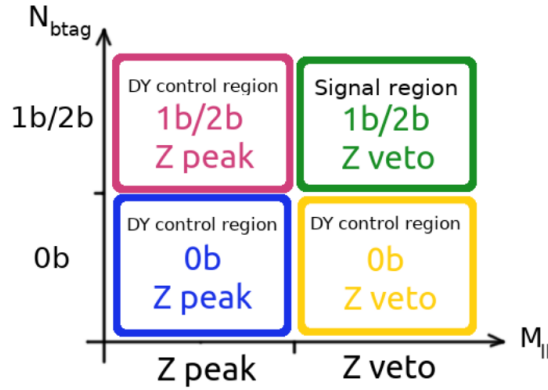


Figure 5.6: A schematic representation of the ABCD data-driven method. Three DY control regions are used to estimate the transfer weight factors and then it is applied in the signal region.

transfer weights can be expressed as:

$$N(x_i|1b/2b, Z - veto) = N(x_i|0b, Z - veto) \frac{N(x_i|1b/2b, Z - peak)}{N(x_i|0b, Z - peak)}, \quad (5.3)$$

where, x is a variable based on which the weight is estimated. i refers to the bin of x that contains N entries of the DY process.

The contribution of the jet faking leptons in the SR is estimated by the Fake Factor method [74]. The idea is to estimate the fake rate of a jet to be considered as a lepton in a

CR, referred as the measurement region (MR). The MR is selected to be a jet enriched region. The measured fake rate is applied in the application region (AR) to extrapolate the contribution of the lepton faking background to the SR. The AR requires that the leptons should pass the fakeable condition, but fail the tight one. The fake factor ($F \equiv f/(1-f)$) is calculated in the MR and then multiplied with the events selected in the AR or analysis side-band. The result is the contribution of fake backgrounds in the SR.

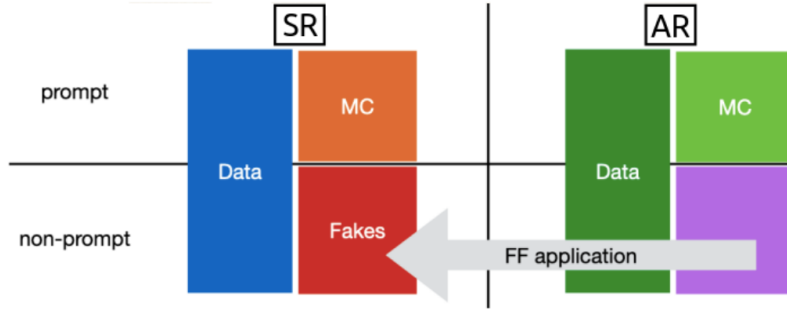


Figure 5.7: The algorithm for application of the fake factor on the non-prompt contribution computed from data and Monte Carlo simulation.

In the analysis, the prompt contribution is computed by matching the fakeable leptons with the generator level information. In AR, the difference between the data and all the prompt Monte Carlo contributions provides the non-prompt contribution in the same region as shown in Figure 5.7 by the purple colored area. The fake factor is applied on those events to estimate the non-prompt fake contribution in the SR as denoted by the red colored area in Figure 5.7.

5.5 Systematic Uncertainties

Several sources of systematic uncertainty are incorporated in the study in the form of nuisance parameters in the maximum likelihood fit. The systematic uncertainties affect both normalization and shape of the final discriminant. The sources of the systematic uncertainties are distributed among three categories described below.

Additional sources systematics are included based on the bin by bin errors of the final

Group	Sources of systematic uncertainty	Prefit yield variations [%] [Background Signal]
Experimental uncertainty	Luminosity Pre-firing Pileup	3.6-4.8 3.6-5.5
	Trigger	1.2-3.4 1.3-3.5
	Electron identification	7.2-8.0 2.8-8.1
	Muon identification	3.0-3.8 2.0-5.0
	Jet PU Id	0.8-4.8 0.8-2.5
	JES JER Unclustered \cancel{E}_T	3.1-19.0 3.5-16.0
	b -tagging	14.5-18.0 9.6-16.15
Data driven uncertainty	DY estimation	9.8-13.2
	Fake estimation	5.6-27.3
Theoretical uncertainty	Branching ratio Event generation	4.0-21.8
	Mass uncertainty EW correction Parton shower Scales $t\bar{t}$ generation	7.6-10.0

Table 5.18: Several sources of systematic uncertainties and their impacts on the event yield. The descriptions are written in Appendix-I 6.5.

discriminant using the Beeston-Barlow light [86] approach due to the small size of the simulated samples used in the statistical inference.

5.6 Statistical Analysis

The final discriminator used in the statistical analysis is the binned DNN response as shown in Figure 5.5. The rates for the non-resonant HH production are determined through a binned simultaneous maximum likelihood fit of the discriminator for each event category in the single lepton and dilepton channels. The categories are motivated by the output classes of the DNN multi-classifiers, although certain processes have been merged to maintain stability in the fit. All the categories on which the likelihood fit is performed are described in Section 5.3.3.

The production rate for the HH signal constitutes the parameters of interest (POI) in the

fit, which is denoted by the symbol μ . Separate fits are performed for each signal model, where one signal model corresponds to a given coupling scenario. The likelihood function (\mathcal{L}) is expressed as:

$$\mathcal{L}(\text{data} | \mu, \theta) = \prod_i P(n_i | \mu, \theta) \prod_k p(\tilde{\theta}_k | \theta_k) \quad (5.4)$$

where the index i refers to individual bins of the distributions in the DNN output and the factor $P(n_i | \mu, \theta)$ represents the probability to observe n_i events in a given bin i where $\nu_i(\mu, \theta)$ events are expected from the sum of signal and background contributions in that bin. The number of expected events is a linear function of the POI:

$$\nu_i(\mu, \theta) = \mu \cdot \nu_i^S(\theta) + \nu_i^B(\theta) \quad (5.5)$$

where the symbols ν_i^S and ν_i^B denote the number of events expected in bin i from the HH signal and the aggregate of contributions expected from background processes in that bin, respectively. It is evident from the expression of $\nu_i(\mu, \theta)$ that the number of events expected from signal and background processes depends on the systematic uncertainties, referred as the nuisance parameters (θ). The probability $P(n_i | \mu, \theta)$ is given by the Poisson distribution:

$$P(n_i | \mu, \theta) = \frac{\nu_i^n(\mu, \theta)}{n_i!} e^{-\nu_i(\mu, \theta)} \quad (5.6)$$

The Individual element of the set of θ is denoted by the symbol θ_k that corresponds to a specific source of systematic uncertainty. The function $p(\tilde{\theta}_k | \theta_k)$ in Equation 5.4 describes the probability to observe a value $\tilde{\theta}_k$ in an auxiliary measurement dedicated for that nuisance, given θ_k as the true value of the parameter. Further details concerning the treatment of systematic uncertainties and the choice of functions like $p(\tilde{\theta}_k | \theta_k)$ are described in References [86–88].

For signal and backgrounds, the corresponding branching fractions are taken into consideration, and therefore, the upper limit on the signal cross section is the limit on the

production cross section of HH . The cross sections and kinematics of the GGF and VBF are dependent on the coupling parameters. The total amplitude of the GGF process is given by $A = \kappa_t \kappa_\lambda \Delta + \kappa_t^2 \square$ and therefore, the cross section can be written as:

$$\sigma(\kappa_t, \kappa_\lambda) \sim |A|^2 = \kappa_t^2 \kappa_\lambda^2 |\Delta|^2 + \kappa_t^4 |\square|^2 + \kappa_t^3 \kappa_\lambda |\Delta^* \square + \Delta \square^*| = c(\kappa_t, \kappa_\lambda)^T \cdot v \quad (5.7)$$

where, Δ denotes the contribution of Figure 2.3 [Left] and \square denotes the contribution of Figure 2.3 [Right]. Here, $c(\kappa_t, \kappa_\lambda) = (\kappa_t^2 \kappa_\lambda^2, \kappa_t^4, \kappa_t^3 \kappa_\lambda)$ is the coupling vector and the contributions of the triangular and box diagrams with their interference is denoted by $|\Delta|^2, |\square|^2, |\Delta^* \square + \Delta \square^*|$. As evident from Equation 5.7, only three sets of couplings $((c_1, c_2, c_3)^T \equiv C)$ and the corresponding cross sections $((\sigma_1, \sigma_2, \sigma_3)^T \equiv \sigma)$ are required to determine $v = C^{-1} \sigma$ fully which allows one to calculate the cross section for any values of the couplings other than the provided ones following:

$$\sigma(\kappa_t, \kappa_\lambda) = c(\kappa_t, \kappa_\lambda)^T C^{-1} \sigma \quad (5.8)$$

Similarly, for the VBF process, six sets of couplings and their cross sections are necessary to fully determine the cross sections of all possible coupling parameter values.

The formulation of Equation 5.7 can also be differential which dictates that any variable x may be used to generate $d\sigma/dx$ at any coupling set from a variety of distributions for which the couplings are known. The interpolation has a limitation that it cannot precisely replicate the shape that was acquired from the generation at one particular coupling point. Fortunately, this limitation has little impact until the required point exhibits more extreme characteristics than those for which the shapes are available. The initial coupling benchmarks must be carefully chosen to achieve maximum kinematic difference in order for events to populate the whole phase space.

The entire HH physics model is implemented in Reference [89] which is included in the ‘‘Inference’’ package [61].

5.6.1 Results

Standard Model Limits

Figure 5.8 demonstrates the upper limit on the signal strength (μ) *i.e.* the ratio of the expected upper limit on $pp \rightarrow HH$ cross section at 95% CL to the SM predicted value, for the SM benchmark point of the $HH \rightarrow b\bar{b}W^+W^-$ signal in all the three eras and for SL, DL and their combination.

The expected upper limit on the SM HH cross section at 95% confidence level is 17 times the SM predicted cross section. Similarly, the upper limit on HH VBF cross section is $255 \times SM$ because of much smaller cross section for the HH VBF production.

EFT Benchmark points

Figure 5.9 illustrates the expected limits for the EFT shape benchmark points. At 95% CL, the upper limits for the benchmarks described in Table 5.4 range from 134 to 1297 fb and from 271 to 725 fb for those in Table 5.5. The variation in the m_{HH} spectra as shown in Figure 8.1 in Appendix 6.5 is the key factor behind the variation in sensitivity for different shape benchmarks. The most sensitive benchmark, labeled as 2, has a two-peak structure and a very broad m_{HH} tail at high values, which makes it simple to detect from the backgrounds, in contrast to benchmark 7, for which the sensitivity is therefore lower. Benchmarks 10 and 12, which are the next in terms of poor sensitivity, also exhibit this, but to a lesser level. The SM point and the upper limit for $\kappa_\lambda = 1$ in Figure 5.10 show the same value.

Coupling Scan

In addition to the SM benchmark cross section study and the EFT benchmarks, it is possible to measure and constrain the coupling modifiers. Each coupling modifier can be

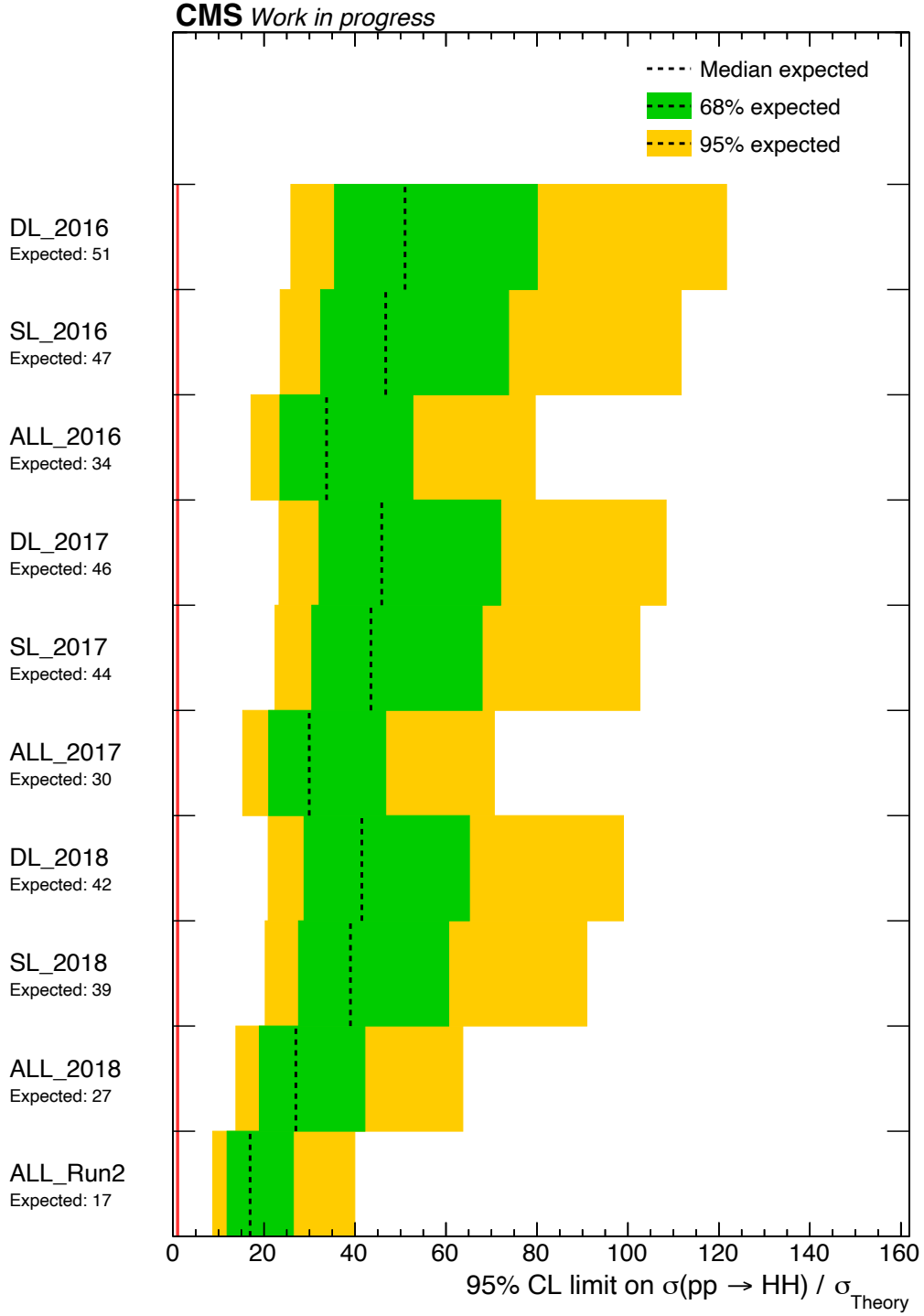


Figure 5.8: 95% CL expected upper limit on the production cross section at $\kappa_\lambda = 1$ (SM case) coupling compared between all three years of DL, SL and the combination of DL and SL. The last row is the combination of all.

scanned keeping the others fixed at their SM value. Figure 5.10 shows the κ_λ scan with the full Run-2 data. The allowed parameter space of κ_λ obtained from the expected upper

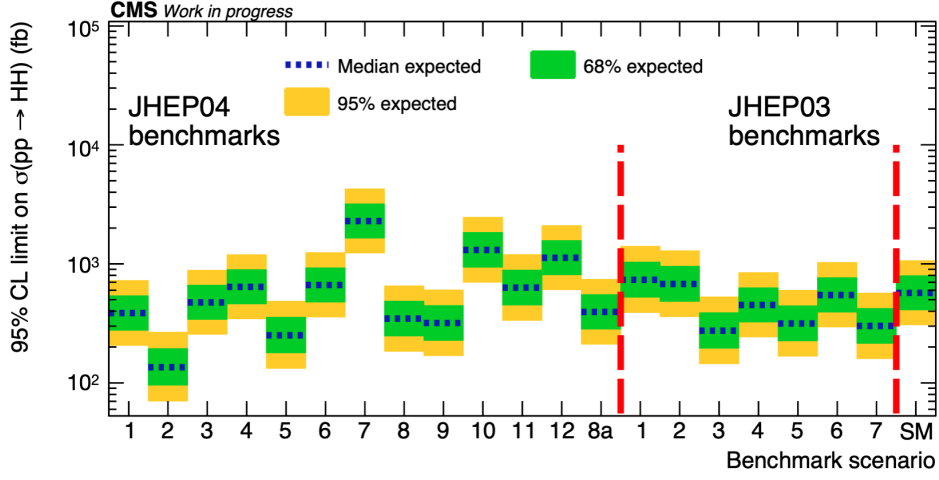


Figure 5.9: For the various benchmark points listed in Table 5.4 (JHEP04) and Table 5.5 (JHEP3), 95% CL expected limits on the cross-section given in fb are presented (right). The SM point gives an upper limit that is equivalent to the actual SM measurement.

limit is $-8.1 < \kappa_\lambda < 14.9$. The variation of cross section with κ_λ also reflects the effect of the destructive interference between the triangular and box diagrams of the HH GGF production mode and, at $\kappa_\lambda = 2.45$, the maximum interference takes place.

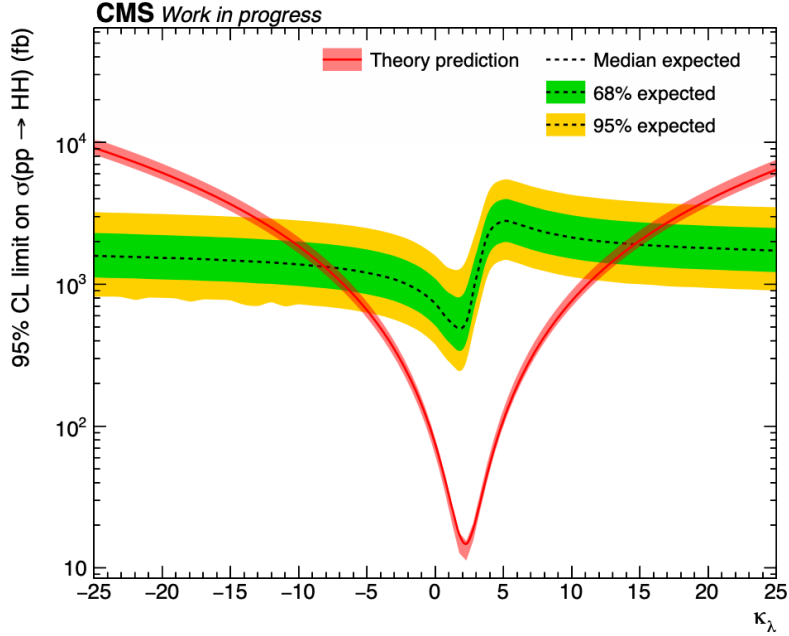


Figure 5.10: The expected upper limit on the HH cross section at 95% confidence level as a function of κ_λ .

Figure 5.11 illustrates the variation of the expected upper limit on VBF HH cross section as a function of the coupling modifier κ_{2V} . The expected allowed region is found to be in

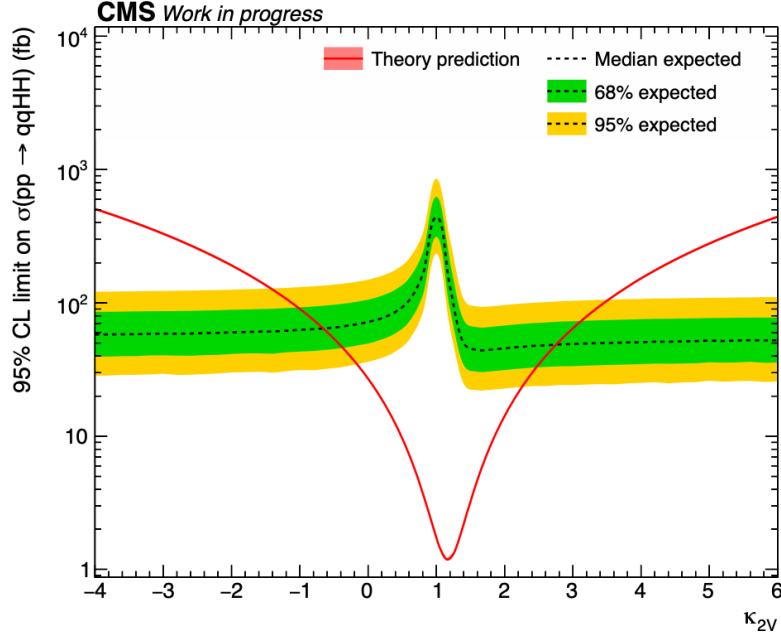


Figure 5.11: The expected upper limit on the VBF HH cross section at 95% confidence level as a function of κ_{2V} .

the range $-0.68 < \kappa_{2V} < 2.8$. The combination of several HH decay modes has already measured $\kappa_{2V} \neq 0$ [24] showing that the $VVHH$ interaction exists in the SM.

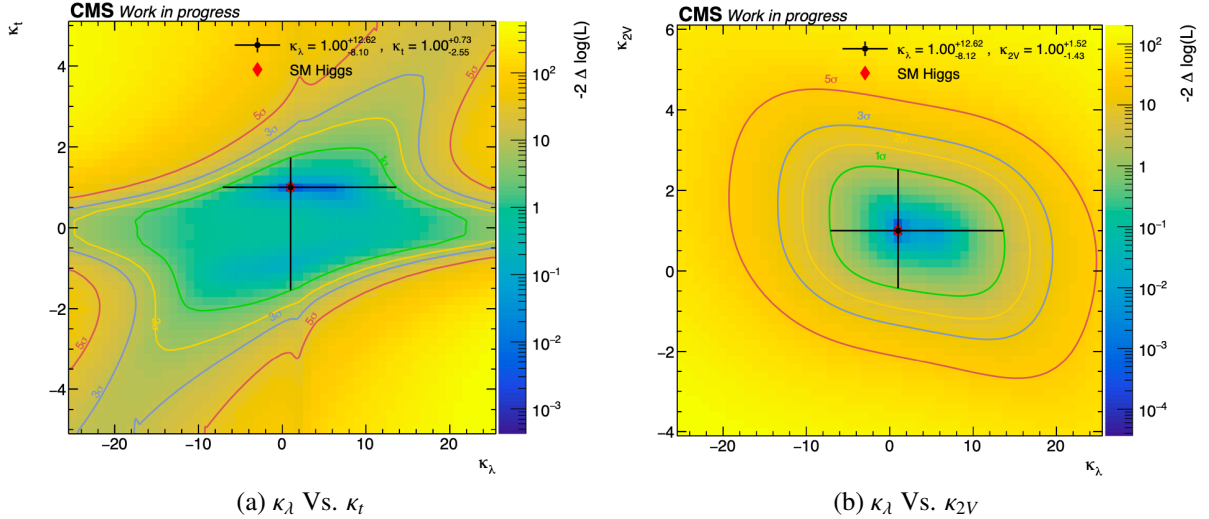


Figure 5.12: Likelihood scans as a function of κ_t , κ_λ and κ_{2V} in the two dimensional plane of these coupling modifiers. The best-fit value is presented as cross-bar, and the uncertainty at various levels on both POIs are displayed as contours.

In Figure 5.12, the two dimensional likelihood scans for κ_λ vs. κ_t and κ_λ vs. κ_{2V} show the best fit values with uncertainty of the respective coupling modifiers. The effect of

$\kappa_\lambda = 2.45$ is visible in Figure 5.12a, but not prominent in Figure 5.12b due to the weaker impact of κ_λ in the VBF process.

Comparison with other HH decay modes

The most three sensitive HH decay channels are $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$ as observed by both CMS and ATLAS. Table 5.19 describes the comparison between $b\bar{b}W^+W^-$ and other decay channels.

Experiment	Channel	Luminosity (fb^{-1})	Observed (Expected)	
			$\sigma_{HH}/\sigma_{HH}^{SM}$	$\sigma_{VBF}/\sigma_{VBF}^{SM}$
ATLAS (Run-1)	Combined [90]	20.3	70 (48)	—
ATLAS (2015-16)	Combined [91]	36.1	6.9 (10)	—
ATLAS (Run-2)	$bbbb$ (VBF) [92]	126	—	840 (550)
	$bb\ell\nu\ell\nu$ [93]	139	40 (29)	—
	$bbbb$ (VBF) [94]	126	5.4 (8.1)	130.5 (133.4)
	$b\bar{b}\gamma\gamma$ [95]	139	4.2 (5.7)	—
	$b\bar{b}\tau\tau$ [96]	139	4.7 (3.9)	—
	Combined [97]	139	2.4 (2.9)	—
CMS (Run-1)	Combined [98, 99]	[17.9, 19.7]	43 (47)	—
CMS (2016)	Combined [100]	35.9	22.2 (12.8)	—
CMS (Run-2)	$bbZZ$ [101]	138	32.4 (39.6)	—
	$WWWW, WW\tau\tau, \tau\tau\tau\tau$ [102]	138	21.3 (19.4)	—
	$b\bar{b}\gamma\gamma$ [103]	138	7.7 (5.2)	225 (208)
	$bbbb$ (resolved) [104]	138	3.6 (7.3)	226 (412)
	$bbbb$ (boosted) [105]	138	9.9 (5.1)	—
	$b\bar{b}\tau\tau$ [76]	138	3.3 (5.2)	124 (154)
	Combined [24]	138	3.4 (2.5)	—
CMS (Run-2)	$b\bar{b}WW$	138	— (17)	— (255)

Table 5.19: Summary of HH production measurements from the CMS and ATLAS collaborations in various decay channels with the integrated luminosity. Upper limits are presented as a multiple of the SM prediction for both the inclusive (mainly GGF) and VBF production processes. ATLAS Run-1 results were obtained by combining $bbbb$, $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, and $WW\gamma\gamma$ (at various integrated luminosity values), while the 2016 combination result (with a small amount of data from 2015) of ATLAS was obtained by combining the same channels with the addition of $b\bar{b}WW$ and $WWWW$. CMS Run-1 results were obtained by merging $bbbb$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau\tau$, with the addition of $b\bar{b}VV$ in 2016. The branching ratios used in each decay channel are taken from the SM alone. The Run-2 combination results are shown in the Run-2 categories combining the processes mentioned above it.

As shown in Table 5.19, the full Run-2 combined limit is two to three times higher the SM prediction. An upper limit on HH signal strength of less than unity will establish the existence of SM HH processes and this expectation requires more data with more decay

channels. $bbWW$ channel is not merged with the full Run-2 combination yet, but it will be included in the next iteration.

Chapter 6

A Phenomenological Study: Search for Exotic Higgs Bosons at the LHC and beyond

The discovery of the Higgs boson at the LHC completes the Standard Model (SM). However, there are several phenomena, like matter-antimatter asymmetry, neutrino mass, existence of dark matter, stability of electroweak potential at very high energy etc. that the SM is unable to explain. Ever since the discovery of the Higgs boson, particle physicists have been studying if more Higgs like scalars exist. If there are more scalars, how they interact with the SM particles. Such questions motivate physicists to hypothesize an underlying extended scalar sector in the BSM framework. One such extension discusses the role of the additional scalars in facilitating explanation to the flavor problem. Specifically, discrete flavor symmetries have been successfully employed to explain the quark and lepton masses and mixing [106–108]. With enlarged scalar spectra, many of these flavor models contain exotic spin-0 states endowed with non-standard couplings to fermions and gauge bosons. The subject matter of this chapter is to search for the exotic scalar (pseudoscalar) states with unconventional couplings at the LHC.

The primary motivation is to study the phenomenology of a class of flavor models based on the S_3 symmetry group that has the ability to extend the scalar states with several exotic scalars with non-standard couplings to the SM particles. However, the analyses have been performed in as much model independent manner as possible. There are three copies of $SU(2)$ doublet $(\phi_{1,2,3})$ Higgs bosons in the S_3 symmetry group, out of which $\phi_{(1,2)}$ form an S_3 doublet and ϕ_3 remains a singlet. A rich scalar spectrum emerges with three CP-even and two CP-odd neutral scalars, and two sets of charged scalars. The details of the minimization of the potential, mass spectra of the scalars (pseudoscalars) and their couplings to the gauge and matter fields can be found in [109, 110]. One of the CP-even neutral scalars turns out to be the SM Higgs, which is denoted by h_{125} . Among the nonstandard states, other than CP-even state (H) and a CP-odd state (χ), the rest may be considered to be sufficiently heavy with couplings to the gauge and matter fields resembling those in the two-Higgs doublet models.

Drawing inspiration from such flavor models with three Higgs doublets and exotic states, no specific model dependency is considered other than two exotic properties of H and χ , namely:

- There are no HVV -type couplings, where $V \equiv W^\pm, Z$. The $H\chi Z$ coupling takes the simple form ($q_\mu \equiv$ momentum transfer):

$$H\chi Z : \left(\frac{-ie}{2 \sin \theta_W \cos \theta_W} \right) q_\mu$$

- $H(\chi)$ has *only* flavor off-diagonal Yukawa couplings. The relevant piece of the Yukawa Lagrangian is

$$Y_{ff'} \bar{f} (i\gamma^5) f' H(\chi) + \text{h.c.}$$

where, $f, f' \equiv e\mu, \mu\tau, e\tau, uc, tc, ut, ds, db, sb$.

Since there is no HVV coupling, neither the LEP2 limit nor the electroweak precision

constraints would apply on the mass of H . The pseudoscalar χ does not couple to VV either. Moreover, since neither H nor χ has any diagonal Yukawa coupling, the usual LHC constraints do not apply on their masses as well. Therefore, both H and χ could be light [109]. Since the choice of their masses would greatly influence the search strategies, it is essential to make a few working assumptions as explained in the next section.

6.1 Signal Processes and Choice of Benchmark Points

The exotic nature of the particles introduced in the previous section with non-standard Yukawa couplings are explored in two separate analyses following two different production mechanisms. The Feynman diagrams in Figures 6.1 and 6.2 show the corresponding processes denoted by “A1” and “A2” in the subsequent discussions.

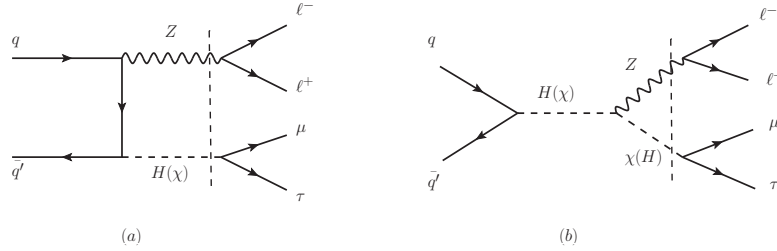


Figure 6.1: Two major Feynman diagrams for the signal process A1 in (a) t -channel and (b) s -channel. The dashed vertical lines are indications of the on-shell production of H/χ in association with Z , and their subsequent decays into $\mu^\pm \tau^\mp$ and $\mu^+ \mu^- / e^+ e^-$, respectively.

H primarily contributes to A1 and the range of m_H is considered to be wide enough to explore the sensitivity of the analysis of A1 in detail. The mass of χ is assumed to be much lighter than H . For A2, The benchmark points are selected assuming $m_H \gtrsim m_t$, so that the top quark does not have a sizable branching fraction into H and a charm quark. As a result, H does not play a significant role for A2. Both the Feynman diagrams in Figure 6.2, show χ only.

The size of the H/χ off-diagonal Yukawa couplings would also impact the search strategies. For any specific flavor symmetry group, the purely off-diagonal Yukawa couplings

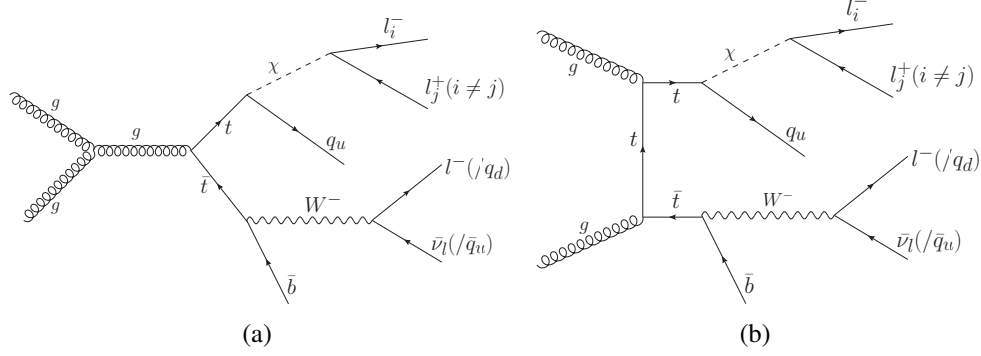


Figure 6.2: The Feynman diagrams for signal events produced from $t\bar{t}$ pair, where one top quark decays to $b/\bar{b}W^\pm$ and the other one to χ and a light flavored jet. W can decay leptonically or to jets and χ decays to a pair of μ and τ .

have a role to play in reproducing the fermion masses and mixing. The available theoretical and experimental constraints are considered before choosing the coupling benchmarks. Extremely tight constraints on Hds and χds couplings from $K_L \rightarrow \mu e$ decays are taken care of, specially when H and χ masses are of $\mathcal{O}(10 - 100)$ GeV [111]. To circumvent this, the off-diagonal Hds and χds couplings are set to zero. The constraints from $D^0 - \bar{D}^0$ mixing is satisfied using the partial cancellation of the contributions from H and χ . Tree level meson mixing amplitude goes as $(\frac{Y_q^H}{m_H^2} - \frac{Y_q^\chi}{m_\chi^2})$, i.e. a scalar and a pseudoscalar contribute with opposite sign (see e.g. [112]), where Huc and χuc couplings are denoted by Y_q^H and Y_q^χ , respectively. For A1, Y_q^χ is scaled automatically by $\approx \frac{m_\chi}{m_H} Y_q^H$ and for A2, Y_q^H is kept at a very low value of 10^{-4} to avoid the constraint. Another off-diagonal Yukawa coupling, Y_{ct}^χ contributes to top quark decay. It plays the main role in A2. The value of Y_{ct}^χ is considered to be within 0.001 to 0.01 to maintain the branching fraction of $t \rightarrow bW$ decay.

In the leptonic sector, non-observation of various lepton flavor violating (LFV) processes, like $\ell_i \rightarrow \ell_j \gamma$ (with $\ell_i \equiv \tau, \mu$ and $\ell_j \equiv \mu, e$), $\mu + N \rightarrow e + N$ (i.e. $\mu - e$ conversion) [113], as well as $e^+ e^- \rightarrow \mu^+ \mu^- (\tau^+ \tau^-)$ put a very strong limit on the product of LFV Yukawa couplings involving the first two generations (e, μ) as a function of m_χ and m_H . To respect these limits, the $H e \mu(\tau)$ and $\chi e \mu(\tau)$ Yukawa couplings are set to tiny values $\mathcal{O}(10^{-9})$, for the range of m_χ and m_H considered in the present analysis. Thus, $\mu^\pm \tau^\mp$ is left as the

dominant leptonic decay mode of $H(\chi)$.

Analysis	Final states	Remarks
A1	$\tau_h + 3\mu$	—
	$\tau_h + \mu + 2e$	—
A2	$b + 3j + \mu + \tau_h$	SL
	$b + j + \mu + \tau_h + \ell(= e, \mu) + \cancel{E}_T$	DL

Table 6.1: *Analysis channels and final state particles.*

Table 6.1 shows the analysis channels under study. For A1, both the final states are analyzed together and for A2, the analysis is divided into two channels, semi-lepton (SL) and double-lepton (DL) following the decay of the pair of W bosons.

Among all the free parameters of the model under consideration, several benchmark points are used for two masses, m_H and m_χ , and for three Yukawa couplings, $Y_{\mu\tau}^\chi$, Y_{ct}^χ and Y_{uc}^H . The benchmark specification for the two analyses are shown in Table 6.2. As ev-

Parameters	A1	A2
m_H (GeV)	140 – 500	160
m_χ (GeV)	20 – 60	20 – 100
Y_{uc}^H	0.001 – 0.01	0.0001
Y_{uc}^χ	$\frac{m_\chi}{m_H} \times Y_{uc}^H$	0.0001
Y_{ct}^χ	—	0.001 – 0.01
$Y_{\mu\tau}^\chi$	0.001 – 0.01	0.005

Table 6.2: *Benchmark points chosen for A1 and A2.*

ident from the table, Y_{ct}^χ is not relevant for A1 as top quark does not have any contribution. Similarly, for A2, negligible strength of Y_{uc}^χ increases the branching ratio of $\chi \rightarrow \mu^\pm \tau^\mp$ and the strength of $Y_{\mu\tau}^\chi$ turns out to be ineffective in overall cross section. The value of $Y_{\mu\tau}^\chi$ is fixed at 0.005.

6.2 Backgrounds

There are several SM processes which constitute the background by imitating the final states of A1 and A2. For all relevant backgrounds, next-to-leading order (NLO) cross section, if they are available in the literature, or the LO cross sections, weighted by the k -factor, are used. The SM background processes used in the analyses are: $Z + jets$, $t\bar{t} + jets$, tV , $t\bar{t}h$, $t\bar{t}V + jets$, $VV + jets$, VVV etc. For A1, $VV + jets$ is the major background whereas, for A2, $t\bar{t} + jets$ turns out to be the dominant SM background. All the background processes with their contribution are shown in Table 6.4 and 6.5.

6.3 Monte Carlo Simulation

All the signal benchmark points and the SM background processes are simulated using MadGraph5_aMC@NLO [114], Pythia-8.2 [58] and Delphes-3.4.2 [115]. For the signal benchmarks the lepton flavor violating (LFV) Yukawa Lagrangian is implemented in FeynRules [116] to generate the Universal FeynRules Output (UFO). The UFO is then interfaced with the event generator MadGraph to generate the signal events at the leading order (LO). The default Standard Model configuration of MadGraph is used to generate the SM background events. To calculate signal and background cross sections, NN23L01 is used as the parton distribution function (PDF) [50]. The parton level events are then passed through Pythia for showering, hadronization and decay. For τ decay the TAUOLA [117] package integrated in the MadGraph is used. To incorporate the detector effects, the resulting events are finally processed through the parametric detector simulation package Delphes-3.4.2 using the default CMS card. In Delphes, the anti- k_T jet clustering algorithm [77] is implemented using the FastJet package [118]. The respective tagging efficiencies for the b and τ -tagged jets have been parametrically incorporated through the default CMS card of Delphes.

6.4 A1: Search for exotic scalars produced in association with a Z boson at the LHC and beyond

Four representative mass benchmark points $(m_\chi, m_H) = (20, 160), (20, 170), (60, 160), (60, 170)$ GeV are first considered. $Y_\ell (Y_{\mu\tau}^X)$ and $Y_q (Y_{uc}^H)$ are varied within the range 0.001 – 0.01 for these mass points. Subsequently, an extended mass benchmark scenario is studied for three representative coupling benchmarks, $Y_\ell = Y_q = 0.003, 0.005, 0.01$ where m_H is varied from 140 to 500 GeV as mentioned in Table 6.2 to assess the effect of systematic uncertainties on the signal significance values.

In the first set of the analysis, a set of acceptance cuts (C0) are applied on all the charged leptons and jets reconstructed in the CMS detector. Several kinematic observables are then constructed and studied for both the signal and backgrounds. Based on the final state composition and distinguishable features of the distributions of kinematic variables for the signal and backgrounds, several baseline selection cuts (C1–C6) are applied in succession to suppress the background contribution.

C0 : This consists of basic selection criteria for e, μ, τ and jets as shown in Table 6.3, where several cuts are applied on the p_T, η of the physics objects to be selected and on the angular separation between the pair of objects, *i.e.* $\Delta R_{i,j} = \sqrt{\Delta\Phi_{i,j}^2 + \Delta\eta_{i,j}^2}$ to keep them isolated.

Objects	Selection cuts
e	$p_T > 10$ GeV, $ \eta < 2.5$, $\Delta R_{e,\mu} > 0.4$
μ	$p_T > 10$ GeV, $ \eta < 2.4$, $\Delta R_{e,\mu} > 0.4$
τ_h	$p_T > 20$ GeV, $ \eta < 2.4$, $\Delta R_{\tau,e/\mu} > 0.4$
Jet	$p_T > 20$ GeV, $ \eta < 4.7$, $\Delta R_{\text{jet},e/\mu} > 0.4$

Table 6.3: Summary of acceptance cuts.

C1 : To ensure the presence of one Z boson, the events are selected with an invariant mass $M_{\ell^+\ell^-}$ close to the Z peak by demanding $|M_{\ell^+\ell^-} - M_Z| < 10$ GeV, where M_Z

is the true Z mass. The same cut has been used to suppress the SM ZZ contribution by rejecting events having more than one Z boson.

C2: As only the leptonic decay of χ , *i.e.* $\chi \rightarrow \mu^\pm + \tau^\mp$ is considered, at least one μ is required to be present.

C3: Presence of three charged leptons in the final state, one μ from χ decay and $e^+e^-/\mu^+\mu^-$ from Z decay, is required.

C4: At least one τ jet (τ_h) needs to be present in the final state.

C5: The μ which is not a decay product of Z , denoted by μ' henceforth, and the τ_h should have opposite charges.

C6: A b jet veto is imposed in the baseline selection to suppress the contributions from the top quark.

Table 6.4 shows the effective cross section (fb) after acceptance and successive pre-selection cuts, C1 to C6, for both signal (a few representative benchmarks) and background, and the last column shows the corresponding number of events at an integrated luminosity of 300 fb^{-1} . The Yukawa couplings Y_{uc}^H and $Y_{\mu\tau}^\chi$ are denoted as Y_q and Y_ℓ .

After the baseline selection, an additional cut on the $\cancel{E}_T < 40$ is applied which completes the cut based analysis for the four representative mass benchmark points mentioned earlier. The final signal significance S is defined in terms of number of signal and background events S and B as,

$$S = \frac{S}{\sqrt{S + B}}, \quad (6.1)$$

where $S(B)$ can be estimated as: $S(B) = \sigma_{S(B)} \times \mathcal{L} \times \epsilon_{S(B)}$, with $\sigma_{S(B)}$, \mathcal{L} and $\epsilon_{S(B)}$ denoting the signal (background) cross section, integrated luminosity and signal (background) selection efficiency, respectively.

In Figure 6.3, the panels (a), (b), (c) and (d) represent signal significances in the $Y_\ell - Y_q$ plane at the 14 TeV LHC with $\mathcal{L} = 300 \text{ fb}^{-1}$ corresponding to $(m_\chi, m_H) = (20, 160), (20,$

CHAPTER 6. A PHENOMENOLOGICAL STUDY: SEARCH FOR EXOTIC HIGGS BOSONS AT THE LHC AND BEYOND

Samples	$\sigma_{\text{prod}} \times$ BR (fb)	Effective cross sections (fb)							Events (300 fb ⁻¹)
		C0	C1	C2	C3	C4	C5	C6	
Signal									
$(m_\chi, m_H) - (Y_\ell, Y_q)$									
(20, 160) – (0.003, 0.001)	2.8	2.708	0.747	0.620	0.367	0.037	0.036	0.035	10.47
(20, 160) – (0.005, 0.005)	66.89	64.799	17.894	14.865	8.809	0.895	0.874	0.845	253.53
(20, 160) – (0.009, 0.007)	132.82	128.59	35.472	29.428	17.403	1.818	1.772	1.725	517.6
(20, 170) – (0.003, 0.001)	2.3	2.238	0.627	0.522	0.310	0.037	0.036	0.035	10.54
(20, 170) – (0.005, 0.005)	55.44	53.945	15.278	12.748	7.493	0.89	0.878	0.845	253.47
(20, 170) – (0.009, 0.007)	110.16	107.19	29.999	25.015	14.718	1.772	1.737	1.667	500.18
(60, 160) – (0.003, 0.001)	2.98	2.89	0.701	0.637	0.464	0.079	0.077	0.074	22.3
(60, 160) – (0.005, 0.005)	53.55	51.914	12.508	11.361	8.297	1.38	1.353	1.304	391.08
(60, 160) – (0.009, 0.007)	114.44	110.88	26.875	24.375	17.768	2.981	2.921	2.82	845.94
(60, 170) – (0.003, 0.001)	2.29	2.231	0.58	0.523	0.368	0.066	0.065	0.063	18.75
(60, 170) – (0.005, 0.005)	42.83	41.717	10.802	9.737	6.887	1.203	1.181	1.142	342.55
(60, 170) – (0.009, 0.007)	93.2	90.751	23.49	21.222	14.988	2.669	2.619	2.521	756.18
SM backgrounds									
Z + jets	6.33×10^6	6.32×10^6	2.9×10^5	1.8×10^5	12.74	0.11	0	0	0
$t\bar{t}$ + jets (2ℓ)	1.09×10^5	1.09×10^5	1522.34	967.5	3.58	0.1	0.03	0.03	9.11
$t\bar{t}W^\pm$ + jets	253.8	253.78	1.125	0.779	0.22	0.013	0.005	0.001	0.43
$t\bar{t}Z$ + jets	240.3	240.3	57.68	39.79	11.86	1.193	0.536	0.141	42.15
$W^\pm Z$ + jets (3ℓ)	2273	2263.6	849.86	614.95	389.99	3.67	1.207	1.144	343.17
$W^\pm Z$ + jets (2ℓ)	4504	4496.3	1220.17	769.65	0.18	0.007	0.002	0.002	0.55
ZZ + jets (4ℓ)	187.3	186.46	71.86	51.86	26.89	2.106	1.286	1.254	376.34
[GGF] ZZ (4ℓ)	14.82	14.476	2.16	1.68	0.92	0.01	0.003	0.002	0.73
[VBF] ZZ (4ℓ)	2.211	2.21	0.32	0.24	0.13	0.003	0.001	0.001	0.28
WWW	236.2	236.07	0.6	0.39	0.08	0	0	0	0
WWZ	188.9	188.75	4.84	3.24	1.0	0.07	0.038	0.034	10.2
WZZ	63.76	63.65	3.036	2	0.46	0.025	0.01	0.009	2.64
ZZZ	15.8	15.73	1.08	0.69	0.06	0.007	0.004	0.003	0.95

Table 6.4: The signal and SM background effective cross sections (fb) after each successive baseline cut (C0-C6) and final event yields for $\mathcal{L} = 300 \text{ fb}^{-1}$ at 14 TeV LHC run. Signal event samples are generated for a few representative values of m_χ and m_H (in GeV) and for a range of Y_ℓ and Y_q . Signal cross sections are calculated at LO, $t\bar{t}$ + jets cross sections at $N^3\text{LO}$, while the other SM backgrounds are estimated at NLO.

170), (60, 160) and (60, 170) GeV, respectively. As can be inferred from the plots, for a given value of m_χ and m_H , the significance increases with increasing Y_ℓ and Y_q because of the functional dependence of the signal cross section on Y_ℓ and Y_q .

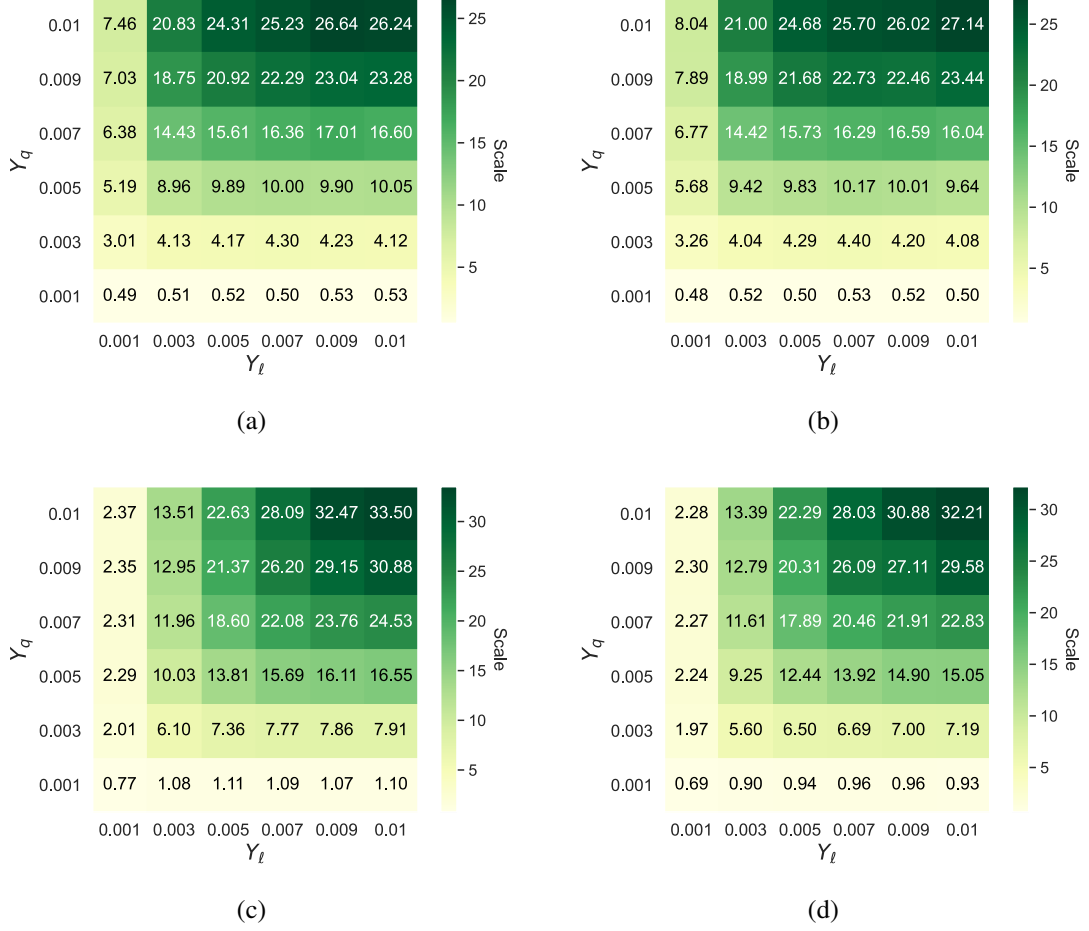


Figure 6.3: Significance plots for the four mass benchmark configuration points at $\mathcal{L} = 300 \text{ fb}^{-1}$ following the cut based analysis. The two plots in the upper panel are for (m_χ, m_H) in GeV : (a) (20, 160), (b) (20, 170). The lower panel contains the other two mass points, in GeV, (c) (60, 160) and (d) (60, 170).

The cut based analysis shows that for all of the four mass benchmark points, the signal significance $\mathcal{S} > 5\sigma$ is achievable for Y_ℓ (Y_q) as low as 0.001 (0.003). By proper scaling one can obtain the signal significance at higher luminosities. For example, by looking at Figure 6.3 for $(Y_\ell, Y_q) = (0.001, 0.001)$, it is evident that the required luminosity for 5σ significance is higher than 3000 fb^{-1} . For the benchmarks with lighter pseudoscalar as shown in Figures 6.3a and 6.3b, we find that the required luminosity for 5σ significance for $(Y_\ell, Y_q) = (0.001, 0.003)$ is $\mathcal{L}_{5\sigma} \sim 800 \text{ fb}^{-1}$. Similarly, for heavier χ shown in Figures 6.3c and 6.3d, $\mathcal{L}_{5\sigma} \sim 2000 \text{ fb}^{-1}$ is required to achieve similar significance with the same set of couplings.

A broader perspective of the analysis is presented employing an extended benchmark region as mentioned in Table 6.2, including possible impact of systematic uncertainties at the HL-LHC. On top of the selection implemented so far, another cut on ΔR between the μ' and τ_h is applied. A comparison between Figures 6.4a and 6.4b shows that for higher

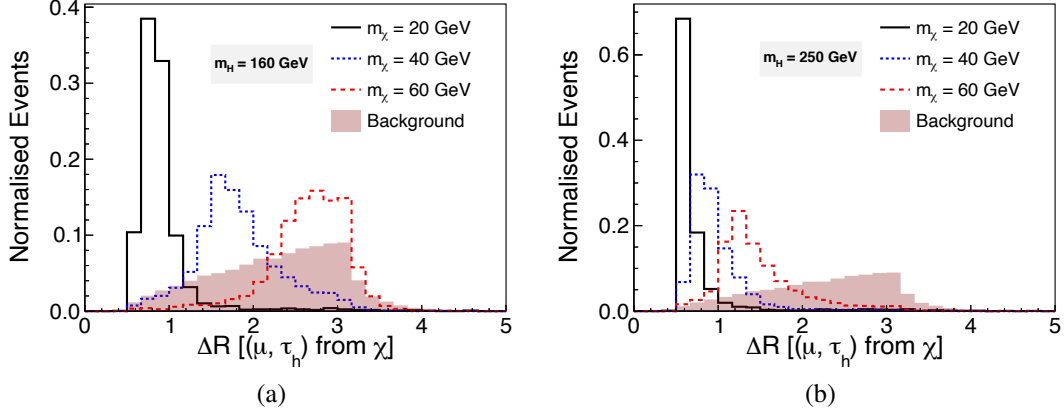


Figure 6.4: $\Delta R(\mu', \tau_h)$ for m_H (in GeV) = (a) 160, (b) 250 with different choices of m_χ .

values of m_H (250 GeV) a uniform cut on ΔR between μ' and τ_h can be applied regardless of the values of m_χ in the given range, while for relatively lighter m_H different values of m_χ require different ΔR cuts. The boost of the decay products from H is the crucial factor here. For $m_H \geq 250$ GeV, a uniform cut $\Delta R \leq 2$, as the best possible choice, is applied to improve the signal significance.

Figure 6.5 describes the impact of including the background only systematic uncertainties by introducing a parameter (α) in the modified expression of signal significance $\mathcal{S} = S / \sqrt{S + B + (\alpha B)^2}$ [119], where α is varied in the range of 0-0.02. Despite the drop in signal significance, for certain parameter choices the significance remains quite promising even after including a 10 – 20% systematic uncertainties.

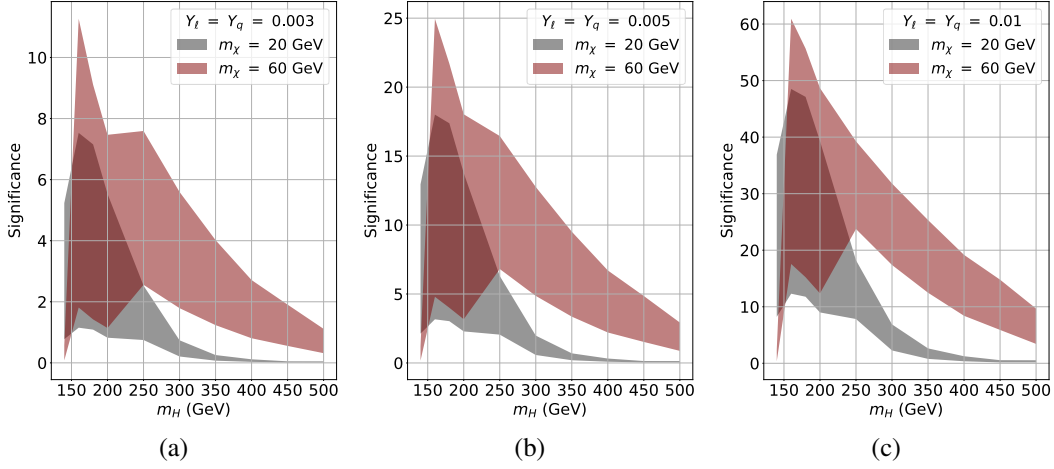


Figure 6.5: Variation of significance with m_H for $m_\chi = 20, 60$ GeV and $Y_\ell = Y_q$ (a) 0.003 (b) 0.005 and (c) 0.01 by scanning α in the range $(0 - 0.2)$ for $\mathcal{L} = 1000 \text{ fb}^{-1}$. For each colored shade, the outer and inner edges correspond to $\alpha = 0$ and 0.2, respectively.

6.5 A2: Searching for exotic Higgs bosons from top quark decays at the HL-LHC

It is expected from the final state topology of A2 that a higher background expectation has to be dealt with. Therefore, the search strategy is different from A1. As mentioned in Table 6.2, the contribution of H and $Y_{uc}^{H(\chi)}$ are negligible here. m_χ is considered to be in the range of 20 – 100 GeV and Y_{ct}^χ in the range of 0.001 – 0.01.

After the simulation of signal benchmarks and relevant SM background events, a few pre-selection cuts are applied to reduce the background contribution to the signal regions of SL and DL channels. Next, instead of the traditional cut based analysis, three multi variate analysis (MVA) methods are used for better discrimination between signals and backgrounds. The pre-selection cuts are described in the following:

C0: In addition to the acceptance cuts described in the C0 selection of Section 6.4, a few conditions like: $p_T > 20$ GeV, $|\eta| < 2.5$ and $\Delta R_{b\text{-jet}, e/\mu} > 0.4$ are imposed to select the b-tagged jets.

C1: In both the DL and SL channels, χ always decays to $\mu^\pm \tau^\mp$. But W decays

leptonically (hadronically) for the DL (SL) channel. Considering the decay of τ -lepton only to hadrons, final state with at least one μ ($e\mu/\mu\mu$) is ensured for DL, whereas for the SL channel, exactly one μ and no e in the final state is required.

C2: Exactly one τ jet i.e. τ_h is required for both channels.

C3: One of the pair produced top quarks in the signal decays to $b(\bar{b})W^-(W^+)$ and the other one to χ and a light jet. So, one b tagged jet is required to be present in the final states of both channels. Apart from that, a cut on the number of light jets is also applied. For DL channel, at least one light jet in the final state is demanded, but for SL analysis, because of the hadronic decay of W , the final state requires to be consisted of minimum three light jets.

C4: In the DL channel, the signal topology does not allow for a pair of opposite sign same flavor (OSSF) leptons in the final state arising out of the decay of Z -boson. Thus to exclude the Z -peak, events with a pair of OSSF leptons having an invariant mass $M_{\ell^+\ell^-}$ within a 10 GeV window of Z mass are vetoed.

C5: The next step is to select the leptons and the jets coming from W boson in the DL and SL channel respectively. As the decay of χ ensures the presence of one μ , the two possible combinations of leptons in the final state are $\mu\mu$ or μe where the second lepton originates from the decay of W in DL channel. While the selection of the μe final state is trivial, it is quite challenging to tag the correct muon coming from χ or W . By looking at the distribution of ΔR between τ_h and μ for the μe combination, it is observed that in most of the signal events, spatial separation between μ coming from χ and τ_h is smaller than that of τ_h and e originating from W . Although this feature is much prominent for the benchmark with the lowest m_χ , but still it can be taken into consideration for the other two benchmarks as well. Now, this nature of ΔR should be the same for $\mu\mu$ scenario. So, the first handle is to choose the μ closer to the τ_h . The second handle is to check the charges of the μ and τ_h and if they are oppositely charged, the μ is considered as the decay product

of χ . Conversely, if they have same charges, the event is not rejected on the basis of that. The τ_h and the distant μ are checked if they have opposite charge and if it is so, the second μ is considered as the decay product of χ and the first μ as the one coming from W . This algorithm does not guarantee the selection of the perfect combination, but it makes the selection more efficient. The μ coming from χ is denoted as μ_χ throughout the rest of the discussion.

For SL channel, it is hard to differentiate among the jets that are coming from t/\bar{t} or W at first sight. At the beginning, the invariant mass of each jet pair is formed out of all three jets. Then for each event the jet pair having invariant mass closest to the W (within a 30 GeV window of the W mass) is chosen as the jet pair to be originated from W . From the remaining jets, the leading one is selected as the light jet coming from t/\bar{t} . At the end, like the DL channel, the μ originated from χ and the τ_h are required to have opposite charges.

Table 6.5 shows the yields for the relevant SM backgrounds and a few representative signal benchmarks after the pre-selections.

Next three different MVA techniques, Decorrelated Boosted Decision Tree (BDTD) [122], Extreme Gradient Boost (XGBoost) [75] and Deep Neural Network (DNN) [123] are used as binary classifiers to extract the contribution of the signals. The same set of input variables are used for all of the three methods and the important ones are tabulated in Table 6.6. Any kinematic variables related to the reconstructed mass of the unknown BSM particles are avoided to keep aside direct model dependence. Before training all of the decision trees and neural network, 75% of the signal and background events is used for training purpose and rest of the events are used for testing the performance of the networks.

The entire algorithm of BDTD is executed within the *Toolkit for Multivariate Data Analysis* (TMVA) framework [124]. In BDTD, “Adaptive Boost” plays a crucial role for robust and efficient classification. To achieve an optimum performance for each benchmark

Process	cross section (pb)	Yields ($\mathcal{L} = 3 \text{ ab}^{-1}$)	
Signals			
(m_χ, Y_{ct}^χ)	NLO ¹ [DL SL]	DL	SL
20, 0.005	0.007 0.022	630	2627
20, 0.01	0.028 0.086	2330	9385
60, 0.005	0.01 0.03	1977	7247
60, 0.01	0.04 0.12	7966	28892
100, 0.005	0.006 0.016	1364	4778
100, 0.01	0.023 0.07	5312	19330
SM Backgrounds			
$t\bar{t} \rightarrow 2\ell + jets$	107.65 [NNLO][119]	954005.17	3891518.73
$t\bar{t} \rightarrow 1\ell + jets$	437.14 [NNLO][119]	8445.99	8709994.85
tW	34.81 [LO]	2463.11	23907.39
$Z \rightarrow \tau^+\tau^- + jets$	803 [NLO]	419.76	20642.35
$t\bar{t}W \rightarrow \ell\nu + jets$	0.25 [NLO]	2113.22	8566.4
$t\bar{t}W \rightarrow qq$	0.103 ² [LO]	207.74	2410.91
$t\bar{t}Z \rightarrow \ell^+\ell^- + jets$	0.24 [NLO][120]	4228.31	11606.32
$t\bar{t}Z \rightarrow qq$	0.206 ² [NLO][120]	404.22	4729.34
$WZ \rightarrow 3\ell\nu + jets$	2.27 [NLO][121]	2404.83	2688.63
$WZ \rightarrow 2\ell 2q$	4.504 [NLO][121]	1275.99	13659.67
$ZZ \rightarrow 4\ell$	0.187 [NLO][121]	169.42	95.15
$t\bar{t}h \rightarrow \tau^+\tau^-$	0.006 ² [LO]	254.85	661.17
$b\bar{b}\tau^+\tau^-$	0.114 ² [LO]	36.89	728.65
WWW	0.236 [NLO]	62.0	439.92
WWZ	0.189 [NLO]	47.6	510.03
WZZ	0.064 [NLO]	22.48	159.66
ZZZ	0.016 [NLO]	3.42	18.83
Total background		976565	12692338

¹ Signal processes are generated at leading order but their cross sections are scaled at next to leading order by multiplying a k-factor of 1.5 [[119](#)]

² Some selections are applied at the generation (i.e. MadGraph) level. p_T of jets (j) and b quarks (b) > 20 GeV, p_T of leptons (ℓ) > 10 GeV, $|\eta|_{j/b} < 5$, $|\eta|_\ell < 2.5$ and $\Delta R_{jj|\ell\ell|j\ell|b\ell} > 0.4$.

Table 6.5: Event yields at $\mathcal{L} = 3 \text{ ab}^{-1}$ after baseline selection for DL and SL channel.

of both DL and SL channels, we adjust the parameters of BDTD as described in Table 6.7. XGBoost is another tree based method like BDTD with some additional advantages. Unlike BDTD, it uses “Gradient Boost” for classification. To reduce over fitting, some additional parameters are used for pruning a decision tree and regularizing the cost function defined as the difference between the true and predicted output. For details of these training parameters, one can refer to the official documentation of XGBoost [[75](#)]. Table 6.8 shows the set of XGBoost hyper parameters used for training the signal and background samples. The last MVA technique which we have tried is DNN. Unlike decision trees, DNN brings in multiple hidden layers with multiple nodes. Non-linear activation

Variables	Description	DL	SL
$p_T^{\tau_h(\tau)}, \eta^{\tau_h(\tau)}$	p_T and η of $\tau_h(\tau)$ jet for DL(/SL)	✓	✓
p_T^{bj}, η^{bj}	p_T and η of leading b tagged jet	✓	✓
\vec{E}_T	Missing transverse energy	✓	✓
p_T^{Wj1}, η^{Wj1}	p_T and η of leading jet from W	✗	✓
$p_T^{\mu_\chi}, \eta^{\mu_\chi}$	p_T and η of μ coming from χ	✓	✓
$p_T^{W_\ell}, \eta^{W_\ell}$	p_T of lepton coming from W	✓	✗
$\vec{p}_T^{\mu_\chi} + \vec{p}_T^{W_\ell}$	Vectorial sum p_T of leptons from χ and W	✓	✗
$\Delta R_{\mu_\chi, W_\ell}$	ΔR between leptons coming from χ and W	✓	✗
$\Delta\phi_{bj, W_\ell}$	$\Delta\phi$ between lepton coming from W and lead b jet	✓	✗
h_T	Scalar sum p_T of all jets	✓	✓
$\Delta R_{\mu_\chi, jets}^{min}$	Minimum ΔR between jets and μ from χ	✓	✓
$\Delta R_{W_\ell, jets}^{min}$	Minimum ΔR between jets and the lepton from W	✓	✗
$\Delta R_{\mu_\chi, \tau_h}$	ΔR between μ and τ_h coming from χ	✓	✓
ΔR_{jets}^{min}	Minimum ΔR between jets	✓	✓
ΔR_{jets}^{max}	Maximum ΔR between jets	✗	✓
$m_T^{\vec{E}_T, W_\ell}$	m_T of lepton from W and \vec{E}_T	✓	✗
$\Delta\phi_{\tau_h, bj}$	$\Delta\phi$ between τ_h and leading b jet	✓	✗
$\Delta R_{Wj1, Wj2}$	ΔR between the two jets from W	✗	✓
$m_{lj, \mu, \tau}^{inv}$	Invariant mass of reconstructed τ, μ and light jet from t/\bar{t}	✗	✓
$\Delta\phi_{lj, bj}$	$\Delta\phi$ between leading b jet and light jet	✓	✓
$\Delta R_{\tau_h, lj}$	ΔR between τ_h and leading light jet	✓	✓
$\sqrt{\hat{s}_{min}}$	Minimum parton level center-of-mass energy	✓	✓

Table 6.6: Important input variables used for all of three MVA methods

Parameters	Description	Benchmarks		
		$m_\chi = 20$ GeV	$m_\chi = 60$ GeV	$m_\chi = 100$ GeV
n_trees	Number of trees	250	250	250
max_depth	Maximum depth of a Decision Tree	2	2	2
boost	Boosting mechanism for training	AdaBoost	AdaBoost	AdaBoost
n_cuts : SL/DL	Number of iteration to find the best split	50/50	46/31	45/40
min_node_size	Minimum events at each final leaf	2.5%	2.5%	2.5%

Table 6.7: Different BDTD parameters used for three different benchmark points.

Parameters	Description	Value
booster	Tree based learner	gbtree
n_estimators	Number of decision trees	auto
max_depth	Maximum depth of a Decision Tree	3
η	Learning rate	0.01
λ	Regularization parameter	0.01

Table 6.8: Description of XGBoost parameters

functions at the nodes help to draw non-linear boundary on the plane of input features to separate signal events from backgrounds. The complete DNN training has been performed using `keras` module of `tensorflow-2.3.0` [125]. All the hyper parameters used for DNN are tabulated in Table 6.9.

Parameters	Description	Value
n_hidden layers	Number of hidden layers	5
n_nodes	Number of neurons in hidden layers	512, 256, 128, 54, 8
activation_func	Function to modify outputs of every nodes	<i>LeakyRelu</i>
loss_function	Function to be minimized to get optimum model parameters	<i>binary_crossentropy</i>
optimizer	Perform gradient descent and back propagation	<i>Adam</i> [126]
eta	Learning rate	0.001
batch_len	Number of events in each mini batch	3000
batch_norm	Normalization of activation output	<i>True</i>
dropout	Fraction of random drop in number of nodes	20%
L2-Regularizer	Regularize loss to prevent over-fitting	0.001

Table 6.9: Summary of DNN parameters

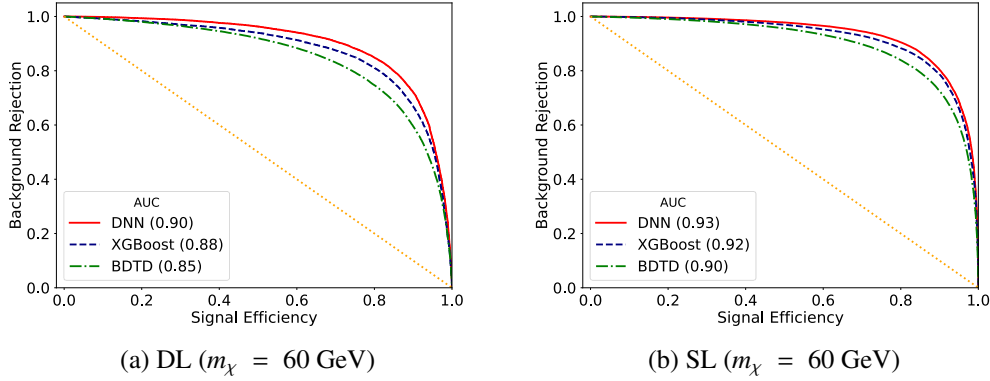


Figure 6.6: ROCs of three MVA techniques for $m_\chi = 60 \text{ GeV}$

These three MVA techniques discussed above deliver somewhat similar performances. In order to compare their responses, the ROC (receiver operating characteristic) curves for all three methods are plotted and the areas under the curve (AUC) of the ROCs are computed as shown in Figure 6.6. The comparison shows that the DNN performs better in both DL and SL channels because of the higher AUC than BDTD and XGBoost. Here in Figure 6.6, the comparison is shown only for $m_\chi = 60 \text{ GeV}$, but for the other two mass points, similar behaviors are observed.

Next, the required luminosity to achieve a 5σ discovery and 2σ exclusion are estimated using the DNN to probe the entire range of the Yukawa coupling between χ and last two generations of quarks, *i.e.* Y_{ct}^χ . The effect of introducing a 5% uncertainty¹ in total number of background events, is also estimated. Apart from the significance, the quantity $\frac{S}{B}$ after applying cut on the DNN response also plays a crucial role here. The benchmark point with $m_\chi = 20$ GeV shows a better sensitivity even after incorporating a 5% uncertainty because of the higher value of $\frac{S}{B}$ than the other two benchmarks. Signal with $m_\chi = 20$ GeV has quite distinguishable kinematic features than the other benchmarks, and therefore, all of the three MVA techniques perform much better for the former one than the other two. In Figure 6.7, a 2σ and 5σ contours are drawn in the integrated luminosity and Y_{ct}^χ plane with and without considering 5% linear-in-background systematic uncertainty. In Figure 6.7a and 6.7b, for $m_\chi = 20$ GeV, the 2σ and 5σ contours drawn with 5% systematic uncertainty are shifted towards the higher values of Y_{ct}^χ , thus requiring more luminosity, for the SL channel relative to the DL channel. Benchmarks with $m_\chi = 60$ GeV and $m_\chi = 100$ GeV are accessible to $2\sigma/5\sigma$ significances only with negligible systematic uncertainties. Therefore, low m_χ provides promising avenues for exploration both for the DL and SL channels at the HL-LHC. Increasing m_χ lowers the search prospect.

In the two analyses presented in this chapter, the existence of light $\mathcal{O}(10 - 100)$ GeV exotic CP-even (H) and CP-odd (χ) states having purely off-diagonal Yukawa couplings have been explored through a resonance signature at the LHC. Both the analyses are performed as much as possible in a model independent way. The mass of the exotic scalars and their purely off-diagonal couplings are scanned over a wide range. The first analysis where χ is produced in association with a Z , shows prospects of high signal significance. Several experimental factors are not taken into account, like jet faking as τ_h and/or leptons, lepton charge misidentification, photon conversions into lepton pairs, uncertainties on luminosity and trigger efficiencies, etc. The promising results encourage

¹In presence of uncertainty the modified significance looks like : $\mathcal{S} = \frac{S}{\sqrt{B + (0.01 \times \theta \times B)^2}}$, where θ is the uncertainty (in %) in the number of background events.

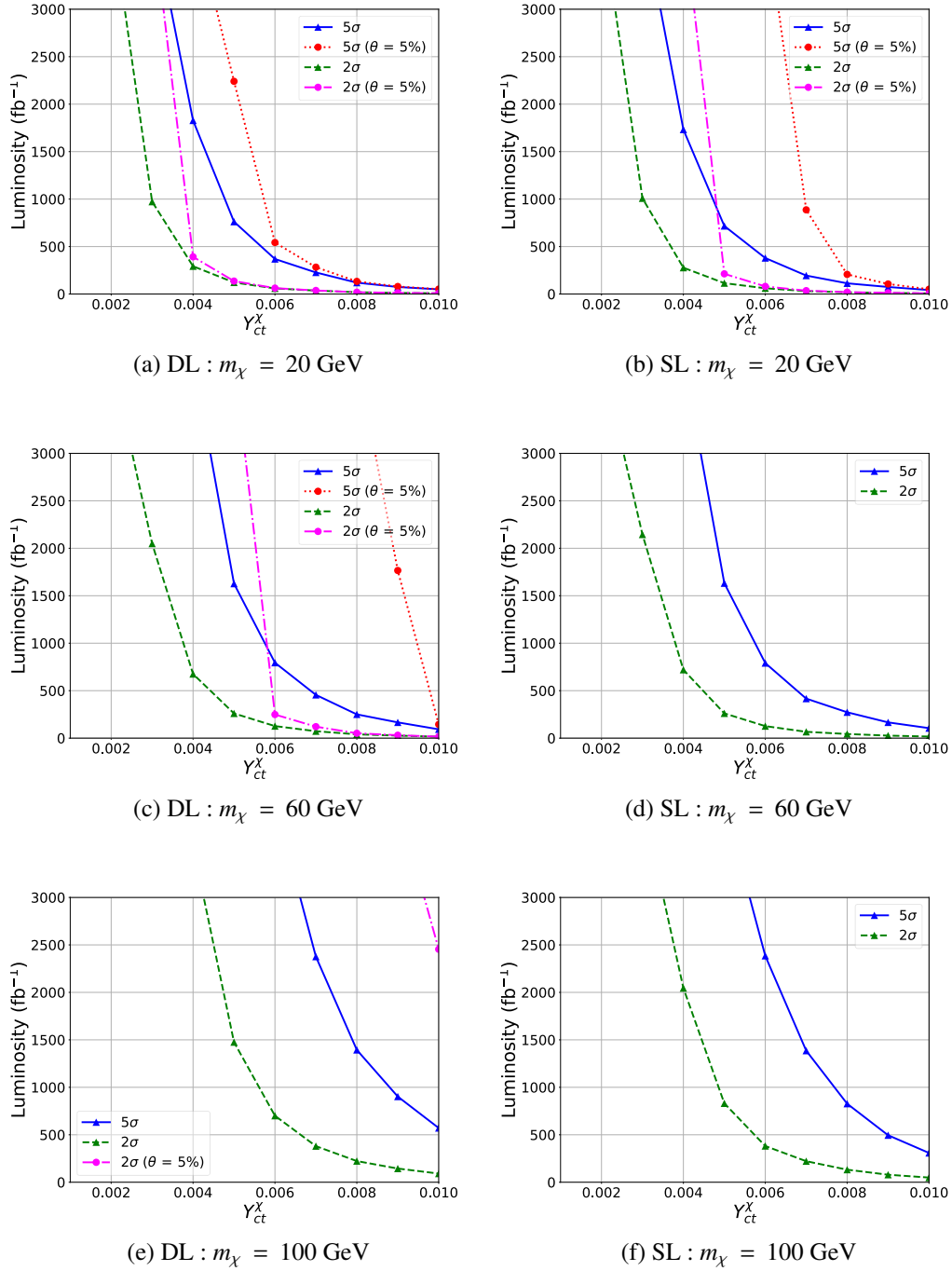


Figure 6.7: Required integrated luminosity to achieve a 2σ exclusion and 5σ discovery. θ is the amount of uncertainties on the backgrounds.

one to take up the studies with LHC Run-2 and Run-3 data.

Chapter 7

Performance of 2S Modules of the CMS Phase-2 Tracker in a Test Beam Environment

Silicon detectors are widely used in particle physics experiments for charged particle tracking produced in collisions. They provide precise position measurements of a charged particle as it passes through the detector in a magnetic field, allowing determination of track momentum with very high resolution. Due to the excellent position resolution, silicon detectors are very useful to determine the location of the primary as well as secondary vertices.

The present CMS tracking system, consisting of silicon pixel and strip detectors, is designed to operate efficiently at an instantaneous luminosity of $\sim 1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, with an average pile-up of 20 – 30 and up to an integrated luminosity (\mathcal{L}) of 500 fb^{-1} . The LHC has successfully delivered a large amount of data in Run-1 and Run-2 during 2010-18. After a shutdown for three years, the Run-3 has started in 2022 and it is expected to operate till the end of 2025 as shown in Figure [7.1](#).

The LHC machine will be upgraded during the third long shutdown (LS3) period (2026-

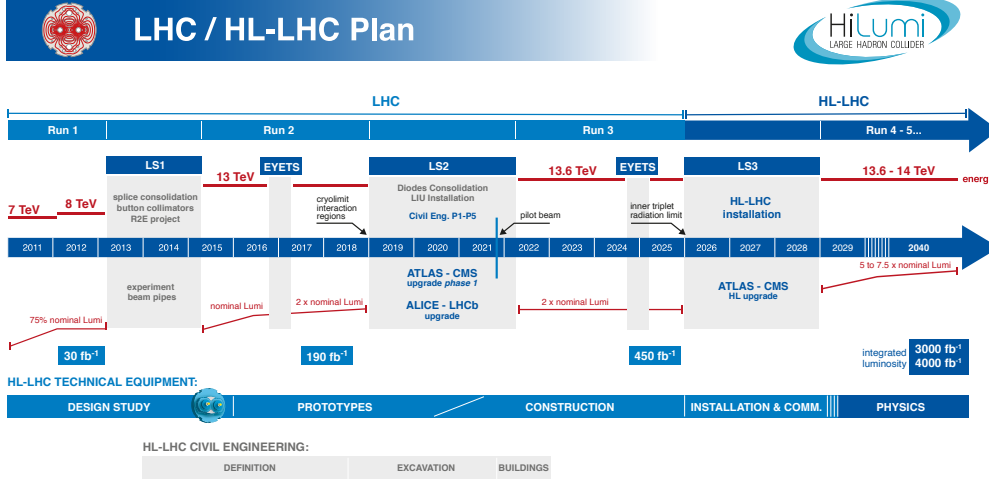


Figure 7.1: Complete LHC/HL-LHC plan

28) and after that the LHC will enter the high luminosity (HL-LHC) phase. At the HL-LHC, the instantaneous luminosity will be 5 – 7.5 times higher than the present value and 140 – 200 collisions per bunch crossing will take place. The CMS experiment is expected to collect a total data of $\mathcal{L} = 3000 - 4000 \text{ fb}^{-1}$ in ten years at the HL-LHC. The present CMS tracker will not be able to cope with the HL-LHC conditions and will be replaced with a new, improved one. The new CMS tracker for the HL-LHC will be more radiation tolerant, have lower material budget, extended tracking acceptance, higher granularity and robust pattern recognition. For the first time in CMS, the tracking information will be used in the level-1 (L1) trigger [127].

7.1 The CMS Phase-2 Tracker

Figure 7.2 shows the layout of the proposed Phase-2 tracker in the $r - z$ plane. The full tracking system is divided into two parts, the inner tracker (IT) and the outer tracker (OT). The IT, which will be equipped with silicon pixel detectors, will consist of four barrel layers that cover a radial region $\sim 30 \text{ mm} < r < \sim 160 \text{ mm}$ and twelve discs at both sides of the barrel region, extending up to $z \sim 2500 \text{ mm}$ from the collision point. The OT

will be built with two types of modules having silicon macro-pixels and micro-strips. It contains six barrel layers, covering a radial region from 21 cm to 112 cm, and five endcap discs, covering $1200 < |z| < 2700$ mm. The IT and OT provide a geometrical coverage up to $|\eta| < 4$ and 2.5, respectively.

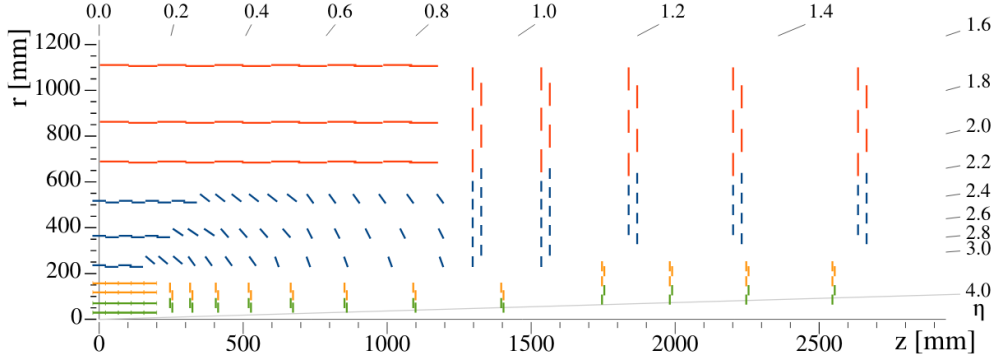


Figure 7.2: One quarter of the CMS Phase-2 tracker layout in the $r - z$ view.

The Phase-2 tracker will contain three types of modules:

Pixel module: The whole inner tracking system will be equipped with pixel modules made of silicon sensors of thickness $100 - 150 \mu\text{m}$ segmented into pixels of size 25×100 or $50 \times 50 \mu\text{m}^2$. Figure 7.2 shows the pixel modules marked in green and yellow colors.

PS module: Three inner barrel layers and the inner part of the endcap discs of the OT will consist of PS modules made of two single-sided, closely spaced, parallel silicon sensors. The sensor towards the interaction point *i.e.* bottom (seed) sensor consists of macro-pixels and the away sensor *i.e.* top (correlated) is built with micro-strips. The PS modules are marked in blue in Figure 7.2.

2S module: Three outer-most barrel layers and the remaining parts of the endcap discs will consist of 2S modules made of two single-sided, closely spaced, parallel sensors containing micro-strips. The 2S modules are marked in red in Figure 7.2. A detailed overview of a 2S module along with the associated front-end electronics is given in Section 7.3.

The main parameters of the proposed OT modules are mentioned in Table 7.1.

	2S module	PS module
Active area	$\sim 2 \times 90 \text{ cm}^2$	$\sim 2 \times 45 \text{ cm}^2$
Bottom sensor	2×1016 strips: $\sim 5 \text{ cm} \times 90 \mu\text{m}$	32×960 macro-pixels: $\sim 1.5 \text{ mm} \times 100 \mu\text{m}$
Top sensor	2×1016 strips: $\sim 5 \text{ cm} \times 90 \mu\text{m}$	2×960 strips: $\sim 2.4 \text{ cm} \times 100 \mu\text{m}$
Sensor spacing	1.8, 4.0 mm	1.6, 2.6, 4.0 mm

Table 7.1: *Some of the main parameters of PS and 2S modules of the CMS Phase-2 tracker.*

Each sensor of the prototype modules has n -type strips on $\sim 250 \mu\text{m}$ thick p -type bulk placed at a distance of $1.8 \mu\text{m}$.

7.2 Concept of p_T Modules

The OT modules are specially designed to discriminate high p_T tracks from low p_T ones. As mentioned in Table 7.1, the two parallel sensors of an OT module are placed at a very small distance depending on the position of the module from the interaction point. In the presence of the 3.8 T magnetic field of CMS, charged particles bend in the detector transverse plane and the bending is inversely proportional to particle p_T . The signal, produced by ionization induced by a charged particle inside the silicon volume and collected by the readout electronics, is termed as hit in the detector. The purpose of a p_T module is to correlate hits produced by a track in the bottom and top sensors. For low p_T tracks, due to a large bending in the magnetic field, the hits in the bottom and top sensors will have large separation in contrary to the tracks with high p_T . A programmable window of $\pm n$ strips can be used to correlate hits of high p_T tracks.

The front-end electronics of the modules are capable of correlating the signals from the two parallel sensors into an entity referred as the “stub”. As depicted in Figure 7.3, if the hit on the top sensor falls outside the acceptable correlation window, stub formation fails.

The correlation window is dependent on the sensor spacing and p_T threshold required to select tracks at the L1 trigger system. A threshold of $p_T > 2 \text{ GeV}$ is required for a stub

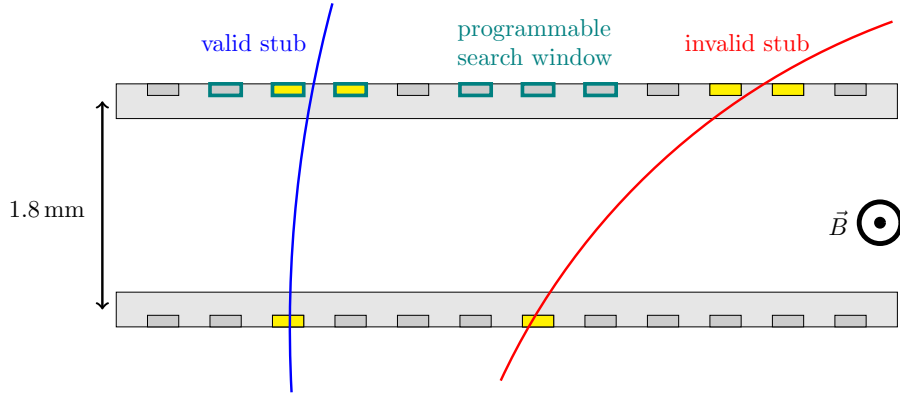


Figure 7.3: An illustration of the stub formation logic used for p_T discrimination in the 2S module of CMS Phase-2 Outer Tracker.

to be accepted and a module with 1.8 mm sensor spacing can satisfy this condition by keeping the correlation window size at ± 5 strips. The stub information can be transmitted to the L1 trigger system at the 40 MHz collision rate. Based on the trigger decision, the complete event information stored in the front-end pipelines is read out for further processing.

7.3 Prototype 2S Modules with CMS Binary Chip (CBC)

A 2S module is equipped with two silicon sensors each having 2×1016 strips read out by the front-end electronic chips (CBC). The module specifications are given in Table 7.1. The strips of each sensor are organized in two columns and connected to the CBCs as shown in Figure 7.4. There are two front-end hybrids (FEH) each with eight CBCs at the two ends of the sensors. A CBC chip provides 254 binary readout channels that connect the strips alternatively from the top and bottom sensors.

A CBC has a pre-amplifier with shaper, post-amplifier and comparator circuits. The pre-amplifier integrates the signal collected from sensor channels and the post-amplifier produces the final signal for the comparator. The amplifiers and shaper take care of the peaking time of the CBC pulse shape to be less than 20 ns and the requirement of 50 ns to return to the DC baseline voltage. Finally, the comparator circuit produces digital sig-

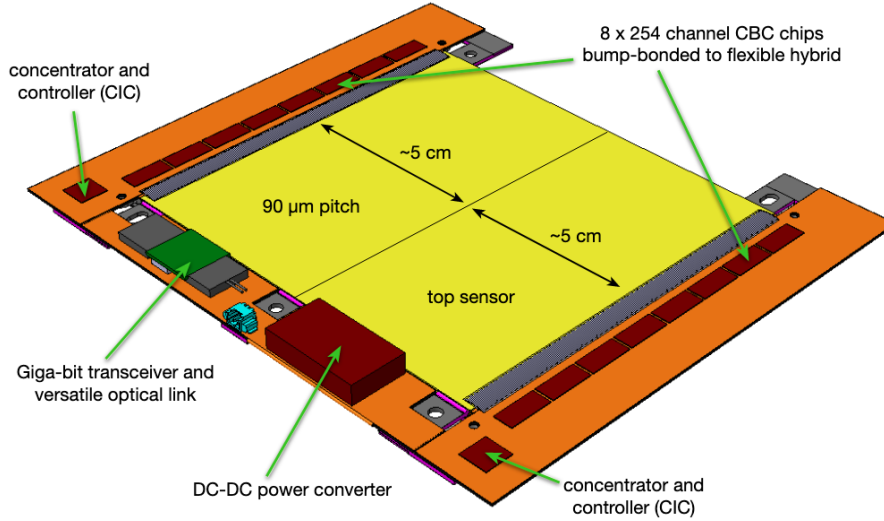


Figure 7.4: Sketch of a 2S module for the CMS Phase-2 Outer Tracker.

nal following a comparator threshold (V_{CTH}). The CBC can work with four hit detection modes [128]. Figure 7.5 schematically shows two modes used in the beam test described in this chapter.

- **40MHz Sampled Mode:** when the comparator output coincides with the rising edge of the 40 MHz sampling clock and the signal is above the threshold the output will be 1 and it will fall to 0 as soon as the signal is below threshold at any of the next rising edge of the clock.
- **Fixed Pulse Width or Latched Mode:** the comparator output will be 1 for a full 25 ns clock period if in the immediately previous clock cycle the signal is above threshold regardless of the width pulse. The output will return to 0 in the next clock cycle if the signal is below threshold in a full 25 ns clock period.

The data gathered by the CBCs are transferred via the concentrator chips (CIC) on each front-end hybrid to the service hybrid for conversion from electrical to optical data transmission.

Different prototypes of the 2S module equipped with the third revision of the CBC have been tested in two test beam facilities. At Fermilab, a mini-module that contains only one

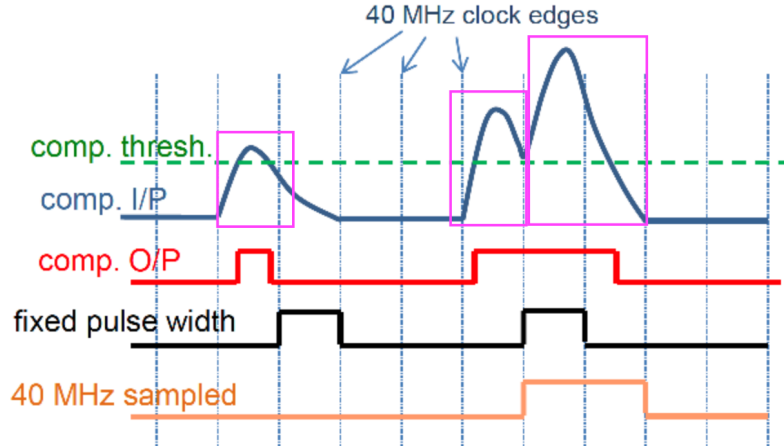


Figure 7.5: Comparator input, output and pulse shapes of two hit detection modes of the CBC.

FEH with two CBCs was tested. In the DESY test beam facility, the performance of two real sized 2S modules were studied thoroughly and the results are presented here. Out of the two modules, one is non-irradiated and the other is built with sensors irradiated with 23 MeV protons up to a 1 MeV neutron equivalent fluence of $4.6 \times 10^{14} \text{ n}_{\text{eq}}\text{cm}^{-2}$ which corresponds 1.2 \times the maximum cumulative radiation the 2S modules closest to beam line module would experience after ten years of HL-LHC operation.

High radiation can cause two types of damage in silicon detectors, a) Surface damage and, b) Bulk damage. In surface damage, charges can be trapped in the oxide layer on the detector surface which increases the noise. In bulk damage, silicon atoms can be displaced from lattice position which can form point and cluster defects. The defects are created in the band gap of silicon detectors that can change depletion voltage which results loss in signal. Therefore, it is extremely essential to study the properties of irradiated modules.

7.4 Beam Test Setup

The two beam test experiments at Fermilab and DESY use similar setup. A schematic layout of a generic test beam setup is shown in Figure 7.6. A 2S module is placed in the middle of the setup referred as the Detector Under Test (DUT). Six layers of pixel

detectors form the “telescope” system which is used for precise reconstruction of tracks. Four scintillators in coincidence configuration are placed at the two ends of the beam test setup to trigger events and a timing detector is placed within the telescope system to provide the timing information. All the detector planes are placed perpendicular to the beam direction.

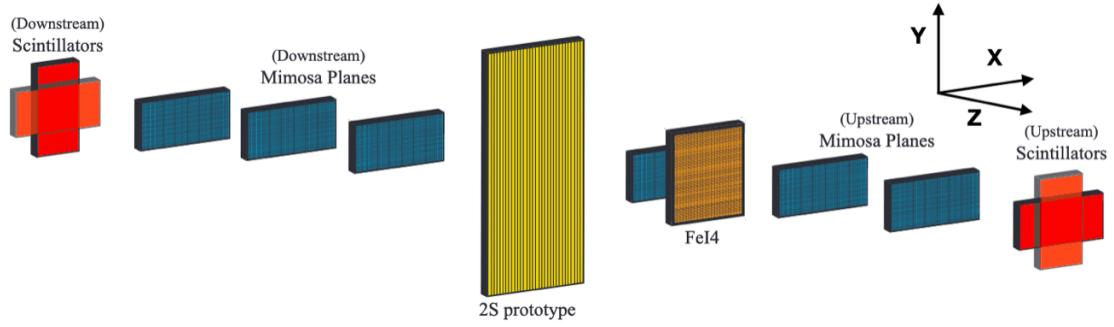


Figure 7.6: A Schematic representation of a typical beam test setup.

A 120 GeV proton beam is used in the Fermilab beam test facility and in the DESY setup, an electron beam with energy up to 6.3 GeV was used to perform the studies. The DESY beam test experiment is discussed in detail in the following sections.

The EUDET-type [129] beam telescopes were used during data taking. The telescope system consists of six MIMOSA-26 active pixel devices with an active area of approximately $1 \times 2 \text{ cm}^2$ that are read out with the EUDAQ [130] framework. The small pixel size of $18.4 \times 18.4 \mu\text{m}^2$ provides a very good spatial resolution of around $10 \mu\text{m}$. The trigger and event number synchronization was performed by an EUDET-type Trigger Logic Unit (TLU) [131]. The timing layer is a FEI4 detector [132] used to provide timing information at a 25 ns granularity.

7.5 Data Acquisition (DAQ) and Track Reconstruction

A data acquisition framework, EUDAQ [CITE], is used for data taking and track reconstruction. The framework is centered around the RunControl module, which provides a

state machine to coordinate all the measurements. Each hardware type is read out by a producer which delivers data to the central DataCollector module. The DataCollector combines various data streams and subsequently data is saved event by event in a raw file for further processing. The data acquisition technique is summarized in the left half of Figure 7.7. The 2S module is configured and read out via the optical link by a micro-

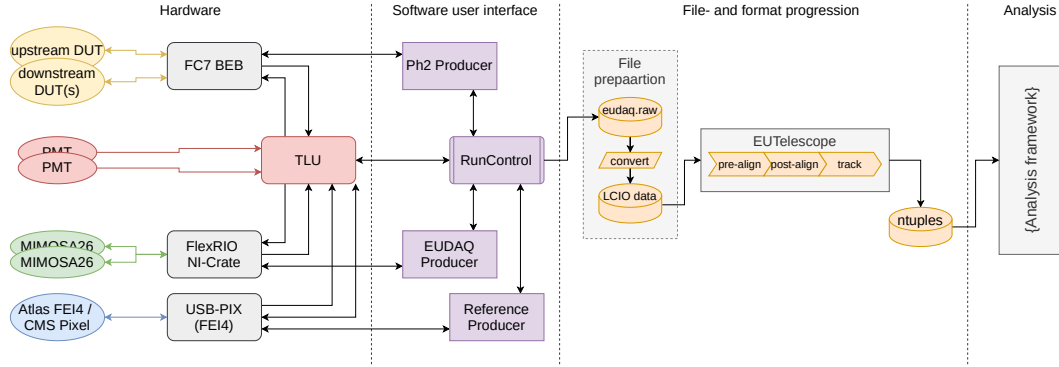


Figure 7.7: Schematic diagram of the data acquisition scheme that consists of a common controller serves the dedicated data producers that ensures the synchronization between the telescope, FEI4 and the DUT.

TCA compatible FPGA board, called the FC7 [133]. The DAQ chain for the 2S module has a custom C++ software package called Phase-2 Acquisition and Control Framework (Ph2_ACF) [134].

The information from the telescope planes is used to reconstruct the particle tracks precisely. The General Broken Line (GBL) [135] technique is utilized for track fitting and alignment after masking hot pixels and finding clusters for each detector layer. The FEI4 plane and the silicon layers of the DUT are utilized in the track fit as passive scatter planes. Following that, the track data is used to perform the alignment of the DUT and the FEI4 plane. Each sensor in the 2S modules is aligned individually, allowing information on the module's assembly precision to be extracted. Finally, for each detector layer, all the information about cluster (see Section 7.6) positions and track intersection points is saved for further analysis. The right half of Figure 7.7 demonstrates all the steps discussed above.

7.6 Data analysis

A charged particle deposits its energy in one or multiple adjacent strips while passing through a sensor. The deposited energy is collected in the form of current which is converted by the front-end electronics to a voltage pulse. If the pulse height crosses the comparator threshold voltage V_{CTH} , the corresponding channel fires and the signal information is stored as binary “1” which is referred as a “Hit”. The position of the hit is referred by the corresponding channel/strip number of the detector module. Hits in consecutive channels are combined together to form a “cluster” and the average of the respective channel numbers in a precision of half-strip pitch is stored as the cluster position. The number of strips in a cluster is termed as cluster width. Clusters with multiple strips provide a better estimation of the position of an incident particle. Clusters from the top and bottom sensors of a module are used to form stubs as mentioned in Section 7.2. The stub position is defined by the position of a cluster on the bottom sensor. Clusters of width < 3 are used for stub formation and the correlation window is kept ± 5 (4.5) strips for the non-irradiated (irradiated) sensors.

Prior to all measurements, a V_{CTH} scan is performed to separate the signal from the pedestal (or noise) in two hit detection modes (Section 7.3). No change is expected in the performance.

Although track reconstruction does not include any information of the DUT, it is mandatory to make sure that the DUT is aligned to the telescope system before any performance study. The alignment step is performed by a χ^2 minimization of the residual between the hit position and the track fit point on the DUT. The offsets measured in both translational and rotational movements of the DUT are applied to keep the DUT aligned.

To estimate the performance of the DUT, only the tracks with a matched cluster on the FEI4 plane fulfilling the conditions: $\Delta x(\text{track, cluster}) \leq 0.2 \text{ mm}$ and $\Delta y(\text{track, cluster}) \leq 0.08 \text{ mm}$, are considered. In addition, the tracks are also required to be isolated by

imposing a condition that no additional track projection on the FEI4 plane should be present within a radius of 0.6 mm. The isolated tracks are used for further analysis.

To study the performance of a 2S module, the cluster and stub efficiencies of the DUT are measured. A cluster is matched to an isolated track if $\Delta x(\text{track}, \text{cluster}) \leq 0.2 \text{ mm}$.

$$\text{Efficiency} = \frac{\text{Number of isolated tracks matched to a cluster / stub}}{\text{Number of isolated tracks}} \quad (7.1)$$

The stub-track matching can be performed in two ways: (a) applying similar condition as above on the stub position, (b) using the information of the clusters of both sensors, where the number of stubs fulfilling the criteria $\Delta x(\text{track}, \text{stub position}) \leq 0.2 \text{ mm}$ and $\Delta x(\text{track}, \text{stub position} + \text{stub bend}) \leq 0.2 \text{ mm}$ are considered matched to the tracks. The performance plots with stub efficiencies are shown in the next section using option (b).

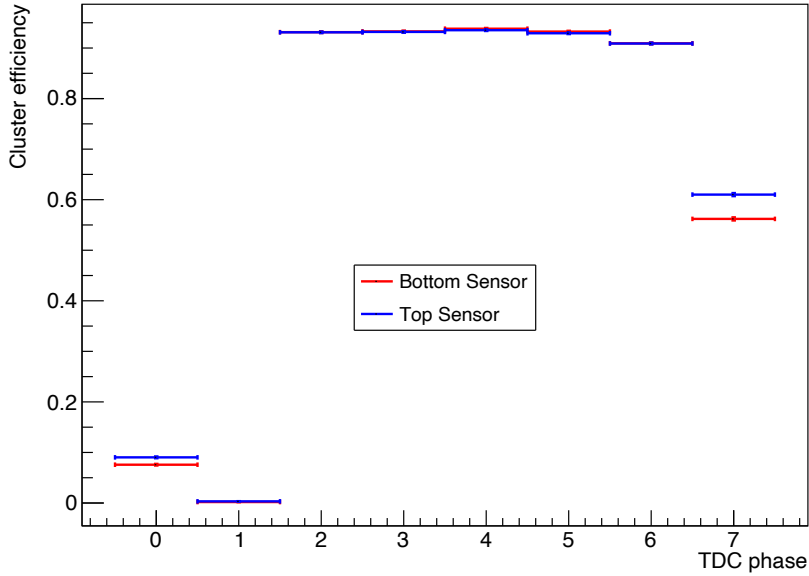


Figure 7.8: Variation of reconstructed cluster efficiency with eight TDC phases for the DUT with non-irradiated sensors.

Unlike the LHC, the clock domains of the DUT and the beam test setup are not synchronized. As a result, particles can arrive at an arbitrary phase with respect to the CBC clock. Because of the fast shaper response, the signal sampling time is not always at the maximum of the pulse amplitude which may reduce the hit detection efficiency. To account

for this effect in the beam test, the FC7 firmware measures the arrival time of the trigger signal from the TLU within $\frac{1}{8} \times 25 = 3.125$ ns of the 40 MHz clock. For each event, the so-called TDC phase between the module clock and particle arrival is recorded. As a result, the optimal phase can be chosen during offline analysis. Figure 7.8 displays the reconstructed cluster efficiency for the non-irradiated 2S module as a function of the TDC phase. In all subsequent evaluations, only the most efficient TDC phases (2, 3 and 4) are chosen for each run independently.

7.7 Results and discussions

The beam test data is taken at a bias voltage of 300 V for the non-irradiated sensors while for the irradiated ones, the bias voltage is kept at 600 V. The irradiated sensors require higher bias voltage to overcome the charge trapping inside the silicon bulk of the sensors due to the effect of cumulative radiation. Performance of non-irradiated and irradiated prototype detectors are presented in the following.

7.7.1 Signal and noise: Threshold scan

In the beam test, various threshold scans were performed. The V_{CTH} can be converted to number of equivalent electrons using the conversion of $1 V_{CTH} = 156 e^-$. The front-end electronics collects electrons from the silicon sensors as the signal and, therefore, higher V_{CTH} represents lower threshold value. Figure 7.9 shows the variation of cluster and stub efficiencies with V_{CTH} . It can be seen that the cluster and stub efficiencies increase with increasing V_{CTH} , reach a plateau, and drop again around 550. Lower threshold accepts more clusters and the overall efficiency increases. However, with further reduction in the threshold value, random noise starts firing the sensor channels. Thus large clusters are formed and the track matching with the clusters fails which causes decrease in efficiencies.

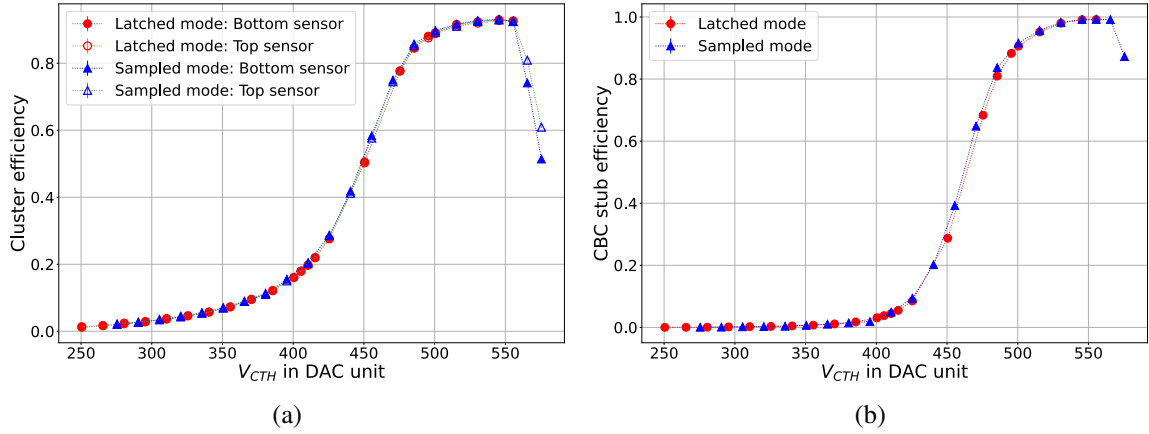


Figure 7.9: Variation of (a) cluster efficiency, and (b) stub efficiency with V_{CTH} for non-irradiated module.

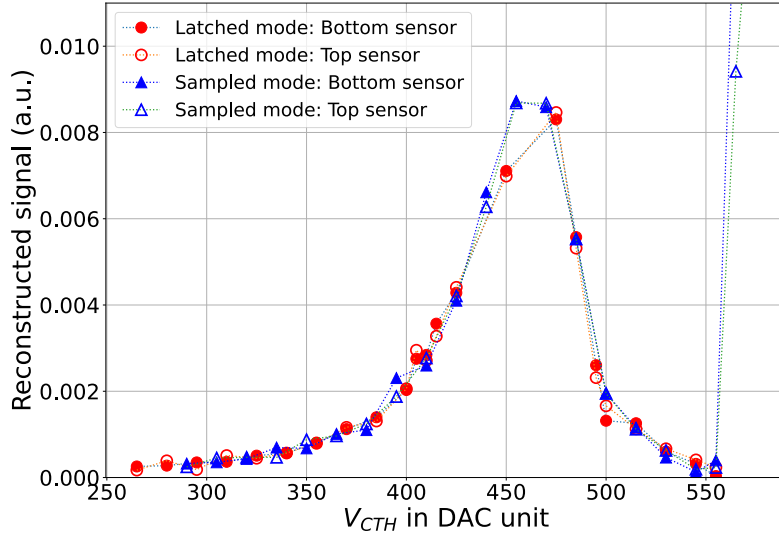


Figure 7.10: Reconstructed signal shape for non-irradiated module.

The differential distribution of the cluster efficiency (Figure 7.10), which follows the inverted Landau distribution, demonstrates the shape of the reconstructed signal. The distribution helps to optimize the threshold to separate signal from noise and the value of V_{CTH} is chosen to be set at 545 for the non-irradiated module. The V_{CTH} scans show similar nature for both the latched and sampled hit detection modes.

7.7.2 Tracking and Module Resolution

The alignment of the 2S module within the beam telescope is optimized by ensuring that the distribution of the residuals of the track and cluster coordinates in x (perpendicular to the strips) is centred around zero. This also provides information regarding the spatial resolution of the 2S module and the uncertainty of the telescope prediction at the DUT plane.

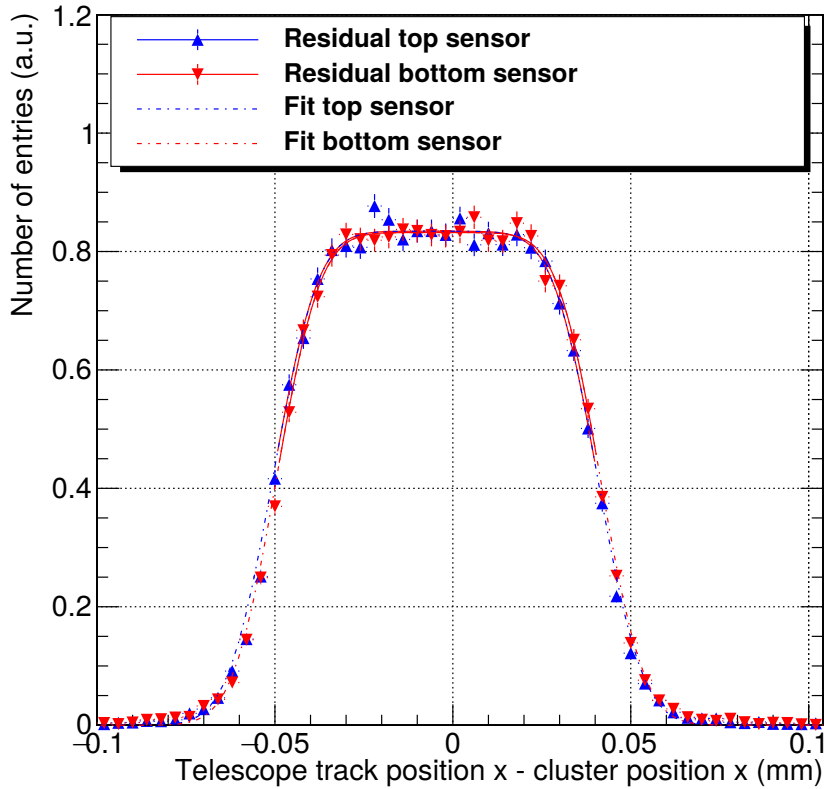


Figure 7.11: *The residual distributions for the two sensors of a non-irradiated 2S module prototype.*

The residual between the expected and physical position of single-strip clusters should ideally satisfy a rectangular function of $90\,\mu\text{m}$ width which corresponds to the strip pitch and binary readout of the 2S module. Finite telescope resolution requires a Gaussian convolution with zero mean and the standard deviation that corresponds to the track uncertainty. Figure 7.11 shows the residual distribution fitted with the Gaussian convolved

rectangular function. The fit results of the left and right rectangle borders are in good agreement with the expectation of $w = 90\mu m$. The fit result is in good agreement with the telescope track uncertainty (σ) estimated by The GBL Track Resolution Simulator [136]. Table 7.2 shows some of the fit results along with the expected values.

Parameters	Top sensor	Bottom sensor	Expectation
Width (μm)	91.3 ± 1.1	91.1 ± 1.1	90
σ (μm)	10.0 ± 0.9	9.7 ± 0.8	$9.4^{+0.3}_{-0.4}$

Table 7.2: Summary of the fit parameters shown in Figure 7.11.

In addition, the resolution of the sensors is calculated to be 26.9 ± 0.3 (27.1 ± 0.3) μm for the top (bottom) sensor which is also in agreement with the expected resolution $90/\sqrt{12} \sim 26\mu m$.

7.7.3 Uniformity of Efficiency

The DUT is mounted on a translation stage, which allows the 2S module to be moved along the strips (y-axis), perpendicular to the strips (x-axis), and rotated around the strip axis. The module is moved along both x and y axes to check the uniformity of the efficiency. Figure 7.12 demonstrates the variation of cluster efficiency with track x-position *i.e.* perpendicular to the strip length. The cluster efficiency is above 99% over the entire tested area on both sensors except for three bins corresponding to two channels on the bottom sensor which are due to the channels disconnected from the readout electronics. This measurement is performed at a bias voltage of 300 V with $V_{cth} = 545$. The variation of efficiency with track x-position ensures the uniformity of the efficiency.

7.7.4 Performance of p_T Discrimination

The prime motivation of the beam test analysis with 2S modules is to study the performance of p_T discrimination through the stub finding logic. There was no magnetic field

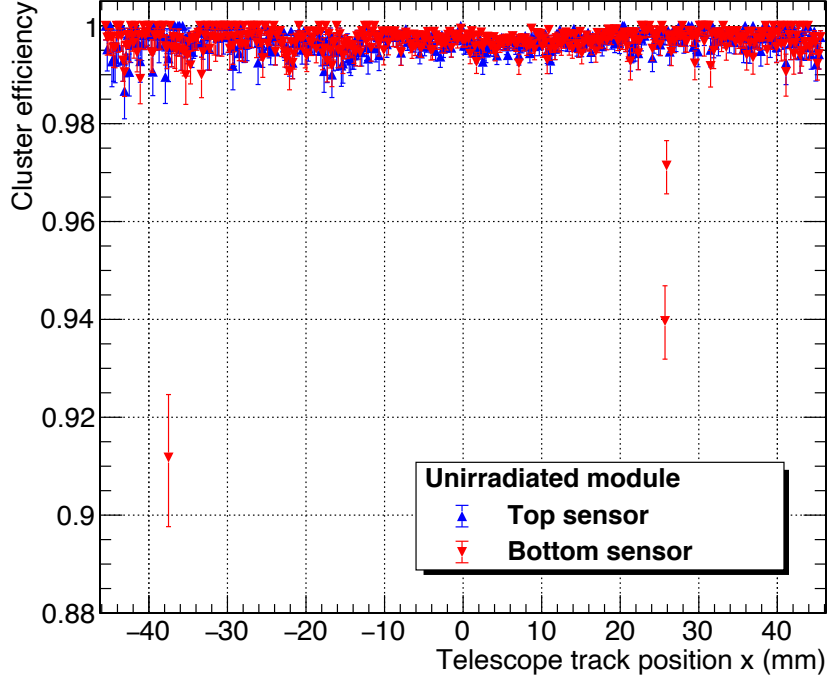


Figure 7.12: Variation of cluster efficiency with track x position, for both non-irradiated sensors of the DUT.

present in the beam test setup, but the effect of bending was emulated by rotating the DUT with respect to the beam axis. In the DESY beam test setup, the DUT is rotated from -40 to 40 degree (α) to study the effect of bending. The correlation window for stub formation is set to ± 5 strips for the non-irradiated module and ± 4.5 strips for the irradiated module. Figure 7.13a shows the variation of stub efficiency with α . It is clearly observed that the stub efficiency sharply falls around 16 (14) degree for the 2S modules with non-irradiated (irradiated) sensors. The variation of stub efficiency is fitted using a parametric error function $f(\alpha) = 1 - \frac{1}{2}(p_0 + p_1 \times \text{erf}(\frac{x-p_2}{p_3}))$, where p_2 denotes the value of α at 50% efficiency. Both values are in accordance with the expectation considering the chosen stub window sizes and the measured sensor distances of the module prototypes of around 1.65 mm. The DUT rotation angle can be emulated in terms of p_T . Assuming the DUT to be at the fourth layer of the tracker barrel *i.e.* at a radial distance (R) of 0.715 m in a magnetic field of $B = 3.8$ T, the emulated p_T would be $0.57 \times \frac{R}{\sin(\alpha)}$. From the variation of stub efficiency with p_T , as shown in Figure 7.13b, a sharp turn-on is observed around $p_T = 1.48$ (1.68)

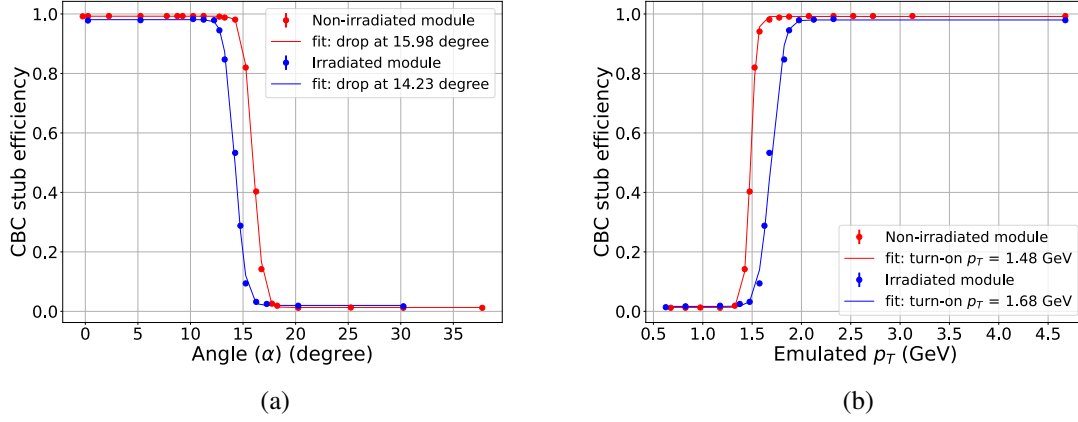


Figure 7.13: Variation of stub efficiency with (a) angle and (b) emulated p_T .

GeV for the 2S modules with non-irradiated (irradiated) sensors obtained by fitting the variation with $1 - f(\alpha)$.

The performance study of the p_T discrimination logic shows that the 2S modules will be able to select tracks with $p_T > 2$ GeV with high efficiency, which is the primary goal of the design of the 2S modules.

Chapter 8

Conclusion

This chapter contains a summary of the studies presented in this thesis.

- The double Higgs production is the ideal process to probe the Higgs boson trilinear self-coupling (λ_{HHH}). A precise measurement of λ_{HHH} may provide hints of any possible deviation from the SM predicted value, which is the primary motivation behind the HH non-resonant analysis presented in this thesis. The analysis has been performed in the $b\bar{b}W^+W^-$ decay channel through both SM and BSM Higgs boson couplings. Two modes of HH production, GGF and VBF, have been analyzed in detail. A DNN based final discriminator is used in the statistical inference. The expected upper limit at 95% confidence level on the HH production cross section with the SM signal has been estimated to be $17 \times \text{SM}$. In addition to the SM HH studies, the coupling modifiers κ_λ and κ_{2V} have been scanned. The corresponding exclusion regions have been estimated from the upper limit on the HH cross section. The expected allowed regions of the respective coupling modifiers are estimated as: $-8.1 < \kappa_\lambda < 14.9$ and $-0.68 < \kappa_{2V} < 2.8$. Finally, the expected upper limits for the EFT shape benchmarks have been estimated. The analysis is currently in unblinding stage and the immediate plan is to look at the observed limits. In the next step, the $HH \rightarrow b\bar{b}W^+W^-$ channel will be combined with other HH decay modes

to achieve better sensitivity. However, because of the very small HH production cross-section at the LHC, a much larger dataset will be required to study the process comprehensively, which will be available only at the HL-LHC.

- A phenomenological study to search for signature of exotic Higgs bosons at the LHC and beyond has been performed. The phenomenology of a flavor model based on the $S3$ symmetry has been considered where the flavor violating signals can be produced in two different mechanisms,
 - Production of a pseudoscalar (χ) from a heavy BSM Higgs (H) in association with a Z boson.
 - Production of χ/H via the $t\bar{t}$ process.

The two different signal processes exhibit different topologies, requiring two separate analyses. The traditional cut based and several multivariate analysis methods namely, BDT, XGBoost and DNN have been used to increase the signal sensitivity. The first analysis shows a high sensitivity even with a 20% linear-in-background systematic uncertainty. For the second process, the multivariate methods improve the signal significance, despite high background contamination.

- The performance of the strip-strip (2S) prototype module of the proposed CMS Phase-2 tracker is studied in FNAL and DESY beam test environment. The module has the capability of separating high p_T tracks from the low p_T ones and for the first time, the tracking information will be used in the Level-1 trigger system. The p_T discrimination logic has been studied by the angular scan of the prototype. It has been found that the stub efficiency remains above 98% for tracks of p_T even as low as 1.7 GeV, which justifies the decision of using the 2 GeV p_T threshold at the level-1 trigger system. The DESY beam test results have been described in this thesis.

Appendix-1

HLT

The HLT paths used for $HH \rightarrow b\bar{b}W^+W^-$ analysis are listed in Table 8.1 for all three eras of Run-2.

Good Lumi JSON Files

The good-lumi or Golden-JSON files for three eras are shown in Table 8.2.

Datasets with Goos Lumi Fractions and Luminosity

The datasets with the runs having good lumisections and the corresponding integrated luminosities are listed for 2016 (Table 8.3), 2017 (Table 8.4), and 2018 (Table 8.5).

Description of the sources of systematic uncertainties

The sources of the systematic uncertainties are described below.

- **Experimental uncertainties**

Single-electron triggers	HLT_Ele25_eta2p1_WPTight_Gsf (✓/-/-) HLT_Ele27_WPTight_Gsf (✓/-/-) HLT_Ele27_eta2p1_WPLoose_Gsf (✓/-/-) HLT_Ele32_WPTight_Gsf (-/✓/✓) HLT_Ele35_WPTight_Gsf (-/✓/-)
Single-muon triggers	HLT_IsoMu22 (✓/-/-) HLT_IsoTkMu22 (✓/-/-) HLT_IsoMu22_eta2p1 (✓/-/-) HLT_IsoTkMu22_eta2p1 (✓/-/-) HLT_IsoMu24 (✓/✓/✓) HLT_IsoTkMu24 (✓/-/-) HLT_IsoMu27 (-/✓/✓)
Double-electron triggers	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (✓/-/-) HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL (-/✓/✓)
Double-muon triggers	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL (✓/-/-) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ (✓/-/-) HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL (✓/-/-) HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ (✓/-/-) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8 (-/✓/-) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8 (-/✓/✓)
Electron + muon triggers	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL (✓/-/-) HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ (✓/✓/✓) HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL (✓/-/-) HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ (✓/-/-) HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ (✓/✓/-) HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ (✓/✓/-) HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL (✓/✓/-) HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (-/✓/-)

Table 8.1: List of HLTs used in the analysis. Checkmarks (✓) and hyphens (-) are added after each HLT path to indicate whether the path was enabled or disabled during the 2016, 2017, and 2018 data-taking periods.

APPENDIX

Data-taking period	JSON file
2016	Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt
2017	Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt
2018	Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt

Table 8.2: List of JSON files used to implement the “good-run” selection in the 2016, 2017, and 2018 data-taking periods.

Dataset name	Run-range	Int. luminosity (fb^{-1})
/SingleElectron/Run2016B-17Jul2018_ver2-v1	273150–275376	5.83
/SingleElectron/Run2016C-17Jul2018-v1	275656–276283	2.62
/SingleElectron/Run2016D-17Jul2018-v1	276315–276811	4.29
/SingleElectron/Run2016E-17Jul2018-v1	276831–277420	4.07
/SingleElectron/Run2016F-17Jul2018-v1	277932–278808	3.14
/SingleElectron/Run2016G-17Jul2018-v1	278820–280385	7.65
/SingleElectron/Run2016H-17Jul2018-v1	281613–284044	8.74
/SingleMuon/Run2016B-17Jul2018_ver2-v1	273150–275376	5.82
/SingleMuon/Run2016C-17Jul2018-v1	275656–276283	2.62
/SingleMuon/Run2016D-17Jul2018-v1	276315–276811	4.29
/SingleMuon/Run2016E-17Jul2018-v1	276831–277420	4.07
/SingleMuon/Run2016F-17Jul2018-v1	277932–278808	3.14
/SingleMuon/Run2016G-17Jul2018-v1	278820–280385	7.65
/SingleMuon/Run2016H-17Jul2018-v1	281613–284044	8.74
/DoubleEG/Run2016B-17Jul2018_ver2-v1	273150–275376	5.83
/DoubleEG/Run2016C-17Jul2018-v1	275656–276283	2.62
/DoubleEG/Run2016D-17Jul2018-v1	276315–276811	4.29
/DoubleEG/Run2016E-17Jul2018-v1	276831–277420	4.07
/DoubleEG/Run2016F-17Jul2018-v1	277932–278808	3.14
/DoubleEG/Run2016G-17Jul2018-v1	278820–280385	7.65
/DoubleEG/Run2016H-17Jul2018-v1	281613–284044	8.74
/DoubleMuon/Run2016B-17Jul2018_ver2-v1	273150–275376	5.82
/DoubleMuon/Run2016C-17Jul2018-v1	275656–276283	2.62
/DoubleMuon/Run2016D-17Jul2018-v1	276315–276811	4.29
/DoubleMuon/Run2016E-17Jul2018-v1	276831–277420	4.07
/DoubleMuon/Run2016F-17Jul2018-v1	277932–278808	3.14
/DoubleMuon/Run2016G-17Jul2018-v1	278820–280385	7.65
/DoubleMuon/Run2016H-17Jul2018-v1	281613–284044	8.74
/MuonEG/Run2016B-17Jul2018_ver2-v1	273150–275376	5.82
/MuonEG/Run2016C-17Jul2018-v1	275656–276283	2.62
/MuonEG/Run2016D-17Jul2018-v1	276315–276811	4.29
/MuonEG/Run2016E-17Jul2018-v2	276831–277420	4.07
/MuonEG/Run2016F-17Jul2018-v1	277932–278808	3.14
/MuonEG/Run2016G-17Jul2018-v1	278820–280385	7.65
/MuonEG/Run2016H-17Jul2018-v1	281613–284044	8.74

Table 8.3: List of 2016 datasets. The /DoubleEG, /DoubleMuon, and /MuonEG datasets are used for the dilepton channel only.

- Luminosity: It is a normalization uncertainty which is applied to all the simulated samples with a value of 1.2% in 2016, 2.3% in 2017 and 2.5% in 2018.

APPENDIX

Dataset name	Run-range	Int. luminosity (fb ⁻¹)
/SingleElectron/Run2017B-31Mar2018-v1	297047–299329	4.79
/SingleElectron/Run2017C-31Mar2018-v1	299368–302029	9.63
/SingleElectron/Run2017D-31Mar2018-v1	302030–302663	4.25
/SingleElectron/Run2017E-31Mar2018-v1	303818–304797	9.31
/SingleElectron/Run2017F-31Mar2018-v1	305040–306460	13.54
/SingleMuon/Run2017B-31Mar2018-v1	297047–299329	4.79
/SingleMuon/Run2017C-31Mar2018-v1	299368–302029	9.63
/SingleMuon/Run2017D-31Mar2018-v1	302031–302663	4.25
/SingleMuon/Run2017E-31Mar2018-v1	303824–304797	9.31
/SingleMuon/Run2017F-31Mar2018-v1	305040–306462	13.54
/DoubleEG/Run2017B-31Mar2018-v1	297047–299329	4.79
/DoubleEG/Run2017C-31Mar2018-v1	299368–302029	9.63
/DoubleEG/Run2017D-31Mar2018-v1	302030–302663	4.25
/DoubleEG/Run2017E-31Mar2018-v1	303818–304797	9.31
/DoubleEG/Run2017F-31Mar2018-v1	305040–306460	13.54
/DoubleMuon/Run2017B-31Mar2018-v1	297047–299329	4.79
/DoubleMuon/Run2017C-31Mar2018-v1	299368–302029	9.63
/DoubleMuon/Run2017D-31Mar2018-v1	302031–302663	4.25
/DoubleMuon/Run2017E-31Mar2018-v1	303824–304797	9.31
/DoubleMuon/Run2017F-31Mar2018-v1	305040–306462	13.54
/MuonEG/Run2017B-31Mar2018-v1	297047–299329	4.79
/MuonEG/Run2017C-31Mar2018-v1	299368–302029	9.63
/MuonEG/Run2017D-31Mar2018-v1	302031–302663	4.25
/MuonEG/Run2017E-31Mar2018-v1	303824–304797	9.31
/MuonEG/Run2017F-31Mar2018-v1	305040–306460	13.54

Table 8.4: List of 2017 datasets. The /DoubleEG, /DoubleMuon, and /MuonEG datasets are used for the dilepton channel only.

Dataset name	Run-range	Int. luminosity (fb ⁻¹)
/SingleMuon/Run2018A-17Sep2018-v2	315257–316995	14.03
/SingleMuon/Run2018B-17Sep2018-v1	317080–319310	7.06
/SingleMuon/Run2018C-17Sep2018-v1	319337–320065	6.90
/SingleMuon/Run2018D-22Jan2019-v2	320500–325175	31.74
/EGamma/Run2018A-17Sep2018-v2	315257–316995	14.03
/EGamma/Run2018B-17Sep2018-v1	317080–319310	7.06
/EGamma/Run2018C-17Sep2018-v1	319337–320065	6.90
/EGamma/Run2018D-22Jan2019-v2	320413–325175	31.74
/DoubleMuon/Run2018A-17Sep2018-v2	315257–316995	14.03
/DoubleMuon/Run2018B-17Sep2018-v1	317080–319310	7.06
/DoubleMuon/Run2018C-17Sep2018-v1	319337–320065	6.90
/DoubleMuon/Run2018D-PromptReco-v2	320500–325175	31.74
/MuonEG/Run2018A-17Sep2018-v1	315257–316995	14.03
/MuonEG/Run2018B-17Sep2018-v1	317080–319310	7.06
/MuonEG/Run2018C-17Sep2018-v1	319337–320065	6.90
/MuonEG/Run2018D-PromptReco-v2	320500–325175	31.74

Table 8.5: List of 2018 datasets. The /DoubleMuon and /MuonEG datasets are used for the dilepton channel only. The /EGamma datasets contains events passing the single-electron and double-electron triggers and replace the former /SingleElectron and /DoubleEG datasets.

- Pre-firing: Pre-firing of Level-1 ECAL trigger primitives correction is applied and the corresponding shape uncertainty applied in 2016 and 2017.
 - Pileup: A shape uncertainty to cover the pileup correction applied in all eras.
 - Trigger: Shape uncertainty with respect to the trigger efficiencies.
 - Jet PUID: Uncertainties about shape related to the jet PU ID criterion used to jets, corresponding to the selection effectiveness and mistagging rate of jets as coming from pileup.
 - JES: Jet Energy Scale uncertainty as shape.
 - JER: Shape uncertainty associated to the Jet Energy Resolution *i.e.* the smearing of jets.
 - b -tagging : Shapes uncertainty associated to the b -jet identification.
 - Unclustered \cancel{E}_T : PF candidate energies that are modified to provide the MET before type-1 adjustments a shape uncertainty.
- **Data-driven uncertainties:** Statistical uncertainties from the measurement regions of the lepton misidentification probabilities are propagated to the respective application regions as systematic uncertainties. In the final fit, they are uncorrelated between the three different years. For DL only, uncertainty on DY background is derived by closure tests. It is applied as shape uncertainty in the background classes, except single Higgs and, applied as normalisation only uncertainty in the HH signal and single Higgs background classes.
 - **Theoretical uncertainties**
 - Branching ratios: Normalization uncertainty of value $\pm 1.25\%$ for $H \rightarrow b\bar{b}$, $\pm 1.52\%$ for $H \rightarrow W^+W^-$ and $\pm 1.64\%$ for $H \rightarrow \tau^+\tau^-$.
 - Event generation: Normalization uncertainty related to QCD scale, PDF and α_s , applied for all the simulated samples including the signals also.

- Mass uncertainties: Normalization uncertainty for top mass correction for $t\bar{t}$ and for total HH mass.
 - EW corrections: Normalization uncertainty for $t\bar{t}V$.
 - Parton shower: ISR and FSR affects the α_s distribution and a shape uncertainty is applied for all the three eras.
 - Scales: Renormalization and factorization scale fluctuations are associated with shape uncertainty.
 - $t\bar{t}$ generation: The $t\bar{t}$ process takes into account a number of shape uncertainties related to theoretical settings during event generation, including the colour reconnection between quarks and gluons in the parton shower in Pythia split into three sources, the damping variable h_{damp} in Powheg, which controls the merging between the matrix element and the parton shower to regulate high- p_T radiation, and the underlying events describing the interaction of the partons that do not enter the hard scattering.
- **Statistical uncertainty:** Additional systematics are incorporated based on the histogram bin errors using the Beeston-Barlow light approach [86] in the statistical inference. This is referred as “Auto MC-Stats” in the Higgs Combined and HH inference tool.

MET Filters

Filter name	Applied to data	Applied to simulation
Flag_goodVertices	✓	✓
Flag_globalSuperTightHalo2016Filter	✓	✓
Flag_HBHENoiseFilter	✓	✓
Flag_HBHENoiseIsoFilter	✓	✓
Flag_EcalDeadCellTriggerPrimitiveFilter	✓	✓
Flag_BadPFMuonFilter	✓	✓
Flag_ecalBadCalibReducedMINIAODFilter [†]	✓	✓
Flag_eeBadScFilter: Applied to data only	✓	—

[†] Applied only in the 2017 and 2018 data-taking periods

Table 8.6: \cancel{E}_T filters applied to events selected in data and to simulated events. A hyphen (—) indicates that the filter is not applied.

Appendix-2

EFT benchmarks

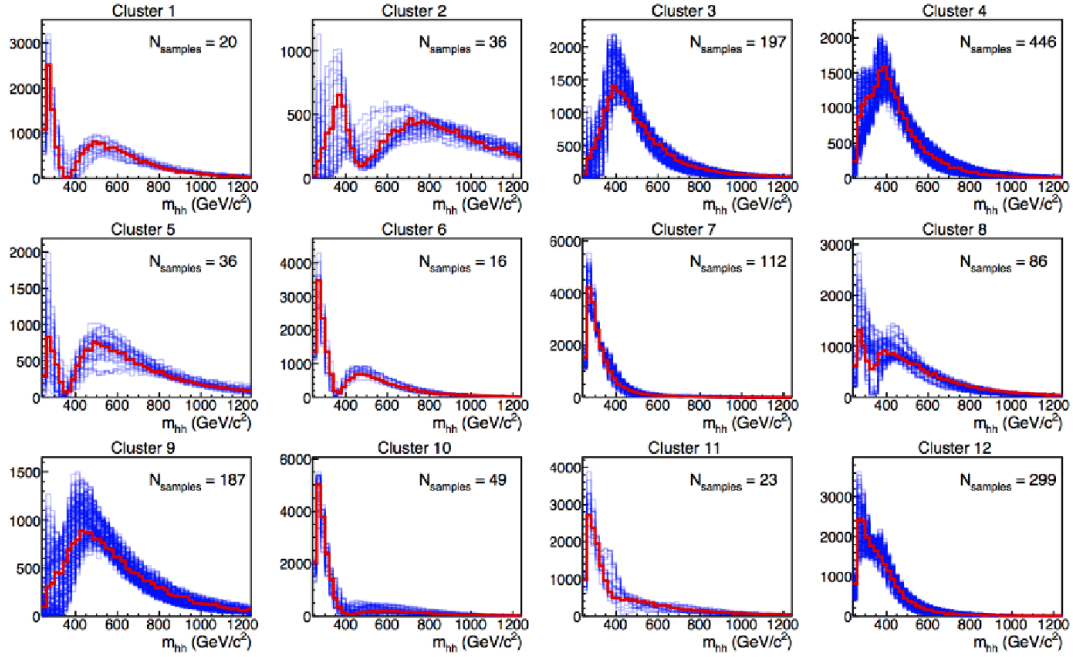


Figure 8.1: For the various benchmarks listed in Reference [2], generator-level distributions of the double Higgs invariant mass m_{HH} . The benchmark that was determined to be the most representative within the cluster is shown by the red line, while all the distributions of the identified cluster are represented by the blue lines.

Bibliography

- [1] J. Alison et al. Higgs boson potential at colliders: Status and perspectives. *Rev. Phys.*, 5:100045, 2020. doi: 10.1016/j.revip.2020.100045.
- [2] Alexandra Carvalho, Martino Dall’Osso, Tommaso Dorigo, Florian Goertz, Carlo A. Gottardo, and Mia Tosi. Higgs Pair Production: Choosing Benchmarks With Cluster Analysis. *JHEP*, 04:126, 2016. doi: 10.1007/JHEP04(2016)126.
- [3] G. Buchalla, M. Capozzi, A. Celis, G. Heinrich, and L. Scyboz. Higgs boson pair production in non-linear Effective Field Theory with full m_t -dependence at NLO QCD. *JHEP*, 09:057, 2018. doi: 10.1007/JHEP09(2018)057.
- [4] Matteo Capozzi and Gudrun Heinrich. Exploring anomalous couplings in Higgs boson pair production through shape analysis. *JHEP*, 03:091, 2020. doi: 10.1007/JHEP03(2020)091.
- [5] Search for ttH production in multilepton final states at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2016. URL <https://cds.cern.ch/record/2141078>.
- [6] Maria Aldaya. Nnlo+nnll top-quark-pair cross sections, 2021. URL <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- [7] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.

BIBLIOGRAPHY

- [8] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.
- [9] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961. ISSN 0029-5582. doi: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2). URL <https://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [10] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. doi: 10.1103/PhysRevLett.19.1264. URL <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [11] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968. doi: 10.1142/9789812795915_0034.
- [12] Murray Gell-Mann. Symmetries of baryons and mesons. *Phys. Rev.*, 125:1067–1084, Feb 1962. doi: 10.1103/PhysRev.125.1067. URL <https://link.aps.org/doi/10.1103/PhysRev.125.1067>.
- [13] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963. doi: 10.1103/PhysRevLett.10.531. URL <https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.
- [14] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, 02 1973. ISSN 0033-068X. doi: 10.1143/PTP.49.652. URL <https://doi.org/10.1143/PTP.49.652>.
- [15] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. doi: 10.1103/PhysRevLett.13.321. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.

BIBLIOGRAPHY

- [16] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. doi: 10.1103/PhysRevLett.13.508. URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [17] Axel Maas. Brout-Englert-Higgs physics: From foundations to phenomenology. *Prog. Part. Nucl. Phys.*, 106:132–209, 2019. doi: 10.1016/j.pnpnp.2019.02.003.
- [18] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli, E. Vryonidou, and M. Zaro. Higgs pair production at the LHC with NLO and parton-shower effects. *Phys. Lett. B*, 732:142–149, 2014. doi: 10.1016/j.physletb.2014.03.026.
- [19] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016. doi: 10.23731/CYRM-2017-002.
- [20] Albert M Sirunyan et al. Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state. *Phys. Lett. B*, 775: 1–24, 2017. doi: 10.1016/j.physletb.2017.10.021.
- [21] Florian Goertz, Andreas Papaefstathiou, Li Lin Yang, and José Zurita. Higgs boson pair production in the D=6 extension of the SM. *JHEP*, 04:167, 2015. doi: 10.1007/JHEP04(2015)167.
- [22] Alexandra Carvalho and Dall’Osso et.al. Analytical parametrization and shape classification of anomalous HH production in the EFT approach. 7 2016.
- [23] T. Plehn, M. Spira, and P. M. Zerwas. Pair production of neutral Higgs particles in gluon-gluon collisions. *Nucl. Phys. B*, 479:46–64, 1996. doi: 10.1016/0550-3213(96)00418-X. [Erratum: Nucl.Phys.B 531, 655–655 (1998)].
- [24] A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607(7917):60–68, 2022. doi: 10.1038/s41586-022-04892-x.

BIBLIOGRAPHY

- [25] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008. doi: 10.1088/1748-0221/3/08/S08002.
- [26] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003.
- [27] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.
- [28] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008. doi: 10.1088/1748-0221/3/08/S08005.
- [29] G. L. Bayatian et al. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. 2006.
- [30] W. Adam et al. The CMS Phase-1 Pixel Detector Upgrade. *JINST*, 16(02):P02027, 2021. doi: 10.1088/1748-0221/16/02/P02027.
- [31] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013. doi: 10.1088/1748-0221/8/09/P09009.
- [32] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003.
- [33] T. Speer, W. Adam, R. Fruhwirth, A. Strandlie, T. Todorov, and M. Winkler. Track reconstruction in the CMS tracker. *Nucl. Instrum. Meth. A*, 559:143–147, 2006. doi: 10.1016/j.nima.2005.11.207.
- [34] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014. doi: 10.1088/1748-0221/9/10/P10009.

BIBLIOGRAPHY

- [35] T. Miao, H. Wenzel, F. Yumiceva, and N. Leioatts. Beam position determination using tracks. 8 2007.
- [36] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *IEEE Proc.*, 86(11):2210–2239, 1998. doi: 10.1109/5.726788.
- [37] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. *J. Phys. G*, 34:N343, 2007. doi: 10.1088/0954-3899/34/12/N01.
- [38] Emilio Meschi, T. Monteiro, Christopher Seez, and Pratibha Vikas. Electron Reconstruction in the CMS Electromagnetic Calorimeter. 6 2001.
- [39] Albert M Sirunyan et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16(05):P05014, 2021. doi: 10.1088/1748-0221/16/05/P05014.
- [40] Albert M Sirunyan et al. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15(09):P09018, 2020. doi: 10.1088/1748-0221/15/09/P09018.
- [41] Gavin P. Salam. Towards Jetography. *Eur. Phys. J. C*, 67:637–686, 2010. doi: 10.1140/epjc/s10052-010-1314-6.
- [42] Serguei Chatrchyan et al. Identification of b-Quark Jets with the CMS Experiment. *JINST*, 8:P04013, 2013. doi: 10.1088/1748-0221/8/04/P04013.
- [43] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018. doi: 10.1088/1748-0221/13/05/P05011.
- [44] Emil Bols, Jan Kieseler, Mauro Verzetti, Markus Stoye, and Anna Stakia. Jet Flavour Classification Using DeepJet. *JINST*, 15(12):P12012, 2020. doi: 10.1088/1748-0221/15/12/P12012.

BIBLIOGRAPHY

- [45] Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. Technical report, CERN, Geneva, 2016. URL <https://cds.cern.ch/record/2196972>.
- [46] A. M. Sirunyan et al. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(10):P10005, 2018. doi: 10.1088/1748-0221/13/10/P10005.
- [47] Met analysis. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkbookMetAnalysis>. Accessed: 2019-06-28.
- [48] Haowen Deng, Youyou Zhou, Lin Wang, and Cheng Zhang. Ensemble learning for the early prediction of neonatal jaundice with genetic features. *BMC Medical Informatics and Decision Making*, 21, 12 2021. doi: 10.1186/s12911-021-01701-9.
- [49] Patrick T. Komiske, Eric M. Metodiev, and Matthew D. Schwartz. Deep learning in color: towards automated quark/gluon jet discrimination. *JHEP*, 01:110, 2017. doi: 10.1007/JHEP01(2017)110.
- [50] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. doi: 10.1007/JHEP04(2015)040.
- [51] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017. doi: 10.1140/epjc/s10052-017-5199-5.
- [52] Juan Rojo et al. The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II. *J. Phys. G*, 42:103103, 2015. doi: 10.1088/0954-3899/42/10/103103.
- [53] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. 2015.
- [54] C. Diez and M. Seidel. Powheg+Pythia8 $t\bar{t}$: tuning α_S^{ISR} and h_{damp} .

BIBLIOGRAPHY

- https://indico.cern.ch/event/567994/contributions/2299854/attachments/1336057/2009649/tt_tuning_mseidel.pdf.
- [55] A. M. Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1), Jan 2020. doi: 10.1140/epjc/s10052-019-7499-4. URL <https://doi.org/10.1140/epjc/s10052-019-7499-4>.
- [56] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070.
- [57] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. doi: 10.1007/JHEP06(2010)043.
- [58] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191: 159–177, 2015. doi: 10.1016/j.cpc.2015.01.024.
- [59] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76(3):155, 2016. doi: 10.1140/epjc/s10052-016-3988-x.
- [60] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1):4, 2020. doi: 10.1140/epjc/s10052-019-7499-4.
- [61] P. Fackeldey B. Fischer T. Lange D. Noll et al. L. Cadamuro, A. Carvalho. Hh inference tools. <https://cms-hh.web.cern.ch/cms-hh/tools/inference/index.html>, 2021.

BIBLIOGRAPHY

- [62] Karl Ehatäht and Christian Veelken. Stitching Monte Carlo samples. *Eur. Phys. J. C*, 82(5):484, 2022. doi: 10.1140/epjc/s10052-022-10407-9.
- [63] <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO>.
- [64] <http://pdglive.lbl.gov/BranchingRatio.action?desig=8&parCode=S043>.
- [65] <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec>.
- [66] https://indico.cern.ch/event/841566/contributions/3565385/attachments/1914850/3185933/Drell-Yan_jets_crosssection_September2019_update.pdf.
- [67] https://indico.cern.ch/event/746829/contributions/3138541/attachments/1717905/2772129/Drell-Yan_jets_crosssection.pdf.
- [68] <https://twiki.cern.ch/twiki/bin/view/CMS/SummaryTable1G25ns>.
- [69] <https://twiki.cern.ch/twiki/bin/view/CMS/HowToGenXSecAnalyzer>.
- [70] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. doi: 10.23731/CYRM-2017-002.
- [71] https://indico.cern.ch/event/448517/contributions/1943045/attachments/1164999/1679225/Long_Generators_WZxsec_05_10_15.pdf.
- [72] CERN Twiki. E/gamma physics object group. <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaPOG>, 2023.
- [73] CERN Twiki. Muon physics object group. <https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG>, 2022.
- [74] Albert M Sirunyan et al. Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 81(4):378, 2021. doi: 10.1140/epjc/s10052-021-09014-x.

- [75] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939785. URL <http://doi.acm.org/10.1145/2939672.2939785>.
- [76] Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. 6 2022.
- [77] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.
- [78] CERN Twiki. Jetmet physics object group. <https://twiki.cern.ch/twiki/bin/view/CMS/JetMET>, 2023.
- [79] CERN Twiki. b tag and vertexing physics object group. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG>, 2022.
- [80] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013. doi: 10.1007/JHEP09(2013)029.
- [81] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100: 242001, 2008. doi: 10.1103/PhysRevLett.100.242001.
- [82] M. Erdmann, E. Geiser, Y. Rath, and M. Rieger. Lorentz Boost Networks: Autonomous Physics-Inspired Feature Engineering. *JINST*, 14(06):P06006, 2019. doi: 10.1088/1748-0221/14/06/P06006.
- [83] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

BIBLIOGRAPHY

- [84] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL <https://doi.org/10.1023/A:1010933404324>.
- [85] Partha Konar, Kyoungchul Kong, and Konstantin T. Matchev. $\sqrt{\hat{s}}_{min}$: A Global inclusive variable for determining the mass scale of new physics in events with missing energy at hadron colliders. *JHEP*, 03:085, 2009. doi: 10.1088/1126-6708/2009/03/085.
- [86] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. In *PHYSTAT 2011*, pages 115–120, 2011. doi: 10.5170/CERN-2011-006.115.
- [87] Serguei Chatrchyan et al. Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 710:26–48, 2012. doi: 10.1016/j.physletb.2012.02.064.
- [88] Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, 2011. URL <https://cds.cern.ch/record/1379837>.
- [89] Custom hh physics model implementing gluon gluon fusion ($ggf/gg hh$), vector boson fusion ($vbf/qq hh$) and ν boson associated production (νhh) modes. https://gitlab.cern.ch/hh/tools/inference/-/blob/master/dhi/models/hh_model.py. Accessed: 2021.
- [90] Georges Aad et al. Searches for Higgs boson pair production in the $hh \rightarrow b b \tau \tau, \gamma \gamma W W^*, \gamma \gamma b b, b b b b$ channels with the ATLAS detector. *Phys. Rev. D*, 92:092004, 2015. doi: 10.1103/PhysRevD.92.092004.
- [91] Georges Aad et al. Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B*, 800:135103, 2020. doi: 10.1016/j.physletb.2019.135103.

- [92] Georges Aad et al. Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 07:108, 2020. doi: 10.1007/JHEP07(2020)108. [Erratum: JHEP 01, 145 (2021), Erratum: JHEP 05, 207 (2021)].
- [93] Georges Aad et al. Search for non-resonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 801:135145, 2020. doi: 10.1016/j.physletb.2019.135145.
- [94] Search for nonresonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 1 2023.
- [95] Georges Aad et al. Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 106(5):052001, 2022. doi: 10.1103/PhysRevD.106.052001.
- [96] Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector. 9 2022.
- [97] Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV. 11 2022.
- [98] Albert M Sirunyan et al. Search for Higgs boson pair production in the $b\bar{b}\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev. D*, 96(7):072004, 2017. doi: 10.1103/PhysRevD.96.072004.
- [99] Vardan Khachatryan et al. Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV. *Phys. Rev. D*, 94(5):052012, 2016. doi: 10.1103/PhysRevD.94.052012.
- [100] Albert M Sirunyan et al. Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 122(12):121803, 2019. doi: 10.1103/PhysRevLett.122.121803.

BIBLIOGRAPHY

- [101] Search for nonresonant Higgs boson pair production in the four leptons plus two b jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. 6 2022.
- [102] Search for Higgs boson pairs decaying to $WWWW$, $WW\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV. 6 2022.
- [103] Albert M Sirunyan et al. Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03:257, 2021. doi: 10.1007/JHEP03(2021)257.
- [104] Armen Tumasyan et al. Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at $s=13$ TeV. *Phys. Rev. Lett.*, 129(8):081802, 2022. doi: 10.1103/PhysRevLett.129.081802.
- [105] Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks. 5 2022.
- [106] Guido Altarelli and Ferruccio Feruglio. Discrete Flavor Symmetries and Models of Neutrino Mixing. *Rev. Mod. Phys.*, 82:2701–2729, 2010. doi: 10.1103/RevModPhys.82.2701.
- [107] Ernest Ma. Non-Abelian discrete flavor symmetries. 5 2007.
- [108] Hajime Ishimori, Tatsuo Kobayashi, Hiroshi Ohki, Yusuke Shimizu, Hiroshi Okada, and Morimitsu Tanimoto. Non-Abelian Discrete Symmetries in Particle Physics. *Prog. Theor. Phys. Suppl.*, 183:1–163, 2010. doi: 10.1143/PTPS.183.1.
- [109] Gautam Bhattacharyya, Philipp Leser, and Heinrich Pas. Novel signatures of the Higgs sector from S_3 flavor symmetry. *Phys. Rev. D*, 86:036009, 2012. doi: 10.1103/PhysRevD.86.036009.
- [110] Gautam Bhattacharyya, Philipp Leser, and Heinrich Pas. Exotic Higgs boson decay modes as a harbinger of S_3 flavor symmetry. *Phys. Rev. D*, 83:011701, 2011. doi: 10.1103/PhysRevD.83.011701.

BIBLIOGRAPHY

- [111] P. A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020. doi: 10.1093/ptep/ptaa104.
- [112] F. J. Botella, G. C. Branco, M. Nebot, and M. N. Rebelo. Flavour Changing Higgs Couplings in a Class of Two Higgs Doublet Models. *Eur. Phys. J. C*, 76(3):161, 2016. doi: 10.1140/epjc/s10052-016-3993-0.
- [113] Manfred Lindner, Moritz Platscher, and Farinaldo S. Queiroz. A Call for New Physics : The Muon Anomalous Magnetic Moment and Lepton Flavor Violation. *Phys. Rept.*, 731:1–82, 2018. doi: 10.1016/j.physrep.2017.12.001.
- [114] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.
- [115] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014. doi: 10.1007/JHEP02(2014)057.
- [116] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.*, 185:2250–2300, 2014. doi: 10.1016/j.cpc.2014.04.012.
- [117] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Was, and Z. Was. Universal Interface of TAUOLA Technical and Physics Documentation. *Comput. Phys. Commun.*, 183:821–843, 2012. doi: 10.1016/j.cpc.2011.12.009.
- [118] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012. doi: 10.1140/epjc/s10052-012-1896-2.
- [119] Glen Cowan. Discovery sensitivity for a counting experiment with background uncertainty. <https://www.pp.rhul.ac.uk/~cowan/stat/medsig/medsigNote.pdf>. Accessed: 2012-05-30.

BIBLIOGRAPHY

- [120] Adam Kardos, Zoltan Trocsanyi, and Costas Papadopoulos. Top quark pair production in association with a Z-boson at NLO accuracy. *Phys. Rev. D*, 85:054015, 2012. doi: 10.1103/PhysRevD.85.054015.
- [121] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 07:018, 2011. doi: 10.1007/JHEP07(2011)018.
- [122] Byron P. Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees, an alternative to artificial neural networks. *Nucl. Instrum. Meth. A*, 543(2-3):577–584, 2005. doi: 10.1016/j.nima.2004.12.018.
- [123] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015. doi: <https://doi.org/10.1038/nature14539>.
- [124] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 3 2007.
- [125] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, and Matthieu Devin et.al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>. Software available from tensorflow.org.
- [126] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014. URL <https://arxiv.org/abs/1412.6980>.
- [127] The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2272264>.
- [128] Mark Prydderch. Cbc3.1 user manual. 2019. URL http://www.hep.ph.ic.ac.uk/ASIC/CBC_documentation/CBC3p1_User_Manual_V1p4.pdf.
- [129] Hendrik Jansen et al. Performance of the EUDET-type beam telescopes. *EPJ Tech. Instrum.*, 3(1):7, 2016. doi: 10.1140/epjti/s40485-016-0033-2.

BIBLIOGRAPHY

- [130] Y. Liu et al. EUDAQ2—A flexible data acquisition software framework for common test beams. *JINST*, 14(10):P10033, 2019. doi: 10.1088/1748-0221/14/10/P10033.
- [131] P. Baesso, D. Cussans, and J. Goldstein. The AIDA-2020 TLU: a flexible trigger logic unit for test beam facilities. *JINST*, 14(09):P09019, 2019. doi: 10.1088/1748-0221/14/09/P09019.
- [132] M. Benoit et al. The FE-I4 Telescope for particle tracking in testbeam experiments. *JINST*, 11(07):P07003, 2016. doi: 10.1088/1748-0221/11/07/P07003.
- [133] M. Pesaresi, M. Barros Marin, G. Hall, M. Hansen, G. Iles, A. Rose, F. Vasey, and P. Vichoudis. The fc7 amc for generic daq amp; control applications in cms. *Journal of Instrumentation*, 10(03):C03036, mar 2015. doi: 10.1088/1748-0221/10/03/C03036. URL <https://dx.doi.org/10.1088/1748-0221/10/03/C03036>.
- [134] The Tracker Group of the CMS Collaboration. Ph2_acf. https://gitlab.cern.ch/cms_tk_ph2/Ph2_ACF. Accessed: 2021.
- [135] Claus Kleinwort. General Broken Lines as advanced track fitting method. *Nucl. Instrum. Meth. A*, 673:107–110, 2012. doi: 10.1016/j.nima.2012.01.024.
- [136] Simon Spannagel and Hendrik Jansen. Gbl track resolution calculator v2.0, April 2016. URL <https://doi.org/10.5281/zenodo.48795>.