



National Taiwan University

Department of Physics

Doctoral Dissertation

**Measurements of the associated VH production
cross section in the $H \rightarrow \gamma\gamma$ channel at $\sqrt{s} = 13$ TeV
in the CMS experiment**

You-Ying Li

Advisor: Stathes Paganis, Ph.D.



August 2021

致謝

此博士論文花了大概七個月的時間撰寫，其內容紀錄了我博士班六年以來，對粒子物理、希格斯玻色子、CMS實驗的了解以及我主要的研究內容，而能夠讓我有如此深入的了解以及研究結果，全都要感謝這六年以來，教導我、與我相處共事並在背後支持我鼓勵我的人。而首先，我想要感謝我的指導教授裴思達，提供我許多獨特且有用的物理觀點與思維，以及值得學習的研究經驗，並幫助我爭取到 VH production in $H \rightarrow \gamma\gamma$ channel 的題目，讓我愛上了 VH production 的美，更不乏提供我難能可貴的機會到CERN移地研究，增加與國外科學家交流的經驗，真誠得感謝您這六年來的幫助。

而在研究方面，非常感謝 Andreas Psallidas 在我博士生涯前期帶領我熟悉 CMS 實驗的分析工具與技術，以及Hgg團隊所使用的 flashgg 架構，讓我在生涯後半段能熟練得運用此架構作分析。也感謝亦師亦友的陳凱風、呂榮祥教授，提供我許多不論是分析上的建議或者是處理事情的方法。另外，非常感謝 Hgg 團隊裡，Julie Malcles 耐心得協助我處理 diphoton vertexing 的工作；Edward Scott 指引我 VH production 分析的大方向；Jonathon Langford 指導我 Higgs combination fit 的技術。也感謝 Joseph Davies 提供 hadronic VH production 的分析資料去補足此博士論文的不足，更特別感謝 Rajdeep Chatterjee 能和我一起討論 VH production 的分析策略，不吝嗇得提供我在光子研究方面的相關經驗，也謝謝其他 Hgg 團隊的夥伴們給予我許多寶貴的建議。

除了研究之外，在我首次的瑞士移地研究時，我一個人努力適應異地生活之際，能夠陪我聊天、分享故事的楊宜知、林子路。更謝謝第二次移地研究時，一起真的是共患難的陳溥生，以及一起聊天、分享有趣的點點滴滴，還一起玩什麼Play什麼Station 4 去拯救某宇宙的余佩容，你們讓我體會到原來身處異地也能夠這麼快樂，有你們這兩個室友，我真的很幸運，我永遠無法忘記與你們相處的那些美好時光！另外，謝謝蔡連勝、林俊廷、陳柏安和我一起在晚上的時候到楓葉世界冒險，還有高裕維、粟星輔、莊振宇、石琳、陳奕安、吳桐、鄭宇晉以及其他高能實驗室的夥伴們，還陪著我在306到晚上，以及常常跟我互相討論物理的林杰還有跟我一起打羽毛球的陳昱潭，真的很謝謝你們！

最後感謝我的父母，沒有你們在背後支持著，在我無助的時候鼓勵我安慰我，我真的沒辦法走到這一步！

Abstract

Measurements of the associated VH production cross section with the Higgs boson decaying to a pair of photons are revealed. The measurements use events collected from the CMS experiment with proton-proton collisions at a center-of-mass energy of 13 TeV from 2016 to 2018, giving an integrated luminosity of 137 fb^{-1} . The analysis strategy uses the simplified template cross section (STXS) framework to improve the sensitivity of the associated VH production and provide a measurement of its kinematic structure. The framework splits the cross section of the associated VH production into several bins with different productions and kinematics. At the same time, the events are divided into different analysis categories to enrich their targeting bins. Boosted Decision Tree technique is also applied in each analysis category to optimize the sensitivity. Eventually, through the state of the art simultaneous fit to the diphoton invariant mass distribution in all the analysis categories, the inclusive signal strength modifier of the associated VH production is measured to be $1.34^{+0.36}_{-0.33}$ with an observed significance of 4.7σ . The cross sections of the associated VH production in different productions and kinematic regions are also measured. The results are totally in agreement with the Standard Model prediction.

Keywords: LHC, CMS, Standard Model, VH, diphoton decay, STXS

Contents

Introduction	1
1 Higgs boson and associated VH production	3
1.1 Standard Model	4
1.1.1 Particle	4
1.1.2 Interaction	7
1.2 Higgs mechanism	13
1.2.1 Spontaneous symmetry breaking and Higgs mechanism	14
1.2.2 Higgs mechanism in the Standard Model	16
1.2.3 Higgs hunting	20
1.3 Higgs boson at the LHC	22
1.3.1 Production of the Higgs boson at the LHC	23
1.3.2 Decay of the Higgs boson	25
1.3.3 What is next ?	29
1.4 Associated VH production and STXS	31
2 The CMS experiment	36
2.1 Large hadron collider	36
2.1.1 Machine design	37
2.1.2 Luminosity and pileup	38
2.1.3 Proton injection and operational cycle	40
2.2 Compact muon solenoid	41
2.2.1 Coordinate system	42
2.2.2 Inner tracker system	43
2.2.3 Electromagnetic calorimeter	44
2.2.4 Hadron calorimeter	47
2.2.5 Solenoid magnet system	49
2.2.6 Muon detector	49
2.3 Trigger	52
2.3.1 Level-1 trigger	53
2.3.2 High Level Trigger	54

3	Event reconstruction and selection	55
3.1	Particle flow element	56
3.1.1	Inner track and vertex reconstruction	56
3.1.2	Calorimeter cluster reconstruction	58
3.2	Muon	60
3.2.1	Muon reconstruction	60
3.2.2	Muon identification and isolation	61
3.3	Photon and Electron	63
3.3.1	Photon/Electron reconstruction	64
3.3.2	Energy calibration	68
3.3.3	Diphoton vertex identification	69
3.3.4	Photon identification	72
3.3.5	Diphoton identification	75
3.3.6	Electron identification	77
3.4	Jet	77
3.4.1	Jet reconstruction	78
3.4.2	Jet identification	79
3.5	Missing energy	82
4	Event categorization	86
4.1	Categorization for the VH STXS stage 0	86
4.1.1	ZH Leptonic category	88
4.1.2	WH Leptonic category	89
4.1.3	VH MET category	96
4.1.4	VH Hadronic category	100
4.2	Multivariate analysis	106
4.2.1	Boost decision tree technique	108
4.2.2	VH Leptonic BDT	112
4.2.3	VH Hadronic BDT	118
4.3	Categorization for the VH STXS stage 1.2	124
4.3.1	Migration of the STXS stage 1.2 bins in the VH Leptonic categories	126
4.3.2	Analysis categories at the stage 1.2	129
5	Signal and background modelling	135
5.1	Signal modelling	135
5.2	Background modelling	136
5.3	Systematic uncertainties	141
5.3.1	Theoretical uncertainties	143
5.3.2	Experimental uncertainties	145

6 Results	148
6.1 Statistical methodology	149
6.2 Signal strength modifier	152
6.3 Simplified template cross sections	154
6.3.1 STXS at Stage 0	156
6.3.2 SXTS at Stage 1.2	156
Conclusion	162
A STXS results	164
Bibliography	166

List of Figures

1.1	Elementary particles in the Standard Model.	5
1.2	Interactions of the quantum chromodynamics.	9
1.3	Interactions of the quantum electromagnetism.	11
1.4	Potential of the complex scalar field.	15
1.5	Interaction of the gauge boson B and the Higgs boson after $U(1)$ spontaneous symmetry breaking.	16
1.6	Three-point interactions of the $W^\pm$, Z and Higgs bosons.	18
1.7	Origin of fermion mass and Yukawa interactions.	20
1.8	Exclusion of the Higgs boson mass from the LEP and Tevatron collaborations. . .	21
1.9	Exclusion of the Higgs boson mass from the Gfitter group by the LEP, Tevatron and other experiments.	22
1.10	Observation of the Higgs boson under the Standard Model from the CMS and ATLAS collaborations.	23
1.11	Four relevant Higgs boson productions at the LHC.	25
1.12	Observation of the Higgs boson in the $\gamma\gamma$, ZZ and WW channels.	28
1.13	Observation of the Higgs boson in the $\tau\tau$, bb channels and $t\bar{t}H$ production. . . .	29
1.14	The best fit estimates of the Higgs boson couplings to fermions and bosons under the κ -framework.	30
1.15	Feynman diagrams of three kinds of the associated VH productions.	32
1.16	Two possible BSM processes in the associated VH production.	33
1.17	Schematic overview of the simplified template cross section framework.	34
1.18	Observed STXS of the associated VH production with the W and Z boson leptonic decay in the $H \rightarrow b\bar{b}$ channel.	35
1.19	Observed signal strength modifiers and STXS of the associated VH production with the W and Z boson leptonic decay in the $H \rightarrow W^+W^-$ channel with the W leptonic decay.	35
2.1	LHC tunnel and beam pipe.	37
2.2	Cross section of the cryostat system, where the dipole magnets are surrounded by the superfluid helium.	38
2.3	Schematic layout of the LHC (Beam 1 - clockwise, Beam 2 anticlockwise) from Ref [48].	39

2.4	Integrated luminosity with variable time and pileup distributions in 2016, 2017 and 2018.	40
2.5	Cross section diagram of the Compact Muon Solenoid with a sandwich structure.	42
2.6	Diagram of the coordinate of the CMS detector, where the x -axis points toward the center of the LHC ring, the y -axis points upwards the sky and the z -axis points toward counterclockwise beam direction.	43
2.7	Arrangement of the Pixel detector after phase 1 upgrade (top) with respect to original one (bottom) [53].	44
2.8	Sketch of one quarter of the silicon strip tracker system in $r - z$ view.	45
2.9	Photogram of a homogeneous lead tungstate crystal (PbWO_4) in the EE, where the end of the crystal is a vacuum phototriode (VPT) photodetector.	45
2.10	Sketch of one quarter of the ECAL in $y - z$ view with the axis parallel to the beam (z) running horizontally and y increasing upward.	46
2.11	Sketch of the tower mapping in $R - z$ of the HCAL.	48
2.12	Diagram of one half of the solenoid magnet system in the CMS detector.	49
2.13	Sketch of one quarter of the CMS detector in $R - z$ view.	50
2.14	Cross section diagram and layout of the drift tubes in the muon detector.	51
2.15	Cut-away diagram and cross-sectional views of a cathode strip chamber.	52
2.16	Schematic view of a generic barrel RPC with 2 roll partitions [60].	52
3.1	Schematic of the Particle Flow algorithm.	55
3.2	The tracking efficiency as a function of the number of primary vertices in the "all tracks" collection.	58
3.3	The efficiency of the "medium" muon ID for the 2017 legacy data and simulation.	62
3.4	Diphoton invariant mass distribution for the $H \rightarrow \gamma\gamma$ simulated sample with $m_H = 125$ GeV.	64
3.5	Overview of an event with converted and unconverted photons, and an electron including bremsstrahlung tangents under the inner tracker system and the ECAL surface.	65
3.6	Distribution of $\Delta\eta = \eta_{\text{seed-cluster}} - \eta_{\text{cluster}}$ versus $\Delta\phi = \phi_{\text{seed-cluster}} - \phi_{\text{cluster}}$ for simulated electrons.	66
3.7	Electron reconstruction efficiency versus η in data (upper panel) and data-to-simulation efficiency ratios (lower panel).	67
3.8	Ratio of the true to the reconstructed electron energy.	69
3.9	Comparison of the dielectron invariant mass spectra in data and simulation after applying energy scale corrections to data and energy smearing to the simulation.	70
3.10	Diagram of the diphoton system in the CMS detector.	71
3.11	Fraction of $ z_{\text{selected}} - z_{\mu\mu} < 1$ cm versus $p_T^{\mu\mu}$ in the $Z \rightarrow \mu\mu$ and simulated $H \rightarrow \gamma\gamma$ events.	72
3.12	EM Shower spread and particle activities near photons.	73
3.13	Architecture of the shower shape and isolation variable corrections.	75
3.14	Comparison of the photon ID BDT distribution in data and simulated events.	76

3.15	Data-to-simulation comparison of the β and PU jet ID BDT output for AK4 jets with $30 < p_T < 50$ GeV in the tracker volume.	81
3.16	Diagram of a heavy-flavour jet with a secondary vertex.	81
3.17	Distribution of the DeepCSV discriminator values for the jets from different flavours in the $t\bar{t}$ events.	82
3.18	Distribution of $\vec{p}_T^{\text{miss}}$ in $Z \rightarrow \mu\mu$ events in data and simulation.	83
3.19	The p_T^{miss} and jet ϕ distributions with and without filtering algorithms applied.	85
4.1	Schematic diagram for the STXS stage 0 bins.	87
4.2	Event flow in the $H \rightarrow \gamma\gamma$ analysis.	88
4.3	The event topology of the $Z(\rightarrow \ell\ell)H$ process in the CMS detector.	88
4.4	Data-and-simulation comparison of the photon kinematics and photon ID BDT distributions in the ZH Leptonic category.	90
4.5	Data-and-simulation comparison of the charged lepton kinematics in the ZH Leptonic category.	91
4.6	Event topology of the $W(\rightarrow \ell\nu)H$ process in the CMS detector.	92
4.7	Data-and-simulation comparison of the diphoton mass before and after two selections in the WH Leptonic category.	92
4.8	Data-and-simulation comparison between of the photon-electron pair and diphoton mass in the WH Leptonic category.	93
4.9	Data-and-simulation comparison of the photon kinematics and photon ID BDT distributions in the WH Leptonic category.	94
4.10	Data-and-simulation comparison of the lepton, missing transverse energy kinematics and other distributions in the WH Leptonic category.	95
4.11	Event topology of the $Z(\rightarrow \nu\nu)H$ process and $W(\rightarrow \ell\nu)H$ process in the CMS detector.	96
4.12	Data-and-simulation comparison of the diphoton mass, $m_{\gamma\gamma}$ in the VH MET category.	97
4.13	The diagram of the signal-like region and the sideband region with a boundaries of -0.7 in the minimal and maximal photon ID BDT scores.	97
4.14	Photon ID BDT distributions with the different p_T (left) and η (right) regions for the fake photons in the γ + jets sample.	98
4.15	Fit result of the fake photon ID BDT PDF.	99
4.16	Data-and-simulation comparison of the maximal photon ID BDT distributions before and after applying the per-event weight in the VH MET category.	100
4.17	Data-and-simulation comparison of the minimal photon ID BDT distributions before and after fitting in the VH MET category.	101
4.18	Data-and-simulation comparison of the photon kinematics and photon ID BDT distributions in the VH MET category.	102
4.19	Data-and-simulation comparison of the jet, missing transverse energy and VH topology associated distributions in the VH MET category.	103
4.20	Event topology of the $V(\text{had.})H$ process in the CMS detector.	104

4.21	Schematic of the data-driven method in the VH Hadronic category.	105
4.22	The fake factors (left) and QCD fractions (right) as a function of photon transverse momenta in different η regions for 2017.	106
4.23	Comparison between data and simulation/data-driven samples of the photon and jet associated distributions in the VH Hadron category.	107
4.24	Structure of the common decision tree.	108
4.25	Schematic of the Boosted Decision Tree. There are N weak trees combined into a BDT.	109
4.26	Comparison of the VH MET BDT performance between the training and test sample in terms of the BDT score distributions and the ROC curves.	117
4.27	Comparison of the WH Leptonic BDT performance between the training and test sample in terms of the BDT score distributions and the ROC curves.	117
4.28	Comparison of the ZH Leptonic BDT performance between the training and test sample in terms of the BDT score distributions and the ROC curves.	118
4.29	Comparison between the $q\bar{q} \rightarrow ZH$ and $gg \rightarrow ZH$ processes with $m_H = 125$ GeV in terms of the ZH Leptonic BDT and VH MET BDT distributions.	118
4.30	Data-and-simulation comparison of the distributions of VH MET BDT, WH Leptonic BDT and ZH Leptonic BDT score.	119
4.31	Comparison between data and simulation/data-driven samples of the diphoton BDT distribution and ROC curve.	122
4.32	Schematic diagram and distribution of the helicity angle in the VH system. . . .	123
4.33	Comparison between data and simulation/data-driven samples of the VH Hadronic BDT distribution and ROC curve.	124
4.34	Schematic of the VH leptonic STXS stage 1.2 bins from the VH leptonic STXS stage 0 bins split by the kinematics such as the p_T^V and jet multiplicity.	125
4.35	Comparison between the ratio of the reconstructed and particle-level W/Z p_T and of the reconstructed diphoton and particle-level W/Z p_T in the VH MET and WH Leptonic and ZH Leptonic categories.	127
4.36	Competition of the STXS stage 1.2 bins with $m_H = 125$ GeV split by the p_T^V of the particle-level W or Z boson.	128
4.37	Schematic of the analysis sub-categories divided by applying the dedicated boundary optimization procedure to the corresponding BDT distribution.	129
4.38	Data-and-simulation comparison of the WH Leptonic BDT distributions in three WH Leptonic categories for the stage 1.2.	132
4.39	Data-and-simulation comparison of the corresponding BDT distributions in three ZH Leptonic, VH MET and VH Hadronic categories.	133
5.1	Signal models and corresponding simulated events with $m_H = 125$ GeV in one of the VH MET and VH hadronic categories.	137
5.2	Signal models and corresponding simulated events with $m_H = 125$ GeV in one of the WH Leptonic and ZH Leptonic categories.	137

5.3	Schematic of the diphoton mass distribution to which there are three parametric functions fitted.	138
5.4	Diagram of the profile likelihood curve and the likelihood curves with nuisance parameters fixed to the best fit and arbitrary values.	139
5.5	Schematic of the discrete profile method.	140
5.6	Background models from four families in the VH MET and VH Hadronic categories for the stage 1.2.	142
5.7	Background models from four families in the WH Leptonic and ZH Leptonic categories for the stage 1.2.	142
6.1	Schematic of the composition between a single fit and a simultaneous fit.	149
6.2	Profile likelihood-ratio curve.	151
6.3	Square of the one random number from the standard normal Gaussian distribution, and the χ^2 distribution with one degree of freedom.	152
6.4	Inclusive best fit of the signal-plus-background model to the diphoton invariant mass distribution by combining all STXS bins, all analysis categories and three years.	153
6.5	Observed results of the signal strength modifiers of four principal Higgs boson production modes.	154
6.6	Main systematic uncertainties of the signal strength modifiers for four principal Higgs boson production modes.	155
6.7	VH best fit of the signal-plus-background model to the diphoton invariant mass distribution in data by combining all STXS bins in all VH analysis categories. . .	155
6.8	Schematic diagram for the STXS stage 0 bins.	156
6.9	Observed results of the STXS at the stage 0 in the $H \rightarrow \gamma\gamma$ channel with $m_H = 125.38$ GeV and without $ y_H > 2.5$ region.	157
6.10	Observed results of the STXS divided by the STXS from the Standard Model prediction at the stage 0 in the $H \rightarrow \gamma\gamma$ channel with profiled m_H and without $ y_H > 2.5$ region for only 2016 year scenario.	158
6.11	Schematic diagram for the STXS stage 1.2 bins [72]. Some bins are merged by placing similar but light color areas behind these STXS bins.	158
6.12	Composition of the analysis categories in terms of a merged set of STXS bins. . .	160
6.13	Observed results of the STXS at the stage 1.2 in the $H \rightarrow \gamma\gamma$ channel with $m_H = 125.38$ GeV and without $ y_H > 2.5$ region.	161

List of Tables

1.1	Representation of the Lorentz group $SU(2)_L \times SU(2)_R$ and corresponding subgroup $SU(2)_{\text{spin}}$	4
1.2	The information of three fundamental interactions in the Strong Model.	7
1.3	Relation between electroweak quantum numbers.	13
1.4	Mass of the Higgs boson, $W^\pm$ and Z bosons, photon and their relations to the electroweak gauge bosons such as W and B.	19
1.5	Branching fraction of the Higgs boson decay channels.	26
1.6	Cross sections and uncertainties of the associated VH productions.	32
2.1	Parameters of the LHC for 2016, 2017 and 2018 in [49].	40
2.2	Lowest unprescaled L1 single e/γ seeds corresponding to E_T thresholds and prescaled L1 double e/γ seeds with asymmetric E_T thresholds for 2016, 2017 and 2018.	53
2.3	Content of the diphoton trigger requirements.	54
3.1	Selection criteria at "medium" working point in the muon identification.	62
3.2	Definition of the shower shape and isolation variables used in the photon ID BDT training.	74
3.3	Requirements of the shower shape and isolation variables in the diphoton selections.	77
3.4	Definition of the track quality and track-cluster matching variables and isolation variables used in the electron ID BDT training.	78
3.5	Content of jet identification criteria.	79
3.6	Definition of the main track-based and jet shape variables used in the pileup jet ID training.	80
3.7	Content of the tight pileup jet identification criteria.	80
3.8	Content of anomalous p_T^{miss} event filter	84
4.1	Four VH analysis categories and corresponding targeted STXS bins.	87
4.2	Pearson correlation coefficients between the prompt/fake photon ID BDT scores and other kinematic variables.	98

4.3	List of all training samples for the VH MET BDT, WH Leptonic BDT, ZH Leptonic BDT.	114
4.4	List of all training variables for the VH MET BDT, WH Leptonic BDT, ZH Leptonic BDT.	116
4.5	Optimized hyperparameters for three AdaBoost classifiers.	117
4.6	Optimized hyperparameters for the diphoton BDT classifiers.	122
4.7	Cross section times branching ratios for the VH leptonic STXS bins.	126
4.8	List of the final VH analysis categories for the step 1 and 2.	131
4.9	Expected number of signal events for $m_H = 125$ GeV in the analysis categories targeting VH production with the W or Z leptonic or hadronic decays.	134
5.1	Collected functions in each VH analysis category for the STXS stage 1.2.	141
6.1	Results of the STXS stage 0.	157
6.2	Results of the STXS stage 1.2.	159
A.1	Results of the STXS stage 0 [72]	164
A.2	Results of the STXS stage 1.2 [72]	165

Introduction

In 2012, the CMS and ATLAS collaborations at the Large Hadron Collider (LHC) independently announced the discovery of the Higgs boson from its diphoton and four lepton decay channels, using data collected from proton-proton collisions with center-of-mass energies of 7 and 8 TeV [1, 2]. Afterward, understanding the Higgs boson properties and testing its compatibility to the Standard Model become more important. A simply method to reveal the Higgs boson properties is to measure the cross sections of the physical processes involving in the Higgs boson coupling to other Standard Model particles. In particular, the Higgs boson coupling to the weak bosons such as the W and Z bosons further provides understanding of the electroweak symmetry breaking under the Standard Model. Physics beyond the Standard Model hidden in the couplings can be further explored at the same time. Therefore, the analysis in this thesis focuses on the cross-section measurements of the associated VH production also known as the Higgs-strahlung process containing the HWW and HZZ couplings.

In this thesis, the measurements use data collected from the CMS experiment with proton-proton collisions at a center-of-mass energy of 13 TeV during 2016-2018, giving a total luminosity of 137 fb^{-1} . The $H \rightarrow \gamma\gamma$ channel with two isolated photons is chosen. This channel has quite good mass resolution in the diphoton invariant mass distribution due to precise measurement of two photons' energies from the ECAL crystals in the CMS detector. Two isolated and energetic photons also provide pretty different features with respect to fake photons from an amount of the QCD events in the proton-proton collisions. These properties help the analysis identify the $H \rightarrow \gamma\gamma$ signal against these backgrounds to achieve the optimal sensitivity of the cross-section measurements. However, the limited branching ratio of 0.23 % [3] for the diphoton channel also makes the measurements be a difficult challenge.

The analysis strategy in this thesis is based on the previous analyses performed by the CMS Collaboration [4, 5]. However, the relatively large size of data combining three years together allows the VH analysis moving on the simplified template cross section (STXS) framework [6]. The goal of the STXS framework is to maximize the sensitivity of cross-section measurements and to reduce the theoretical dependence at the same time. It splits the VH production process into different STXS bins according to different production mechanisms and kinematics of the weak boson under the particle level. Simultaneously, the analysis designs several analysis categories to enrich these bins through the same definitions under the reconstruction level in order to reduce the theoretical dependence due to the bin mixture in different categories. Besides, this framework allows the analysis using large acceptance correction like multivariate techniques

for background subtraction and exclusion of non-target STXS bins. For the VH analysis, Boost Decision Tree (BDT) classifiers are performed. Finally, the observed cross sections for different STXS bins are estimated by a state-of-the-art simultaneous fit to the diphoton invariant mass distribution in all analysis categories. Through such splitting in the STXS bins, the cross sections in different VH production mechanisms and kinematics of the weak boson can be revealed and the Standard Model can be tested. Results of high- p_T bins further allow us probing the BSM effect.

The thesis is organized as follows. Some theoretical frameworks under particle physics and Higgs boson such as the Standard Model, electroweak symmetry breaking and Higgsstrahlung, the profile of the Higgs boson studies at the LHC and an introduction of the simplified template cross section framework are presented in Chapter 1. The overview of the LHC and CMS detector are used for proton-proton collision and detection introduced in Chapter 2. The overall event reconstruction involving in the VH production process is introduced in Chapter 3, containing the final states of the VH production like muons, electrons, photons, jets and missing transverse energy. The VH production under the STXS and corresponding event categorization are shown in Chapter 4. Signal, background modelling and considered all systematic uncertainties are given in Chapter 5. The observed results of the VH production with $H \rightarrow \gamma\gamma$ are presented in Chapter 6, containing the measurements of the cross sections in different kinematics and of the signal strength modifier for the inclusive VH production.

Chapter 1

Higgs boson and associated VH production

Particle physics is a branch of physics which describes phenomena originating from fundamental particles in Nature and also helps us understand the evolution of our universe from Big Bang. In particle physics, as soon as new particle phenomena are observed from experiments or Nature, physicists try to create new models through mathematical language to explain their underlying mechanism and predict undiscovered ones. Afterwards, they design new experiments to verify the predictions given by these models for either acceptance or rejection and these experimental results give hints to build new models again. After continuously repeating the above processes, currently the most worldwide model supported by physicists is called as *Standard Model*.

The Standard Model is constructed through the language of quantum field theory and a variety of physical parameters from experiments like particle mass and interaction strength constants and so on. It successfully explains almost most of particle physics phenomena and is experimentally supported. Among the Standard Model, a crucial key, which is considered as a holy grail, is *Higgs mechanism*. The Higgs mechanism solves the most important problem from the gauge theory in the Standard Model and also explains the origin of mass for elementary particles. In 2012, the discovery of the Higgs boson predicted by the Higgs mechanism is further a milestone for particle physics. Until now, the latest results of the measurements for the Higgs boson properties in the LHC collaborations almost confirm that the observed Higgs boson satisfies the Standard Model prediction. However, the Higgs mechanism is still insufficient to explain more fundamental questions like the hierarchy problem involving in huge strength difference between the electroweak and gravitational interactions. Therefore, the Higgs mechanism under the electroweak physics needs to be further understood. Therein, the associated VH production plays an important role in understanding this Higgs mechanism.

The following sections will introduce the content of the Standard Model in Section 1.1 and Higgs mechanism in Section 1.2 by referring the textbooks [7, 8] and my understanding, all physical parameters given by Ref [3]. Section 1.3 gives a profile for the recent Higgs boson

searches and measurements at the LHC and finally touches the physics of the associated VH production and the simplified template cross section framework.

1.1 Standard Model

The Standard Model is composed of 18 elementary particles, three kinds of interactions described by gauge theories and all physical parameters given by experiments. The following will individually introduce the particles in Section 1.1.1 and interactions in Section 1.1.2.

1.1.1 Particle

In quantum field theory, particles can be considered as a quantized excitation of physical fields under vacuum states. Propagation and interaction of these particles in a given physical system can be described by a Lorentz invariant quantity or function called as *Lagrangian*. The Lagrangian actually satisfies a basic requirement in physics, *principle of relativity*, namely, the laws of physics are the same in all inertial frames of reference. However, to make the Lagrangian invariant under the Lorentz transformation, fields inside the Lagrangian also have to perform the Lorentz transformation with specific representations under the Lorentz group. From another point of view, this specific representation actually determines the fields' types or, more generally, particles are defined by the Lorentz group.

The Lorentz group contains rotation and boost parts described by a $SO(3, 1)$ group. Furthermore, a general representation of the Lorentz group is expressed with a cross of two irreducible unitary groups of $SU(2)_L \times SU(2)_R$ where L and R are the left-handed and right-handed chirality. Table 1.1 shows the representation of the Lorentz group and its subgroup, $SU(2)$, which defines particles' spin.

Representation of $SU(2)_L \times SU(2)_R$	$(0, 0)$	$(\frac{1}{2}, 0)_L$	$(0, \frac{1}{2})_R$	$(\frac{1}{2}, \frac{1}{2})$
Representation of $SU(2)_{\text{spin}}$	0	$\frac{1}{2}$	$\frac{1}{2}$	1

Table 1.1: Representation of the Lorentz group $SU(2)_L \times SU(2)_R$ and corresponding subgroup $SU(2)_{\text{spin}}$.

Hence, a field following the $(0, 0)$ transformation under the Lorentz group, which is trivial, is regarded as a spin-0 field called as *scalar field* corresponding to a scalar boson; following the $(\frac{1}{2}, 0)_L$ or $(0, \frac{1}{2})_R$ transformation is a spin-1/2 field called as *spinor* with either left-handed or right-handed chirality corresponding to a left-handed and or right-handed fermion¹ which respects the chiral symmetry; following the $(\frac{1}{2}, \frac{1}{2})$ transformation is a spin-1 field called *vector field* corresponding to a vector boson. The following section will sequentially introduce the elementary particles with spin-0, spin-1/2 and spin-1 in the Standard Model summarized in Figure 1.1.

¹Since currently there is no elementary fermion with spin more than $\frac{1}{2}$ in the Standard Model, the following contents no longer emphasize "spin- $\frac{1}{2}$ fermions" but "fermions".

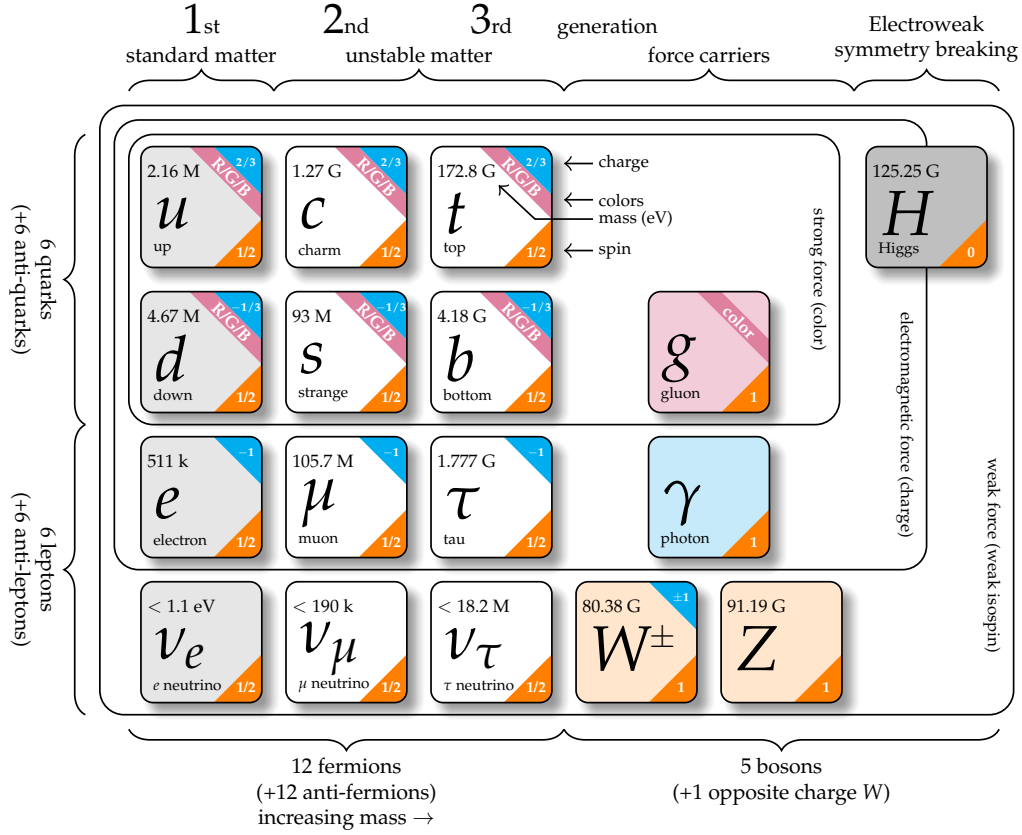


Figure 1.1: Elementary particles in the Standard Model.

Scalar boson (Scalar field)

In the Standard Model, only one fundamental scalar boson observed experimentally is called as *Higgs boson* (H) with a mass of near 125 GeV and without any electric charge, colour charge and parity. It is a product of the spontaneous symmetry breaking occurring under the electroweak symmetry. Hence, measuring the Higgs boson properties is quite important to understand the electroweak broken symmetry. More importantly, only massive particles can directly couple with the Higgs boson with their coupling strength related to the mass of the particles.

Spin-1/2 fermion (Spinor)

In the Standard Model, the fundamental fermions are the main components which construct matter in the universe. According to experimental observations, the fundamental fermions contain 6 quarks and 6 leptons, total 12 particles under the Standard Model. The 6 quarks can further split into three generations, each of which is composed of an up-type and down-type quarks with electric charges of $+2/3$ and $-1/3$, respectively. The leptons can be also separated into charged leptons with a charge of -1 and neutral leptons called neutrinos. It's also observed

that each fermion has different mass (See Figure 1.1). This observation, in principle, violates the global chiral symmetry of the fermions which forbids massive fermions. Therefore, the origin of mass of the fermions implies that there should be a more fundamental theory which is the Higgs mechanism. On the other hand, all fermions have their own anti-particles with opposite sign electric charges but same mass and flavours. The anti-quarks even contain anti-colour charges.

Since the quarks contain three kinds of quantum numbers such as electric charges, colour charges and flavours, they join three kinds of interactions, electromagnetic, weak and strong interactions. The flavours of the up-type quarks in three generations are the up (u) quark, charm (c) quark and top (t) quark, respectively, while the down-type quarks are the down (d) quark, strange quark (s) and bottom (b) quark. Due to the QCD confinement, a particle carrying the colour charge can not exist alone in space, namely, the colourful quarks are never as free particles but bound by the strong interaction at lower energy scale. This implies only colourless particles can get rid of the binding of the strong interaction. Therefore, in Nature or experiments, only colourless particles as *mesons* with quark-antiquark pairs and *baryons* composed of three quarks with the red, green and blue colour charges can be observed. The mesons and baryons can be named as *hadron*, whose electric charges depend on their constituents. However, most of the hadrons are unstable and immediately decay to other particles. Only protons (lifetime $> 10^{29}$ year) and neutrons (lifetime = 879.4 s) are relatively stable. Nevertheless, these quarks still allow asymptotic freedom in high energy with very short distance, where the strength of the strong force is reduced. This is also the reason why protons can be stably accelerated at the LHC and inelastic scatterings where quarks or gluons inside the protons become free-like particles and interact with each other.

On the other hand, in the Standard Model, leptons only join the electromagnetic and weak interactions due to no colour charge. There are three different flavours of charged leptons, the electron (e^-), muon (μ^-) and tau (τ^-), which are sequentially located on three generations, respectively. The electron is a stable particle and dominates the electromagnetic interaction in our life. The muon is heavier than the electron and relatively unstable (lifetime = 2.20×10^{-6} s) and almost decays to a electron. The tau is the heaviest but very unstable lepton with a lifetime of 2.9×10^{-13} s, which does not almost exist in Nature. Since the mass of the tau is larger than some hadrons, it can decay to both charged leptons like electrons and muons, and hadrons like pions and kaons. There are also three neutral leptons without the electromagnetic interaction called electron neutrino (ν_e), muon neutrino (ν_μ) and tau neutrino (ν_τ) in three generations. Since neutrinos only interact with other particles through the weak interaction, they almost penetrate matters without any interaction. It's observed that the flavour oscillation phenomenon of the neutrinos, i.e. time-variation neutrino flavours during the neutrino propagation. This phenomenon originates from the difference between the Hamiltonian (mass) eigenstates as physical states and flavour eigenstates involved in the weak interaction. This verifies that the neutrinos are massive particles but their mass is still ambiguous.

Vector boson

Vector bosons play the role in the force-carry mediums under three kinds of interactions in the Standard Model. When the fermions interact with each other, they actually exchange these force-carry vector bosons and change their physical quantities. the charged fermions under electromagnetic interaction are to exchange a massless and neutral particle called as *photon* (γ), the quarks interacting through the strong force exchange a massless and colourful particle called as *gluon* (g) and the weak interaction of the fermions is to exchange the massive charged bosons $W^\pm$ or massive neutral boson Z .

Since each interaction respects gauge symmetry, these vector bosons are also called as *gauge boson*. Under the gauge symmetry, massive gauge bosons are forbidden. This is true for the photons and gluons which are massless. However, the $W^\pm$ and Z boson as massive gauge bosons explicitly violate the gauge symmetry under the weak interaction. It seems that there should be a more fundamental theory to explain why the gauge bosons can acquire mass. It's *spontaneous symmetry breaking*, introduced in Section 1.2.

1.1.2 Interaction

In the Standard Model, there are three kinds of fundamental interactions (forces), arranged by force strength from strong to weak such as the strong, electromagnetic (EM), weak interactions, that interact between matter particles, i.e. fermions. As Section 1.1.1 has mentioned, from microscopic point of view, the matter particles under the interactions are actually to exchange force-carry mediums called as spin-1 gauge bosons as photons in the EM interaction, $W^\pm$ and Z bosons in the weak interaction and gluons in the strong interaction. Since the strong interaction strengthens when going to lower energy scale, equivalent to long distance, particles which can join this interaction can't escape the binding of the strong force, known as *colour confinement*. The strong force thus interacts only in a very short distance less than 10^{-15} m about a proton radius. Moreover, the massive and unstable $W^\pm$ and Z bosons make them decay immediately and the interaction range is even shorter than the strong interaction. Only the electromagnetic interaction with massless photons and medium strength can be the long-distance force. Table 1.2 shows the summary of these three interaction forces.

Interaction	Strength (relative)	Gauge boson	Range (m)
Strong	1	g	10^{-15}
Electromagnetism	10^{-3}	γ	∞
Weak	10^{-8}	$W^\pm, Z$	10^{-18}

Table 1.2: The information of three fundamental interactions in the Strong Model.

In fact, each interaction also respects its gauge symmetry corresponding to its unitary group, that is, the electromagnetic interaction follows the $U(1)_{EM}$ symmetry, the weak interaction is the $SU(2)_L$ symmetry and the strong interaction is the $SU(3)_c$ symmetry. Moreover, the electromagnetic and weak forces are unified to a so-called electroweak interaction with equal interaction strength at energy scale of rough hundreds GeV, satisfying the $SU(2)_L \times U(1)_Y$ symmetry.

Therefore, overall interactions under the Standard Model follow the gauge symmetry as

$$SU(3)_c \times SU(2)_L \times U(1)_Y.$$

The following section simply introduces these interactions under the gauge theory through the language of quantum field theory.

Strong interaction

Strong interaction in quantum field theory is encoded by a theory of so-called *quantum chromodynamics* (QCD) which describes the interaction between quarks and gluons in the Standard Model and respects the $SU(3)$ symmetry. The $SU(3)$ group is a *non-Abelian* group with eight free parameters corresponding to eight generators, where the minimal irreducible representation is 3×3 matrices, known as *Gell-Mann matrices*. Hence, quarks, ψ , under the $SU(3)$ transformation are given by the following

$$\psi \rightarrow \psi' = U\psi = e^{ig_s\alpha^a T^a} \psi \approx (\mathbf{1} + ig_s\alpha^a T^a)\psi \quad (1.1)$$

where g_s is a real value interpreted as the coupling strength of the strong interaction and α_a is eight arbitrary phases carrying eight generators T^a with $a = 1, 2, \dots, 8$. Since T^a is a 3×3 matrix operating to the quark fields, the quark fields are extended by a new quantum space with three degree of freedoms as a triplet, which are called as *colours* such as red, green and blue. Therefore, here the $SU(3)$ is also expressed as $SU(3)_c$ where c emphasizes colour spaces.

In general, free fermions carrying mass can be described by a *Dirac* Lagrangian given by

$$\mathcal{L}_{\text{free}} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi. \quad (1.2)$$

which is a Lorentz invariance form. For the global $SU(3)$ transformation where α^a is independent of spacetime, the Lagrangian (1.2) is always invariant. This gives that the colour charge is conserved according to Noether's theorem. However, if $\alpha^a = \alpha^a(x)$, namely doing a local (or say "gauge") $SU(3)$ transformation, the Lagrangian will create additional terms which are unphysical and violate the $SU(3)$ gauge symmetry. In principle, the gauge transformation is to regulate redundant degree of freedoms in the Lagrangian by some specific mathematical formalism. Physical observables affected by the gauge transformation, that is, the choice of gauge, are physically forbidden. To establish the gauge-invariance Lagrangian based on Eq. (1.2) under the gauge transformation, a set of spin-1 gauge fields, called as *gluon*, encoded by eight $SU(3)$ generators is introduced as

$$G_\mu(x) = G_\mu^a(x)T^a \quad (1.3)$$

Moreover, we define the covariant derivative of the QCD to replace the partial derivative in Eq. (1.2) through

$$\partial_\mu \psi \rightarrow D_\mu \psi = (\partial_\mu + ig_s G_\mu^a T^a)\psi \quad (1.4)$$

The additional term $ig_s G_\mu^a T^a \psi$ implies the strong interaction only occur if the fermions, ψ , own the $SU(3)_c$ triplet, that is, carry the colour charge. At the same time, the gauge fields also have to follow another gauge transformation as the following form under the $SU(3)$ group

$$G_\mu \rightarrow G'_\mu = U G_\mu U^\dagger + \frac{i}{g_s} (\partial_\mu U) U^\dagger \approx G_\mu^a - \partial_\mu \alpha^a - g_s f_{abc} \alpha^b G_\mu^c \quad (1.5)$$

where the right formula is the result under the infinitesimal transformation; the f_{abc} are the structure constants of the $SU(3)$ group, defined by the commutation relation $[\lambda_a, \lambda_b] = 2if_{abc}\lambda_c$ where $\frac{\lambda_a}{2} = T_a$. Therefore, under the $SU(3)$ transformation, the covariant derivative (1.4) performs

$$D_\mu \psi \rightarrow \partial_\mu (U\psi) + ig_s (U G_\mu U^\dagger + \frac{i}{g_s} (\partial_\mu U) U^\dagger) U\psi = U D_\mu \psi \quad (1.6)$$

which lets the modified Lagrangian based on (1.2) invariant.

On the other hand, the kinetic term of the gluons can be described by a Lorentz invariant form donated by

$$-\frac{1}{4} G^{(a)\mu\nu} G_{\mu\nu}^{(a)}, \quad \text{where } G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a.$$

However, the $SU(3)$ as a non-Abelian group causes that the $G_{\mu\nu}^a$ has to add an additional term of $-g_s f_{abc} G_\mu^b G_\nu^c$ to make a gauge-invariance kinetic term of

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_s f_{abc} G_\mu^b G_\nu^c \quad (1.7)$$

This additional term implies that the gluons can have three- and four-point self-interactions. These two interactions are permitted because physically the gluons also carry colour charges and mathematically they own octuplets under the $SU(3)$ adjoint representation defined by the structure constants. Therefore, a complete Lagrangian of the QCD is given by

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4} G^{(a)\mu\nu} G_{\mu\nu}^{(a)} \quad (1.8)$$

to describe the interactions as Figure 1.2.

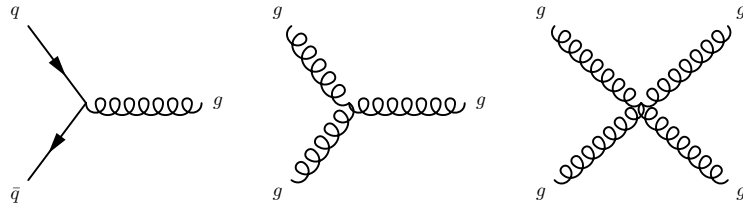


Figure 1.2: Interactions of the quantum chromodynamics. The left diagram is the quark and gluon coupling, and the right two diagrams are the self-interactions of the gluon.

Electroweak interaction

Electroweak interaction is the unification of the electromagnetic and weak interaction at hundreds GeV scale. However, in lower energy scale below Z boson mass, the dominant interaction is the electromagnetism as the strength of the EM interaction below this scale increases while that of the weak interaction decreases.

The electromagnetic interaction is described by the quantum electrodynamics (QED), which respects the $U(1)$ gauge symmetry. The $U(1)$ group is an Abelian and trivial group with only one free parameter carrying one trivial generator. Fermions under the $U(1)$ transformation can be written by

$$\psi \rightarrow \psi' = e^{iQe\alpha} \psi \sim (1 + iQe\alpha) \psi \quad (1.9)$$

where Q is the charge of the fermions and e is a real value interpreted as the coupling strength of the electromagnetic force. Similarly to the QCD framework, although the global phase transformation of the $U(1)_{EM}$ group is conserved, giving electric charge conservation, the gauge symmetry is not satisfied in the free fermions in Lagrangian (1.2). Only one trivial generator corresponds only one spin-1 field A_μ which is *photon* to construct a gauge-invariance Lagrangian form based on Lagrangian (1.2) with a covariant derivative

$$\partial_\mu \psi \rightarrow D_\mu \psi = (\partial_\mu + iQeA_\mu) \psi \quad (1.10)$$

with the photon under the $U(1)$ gauge transformation form as

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{i}{Qe}(\partial_\mu U)U^\dagger \approx A_\mu - \partial_\mu \alpha \quad (1.11)$$

The additional terms $iQeA_\mu \psi$ also implies the EM interaction only occur if the fermions, ψ , carry the electric charge. Adding the kinetic term of the photon, an overall Lagrangian of the QED satisfying the $U(1)$ gauge symmetry is given by

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (1.12)$$

where $D_\mu = (\partial_\mu + iQeA_\mu)$ is the covariant derivative and $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the photon's field stress tensor. Due to no structure constant in the $U(1)$ group, the photon which is a neutral particle only interact with charged fermions such as Figure 1.3.

On the other hand, the weak interaction follows a non-Abelian $SU(2)$ gauge symmetry. Before discussing about that gauge theory, an important observation were revealed from experiments : *Parity is violated in the weak interaction*. This particular feature provides a hint that the interaction of fermions and weak gauge bosons should be a vector-axial structure of

$$\underbrace{\bar{\psi}\gamma^\mu\psi}_{\text{Vector}} - \underbrace{\bar{\psi}\gamma^\mu\gamma^5\psi}_{\text{Axial}} = \bar{\psi}\gamma^\mu(1 - \gamma^5)\psi \quad (1.13)$$

so that this amplitude violates the parity symmetry which never occurs in the electromagnetic

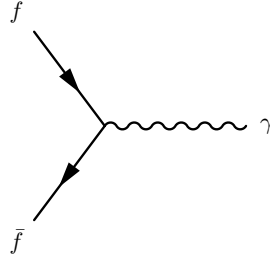


Figure 1.3: Interactions of the quantum electromagnetism.

and strong interactions. As we mentioned in Section 1.1.1, the fermions can be split into left-handed and right-handed particles. The $(1 - \gamma^5)$ in Eq. (1.13), in fact, is a projection operator to project the fermions to their left-handed particles, described as

$$P_L \psi = \frac{1}{2}(1 - \gamma^5)\psi = \psi_L$$

This results in an important physical feature : only left-handed particles and right-handed antiparticles join the weak interaction. Therefore, a $SU(2)_L$ gauge-invariance Lagrangian to describe the weak interaction is written as

$$\mathcal{L}_{weak} = i\bar{\psi}_L \gamma^\mu D_\mu \psi_L - \frac{1}{4} W^{(a)\mu\nu} W_{\mu\nu}^{(a)} \quad (1.14)$$

where the subscript L emphasizes only the left-handed fermions join the weak interaction and the covariant derivative D_μ is given by

$$D_\mu = \partial_\mu + ig_W W_\mu^a \sigma^a \quad (1.15)$$

where three generators σ^a are the *Pauli matrices* corresponding to three spin-1 fields W_μ^a and g_W is the strength of the weak interaction; the right term is a kinetic term of the W fields with $W_{\mu\nu}^a$ defined by

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g_W f_{abc} W_\mu^b W_\nu^c \quad (1.16)$$

where the structure constant f_{abc} implies the $SU(2)_L$ gauge boson can have self-interactions. Since the Pauli matrices operate on these left-handed fermions, these fermions are extended by an additional quantum space, or say weak eigenstate, called as *flavour* which are the $SU(2)_L$ doublet forms. In the Standard Model, these $SU(2)_L$ doublets are the left-handed quarks (Q) and leptons (L) which can be expressed by

$$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \quad L = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_{\tau L} \\ \tau_L \end{pmatrix} \quad (1.17)$$

where each doublet corresponds to each generation of the fermions, mentioned in Section 1.1.1. Namely, the weak interaction only allows flavour changing within the same doublet. Moreover,

a quantum number called as *weak isospin*, $I_W^{(3)}$, is donated by $\sigma^{(3)}$ to each doublet with $I_W^{(3)} = \frac{1}{2}$ for the upper component and $I_W^{(3)} = -\frac{1}{2}$ for the lower component in each doublet. As for all right-handed fermions, they are considered as $SU(2)$ singlets with $I_W = 0$ as they don't join the weak interaction.

For the weak gauge bosons, the physical gauge fields are actually the $W^\pm$ and Z bosons. To change the flavour from the upper component to lower component or vice versa in given doublets, raising and lowering operators $\sigma_\pm = \frac{1}{2}(\sigma_1 \pm i\sigma_2)$ are regarded as physical operators. According to this reason, the $W^\pm$ boson related to the $W^{(1)}$ and $W^{(2)}$ can be expressed as

$$W_\mu^\pm = W_\mu^{(1)} \mp iW_\mu^{(2)}. \quad (1.18)$$

Naively, the rest gauge field, $W_\mu^{(3)}$, can represent the Z boson which can't change the flavour. Nevertheless, the Z boson can experimentally couple both left-handed and right-handed particles and this is not consistent with the properties of the weak interaction. It seems that the Z boson results from a linear combination of the $W_\mu^{(3)}$ and some field. Since the photon A_μ is also a neutral particle and interact with left- and right-handed particles, in 1961, Sheldon Glashow proposed a theory of electroweak unification combining the electromagnetic and weak interactions. This theory donates a new $U(1)_Y$ gauge symmetry with a new gauge field B_μ to replace the A_μ field. The $U(1)_Y$ provides a new quantum number, called as *weak hypercharge*, so that the fermions under the $U(1)$ transformation are defined as

$$\psi \rightarrow \psi' = e^{ig' \frac{Y}{2} \alpha} \psi \sim (1 + ig' \frac{Y}{2} \alpha) \psi \quad (1.19)$$

hence, a complete electroweak Lagrangian can be described by

$$\mathcal{L}_{EW} = i\bar{\Psi}_L \gamma^\mu D_\mu \Psi_L + i\bar{\psi}_R \gamma^\mu D'_\mu \psi_R - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^{(a)} W^{(a)\mu\nu} \quad (1.20)$$

where the covariant derivative for the left-handed fermions Ψ is

$$D_\mu = \partial_\mu + ig' \frac{Y}{2} B_\mu + ig_w W_\mu^a \sigma^a \quad (1.21)$$

and that for the right-handed fermions ψ is

$$D'_\mu = \partial_\mu + ig' \frac{Y}{2} B_\mu. \quad (1.22)$$

In the unified electroweak model, the photon and Z boson now can be expressed by a set of linear combinations of the $W_\mu^{(3)}$ and B_μ field as

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

$$Z_\mu = -B_\mu \sin \theta_W + W_\mu^{(3)} \cos \theta_W \quad (1.24)$$

where θ_W is the weak mixing angle. This relation connects between the electromagnetic and

weak interactions and derives two important relations by

$$Y = 2(Q - I_W^{(3)}) \quad (1.25)$$

and

$$e = g_W \sin \theta_W = g' \cos \theta_W. \quad (1.26)$$

Table 1.3 shows the relation between electroweak quantum numbers such as the electric charge, weak hypercharge and isospin for the left-handed and right-handed fermions in the Standard Model. It's worthy to note that the electroweak quantum numbers of the right-handed neutrinos are zero, that is, the right-handed neutrinos do not join any electroweak interaction. That is why the right-handed neutrinos are never observed.

Fermion	Left-handed			Right-handed	
	Q	$I_W^{(3)}$	Y	$I_W^{(3)}$	Y
u, c, t	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$	0	$\frac{4}{3}$
d, s, b	$-\frac{1}{3}$	$-\frac{1}{2}$	$\frac{1}{3}$	0	$-\frac{2}{3}$
ν_e, ν_μ, ν_τ	0	$\frac{1}{2}$	-1	0	0
e^-, μ^-, τ^-	-1	$-\frac{1}{2}$	-1	0	-2

Table 1.3: Relation between electroweak quantum numbers such as the electric charge, weak hypercharge and isospin for the left-handed and right-handed fermions in the Standard Model.

The electroweak successfully unifies the electromagnetic and weak interactions with an elegant mathematical form which is also experimentally supported. However, it's observed that the $W^\pm$ and Z boson are massive particles which are physically forbidden by the gauge theory. Therefore, there should be a more fundamental reason to solve this problem, introduced in the next section.

1.2 Higgs mechanism

A crucial point for gauge symmetry in particle physics is that gauge bosons, in principle, must be massless particles. The reason is that when performing arbitrary gauge transformations, a Lagrangian with mass terms describing gauge bosons can create additional but unphysical terms which violate gauge symmetries. This is true for photons and gluons as massless particles satisfying their gauge symmetries. However, in experimental observations, the $W^\pm$ and Z bosons which are the gauge bosons of the weak interaction are massive particles. This violates the gauge symmetries that should not happen. In 1964, Peter Higgs et al. propose a concept of spontaneous symmetry breaking to the gauge theory in particle physics to explain how gauge bosons acquire mass and even predict a new scalar boson. In 1967, after Steven Weinberg and Abdus Salam introduce the Higgs mechanism to the electroweak unified theory, the problem from massive gauge bosons is theoretically solved. More importantly, the Higgs mechanism predicts a new single scalar boson is called as *Higgs boson*.

On the other hand, a problem also occurs in the elastic scattering of the longitudinal $W^\pm$ bosons, $W_L W_L \rightarrow W_L W_L$. It's calculated that the amplitude of this scattering creates quadratic divergence which violates the perturbative unitarity bound when the energy scale for this process grows up. Namely, the weak interaction will become very strong and enter the non-perturbative regime. However, this quadratic divergence can be canceled if the Higgs boson also join this process which makes an important boundary of the Higgs boson mass as

$$m_H^2 \leq \frac{4\pi\sqrt{2}}{G_F} \approx (1.2\text{TeV})^2 \quad (1.27)$$

Hence, to solve this problem, the Higgs boson must exist and its mass has to be below around 1.2 TeV, proposed in Ref [9]. The Higgs boson is not discovered and this theory isn't completely supported until 2012. In the following sections, we will take an $U(1)$ gauge symmetry to introduce the concept of the spontaneous symmetry breaking in Section 1.2.1 and the spontaneous symmetry breaking really happens in the electroweak gauge symmetry, described in Section 1.2.2. Section 1.2.3 gives a brief of the Higgs boson hunting.

1.2.1 Spontaneous symmetry breaking and Higgs mechanism

The concept of spontaneous symmetry breaking can be simply explained by a scalar field under the $U(1)$ symmetry. Consider a generic Lagrangian for a complex scalar field which satisfies the global $U(1)$ symmetry is mounted by an additional potential $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$ as

$$\mathcal{L} = (\partial_\mu\phi)^*(\partial^\mu\phi) - V(\phi). \quad (1.28)$$

When $\mu^2 > 0$, the physical vacuum state in the potential V is always at $V = 0$. This is conserved under the global $U(1)$ transformation as a rotation under a point is always invariant. When somehow the μ^2 becomes negative, the shape of the potential deforms to a Mexican hat where the vacuum states are no longer at the original point but located in the red dashed line circle, described in Figure 1.4, with non-zero vacuum expectation value of v , defined by $\phi^1 + \phi^2 = -\frac{\mu^2}{\lambda} = v^2$.

At this moment, the field standing at the original vacuum state which is now unstable must falls down to one of the new vacuum states (red dashed circle) which is stable in this potential. Once the field starts doing "fall-down" action, the global $U(1)$ symmetry is spontaneously broken by the non-zero vacuum expectation value as the $U(1)$ transformation makes the field rotate along the dashed line to another fields. This does not mean that the $U(1)$ symmetry is explicitly violated in the scalar Lagrangian, that is, Eq. (1.28) under the spontaneous symmetry breaking since this Lagrangian change nothing. Instead, it is broken near the vacuum states with the non-zero vacuum expectation value. This concept can be promoted to the $U(1)$ gauge symmetry, known as *Higgs mechanism*.

Now, consider a scalar field Lagrangian with the $U(1)$ local gauge symmetry as

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^*(D^\mu\phi) - \mu^2\phi^2 - \lambda\phi^4 \quad (1.29)$$

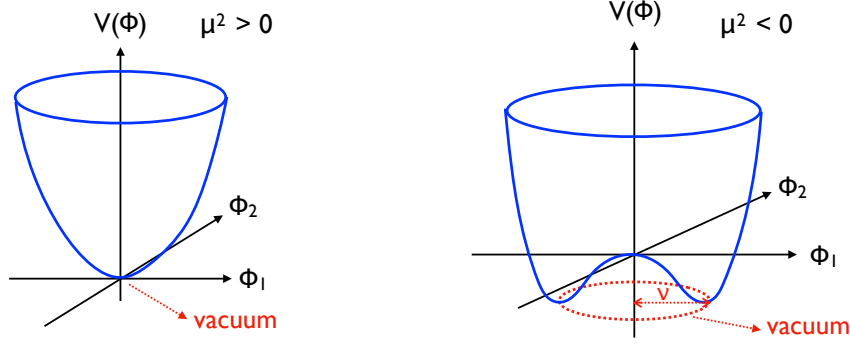


Figure 1.4: Potential, $V(\phi)$, for the complex scalar field for the $\mu^2 > 0$ case (left) and $\mu^2 < 0$ case (right). The red dashed circle in the right potential is the infinite degenerate vacuum states with non-zero vacuum expectation value of v .

where D_μ is the covariant derivative given by $D_\mu = \partial_\mu + igB_\mu$ and $F_{\mu\nu}$ is the field stress tensor given by $F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. To explore the physics near the vacuum states after spontaneous symmetry breaking, without loss of generality, the physical vacuum state is chosen at $\phi_1 + i\phi_2 = v$ and we consider the excitation of the complex scalar field near this vacuum state like the disturbance of this scalar field, given by

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\omega(x)) \quad (1.30)$$

By substituting Eq. (1.30) into Lagrangian (1.29), we can get

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \lambda v^2 \eta^2}_{\text{massive } \eta} + \underbrace{\frac{1}{2}(\partial_\mu \omega)(\partial^\mu \omega)}_{\text{massless } \omega} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2 v^2 B_\mu B^\mu + gvB_\mu \partial^\mu \omega}_{\text{massive gauge field}} + V_{int} \quad (1.31)$$

where V_{int} contains three- and four-point interactions of the η , ω and B field. This Lagrangian contains a massive scalar field η , a massless scalar field ω , called as a *Goldstone boson* and the massive gauge field B^μ . However, the term $gvB_\mu \partial^\mu \omega$ in this Lagrangian is an unphysical interaction between a spin-1 boson and a spin-0 boson. It's observed that this term can be eliminated by writing down

$$\frac{1}{2}(\partial_\mu \omega)(\partial^\mu \omega) + gvB_\mu \partial^\mu \omega + \frac{1}{2}g^2 v^2 B_\mu B^\mu = \frac{1}{2}g^2 v^2 \left[B_\mu + \frac{1}{gv} \partial_\mu \omega \right]^2 \quad (1.32)$$

where $\frac{1}{gv} \partial_\mu \omega$ can be considered as an $U(1)$ gauge transformation on the B_μ field. Namely, the ω can be absorbed into the B_μ field to another B'_μ field which doesn't change any physical prediction². For this reason, $\phi(x)$ can be rewritten by

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\omega(x)) \rightarrow \phi'(x) = \frac{1}{\sqrt{2}}(v + h(x)) \quad (1.33)$$

²In the global symmetry breaking, this Goldstone boson ω really exists.

where $h(x)$ called as Higgs field is a physical field for this gauge transformation. Finally, Lagrangian (1.31) can be rewritten by the following Lagrangian as

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu h)(\partial^\mu h)}_{\text{massive } \eta} - \underbrace{\lambda v^2 h^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{massive gauge field}} + \underbrace{g^2 v B_\mu B^\mu h + \frac{1}{2}g^2 B_\mu B^\mu h^2}_{h, B \text{ interactions}} - \underbrace{\lambda v h^3 - \frac{1}{4}\lambda h^4}_{h \text{ self-interactions}} \quad (1.34)$$

with the corresponding interactions shown in Figure 1.5.

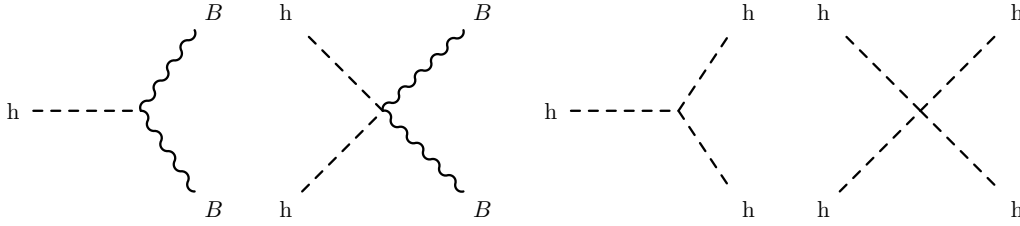


Figure 1.5: Interaction of the gauge boson B and the Higgs boson after the spontaneous symmetry breaking happening in the $U(1)$ gauge symmetry. The first two are the three- and four-point interactions of the gauge boson B and the Higgs boson; the last two are the self-interactions of the Higgs boson.

As a result, after spontaneous symmetry breaking, a Goldstone boson generated by a degree of freedom from the broken $U(1)$ symmetry is “eaten” by the spin-1 gauge boson B_μ . The additional degree of freedom becomes the longitudinal degree of freedom of the B_μ and makes the B_μ massive. However, the $U(1)$ gauge symmetry is not explicitly violated in this Lagrangian (1.34), but is hidden. More importantly, the Higgs mechanism predicts a new massive scalar particle, the Higgs boson, as the disturbance of the Higgs field.

1.2.2 Higgs mechanism in the Standard Model

From the previous section, the Higgs mechanism explains why the $U(1)$ gauge symmetry under spontaneous symmetry breaking gives the $U(1)$ gauge boson mass and predicts a new spin-0 particle called the Higgs boson at the same time. The experimental observations have showed three massive gauge bosons of the $W^\pm$ and Z from the “broken” $SU(2)_L$ gauge symmetry and a massless photon from the $U(1)_{EM}$ gauge symmetry. Namely, in the Standard Model, the Higgs mechanism should occur in the electroweak symmetry and make

$$SU(2)_L \times U(1)_Y \xrightarrow{\text{break}} U(1)_{EM}$$

where only the $U(1)_{EM}$ symmetry is kept and protects the photon as a massless particle.

To derive the electroweak symmetry breaking, we consider a Higgs field with an electroweak doublet form H which is composed of a charged h^+ and a neutral h^0 field for providing

the longitudinal degree of freedom to three gauge bosons, $W^\pm$ and Z .

$$H = \begin{pmatrix} h^+ \\ h^0 \end{pmatrix} \quad (1.35)$$

The Lagrangian of the Higgs field respecting the electroweak symmetry can be written as

$$\mathcal{L} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^{(a)}W^{(a)\mu\nu} + (D_\mu H)^\dagger(D^\mu H) - \mu^2 H^\dagger H - \lambda(H^\dagger H)^2 \quad (1.36)$$

with the Higgs potential $\mu^2 H^\dagger H + \lambda(H^\dagger H)^2$ embedded. After spontaneous symmetry breaking, $\mu^2 < 0$, the potential gives a set of infinite degenerate vacuum states satisfying

$$H^\dagger H = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda} \quad (1.37)$$

Since the photon must be massless and three degree of freedom in the H have to be absorbed by the $W^\pm$ and neutral Z bosons, without loss of generality, the vacuum state and H is chosen by

$$\langle 0|H|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \quad (1.38)$$

By substituting the Higgs field in Eq. (1.38) to the Lagrangian (1.36), we can extend this term as

$$\begin{aligned} (D_\mu H)^\dagger(D^\mu H) &= \frac{1}{8}v^2 g_W^2 (W_\mu^{(1)}W^{(1)\mu} + W_\mu^{(2)}W^{(2)\mu}) \\ &\quad + \frac{1}{8}v^2 (g_W W_\mu^{(3)} - g' B_\mu)(g_W W^{(3)\mu} - g' B^\mu) + \dots \end{aligned} \quad (1.39)$$

The first two terms in the right equation indicate the mass terms of the $W^{(1)}$ and $W^{(2)}$ field with the same mass of $m_W = \frac{1}{2}g_W v$ which is equal to the $W^\pm$ boson mass. The third term can be written as a non-diagonal mass matrix for $W^{(3)}$ and B mass, allowing the $W^{(3)}$ and B field mixing.

$$\frac{1}{8}v^2 (g_W W_\mu^{(3)} - g' B_\mu)(g_W W^{(3)\mu} - g' B^\mu) = \frac{1}{8}v^2 \begin{pmatrix} W_\mu^{(3)} & B_\mu \end{pmatrix} \begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix} \begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix} \quad (1.40)$$

However, the physical boson fields propagating as independent eigenstates of the free particle Hamiltonian must have a diagonal mass matrix. For this reason, the mass eigenstates can be

transformed to the diagonal form as the following

$$\frac{1}{8}v^2 \begin{pmatrix} A_\mu & Z_\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & g_W^2 + g'^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \frac{1}{2} \begin{pmatrix} A_\mu & Z_\mu \end{pmatrix} \begin{pmatrix} m_A^2 & 0 \\ 0 & m_Z^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} \quad (1.41)$$

where A_μ and Z_μ are the physical fields as the mass eigenstates of

$$A_\mu = \frac{g' W_\mu^{(3)} - g_W B_\mu}{\sqrt{g_W^2 + g'^2}} \quad (1.42)$$

$$Z_\mu = \frac{g_W W_\mu^{(3)} + g' B_\mu}{\sqrt{g_W^2 + g'^2}} \quad (1.43)$$

which correspond to the relation in Eq. (1.24) with $\tan \theta_w = \frac{g'}{g_W}$. The corresponding eigenvalues $m_A = 0$ and $m_Z = \frac{1}{2}v\sqrt{g_W^2 + g'^2}$ are the mass of the A_μ and Z_μ , respectively. This massless neutral boson A_μ and massive neutral boson, Z_μ are actually identical to the photon and Z boson, respectively, which are the massless gauge boson mixtures of the $SU(2)_L$ and $U(1)_Y$. Finally, the Lagrangian can be integrated as

$$\begin{aligned} \mathcal{L} = & \underbrace{\frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \frac{1}{2}m_h^2 h^2}_{\text{Higgs boson}} + \underbrace{-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{massless photon}} + \underbrace{-\frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu}_{\text{massive Z boson}} + \underbrace{-\frac{1}{4}W_{\mu\nu}^+ W^{\mu\nu-} + m_W^2 W_\mu^+ W^{\mu-}}_{\text{massive W boson}} \\ & + \underbrace{2\frac{m_W^2}{v}hW_\mu^+ W^{\mu-} + \frac{m_Z^2}{v}hZ_\mu Z^\mu}_{\text{Three-point Higgs-V interaction}} + \underbrace{-\frac{gm_h^2}{4m_W}h^3}_{\text{Three-point Higgs self-interaction}} - V_{int} \end{aligned} \quad (1.44)$$

where V_{int} contains three- and four-point interactions of the photon, $W^\pm$, Z and Higgs bosons, described in Figure 1.6, with their mass given by Table 1.4.

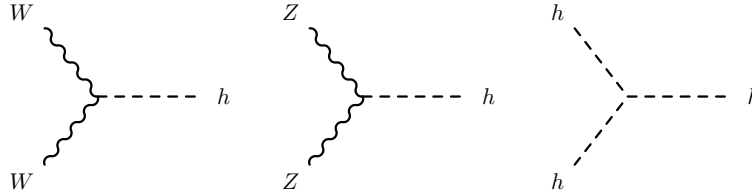


Figure 1.6: Three-point interactions of the $W^\pm$, Z and Higgs bosons. The first two diagrams are the HWW and HZZ couplings, and the last one is the self-coupling of the Higgs boson.

As the previous section mention, three Goldstone bosons as three degree of freedom from spontaneously broken $SU(2)_L$ symmetry are absorbed into three gauge bosons. These additional degree of freedoms are considered as their longitudinal degree of freedom which give them mass

Boson	Relation	Mass
h	—	$\sqrt{2}\lambda v$
$W_\mu^\pm$	$\frac{1}{\sqrt{2}}(W_\mu^{(1)} \mp iW_\mu^{(2)})$	$\frac{1}{2}g_W v$
A_μ	$\cos \theta_W B_\mu + \sin \theta_W W_\mu^{(3)}$	0
Z_μ	$-\sin \theta_W B_\mu + \cos \theta_W W_\mu^{(3)}$	$\frac{1}{2}v\sqrt{g_W^2 + g'^2}$

Table 1.4: Mass of the Higgs boson, $W^\pm$ and Z bosons, photon and their relations to the electroweak gauge bosons such as W and B .

but keep the photon as a massless particle, shown by Lagrangian (1.44). More importantly, the Lagrangian predicts the existence of a new spin-0 massive particle known as the Higgs boson under the Standard Model.

On the other hand, a Lagrangian which describes the interaction between left, right handed fermions and the Higgs field also respects the $SU(2)_L \times U(1)_Y$ symmetry like (take the third generation of quarks for example)

$$\mathcal{L}_f = -y_f \left[\begin{pmatrix} \bar{t} & \bar{b} \end{pmatrix}_L \begin{pmatrix} h^+ \\ h^0 \end{pmatrix} b_R + \bar{b}_R \begin{pmatrix} h^{+*} & h^{0*} \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}_L \right] \quad (1.45)$$

where y_f is a constant called as *Yukawa* coupling constant of the b quarks and the Higgs field. After electroweak symmetry breaking, by using the Higgs doublet given in Eq. (1.38), Lagrangian (1.45) can be derived as

$$\mathcal{L}_f = -\frac{y_f}{\sqrt{2}}v(\bar{b}_L b_R + \bar{b}_R b_L) - \frac{y_f}{\sqrt{2}}h(\bar{b}_L b_R + \bar{b}_R b_L) \quad (1.46)$$

The first term in Lagrangian (1.46) is obviously a gauge-invariance form of the b quark mass term, hence, the Yukawa coupling can connect to the observed b quark mass given by

$$y_b = \sqrt{2}\frac{m_b}{v}$$

with a proportional relation. In this case, the Lagrangian (1.46) can be rewritten by

$$\mathcal{L}_f = -m_b \bar{b}b - \frac{m_b}{v} \bar{b}bh \quad (1.47)$$

The first term can be interpreted as the origin of mass of fermions originates from the non-zero vacuum expectation value coupling to massless left-handed and right-handed fermions. Therefore, the mass scale is, in fact, driven by the strength of fermions coupling to the Higgs field. The second term is that the Higgs boson couples to fermions through the Yukawa interaction, whose coupling strength is proportional to the Yukawa coupling constant or say the mass of fermions. These two diagrams can be described in Figure 1.7. The Higgs mechanism not only successfully solves the problem about why the weak gauge bosons can acquire mass but gives an explanation about the origin of fermions' mass. All reasons point to this non-zero vacuum expectation value.

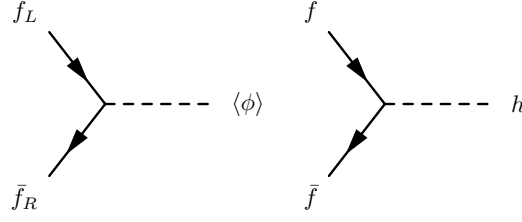


Figure 1.7: Origin of fermion mass and Yukawa interactions. The left diagram is the origin of fermion mass considered as the left- and right-handed fermions coupling to the vacuum expectation value, $\langle\phi\rangle$.

1.2.3 Higgs hunting

In the previous section, the spontaneous symmetry breaking successfully explains why three gauge bosons, $W^\pm$ and Z related to the electroweak symmetry, have mass and what the origin of mass of fermions in the Standard Model is. In fact, the electroweak unified theory including the Higgs mechanism isn't strongly highlighted, until the massive $W^\pm$ and Z gauge bosons are discovered in 1983 by the UA1 and UA2 collaborations. Afterward, the Higgs boson physics becomes one of the most important topics in particle physics. Even if the vacuum expectation value in the Higgs mechanism was extrapolated through other physical observables at that time as

$$v \simeq 246 \text{ GeV},$$

the Higgs boson mass is still a final missing parameter in the electroweak theory. Hence, prior to the Higgs discovery by the CMS and ATLAS experiments at the LHC in 2012, some experiments provide the mass constraints of the Higgs boson through direct and indirect searches.

For direct searches, the four LEP collaborations ALEPH, DELPHI, OPAL, and L3 collect data from e^+e^- collisions with a range of center-of-mass energy between 189 and 209 GeV and the total luminosity of 2461 pb^{-1} from 1996 to 2000 and search for all Higgs decay channels under the $e^+e^- \rightarrow ZH$ production. The final combination results published in 2003 give no evidence of the Higgs boson under the Standard Model but make a lower boundary of 114.4 GeV, at the 95 % confidence level, on the Higgs boson mass [10], shown by the left-handed figure 1.8. At the same time, the Tevatron collider in the Fermilab also starts focusing the Higgs boson search in proton-antiproton collisions with a center-of-mass energy up to 1 TeV after the discovery of the Top quark in 1995. In 2011, the preliminary combination results from the CDF and D0 collaborations collecting an approximate luminosity of 8.6 fb^{-1} give an exclusion region of the Higgs mass between 156 to 177 GeV and between 100 and 108 GeV covered by the LEP results, at the 95 % confidence level, without any signal evidence [11], shown in the right-handed figure 1.8.

On the other hand, exploring radiative corrections like theoretically including a new particle to vacuum polarization loops to correct the mass of the SM particles also indirectly provides the potential constraints of new physics at higher energy scales than what experiments can achieve.

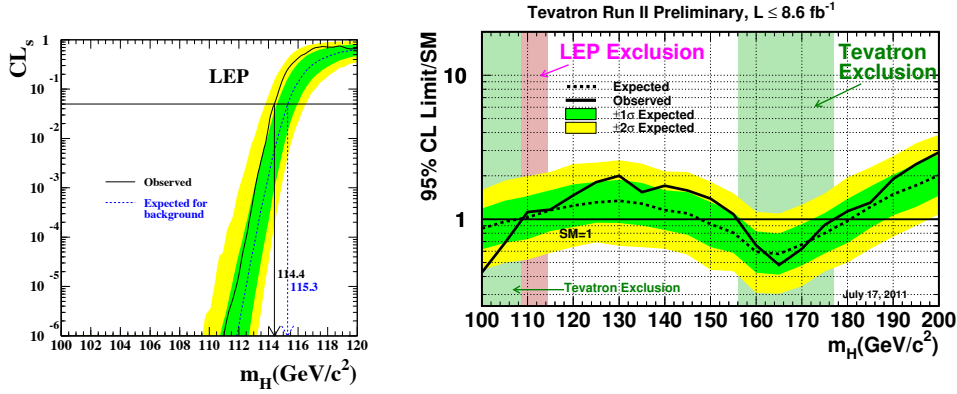


Figure 1.8: The left is the exclusion plot through the ratio $CL_s = CL_{s+B}/CL_s$ for the signal plus background hypothesis given by the LEP collaborations [10]. The blue dashed line is the median background expectation corresponding to 1σ (green band) and 2σ (yellow band) uncertainties and the black solid line is the observed result. The 95% C.L. lower limit boundary is labeled by a horizontal line of $CL_s = 0.01$ which gives the observed(expected) limit of $m_H > 114.4(115.3)$ GeV. The right is the observed (black solid curve) and expected upper limits (median background expectation, black dashed curve) on the ratios to the SM cross section at the 95 % confidence level with the Higgs boson mass range of 100 to 200 GeV from the Tevatron collaborations [11] including the exclusion region from the LEP results. The region for which the black solid curve dips below the horizontal line at the value of 1 is excluded at a 95% confidence level (CL) corresponding to the green rectangle areas.

For this reason, the indirect constraint of the Higgs boson mass can be obtained by performing a global fit of the electroweak radiative corrections. In 2011, a published result from the Gfitter group using the global fit reveals a stringent constraint of the Higgs boson mass [12]. This global fit performs a state of the art of the theoretical calculations by the radiative corrections and inputs the most precise electroweak observables such as the mass and nature width of the Z, W bosons, the top quark mass and Z-pole observables from the LEP, SLC and Tevatron precision results at that time. Only the mass of the Higgs boson is considered as a free parameter. The result, shown in Figure 1.9, gives a upper bound of 143 GeV for the mass of the Standard Model Higgs boson, at the 95 % confidence level, from the complete fit where the direct searches of the Higgs boson from the LEP and Tevatron are also included. This result shows that the Standard Model prefers a light Higgs boson with respect to 1 TeV scale given by Eq. (1.27). Therefore, until 2011, the mass constraint of the Standard Model Higgs boson fall on a range of $114.4 < m_H < 143$ GeV which gives a exciting hint to the LHC collaborations which were taking data then.

Finally, in 2012, the CMS and ATLAS collaborations which are two main experiments of the LHC at CERN independently discover the Higgs boson from direct searches with significance more than 5 standard deviations in diphoton and four charged leptons channels using part of Run I data with $\sqrt{s} = 7$ and 8 TeV [1, 2]. The combined results of the CMS and ATLAS, showed

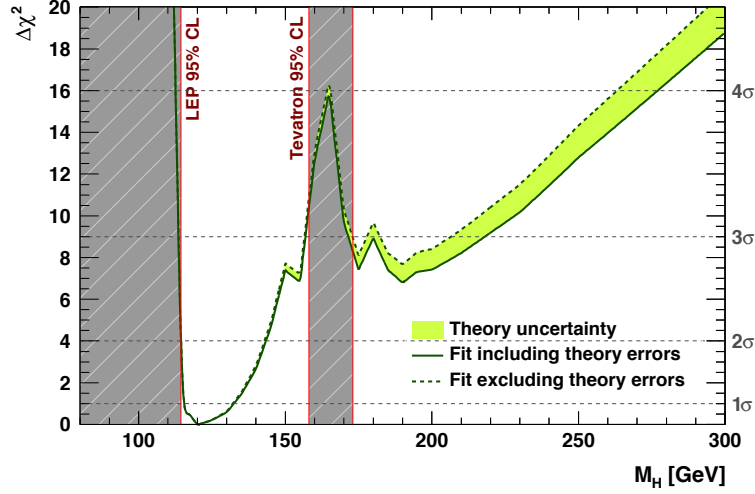


Figure 1.9: Indirect determination [12] of the Higgs boson mass by performing a global fit with the most precise electroweak parameters observed from other experiments as the input parameters and the Higgs mass as a free parameter. The direct searches from the LEP and Tevatron experiments are included to the global fit as constraints. The green solid and dashed line are the profile $\Delta\chi^2$ scan in terms of the Higgs boson mass with and without theoretical uncertainties. The gray areas are the exclusion areas given by the LEP and Tevatron experiments. The best fit value ($\Delta\chi^2 = 0$) falls on the $m_H = 120$ GeV.

in Figure 1.10, reveal the Higgs boson mass of around 125 GeV, and signal strength modifier, defined by the ratio of observed cross section w.r.t the Standard Model prediction, is about 1 that satisfies the Standard Model prediction [13]. This result also formally announced that the last key in the Standard model which is the Higgs boson was found.

1.3 Higgs boson at the LHC

Since the Higgs boson was discovered from the CMS and ATLAS collaborations in 2012 and the mass of the Higgs boson was revealed at an approximate value of 125 GeV, all parameters of the electroweak theory under the Standard Model are completely established. Afterwards, measuring the properties of the Higgs boson and testing its compatibility to the Standard Model become more important. The Higgs boson properties reflect on its coupling to other particles, even itself as a self-interaction. Therefore, to precisely measure these couplings and search for those of which have not been observed yet are the principal works at the recent LHC. The most straightforward method to understand the Higgs boson couplings is to measure the cross section of its related processes. For a given process of the Higgs boson, the cross section can be written by

$$\sigma(\vec{\kappa}, \vec{\kappa}') = \sigma_{pp \rightarrow H+X}(\vec{\kappa}) \times BR_{H \rightarrow ab}(\vec{\kappa}') \times \vec{\theta}_{X,a,b} \quad (1.48)$$

where $\sigma_{pp \rightarrow H+X}$ is the cross section of the Higgs boson production with its coupling parameters

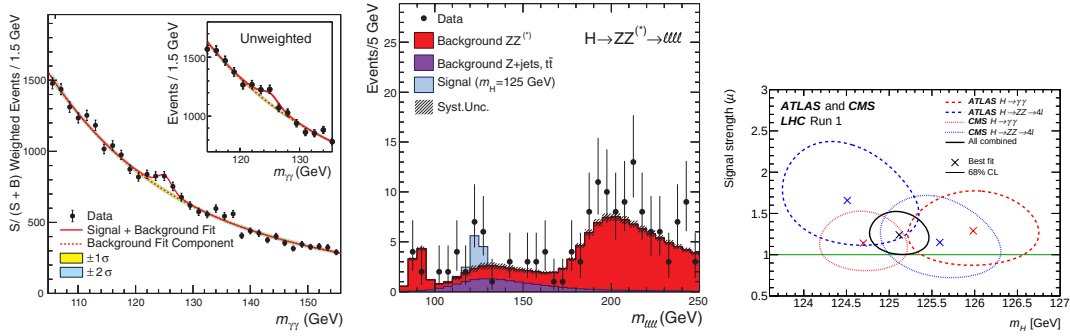


Figure 1.10: The first observation of the Higgs boson from the $H \rightarrow \gamma\gamma$ (left), $H \rightarrow 4\ell^\pm$ (middle) channels in the CMS and ATLAS collaborations during the LHC Run I period [1, 2]. The left is the best-fit signal-plus-background distribution in $m_{\gamma\gamma}$ where the uncertainties from background model choice are given by green (1σ) and yellow (2σ) bands and all analysis categories are weighted by $S/(S+B)$ for stacking and visibility. The bottom panel is the background subtraction.. The middle is the post-fit reconstructed $m_{4\ell}$ with background components given by difference color solid histograms stacked together and the signal on the top of the background with a light blue solid histogram. The right is the combination results of the signal strength modifier and the Higgs boson mass of the $H \rightarrow \gamma\gamma$ and $H \rightarrow 4\ell^\pm$ channels in the CMS and ATLAS collaborations in the LHC Run I period [13]. The green horizontal line labels the signal strength under the Standard model prediction.

$\vec{\kappa}$ to other particles; $BR_{H \rightarrow ab}$ is the branching ratio of the Higgs boson decaying to a and b with coupling parameters $\vec{\kappa}'$ of the Higgs boson; $\vec{\theta}_{X,a,b}$ is the rest parameters like branching fractions of X , a and b which are fixed to the Standard Model. Hence, the $\vec{\kappa}$ and $\vec{\kappa}'$ parameters can be extrapolated through cross section measurements and the Higgs boson properties can be revealed.

1.3.1 Production of the Higgs boson at the LHC

In proton-proton collision, the cross-section of the Higgs boson productions can be expressed as

$$\sigma_{p_1 p_2 \rightarrow H+X}(\vec{\kappa}) = \sum_{i,j} \int dx_1 dx_2 f_{i/p_1}(x_1, \mu_F^2) f_{j/p_2}(x_2, \mu_F^2) \sigma_{i,j \rightarrow H+X}(x_1 x_2 s, \mu_R^2, \mu_F^2, \vec{\kappa}) \quad (1.49)$$

where i, j are the partons from the proton 1 (p_1) and proton 2 (p_2); f_{i/p_1} and f_{j/p_2} are the parton density functions as a function of Bjorken x_1, x_2 and the factorization and renormalization scales, μ_F and μ_R . According to the parton density function, there are four relevant Higgs production mechanisms in proton-proton collisions such as the gluon-fusion, vector boson fusion, associated VH production and associated production with heavy quarks, introduced in the following list and Figure 1.11.

- **Gluon-fusion (ggH):** The ggH is the most dominant production mechanism with a cross-

section of 48.6 pb for $m_H = 125$ GeV. Therefore, the first observations of the Higgs boson coupling with other particles almost relies on this production. In fact, gluons can't directly couple to the colourless Higgs bosons under the Standard Model but have to go through a fermion triangle loop to produce it. Although the loop structure, in principle, suppresses the cross section of this production, the dominant gluons inside protons in all region of Bjorken x make the cross section raised up, especially, in lower Bjorken x . Besides, since the main contribution within the triangle loop is expected to be the top quarks with the strongest Yukawa coupling to the Higgs boson, to study the ggH production provides indirect measurements of the top-quark Yukawa coupling and even probes physics beyond the Standard Mode hidden in this loop.

- **Higgs boson associated production heavy quarks ($t\bar{t}H, tH$)** : The main fermion joining this production is the top quarks³ with the strongest Yukawa coupling in the Standard Model. According to the number of the top quarks, the production can be split into a $t\bar{t}$ case ($t\bar{t}H$) and a single top case (tH). The $t\bar{t}H$ production originates from either the $t\bar{t}$ pair productions with a single Higgs boson emission from one of these two top quarks or four top productions with two top quarks fusing into a single Higgs boson. The tH production is the single top production with a single Higgs boson emission from the top quark or the intermediate W boson. The $t\bar{t}H$ production cross section, given by 0.507 pb, is the smallest in the four relevant productions due to the top quark with the heaviest mass. The tH production is even suppressed due to the initial-state b quark, by the parton density function, giving a cross section of 0.077 pb. However, the critical point for these productions is that they give direct measurements of the top-quark Yukawa interaction.
- **Vector boson fusion (VBF)** : The VBF is the second dominant production with a cross section of 3.78 pb. It occurs when two initial-state quarks emit W^+W^- or ZZ and then these two bosons fuse to a single Higgs boson. The additional two energetic jets which tend to forward directions analytically provide a strong feature to reduce most of background processes with their jets in the central region. Physically, the process not only provides a indirect measurement of the HVV coupling under the electroweak interaction but also explores the potential BSM inside the longitudinal W^+W^- and ZZ scattering. Namely, the discovered Higgs boson could not completely solves the unitarity violation of the longitudinal vector boson scattering, mentioned in Section 1.2. However, the measurements of the VBF production have to face significant contamination originating from the ggH production with large colour-flow activities which produces two additional jets.
- **Associated VH production (VH)** : The production aka *Higgs-Strahlung* happens from a W or Z boson production but emits a Higgs boson which gives a cross section of 2.257 pb. Measuring this production not only estimates the electroweak parameters directly from the Higgs boson but constrains the potential physics beyond the electroweak symmetry breaking hidden in the HVV coupling or in the intermediate propagator. Analytically, the

³In fact, the $b\bar{b}H$ production can be also put in this class. However, analytically, the $b\bar{b}H$ production seriously mixes with the ggH production. Instead, it is merged into the ggH production with additional two b quarks from a gluon radiation.

difference of the W and Z boson leptonic decay modes allows analyses separating the VH production into the WH and ZH processes for individual measurements. Moreover, the lepton final states reduce an amount of QCD backgrounds. That is why most of analyses involving in the VH production focus on the leptonic decay channels of the W and Z bosons. More details can be shown in Section 1.4.

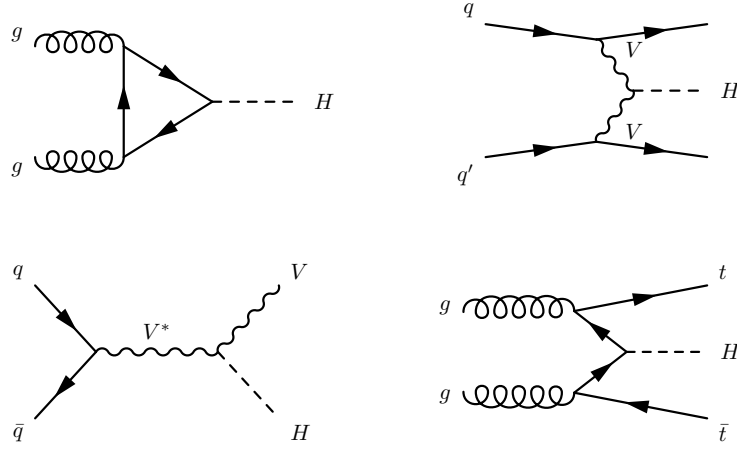


Figure 1.11: Four relevant Higgs boson productions at the LHC such as the gluon-fusion (top left), vector boson fusion (top right), associated VH production (button left) and associated production with top quarks (button right).

1.3.2 Decay of the Higgs boson

Since the Higgs boson is expected to have a nature width of 4.1 MeV for $m_H = 125$ GeV corresponding to a lifetime of 1.56×10^{-22} s, such short lifetime lets the Higgs boson decays immediately before going to detectors. Nevertheless, a variety of the Higgs boson decays provides a lot of information about the Higgs boson couplings, described by Eq. (1.48). The main decay channels of the Higgs boson and corresponding branching fractions at $m_H = 125$ GeV shown in Table 1.5.

Analytically, the sensitivity of the Higgs boson measurements depends on the production cross sections, branching ratios, the reconstructed mass resolution of the Higgs boson and the level of background for given final states. In the LHC, the proton-proton collision producing sizable QCD activities causes that measuring rare events such as the Higgs boson processes actually suffer a lot of challenges. To effectively gain the sensitivity of measurements, the strategy of the Higgs boson studies usually chooses the processes with energetic photons, charged leptons or particular features against those multi-jet backgrounds. Nevertheless, these processes usually have relatively rare events or missing pieces like neutrinos which would reduce the sensitivity. The above statement reflects on the Higgs boson decaying to electroweak related bosons like $W^\pm$, Z and γ :

Channel	Branching Ratio [%]
$b\bar{b}$	58.24
W^+W^-	21.37
gg	8.187
$\tau^+\tau^-$	6.272
$c\bar{c}$	2.891
ZZ	2.62
$\gamma\gamma$	0.227
$Z\gamma$	0.1533
$\mu^+\mu^-$	0.02176

Table 1.5: Branching fraction of the Higgs boson decay channels.

- $H \rightarrow \gamma\gamma$: Since the Higgs boson in the Standard Model is neutral and photons are massless particles, the Higgs boson coupling to the photons has to go through a top quark or W boson loop to realize the $H \rightarrow \gamma\gamma$ decay. The branching fraction is thus seriously suppressed by the loop structure. However, two energetic photons with well-measured energy provide a narrow peak lying on the smooth background in the diphoton invariant mass distribution, and their clear topology in events gives the strong separation power between signal and background. This is why the Higgs boson discovery from the $H \rightarrow \gamma\gamma$ channel is called as a "golden channel".
- $H \rightarrow ZZ^* \rightarrow 4\ell^\pm$: Due to $m_H < 2m_Z$, only one Z boson close to its on-shell mass is allowed in the $H \rightarrow ZZ$ channel ⁴. Moreover, to benefit the sensitivity by avoiding large QCD contribution, the $H \rightarrow ZZ^*$ with two Z bosons decaying leptonically, where the τ and neutrinos are omitted, is only allowed for the measurements. Hence, in this $H \rightarrow 4\ell^\pm$ channel, four charged-lepton final states kill all jet associated backgrounds and leave leptonic ZZ , $Z\gamma^*$ and $Z + X$ processes as the main background. Furthermore, well-measured energy of charged leptons forms a narrow peak in the invariant mass spectrum of the four charged leptons. This channel is thus another contribution for the Higgs boson discovery and another golden channel for the Higgs boson. However, the limited branching ratio of the $H \rightarrow 4\ell^\pm$, even lower than that of the $H \rightarrow \gamma\gamma$ channel, makes sensitive only to the ggH and VBF production.
- $H \rightarrow W^+W^- \rightarrow l^+\nu_l l'^-\nu_{l'}$: With similar situation to the $H \rightarrow ZZ^*$, the $H \rightarrow W^+W^-$ channel focuses on leptonic final states. Two energetic charged-leptons and large missing transverse momentum from undetectable neutrinos leave only the leptonic W^+W^- production and top-quark backgrounds as the dominant backgrounds. In principle, this channel has relatively large branching fraction than the $H \rightarrow \gamma\gamma$ and $H \rightarrow 4\ell^\pm$. Due to the presence of neutrinos, only the transverse mass m_T with poor resolution can reflect the Higgs mass. This causes that the sensitivity of the leptonic $H \rightarrow W^+W^-$ channel is worse than two golden channels and is finally discovered in the CMS and ATLAS exper-

⁴One of both Z bosons is usually labeled by a star at its superscript to emphasize the off-shell Z boson.

iments [14, 15] during the LHC Run I period. Therefore, the latest $H \rightarrow W^+W^-$ analysis moves on the MVA discriminator with better signal and background separation to replace the transverse mass for measurements [16].

On the other hand, for the Higgs boson decaying to fermions, only the third generation fermions were discovered at the LHC Run II period due to relatively strong Yukawa coupling which is proportional to the mass of fermions:

- $H \rightarrow b\bar{b}$: This decay channel dominates overall the Higgs boson decay with a branching ratio of almost 60%. Although the b -jet tagging technique according to particular features of b quarks allows searching for the $H \rightarrow b\bar{b}$ events, sizable QCD background is still dominant. For this reason, the $H \rightarrow b\bar{b}$ measurements aim at the associated VH production with the W or Z boson decaying leptonically. The transverse momentum of this W or Z boson is even selected with a requirement of $P_T^V > 75$ and even 150 GeV. Even so, a lot of the top quarks, Drell-Yan and W + jets events still contaminate the $H \rightarrow b\bar{b}$ measurements. Moreover, the poor $m_{b\bar{b}}$ resolution originating from poorly-measured jet energy makes the signal and background become ambiguous. The sensitivity of $H \rightarrow b\bar{b}$ measurements are limited. Until 2018, the $H \rightarrow b\bar{b}$ signal was discovered and the b -quark Yukawa coupling was first measured with a sufficient luminosity of approximate 80 fb^{-1} and more powerful b -jet tagging technique from both the CMS and ATLAS collaboration [18, 19].
- $H \rightarrow \tau^+\tau^-$: Although the τ is the heaviest lepton, it is quite unstable and immediately decay to other leptons ($BR \approx 35\%$) and hadrons ($BR \approx 65\%$). The $H \rightarrow \tau^+\tau^-$ analysis usually focuses on at least a tau decaying hadronically which covers about 94% $H \rightarrow \tau^+\tau^-$ events. Therefore, the major background of the $H \rightarrow \tau^+\tau^-$ channel is the QCD multi-jets, Drell-Yan $Z \rightarrow \tau^+\tau^-$ and top-quark pair productions. However, the undetectable neutrinos for the τ decay cause the poor resolution of $\tau\tau$ invariant mass like the $H \rightarrow b\bar{b}$ channel. That is also the reason why this channel is not independently discovered in the CMS and ATLAS collaborations until 2016, combining data in Run II (2016) and all Run I [20, 21].

As for the top quark with the largest Yukawa coupling to the Higgs boson in the Standard Model, the $H \rightarrow t\bar{t}$ decay channel is totally suppressed by two off-shell top quarks arising from $m_H < m_t$. Nevertheless, in 2018, by combining all Higgs decay channels, the CMS and ATLAS collaborations independently observe the $t\bar{t}H$ production with a standard deviation of 5.2 and 6.3 with respect to the background-only hypothesis [22, 23]. This announces the Yukawa couplings of all third generation fermions without ν_τ are measured directly, and also confirms the indirect measurement of the t -quark triangle loop in the ggH production and $H \rightarrow \gamma\gamma$ decay.

For the second generation, the $H \rightarrow c\bar{c}$ is compatible to $H \rightarrow \tau^+\tau^-$. However, since jets originating from the fragmentation and hadronization of the c quarks has no significant characteristic for background reduction like that of b quarks, an amount of QCD contribution leads to no sensitivity for observation in the $H \rightarrow c\bar{c}$ channels [25, 26]. In 2020, the CMS collaboration shows the evidence of the $H \rightarrow \mu^+\mu^-$ channel with a standard deviation of 3σ w.r.t

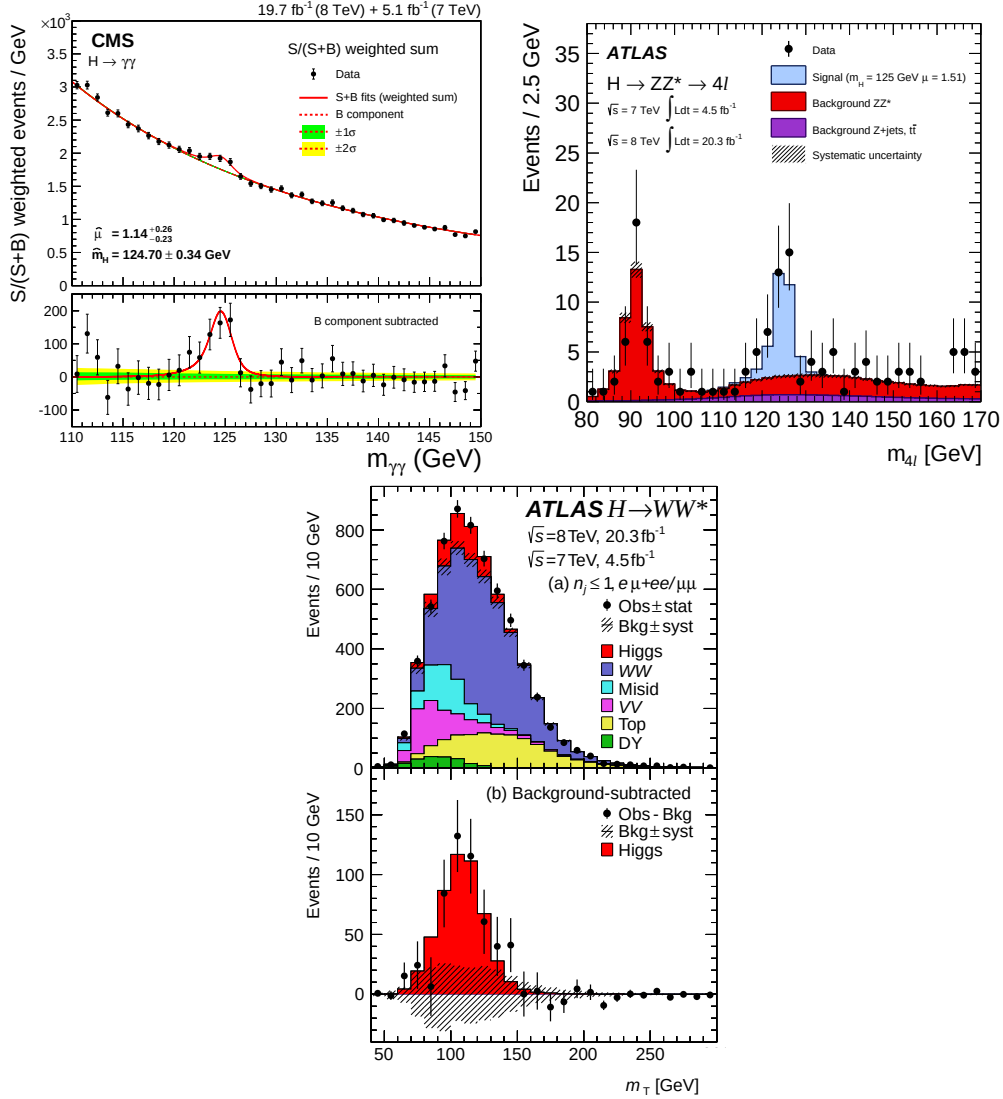


Figure 1.12: Observation of the Higgs boson from the $H \rightarrow \gamma\gamma$ (top left) [4], $H \rightarrow 4\ell^\pm$ (top right) [17] and $H \rightarrow W^+W^-$ channels (bottom) [15] in the CMS and ATLAS collaborations during the LHC Run I period. The top left is the best-fit signal-plus-background distribution in $m_{\gamma\gamma}$ where the uncertainties from background model choice are given by green (1σ) and yellow (2σ) bands and all analysis categories are weighted by $S/(S+B)$ for stacking and visibility. The bottom panel is the background subtraction. The top right is the post-fit reconstructed $m_{4\ell}$ with background components given by difference color solid histograms stacked together and the signal on the top of the background with a light blue solid histogram. The bottom is the post-fit reconstructed m_T with background components given by difference color solid histograms stacked together and the signal on the top of the background with a red solid histogram. The bottom panel is also the background subtraction.

background-only hypothesis [27]. This result announces the LHC formally touches the measurements of the second generation fermions.

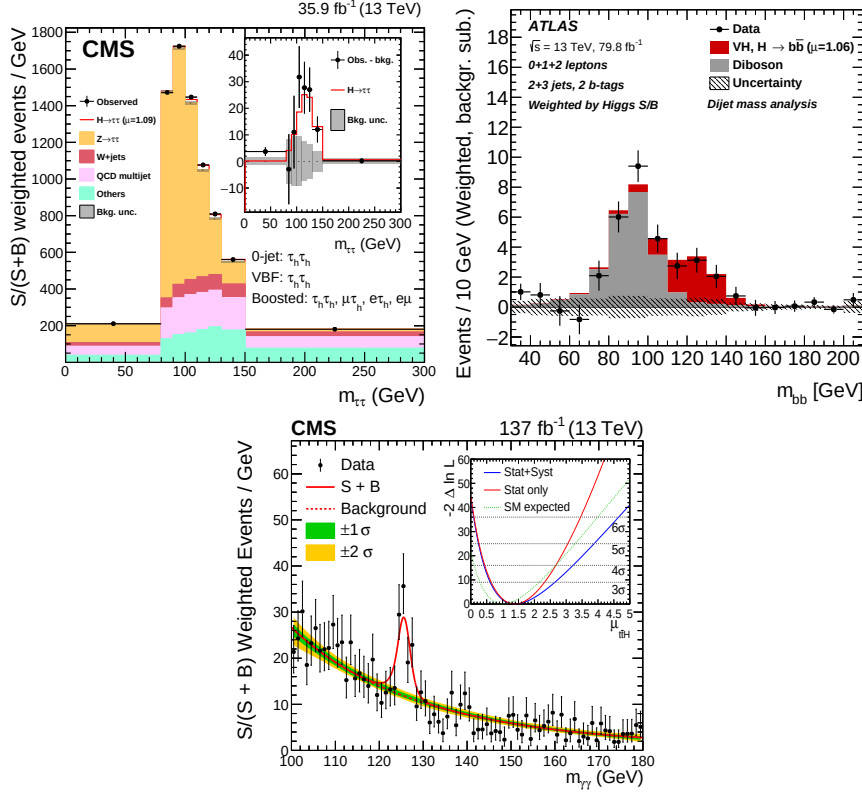


Figure 1.13: Observation of the Higgs boson from the $H \rightarrow \tau^+\tau^-$ (top left) [20], $H \rightarrow b\bar{b}$ (top right) [19] channels and the $t\bar{t}H$ with $H \rightarrow \gamma\gamma$ channel (bottom) [24] in the CMS and ATLAS collaborations during the LHC Run II period. The top left is the post-fit reconstructed $m_{\tau\tau}$ with background components given by difference color solid histograms stacked together and the signal on the top of the background with a red line histogram and total background uncertainty with gray area and background subtraction with small panel. The top right is the post-fit reconstructed $m_{b\bar{b}}$ with the leptonic VH signal and diboson background stacked by red and gray solid histograms where other background components are subtracted. The bottom is the best-fit signal-plus-background distribution in $m_{\gamma\gamma}$ where the uncertainties from background model choice are given by green (1σ) and yellow (2σ) bands. The profile likelihood scan for signal strength modifier of the $t\bar{t}H$ is also put with a minor panel with expected (green curve) and observed (blue and red) scenarios. All analysis categories are weighted by $S/(S+B)$ or S/B for visibility in these three plots.

1.3.3 What is next ?

After the evidence of the $H \rightarrow \mu^+\mu^-$ channel is shown, the state of the art combination results of the Higgs boson couplings by taking the results of the Higgs productions and decays from the CMS experiment [27], showed in Figure 1.14, reveal that the strength of all couplings aligns on the Standard Model prediction. On the other hand, the latest studies [24, 28–34] exclude the CP structure of the Higgs boson and non-zero spin more than 95% C.L. with respect to a spin-parity of $J^P = 0^+$. However, this does not mean that the Higgs boson completely respects the

SM prediction. We still need more observations, more stringent constraints in the parameter of the Higgs boson couplings. For example, the Higgs boson self-interaction which is the core of the Higgs mechanism, described in Eq. (1.44), is not observed yet [35–39].

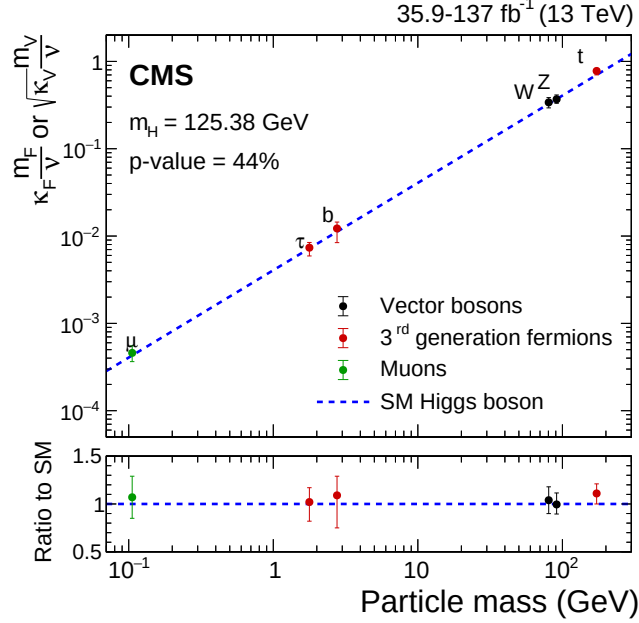


Figure 1.14: The best fit estimates for the reduced coupling modifiers extracted for fermions and weak bosons from the resolved κ -framework compared to their corresponding prediction from the SM. The error bars represent 68% CL intervals for the measured parameters. In the lower panel, the ratios of the measured coupling modifiers values to their SM predictions are shown.

Interestingly, the discovery of this “light” Higgs boson, in some sense, produces more problems. One of the most important question for the observed Higgs boson is : *why the Higgs boson is so light?* In quantum field theory, a particle propagator actually carry a important parameter : mass. However, according to perturbation theory, the particle propagator is actually composed of its pure propagator and propagators with vacuum polarizations. The former provides a so-called *bare mass*⁵ of the particle and the later gives quantum corrections to the bare mass such that the consequence is consistent with the physical mass, or so-called *pole mass*, given by experimental observations. For the Higgs boson propagator, its vacuum polarizations are mainly contributed by the $W^\pm$, Z bosons, top quark and Higgs boson itself.

Suppose the Standard Model is valid until a cutoff energy scale of Λ . The Higgs boson propagator with quantum correction gives a relation of

$$m_P^2 = m_0^2 + \delta m^2 \quad (1.50)$$

where m_P is the physical mass, m_0 is the bare mass and δm^2 is the quantum corrections given by

⁵Because the vacuum polarizations is always exist, this implies the bare mass is never known and measured.

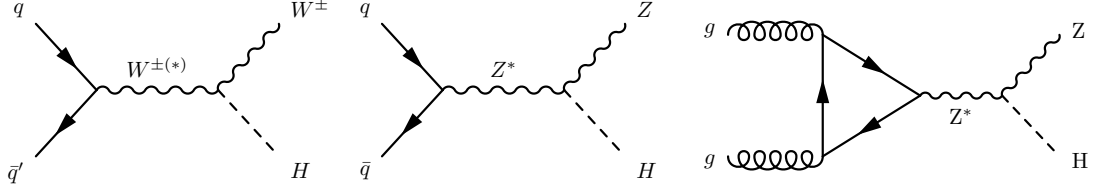


Figure 1.15: The left and middle diagrams are the VH production with a structure of the Drell-Yan process and the right diagram is a loop induced structure involving in the HZZ coupling. The star symbol on the V represents off-shell.

The state of the art calculations [40] for three VH productions are given by the following method

$$\begin{aligned}\sigma_{WH} &= \sigma_{\text{NNLO QCD}}^{\text{WH, DY}}(1 + \delta_{\text{EW}}) + \sigma_{\text{t-loop}} \\ \sigma_{ZH} &= \sigma_{\text{NNLO QCD}}^{\text{ZH, DY}}(1 + \delta_{\text{EW}}) + \sigma_{\text{t-loop}} + \sigma^{gg \rightarrow ZH}\end{aligned}\quad (1.51)$$

where $\sigma_{\text{NNLO QCD}}^{\text{VH, DY}}$ is the cross section of the Drell-Yan-like part of the VH production corrected to the NNLO QCD by VH@NNLO and δ_{EW} is a NLO EW correction given by HAWK. At the NNLO level, the WH and ZH production could create non Drell-Yan-like processes where a induced top-quark loop emits a Higgs boson, and this additional contribution, expressed by $\sigma_{\text{t-loop}}$, is thus added. On the other hand, the $\sigma^{gg \rightarrow ZH}$ given by $gg \rightarrow ZH$ with the loop induced structures is corrected to NLO QCD by VH@NNLO. The results are given by Table 1.6.

Production	Cross section [pb]	QCD scale [%]	PDF [%]	α_S [%]
$pp \rightarrow WH$	1.373	+0.5 / -0.7	± 1.7	± 0.9
$pp \rightarrow ZH$	0.7612	+0.5 / -0.6	± 1.7	± 0.9
$gg \rightarrow ZH$ (loop-induced)	0.1227	+25.1 / -18.9	± 1.8	± 1.6

Table 1.6: Fixed-order cross sections of three kinds of the associated VH production and corresponding theoretical uncertainties from the QCD scale (renormalization and factorization scales), PDF and strong coupling strength (α_S). Among these quantum corrections, the factorization (μ_F) and renormalization (μ_R) scales for the QCD and EW parts are set to $M_{\text{VH}} = (p_V + p_H)^2$ and $M_{\text{VH}} + M_H$, respectively. The set of the parton density function is PDF4LHC15_nnlo_mc for the QCD part and NNPDF2.3QED for the EW part.

The VH production is of particular importance since it allows a direct measurement of two important Higgs sectors, i.e. HWW and HZZ couplings, each of which contains crucial electroweak parameters, shown in Eq. (1.44). The measurements further test the compatibility of the VH production to the Standard Model. For example, in Figure 1.16, heavy vector bosons like the W' or Z' bosons decaying to the WH or ZH [41–43] can contribute the observed cross section of the VH production. The HVV anomalous couplings like CP structure [32] could also give strange kinematic behaviours on the VH production with respect to the SM. Moreover,

since these processes beyond the Standard Model usually occur at high energy scale which causes a boosted W/Z or Higgs boson, the cross-section measurements for the high- p_T W/Z or Higgs thus become more important.

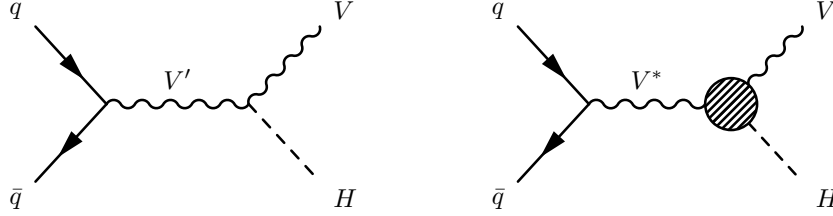


Figure 1.16: Two possible BSM processes in the associated VH production. The left is the new heavy resonances decaying to the VH final states and the right is new physics hidden in the HVV coupling where the star symbol on the V represents off-shell.

For a variety of the production mechanisms and decay channels of the Higgs boson and the relatively small cross section of the VH production, the cross section measurements are pretty difficult and complicated. Measuring a single fiducial cross section or different cross sections of the Higgs boson production might be a method. However, to avoid losing much information and keep exact fiducial volume, this kind of analysis usually only allows minimal acceptance corrections and is agnostic to the production mode. This could sacrifice sensitivity, especially those rare productions like the associated VH production. Instead, the cross section measurements of the VH productions more appropriately adopt another method called as *simplified template cross sections* (STXS) framework [44].

Simplified template cross sections framework

The main purpose of the STXS framework is to maximize the sensitivity of the cross section measurements for the Higgs boson processes and reduce the theoretical dependence at the same time. Under this framework, the Higgs production cross section is decoupled to some mutually exclusive regions called *STXS bins* according to the different Higgs boson productions and simple kinematics in particle level, establishing a set of the simplified regions with respect to exact fiducial volume. Events are also divided into different analysis categories under reconstruction level, each of which is designed to enrich these STXS bins and achieve orthogonality between the analysis categories and the STXS bins as much as possible, described in Figure 1.17. Such method can avoid strong correlation between multiple STXS bins in a single category which leads to large theoretical dependence in the STXS measurements. For each analysis category, the STXS framework allows using larger acceptance correction such as multivariate techniques to maximize the sensitivity of targeted STXS bins by excluding non-resonance backgrounds and non-target STXS bins. The cross section measurements of the Higgs boson can be optimized through this framework. Besides, the STXS further provides the understanding of the kinematic structures of the Higgs boson productions and probes the BSM effect from the STXS bins with higher energy scale.

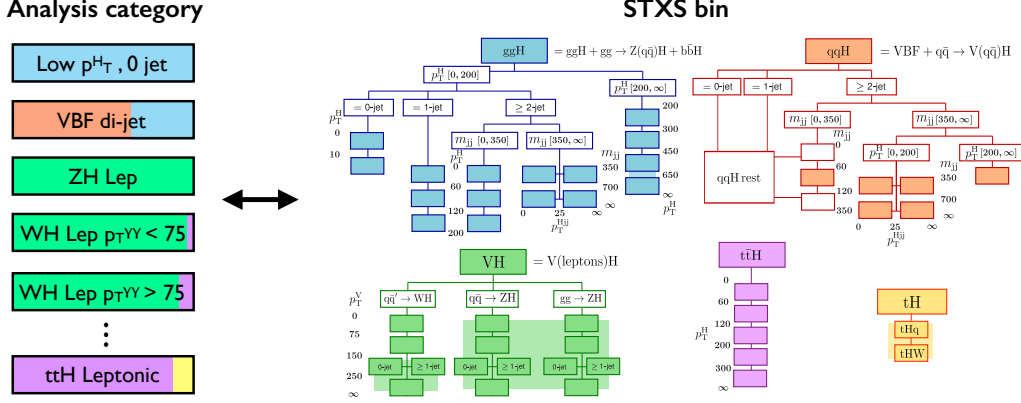


Figure 1.17: Schematic overview of the simplified template cross section framework. The left side represents analysis categories, each of which contains single or multiple STXS bins with different colors following the right part as the sketch of the STXS bins with several relevant Higgs boson productions.

The VH STXS bins with the W and Z leptonic decays are divided by three production modes, $q\bar{q}' \rightarrow WH$, $q\bar{q} \rightarrow ZH$ and $gg \rightarrow ZH$, described in Figure 1.15. Each mode is further split through the transverse momentum of the W/Z bosons and number of jets like the green STXS schematic in Figure 1.17. The VH STXS bin with the W and Z hadronic decays is only one bin located in the $60 < m_{jj} < 120$ GeV of the qqH bins, i.e. the orange STXS schematic in Figure 1.17. However, some analysis categories to enrich such fine bins could lead to leak statistics and reduce the sensitivity of the STXS measurements. Also, large correlation between STXS bins in some given analysis categories cause STXS results with huge theoretical dependence. Merging some STXS bins, analysis categories or both are thus necessary to improve this analysis. In our VH analysis, this action will be performed in Chapter 4.

From the latest STXS measurements of the VH production given by the CMS and ATLAS collaborations, only the $H \rightarrow b\bar{b}$ and $H \rightarrow W^+W^-$ channels are sensitive. In the ATLAS Experiment, the VH measurement through the $H \rightarrow b\bar{b}$ channel are split into two scenarios: resolved two b -jets for the Higgs boson with relatively low energy scale [45] and merged one from the highly boosted Higgs boson for exploring the high energy VH production which is BSM sensitive [46]. Both scenarios perform the STXS framework to isolate the VH cross sections with different kinematics to understand the kinematic structure of the VH production, shown in Figure 1.18. The results for both scenarios satisfy the Standard Model prediction. Similarly, the cross-section measurement of the VH production through the $H \rightarrow W^+W^-$ channel from the CMS collaboration [16] also uses the STXS framework. The target observable is no longer the transverse mass but the MVA classifiers to improve the sensitivity. Although the excess with respect to the Standard Model prediction is observed in some kinematic regions from its preliminary results in Figure 1.19, interpreted with the signal strength modifier and STXS, huge uncertainties could dilute the BSM effect. Therefore, the current results show that the associated

VH production respects the Standard Model prediction in these two channels.

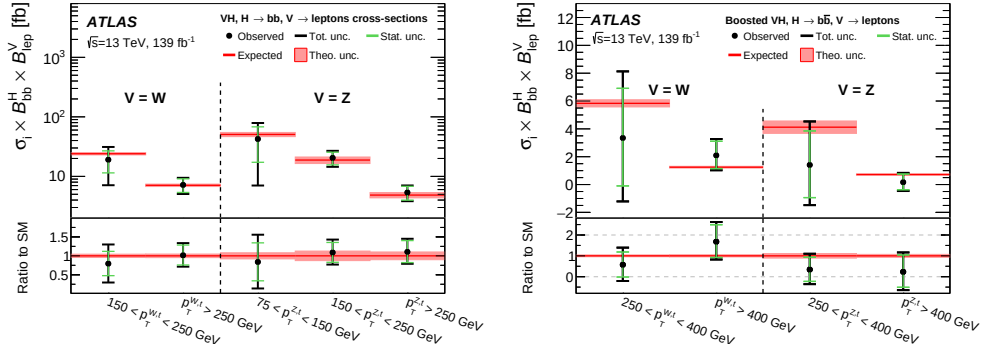


Figure 1.18: Observed STXS of the associated VH production with the W and Z boson leptonic decay in the $H \rightarrow b\bar{b}$ channel with the resolved b -jets (left) and boosted b -jet (right) scenarios. The uncertainties for both scenarios are split into the total uncertainties (black error bar) and statistical-only uncertainties (green error bar). The red lines and blocks represent the cross sections and their uncertainties with respect to the Standard Model prediction.

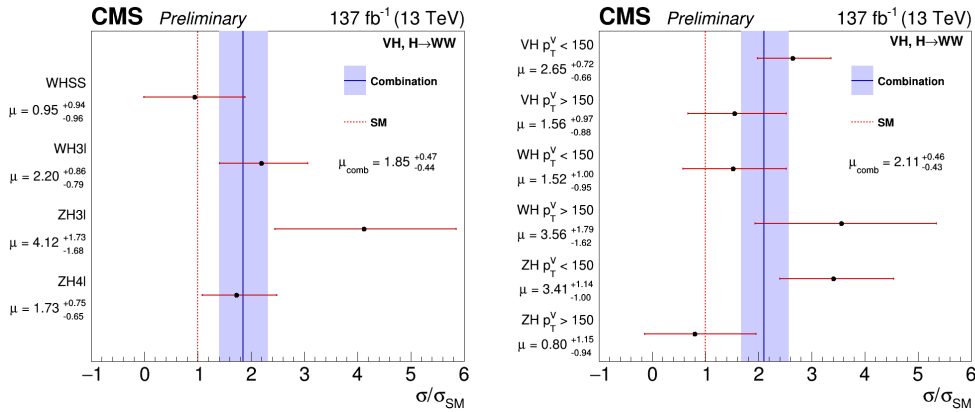


Figure 1.19: Observed signal strength modifier and STXS divided by the Standard Model prediction of the associated VH production with the W and Z boson leptonic decay in the $H \rightarrow W^+W^-$ channel with the W leptonic decay. The red error bars are the total uncertainties and the band with a violet color is the combined results.

As Section 1.3 have mentioned, the $H \rightarrow \gamma\gamma$ channel provides higher mass resolution and less background than those in the $H \rightarrow b\bar{b}$ and $H \rightarrow W^+W^-$ channels. However, the limited branching fraction for the diphoton channel makes the measured results of the VH production cross sections flooded by statistical uncertainties. However, sizable data with an integrated luminosity of 137 fb^{-1} collected by the CMS detector is sufficient to significantly reduce the uncertainty. The STXS framework further optimizes the VH analysis with better sensitivity. Hence, this analysis will reveal the state-of-the-art STXS measurements of the VH production through the $H \rightarrow \gamma\gamma$ channel.

Chapter 2

The CMS experiment

In the Standard Model, the associated VH production can occur if and only if the interaction energy achieves at least $m_H + m_V$. In the LEP, although electron-positron collisions provide quite strong physics conditions for the ZH production and very clean environment for precision measurements, the collision energy is at most 207 GeV which is lower than $m_H + m_Z$, i.e. 216 GeV. There is no VH signal observed. Even if the Tevatron collider performs proton-antiproton collisions with the collision energy up to 1.96 TeV and the instantaneous luminosity is $4 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$, the Higgs boson signal is not significant, not to mention the VH production.

The Large Hadron Collider provides an extraordinary collision energy of 7 TeV for proton-proton collisions and the collision energy is up to 13 TeV in the LHC Run II period (2015-2018). Besides, the instantaneous luminosity further achieves $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ which is 50 times than the Tevatron. Both reasons significantly enhance the events of the associated VH production. Among the detectors on the LHC, the electromagnetic calorimeter in the Compact Muon Solenoid detector provides excellent energy resolution of photons which gives a perfect diphoton mass resolution to significantly increase the sensitivity of the $H \rightarrow \gamma\gamma$ signal. From these reasons, our VH analysis under the $H \rightarrow \gamma\gamma$ channel uses data produced from the proton-proton collision in the LHC and collected from the CMS detector. This chapter will introduce the LHC in Section 2.1 and the CMS detector in Section 2.2. The trigger system in the CMS experiment is also described in Section 2.3.

2.1 Large hadron collider

Large Hadron Collider (LHC) at European Organization for Nuclear Research (CERN) is the most complex and the highest-energy particle collider machine in the world. It is located 45 to 170 metres underground France-Switzerland border and installed in the 27 kilometre circular tunnel constructed from 1984 to 1989 for the Large Electron-Positron (LEP) collider. The main goal is to provide proton-proton collisions with a center-of-mass energy of 13 TeV since 2015 and high luminosity for the exploration of rare events, especially, the Higgs boson production and physics beyond the Standard Model.

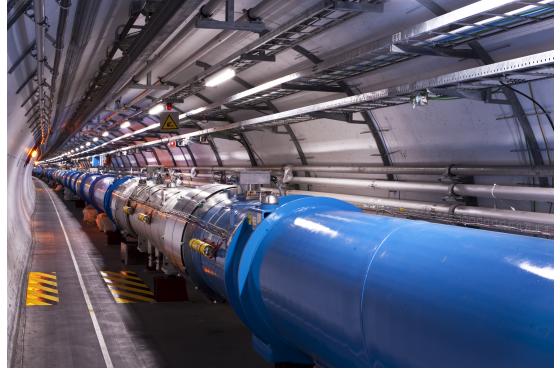


Figure 2.1: LHC tunnel and beam pipe.

2.1.1 Machine design

The LHC [47] is a two-ring-superconducting-hadron accelerator and collider where two beam pipes with magnetic flux in opposite directions are surrounded by a common cold mass and cryostat. Due to the limitation of the geometrical space in the tunnel, two rings are designed as "twin-aperture" magnets with two contra-rotating beams inside. Each beam pipe also gives a vacuum pressure of 10^{-7} Pa to avoid those accelerated protons colliding the air molecules before proton-proton collision. Among the LHC, there are two components constructing most of the LHC, superconducting magnets and RV cavities.

All superconducting magnets inside the LHC are electromagnets. Among the magnets, total 1232 main dipoles magnets, each 15 metres long, generating a powerful magnetic field of 8.3 T, drive the bending of high-energy proton beams (TeV scale) in a nearly circular trajectory. Each dipole magnet is operated by superconducting coils with a current of 11080 amperes (11.08 kA), which is composed of NbTi/Cu Rutherford type cables cooled down to 1.9 K by superfluid helium, shown in Figure 2.2. Total 392 quadrupole magnets, where generated magnetic field grows quickly with the radial distance from its longitudinal axis, arrange symmetrically around the beam pipe to focus the beam in the transverse plane either vertically or horizontally. Especially, ones with stronger magnetic field close to intersection points provide better focusing and alignment to maximize the collision luminosity.

For beam acceleration within the beam pipes, there are 8 radiofrequency (RF) cavities per beam (total 16 RF cavities) on the LHC. Each RF cavity is operated by high-power RF electromagnetic fields generated from klystrons and transferred via waveguides. These electromagnetic waves in the designed cavities form standing waves with 400 MHz. Their resonant frequency and accelerating electromagnetic fields provide a maximum voltage of 2 megavolts (MV) per cavity (total 16 MV per beam) in the longitudinal direction. In general, each injection beam with 450 GeV energy can bring 16 MeV through RF cavities once and achieves 6.5 TeV in 20 minutes. Since the accelerating electromagnetic field oscillates in the longitudinal direction, controlling the arrival time of protons to the cavities is quite important. The protons achieving the desired energy can not be accelerated, while those with slightly different energy arriving earlier or later to the cavities will be accelerated or decelerated so that they stay close to the

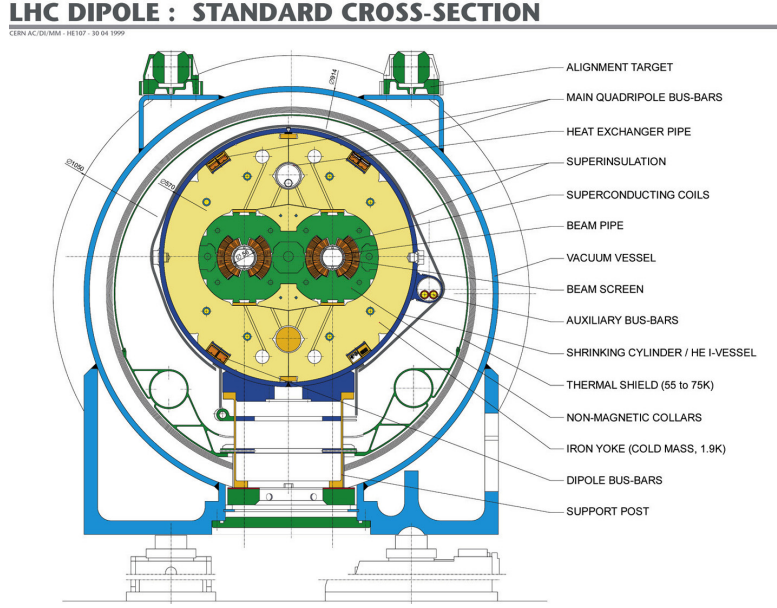


Figure 2.2: Cross section of the cryostat system, where the dipole magnets are surrounded by the superfluid helium.

desired energy. Through this way, the proton beams form proton packs called "bunches".

For the layout of the LHC, it is made of eight arcs and eight straight sections, shown in Figure 2.3. Each straight section with around 528 m long is either an experimental or utility insertion. Two high luminosity experimental insertions like the ATLAS experiment and CMS experiment are located at Point 1 and Point 5, respectively, while two further experimental insertions like the ALICE experiment and LHCb experiment are located at Point 2 and Point 8, respectively. Point 3 and Point 7 contain collimation systems to reduce quenching of superconducting magnets and damage of accelerator components as a 10^{-6} fraction of normal LHC proton beams can damage copper. Point 4 has the RV cavity systems for particle acceleration and Point 6 is a beam dumping system which extracts beams from each ring of the collider with minimal loss and terminates beams by an external dump block.

2.1.2 Luminosity and pileup

Luminosity is an important quantity to reflect the number of proton-proton collisions per unit second. Therefore, the number of events produced per second from a given X production (N_X) in the proton-proton collision can be calculated by:

$$N_X = L_0 \sigma_{pp \rightarrow X} \quad (2.1)$$

where L_0 is the instantaneous luminosity and $\sigma_{pp \rightarrow X}$ is the cross section of the X production. Assume that the proton beam density function follows a Gaussian distribution and two colliding beams have equal beam parameters, then the instantaneous luminosity can be expressed as

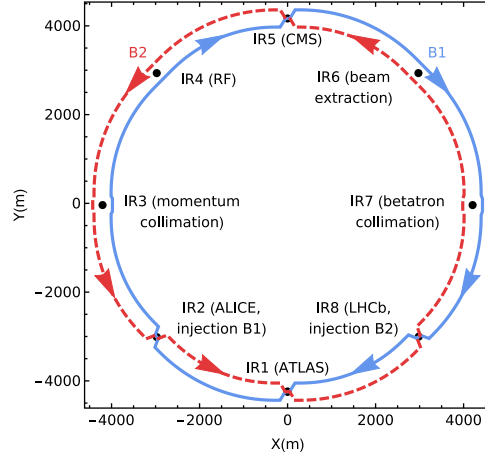


Figure 2.3: Schematic layout of the LHC (Beam 1 - clockwise, Beam 2 anticlockwise) from Ref [48].

the following formula

$$L_0 = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \hat{\epsilon} \beta^*} F \quad (2.2)$$

where N_b is the number of particles per bunch, the number of bunches per beam n_b , the revolution frequency f_{rev} , the relativistic gamma factor γ_r , the normalized transverse beam emittance $\hat{\epsilon}$, the amplitude β function at the collision point β^* and the geometric luminosity reduction factor F defined by

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right)^{-\frac{1}{2}} \quad (2.3)$$

with θ_c as the full crossing angle at the interaction point, the RMS bunch length σ_z , and the transverse RMS beam size at the interaction point σ^* .

However, pursuing high instantaneous luminosity in the LHC has to face an important phenomenon, the pileup (PU), μ , defined as the number of inelastic proton-proton collisions per bunch crossing. The average PU, $\langle \mu \rangle$, is straightforward estimated by a formula given by

$$\langle \mu \rangle = \frac{\mathcal{L}_{\text{inst.}} \sigma_{pp}^{\text{inel.}}}{n_b f_{\text{rev.}}} \quad (2.4)$$

where $\sigma_{pp}^{\text{inel.}}$ is the inelastic proton-proton cross section, giving a value of 69.2 mb at $\sqrt{s} = 13$ TeV. The higher pileup that more particles are produced at the same time makes the particle identification become quite important. Table 2.1 shows the parameters of the LHC operation for 2016, 2017 and 2018 and Figure 2.4 shows the integrated luminosity with various time during the LHC Run II period (left) and pileup distributions for 2016, 2017 and 2018, individually (right). The integrated luminosity for physics analysis from 2016 to 2018 is 137 fb^{-1} and the collisions in 2017 and 2018 have relatively higher medium pileup and instantaneous luminosity

than in 2016.

Symbol	Parameter	2016	2017	2018
$\sqrt{s}$	Centre-of-mass energy (TeV)	13	13	13
Δt	Bunch spacing (ns)	25	25	25
$\mathcal{L}_{inst}$	Instantaneous luminosity peak ($10^{34} \text{ cm}^{-2} \text{ s}^{-1}$)	1.4	2.1	2.1
$\langle \mu \rangle$	Average pileup	25	38	37
n_b	Max. No. bunches per beam	2220	2556/1868	2556
N_b	No. protons per bunch (stable beam) (10^{11})	1.0-1.25	1.0-1.25	1.0-1.25
ε	Transverse emittance (stable beam) (μm)	2.2	2.2	1.9
f_{rev}	Revolution frequency (Hz)	11245	11245	11245
β^*	Beta function (cm)	40	40-30	30-25

Table 2.1: Parameters of the LHC for 2016, 2017 and 2018 in [49].

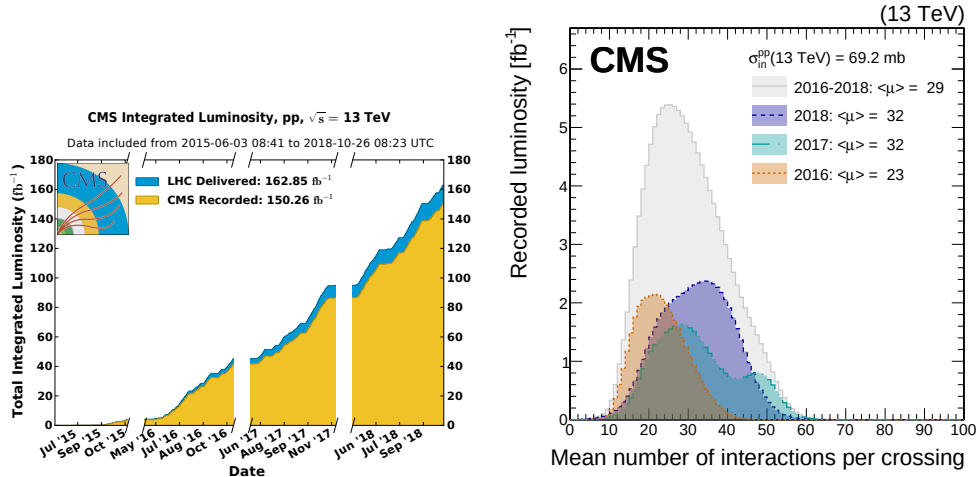


Figure 2.4: The left is the integrated luminosity with variable time during the LHC Run II period. The integrated luminosity is delivered from the LHC with the blue histogram and recorded by the CMS with the yellow histogram. The total valid luminosity for physics is 137 fb^{-1} . The right [50] is the distribution of the mean number of inelastic interactions per crossing (pileup) in data for pp collisions in 2016 (orange histogram), 2017 (light blue histogram), 2018 (navy blue histogram), and integrated over 2016-2018 (grey histogram). A minimum bias cross section of 69.2 mb is chosen. The mean number per bunch crossing and year of inelastic interactions is provided in the legend.

2.1.3 Proton injection and operational cycle

In reality, the proton is not originally accelerated from the LHC. The proton source comes from a bottle of hydrogen gas at one end of Linear accelerator 2 (Linac 2). Initially, electrons on given hydrogen molecules are removed by electric field and then protons enter the Linac 2 accelerator. When the protons reach a energy of 50 MeV, they are injected into the Proton Synchrotron

Booster which is the next step of CERN's accelerator chain. The Proton Synchrotron Booster (PSB), containing four superimposed rings with a radius of 25 metres, allows the protons from the Linac 2 to be accelerated up to 1.4 GeV. Then they enter the Proton Synchrotron (PS). The Proton Synchrotron with a circumference of 628 metres accelerates the protons to 25 GeV and provides the bunches with 25 ns to the Super Proton Synchrotron. The SPS, as the second-largest machine in CERN's accelerator complex, accelerates protons from the PS up to 450 GeV, and finally injects these proton bunches to the LHC.

Successively, the LHC starts controlling these proton bunches with several different operation steps. The operation of the LHC can be separated as different modes (phases), called "beam modes". As "setup" step is the preparation for each configuration like dipole magnets in the LHC to handle the incoming injected beams with an energy of 450 GeV from the injector complex. During "Injection", two beams (beam 1 and beam 2) are injected into two different pipes with opposite directions from SPS until the number of bunches achieves a specified value. Then, in "ramp" mode, the beam energy raises through the RF system and the magnetic field produced by the main dipoles increases for bending of the beams at the same time (The procedure basically needs to take around 20 minutes.) Once the two beams are up to 6.5 TeV, overall machine configuration are changed from injection to collision version. The procedure thus transfers to "Squeeze" mode. In this mode, the focusing triplets composed of the quadrupole magnets near the interaction point help the beams squeeze for low- β^* . The "adjust" mode at the end of "Squeeze" mode then brings and adjusts the beams into collision at interaction points. Finally, the stable beams are achieved and data taking starts.

2.2 Compact muon solenoid

The Compact Muon Solenoid (CMS) detector [51] is a general purpose detector, covering almost a solid angle of 4π steradians around interaction points except the two holes near the beam line, to detect all stable particles originating from proton-proton collisions as much as possible. It provides high-granularity subdetectors with high spatial, time resolutions and high radiation resistance against high luminosity environment. These subdetectors are sandwiched in the CMS detector, shown in Figure 2.5, and the details are described in the following sections. From the innermost to outermost parts, the inner tracker system with a pixel detector (inner) and a silicon strip tracker (outer) provides the precise and efficient measurements of charged particle trajectories from collision points in Section 2.2.2; two kinds of calorimeters measure the energy of the specific particles and block them for particle identification as the electromagnetic calorimeter (ECAL) focusing on photons and electrons in Section 2.2.3, and the hadron calorimeter (HCAL) focusing on hadron particles in Section 2.2.4; the outermost muon detector further provides the most precise momentum measurement and the best time resolution of muons in Section 2.2.6.

A huge superconducting solenoid magnet is also mounted between the HCAL and muon detectors, introduced in Section 2.2.5. This magnet provides an average magnetic field of 3.8 Tesla for bending of charged particles in the inner tracker system. The magnetic field is then

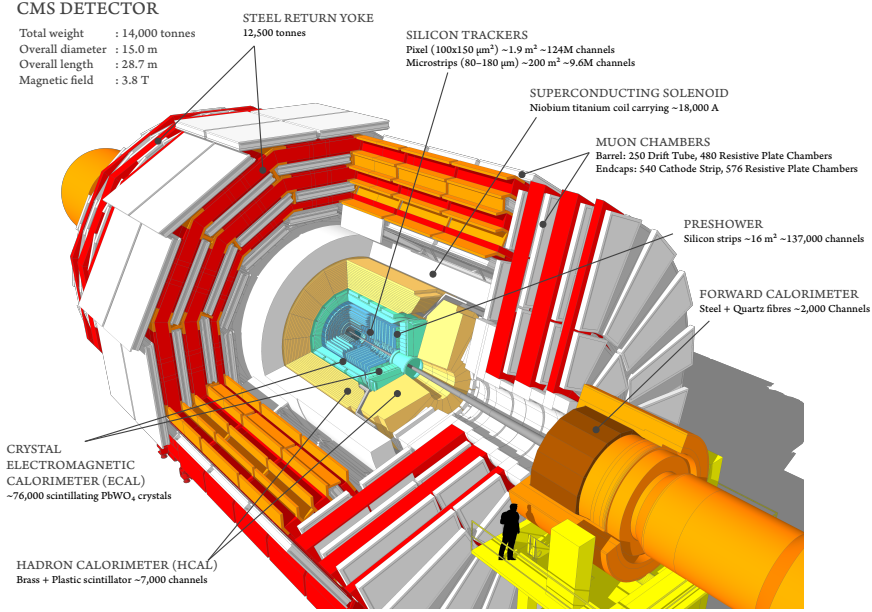


Figure 2.5: Cross section diagram of the Compact Muon Solenoid with a sandwich structure.

confined by the iron return yokes between muon detector, giving better measurement for muon momenta. Since the solenoid magnet and steel yokes contribute a weight of 12,500 tonnes in the CMS detector, the CMS detector is weighted up to 14,000 tonnes, more than the weight of Eiffel Tower.

2.2.1 Coordinate system

The CMS experiment uses a right-handed coordinate system, where the x -axis points toward the center of the LHC ring, the y -axis points upwards perpendicularly to the plane of the LHC ring and the z -axis points toward counterclockwise beam direction, described in Figure 2.6. The original point is located at the center of the CMS detector. By convention, the coordinate is separated into the transverse plane, i.e. xy -plane, and the longitudinal part as z -axis. Since the structure of the CMS detector is shaped as a cylinder, the cylinder coordinate thus provides more convenient description, where ϕ is an azimuthal angle in the transverse plane and θ is a polar angle between the z -axis and original $\vec{r}$ vector.

In proton-proton collisions, although two protons collide from opposite directions with the same momenta at the z component, two partons joining interactions inside the protons give unknown momenta at the z component due to the parton density function. Each product from the collisions could be boosted along the z -axis but the exact boost quantities are unknown. For this reason, the polar angle of the products in the CMS experiment is expressed by an observable called *rapidity*, defined by

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (2.5)$$

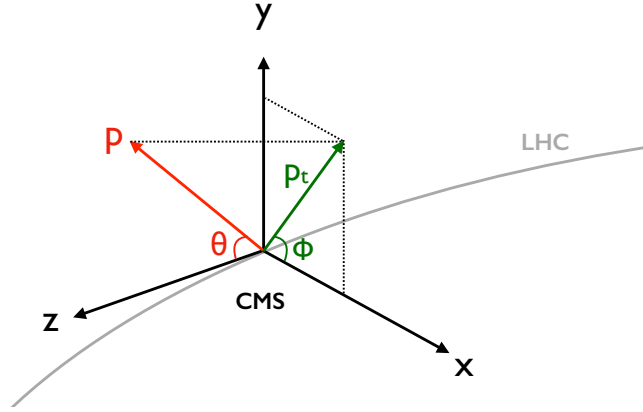


Figure 2.6: Diagram of the coordinate of the CMS detector, where the x -axis points toward the center of the LHC ring, the y -axis points upwards the sky and the z -axis points toward counterclockwise beam direction.

The advantage of this physical observable is that the rapidity difference is an invariant quantity under the longitudinal boost. Moreover, since the mass of the products with high energy can be negligible, the rapidity can approximate to another observable so-called *pseudorapidity* given by

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \quad (2.6)$$

where θ is the polar angle. Besides, the transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ and ϕ in the xy -plane are also invariant under the longitudinal boost. That is why the four-momentum of the products is usually described by (p_T, η, ϕ, E) for convenience. Other common physical quantities such as the spatial angle distance between two products, defined by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, or the transverse energy, $E_T = \sqrt{m^2 + p_T^2}$, are invariant under the longitudinal boost.

2.2.2 Inner tracker system

Inner tracker system in the CMS detector plays an important role in the reconstruction of charged particle trajectories originating from proton-proton collisions. From these reconstructed charged tracks, the interaction vertices and secondary vertices can be further reconstructed. These vertices provide quite important information for physical object reconstruction, identification and heavy flavour tagging. The transverse momentum, p_T , of each charged particle under the magnetic field can be also calculated by

$$p_T = qrB \quad (2.7)$$

where q is the electric charge of the charged particle and r is the radius of its trajectory under the magnetic field B . To gain the most precise and efficient reconstruction with minimum energy loss for charged tracks against the environment with high instantaneous luminosity

($\sim 10^{34} \text{ cm}^{-2}$), the inner tracker system with high-granularity and thin silicon sensors is placed in the innermost of the CMS detector, surrounding the collision area and covering the $|\eta| < 2.5$ region.

According to different granularity, the inner tracker system is split into the inner and outer part, i.e. a higher-granularity pixel detector and a lower-granularity silicon strip tracker, respectively. The pixel detector is composed of three barrel layers (BPix) with two endcap disks (FPix) with a arrangement shown in Figure 2.7 (bottom). This arrangement offers 3 tracking points in $|\eta| < 2.5$ region. Since the phase 1 upgrade in 2017 [52], the layout of the pixel detector was changed to four layers of the barrel part and three disks in the endcap part with a new arrangement shown in Figure 2.7 (top). This geometry provides full 4 tracking points in $|\eta| < 2.5$ region against higher pileup environment. The pixel detector adopts silicon sensors consisting of 66 million high dose n-type pixels with a size of $100 \mu\text{m}$ by $150 \mu\text{m}$ implanted into a high resistance n-substrate bulk with a thickness of $285 \mu\text{m}$. A high dose p-type backplane with reverse bias voltage is placed below the bulk. When charged particles with enough energy go through a pixel, they ionize electron-hole pairs in the bulk. These electron-hole pairs are carried out to surface through the reverse bias voltage and induce small electric signals. The electric signals are finally delivered to the read-out plane.

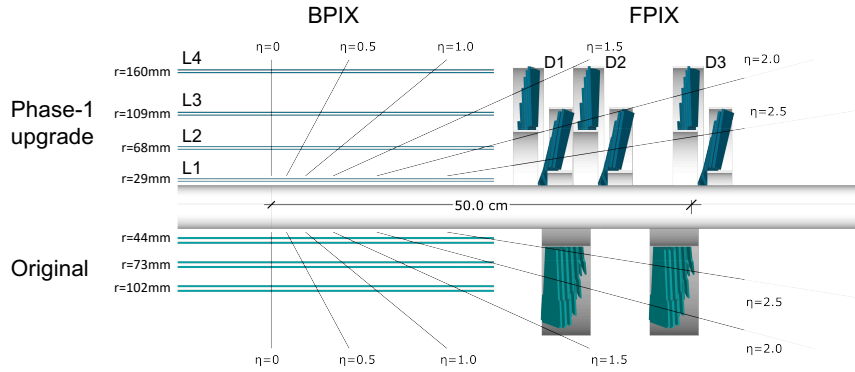


Figure 2.7: Arrangement of the Pixel detector after phase 1 upgrade (top) with respect to original one (bottom) [53].

The silicon strip tracker outside the pixel detector has a length of 5.5 m, a diameter of 2.4 m and the coverage up to $|\eta| = 2.5$. It contains two barrel parts, called Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB), with total 10 layers. There are two Tracker Inner Disks (TID) with 3 wheels per side and two Tracker End Caps (TEC) with 9 wheels per side described in Figure 2.8. The silicon strip tracker contains 9.6 million p^+ strips which are implanted into n-type bulks with thickness of $320 \mu\text{m}$ ($500 \mu\text{m}$). The n-type backplane of the bulk is connected to a positive voltage to form a reverse bias voltage between the backplane and p^+ strip.

2.2.3 Electromagnetic calorimeter

Electromagnetic Calorimeter (ECAL) [51] in the CMS is mounted outside the inner tracker system with $|\eta|$ up to 3. The ECAL is designed in order to precisely measure the electromagnetic

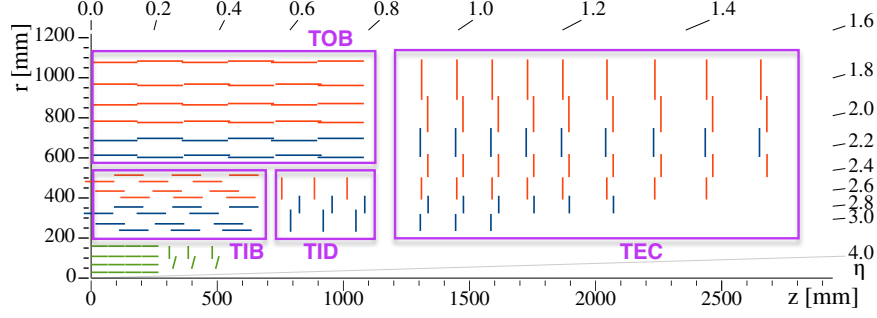


Figure 2.8: Sketch of one quarter of the silicon strip tracker system in $r - z$ view with the axis parallel to the beam (z) running horizontally and radius (r) increasing upward [54]. The interaction point is at the lower left corner. The single-sided and double-sided strip modules are colored with red and blue segments, respectively. In addition, the pixel detector is shown with green segments.

deposit energy of photons from the $H \rightarrow \gamma\gamma$ channel and electrons from the $H \rightarrow 4\ell$ channel. It's made of 75,848 homogeneous lead tungstate crystals (PbWO_4). These crystals have the properties of the high density (8.28 g/cm^3), short radiation length (0.89 cm) and small Molière radius (2.2 cm) to provide the ECAL with high energy resolution and high granularity. Moreover, the crystals also provide high time resolution. Their scintillation decay time is the same order of the LHC bunch crossing time (80% of the light emitted is within 25 ns and 99 % is collected in 100 ns) for avoiding signal overlapping between events. However, relatively low light outputs for the PbWO_4 crystals (4.5 photoelectrons per MeV) need photodetectors with internal gains which are operable in the magnetic field of 3.8T for photon collection and signal amplification, shown in Figure 2.9.



Figure 2.9: Photogram of a homogeneous lead tungstate crystal (PbWO_4) in the EE, where the end of the crystal is a vacuum phototriode (VPT) photodetector.

The ECAL is divided into a central barrel section (EB) covering $|\eta| < 1.479$ and two endcaps (EE) in the forward and backward regions extending $|\eta|$ up to 3, shown in Figure 2.10. The EB part contains 61,200 crystals. Each crystal has a cross-section of $22 \times 22 \text{ mm}^2$ or roughly equivalent to $0.0174 \times 0.0174 (\eta-\phi)$ at the front face, $26 \times 26 \text{ mm}^2$ at the rear face and a length of 230 mm corresponding to $25.8 X_0$. Since the truncated pyramid shaped crystals are placed in

a quasi-projective geometry, their axes have 3° difference with respect to the line extended to the center of the CMS detector. In the EE part, the longitudinal distance between the front face of each endcap and the center of the CMS detector is 315.4 cm. Each endcap is separated into two halves or Dees where 3662 crystals are arranged in a rectangular x-y grid. These crystals point at a focus 1300 mm beyond the center of the CMS detector, giving off-pointing angle ranging from 2 to 8 degrees. Each crystal in two endcaps has a cross-section of $28.62 \times 28.62 \text{ mm}^2$ at the front face, $30 \times 30 \text{ mm}^2$ at the rear face and a length of 220 mm corresponding to $24.7 X_0$. The preshower detector (ES) is a high-granularity photon-sensitive detector covering $1.65 < |\eta| < 2.5$ in front of the crystals in the endcaps. The main purpose is to separate a single photon against a boosted π_0 , where its diphoton decay close to each other can be mis-identified like a single photon. Therefore, the ES adopts a sampling calorimeter composed of two layers with highly spatial resolution. Each layer is composed of a lead absorber with $2X_0$ and a silicon strip sensor with X_0 placed behind the absorber in order to measure the energy deposit and the high-granularity transverse shower profiles. These informations provide additional features to remove the main background contributions like the QCD multi-jets and $\gamma + \text{jets}$ events in the $H \rightarrow \gamma\gamma$ analysis.

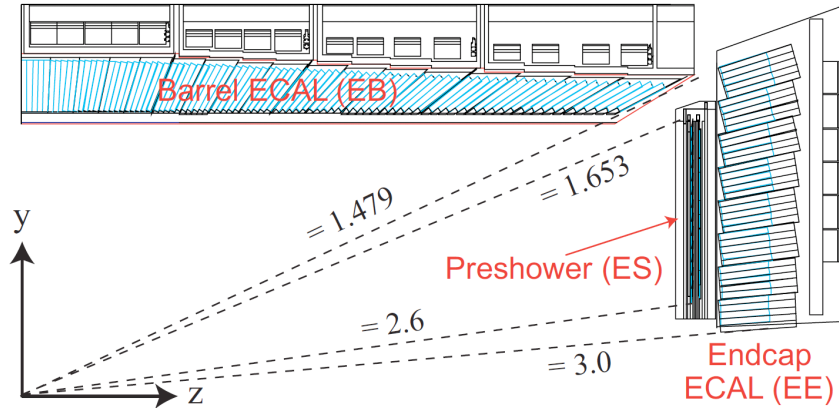


Figure 2.10: Sketch of one quarter of the ECAL in $y - z$ view with the axis parallel to the beam (z) running horizontally and y increasing upward [55]. The interaction point is at the lower left corner. The EB covers the $|\eta| < 1.479$ region, the EE extends up to $|\eta| = 3.0$ and the ES is mounted in front of the EE with $1.65 < |\eta| < 2.5$.

For a simple workflow of the ECAL, the scintillation light produced from crystals with energy deposited is collected by photodetectors, as avalanche photodiodes (APDs) in the EB and vacuum phototriodes (VPTs) in the EE. Then, the electronic signals from the photodetectors are amplified and shaped by the front-end electronics. This Multi-Gain Pre-Amplifier electronics provides three simultaneous analogue outputs with three different gain settings (Gain 1, Gain 6 and Gain 12) and digitized at 40-MHz by a 12-bit analog-to-digital converter(ADC) which provides a discrete set of amplitude measurement. The integrated logic finally selects the largest unsaturated digitization from these three ADCs. These digitized and selected signals are stored

in buffers until a Level-1 trigger is received and finally transmitted to off-detector (beyond 100 m) for further management.

The ECAL barrel energy (E) resolution, reflecting the ECAL performance, has been measured from an electron test beam experiment (beam energy from 20 to 250 GeV) [56] and decoupled into three terms

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

where the electron energy is reconstructed by the shower in 3×3 crystals. The right first term is the stochastic term coming from shower containment, number of photoelectrons and fluctuations in the gain process. The second one is the noise term contributed by the photodetector electronics. The constant term, which dominates in high energy photon and electron showers, depends on non-uniformity of the longitudinal light collection, energy leakage from back of the calorimeter, single-channel response uniformity and stability. In reality, these values are worse than the test beam results due to detector instabilities, magnetic field, pileup during the CMS operation.

2.2.4 Hadron calorimeter

The hadron calorimeter (HCAL) [51] outside the ECAL is a hermetic calorimeter covering $|\eta|$ up to 5 for measuring the energy of hadrons and indirectly estimating the presence of weakly interacting particles like neutrinos. This subdetector drives the reconstruction of the jets originating from quark/gluon fragmentation and hadronization. Therefore, the reconstruction of the dijet invariant mass from the Z/W hadronic decays in the VH production relevantly relies on the HCAL. The HCAL is a sampling calorimeter which is mainly made of repeating layers of dense absorber for producing hadronic showers and tiles of plastic scintillator for producing rapid light pulses when hadronic showers pass through these scintillators. The absorber is composed of flat brass alloy absorber plates with 70% Cu/30% Zn, known for a radiation length of 1.49 cm and a interaction length (λ_I) of 16.42 cm.

For the layout, the HCAL can be separated into several parts: a barrel part (HB), two endcap parts (HE), outer part (HO), shown in Figure 2.11. The HB covering the $|\eta| < 1.4$ region is composed of 17 layers of active plastic scintillator tiles interspersed between these absorber plates, giving a granularity of 0.087×0.087 in the HB which is lower but spatially matches with the granularity of the EB. A set of scintillators with the same value of η and ϕ in different layers is grouped together as a tower. The innermost and outermost absorber layers are made of stainless steel for structural strength. The HB effective thickness absorbers increase with the polar angle from $5.82 \lambda_I$ to $10.6 \lambda_I$. To increase sufficient containment in the central region ($|\eta| < 1.26$) which can avoid the energy leakage due to the limit of HB geometry, additional layers of plastic scintillators, called as HO, with the same position and granularity as the HB, is placed outside the solenoid coil to measure the late shower energy deposited after the HB. Since the absorber in the HO are directly the solenoid coil itself, the total effective length can be extended to a maximum of $11.8 \lambda_I$ in the central region. The HE covers $1.3 < |\eta| < 3.0$ with HB

and HE overlapping region in $1.3 < |\eta| < 1.4$. There are 19 layers of active plastic scintillator interleaved in these brass alloy absorbers, giving two kinds of granularities of 0.087×0.087 in $1.3 < |\eta| < 1.74$ and $(0.09 \sim 0.35) \times 0.087$ in $1.74 < |\eta| < 3.0$. The total effective length including the EE is approximately $10 \lambda_I$.

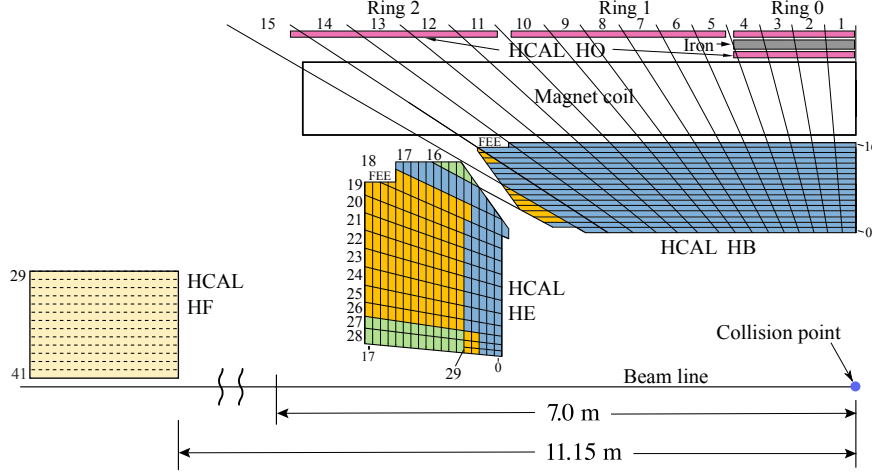


Figure 2.11: Sketch of the tower mapping in $R - z$ of the HCAL barrel part (HB), two endcap parts (HE), outer part (HO) and forward part (HF) with the axis parallel to the beam (z) running horizontally and radius (R) increasing upward [57]. The interaction point is at the lower right corner.

An additional calorimeter called forward calorimeters (HF) are located 11.2 m from the center of the CMS detector and cover $3.0 < |\eta| < 5.2$ region. It's made of steel absorbers like a cylindrical structure (1.65 m long, 12.5 cm inner radius and, 130.0 cm outer radius) with radiation hard quartz fibers inserted, which provide the collection of the Cherenkov lights produced by charged shower particles. These fibers running parallel to the z -axis and are bundled to form a granularity of 0.175×0.175 . Among these fibers, half of them run over the full depth of the absorber ($10\lambda_I$) and half start a depth of 22 cm from the front of the detector. Such arrangement is able to identify the photons and electrons which have large fraction of energy deposited in the first 22 cm.

For a simple workflow of the HCAL, a hadron particle initially goes through successive layers of absorber and produces an amount of secondary particles. These secondary particles pass through the alternating layers of active scintillation which emit blue-violet light collected by optical wavelength-shifting fibres, shifting the blue-violet light to the green one. Then, the normal optic cables send the green optic signal to the front-end electronics where the optical signal is transformed into fast electronic signals by Hybrid Photodiodes (HPDs) and the signals in the same tower are summed. Finally, the signals are sent to data acquisition system for event triggering. The combined calorimeter (ECAL + HCAL) energy resolution was also measured from a test beam experiment (beam energy from 2 to 350 GeV) [58]

$$\frac{\sigma_E}{E} = \frac{84.7\%}{\sqrt{E(\text{GeV})}} \oplus 7.4\% \quad (2.9)$$

where the right first term is a stochastic term and the second term is a constant term.

2.2.5 Solenoid magnet system

As the name of the CMS, the CMS detector adopts a solenoid magnet shaped as a cylinder with 6.4 m diameter and 12.5 m length, containing four layers of the coils made by stabilised reinforced NbTi conductor. These coils are surrendered by the 220 tonne cold mass cooled down to a temperature of 1.8 K which makes these coils as superconducting with a current of 19.14 kA, corresponding to a stored energy of 2.6 GJ.

The solenoid magnet provides a homogeneous magnetic field of 3.8 Tesla inside this cylinder structure including the inner track system, ECAL and HCAL. The main goal for the magnetic field is to bend charged particle trajectories for measuring their charges and momenta. However, the return magnetic field outside the solenoid forms non-uniform and sparse flux which reduces its strength in the muon detector. This could leads to ambiguous bending of muon trajectories and worse momentum resolution, especially high-energy muons. To strengthen the magnetic field in the muon detector, a steel flux-return yoke with a weight of 12,500 tonne comprising 5 wheels in the barrel part and 2 endcaps is interspersed between the muon chambers. This increases the strength of magnetic field to 2 Tesla which gives better resolution for measured muon momenta. A complete distribution of the magnetic field strength for the magnet is shown in Figure 2.12.

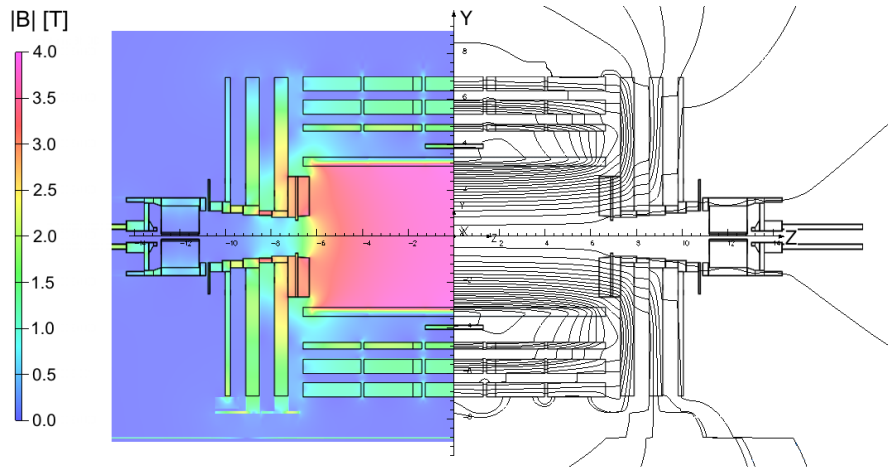


Figure 2.12: Diagram of one half of the solenoid magnet system in the CMS detector with the distribution of the magnetic field strength expressed by color in the left half and magnetic field lines in the right half [59].

2.2.6 Muon detector

Since muons are approximately 200 times heavier than electrons, the bremsstrahlung radiation for the muons from a few to several hundreds GeV energy is relatively small so that they can pass through all matters in the CMS detector. Therefore, a gas-chamber type of the muon detec-

tor [60] covering the $|\eta| < 2.4$ is located outside the solenoid to reconstruct the muon tracks and provide precise momentum measurement, high efficiency muon identification and triggering. This significantly helps the VH analysis extract the rare VH events with the W/Z decaying to muons from an amount of QCD multi-jets events. It's mainly made of three kinds of gas ionization chambers : drift tube chambers (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs), mounted on these iron yokes, shown in Figure 2.13.

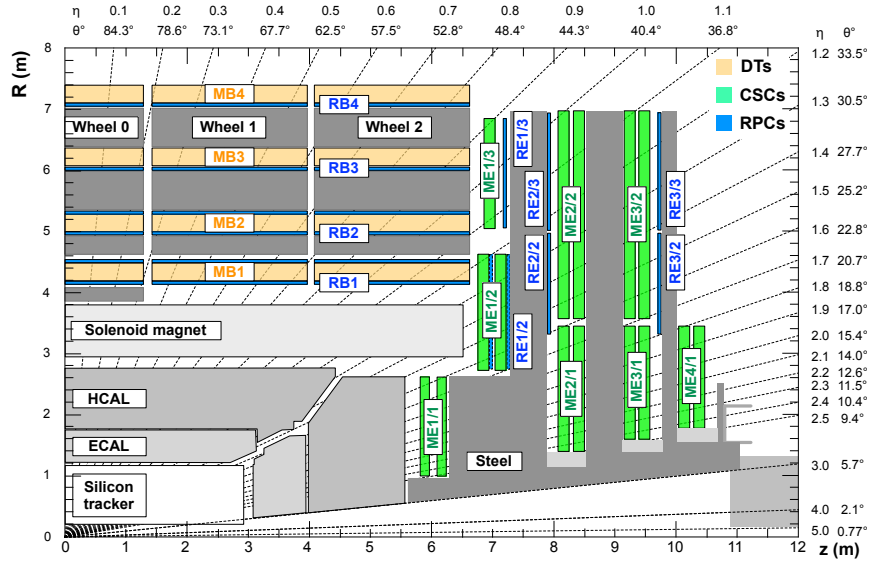


Figure 2.13: Sketch of one quarter of the CMS detector in $R - z$ view with the axis parallel to the beam (z) running horizontally and radius (R) increasing upward [60]. The interaction point is at the lower left corner. The 4 drift tube (DT, in light orange) stations are labeled MB (muon barrel) and the cathode strip chambers (CSC, in green) are labeled ME (muon endcap). Resistive plate chambers (RPC, in blue) are in both the barrel and the endcaps of CMS, where they are labeled RB and RE, respectively.

In the barrel part, the drift chambers with standard rectangular cells and sophisticated electrical field shaping are used in the barrel muon system covering the $|\eta| < 1.2$ region. The basic element of the DTs is a drift tube cell with a transverse size of $42 \times 13 \text{ mm}^2$ and a length of 2.4 m. A $50\text{-}\mu\text{m}$ -diameter and 2.4-m-length gold-plated stainless-steel anode wire is centered inside each tube, where there are four electrodes for shaping the effective drift field shown as Figure 2.14 (left). Inside the cell, a mixture of the Ar (85%) and CO_2 (15%) gas is filled to give induced electrons a saturated drift velocity of about $55\mu\text{m/ns}$ towards the anode when muons go through the cell and ionize the gas. For these tubes, four layers of these drift tubes form a superlayer and three superlayers form a drift chamber as a station. Among 3 superlayers, the wires in the 2 outer superlayers are parallel to the beam line for the muon track measurement in the $r - \phi$ coordinate where r is the distance from the beam line, while the middle superlayer are orthogonal to the beam line for the $r - z$ coordinate measurement except MB4. Such design gives a better spatial resolution than $250 \mu\text{m}$ in MB1, MB2 and MB3 and $300 \mu\text{m}$ in MB4 for the ϕ superlayers and a spatial resolution of about 250 to $600 \mu\text{m}$ in the z superlayers except in the

outer wheels of MB1 [61].

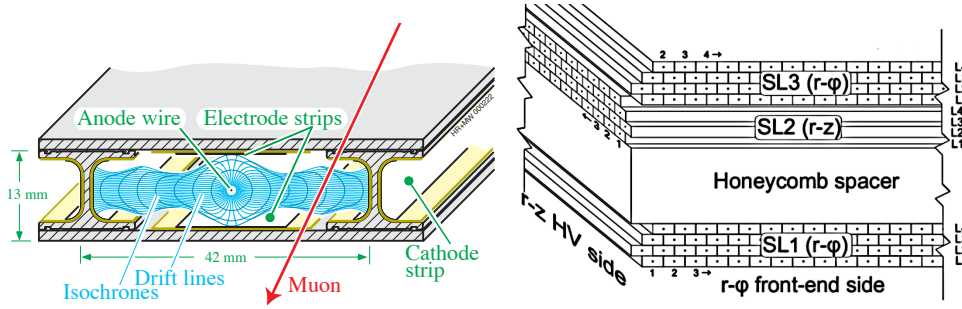


Figure 2.14: The left is the cross section of a drift tube cell. There are four electrodes for shaping the effective drift field: two on the side walls of the tube as cathode strips with 1200 V, and 2 above and below the wires on the ground planes between the layers with +1800 V. The center anode wire applies a high voltage of +3600 V. The right is the layout of the drift tubes where the 2 outer superlayers (SL) are parallel to the beam line for the $r - \phi$ coordinate measurement and the middle superlayer is orthogonal to the beam line the $r - z$ coordinate measurement [60].

In the endcaps of the CMS detector, since muon rates and background level are higher and the magnetic field is not uniform, the muon detector in the endcap part adopts cathode strip chambers (CSCs) with fast response time, covering a region of $0.9 < |\eta| < 2.4$, which provide good timing and spatial performance. Each CSC consists of 6 anode wire planes interleaved among 7 cathode panels. Each cathode panel contains 80 cathode strips arranged like a radial structure with a constant ϕ width and project to the beam line. The wires with a diameter of $50 \mu\text{m}$ (except ME1/1, $30 \mu\text{m}$ spaced by 2.5 mm) on the anode planes run azimuthally and are spaced by 3.16 or 3.12 mm, almost perpendicular to the strips. Among these layers, there are 6 gas gaps corresponding to 6 anode wire planes, each of which is filled by a mixture of the CO_2 (50%), Ar (40%) and CF_4 (10%) gas. This structure makes the CSCs as standard multi-wire proportional counters, but with a cathode strip readout that precisely measures the position of a muon when the muon transverses across the gas volume. It gives a spatial resolution of $75 \mu\text{m}$ for ME1/1 chambers and $150 \mu\text{m}$ for the others [61].

Besides the muon chambers as the DTs and CSCs, a triggering detector system called resistive plate chambers (RPCs) mounted in both barrel and endcaps provides excellent time resolution for a fast, independent trigger. Each RPC has two gaps, each of which is composed of two 2-mm-thick resistive Bakelite plates with a 2-mm-thick gas gap filled by a mixture of the $\text{C}_2\text{H}_2\text{F}_4$ (95.2%), $i\text{-C}_4\text{H}_{10}$ (4.5%) and SF_6 (0.3%) gas. There is a thin conductive graphite layer on the outer surface of the resistive plates applied a high voltage of 9.6 kV which provides avalanches when charged particles go through a RPC and ionize the gas. The avalanches quickly send to the readout strips and form a signal.

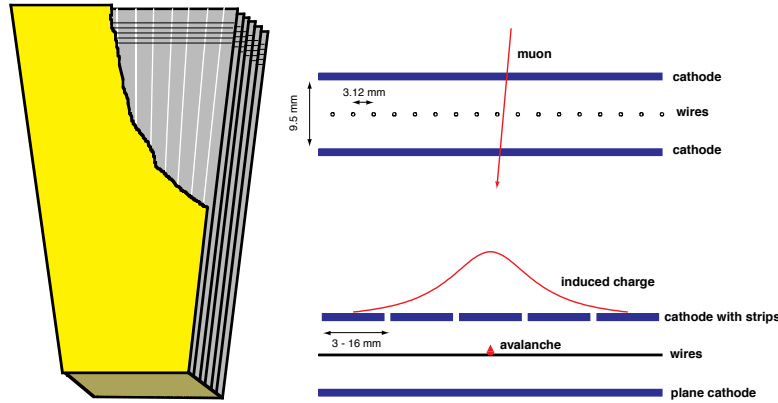


Figure 2.15: The left is the cut-away diagram of a cathode strip chamber with the 6 layers, where the 80 strips are arranged with a radial structure and are spaced by the white lines, while the black and horizontal lines are the wires (not all shown). The right is cross-sectional views of the gas gap in a CSC gap with the anode wires and cathode planes, and a schematic of the gas ionization avalanche and induced charge distribution on the cathode strips [60].

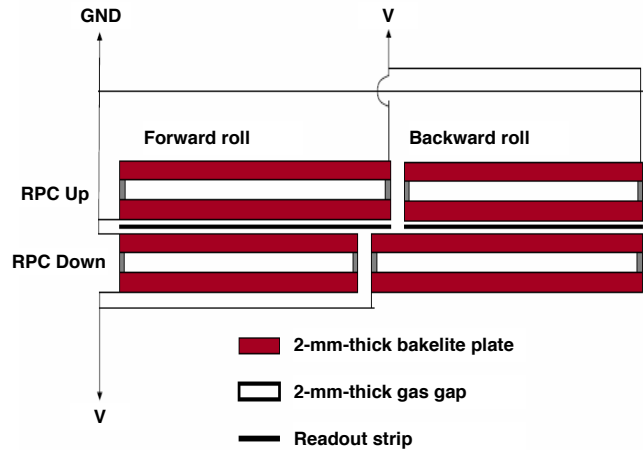


Figure 2.16: Schematic view of a generic barrel RPC with 2 roll partitions [60].

2.3 Trigger

The LHC provides high-luminosity proton-proton collisions with a bunch crossing rate of 40 MHz for producing rare interactions like Higgs productions and physics beyond the Standard Model. However, due to the limit of CPU time consumption, buffer size and disk storage during data collection, recording the information of all collisions is impossible and also unnecessary as most of collisions are not the physics of interest. Therefore, the CMS experiment performs a Level-1 trigger system using customized hardware electronics to select the most interesting events against background events, with a maximal rate of 100 kHz from the original rate of 40

MHz. Then, events pass the High Level Trigger (HLT) used by the CMS offline reconstruction software in commercial PCs which further reduces the rate to 1 kHz. Finally, these events are sent to the CERN Tier system for reconstruction, calibration and storage.

2.3.1 Level-1 trigger

Facing an amount of events from proton-proton collision with the rate of 40 MHz, the Level-1 trigger system [62] in the CMS experiment uses information only from the muon and calorimeter detectors with coarse granularity and precision to immediately but roughly reconstruct physical objects like photons/electrons, muon and missing transverse energy and so on. The trigger algorithm then uses simple criteria like transverse energy E_T or pseudorapidity η of the physical objects to select interesting events like the Higgs boson productions immediately. In the Level-1 trigger system, a few criteria are grouped to form a so-called "seed" and all seeds are collected in a "menu". Once an event satisfies the requirements of at least one seed in the menu, its complete detector information stored in the front-end buffers will be read out and sent to the HLT. Due to limited bandwidth in electronics, the 70% seeds focusing on calibration or monitoring are, in fact, prescaled with a prescale factor of N which means only one event is read out for N events passing this seed, while the 30% seeds related to physics analysis are unscaled.

The $H \rightarrow \gamma\gamma$ analysis needs events to pass a Level-1 seed with at least one electromagnetic candidate (L1.SingleEG, called single electron/photon) to avoid any loss of possible $H \rightarrow \gamma\gamma$ events. However, the bandwidth limit makes the lowest unscaled L1.SingleEG seed have to satisfy a requirement of $E_T > 40$ GeV that gives inefficiency for photons with the lowest transverse energy requirement in the $H \rightarrow \gamma\gamma$ analysis. To recover data with photons in the lower energy region, a prescaled Level-1 seed with a pair of electromagnetic candidates, where transverse momenta of both EM candidates are different, are included as "OR" in the analysis. It is observed that the rate of events passing at least one of the lowest unscaled L1.SingleEG or L1.DoubleEG seeds is up to 99%. The impact of 1% loss for the prescaled L1.DoubleEG is found to be negligible when the effect propagates to the diphoton high level trigger. The lowest unscaled L1.SingleEG and prescaled L1.DoubleEG used in the analysis are listed in Table 2.2.

Year	L1 seed	E_T threshold (GeV)
2016	Single e/γ	40
	Single e/γ (isolated)	30
	Double e/γ	23, 10
2017, 2018	Single e/γ	40
	Single e/γ (isolated)	32
	Single e/γ (isolated and with $ \eta < 2.1$)	30
	Double e/γ	25, 14

Table 2.2: Lowest unscaled L1 single e/γ seeds corresponding to E_T thresholds and prescaled L1 double e/γ seeds with asymmetric E_T thresholds for 2016, 2017 and 2018.

2.3.2 High Level Trigger

Events passing the Level-1 trigger will send to the High Level Trigger system. The High Level Trigger (HLT) [63] in the CMS experiment is overall a software system running under PCs with about 32000 available CPU cores supported. The trigger algorithm is composed of a sequence, called as HLT Path, of the offline reconstruction with light version and filter modules using complete detector information which allows more complete and precise reconstruction. Therefore, not only simple physical quantities like E_T or η but more sophisticated variables like shower shape, isolation and track-vertex related variables are considered as criteria in the menu. Although the maximal processing time per event is up to approximately 320 ms, the HLT provides higher performance to identify the most interesting events and to reduce to a rate of 1 kHz from 100 kHz given by the Level-1 trigger.

The HLT criteria in the $H \rightarrow \gamma\gamma$ analysis are composed of general variables like E_T , η , diphoton invariant mass $m_{\gamma\gamma}$ and shower shape variables like R_9 , $\sigma_{i\eta i\eta}$, H/E . The isolation variables like the ECAL Particle-Flow cluster isolation with the mean energy density, ρ , correction $\mathcal{I}_{\text{EM PF}}^{\text{HLT}}$ and the tracker isolation $\mathcal{I}_{\text{trk}}^{\text{HLT}}$ within $0.06 < \Delta R < 0.3$ are included as well. Every Level-1 seeded leg in the diphoton HLT is required to pass the selections of $E_T > 30$ GeV and $|\eta| < 2.5$. A set of selection requirements, shown in Table 2.3 without $\mathcal{I}_{\text{trk}}^{\text{HLT}}$, is also applied to the seeded legs. If events are found to have at least a seeded leg passing the above requirements, a complete HLT ECAL cluster algorithm is performed. Then, the events with at least one of the unseeded clusters with $E_T > 22(18)$ GeV and $|\eta| < 2.5$ passing the criteria in Table 2.3 and satisfying $m_{\gamma\gamma} > 90$ GeV are kept for later event reconstruction.

	R_9	H/E	$\sigma_{i\eta i\eta}$	$\mathcal{I}_{\text{EM PF}}^{\text{HLT}}$ (GeV)	$\mathcal{I}_{\text{trk}}^{\text{HLT}}$ (GeV)
Barrel	> 0.85	< 0.12	-	-	-
	$[0.50, 0.85]$	< 0.12	< 0.015	$< 6.0 + 0.012E_T$	$< 6.0 + 0.002E_T$
Endcaps	> 0.90	< 0.1	-	-	-
	$[0.80, 0.90]$	< 0.1	< 0.035	$< 6.0 + 0.012E_T$	$< 6.0 + 0.002E_T$

Table 2.3: Content of the diphoton trigger requirements.

Chapter 3

Event reconstruction and selection

In the CMS experiment, event reconstruction relies on an so-called *Particle-Flow algorithm*. The Particle-Flow (PF) algorithm [64] aims at the reconstruction and identification of all stable particles i.e. muons, electrons, photon, charged hadrons and neutron hadrons from collisions by combining the information from the different subdetectors, described by Figure 3.1.

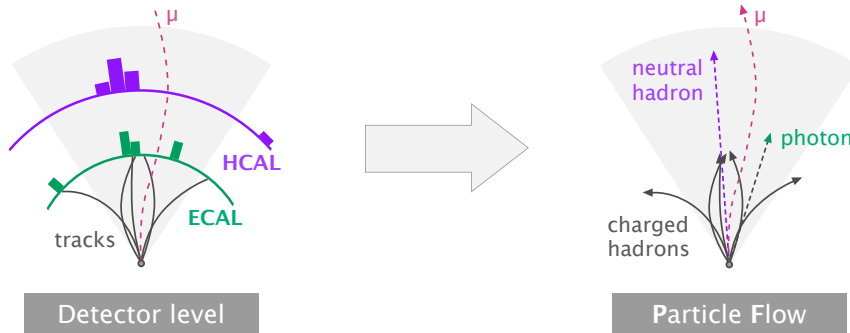


Figure 3.1: Schematic of the Particle Flow algorithm from detector level with tracks and clusters to the stable particle reconstruction.

The algorithm starts from basic object reconstruction, i.e. tracks and clusters. The tracks (trajectories) of charged particles are reconstructed by hits on sensitive layers from either the inner tracker system or the muon detector. According to the reconstructed tracks, the positions of vertices associated with these tracks are further built in order to reconstruct interaction points of proton-proton collisions. The position and energy deposited of electromagnetic and hadronic showers from the ECAL and HCAL are reconstructed by dedicated clustering algorithms. These tracks and clusters are finally collected to form a set of the PF elements.

Once the PF elements are reconstructed, the PF algorithm then performs dedicated link algorithms corresponding to targeted stable particles to determine the relationship between different PF elements. The link algorithm will spatially connect the vertices, tracks and showers together and identify to some stable particles. The corresponding energy and propagation direction of each stable particle can be optimally determined. Once the targeted particles are

reconstructed, the associated PF elements are masked from the list of PF elements against the following reconstruction, that is, are not used for other object building. For the PF algorithm in the CMS experiment, the reconstruction procedure follows a priority depending on the level of significant features for given objects. It begins from the muon reconstruction with the most significant features against other stable particles. Then, the algorithm aims at the isolated photons and electrons and finally the rest non-isolated photon, charged and neutron hadrons are reconstructed. These physical objects are finally collected as a set of PF candidates. With the PF candidates, particle-based physical objects as jets and missing transverse energy can be further reconstructed.

The VH production in the $H \rightarrow \gamma\gamma$ channel involves in most of the physical objects, especially, photons which are the most important objects for the Higgs boson reconstruction. Moreover, each used physical objects in the VH analysis uses dedicated identification strategies in order to reject objects due to misreconstruction or low quality. Therefore, this chapter will introduce the reconstruction and identification of the PF elements in Section 3.1 and each physical object used in the VH analysis is introduced in Section 3.2 for muons, photons and electrons in Section 3.3 and jets and missing energy in Section 3.4 and Section 3.5, respectively.

3.1 Particle flow element

Inner tracks and clusters are the most basic elements in the Particle-Flow algorithm. They determine trajectories of charged particles, interaction points, hit positions on the subdetectors and energy of particles and support more precise energy-momentum measurement of the charged particles from track-cluster connection. The reconstruction and identification of high-level physical objects, i.e. all stable particles, and particle-based objects such as jets further rely on these PF elements. Therefore, the inner tracks and clusters plays an important role in overall event reconstruction. This section will introduces both inner track and cluster reconstruction.

3.1.1 Inner track and vertex reconstruction

Track reconstruction in the CMS experiment uses a dedicated tracking software called Combinatorial Track Finder (CTF) aka iterative tracking which runs several iterations of combinatorial Kalman filter [65] integrating pattern recognition and track fitting with respect to a normal Kalman filter (KF) algorithm. The iteration initial looks for the most significant tracks with relatively high p_T or close to interaction points. After this iteration is finished, the hits associated with these reconstructed tracks are masked from a general hit list to reduce the combinatorial complexity when moving on the next iterations which need to search for tracks with lower significance. In the CMS experiment, there are several iterations for the tracking process, each of which has different configuration and targeted tracks. The first iteration aims at the prompt tracks with $p_T > 0.8$ GeV and three (four) pixel hits which are the dominant source in all reconstructed tracks. The next iteration recovers the prompt tracks with only two (three) pixel hits. The later iterations will search for lower p_T tracks or tracks not originating from interac-

tion points or not found in the previous iterations. Each tracking iteration is composed of four steps: seed generation, track finding, track fitting and track selection described in Ref [66].

1. **Seed generation** : Seeds are defined by 2 or 3 hits compatible to a helical trajectory hypothesis from an origin like beam spot area under the quasi-uniform magnetic field. These seeds determine the initial trajectory parameters and their uncertainties.
2. **Track finding** : The KF algorithm starts from the trajectory parameters with rough estimate provided by those track seeds and constructs track candidates by including more hits from successive tracker layers, updating the trajectory parameters at each layer. During this procedure, the effects of multiple Coulomb scattering, inhomogeneous magnetic field and energy loss as a Gaussian distribution are also included. The procedure is repeated layer by layer from the seeds until termination conditions are satisfied. The trajectory candidates are formed.
3. **Track fitting** : Once the hits are collected, the Kalman filter and smoother then refits to each track candidate to provide the optimal track parameters at any point, including the innermost and outermost hits on the track which are used to extrapolate the trajectory to interaction points and to outermost sub-detectors, respectively. The refitting iterates the full collected hits from inside outwards to update the track trajectory estimate sequentially with each hit. Each iteration scales up the covariance matrix with a large factor to avoid the bias from the beam spot constraint.
4. **Track selection** : These refitted trajectories are selected by quality requirements such as good $\chi^2/\text{d.o.f}$ and number of layers for reducing fake tracks, defined by a reconstructed track not associated with a charged particle.

Finally, the tracks found by each of the these iterations are merged into a single collection. Figure 3.2 shows the taking efficiency can achieve at least 98% in the CMS experiment.

Vertex reconstruction

Once the tracks are built, vertices associated with these tracks can be further reconstructed. The purpose of the vertex reconstruction is to measure the position and associated uncertainty of the interaction points from proton-proton collections, including the primary vertex from which the physics of interest come and vertices from pileup collisions. The vertices, especially the primary vertex, are quite important information for further physical object reconstruction and identification.

There are three steps in the vertex reconstruction. The first step selects the tracks originating from the interaction region. Therefore, some requirements such as impact parameters with respect to the beam spot area and simple track qualities without the p_T criterion are needed. Next step is the grouping of the tracks that originate from the same interaction vertex. This step performs a deterministic annealing (DA) algorithm [67] and uses the z-coordinates of the selected tracks and the beamspot position to optimally group the selected tracks. Finally, each

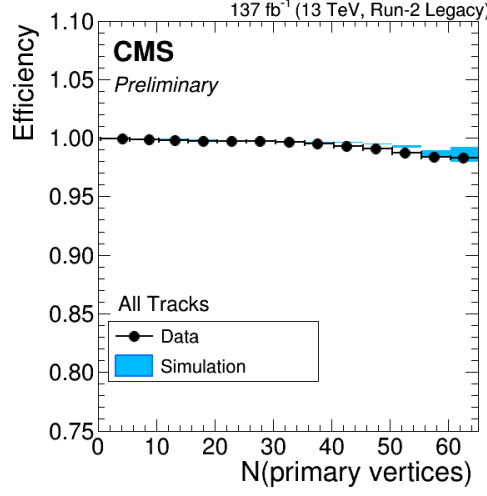


Figure 3.2: The tracking efficiency as a function of the number of primary vertices in the “all tracks” collection. Data (black dots) and Madgraph DY (light blue rectangles). The uncertainties shown are statistical.

vertex is reconstructed by fitting a group with at least two tracks to obtain the best estimate of the vertex parameters like its x , y and z position. The vertex with the maximum scalar sum of the associated tracks p_T is eventually considered as the primary vertex.

3.1.2 Calorimeter cluster reconstruction

Cluster reconstruction initially looks for cluster seeds in the ECAL or HCAL cells with local maximal energy with respect to the adjacent cells, larger than the seed threshold. Then, topological clusters are grown from these seeds by merging neighbouring cells with an energy larger than a cell threshold of twice the noise level. From a given topological cluster, an expectation-maximization algorithm based on a Gaussian-mixture model is performed reconstructing calorimeter clusters. The Gaussian-mixture model supposes that the energy deposits in the M individual cells of the topological cluster originating from N calorimeter clusters (the same as the number of seeds) with the distribution of energy deposits as Gaussian functions in the (η, ϕ) plane. Therefore, in the algorithm, an expected function, f_{ji} , is defined as energy E_j in the cell with position $\vec{c}_j$ arising from the i -th Gaussian energy deposit, calculated by

$$f_{ji} = \frac{A_i e^{-(\vec{c}_j - \vec{\mu}_i)^2 / (2\sigma^2)}}{\sum_{k=1}^N A_k e^{-(\vec{c}_j - \vec{\mu}_k)^2 / (2\sigma^2)}} \quad (3.1)$$

where A_i and $\vec{\mu}_i$, which are free parameters of this model, represent the energy and position of the i -th cluster. The energy and position of the seeds are used to initialize the parameters in the Gaussian-mixture model, and the free parameters can be determined by an analytical

maximum likelihood fit yielding

$$A_i = \sum_{j=1}^M f_{ji} E_j \quad (3.2)$$

$$\vec{\mu}_i = \sum_{j=1}^M f_{ji} E_j \vec{c}_j \quad (3.3)$$

The expectation maximization cycle is repeated until stable convergence, and the best fit positions and energies in the Gaussian functions are taken as the clusters' parameters.

Calorimeter cluster calibration

In the PF algorithm, neutral particles are only reconstructed from calorimeter clusters in the ECAL and HCAL. However, if neutral-particle energy deposits in calorimeters are overlapping with charged-particle ones, these neutron particles can only be identified from the energy excess with respect to the momentum of the charged-particle tracks. An accurate energy calibration for calorimeter clusters thus effectively raises the efficiency of the neutral particle identification. More importantly, it helps the $H \rightarrow \gamma\gamma$ analysis obtain better resolution of the $m_{\gamma\gamma}$ distribution.

ECAL cluster calibration [64] mainly aims at the energy loss resulting from certain thresholds during the cluster reconstruction. To recover the clusters' energy, especially for low-energy photons, the calibration uses a single photon without any conversion¹ for focusing on a clean single cluster. In this case, the clusters located on the EB and EE uses different method. In the EB, an analytical function,

$$f(E, \eta) = g(E)h(\eta) \quad (3.4)$$

as a function of the clusters' energy without any correlation with their pseudorapidity, is fitted to the two-dimensional distribution of $\langle E^{\text{true}}/E \rangle$ in (E, η) plane. In the EE, since there are additional two layers of the preshower detector in front of the ECAL crystals, the calibrated energy for the clusters in the EE can be expressed as

$$E^{\text{calib}} = \alpha(E^{\text{true}}, \eta^{\text{true}})E_{\text{ECAL}} + \beta(E^{\text{true}}, \eta^{\text{true}})[E_{\text{PS1}} + \gamma(E^{\text{true}}, \eta^{\text{true}})E_{\text{PS2}}] \quad (3.5)$$

where α, β, γ are the calibration parameters depending on the energy and pseudorapidity of the particle-level photons and $E_{\text{ECAL}}, E_{\text{PS1}}$ and E_{PS2} represent energy deposits of the ECAL clusters, the first and second preshower layers, respectively. These calibration parameters are given in each $(E^{\text{true}}, \eta^{\text{true}})$ bin by a χ^2 minimization as

$$\chi^2 = \sum_{i=1}^{N_{\text{events}}} \frac{(E^{\text{calib}}_i - E^{\text{true}}_i)^2}{\sigma_i^2} \quad (3.6)$$

where σ_i is the expected calorimetric energy resolution of the i -th single photon which is also energy dependence.

¹The conversion arises from that photons hit on atomic nuclei in the inner tracker system and convert to e^+e^- .

Hadron cluster energy deposits rely on the clusters from both the ECAL and HCAL. Although the ECAL clusters have been calibrated by Eq. (3.4-3.5), the detector response for hadrons in the ECAL is not the same as that for photons. The ECAL and HCAL cluster energies are therefore recalibrated to get an estimate of the true hadron energy. The calibrated energy for a hadron is given by

$$E^{\text{calib}} = a + b(E^{\text{true}})f(\eta^{\text{true}})E_{\text{ECAL}} + c(E^{\text{true}})g(\eta^{\text{true}})E_{\text{HCAL}} \quad (3.7)$$

where a, b, c are the calibration coefficients as well as the functions f and g ; E^{true} and η^{true} are the true energy and pseudorapidity of the hadron; E_{ECAL} is already calibrated from Eq. (3.4-3.5). The event-independent coefficient a (GeV) is to recover the loss from the energy thresholds during the cluster reconstruction, so this constant is given according to different conditions of the hadron cluster. The calibration uses a lot of simulated samples of a single neutral particle by minimizing a χ^2 method

$$\chi^2 = \sum_{i=1}^N \frac{(E^{\text{calib}} - E_i^{\text{true}})^2}{\sigma_i^2} \quad (3.8)$$

where σ_i is the expected calorimetric energy resolution of the i -th single hadron which is also energy dependence. The calibration starts from the estimation of b, c with energy dependence, and the $f(\eta^{\text{true}})$ and $g(\eta^{\text{true}})$ are thus set to unity. Once the b and c are determined from this χ^2 minimization, the relatively small residual with η dependence can be recalibrated by the same χ^2 minimization, defined by Eq. 3.8, with the second order polynomials of $f(\eta^{\text{true}})$ and $g(\eta^{\text{true}})$.

3.2 Muon

In the associated VH production, muons originate from the W or Z boson decay. However, since the muon decay from the W and Z boson is relatively small, the performance of muon reconstruction and identification becomes quite important for reconstructing more muons with high quality and rejecting more fake muons. It can help the VH analysis achieve better sensitivity. The following section will introduce the muon reconstruction and identification in the CMS experiment.

3.2.1 Muon reconstruction

Muons are first reconstructed under the Particle-flow algorithm. Since muons only leave the energy on the trajectories as hits in all sub-detectors, the muon reconstruction is thus based on muon track reconstruction. In the CMS experiment, there are three kinds of reconstruction method [61] for muons:

- **Standalone muon** : Muon tracks are reconstructed by combining hit informations only from the CSC, DT, and RPC in the muon detector. The reconstruction starts from the seed generation made up of the DT or CSC segments and then uses a Kalman-filter technique to build the muon trajectories.

- **Tracker muon** : Muon tracks are built "inside-out" by extrapolating the inner-tracker tracks, introduced in Section 3.1.1, with $p_T > 0.5$ GeV and a total momentum $p > 2.5$ GeV to the muon system with loosely matching to the DT or CSC segments. If at least one muon segment matches with the extrapolated track, the inner track is considered as a tracker muon track.
- **Global muon** : Muon tracks are built "outside-in" by matching standalone-muon tracks and inner-tracker tracks. Once the matching is qualified by comparing the parameters given by both two kinds of tracks propagating to some common surface, a combined fit by the Kalman-filter technique is performed constructing a global muon track from the information of both the standalone-muon track and inner-tracker track.

Due to high efficiency of the tracker track and muon segment reconstruction, about 99 % muons fallen in the detector acceptance are reconstructed by both tracker muon tracks and global muon tracks. If the global muon track actually matches with the tracker muon track, both two tracks will be merged as a single object. Once the muons are reconstructed, the associated PF elements are masked against the following physical object reconstruction.

On the other hand, to obtain the most precise momentum of the muons in different situations when the muons propagate from interaction points and go through the multiple detectors, the momentum measurement performs a *Tune-P algorithm* to decide a kind of p_T measurement from the following four algorithms by goodness-of-fit and $\sigma(p_T)/p_T$ criteria.

- **Inner-Track fit** : The momentum is determined by using the information only from the inner tracker. For muons with $p_T < 200$ GeV, the information from the muon detector has less contribution for momentum reconstruction. The Tune-P algorithm thus prefers the method for low p_T muons.
- **Tracker-Plus-First-Muon-Station (TPFMS) fit** : Since the innermost muon station provides the best information for momentum measurement among the muon detector, the fit performs a refit to the hits of the global muon track from the inner tracker and innermost muon station.
- **Picky fit** : Similarly to TPFMS fit, the picky fit also takes the hits from the global muon track but focuses on the case that a large hit occupancy like a shower happening in the muon chambers. In this large hit occupancy, only the hits that are compatible to the extrapolated trajectory (based on χ^2) are added into the refit work.
- **Dynamic-Truncation fit** : It aims cases that the trajectory of the muon is significantly bent due to energy loss. The algorithm extrapolates the inner-tracker track to the innermost station and performs a refit adding compatible hits from the segment closest to the extrapolated trajectory.

3.2.2 Muon identification and isolation

In the VH analysis, muon candidates are defined according to a set of the identification criteria with a "medium" working point provided by the CMS collaboration (CMS Muon POG) and re-

quire $p_T > 10$ GeV. This working point is optimized for identifying prompt muons and muons from heavy flavor decays. The selection criteria within this working point include the track fit χ^2 during the muon reconstruction, the level of matching between the tracker track and the standalone-muon track, number of hits in the inner tracker and muon detector, and segment compatibility as a quality when propagating the tracker track to the muon detector. A kink-finding algorithm are also donated for the muon identification. The algorithm splits the tracker track into two individual tracks at several places along the muon trajectory and compares between each other through a χ^2 method. A large χ^2 gives the fact that the two individual tracks does not belong the single track. The summary about the "medium" working point is shown in Table 3.1. The corresponding efficiency gives a value of more than 98% for muons with $p_T > 15$ GeV, shown in Figure 3.3. This is also true for muons with $10 < p_T < 15$ GeV.

Muon type	Description	Requirement
—	PF muon	Yes
	Tracker muon	Yes
Global muon	χ^2/ndof of global-muon track fit	< 3
	χ^2 of position match between the tracker muon and standalone muon	< 12
	Max. χ^2 given by the kink-finding algorithm	< 20
	Segment compatibility	> 0.303
Non-global muon	Segment compatibility	> 0.451

Table 3.1: Selection criteria at "medium" working point in the muon identification.

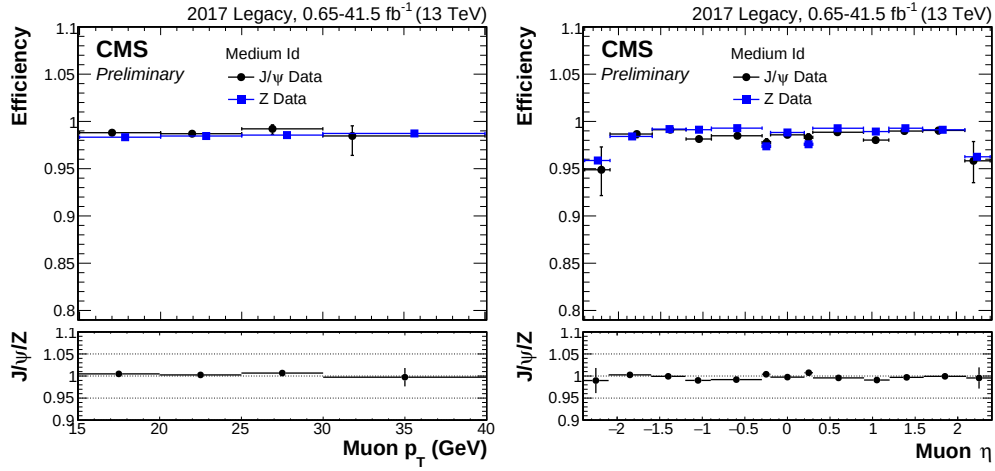


Figure 3.3: The efficiency of the "medium" ID for the 2017 legacy data and simulation is presented as a function of muon p_T and η . The denominator is tracker muons. The efficiencies obtained from Z (blue) and J/ψ (black) resonances show good agreement within 1% difference, shown in the bottom panel [68].

In addition, to distinguish prompt muons originating from the W and Z boson decay against those from heavy flavor decays within jets, an isolation criterion is also applied to each muon.

The criterion uses a relative isolation variable ($\mathcal{I}_{\text{rel},\mu}$) defined by the energy scale sum of all other particles within a geometrical cone, $\Delta R = 0.4$, surrounding the muon

$$\mathcal{I}_{\text{rel},\mu} = \frac{1}{p_{T,\mu}} \left[\sum p_T^{\text{ch.had.}} + \text{Max.} \left(0, \sum p_T^{\text{neutr.had.}} + \sum p_T^\gamma - \Delta\beta \sum p_T^{\text{ch.had.,PU}} \right) \right] \quad (3.9)$$

where $p_{T,\mu}$, $p_T^{\text{ch.had.}}$, $p_T^{\text{neutr.had.}}$, p_T^γ represent the transverse momentum of the muon, charged hadrons, neutral hadrons and photons, respectively. Since the neutral particles from PU vertices can not be eliminated within the tracker acceptance, the energy of the neutral particles in the cone is overestimated. According to simulations, the ratio of neutral particle and charged hadron production in inelastic proton-proton collisions is approximately 0.5. A PU correction is thus applied to the relative isolation variable with $\Delta\beta = 0.5$ [69] to the energy scale sum of the charged hadrons from PU vertices ($\sum p_T^{\text{ch.had.,PU}}$). The VH analysis chooses a "loose" working point $\mathcal{I}_{\text{rel},\mu} < 0.25$ whose efficiency is more than 98%.

3.3 Photon and Electron

In the $H \rightarrow \gamma\gamma$ analysis, photon reconstruction plays an crucial role in the resolution of the diphoton invariant mass distribution. A good resolution provides better sensitivity and more precise results for measurements. Since the diphoton mass reconstruction is determined by the energy and opening angle of two photons, following a formula of

$$m_{\gamma\gamma} = \sqrt{2E_{\gamma 1}E_{\gamma 2}(1 - \cos \alpha)} \quad (3.10)$$

where α is the opening angle of two photons, the corresponding experimental mass resolution can be decoupled to the resolution of the measurement of these two photon energies and their opening angle as

$$\frac{\delta m_{\gamma\gamma}}{m_{\gamma\gamma}} = \frac{1}{2} \left(\frac{\delta E_{\gamma 1}}{E_{\gamma 1}} \oplus \frac{\delta E_{\gamma 2}}{E_{\gamma 2}} \oplus \frac{\delta \alpha}{\tan(\alpha/2)} \right) \quad (3.11)$$

The photon energy resolution relies on the ECAL performance and calibration, photon reconstruction procedure and energy correction. These aspects are introduced in Section 3.3.1 and 3.3.2. The opening angle depends on the choice of the reconstructed interaction vertex. It's observed that the contribution of the angle effect to the diphoton mass resolution can be actually ignored with respect to photon energy resolution when the longitudinal distance, $|dz|$, between the chosen diphoton vertex and the Higgs interaction point is less than 1 cm. Figure 3.4 shows the resolution of the $m_{\gamma\gamma}$ distribution between the $|dz| < 1$ and $|dz| < 0.5$ cm cases has no any difference, while that between the $|dz| > 1$ and $|dz| > 2$ cm cases is sensitive and worse. Thus, to choose a reconstructed vertex matching to the Higgs interaction point with $|dz| < 1$ cm is quite important to eliminate the angle effect. The technique of diphoton vertexing is introduced in Section 3.3.3. Besides, the limited branching fraction for the diphoton channel makes photon identification become quite important. Hence, the $H \rightarrow \gamma\gamma$ analysis designs a dedicated multi-

variate technique to identify genuine photons against fake photons, introduced in Section 3.3.4. A set of cut-based diphoton selections is also donated to preliminarily clean fake diphoton pairs in each event, described in Section 3.3.5.

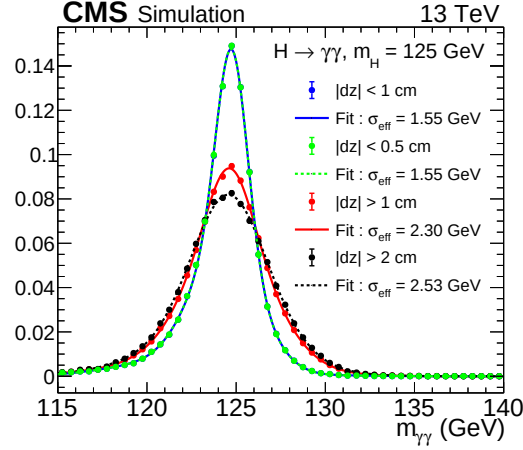


Figure 3.4: Diphoton invariant mass distribution for the $H \rightarrow \gamma\gamma$ simulated sample with $m_H = 125$ GeV. The dots with four different colors represent the simulated events with four different longitudinal-distance of the chosen diphoton vertex with respect to the Higgs interaction point. The line shapes with different colors are given by performing a fit to the dots with the corresponding colors. The effective width of these line shape distributions, defined by 68% simulated events fallen into this region with a center at 125 GeV, are provided.

On the other hand, electrons are also important physical objects, especially in the VH analysis. The electrons in the VH production come from the W and Z boson decays which also have a relatively small branching fraction, equal to the muon channel. Although both electrons and muons are charged particles with track information, the muon properties and muon detector increase the performance of the muon reconstruction and identification than the electron case. Moreover, since the photons and electrons make similar and undistinguishable responses in the ECAL, the electron reconstruction strategy shares with the photon case, except the electron track reconstruction, introduced in Section 3.3.1 and 3.3.2. The electron identification is introduced in Section 3.3.6.

3.3.1 Photon/Electron reconstruction

Photon and electron reconstructions [70] in the CMS experiment share the same algorithm due to the similar response in the ECAL detector. Differently, the electron reconstruction performs an additional algorithm to reconstruct electron tracks. The electron tracks not only provide more precise measurement from electron momenta but also identify electrons and photons during the reconstruction. However, when photons or electrons travel from interaction points to the ECAL crystals, they could interact with the inner tracker materials and produce many different situations such as bremsstrahlung and conversions, demonstrated in Figure 3.5. The connection and integration between these situations and main photons or electrons are an chal-

length in order to recover original ones. The following will introduce the e/γ object reconstruction in the CMS experiment.

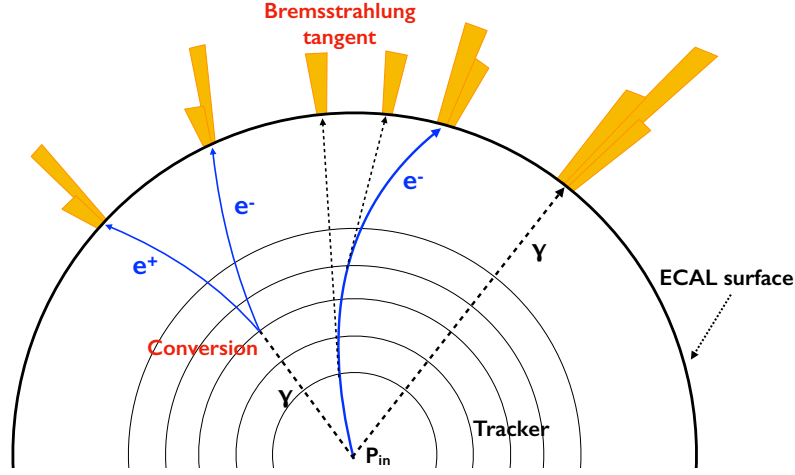


Figure 3.5: Overview of an event with converted and unconverted photons, and an electron including bremsstrahlung tangents under the inner tracker system and the ECAL surface.

EM supercluster reconstruction

Electron and photon reconstructions start from the clusters in the ECAL given by the PF cluster algorithm, described in Section 3.1.2. Since these clusters could overlap and share the same ECAL energy deposit, the shared ECAL energy is split to these clusters by some fractions. Moreover, the bremsstrahlung and photon conversion significantly happen when original electrons or photons propagate through materials from interaction points before hitting on the ECAL. Therefore, an ECAL cluster is insufficient to recover the energy of original electrons or photons. Instead, these associated ECAL clusters have to be combined to a signal cluster called *supercluster*. The supercluster reconstruction in the CMS experiment adopts a so-called *mustache* algorithm, including information from the ECAL and preshower detectors. The algorithm starts from a cluster called a seed cluster above a given threshold. The additional clusters located in mustache-like region in the $\eta - \phi$ plane with respect to this seed cluster are added to form a supercluster like Figure 3.6. The design is to collect clusters under the strong magnetic field provided by the solenoid where their energy deposits are spread along ϕ direction.

GSF track reconstruction

Besides the supercluster reconstruction, electrons in the CMS experiment are also built by track reconstruction. However, the electron energy loss mainly originates from bremsstrahlung instead of ionization in the inner tracker layers. Instead of following the normal KF algorithm with energy loss distribution as a Gaussian function, the electron track reconstruction performs using a Gaussian-sum filter (GSF) algorithm to deal with the bremsstrahlung energy loss with a non-Gaussian form.

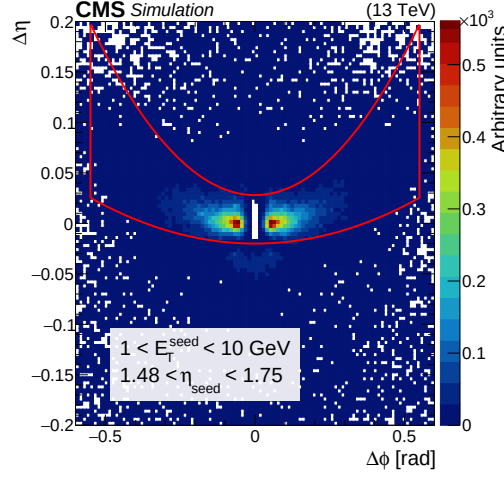


Figure 3.6: Distribution of $\Delta\eta = \eta_{\text{seed-cluster}} - \eta_{\text{cluster}}$ versus $\Delta\phi = \phi_{\text{seed-cluster}} - \phi_{\text{cluster}}$ for simulated electrons with $1 < E_T^{\text{seed}} < 10$ GeV and $1.48 < \eta_{\text{seed}} < 1.75$ [70]. The color represents the occupancy of the number of PF clusters matched with the simulation (requiring to share at least 1% of the simulated electron energy) around the seed. The red line contains approximately the set of clusters selected by the mustache algorithm. The white region at the centre of the plot represents the $\eta - \phi$ footprint of the seed cluster.

The GSF track reconstruction starts from the electron trajectory seeding by either ECAL-driven or tracker-driven. The “ECAL-driven” seeding method chooses the mustache superclusters with $E_{\text{SC},T} > 4$ GeV and $H/E_{\text{SC}} < 0.15$, where $E_{\text{SC},T}$ and E_{SC} are the transverse energy and total energy of the superclusters and H indicates the HCAL tower energy deposited in a cone size of $\Delta R < 0.15$ with the supercluster positions as the center. For each selected supercluster, its energy-weighted position, transverse energy and the magnetic field are used to extrapolate the trajectory toward the collision vertex (outside-in) with the assumption of the helical trajectory and no radiative emission. If the first two hits of a given inner-track seed are located on the pixel layers or endcap tracker and are matched to the extrapolation spatially, this seed will be considered as a electron track seed. On the other hand, the “tracker-driven” seeding approach aims at all inner tracks. If a given inner track is compatible to an ECAL cluster by some criteria using track quality and track-cluster matching variables, the inner track seed is considered as a GSF track seed.

The electron track seeds from “ECAL-driven” and “tracker-driven” methods are then grouped together as the initiation of the electron track reconstruction. For each given seed, the track finding follows Section 3.1.1 using the KF algorithm but the energy loss dominated by bremsstrahlung is modeled using a BetheHeitler distribution. Once the electron track candidates with their hits are collected, the track fitting then estimates the track parameters. The fractional energy loss of an electron, coming from not only the material but the bremsstrahlung emission, has a distribution described by the Bethe-Heitler formula which is non-Gaussian distribution. This distribution is actually unsuitable in the Kalman filter algorithm. Instead, the GSF fit uses a sum of multi-Gaussian functions as an approximation of the Bethe-Heitler formula and estimates the

parameters of the electron trajectories under the assumption of a homogeneous magnetic field.

Track-cluster Integration

Once the superclusters and GSF tracks are reconstructed, all ECAL clusters, inner tracks, mustache superclusters and GSF tracks will be linked and integrated together. First of all, the mustache superclusters must be associated with a GSF track or pass $E_{\text{SC},T} > 10$ GeV and $H/E_{\text{SC}} < 0.5 \parallel E_{\text{SC},T} > 100$ GeV. Then, conversions and bremsstrahlung radiations are reconstructed to recover the missing pieces. The conversion reconstruction uses the information from the inner tracks, conversion ECAL-seeded tracks and a 3D-constrained kinematic vertex fitter to build the conversion vertices, associated tracks and link them to the ECAL clusters. The bremsstrahlung photons are reconstructed by linking the GSF track at each tracker layer to the compatible ECAL clusters through a straight-line extrapolation as a bremsstrahlung tangent. These conversion tracks and bremsstrahlung tangents are then linked to the mustache superclusters or GSF tracks, and the ECAL clusters linked to both are added to mustache superclusters to form so called refined superclusters.

Afterward, all ECAL clusters, general tracks, refined superclusters and GSF tracks are grouped to form e/γ hypotheses. Some loose selections based on the BDT classifier are applied to the e/γ objects to kill the fake ones from jet fragmentation and hadronization. Finally, each electron candidate is defined as an e/γ object with a GSF track and a refined supercluster, while a photon candidate is that with only a refined supercluster.

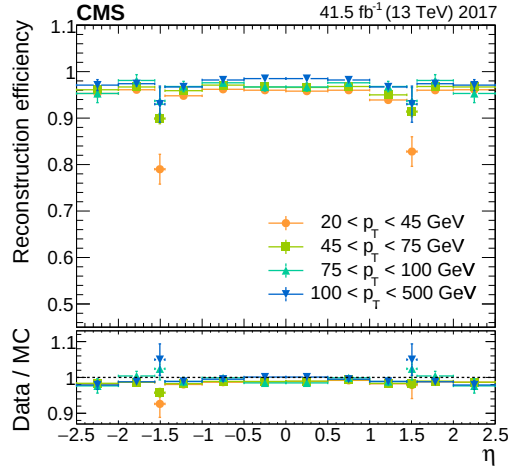


Figure 3.7: Electron reconstruction efficiency versus η in data (upper panel) and data-to-simulation efficiency ratios (lower panel) for the 2017 data taking period [70]. The vertical bars on the markers represent the combined statistical and systematic uncertainties. The region $1.44 < |\eta| < 1.57$ corresponds to the transition region between the barrel and endcap regions of ECAL and is not considered in physics analyses.

3.3.2 Energy calibration

In the energy measurement, the ECAL crystals provide quite good energy resolution for photons and electrons, and the photon and electron reconstructions in the previous section also try to optimally collect the energy deposits from these crystals. The reduction of the ECAL performance due to the crystal transparency loss during data taking could still result in serious energy loss. Several dedicated calibration approaches like crystal transparency correction by the laser monitoring system or intercalibration have tried to recover the e/γ energy [71]. Energy loss arising from the pileup effect or ungrouped clusters from conversions and bremsstrahlung still needs to be accounted for. To obtain the optimal energy resolution, especially for the diphoton mass reconstruction in the $H \rightarrow \gamma\gamma$ analyses, a regression technique based on the BDT algorithm is implemented. The regression provides a per- e/γ estimate as a function of input features to correct the reconstructed energy of the e/γ object to its true energy as close as possible.

In this regression architecture, the target y is designed by a ratio between the true energy of an e/γ object and its reconstructed energy. The input features, denoted by $\vec{x}$, include some shower shape variables ($R_9, \sigma_{i\eta i\eta}, \sigma_\eta, \sigma_\phi, H/E$) and the energy and position of the supercluster which correlate with the target. The number of primary vertices and the median energy density ρ are considered for the energy loss from the pileup effect. The regression is established by a gradient Boosted Decision Tree with a likelihood function defined by

$$\mathcal{L} = - \sum_{\text{MC } e/\gamma \text{ objects}} \ln p(y|\vec{x}) \quad (3.12)$$

where $p(y|\vec{x})$ is the estimated probability for the observed value y under given the input features $\vec{x}$ and the sum runs overall all e/γ objects in simulated samples with known true energy for each object. The probability density function is a double-sided Crystal Ball (DSCB) function composed of a Gaussian distribution with power law tails on both sides. Here, the target y and input features $\vec{x}$ is encoded by

$$p(y|\vec{x}) = \text{DSCB}(y; \mu(\vec{x}), \sigma(\vec{x}), \theta(\vec{x})) \text{ with a Gaussian structure of } \exp\left(-\frac{(y - \mu(\vec{x}))^2}{2\sigma(\vec{x})^2}\right) \quad (3.13)$$

where μ, σ and θ as a function of $\vec{x}$ is the mean and width of the Gaussian structure and other parameters in this DSCB, respectively. After regression optimization by minimizing the likelihood function through the gradient BDT, the μ and σ is the final collection estimate and energy resolution of the e/γ object with the feature $\vec{x}$.

In fact, the electron energy is sequentially corrected by three iterations corresponding to three regression estimators. The step 1 focuses on the supercluster energy correction, the step 2 provides an estimate of the supercluster energy resolution and the step 3 is to correct the combined energy between the supercluster and the electron track information. The photon energy is corrected by only the step 1 and 2. Figure 3.8 shows that the distribution of the ratio between the true energy and reconstructed one of the supercluster concentrate on 1 after the

regression correction applied.

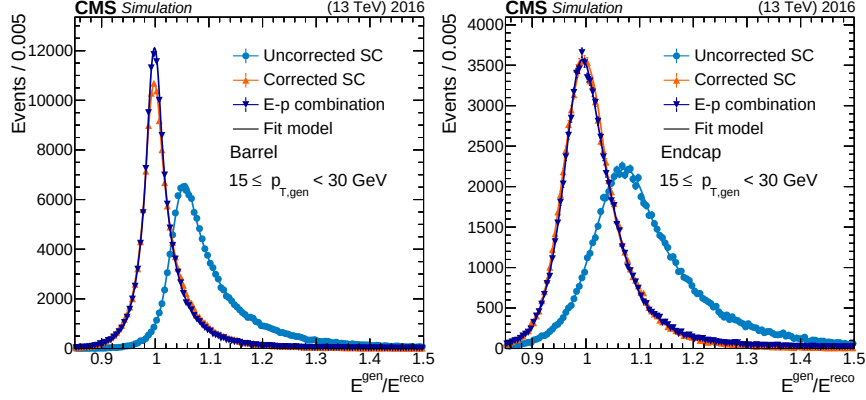


Figure 3.8: Ratio of the true to the reconstructed electron energy in the p^T range 15-30 GeV with and without regression corrections, with a DSCB function fit overlaid, in 2016 MC samples for barrel (left) and endcap (right) electrons [70]. Vertical bars on the markers represent the statistical uncertainties of the MC samples.

After applying the regression correction, it's observed the small difference still remains between data and simulation in both the electron and photon energy scales and resolutions. In particular, the energy resolution in simulation is expected to be better than that in data. The energy in data is thus re-corrected by a correction factor (scale) as a function of the time, $|\eta|$, R_9 and different gains of the ECAL photodetector amplifiers, while the simulated samples are corrected by the Gaussian smearing with $|\eta|$, R_9 bins to better describe data. These factors (scale and smear) are derived in $Z \rightarrow ee$ events. Figure 3.9 shows that the comparison of the m_{ee} distribution for $Z \rightarrow ee$ events in data and simulation gets good agreement after these residual corrections.

3.3.3 Diphoton vertex identification

Apart from photon energy resolution, diphoton vertex choice, affecting the opening angle of diphoton pairs, is another important source to affect the diphoton mass resolution. However, no charged track in the inner tracker for two photons without conversion causes the interaction vertex for the Higgs production can't be determined from those vertices introduced in Section 3.1.1. The opening angle given by wrong vertex choice reduces the mass resolution. In fact, most of Higgs productions are associated with other charged particles like charged leptons or jets with relatively large energy scale. This reason enables the analysis to directly assign the primary vertex, which contains the maximal scalar sum of the transverse momenta of the associated tracks, as the Higgs interaction vertex. Nevertheless, the ggH production, the most dominant Higgs production mechanism, which has no any energetic charged track associated production is not able to confirm the primary vertex as the diphoton vertex. To correctly identify the diphoton vertex from the vertices, the $H \rightarrow \gamma\gamma$ analysis provides a dedicated approach called diphoton vertex identification derived by a gradient Boost Decision Tree classifier.

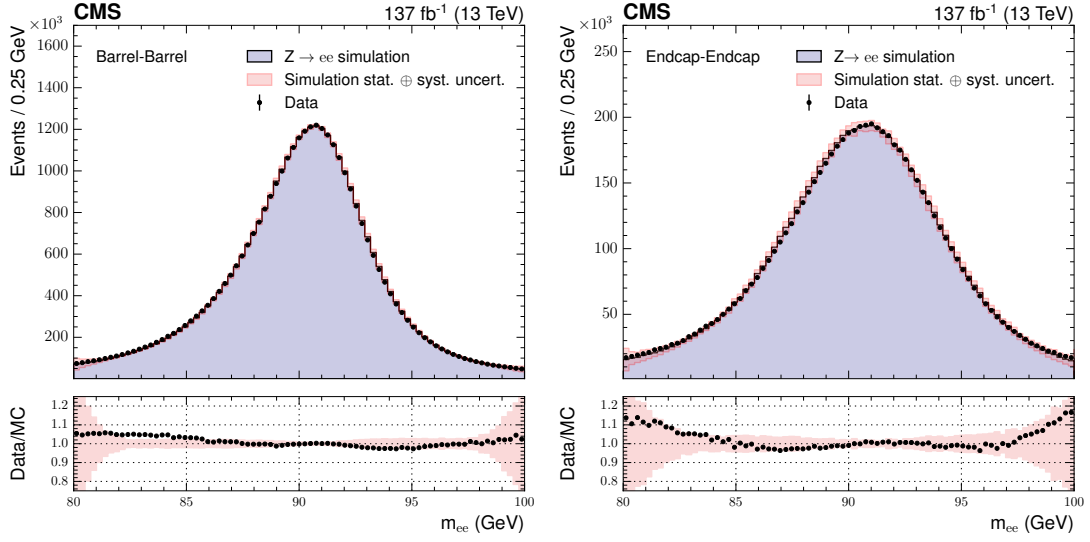


Figure 3.9: Comparison of the dielectron invariant mass spectra in data (black points) and simulation (blue histogram), after applying energy scale corrections to data and energy smearing to the simulation, for $Z \rightarrow ee$ events with electrons reconstructed as photons. The statistical and systematic uncertainty on the simulation is shown by the pink band. The comparison is shown for events where both electrons are reconstructed in the ECAL barrel (left), and both in the ECAL endcaps (right). The lower panels show the ratio of the data to the MC simulation in black points, with the uncertainty on the ratio represented by the pink band. The full data set collected in 2016-2018 and the corresponding simulation are shown [72].

This diphoton vertex ID relies on the information from the kinematic properties of the diphoton system and its correlations with the recoiling tracks, and gives scores with a range of $(-1, 1)$ to all vertices for each diphoton pair in each event. The vertex with the largest score is assigned as the diphoton vertex for this diphoton pair. This design arises from that the vertex with large track recoil expected as the interaction point of the Higgs boson, i.e. the blue point described in Figure 3.10. There are three variables designed as input features to describe the kinematics of the diphoton and recoiling tracks associated with a given vertex:

- $\sum_i |\vec{p}_T^i|^2$
- $-\sum_i \left(\vec{p}_T^i \frac{\vec{p}_T^{\gamma\gamma}}{|\vec{p}_T^{\gamma\gamma}|} \right)$
- $\left(|\sum_i \vec{p}_T^i| - p_T^{\gamma\gamma} \right) / \left(|\sum_i \vec{p}_T^i| + p_T^{\gamma\gamma} \right)$

The first is the sum of the squared transverse momenta of the charged particle tracks associated with the vertex. Later two variables describe the vector and scalar balance of the diphoton system and the charged tracks associated for the vertex.

Moreover, the case with at least one photon converted provides an additional information for the diphoton vertex ID. The conversion photons carrying electron tracks enable the extrapolation of the z position of the diphoton vertex, denoted by z_{PV}^{conv} . Therefore, an additional input

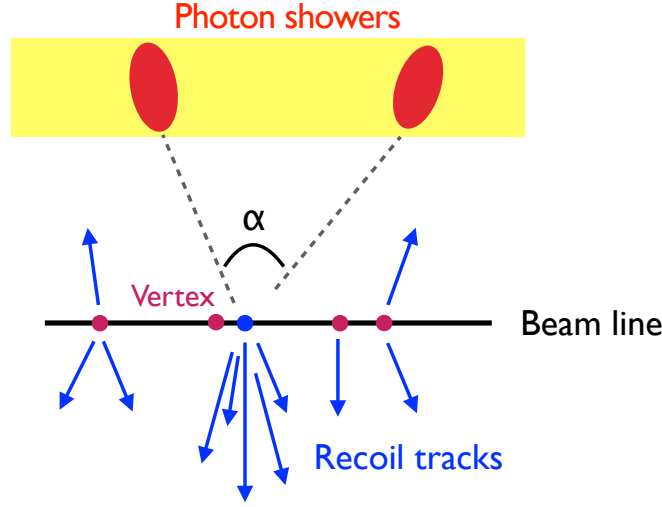


Figure 3.10: Diagram of the diphoton system in the CMS detector. The vertex with large track recoil is expected as the interaction point of the Higgs boson like the blue point with the corresponding opening angle of two photons (α).

feature is given by $|z - z_{\text{PV}}^{\text{conv}}|/\sigma_{\text{conv}}$ as the pull between the z position of the tested vertex and that of the extrapolated primary vertex with the corresponding uncertainty, σ_{conv} . There are two method computing $z_{\text{PV}}^{\text{conv}}$. One solely uses the direction of conversion momentum given by the conversion tracks refitted with the conversion vertex constraint, given by

$$z_{\text{PV CO}}^{\text{conv}} = z_{\text{conv}} - r_{\text{conv}} \cot(\alpha_{\text{conv}}) \quad (3.14)$$

where z_{conv} is the z position of the conversion vertex, r_{conv} is the distance between the conversion vertex and the beam line and α_{conv} is the angle between the conversion momentum and the beam line. The other method combines the position information from the conversion vertex and supercluster, given by

$$z_{\text{PV SC}}^{\text{conv}} = z_{\text{conv}} - r_{\text{conv}} \frac{z_{\text{SC}} - z_{\text{conv}}}{r_{\text{SC}} - r_{\text{conv}}} \quad (3.15)$$

where z_{SC} is the z position of the supercluster and r_{SC} is the distance between the supercluster and beam line. For the training, samples simulating four relevant Higgs boson productions with $H \rightarrow \gamma\gamma$ and $m_H = 125$ GeV, i.e. ggH , VBF , VH and $t\bar{t}H$, are considered for overall $H \rightarrow \gamma\gamma$ analysis. The vertex matching with the Higgs interaction point ($|dz| < 0.1$) is regarded as a signal class, while a background class randomly chooses one of the vertices except the signal vertex. Once the diphoton vertex ID is trained and applied for choosing the diphoton vertex, each photon momentum is then recalculated according to the chosen vertex and its original supercluster position.

The second vertex-related classifier, called as the vertex probability BDT, is also prepared to estimate the probability that the vertex, chosen by the diphoton vertex ID, is within 1 cm

of the Higgs interaction point in terms of the longitudinal distance. The probability provides the $H \rightarrow \gamma\gamma$ analysis an estimate to either split events into several analysis categories with different diphoton mass resolutions or drop events with the worst vertex choice for benefiting sensitivity. This classifier is trained on the same simulated samples in the vertex ID training and uses the most relevant features such as the p_T of the diphoton system, number of vertices, three highest vertex ID scores from the vertices and the dz between them. The training signal and background are defined by events with the chosen vertex within and outside 1 cm of the Higgs interaction point, respectively. The probability is then evaluated with a fourth-order polynomial function of the vertex probability BDT score, where the function parameters are determined by performing a fit to the fraction of $|z_{\text{chosen vtx.}} - z_{\text{Higgs point}}| < 1 \text{ cm}$ versus this score in simulated samples. Figure 3.11 shows that the performance of the diphoton vertex ID validated by an unbiased set of the $Z \rightarrow \mu\mu$ events and of the vertex probability BDT by $H \rightarrow \gamma\gamma$ simulated samples have good agreement between data and simulation, and between the probability and the fraction of $|z_{\text{selected}} - z_{\text{true}}| < 1 \text{ cm}$ in terms of the fraction of $|z_{\text{selected}} - z_{\text{true}}| < 1 \text{ cm}$ versus $p_T^{\gamma\gamma}$, respectively.

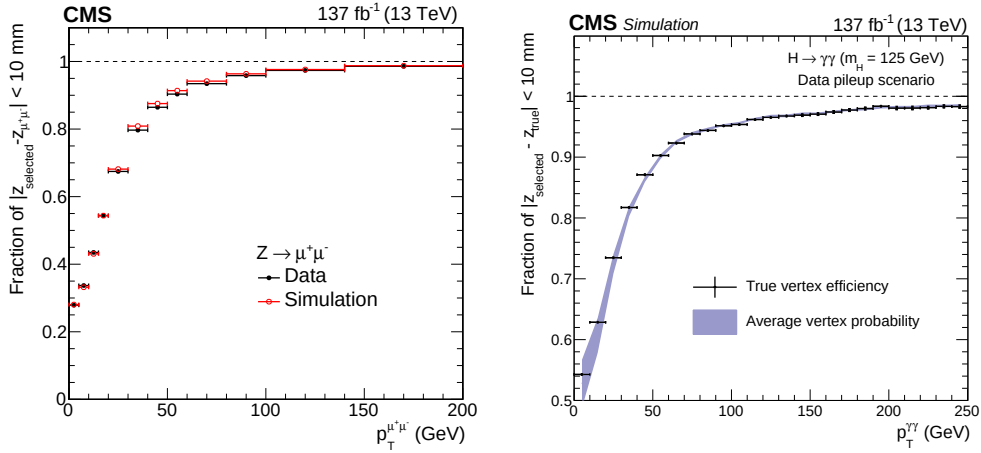


Figure 3.11: The left is the fraction of $|z_{\text{selected}} - z_{\mu\mu}| < 1 \text{ cm}$ versus $p_T^{\mu\mu}$ in the $Z \rightarrow \mu\mu$ events combining full data set and simulation in 2016-2018. The vertices are refitted by overall charged tracks with without two muon ones to mimic the diphoton system. The chosen vertex is selected from these refitted vertices through the diphoton vertex ID and the dimuon vertex is given by the original vertex matching with two muon tracks. The right is the fraction of $|z_{\text{selected}} - z_{\text{true}}| < 1 \text{ cm}$ versus $p_T^{\gamma\gamma}$ in the $H \rightarrow \gamma\gamma$ simulated events, where “true” points the Higgs interaction point. The purple area is the average vertex probability given by the vertex probability BDT. Overall simulated events are weighted by the pileup reweighting factors.

3.3.4 Photon identification

In the energy scale of the Higgs boson physics, light neutral hadrons like π_0 and η from jet fragmentation become boosted and their diphoton final states almost collinearly propagate on the ECAL surface. These two clusters close to each other could let the photon reconstruction

misidentify to a single cluster from a single photon. This is a main background source in the $H \rightarrow \gamma\gamma$ channel. However, fake photons in jets gives relatively low energy, widely spread EM showers in the ECAL crystals and more particle activities near them with respect to genuine photons, illustrated in Figure 3.12. With these reasons, some shower shape, isolation variables to describe the EM shower spread and particle activities near photons are designed for the photon identification.

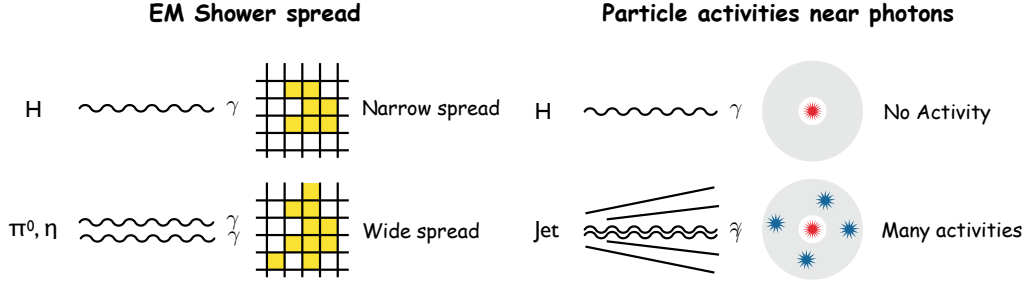


Figure 3.12: EM Shower spread and particle activities near photons for the photons from the Higgs boson are different from that from the π_0 and η . The EM shower for the collinear photons from the π_0 and η decays is wider than that from the Higgs boson decay. On the other hand, there are also a lot of particle activities near the photons from the π_0 and η decays.

In the $H \rightarrow \gamma\gamma$ analysis, a photon identification is performed using a multivariate technique called *photon ID BDT* which provides a powerful classification for the genuine photons against those fake photons. The training uses some designed shower shape and isolation variables, described in Table 3.2, and energy, pseudorapidity η of photon superclusters with significant separation power. The median energy density ρ is also considered as an input feature against the pileup effect.

The photon ID BDT is trained using γ + jets simulated samples where prompt photons matching with true photons are considered as the signal and non-prompt photons from jets as the background under the gradient Boosted Decision Tree (BDT) algorithm. In addition, both prompt and non-prompt photons are required to pass a set of the diphoton selections, described in Section 3.3.5. To eliminate the dependence of the photon kinematics during the BDT training, the p_T and supercluster η distributions of the prompt photon are two-dimension weighted to match those of the non-prompt photon.

However, the significant discrepancy between data and simulation is observed in the photon shower shape and isolation variables. This disagreement can further propagate to the final photon ID BDT distribution and make a significant systematic uncertainty for the $H \rightarrow \gamma\gamma$ analysis. A chained quantile regression (CQR) [73] method is thus performed correcting the simulated shower shape and isolation variables to agree with data. The correction is derived by an unbiased set of electrons from $Z \rightarrow ee$ events. The CQR contains a set of BDTs which predict the cumulative distribution function (CDF) of a given shower shape variable. This prediction is conditional upon three electron kinematic variables, p_T , $|\eta|$, ϕ and the median energy density ρ . The CDFs from data and simulated events given by this method are then used to derive a

Variable	Definition
R_9	The ratio between the energy sum of the 3 by 3 crystals around the most energetic crystal in the supercluster and the energy of the supercluster is expressed by $E_{3 \times 3} / E_{SC}$.
$\sigma_{i\eta i\eta}$	The log-weighted distribution of crystal energies in η direction computed in the 5×5 crystal matrix around the most energetic crystal in the supercluster and rescaled to units of crystal size. The mathematical expression is given below: $\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5 \times 5} w_i (\eta_i - \bar{\eta}_{5 \times 5})^2}{\sum_i^{5 \times 5} w_i}} \quad (3.16)$ where η_i is the pseudorapidity of the i-th crystal, $\bar{\eta}_{5 \times 5}$ is the pseudorapidity mean position in the 5×5 crystal matrix; w_i is a weighted factor defined by $w_i = \max(0, 4.7 + \ln(E_i / E_{5 \times 5}))$. The weighted factor is considered as a selection requirement to reject the ECAL noise, that is, crystal energy deposits must be at least 0.9% of $E_{5 \times 5}$. The distribution of $\sigma_{i\eta i\eta}$ is expected to be narrow for single-photon showers, while one for two-photon showers from neutral meson decays is broad.
$\sigma_{\eta(\phi)}$	The logarithmic energy weighted standard deviation of single crystal $\eta(\phi)$ within the supercluster.
H/E	The ratio of the energy deposited in the HCAL in a cone of radius $\Delta R = 0.15$ around the supercluster direction and the energy of the supercluster.
Preshower σ_{RR}	The standard deviation of the shower spread in the x and y planes of the preshower (only defined in the endcap).
$\mathcal{I}_\gamma, \mathcal{I}_{\text{trk}}, \mathcal{I}_{\text{ch. had.}}$	Sum of the transverse momenta of all photons, tracks and charged hadrons in a cone of radius $R = 0.3$ around the photon candidate direction, excluding the objects in an inner cone of size $R = 0.04$.

Table 3.2: Definition of the shower shape and isolation variables used in the photon ID BDT training.

correction factor to be applied to simulated electrons. The correction factors morph the CDF of the simulated shower shape onto the one in data through quantile matching, demonstrated in Figure 3.13 (left).

Moreover, the CQR adopts a sequential chain structure of the shower shape variables with a given order. In this chain, the CDF of the first shower shape variable is predicted solely by three electron kinematic variables and ρ . The corrected shower shape variables by previous procedures are considered as input features of later predictions, demonstrated in Figure 3.13 (right). The order of the shape shape variables is optimized to minimize the discrepancy of the final photon ID BDT distribution between data and simulation.

On the other hand, the isolation variables are not included in the chain because their correlation to those shower shape variables is negligible. Moreover, since there is a p_T threshold applied

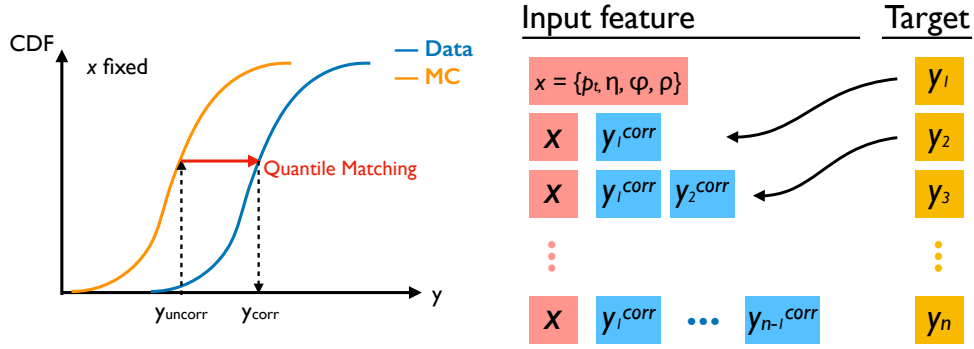


Figure 3.13: The left is the CDF of the given shower shape variable, y , in the data (blue curve) and simulation (orange curve) with three electron kinematic variables and ρ fixed to given values. The morphing from simulation to data is realized by matching the quantile given by some uncorrected shower shape value, y_{uncorr} , to data with the same quantile where the corresponding shower shape value, y_{corr} , is considered as the correlation of the y_{uncorr} . The architecture of the chained quantile regression with a sequential chain structure is shown in the right figure. The CDF prediction of the first shower shape variable, y_1 , is trained by only x , i.e. those three electron kinematic variables and ρ . The training of the CDF prediction of later shower shape variables considers the previous corrected shower shape variables as input features.

during the calculation of the isolation variables, a spike in the region between zero to the beginning of the tail is observed in the isolation variable distributions. Therefore, the isolation variable correlation is divided into two cases. One is the spike region corrected by a constant of the CDF, and the other is the tail part corrected following the same correction technique used in the shower shape correction. Figure 3.14 (left) shows better agreement between data and simulation in the $Z \rightarrow ee$ events after the corrections. The final photon ID BDT distribution is shown in Figure 3.14 (right).

3.3.5 Diphoton identification

Once the photon energy correction is applied and the diphoton vertex is reassigned, diphoton candidates are defined by a series of cut-based selections to remove the diphoton pairs with at least one fake photon originating from neutral hadrons like π_0 and η in jets. Following the similar rejection strategy to the design of the photon ID BDT, the selections contain several kinematic properties, shower shape and isolation variables of a supercluster shown in Table 3.2. To compromise the signal efficiency and background reduction and avoid biasing diphoton kinematics to a specific phase space, the selections mimic the requirements in the diphoton high-level trigger, shown in Section 2.3.2, that is the least limitation event filter, but their thresholds are raised slightly. The diphoton selections not only play the role of trigger safety selections to remove events accidentally mis-triggered by the diphoton HLT but also simultaneously kill those fake diphoton pairs in each event. Besides, a very loose cut in the photon ID BDT is also added in the diphoton selections to drop the most background-like photons. Eventually, the

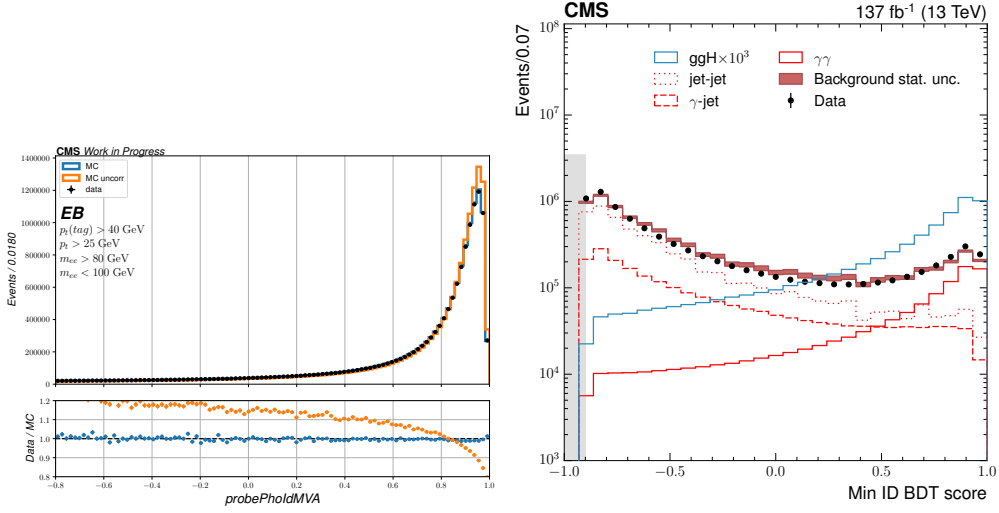


Figure 3.14: The left is the comparison of the photon ID BDT distribution in data and simulated events with (blue) and without (orange) the shower shape and isolation correction under the $Z \rightarrow ee$ events. The selections are shown in this figure. The bottom panel is the ratio of data to corrected and uncorrected simulations. The right is the comparison of the photon ID BDT distribution in data and simulated events collected from 2016-2018 without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV. The non-resonance $\gamma\gamma$, γ + jets and jet-jet (QCD) simulated background components are expressed by three overlapping histograms with one red solid and two dashed line with different styles. Total background and its statistic uncertainties combining these background components are described with a brown solid line and box, respectively. The ggH process with $m_H = 125$ GeV is shown with a blue solid line, rescaled 1000 times than the expected yield for visibility [72].

diphoton pairs with the largest scalar sum of transverse momenta of two photons in each event is treated as the diphoton candidate. The selections give a signal efficiency of approximately 44% for the inclusive VH production in the $H \rightarrow \gamma\gamma$ channel. The details of the diphoton selection criteria are shown in the following list :

- $p_T^{\gamma_1} > 35$ GeV, $p_T^{\gamma_2} > 25$ GeV where γ_1 and γ_2 are the leading and subleading photons, respectively.
- Pseudorapidity of the photons $|\eta| < 2.5$ without the barrel-endcap gaps in $1.44 < |\eta| < 1.57$. The geometrical acceptance requirement is applied to both the supercluster position and the photon momentum after the diphoton vertex assignment.
- Electron veto, which rejects the photon candidate if its supercluster is matched to a track.
- Shower shape variables for removing non-single isolated electromagnetic showers from neutral mesons ($R_9, \sigma_{\text{in}ij}$) and for rejecting hadrons (H/E); isolation variables like $\mathcal{I}_\gamma$ and $\mathcal{I}_{\text{trk}}$ for loose selections. The complete criteria are shown in Table 3.3.
- Either $R_9 > 0.8$ and $\mathcal{I}_{\text{ch. had.}} < 20$ GeV or $\mathcal{I}_{\text{ch. had.}}/p_T^\gamma < 0.3$.

- Photon ID BDT score > -0.9 for both photons.

	R_9	H/E	$\sigma_{i\eta i\eta}$	$\mathcal{I}_\gamma$ (GeV)	$\mathcal{I}_{\text{trk.}}$ (GeV)
Barrel	> 0.85	< 0.08	-	-	-
	$[0.50, 0.85]$	< 0.08	< 0.015	< 4.0	< 6.0
Endcaps	> 0.90	< 0.08	-	-	-
	$[0.80, 0.90]$	< 0.08	< 0.035	< 4.0	< 6.0

Table 3.3: Requirements of the shower shape and isolation variables in the diphoton selections.

3.3.6 Electron identification

Electron candidates in the VH analysis are selected by $p_T > 10$ GeV, $|\eta| < 2.4$ without the barrel-endcap gaps in $1.44 < |\eta| < 1.57$ and require to pass a dedicated BDT-based identification with a working point of 90% efficiency, given by CMS collaboration (CMS EGM POG). The goal of the electron ID BDT is to reject electrons' background arising from photon conversions, hadrons misidentified as electrons, and semileptonic decay of b or c quarks. Therefore, for the electron ID BDT training, apart from some shower-shape variables of electrons which are similar to Table 3.2 given by photons, there are more variables related to track quality and track-cluster matching for electron tracks, described in Table 3.4. The fractional difference between the track momentum at the innermost tracker layer and at the outermost tracker layer, f_{brem} , is also included. This BDT training uses electron candidates in DY + jets simulated events with $E_T > 5$ GeV and $|\eta| < 2.5$ under a gradient-Boost Decision Tree. Signal electrons are defined as reconstructed electrons matching true electrons from the Z boson decay within $\Delta R = 0.1$ and background electrons are assigned to reconstructed electrons from non-prompt electrons given by hard jets. The training is split at $E_T = 10$ GeV for different background components and at $|\eta| = 0.8, 1.4$ for the different inner tracker arrangement.

3.4 Jet

In the VH production, quarks or gluons can be produced via the W and Z boson decays or from initial state radiation. These quarks or gluons can emit soft gluons that can further produce quark and anti-quark pairs or gluons, called fragmentation during their travelling. Due to the colour confinement in the QCD, these produced partons from the fragmentation form colourless particles, hadrons. Some unstable hadrons immediately decay to stable particles like other hadrons, photons or leptons. At the energy scale of the Higgs physics, these stable particles are boosted to the similar direction of the initial quark or gluon within some cone size. This collimated spray structure of the stable particles is called as *jet*.

Variable	Definition
$ 1/E - 1/p $	Energy compatibility between the supercluster and track of electrons. The E is the supercluster energy and p is the track momentum at the point of closest approach to the primary vertex.
$ \Delta\eta_{\text{in}}^{\text{seed}} $	$ \eta_{\text{seed}} - \eta_{\text{track}} $ where η_{seed} is the η position of the seed cluster in the supercluster and η_{track} is the track η extrapolated from the innermost track position.
$ \Delta\phi_{\text{in}} $	$ \phi_{\text{SC}} - \phi_{\text{track}} $ where ϕ_{SC} is the ϕ energy-weighted position of the supercluster and ϕ_{track} is the track ϕ extrapolated from the innermost track position.
f_{brem}	The fractional difference between the track momentum at the innermost tracker layer and at the outermost tracker layer.
$\mathcal{I}_{\gamma}$ $\mathcal{I}_{\text{ch. had.}}$ $\mathcal{I}_{\text{neutr. had.}}$	Scalar sum of the transverse momenta of all photons, charged and neutral hadrons in a cone of radius $R = 0.3$ around the electron candidate direction, excluding the objects in an inner cone of size $R = 0.04$.

Table 3.4: Definition of the track quality and track-cluster matching variables and isolation variables used in the electron ID BDT training.

3.4.1 Jet reconstruction

In the CMS experiment, jets are reconstructed by clustering overall Particle-flow candidates. For the VH analysis, the clustering algorithm is performed using an anti-kt algorithm [74] with $R = 0.4$ (AK4) in the fastjet package [75]. This algorithm aims at non-boosted jets expected from the associated VH production under the Standard Model. It defines *entity* as either particles or pseudo-jets², and the distances between entities i and j and between entity i and the beam line (B) donated by the following

$$d_{ij} = \min \left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2} \right) \times \frac{\Delta_{ij}^2}{(R = 0.4)^2} \quad (3.17)$$

$$d_{iB} = \frac{1}{p_{Ti}^2} \quad (3.18)$$

which $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and p_{Ti} , y_i and ϕ_i are the transverse momentum, rapidity and azimuthal angle of the i -th entity. The method first creates a set $\{d_{ij}, d_{iB}\}$ from all particle-flow candidates. If the d_{ij} is the minimum, i and j particles are merged as a pseudo-jet in a list of entities; if the d_{iB} is the minimum, the i -th entity will be considered as a final jet and is removed from the list. This procedure is repeated until no entity is left in the list.

However, a part of particle-flow candidates originating from pile-up (PU) vertices leads to incorrect energy reconstruction of jets. To eliminate the pileup effect, a PU mitigation called charged-hadron subtraction (CHS) algorithm [50] is applied to the particle-flow candidates

²Intermediate jets, not final ones during the clustering.

before jet clustering. The CHS algorithm takes the track information to remove the charged-particle candidates which are used for the reconstruction of the PU vertices. In the $H \rightarrow \gamma\gamma$ analysis, the diphoton vertex identification already redefine the diphoton vertex, other vertices beyond the diphoton vertex are regarded as the PU vertices. For those which are not associated with any PU vertex and neutral particles are kept. Finally, the jets called AK4CHS jets are reconstructed and used in the VH analysis. For the AK4CHS jets, the efficiency is larger than 95% in the overall detector region up to $|\eta| < 5$ and the purity is about 70% and 18% at 50 interactions for the regions of $|\eta| < 2.5$ and $|\eta| > 2.5$, respectively.

3.4.2 Jet identification

In the VH analysis, jet candidates are defined by $p_T > 20$ GeV and $|\eta| < 2.4$ in order to remove jets from soft gluon radiation and to reduce the pileup effect. In addition, the dedicated jet identifications for the rejection of noise jets and pileup jets are applied. A dedicated ML-based identification is also used in the VH analysis to veto events with jets from b quark fragmentation and hadronization. The brief introduction for these jet identifications is in the following sections.

Noise jet rejection

Jets are clustered by particle-flow candidates reconstructed through signals from certain sub-detectors like the ECAL and HCAL. Either some noises originating from the ECAL and HCAL or jet reconstruction failure could cause that noise jets are accidentally reconstructed. To reject these noise jets after detector signal processing and jet clustering, a set of jet ID criteria based on constituent energy fractions and multiplicities, shown in Table 3.5, is applied to jets. The jet ID not only keeps more than 98–99 % of physical jets for all η regions but achieves a fraction of the noise jet rejection of more than 98% for $|\eta| < 3$.

Content	Requirement
Neutral hadron energy fraction	< 0.9
Neutral EM energy fraction	< 0.9
Number of neutral particles	–
Charged hadron energy fraction	> 0
Charged multiplicity	> 0
Number of constituents	> 1

Table 3.5: Content of jet identification criteria.

Pileup jet rejection

Pile-up jets are reconstructed by particles from either a pileup vertex or stochastically multiple pileup vertices. These jets not joining the VH production should be removed from events. The CHS algorithm already removes most of the charged PF candidates originating from those

pileup vertices, reducing roughly 1/3 of the original p_T of pileup jets. Hence, the reconstructed pile-up jets can be simply removed by their p_T . To maximize the rejection of the pileup jets, a Boosted Decision Tree classifier, called as pileup jet ID [50], is performed in the VH analysis.

It's observed that the majority of tracks within pileup jets do not originate the primary vertex³. Moreover, the shape of pileup jets constructed by particles from multiple pileup vertices is boarder than that of jets from the primary vertex. The pileup jet classifier, based on these observations, adopts track-based variables and jet shape variables as the input feature for training. Table 3.6 shows mainly input features. The PV $\sum p_T$ fraction is small for pileup jets and available for only charged particles within the tracker acceptance. The $\langle \Delta R^2 \rangle$ and f_{ringX} define the jet shape which is expected to tend to larger for pileup jets. Since around 92% pileup jets tend to happen at $p_T < 50$ GeV, the classifier is designed to act on jets with this scale.

Variable	Definition
PV $\sum p_T$ fraction = (β)	Fraction of p_T from charged particles associated with the primary vertex, defined as $\sum_{i \in \text{PV}} p_{T,i} / \sum_i p_{T,i}$, where i iterates over all charged PF particles in a given jet.
$\langle \Delta R^2 \rangle$	Square distance from the jet axis weighted by p_T^2 average of jet constituents: $\sum_i \Delta R^2 p_{T,i}^2 / \sum_i p_{T,i}^2$
f_{ringX}	Fraction of p_T of the constituents ($\sum_i p_{T,i} / p_T^{\text{jet}}$) in the region $R_i < \Delta R < R_{i+1}$ around the jet axis, where $R_i = 0, 0.1, 0.2, 0.3$ for $X = 1, 2, 3, 4$.

Table 3.6: Definition of the main track-based and jet shape variables used in the pileup jet ID training.

The distributions of the β and pileup jet ID score, shown in Figure 3.15, give robust separation power between pileup jets and non-pileup jets with $30 < p_T^{\text{jet}} < 50$ GeV and within the tracker acceptance. For the VH analysis, a tight working point of the pileup jet ID, described in Table 3.7, is applied to the jets. Although this working point gives an relatively low efficiency of 80%, the purity can achieve more than 98%.

$ \eta < 2.5$	$20 \text{ GeV} < p_T^{\text{jet}} < 30 \text{ GeV}$	$30 \text{ GeV} < p_T^{\text{jet}} < 50 \text{ GeV}$
Pileup jet ID	> 0.69	> 0.86

Table 3.7: Content of the tight pileup jet identification criteria.

b quark jet identification

Since the VH production with the W/Z boson leptonic decay has no any b quark final state, to exclude b -jets originating from the b quark fragmentation and hadronization helps the analysis reduce the contaminations from the top-pair and ttH productions with at least one b -jet. In principle, the b quark can decay into either an u quark or c quark via the weak interaction.

³Here is the diphoton vertex.

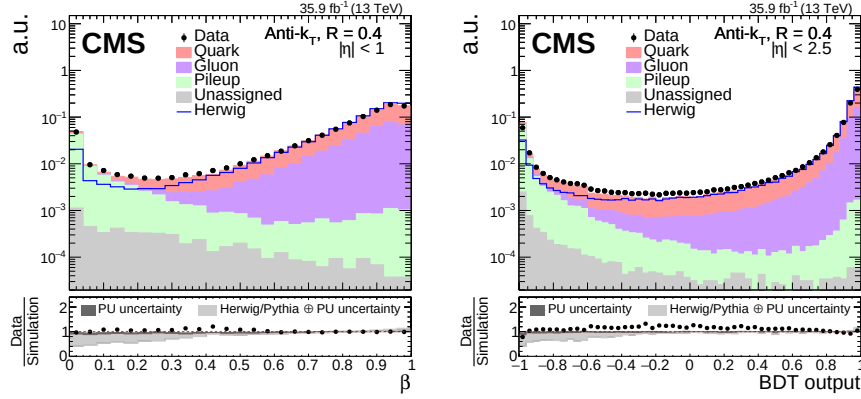


Figure 3.15: Data-to-simulation comparison of the β (left) and PU jet ID BDT output (right) for AK4 jets with $30 < p_T < 50$ GeV within the tracker acceptance. Black markers represent the data while the colored areas are Z+jets simulation events. The simulation sample is split into jets originating from quarks (red), gluons (purple), PU (green), and jets that could not be assigned (gray). The distributions are normalized to unity. The shape of a sample showered with HERWIG++ is superimposed and included in the total uncertainty band in the ratio panel (light gray). Also included in the ratio panel is the PU rate uncertainty (dark gray) [50].

However, both decay processes are suppressed by the elements V_{ub} and V_{cb} in the CKM matrix. According to this reason and the Lorentz time-dilation, hadrons containing b quarks in the jets originating from the b quark hadronization has relatively longer lifetime $\mathcal{O}(10^{-12})$ s, causing that these b hadrons can propagate a distance of a few mm to one cm. With respect to the primary vertex, the displacement creates a secondary vertex (SV) reconstructed by displaced tracks from these b hadron decays, illustrated by Figure 3.16.

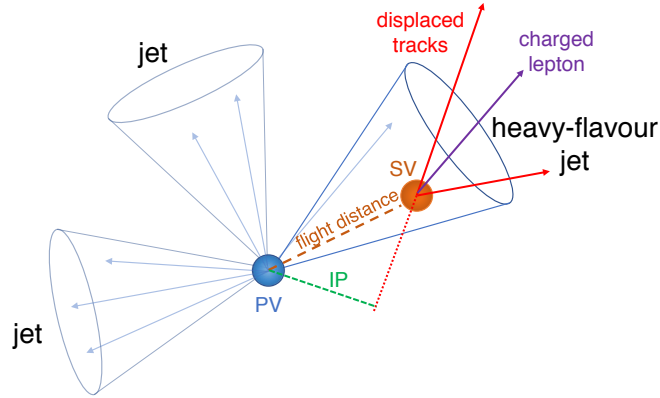


Figure 3.16: Diagram of a heavy-flavour jet with a secondary vertex (SV) from the decay of a b or c hadron resulting in charged-particle tracks (including possibly a soft lepton) that are displaced with respect to the primary interaction vertex (PV), and hence with a large impact parameter (IP) value [76].

Besides, the b quarks have a larger mass and harder fragmentation with respect to light quarks

and gluons. The decay products from the b hadrons thus have relatively larger p_T and angle between them and the original b hadrons with respect to other jet constituents. Moreover, muons or electrons are produced in the decay chain of the approximate 20% b hadrons. These features give quite powerful b -jet identification.

In the CMS experiment, the ML-based combined secondary vertex (CSV), based on these features, had been successfully derived for the b -jet identification in the CMS Run I analysis. To optimize the performance of the b -jet identification in the Run II analysis with higher pileup environment, a deep neural network (DNN) is performed to improve the CSV method, called as deepCSV [76]. The deepCSV uses many track-based variables such as the impact parameters with respect to the PV, SV and jet axis related variables like the displacement, secondary vertex mass. The output score and performance of the deepCSV for the b -jet discriminator is shown in Figure 3.17. The output deepCSV variable in the VH analysis will be used in the training of the VH multivariate technique for the b -jet veto.

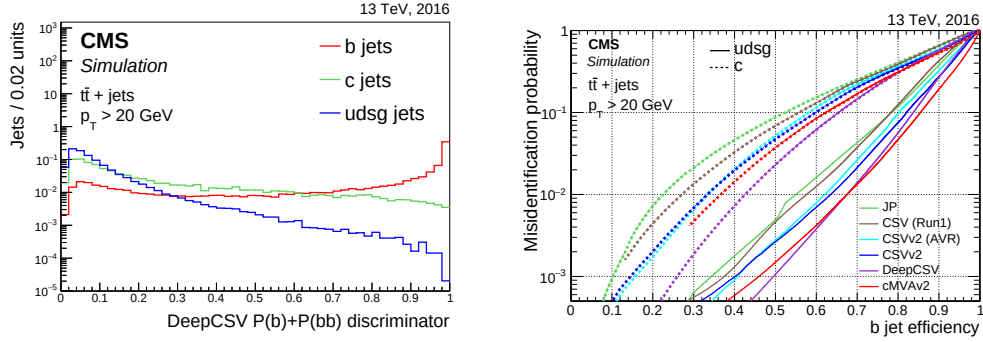


Figure 3.17: Distribution of the DeepCSV ($P(b)+P(bb)$) discriminator values for the jets from different flavours in the $t\bar{t}$ events. The distributions are normalized to unit area. The right is the misidentification probability for c and light-flavour jets versus b jet identification efficiency for various b -tagging algorithms applied to jets in the $t\bar{t}$ events [76].

3.5 Missing energy

The Particle-flow algorithm in the CMS experiment tries to reconstruct all physical objects in each event as complete as possible. However, some particles like neutrinos or other weakly interacting objects leave no signal in the detector. In particular, the neutrinos are an important final states which can extract events originating from the VH production with $Z \rightarrow \nu\nu$. The information loss of the neutrinos make this VH process undetectable. Although the parton density function of protons leads to unknown total-momentum at the z direction for proton-proton collisions, the sum of momentum in the transverse plane is zero according to momentum conservation. For this reason, these missing objects can produce a momentum imbalance in the transverse plane which needs an offset transverse momentum for balance. This offset transverse momentum is called as *missing transverse energy* from the missing objects, giving a feature of the neutrinos in the VH analysis.

By definition, the missing transverse energy, p_T^{miss} , is calculated by an negative vectorial sum of all reconstructed particle-flow candidates, namely, all observed final-state particles

$$\vec{p}_T^{\text{miss, raw}} = - \sum_{i=1}^{N_{\text{particles}}} \vec{p}_{T,i} \quad (3.19)$$

where $\vec{p}_{T,i}$ is the transverse momentum of the i -th reconstructed particle-flow candidate. However, the minimum energy thresholds in the calorimeters, p_T thresholds and inefficiencies in the track reconstruction could leads to an inaccurate estimation of the $\vec{p}_T^{\text{miss}}$. Namely, the $\vec{p}_T^{\text{miss}}$ is particularly sensitive to the reconstruction performance of the particle-flow candidates. To obtain more precise $\vec{p}_T^{\text{miss}}$, the correction of these physics objects also have to propagate to correct the $\vec{p}_T^{\text{miss}}$. The precise measurement of photons, electrons and muons makes their energy correction negligible for the $\vec{p}_T^{\text{miss}}$. Instead, relatively large corrections for jet energy is dominant in the $\vec{p}_T^{\text{miss}}$ correction. In the CMS experiment, a Type-1 correction [77] is thus applied to $\vec{p}_T^{\text{miss, raw}}$ by

$$\vec{p}_T^{\text{miss}} = \vec{p}_T^{\text{miss, raw}} - \sum_{\text{jets}} (\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}^{\text{raw}}) \quad (3.20)$$

where $\vec{p}_T^{\text{miss, raw}}$ is the uncorrected $\vec{p}_T^{\text{miss}}$ and $\vec{p}_{T,\text{jet}}^{\text{corr}}$, $\vec{p}_{T,\text{jet}}^{\text{raw}}$ are the transverse momentum of the corrected and uncorrected jets. To avoid soft/pileup jets and the overlapping between the jets and electrons, photons and muons, the jets with $p_T < 15$ GeV and EM energy fraction greater than 90% are removed. The jet energy smearing for simulated events more close to data is also applied for the $\vec{p}_T^{\text{miss}}$ correction. Figure 3.18 shows good agreement between data and simulation in the p_T^{miss} distribution under the $Z \rightarrow \mu\mu$ events.

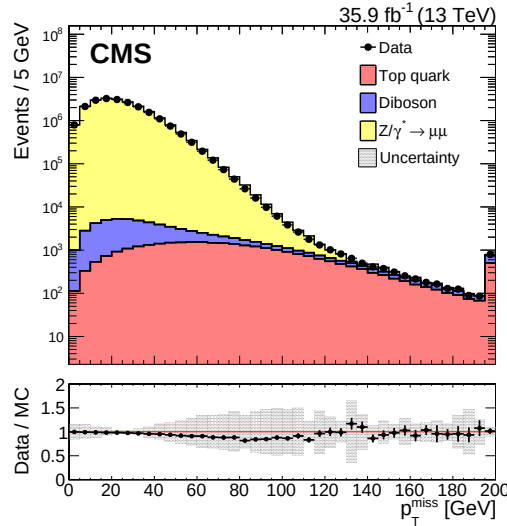


Figure 3.18: Upper panel: Distribution of $\vec{p}_T^{\text{miss}}$ in $Z \rightarrow \mu\mu$ events in data and simulation [77]. The last bin includes all events with $\vec{p}_T^{\text{miss}} > 195$ GeV. Lower panel: Data to simulation ratio. The systematic uncertainties due to the JES, the JER, and variations in the unclustered energy are added in quadrature and represented by the shaded band.

On the other hand, a variety of reconstruction failures and malfunctioning detectors leads to unexpected energy deposits so that anomalous p_T^{miss} is accidentally reconstructed, especially, high- p_T^{miss} region, where the VH production with $Z \rightarrow \nu\nu$ and physics beyond the Standard Model are sensitive. For example, a high- p_T muon track with a low quality reconstruction could contribute to p_T^{miss} as a poorly reconstructed PF muon. Some dead cells or channels in the sub-detectors could happen due to non-functioning electronics and this will result in the underestimation of energy deposits. Spurious signals also come from noise in readout electronics or in photodetectors, as well as from direct particle interaction with signal transition fibers or amplifiers. Moreover, those beam halo particles which propagate almost along the collision axis could hit on calorimeters and produce unexpected signals. These anomalous signals are already removed in the stage of the CMS trigger system and event reconstruction through dedicated algorithms according to topological and timing information. The remaining will be removed by a series of dedicated noise filters, shown in Table 3.8. Figure 3.19 shows agreement between data and simulation after these event filters applied under dijet and monojet events.

Filter	2016	2017	2018
Primary vertex filter	✓	✓	✓
Global super-tight halo filter	✓	✓	✓
HBHE noise filter	✓	✓	✓
HBHE iso noise filter	✓	✓	✓
Ecal dead cell trigger primitive filter	✓	✓	✓
Bad PF muon filter	✓	✓	✓
EE bad SC noise filter (only for data)	✓	✓	✓
ECAL bad calibration filter	✗	✓	✓

Table 3.8: Content of anomalous p_T^{miss} event filter

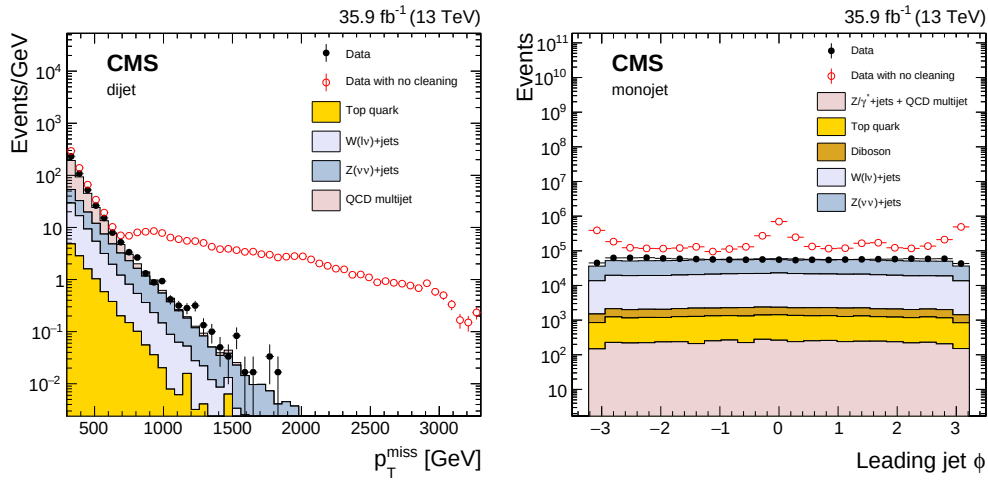


Figure 3.19: The p_T^{miss} (left) and jet ϕ (right) distributions for events passing the dijet (left) and monojet (right) selection with the event filtering algorithms applied, including that based on jet identification requirements (filled markers), without the event filtering algorithms applied (open markers), and from simulation (solid histograms) [77].

Chapter 4

Event categorization

In this chapter, event categorization for the VH production in $H(\rightarrow \gamma\gamma)$ channel follows up the STXS framework to optimize the VH analysis and measure its kinematic structures, described in Section 1.4. Since the VH STXS bins with a leptonically decaying W or Z boson are split by different productions and the transverse momenta of the W and Z bosons, the analysis categories have to be designed to enrich these STXS bins. However, analysis categories with fine granularity could reduce the sensitivity of their targeted STXS bins due to low statistics in some categories constrained by the diphoton channel with the limited branching fraction. Merging these analysis categories and corresponding STXS bins is necessary. Hence, the VH analysis in $H \rightarrow \gamma\gamma$ channel adopts a sequential approach with two steps corresponding to two bin merging schemes, i.e. stage 0 and stage 1.2 in the STXS framework. In the step 1, the VH analysis categories are defined to enrich the STXS stage 0 bin, which separates the associated VH production into the WH and ZH STXS bins with leptonic and hadronic cases, given by Section 4.1. In each analysis category, a multivariate classifier is built to optimize the signal and background separation, introduced in Section 4.2. In the step 2, the analysis categories defined by the step 1 are further split to enrich the STXS stage 1.2 with higher granularity according to different transverse momenta of the W and Z bosons on the top of the STXS stage 0, described in Section 4.3.

4.1 Categorization for the VH STXS stage 0

The STXS stage 0 bins are mainly defined according to the different Higgs production mechanisms. For the associated VH production with the V hadronic decay, the distinction between the $q\bar{q} \rightarrow VH$ and VBF processes, and that between the $gg \rightarrow ZH$ and ggH production is ambiguous in reconstruction level. Therefore, the associated VH production in the STXS stage 0 only focuses on the V decaying leptonically, while the $q\bar{q} \rightarrow V(\rightarrow jj)H$ process is grouped to the electroweak qqH production bin with the VBF production together. Similarly, the $gg \rightarrow Z(\rightarrow jj)H$ and the ggH production are merged to a STXS bin called as ggH . Moreover, the VH leptonic bin is also divided into three bins by the different processes, i.e. $q\bar{q}' \rightarrow WH$,

$q\bar{q} \rightarrow ZH$ and $gg \rightarrow ZH$. Since the $q\bar{q} \rightarrow ZH$ and $gg \rightarrow ZH$ processes are not also distinguishable in reality, the two STXS stage 0 bins are merged to one bin. A complete schematic diagram for the STXS stage 0 is shown in Figure 4.1.

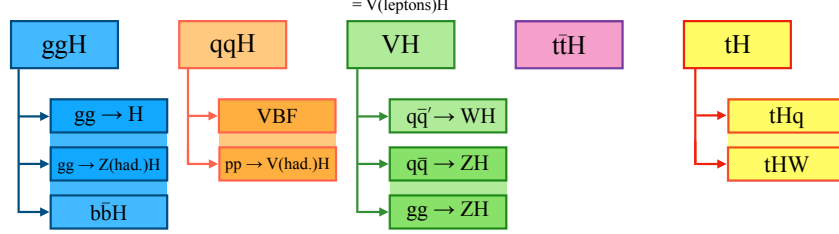


Figure 4.1: Schematic diagram for the STXS stage 0 bins. Some bins are merged by placing similar but light color areas behind these STXS bins.

In the $H \rightarrow \gamma\gamma$ analysis, the categorization strategy in the step 1 refers to the STXS stage 0 and designs a series of analysis categories. These categories are defined according to the final states of targeted STXS bins in order to avoid the significant migration of the STXS bins from one to other categories and the large correlation between multiple STXS bins inside a single category at the same time. In the VH analysis, there are four analysis categories designed to enrich the VH STXS stage 0 bins including the hadronic VH bin from the qqH bins, totally three leptonic and one hadronic categories with the targeted VH STXS bins shown in Table 4.1.

Category	W/Z decay	Target STXS bin
ZH Leptonic	2ℓ	$q\bar{q}(gg) \rightarrow ZH$
WH Leptonic	1ℓ	$q\bar{q}' \rightarrow WH$
VH MET	0ℓ	$q\bar{q}(gg) \rightarrow ZH$ or $q\bar{q}' \rightarrow WH$
VH Hadronic	hadron	$q\bar{q} \rightarrow V(\text{had.})H$

Table 4.1: Four VH analysis categories and corresponding targeted STXS bins. The ℓ represents the charged lepton such as e, μ .

Furthermore, to establish a series of mutually exclusive analysis categories, the $H \rightarrow \gamma\gamma$ analysis performs an event flow where these categories are arranged with a given precedence, shown in Figure 4.2. In this event flow, events passing the requirement of a given category are kept, while the rest events will be propagated to the next category. Events not kept in any category are eventually dropped from the $H \rightarrow \gamma\gamma$ analysis. Since it's observed that events with special final state objects like b -jets or charged leptons have relatively small statistics but are easier caught, the categories targeting these kinds of events have higher priority. Through this event flow, these rare events can avoid mixing with other common events and overall analysis categories also achieve the effect of exclusion without any event overlapping. Therefore, the VH Leptonic categories have higher priority but the VH Hadronic one is arranged almost at the end of the event flow. The following sections will introduce four VH enriched categories.

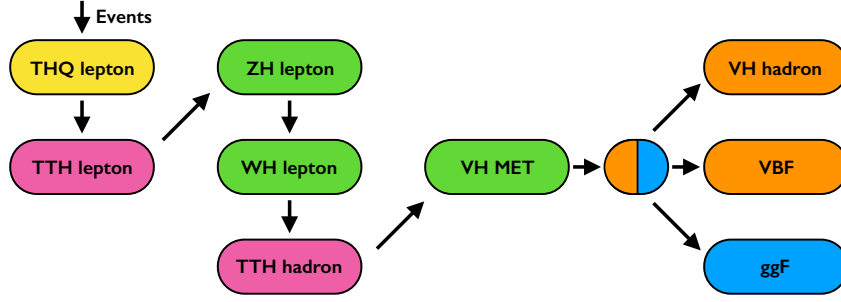


Figure 4.2: The event flow is composed of different categories. The analysis categories targeting events with special final state objects like b jets or charged leptons have Higher priority. The color in the analysis categories stand for the targeted STXS bins from Figure 4.1.

4.1.1 ZH Leptonic category

The ZH Leptonic category aims at the $q\bar{q}(gg) \rightarrow Z(\rightarrow \ell\ell)H$ process with the event topology in the CMS detector shown in Figure 4.3. Events with two isolated photons and two isolated, same-flavour charged leptons without tau are included in this category.

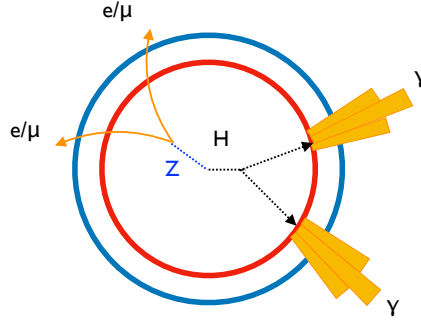


Figure 4.3: The event topology of the $Z(\rightarrow \ell\ell)H$ process in the CMS detector.

In this category, since two additional charged leptons are required in the events with the diphoton candidate, the main backgrounds for the diphoton final state like the QCD multi-jets, γ +jets and $\gamma\gamma$ non-resonance events with the huge cross sections are almost removed. The rest background processes are the $Z(\rightarrow \ell\ell) + \gamma$ with one fake photon ¹ and few $t\bar{t}H$ production with the two top quarks decaying leptonically. On the top of the physical object selections in Chapter 3, the loosest selection requirements are applied to the ZH Leptonic category to obtain the ZH lepton-like events and keep the statistics for further multivariate training at the same time :

- Exact two same-flavour charged leptons passing the lepton ID/ISO

¹In fact, the $Z(\rightarrow \ell\ell) + \gamma\gamma$ process has the same final states like $Z(\rightarrow \ell\ell)H$ process, but its limited cross section with respect to $Z(\rightarrow \ell\ell) + \gamma$ makes its contribution ignored in the ZH Leptonic category.

- $60 < m_{\ell\ell} < 120$ GeV
- $\Delta R(\gamma, \ell) > 0.2$ for the photons and charged leptons

After the above selections, the data-and-simulation comparisons without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV, given in Figure 4.4 and 4.5, show that the kinematic behaviors of the photons and leptons in data are well-described by those in the background simulation samples. These results can extrapolate that the background simulation samples are also able to describe the background of data in $m_{\gamma\gamma} \in [120, 130]$ GeV region under the ZH Leptonic category. This is an important key for the construction of later multivariate classifiers for signal and background separation, described in Section 4.2, where all the simulation samples including signal and background are used as training samples.

4.1.2 WH Leptonic category

Events failing the criteria of the ZH Leptonic category will enter the WH Leptonic category with the event topology in the CMS detector shown in Figure 4.6. The WH Leptonic category mainly targets the $q\bar{q}' \rightarrow W(\rightarrow \ell\nu)H$ process, hence, events with two isolated photons and exact one isolated charged lepton without any tau lepton included will be kept in this category. Since the neutrino is undetectable due to the limitation of the CMS detector, the missing transverse energy providing a part of the neutrino information is also taken into account. The undetectable neutrino could produce higher missing transverse energy with respect to processes without any neutrino.

The main background processes in this category are expected to be the $W(\rightarrow \ell\nu) + \gamma$, $Z(\rightarrow \ell\ell) + \gamma$ ² and Drell-Yan process with fake photons. Moreover, it's observed that the $t\bar{t}H$ production with at least one top quark decaying leptonically and $Z(\rightarrow \ell\ell)H$ with a charged lepton failing the lepton criteria are compatible in the WH Leptonic category. Following the goal of the STXS framework, these should be removed from this category to avoid large correlation between these STXS bins. For this task will be taken care during the construction of the multivariate classifier in the WH Leptonic category.

However, in the $m_{\gamma\gamma} \notin [120, 130]$ GeV region, dominated by non-resonance backgrounds, it's observed that the QCD multi-jets events have significant contribution in this category but are poorly-described by the corresponding simulation sample, shown in Figure 4.7. This observation gives an essential problem that the sample simulating the QCD process has no ability to extrapolate the QCD events in data. It even reduces the performance of the later multivariate classifier which rely on these simulated samples. The rejection of the QCD contribution in the WH Leptonic category is thus necessary. In principle, these fake charged leptons originating from hadron decays in jets carry relatively low momenta than prompt ones. Raising the threshold of the charged lepton p_T from 10 GeV to 15 GeV is reasonable to remove these QCD multi-jets events and keep a compatible signal efficiency in this category. Additionally, leading and sub-leading photons are selected with the photon ID BDT scores of at least -0.4 to further

²The $W(\rightarrow \ell\nu) + \gamma\gamma$ process has the same final states like $W(\rightarrow \ell\nu)H$ process, but its limited cross section with respect to $W(\rightarrow \ell\nu) + \gamma$ and $Z(\rightarrow \ell\ell) + \gamma$ makes its contribution ignored in the WH Leptonic category.

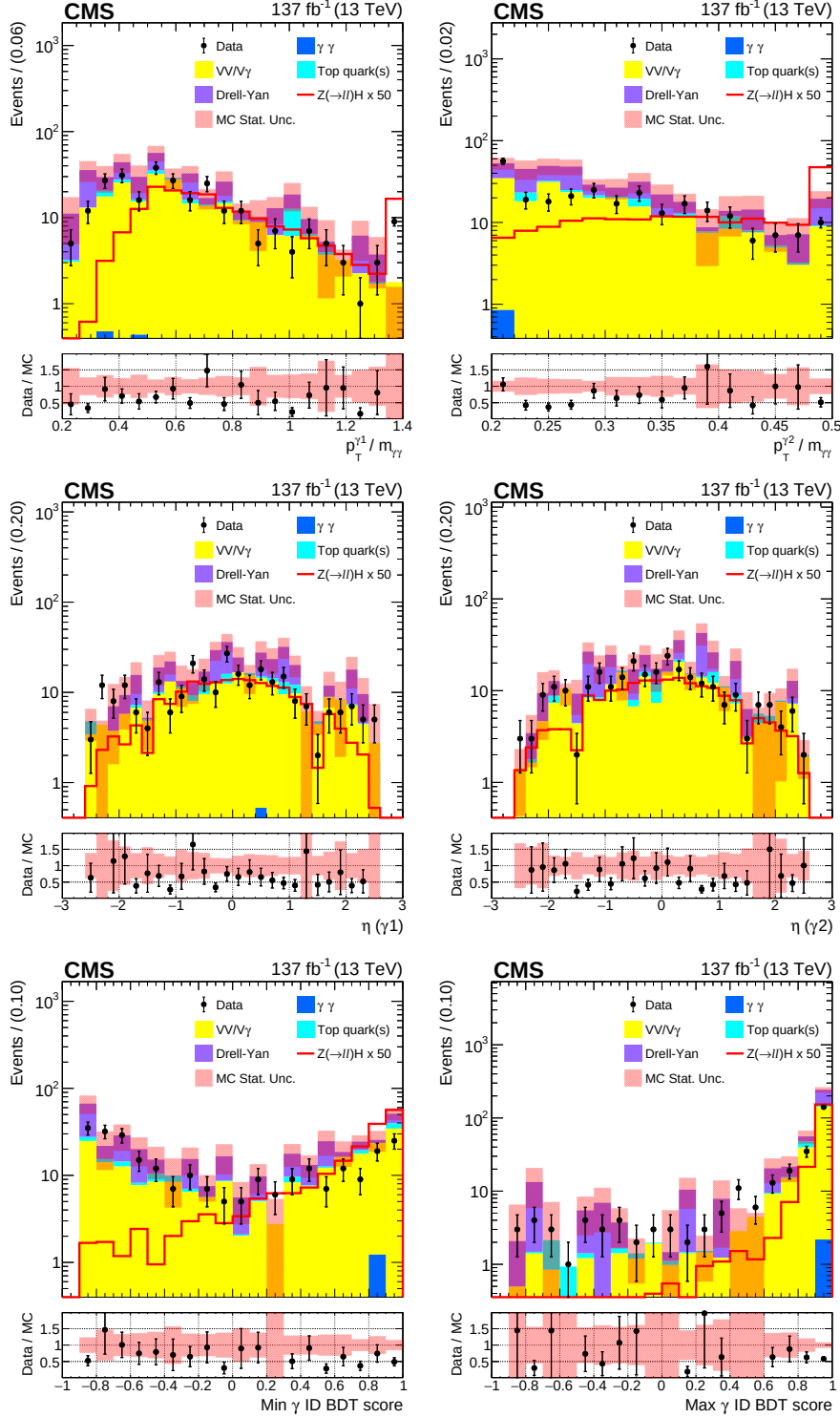


Figure 4.4: Data-and-simulation comparison of the photon kinematics and photon ID BDT distributions with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the ZH Leptonic category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. The red line is the sample simulating the $q\bar{q} \rightarrow Z(\rightarrow \ell\ell)H$ process with $m_H = 125$ GeV and is rescaled 50 times than the expected yield for visibility. The bottom panel in each figure is the data-to-simulation ratio.

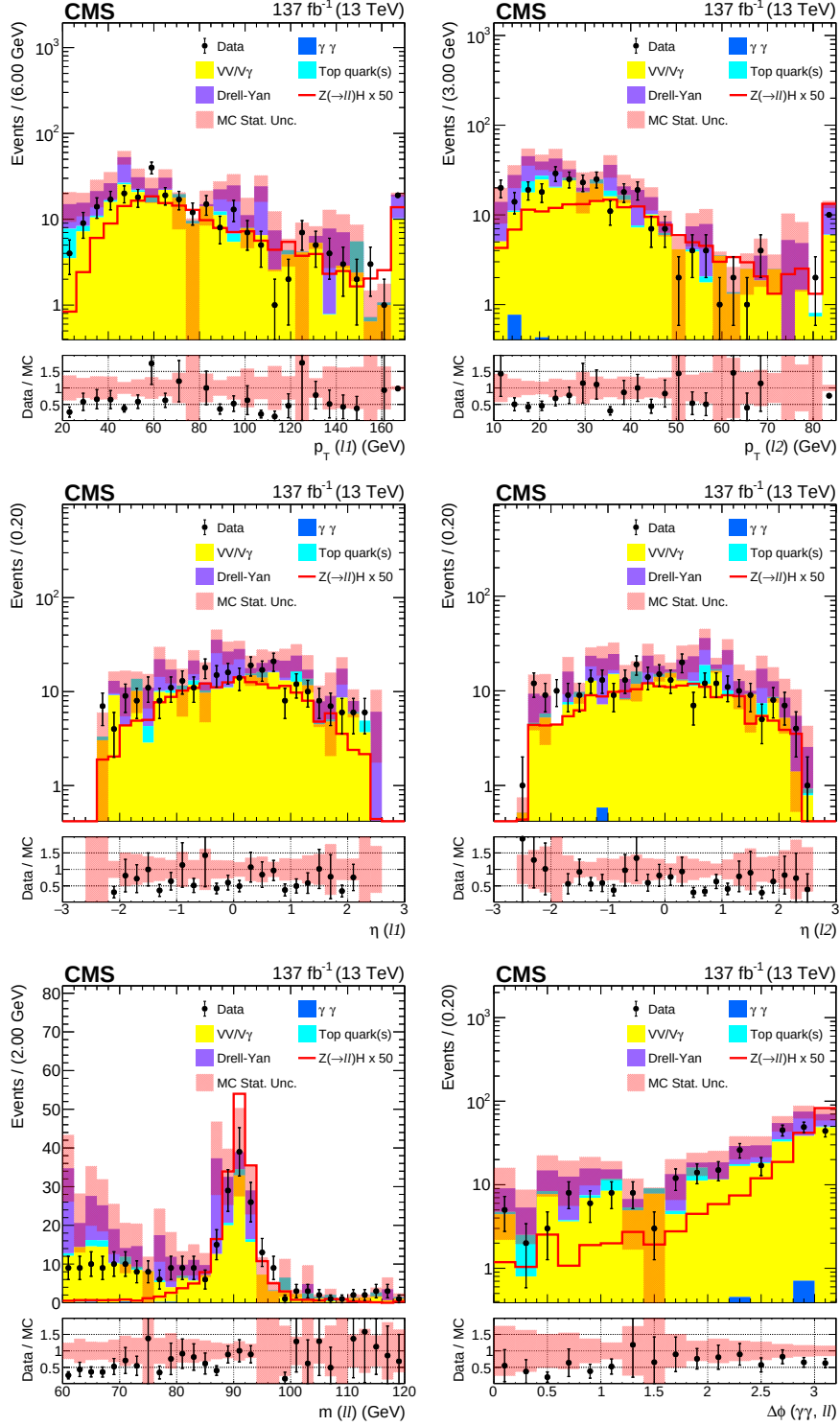


Figure 4.5: Data-and-simulation comparison of the charged lepton kinematics with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the ZH Leptonic category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. The red line is the sample simulating the $q\bar{q} \rightarrow Z(\rightarrow \ell\ell)H$ process with $m_H = 125$ GeV and is rescaled 50 times than the expected yield for visibility. The button panel in each figure is the data-to-simulation ratio.

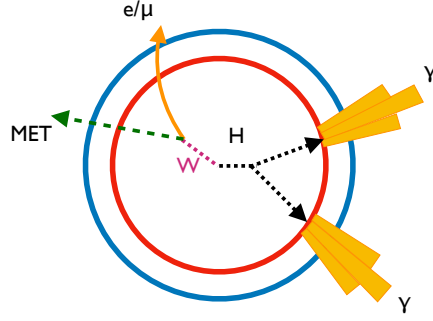


Figure 4.6: The event topology of the $W(\rightarrow \ell\nu)H$ process in the CMS detector. The undetectable neutrino from W boson decay is replaced by the missing transverse energy marked with a green dashed arrow.

remove the fake photons from jets. With these two selections, all of QCD multi-jets events are almost removed, shown in Figure 4.7 (right) and the corresponding signal efficiency is up to 90% which is acceptable.

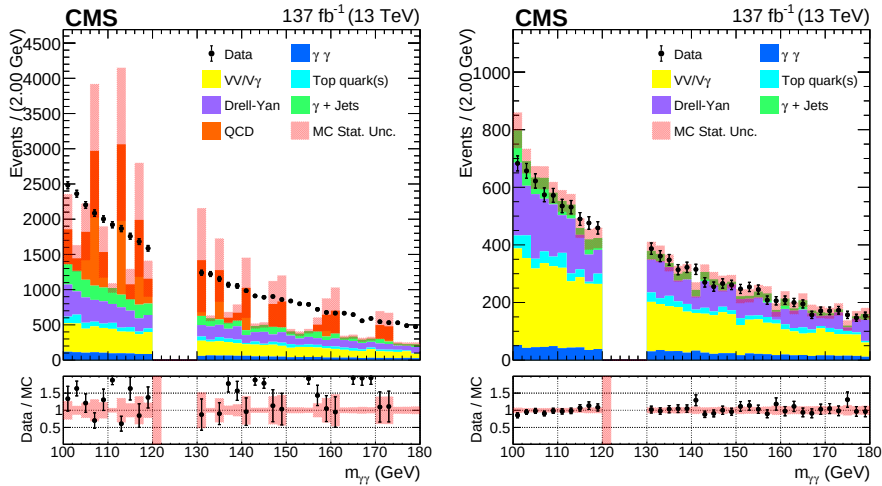


Figure 4.7: Data-and-simulation comparison of the diphoton mass before (left) and after (right) two selections with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the WH Leptonic category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty for the background samples is also shown with pink boxes. The bottom panel in each figure is the data-to-simulation ratio.

On the other hand, during event reconstruction, electrons could be misidentified as photons when the low-quality electron tracks are removed by certain track-quality thresholds so that no GSF track matches with superclusters in the ECAL detector. Therefore, $Z(\rightarrow e^+e^-) + \gamma$ and Drell-Yan processes with an electron misidentified as a photon are included in the category. Figure 4.8 shows that there is a Z boson resonance given by these two processes in the photon-plus-electron invariant mass distributions, while the flat distribution is expected in the $W(\rightarrow$

$\ell\nu)H$ process.

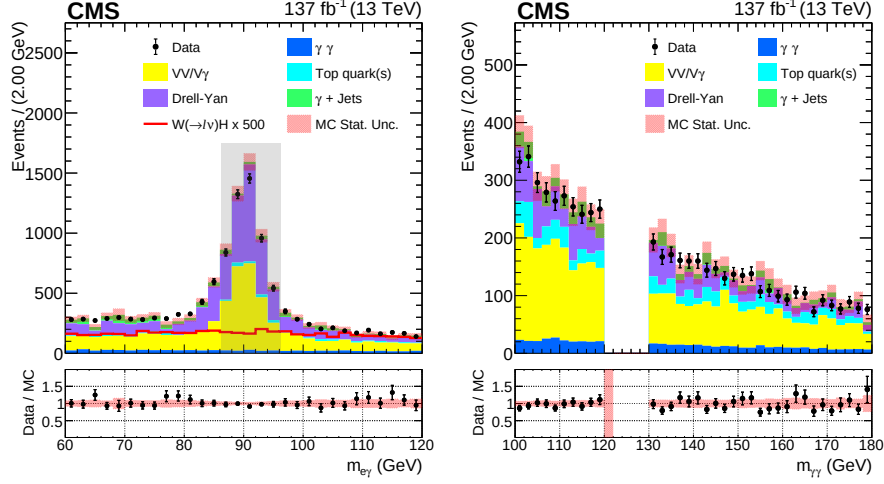


Figure 4.8: Data-and-simulation comparison of the photon-electron pair (left) and diphoton (right) mass with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the WH Leptonic category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. The red line is the sample simulating the $W(\rightarrow \ell\nu)H$ process with $m_H = 125$ GeV and is rescaled 500 times than the expected yield for visibility. The gray area is to kill the Z boson mass window where the electron is misidentified as the photon. The right is the diphoton mass after killing. The bottom panel in each figure is the data-to-simulation ratio.

In order to kill this kind of misidentification, events with the photon-plus-electron invariant mass falling in a range of ± 5 GeV with respect to Z boson mass will be rejected from this category. This requirement also leads to a reduction of 6% in terms of the expected $W(\rightarrow \ell\nu)H$ events. In summary, an entire set of selections in the WH Leptonic category on the top of physical object selections is listed in the following :

- Photon ID BDT scores for both photons > -0.4
- Exact one charged lepton passing the lepton ID/ISO
- $p_T^\ell > 15$ GeV for the charged lepton
- $\Delta R(\gamma, \ell) > 0.2$ for both photons and the charged lepton
- $|M_{\gamma e} - M_Z| < 5$ GeV

After the above selections, the data-and-simulation comparisons without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV, given in Figure 4.9 and 4.10, show that the kinematic behaviors of two photons, one lepton and missing transverse energy in data are well-described by those in the background simulated samples. Therefore, it's trusted that the samples well simulate the background within $m_{\gamma\gamma} \in [120, 130]$ GeV region under the WH Leptonic category.

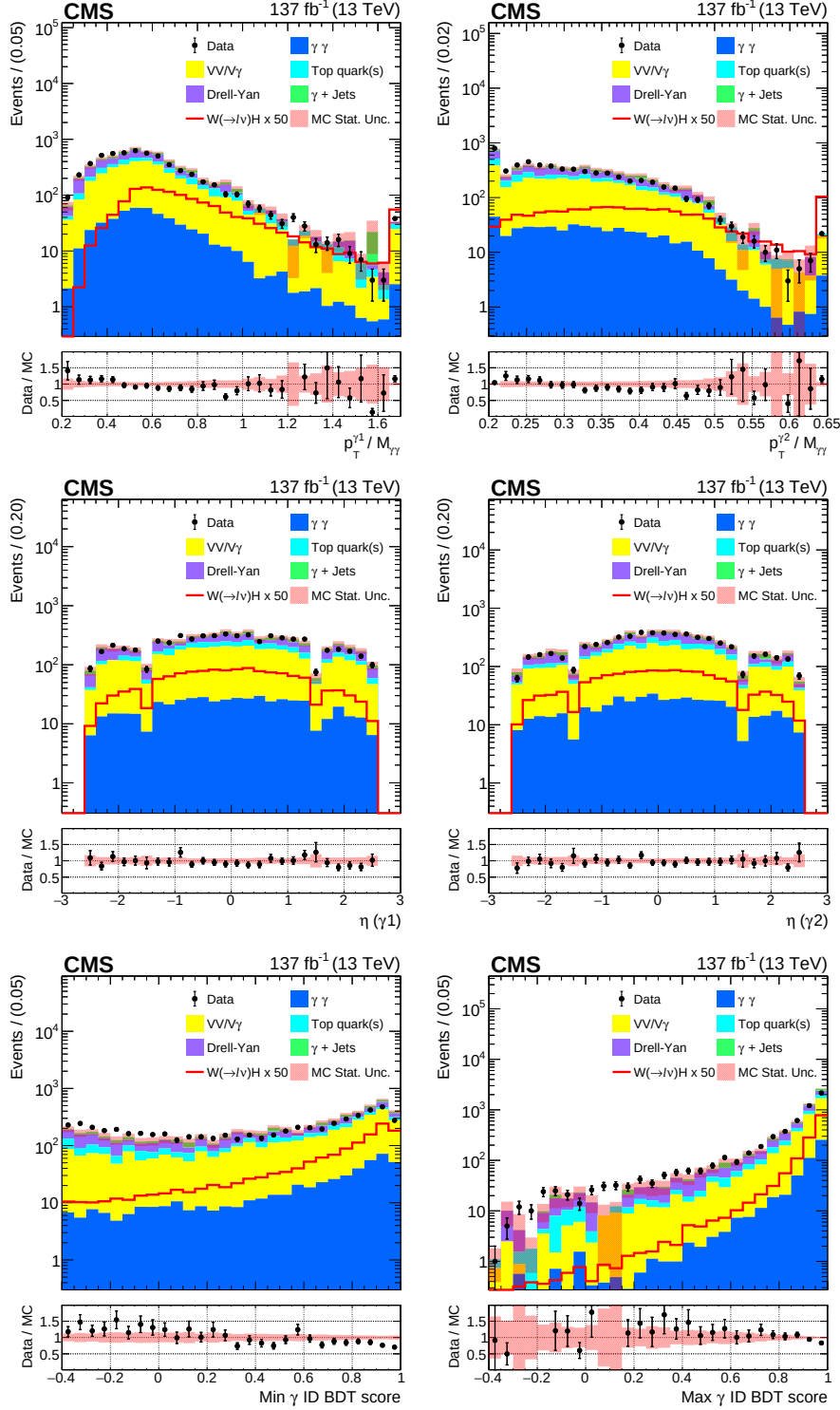


Figure 4.9: Data-and-simulation comparison of the photon kinematics and photon ID BDT distributions with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the WH Leptonic category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. The red line is the sample simulating the $W(\rightarrow \ell\nu)H$ process with $m_H = 125$ GeV and is rescaled 50 times than the expected yield for visibility. The bottom panel in each figure is the data-to-simulation ratio.

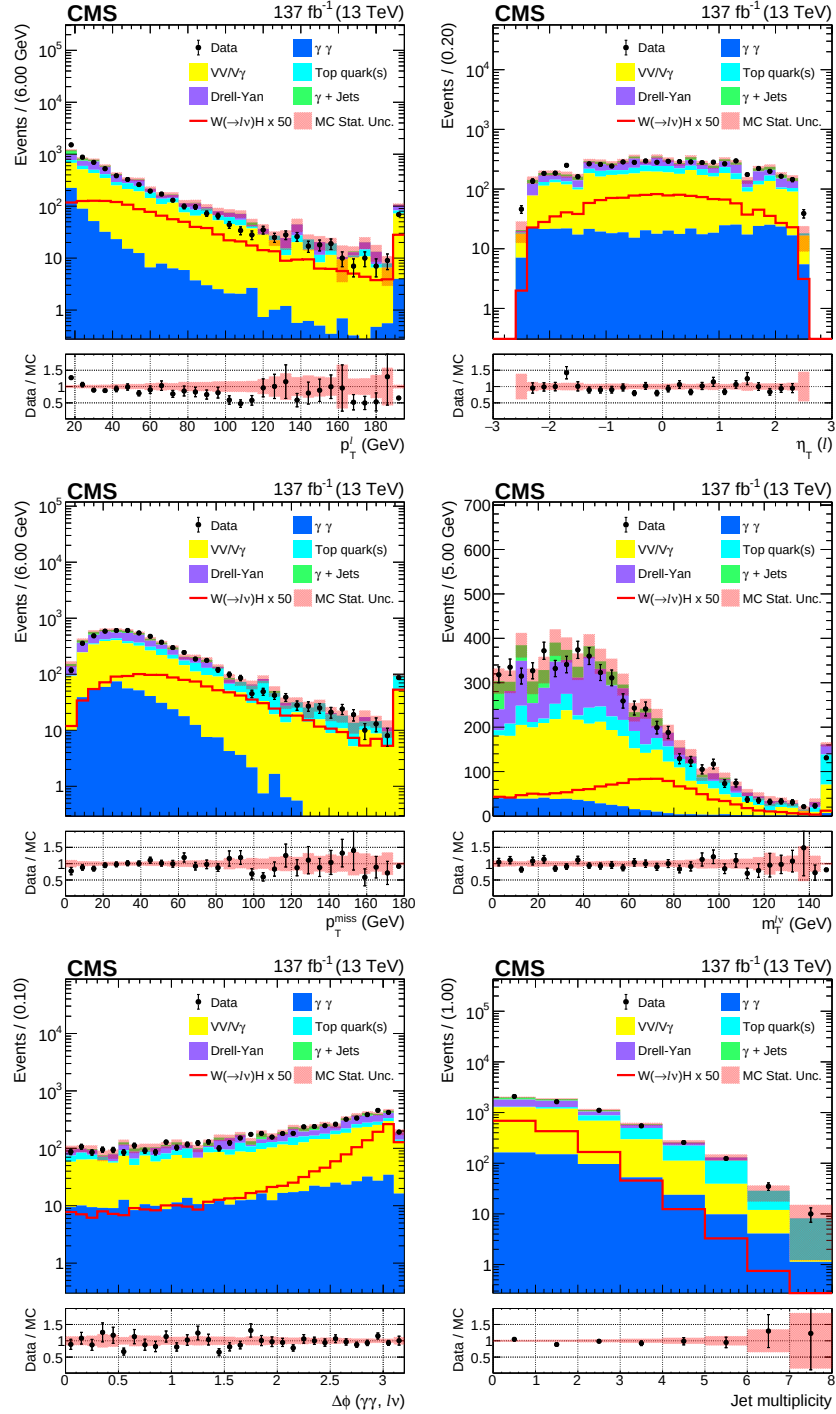


Figure 4.10: Data-and-simulation comparison of the lepton, missing transverse energy kinematics and other distributions with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the WH Leptonic category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. The red line is the sample simulating the $W(\rightarrow \ell\nu)H$ process with $m_H = 125$ GeV and is rescaled 50 times than the expected yield for visibility. The bottom panel in each figure is the data-to-simulation ratio.

4.1.3 VH MET category

The main purpose for the VH MET category is to hunt the $q\bar{q}(gg) \rightarrow Z(\rightarrow \nu\nu)H$ process and the rest $W(\rightarrow \ell\nu)H$ process where the charged lepton does not pass the lepton requirements. The event topology of both processes in the CMS detector looks like Figure 4.11. Since the neutrino(s) can not be reconstructed due to the limitation of the CMS detector, events with two isolated photons can be collected in this category. The missing transverse energy providing a part of the neutrino information is also taken into account as the neutrino(s) could produce higher missing transverse energy with respect to processes without any neutrino.

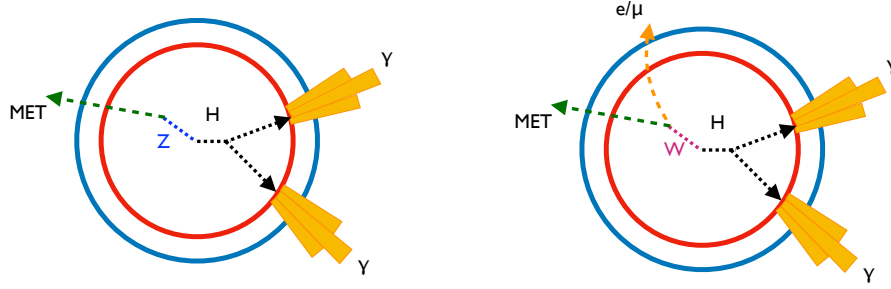


Figure 4.11: The event topology of the $Z(\rightarrow \nu\nu)H$ process (left) and $W(\rightarrow \ell\nu)H$ process (right) in the CMS detector. The undetectable neutrino(s) from the W and Z boson decays are replaced by the missing transverse energy marked with a green dashed arrow.

Without charged leptons supported, the QCD multi-jets, γ + jets and $\gamma\gamma$ non-resonance events with huge cross sections become dominant backgrounds³. Moreover, an amount of the VBF , gluon fusion and hadronic VH processes with the similar final states are also kept in this category. These processes should be removed from this category, otherwise, strong correlation between the targeted VH processes and the other STXS bins are produced in this category. The removal is performed in the later multivariate classifier in the VH MET category.

Similarly, the samples simulating the dominant QCD multi-jets and γ + jets events in the VH MET category have poor description to data, even if tighter selections are applied to the category to reduce these background contributions, shown in Figure 4.12.

To solve this problem, the analysis for the VH MET category adopts a dedicated data-driven background estimation which is successfully derived in Ref [24]. This data-driven technique replaces two kinds of poorly-simulated samples, i.e. QCD multi-jets and γ + jets, by data from background-like regions with similar kinematic behavior. Before performing the data-driven method, two very loose cuts are applied to events :

- Charged-lepton veto : veto charged leptons passing the lepton selections described in Chapter 3
- $\Delta\phi(\gamma\gamma, \vec{p}_T^{\text{miss}}) > 2$

³Although $Z(\rightarrow \nu\nu) + \gamma$ events with a fake photon have similar final states, its cross section is not compatible to those QCD events

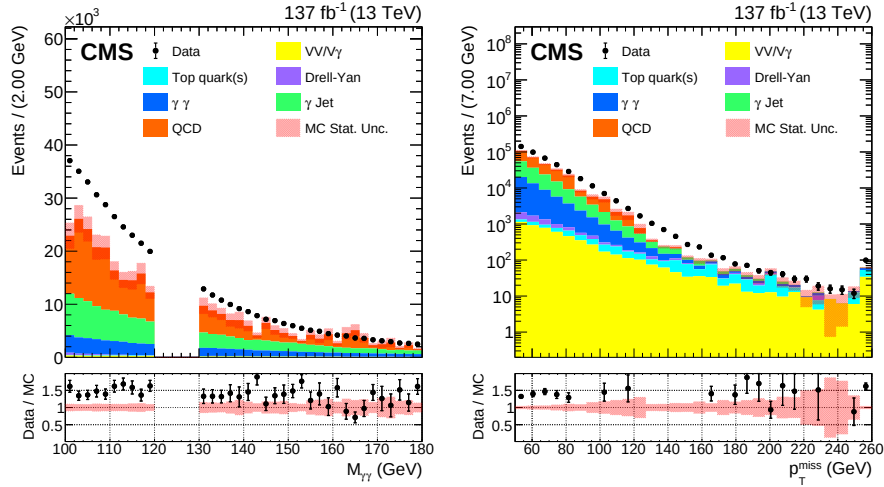


Figure 4.12: Data-and-simulation comparison of the diphoton mass, $m_{\gamma\gamma}$, with the 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the VH MET category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. The bottom panel in each figure is the data-to-simulation ratio.

The data-driven method defines two regions : "signal-like region" with the minimal and maximal photon ID BDT scores of two photons > -0.7 where the signal is almost located and "sideband region" with the minimal photon ID BDT score < -0.7 , maximal photon ID BDT score > -0.7 , as described in Figure 4.13.

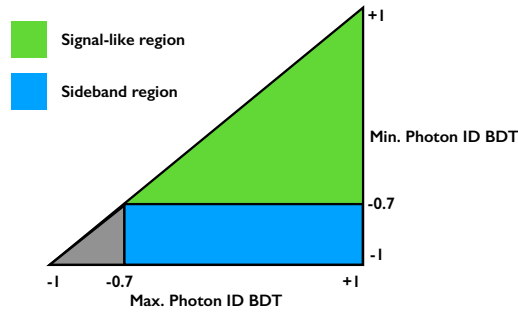


Figure 4.13: The diagram of the signal-like region (green) and the sideband region (blue) with boundaries of -0.7 in the minimal and maximal photon ID BDT scores.

To validate the data-driven description under the VH MET category, there are three assumptions that need to be satisfied :

1. **The sideband region is dominated by the QCD multi-jets and γ + jets events.** : The assumption is to make sure that events in the sideband region can replace the QCD multi-jets, γ + jets samples in the signal-like region. It's observed that approximate 98 % events in the sideband region originate from these two backgrounds.

2. **The correlation between the photon ID BDT score and kinematics is small.** : Similar kinematic behaviors in different photon ID BDT score regions will verify that these QCD multi-jets and γ + jets samples in the signal-like region can be replaced by data from the sideband region. A simple way to check the above statement is to calculate the correlation between the photon ID BDT score and kinematic variables. Small correlation will confirm that the kinematic distributions in both regions are similar. Table 4.2 shows that the Pearson correlation coefficients between the photon ID BDT score and the kinematic variables in the γ + jets sample are less than 5 % which is small except the p_T of the leading photon. However, small correlation between the fake photon ID BDT score and p_T is revealed in Figure 4.14 where the fake photon ID BDT score distributions in different p_T and η regions are similar.

Variables	Fake photon ID BDT	Prompt photon ID BDT
$p_T^{\gamma 1}/m_{\gamma\gamma}$	-0.127	0.077
$p_T^{\gamma 2}/m_{\gamma\gamma}$	0.04	-0.04
$\eta^{\gamma 1}$	0.002	0.007
$\eta^{\gamma 2}$	0.002	0.007
$\cos(\Delta\phi(\gamma\gamma))$	0.01	0.02
p_T^{miss}	0.003	-0.049
Jet multiplicity	0.02	0.005
p_T^{jet1}	-0.01	0.04
$\Delta\phi(\gamma\gamma, p_T^{\text{miss}})$	0.003	-0.008

Table 4.2: Pearson correlation coefficients between the prompt/fake photon ID BDT scores and other kinematic variables.

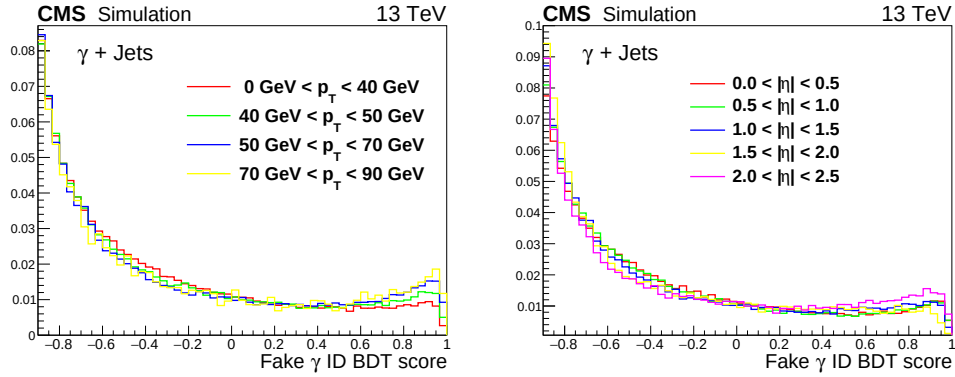


Figure 4.14: Photon ID BDT distributions with the different p_T (left) and η (right) regions for the fake photons in the γ + jets sample.

3. **At least one photon is fake for events in the sideband region.** : The assumption is always true for the QCD events with two fake photons, and there is 92.1 % 1 prompt + 1 fake photons for the γ + jets events. This can assume that a photon with the minimal photon ID BDT score is always fake for the QCD and γ + jets events in both regions.

Although the data-driven method allows replacing the QCD multi-jets and γ + jets samples in the signal-like region by data in the sideband region, the photon ID BDT scores themselves will be forbidden due to the definition of the sideband region. However, for the construction of the later multivariate classifier, the minimal and maximal photon ID BDT scores provide the most discriminant features for rejecting events with fake photons. With the third assumption of the data-driven description, the method can fit the fake photon ID BDT distribution from the γ + jets sample by the 7-th order polynomial, shown in Figure 4.15.

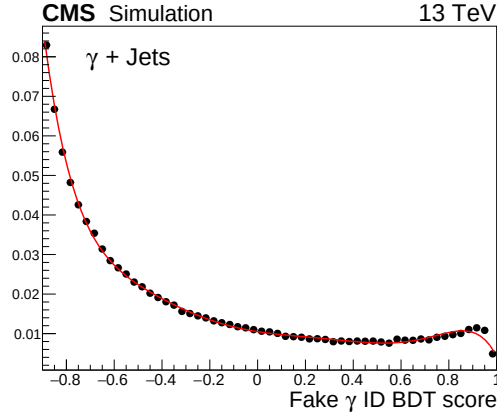


Figure 4.15: Fit result of the fake photon ID BDT PDF. The red line is the 7-th order polynomial form of the fake photon ID BDT PDF and the black dots are the fake photon ID BDT distribution in the γ + jets sample.

Then, for each maximal photon ID BDT score given by data in the sideband region, the minimal photon ID BDT score can be generated by this minimal photon ID BDT PDF in a range of [sideband cut value = -0.7, maximal photon ID BDT score] and mapped to the signal-like region. On the other hand, the maximal photon ID BDT score is directly mapped to the signal-like region from the sideband region. Since the maximal photon ID BDT scores in the signal-like region are expected to tend to higher values than ones in the sideband region, a per-event weight (w) defined by Eq. (4.1) is applied for correcting the maximal photon ID BDT distribution :

$$w = \frac{\int_B^{\max \text{ photon ID}} \text{fake PDF}}{\int_A^B \text{fake PDF}} \quad (4.1)$$

where A, B are the lower limit of the photon ID BDT score in the sideband region and the signal-like region, i.e. -0.9 and -0.7 , respectively. The maximal photon ID BDT distribution of the data-driven sample after applying the per-event weight has better description to data in the signal-like region, shown in Figure 4.16.

Once the data-driven description is done in the signal-like region, a set of loose selections are then applied to the signal-like region to kill the most background-like events :

- $p_T^{\gamma^1} / m_{\gamma\gamma} > 1/3$
- $p_T^{\gamma^2} / m_{\gamma\gamma} > 1/5$

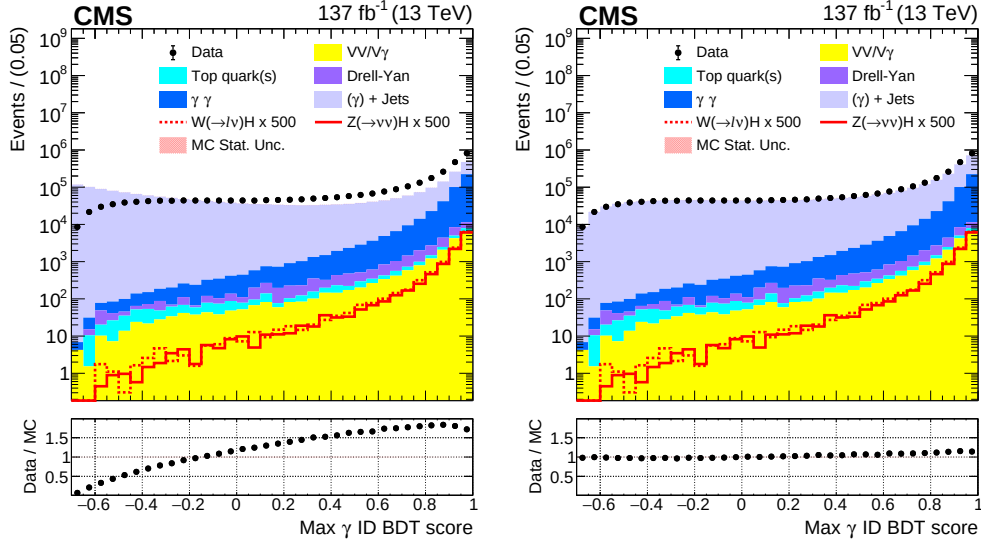


Figure 4.16: Data-and-simulation comparison of the maximal photon ID BDT distributions before (left) and after (right) applying the per-event weight with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the VH MET category. The different background and data-driven components are stacked with different color and solid histograms. The total statistical uncertainty given by the background simulated and data-driven samples is also shown with pink boxes. The bottom panel in each figure is the data-to-simulation ratio.

- $p_T^{\text{miss}} > 50$ GeV

However, since inconsistent overall normalization for the sideband and signal-like region is expected. In the signal-like region, the data-driven method performs a simultaneous binned likelihood fit of the minimal and maximal photon ID BDT to data. There are two shared and floated parameters scaling the data-driven and $\gamma\gamma$ non-resonance samples, while the yields of the other background samples are fixed to the total luminosity. Finally, Figure 4.17 shows better agreement between the data and simulated samples plus data-driven samples under the minimal photon ID BDT distribution after fitting.

After applying the data-driven technique, Figure 4.18 and 4.19 show the kinematics of the photons and jets, missing transverse energy and VH topology in data are well-described by those of the background simulation and data-driven sample in the $M_{\gamma\gamma} \notin [120, 130]$ GeV region. This is also expected to be true in $M_{\gamma\gamma} \in [120, 130]$ GeV region.

4.1.4 VH Hadronic category

Once events fail the requirements of the VH Leptonic categories, these events enter the VH Hadronic category which focuses on the $W(\rightarrow q\bar{q}')H$ and $q\bar{q} \rightarrow Z(\rightarrow q\bar{q})H$ processes⁴. However, since the energy of jets originating from quark fragmentation and hadronization is not well-measured, the W and Z boson resonances reconstructed by two jets are indistinguishable.

⁴Since the $gg \rightarrow Z(\rightarrow q\bar{q})H$ is merged into the ggH STXS bins, this process is not considered as the target.

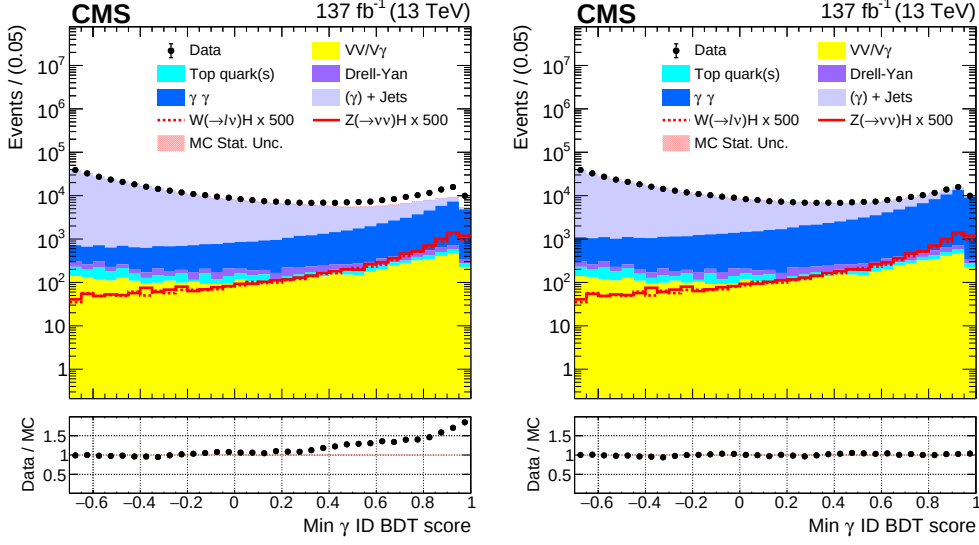


Figure 4.17: Data-and-simulation comparison of the minimal photon ID BDT distributions before (left) and after (right) fitting with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the VH MET category. The different background and data-driven components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. The bottom panel in each figure is the data-to-simulation ratio.

The goal of the VH Hadronic category is to collect events with two isolated photons and at least two hadronic jets from the W and Z bosons. The corresponding event topology in the CMS detector looks like Figure 4.20.

The main background components in the VH Hadronic category are expected to be the common QCD multi-jets, γ + jets and $\gamma\gamma$ non-resonance events like those in the VH MET category. To reduce these QCD multi-jets and γ + jets contributions, the $p_T^\gamma/m_{\gamma\gamma}$ and photon ID BDT score are tuned to tighter thresholds than those in the VH Leptonic categories. The transverse momentum of two jets are also selected with higher values than the default requirement, described in Chapter 3, in order to match with the definition of the particle-level jets in the STXS framework.⁵ The dijet invariant mass is restricted in a range of $[60, 120]$ GeV corresponding to the reconstructed W and Z resonances. The details of the selections in the VH Hadronic category on the top of Chapter 3 are shown in the following :

- $p_T^{\gamma 1}/m_{\gamma\gamma} > 1/3$ and $p_T^{\gamma 2}/m_{\gamma\gamma} > 1/4$
- Photon ID BDT score > -0.2
- $60 \text{ GeV} < m_{jj} < 120 \text{ GeV}$
- At least two jets with $p_T^{\text{jet}} > 30 \text{ GeV}$

⁵In fact, this is also for rejecting the pile-up jets which are expected to tend to lower energy.

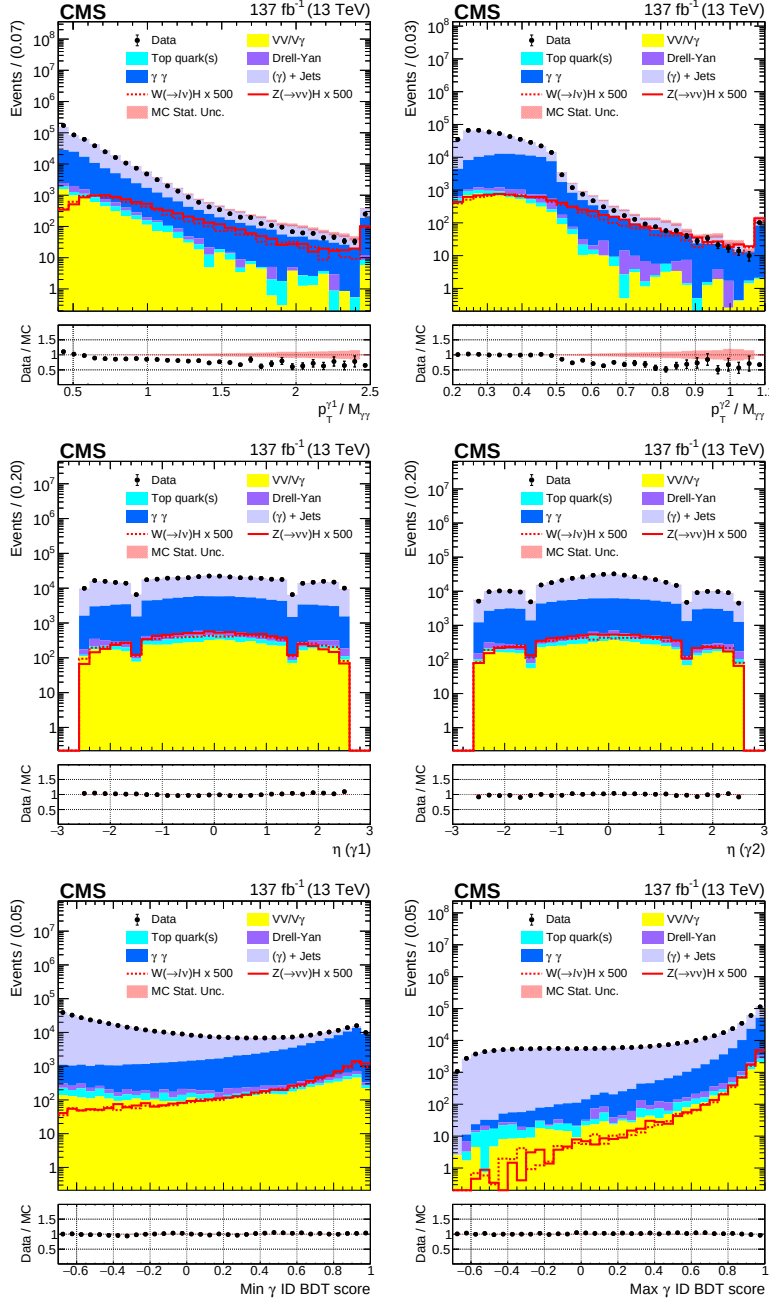


Figure 4.18: Data-and-simulation comparison of the photon kinematics and photon ID BDT distributions with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the VH MET category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. Two overlapping histograms with the red solid and dashed lines represent the samples simulating the $q\bar{q}(gg) \rightarrow Z(\rightarrow \nu\nu)H$ and $W(\rightarrow \ell\nu)H$ processes with $m_H = 125$ GeV, respectively, rescaled 500 times than the expected yield for visibility. The bottom panel in each figure is the data-to-simulation ratio.

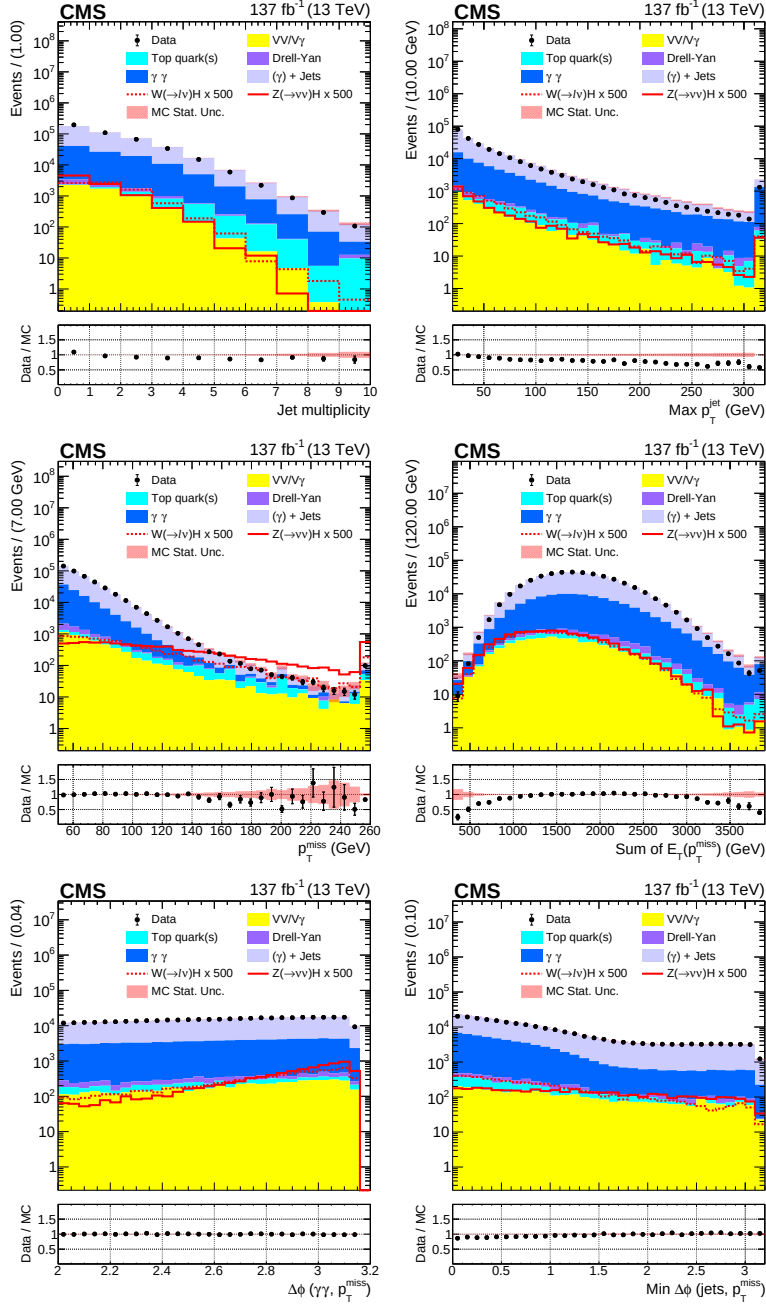


Figure 4.19: Data-and-simulation comparison of the jet, missing transverse energy and VH topology associated distributions with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the VH MET category. The different background components are stacked with different color and solid histograms. The total statistical uncertainty given by the background samples is also shown with pink boxes. Two overlapping histograms with the red solid and dashed lines represent the samples simulating the $q\bar{q}(gg) \rightarrow Z(\rightarrow \nu\nu)H$ and $W(\rightarrow \ell\nu)H$ processes with $m_H = 125$ GeV, respectively, rescaled 500 times than the expected yield for visibility. The bottom panel in each figure is the data-to-simulation ratio.

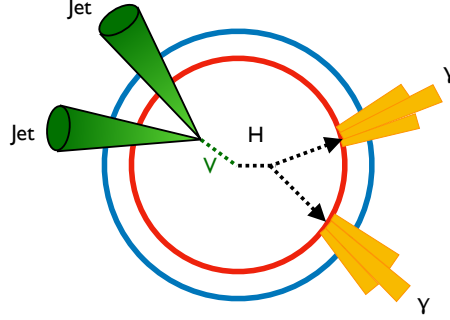


Figure 4.20: The event topology of the $V(\text{had.})H$ process in the CMS detector as the W and Z are regarded as the same boson. Two green cones are two jets from the W and Z boson decays.

Even though applying these selections, the QCD multi-jets and $\gamma + \text{jets}$ events can not be removed completely. Moreover, the same situation in the VH MET category also happens in the VH Hadronic category as the QCD multi-jets, $\gamma + \text{jets}$ samples are not able to well model background in data, especially in $m_{\gamma\gamma} \in [120, 130]$ GeV region. A dedicated data-driven background description is also performed in the VH Hadronic category. Since the later multivariate classifier in the VH Hadronic category does not use the photon ID BDT score as input variables, this desired data-driven method will be different from that in the VH MET category.

The data-driven method in the VH Hadronic category defines the signal-like region and control regions like :

- **Signal-like region (SR)** : both photons pass the photon ID BDT score > -0.2 , mainly populated by the $H \rightarrow \gamma\gamma$ signals and $\gamma\gamma$ process.
- **Control regions (CR)** : prompt-fake (PF), fake-prompt (FP) and fake-fake (FF) regions, where "prompt" is defined by requiring the photon ID BDT score > -0.2 and "fake" is the photon ID BDT score < -0.4 . The PF and FP regions are populated by the $\gamma + \text{jets}$ process and the FF region is populated by the QCD multi-jets process.

According to this definition, the data-driven method takes data from the PF CR, FP CR and FF CR to model the QCD multi-jets and $\gamma + \text{jets}$ samples in the signal-like region (PP SR), shown in the schematic of Figure 4.21.

For implementation of the data-driven method, data in each control region is mapped to the signal-like region through a data-driven event weight, $w_{\text{data-driven}}$, composed of a fake factor and a QCD fraction. The fake factor, $f(\eta^\gamma, p_T^\gamma)$, is a transformation coefficient from a fake photon in the control region to a prompt photon in the signal-like region, calculated by the ratio between simulated events in the signal-like region and ones in the control region like

$$f(\eta^\gamma, p_T^\gamma) = \frac{N^{\text{SR}}(\eta^\gamma, p_T^\gamma)}{N^{\text{CR}}(\eta^\gamma, p_T^\gamma)} \quad (4.2)$$

which is binned by four p_T bins of (0-30, 30-50, 50-100, >100 GeV) and two η bins of (EE, EB) of a given photon. On the other hand, according to the purpose of the data-driven description,

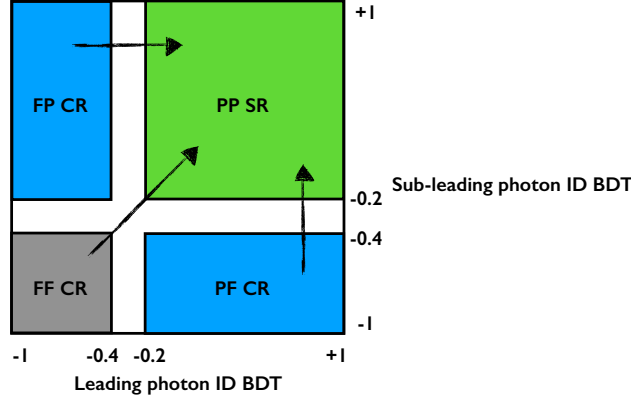


Figure 4.21: The schematic of the data-driven method in the VH Hadronic category. The PF CR and FP CR are highlighted by the blue block, the FF CR is the gray one and the signal-like region is the green one. All regions are defined by both photon ID BDT scores with two boundaries of -0.4 and -0.2 . Three black arrows represent data in three control regions that replaces the QCD multi-jets and γ + jets samples in the signal-like region.

only QCD multi-jets and γ + jets samples in the signal-like region need to be replaced instead of the $\gamma\gamma$ non-resonance sample. The data-driven event weight thus depends on a QCD fraction $p_{\text{QCD}}(\eta^\gamma, p_T^\gamma)$ with the same binning as that in the fake factor and defined by

$$p_{\text{QCD}}(\eta^\gamma, p_T^\gamma) = \frac{N_{j\gamma}^{\text{CR}} + N_{jj}^{\text{CR}}}{N_{\gamma\gamma}^{\text{CR}} + N_{j\gamma}^{\text{CR}} + N_{jj}^{\text{CR}}} \quad (4.3)$$

which can correct the number of events in the control region where the $\gamma\gamma$ non-resonance events can contribute. The fake factors and QCD fractions for 2017, for example, are summarized in Figure 4.22.

Eventually, the complete data-driven background can be defined as

$$N_{\text{data-driven}} = w_{\text{data-driven}}(\text{FP}) \times N_{\text{FP}} + w_{\text{data-driven}}(\text{PF}) \times N_{\text{PF}} - w_{\text{data-driven}}(\text{FF}) \times N_{\text{FF}} \quad (4.4)$$

where the data-driven event weights as a function of the leading and the sub-leading photon kinematics in the PF, FP and FF regions can be expected by

$$w_{\text{data-driven}} = \begin{cases} +p_{\text{QCD}}(\eta^{\gamma_1}, p_T^{\gamma_1}) \times f(\eta^{\gamma_1}, p_T^{\gamma_1}), & \in \text{FP} \\ +p_{\text{QCD}}(\eta^{\gamma_2}, p_T^{\gamma_2}) \times f(\eta^{\gamma_2}, p_T^{\gamma_2}), & \in \text{PF} \\ -p_{\text{QCD}}(\eta^{\gamma_1}, p_T^{\gamma_1}) \times f(\eta^{\gamma_1}, p_T^{\gamma_1}) \times p_{\text{QCD}}(\eta^{\gamma_2}, p_T^{\gamma_2}) \times f(\eta^{\gamma_2}, p_T^{\gamma_2}), & \in \text{FF}. \end{cases} \quad (4.5)$$

After applying the selections and the data-driven method, Figure 4.23 shows that the photon and jet kinematic behaviors of data is well-described by the $\gamma\gamma$ non-resonance and data-driven sample.

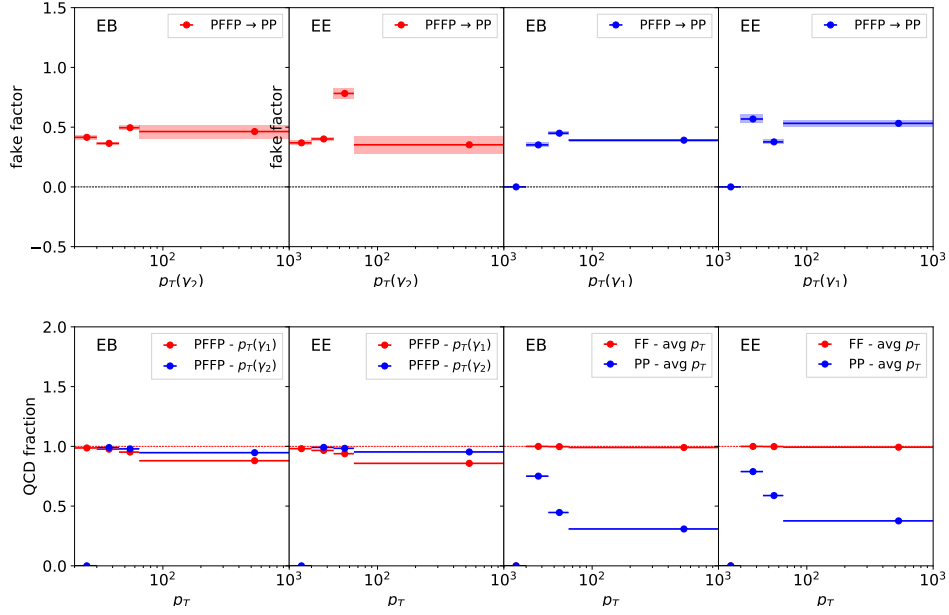


Figure 4.22: The fake factors (left) and QCD fractions (right) as a function of photon transverse momenta in different η regions for 2017.

4.2 Multivariate analysis

The STXS framework tries to maximize the sensitivity and reduce the theoretical dependence in measurements at the same time. It allows the analysis to use larger acceptance correction for reducing non-resonance background events and non-target STXS bins in analysis categories. The cut-based analysis by constructing a series of selection requirements could be a method to realize this goal. Nevertheless, in order to optimize the selection requirements which could have highly dimensional relations, the VH analysis under the STXS framework adopts a widely-used multivariate technique called Boosted Decision Tree (BDT) to achieve optimal discrimination between signal and background.

The Boosted Decision Tree is a kind of machine learning technique and its performance relies on training samples during BDT training. A simulated sample with good description to some realistic situation will provide correct information for classification. Namely, in the VH analysis, once it's confirmed that the background and data-driven samples can well model background events in data for each analysis category from the previous section, the BDT classifiers are able to be prepared through these background samples and given signal samples to achieve correct signal and background classification. Hence, this section will introduce the BDT technique in Section 4.2.1 and the BDT classifiers in the VH Leptonic categories in Section 4.2.2 and in the VH Hadronic category in Section 4.2.3.

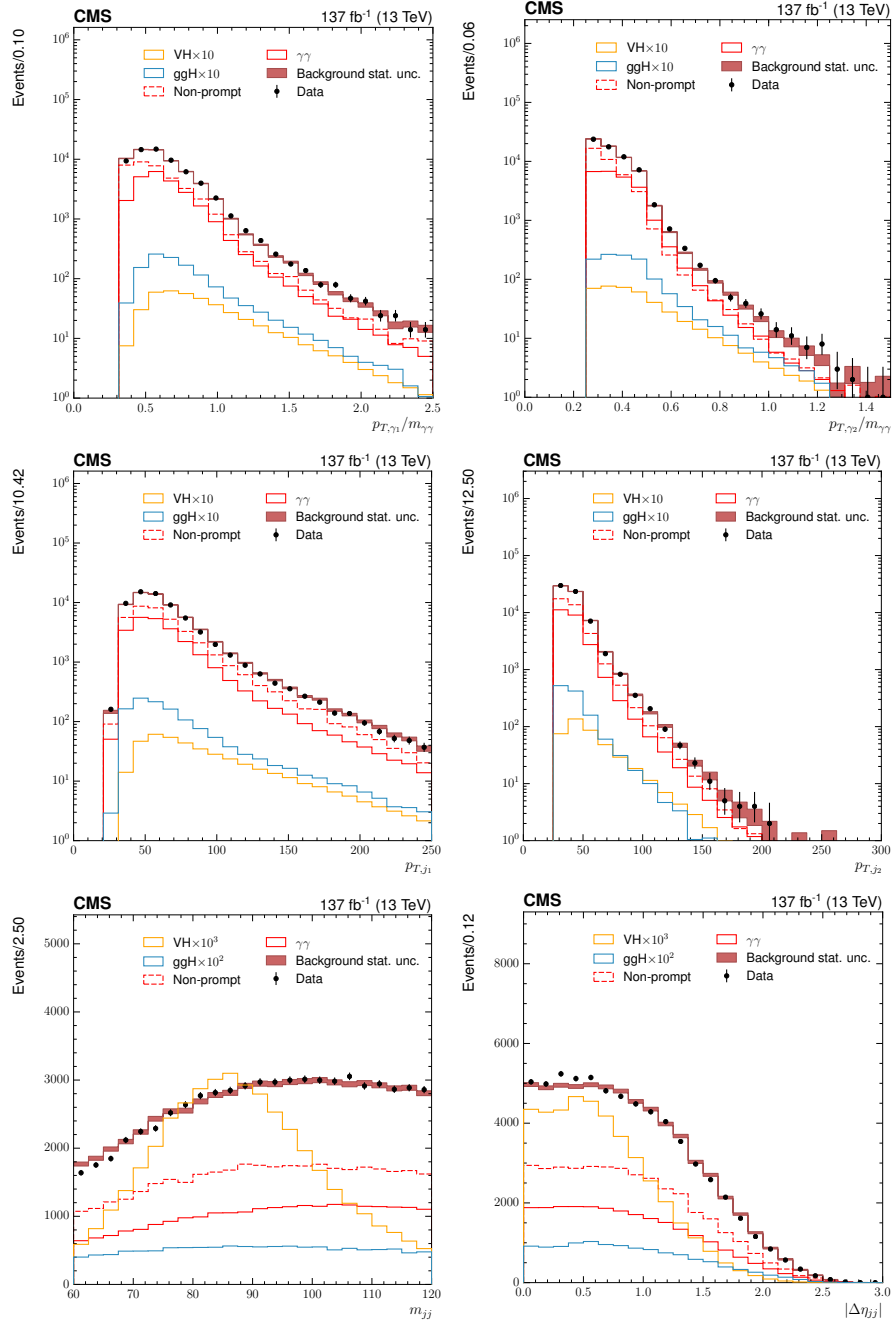


Figure 4.23: Comparison between data and simulation/data-driven samples of the photon and jet associated distributions with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the VH Hadronic category. The $\gamma\gamma$ non-resonance and (non-prompt) data-driven components are expressed by two overlapping histograms with the red solid and dashed lines. The total background and its statistical uncertainty combining these two background components are described with a brown solid line and box, respectively. Two overlapping histograms with the yellow and blue solid lines represent the samples simulating the $V(\text{had.})H$ and ggH processes with $m_H = 125$ GeV, respectively, rescaled 10-1000 times than the expected yield for visibility.

4.2.1 Boost decision tree technique

A common decision tree is a binary classifier with some selection rules encoded as a tree structure in order to separate a dataset into two classes. For data analysis in high energy physics, the decision tree is usually used to classify the targeted signal against background. Suppose there is a tree and a HEP dataset. The dataset starts from so-called *root node* in the tree and a selection rule given by some physical quantity then split the dataset to next two nodes with different phase spaces. The iteration continues until it arrives final nodes, called *leaf nodes*, which are labeled as either signal or background. Finally, the dataset is split to the signal and background classes. A diagram of a simple tree structure is shown in Figure 4.24. Of course, this doesn't mean that events falling in the signal leaf nodes must be signal. There is still a misclassification rate which depends on the decision tree training.

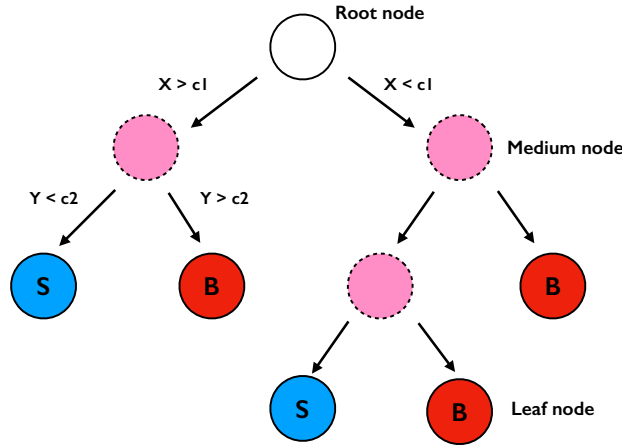


Figure 4.24: The decision tree starts from the root node (white node) and the root node is split by a selection criterion, i.e. $X > c1$ and $X < c1$, to next two nodes. The split will continue until the leaf nodes are labeled as signal "S" (blue nodes) and background "B" (red nodes). Those which can be split to next two nodes are called "medium nodes" (purple nodes).

However, a decision tree training is easy to create overtraining, namely, bias to a special phase space due to a huge and complex structure of the tree. This could reduce the performance of this decision tree in the application step. To solve the overtraining issue, the decision tree is modified through *boosting* technique. The concept of the boosting [78] is to construct many weak and simple classifiers and combine these classifiers to form a strong classifier. Besides, during the boosting training, instead of producing all weak classifiers at the same time, the boosting algorithm usually uses *additive training*. Namely, each training iteration still keeps existing weak classifiers and adds a new weak classifier on the top of them. This new weak classifier is always to reduce the residual between truth values and predicted values given by the sum of existing weak classifiers. This implies that the construction of new weak classifiers, in fact, depends on existing ones. Hence, the boosting approach not only enhances the classification power but also prevent overtraining due to complex structure. Applying the boosting

algorithm to the decision tree will create a more robust classifier called as *Boosted Decision Tree* (BDT) by constructing many weak and mutually related trees with low complexity and summing them together, described by Figure 4.25.

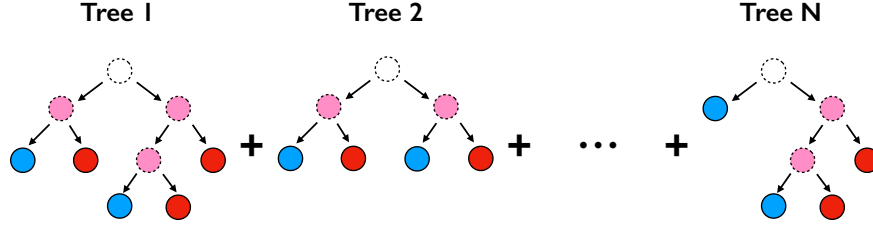


Figure 4.25: Schematic of the Boosted Decision Tree. There are N weak trees combined into a BDT.

In fact, there are many kinds of boosting algorithms used in data analysis. For the VH analysis, the *Adaboost* and *XGBoost* algorithms are performed extracting the targeted VH STXS bins against other events in the VH leptonic categories and VH Hadronic category, respectively. The following section will introduce these two algorithms.

AdaBoost

The AdaBoost algorithm is one of the most classic BDT algorithms which is supported by many different machine learning packages. In the VH analysis, this algorithm is provided by the TMVA toolkit ⁶ [79] which is widely used in high energy physics. Conceptually, suppose there is a training dataset with n events, each of which carries a set of features, $\mathbf{x}_i$, labeled by a true value, either $y_i = 1$ as signal or $y_i = -1$ as background, where i specifies the i -th event. And the k -th tree can be parametrized by $f_k(\mathbf{x}_i) = \{-1, 1\}$ where the i -th event $\mathbf{x}_i$ is classified as signal ($f_k(\mathbf{x}_i) = 1$) or background ($f_k(\mathbf{x}_i) = -1$). Namely, each leaf node in each tree only carries two binary-like values, i.e. 1 and -1 . Before the first decision tree starts training, a weight is initially assigned to each event for the AdaBoost algorithm. In general, this weight can be given from a weighted dataset or be assigned to $1/N$ for an unweighted dataset where N is the number of events. Then, the next tree is trained by the weight-modified events where the weights of the previously misclassified events are multiplied by a common *boost weight* α , defined by

$$\alpha = \frac{1 - \text{err}}{\text{err}} \quad (4.6)$$

where "err" is the misclassification rate ⁷ of the previous tree. In other words, when learning next tree (let's say the $(k + 1)$ -th tree), the event weight is modified by the boost weight α as

$$W_i^{(k+1)} = \begin{cases} W_i^{(k)} \cdot \alpha & \text{if } y_i \neq f_k(\mathbf{x}_i), (\text{misclassified}) \\ W_i^{(k)} & \text{if } y_i = f_k(\mathbf{x}_i), (\text{classified}) \end{cases} \quad (4.7)$$

⁶Although the AdaBoost algorithm for each machine-learning package, in fact, has slightly different, the concept is the same.

⁷The misclassification rate is always less than 0.5 as 0.5 is a random state.

where the sum of $W_i^{(k+1)}$ is normalized to 1. It implies those misclassified events in the current learning iteration are highlighted more in the next learning iteration. Besides, for some cases, the performance can be further enhanced by slowing down the speed of learning to allow more steps during training. In these cases, the boost weight α is transformed to $\alpha \rightarrow \alpha^\beta$ where β is considered as a *learning rate* which is less than 1 for "slow learning mode" in the AdaBoost algorithm.

Once the learning is finished due to satisfying some conditions like maximal number of trees, an output score for the i -th event $\mathbf{x}_i$ is computed by a sum of all single tree classifiers as

$$\hat{y}_i = \frac{1}{N_{\text{tree}}} \sum_{k=1} \ln(\alpha_k) \cdot f_k(\mathbf{x}_i) \quad (4.8)$$

where a small (large) $\hat{y}_i$ indicates a background-like (signal-like) event. Through this method, the learning not only concentrates on the classification of those misclassified events from the previous iteration but reduces the importance of those trees with large misclassification rates at the same time.

On the other hand, a tree construction in the AdaBoost algorithm relies on *greedy algorithm* which chooses the optimal solution at the current step. Namely, a tree structure is built node by node instead of searching for the optimal structure in all possible tree structures which is inefficient. Therefore, starting from the root node, the learning first picks up all possible features and corresponding cuts to build all possible rules of the splitting. For a given rule, the root node is split to the "right" node and "left" node with the event sets I_R and I_L , respectively. The AdaBoost algorithm then computes a indicator called as *Gini Index* of each node with two classes, signal and background, defined by

$$\text{Gini} \sim p(1 - p) \quad (4.9)$$

where p is the purity of a given node as the maximal purity = 0.5 is completely mixed. Finally, the AdaBoost algorithm calculates the information gain of the right node and left node with respect to the root node, defined by

$$\text{Gain} = \text{Gini}(I) - \text{Gini}(I_R) - \text{Gini}(I_L) \quad (4.10)$$

and picks up the splitting with the maximal gain as a new rule between the root node and the next two nodes. The iteration continues to next nodes until the information gain gets negative value or the termination conditions are satisfied like the maximal depth or minimal node size which is the percentage of the dataset in some node with respect to the overall dataset for learning.

XGBoost

Differently from the AdaBoost algorithm, each leaf node in the XGBoost algorithm [80] carries a various weight w instead of simple 1 or -1. Suppose there is a training dataset with n events,

each of which carries a set of features $\mathbf{x}_i$ and a true value y_i ⁸. For the i -th event $\mathbf{x}_i$, an output score under a given XGBoost classifier can be computed by

$$\hat{y}_i = \sum_{k=1} f_k(\mathbf{x}_i) \quad (4.11)$$

where f_k itself represents the k -th tree and $f_k(\mathbf{x}_i)$ can be parameterized by $f_k(\mathbf{x}_i) = w_{q(\mathbf{x}_i)}^{(k)}$ where $q(\mathbf{x}_i)$ is the leaf node index for the $\mathbf{x}_i$.

The main goal of the XGBoost algorithm is to reduce the residual difference between the predicted value $\hat{y}_i$ given by the existing trees and the true value y_i as much as possible by adding a new tree with optimized weights to Eq. (4.11). To express the residual difference between the predicted value $\hat{y}_i$ and true value y_i , the XGBoost algorithm defines a regularized objective function:

$$Obj = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \quad (4.12)$$

which is composed of a differentiable loss function L for each event and a regularization term Ω for each tree. For the XGBoost, the default loss function for classification is a logarithmic loss function. To avoid the overtraining issue, this added regularization term as a penalty of the model contains the complexity of the tree structure defined by

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (4.13)$$

where T is the number of leaves and w_j is the weights of all leaf nodes in the k -th tree; γ and λ are adjustable hyperparameters. Now suppose there is a new tree (let's say the $(k+1)$ -th tree) with known structure added into the trees. To optimal the weights in the $(k+1)$ -th tree which let the prediction more close to the true value, we need to minimize the objective function with the $(k+1)$ -th tree added. Therefore, Eq. (4.12) can be modified and simplified by Taylor expansion by ignoring high-order terms as

$$\begin{aligned} Obj^{(k+1)} &= \sum_{i=1}^n L(y_i, \hat{y}_i^{(k)} + f_{k+1}(x_i)) + \Omega(f_{k+1}) \\ &= \sum_{i=1}^n [L(y_i, \hat{y}_i^{(k)}) + g_i f_{k+1}(x_i) + \frac{1}{2} h_i f_{k+1}^2(x_i)] + \Omega(f_{k+1}) \end{aligned} \quad (4.14)$$

where $g_i = \partial_{\hat{y}_i^{(k)}} L(y_i, \hat{y}_i^{(k)})$ and $h_i = \partial_{\hat{y}_i^{(k)}}^2 L(y_i, \hat{y}_i^{(k)})$ are the first and second order gradient statistics of the loss function. Since the events falling on the same leaf node have the same leaf weight, we can remove the constant term which is useless and transform Eq. (4.14) in the event

⁸Since the XGBoost allows multiclass classification, the y_i can be multiple. However, in the signal and background separation, the y_i is usually chosen by tow binary-like values as $\{-1, 1\}$ or $\{0, 1\}$.

base to Eq. (4.15) in the leaf node base.

$$\begin{aligned}
 Obj^{(k+1)} &= \sum_{i=1}^n [g_i f_{k+1}(x_i) + \frac{1}{2} h_i f_{k+1}^2(x_i)] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \\
 &= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T
 \end{aligned} \tag{4.15}$$

Finally, minimize Eq. (4.15) and we can get the optimal weight w_j^* of the leaf j by

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{4.16}$$

and put Eq. (4.16) back to Eq. (4.15) as

$$Obj^{(k+1)} = - \frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \tag{4.17}$$

The above weight optimization is, in fact, derived for a known tree structure. However, in reality, it's impossible to know the exact tree structure. Therefore, similarly to the AdaBoost algorithm, the greedy algorithm is also applied for the tree construction in the XGBoost algorithm. Given a root node, the XGBoost algorithm first picks up all possible features and corresponding cuts as the rule of the splitting. For a given rule, the root node is split to "right" node and "left" node with the event sets, I_R and I_L , respectively. Let $I = I_R + I_L$, we can get the loss reduction after the split by Eq. (4.17) as

$$\begin{aligned}
 L_{\text{split}} &= Obj(I) - Obj(I_R) - Obj(I_L) \\
 &= \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} g_i \right)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{\left(\sum_{i \in I_R} g_i \right)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{\left(\sum_{i \in I} g_i \right)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma
 \end{aligned} \tag{4.18}$$

Finally, the XGBoost algorithm find out the rule with the largest loss reduction as a new rule between the root node and the next two nodes. The tree can grow until it touches given thresholds.

4.2.2 VH Leptonic BDT

In the VH analysis, three VH Leptonic categories are performed using the AdaBoost algorithm to increase the sensitivity of the STXS measurements. Since the training of these BDT classifiers always relies on simulated samples, samples well simulating background in data can help the classifiers obtain better performance for signal and background classification in data. Section 4.1.1-4.1.3 have shown the simulated samples describe data well in three VH Leptonic categories by either applying tighter selections or using dedicated data-driven methods. Hence,

this confirms the feasibility of the AdaBoost training in three VH Leptonic categories.

The AdaBoost classifier for each VH Leptonic category is named as ZH Leptonic BDT for the ZH Leptonic category, WH Leptonic BDT for the WH Leptonic category and VH MET BDT for the VH MET category. Their trainings adopt different configurations according to different situations of three categories such as different lepton final states. In the training samples, the samples simulating the associated VH production with the W and Z boson leptonic decay are considered as signal. In order to gain more statistics, the AdaBoost trainings collect five different signal samples corresponding to five Higgs boson mass points ($m_H = 123, 124, 125, 126, 127$ GeV). However, the sample with $m_H = 125$ GeV is required to use Powheg generator to avoid biasing the signal modelling which uses Madgraph5_aMC@NLO generator, while the other mass points keep the Madgraph5_aMC@NLO samples as usual. As for background, common background processes of the $H \rightarrow \gamma\gamma$ analysis such as the $\gamma + \text{jets}$ and non-resonance $\gamma\gamma$ processes are used. Moreover, Top, diboson production and $V + \gamma$ processes with lepton final states are also considered as background in the VH Leptonic categories. In particular, the VH MET category adopts the data-driven sample to replace the QCD multi-jets and $\gamma + \text{jets}$ samples with poor description. Some Higgs production processes such as the ggH , VBF and $t\bar{t}H$ not targeted by the VH Leptonic categories are also regarded as background during the training. This can avoid large theoretical dependence from strong correlation between targeted and non-target STXS bins in a single category. Table 4.3 lists the details of all training samples are used in three AdaBoost classifiers corresponding to three VH Leptonic analysis categories.

Type	Process	VH MET BDT	WH Leptonic BDT	ZH Leptonic BDT
Signal				
V(leptons)V	$Z(\rightarrow \nu\nu)H$	✓		
	$W(\rightarrow \ell\nu)H$	✓	✓	
	$Z(\rightarrow \ell\ell)H$			✓
Background				
Prompt/Fake γ	$\gamma\gamma(+\text{jets})$	✓	✓	✓
	$\gamma + \text{jets}$		✓	
	Data driven	✓		
Drell-Yan	$Z(\rightarrow \ell\ell) + \text{jets}$	✓	✓	✓
V + γ	$Z(\rightarrow \nu\nu) + \gamma$	✓		
	$W(\rightarrow \ell\nu) + \gamma$	✓	✓	
	$Z(\rightarrow \ell\ell) + \gamma$	✓	✓	✓
VV	WW, WZ, ZZ	✓	✓	✓
Top quark	$t\bar{t} + X(X = \gamma, \gamma\gamma, \text{jets})$	✓	✓	✓
SM Higgs	ggH	✓	✓	
	VBF	✓	✓	
	$t\bar{t}H$	✓	✓	✓
	$Z(\rightarrow \ell\ell)H$	✓	✓	
	$V(\rightarrow jj)H$	✓	✓	

Table 4.3: List of all training samples for the VH MET BDT, WH Leptonic BDT, ZH Leptonic BDT. The tick mark represents that the sample is used in the corresponding category.

On the other hand, although each category has different kinds of final states, the choice of input features can be summarized in the following observations:

- The diphoton system from the Higgs boson decay has higher energy scale with respect to the non-resonance processes.
- The charged lepton(s) and missing energy originating from the W and Z boson decay in the VH production tend to higher energy scale with respect to the non-resonance backgrounds and some Higgs boson productions such as the VBF, ggH .
- No QCD activity within the VH production process causes lack jet multiplicity.

- No b -quark final state for the VH production with the W and Z boson leptonic decay gives no b -jet object in events.
- For the VH kinematic topology, the momentum of the W/Z boson and Higgs boson from VH production tends to the same scale but opposite direction.

According to these observations, the AdaBoost classifiers in the VH Leptonic categories include the kinematic variables of the corresponding final states to achieve the best signal-and-background separation power. For example, including the kinematic variables of two photons, leading charged lepton and missing energy can kill the fake photon and fake charged lepton backgrounds in the WH Leptonic category. Furthermore, jet and VH kinematic topology associated variables are taken into account in the AdaBoost trainings for removing the top-quark relative backgrounds and non-target Higgs productions like the ggH , VBF , ttH and $V(\text{had.})H$. The complete description of the input features for three AdaBoost classifiers is shown in Table 4.4.

In each VH Leptonic category, all training samples are normalized to their cross sections and given luminosity. Due to the limited events of the VH production with the Z and W boson decaying leptonically, it's expected that the statistical uncertainties dominate the final results rather than the experimental uncertainties arising from different configurations and the CMS detector status in 2016-2018 years. From this reason, instead of preparing an AdaBoost classifier for each of the years separately, the training combines all three-year simulated samples together to generate an inclusive AdaBoost classifier for each VH Leptonic category. For the training architecture, the weights of signal events are scaled such that the sum of weights of all signal events is equal to the sum of weights of all background events in order to prevent imbalanced classes from reducing the performance. Furthermore, to implement the overtraining test, each training splits the samples to training and test ones with a ratio of 0.5. The splitting uses alternate event extraction to prevent some low-statistics samples from inhomogeneously arranged to the training and test samples. Simultaneously, events in the WH Leptonic and ZH Leptonic categories are divided into the electron and muon classes which are expected to have different performance in the CMS detector. On the other hand, the hyperparameter tuning is also performed for achieving the optimized performance. The tuning scans the space composed of the most sensitive hyperparameters in the AdaBoost algorithm like number of trees, maximal depth and minimal node size to look for the best ROC curve. Since the training samples for the ZH Leptonic BDT have few statistics, the learning rate is tuned to 0.5, while the others are 0.1. The optimized hyperparameters for three AdaBoost classifiers are shown in Table 4.5.

Once the hyperparameter optimization is finished, the training results give three AUCs of 0.960, 0.912, 0.951 corresponding to the VH MET BDT, WH Leptonic BDT and ZH Leptonic BDT, respectively, giving strong separation power for the VH leptonic events against the background events in the VH Leptonic categories. There is also no significant overtraining as the BDT score distributions and ROC curves of the training samples are compatible to those of the test samples for three VH Leptonic categories, shown in Figure 4.26, 4.27, 4.28. In fact, the $gg \rightarrow ZH$ process is not included in the ZH Leptonic BDT and VH MET BDT training due to limited schedule. However, after applying the BDT classifiers to this simulation, Figure 4.29 checks that the ZH

Table 4.4: The list of all training variables for the VH MET BDT, WH Leptonic BDT, ZH Leptonic BDT. The tick mark represent the variable is used in the corresponding category.

Type	Input Variable	VH MET BDT	WH Leptonic BDT	ZH Leptonic BDT
Photon	$p_T^{\gamma 1(\gamma 2)} / m_{\gamma\gamma}$	✓	✓	✓
	$\eta^{\gamma 1(\gamma 2)}$	✓	✓	✓
	$\cos(\Delta\phi(\gamma\gamma))$	✓	✓	✓
	Max. γ ID BDT score	✓	✓	✓
	Min. γ ID BDT score	✓	✓	✓
	Pixel seed veto ($\gamma 1, \gamma 2$)		✓	
Lepton	$p_T^{\ell 1}$		✓	✓
	$p_T^{\ell 2}$			✓
	$\eta^{\ell 1}$		✓	✓
	$\eta^{\ell 2}$			✓
	$\cos(\Delta\phi(\ell\ell))$			✓
	$\Delta R(\gamma 1(2), \ell 1)$		✓	✓
	$\Delta R(\gamma 1(2), \ell 2)$			✓
	$m_{\ell\ell}$			✓
Missing energy	p_T^{miss}	✓	✓	
	SumEt (p_T^{miss}) ^a	✓	✓	
	$m_T^{\ell\nu}$		✓	
Jet	Jet multiplicity	✓	✓	✓
	p_T^{jet}	✓	✓	✓
	Max. DeepCSV score	✓	✓	✓
VH topology	$\Delta\phi(\gamma\gamma, \vec{p}_T^{\text{miss}})$	✓		
	$\Delta\phi(\gamma\gamma, \ell\nu)$		✓	
	$\Delta\phi(\gamma\gamma, \ell\ell)$			✓
	$\text{Min}(\Delta\phi(\vec{p}_T^{\text{miss}}, \text{jets}))$ ^b	✓		
	$\text{Min}(\Delta\phi(\ell\nu, \text{jets}))$ ^c		✓	
	$p_{T, \text{balance}}$ ^d	✓	✓	✓

^a Scalar sum of E_T of the Particle-Flow candidates which are used to calculate $\vec{p}_T^{\text{miss}}$.

^b Minimum of the azimuthal angle between the $\vec{p}_T^{\text{miss}}$ and one of the jets.

^c Minimum of the azimuthal angle between the $\ell + \nu(\vec{p}_T^{\text{miss}})$ and one of the jets.

^d $(p_T^{\gamma\gamma} - p_T^{\text{miss}}) / p_T^{\gamma\gamma}$ for the VH MET BDT and $(p_T^{\gamma\gamma} - p_T^{\ell\nu}) / p_T^{\gamma\gamma}$ for the WH Leptonic BDT.

Category	Trees	Max. Depth	Min. Node Size	Learning rate
VH MET BDT	600	8	5.0%	0.1
WH Leptonic BDT	500	5	5.0%	0.1
ZH Leptonic BDT	160	7	4.5%	0.5

Table 4.5: Optimized hyperparameters for three AdaBoost classifiers.

Leptonic BDT and VH MET BDT distributions in the $gg \rightarrow ZH$ process are similar or more close to the signal-like region (close to 1) than those in the $q\bar{q} \rightarrow ZH$ process. Hence, these two classifiers still keep the sensitivity of the $gg \rightarrow ZH$ process.

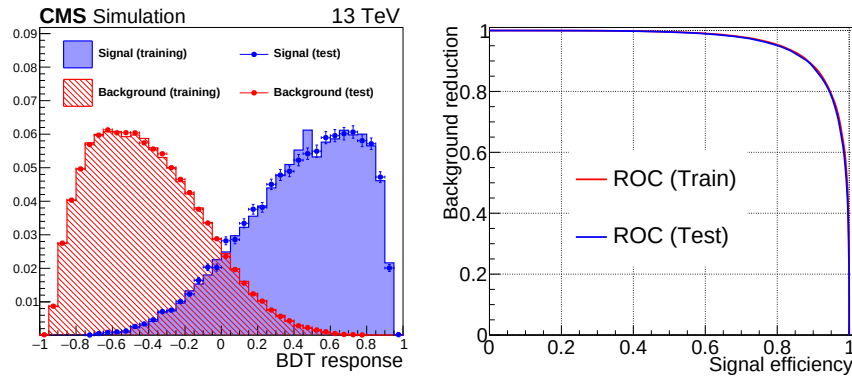


Figure 4.26: The comparison of the VH MET BDT performance between the training and test sample in terms of the BDT score distributions (left) and the ROC curves (right). In Figure (left), the histogram style is the training sample; the dot point is test one; the red color and blue one are signal and background, respectively. In Figure (right), the red and blue curves are the training and test with the corresponding AUC of 0.960.

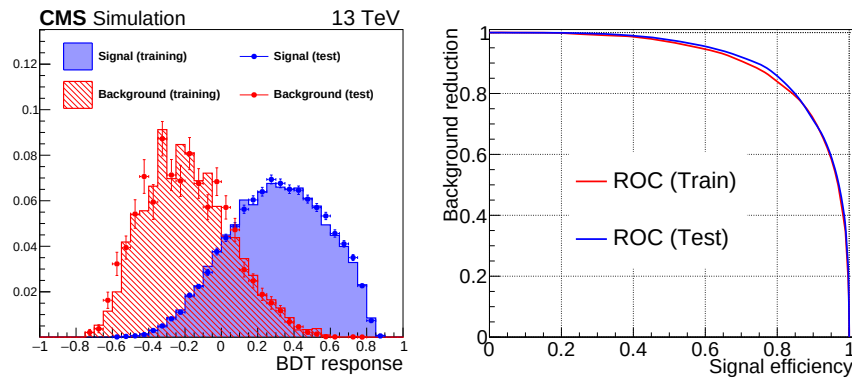


Figure 4.27: The comparison of the WH Leptonic BDT performance between the training and test sample in terms of the BDT score distributions (left) and the ROC curves (right). In Figure (left), the histogram style is the training sample; the dot point is test one; the red color and blue one are signal and background, respectively. In Figure (right), the red and blue curves are the training and test with the corresponding AUC of 0.912.

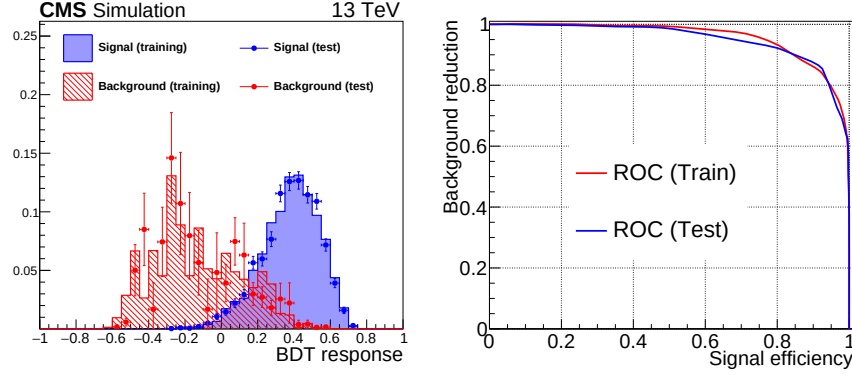


Figure 4.28: The comparison of the ZH Leptonic BDT performance between the training and test sample in terms of the BDT score distributions (left) and the ROC curves (right). In Figure (left), the histogram style is the training sample; the dot point is test one; the red color and blue one are signal and background, respectively. In Figure (right), the red and blue curves are the training and test with the corresponding AUC of 0.951.

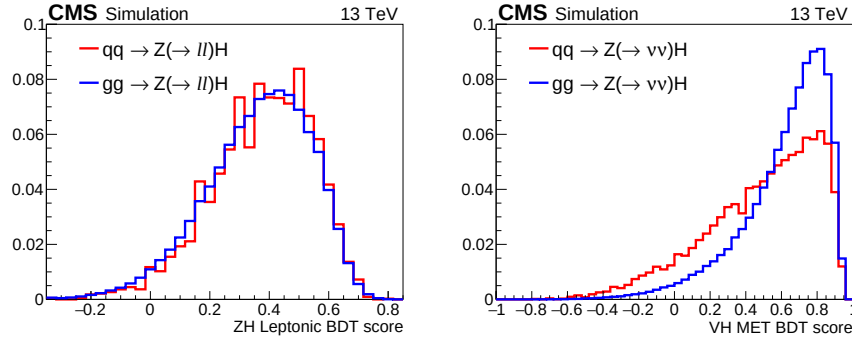


Figure 4.29: The comparison between the $q\bar{q} \rightarrow ZH$ (red histogram) and $gg \rightarrow ZH$ (blue histogram) processes with $m_H = 125$ GeV in terms of the ZH Leptonic BDT (left) and VH MET BDT (right) distributions. All histograms are normalized to 1.

After applying three AdaBoost classifiers to the corresponding categories, Figure 4.30 shows the comparison between data and simulation samples in terms of the VH MET BDT and WH Leptonic BDT distributions is compatible, removing the $m_{\gamma\gamma} \in [120, 130]$ GeV. Although the ZH Leptonic BDT distribution is inconsistent between data and simulated samples due to the limited statistics, since the inconsistency occurs in the background-like part of the distribution, this does not have any significant impact on the training performance.

4.2.3 VH Hadronic BDT

Signal extraction in the VH Hadronic category is performed using the XGBoost algorithm for the signal and background classification. However, instead of using a single BDT classifier like those in the VH Leptonic categories, there are two XGBoost classifiers, i.e. diphoton BDT and VH Hadronic BDT, which aim at different kinds of background processes. In Section 4.1.4, it

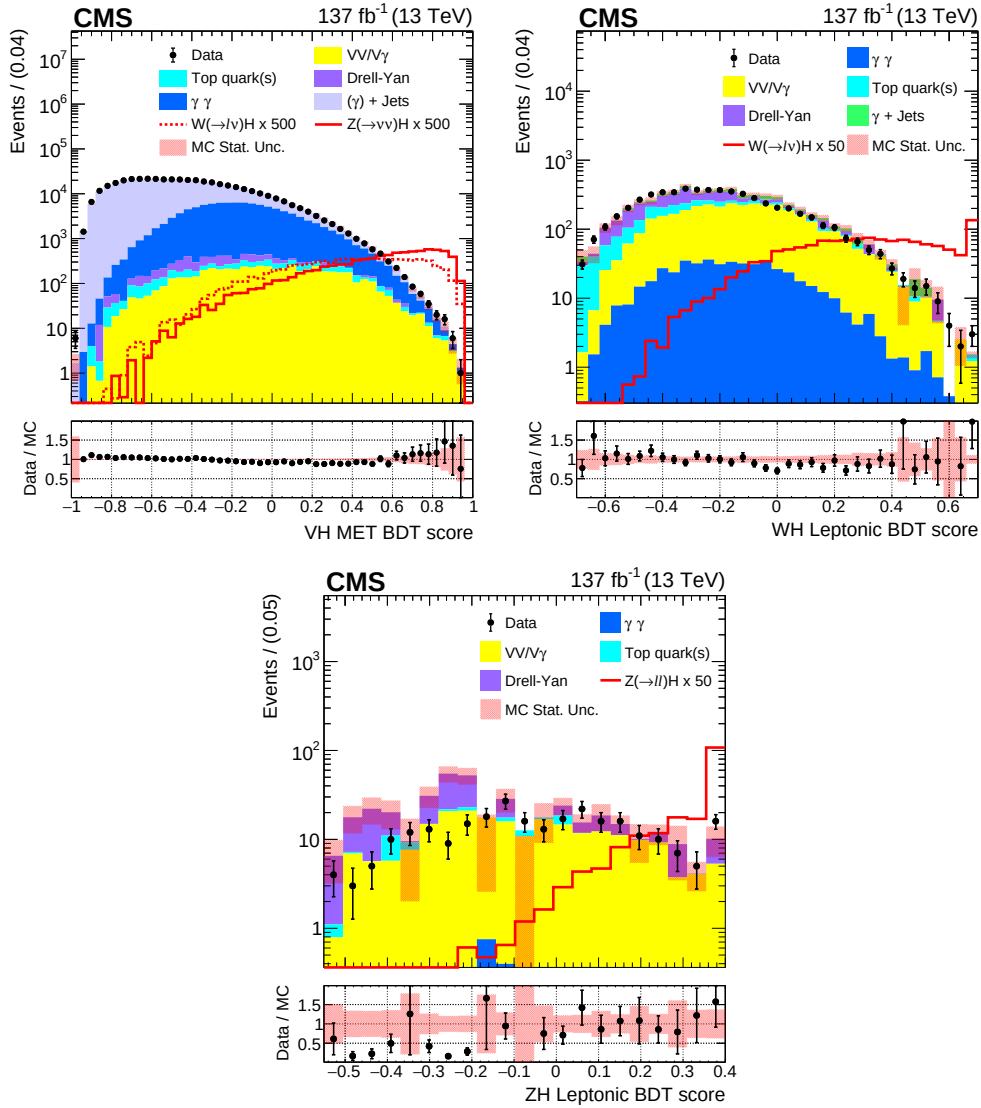


Figure 4.30: Data-and-simulation comparison of the distributions of VH MET BDT (top left), WH Leptonic BDT (top right) and ZH Leptonic BDT score (bottom) with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV in the VH MET category. The different background components are stacked with different color and solid histograms. The corresponding statistical uncertainty given by background samples is also shown with pink boxes. The histograms with red lines in these three figures represent the samples simulating $q\bar{q}(gg) \rightarrow Z(\rightarrow \nu\nu)H$, $W(\rightarrow \ell\nu)H$ and $q\bar{q}(gg) \rightarrow Z(\rightarrow \ell\ell)H$ processes with $m_H = 125$ GeV. Additionally, $W(\rightarrow \ell\nu)H$ process with $m_H = 125$ GeV made by a histogram with a red dashed line is also overlapped in the bottom figure. All those VH processes are rescaled 50 times (WH Leptonic BDT and ZH Leptonic BDT) and 500 times (VH MET BDT) than the expected yields for visibility. The bottom panel in each figure is the data-to-simulation ratio.

has been confirmed that the background simulated samples in the VH Hadronic category have ability to model the background events in data by applying the data-driven method and the XGBoost classifiers are thus permitted in this category. The following sections will individually introduce these two XGBoost classifiers.

Diphoton BDT

The main goal for the Diphoton BDT, in fact, is to reduce the common background contributions in the $H \rightarrow \gamma\gamma$ analysis, namely, fake photon backgrounds like the QCD multi-jets and γ + jets events or prompt photon background with lower energy scale like the non-resonance $\gamma\gamma$ background. Therefore, the samples simulating four main Higgs boson productions, i.e. the ggH , VBF , VH and ttH productions, with $m_H = 125$ GeV are cross-section weighted and mixed together as signal. These signal samples are simulated with the Powheg generator to avoid any bias in the later signal modelling. The QCD multi-jets, γ + jets and non-resonance $\gamma\gamma$ simulation samples are considered as background with cross-section weighted. Only one inclusive classifier by combining all 2016-2018 simulation samples is produced⁹. All samples are required to pass the diphoton selections, described in Section 3.3.5, before the diphoton BDT training.

In order to extract events with signal-like kinematics, prompt-like photons and even better diphoton mass resolution, the diphoton BDT classifier adopts the kinematic variables of the diphoton system, photon ID BDT score and a per-event estimate of the diphoton mass resolution as the input features, respectively. These variables are designed as dimensionless ones in order to avoid the choice of the diphoton BDT score biasing the diphoton mass. The following list shows the details of the variables used to the classifier :

- the transverse momenta for both photons, rescaled for the diphoton mass, $p_T^{1,(2)} / m_{\gamma\gamma}$;
- the pseudorapidities of both photons, $\eta^{1,(2)}$;
- the cosine of the angle between the two photons in the transverse plane, $\cos(\Delta\phi)$;
- the identification BDT score for both photons;
- the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using the correct primary vertex (σ_{rv});
- the per-event relative mass resolution estimate, under the hypothesis that the mass has been reconstructed using an incorrect primary vertex (σ_{wv});
- the per-event probability estimate that the correct primary vertex has been used to reconstruct the mass, based on the event-level vertex selection MVA (p_{vtx}).

⁹It's confirmed that the individual BDT training with year split reduces 1% performance with respect to the single classifier inclusively.

According to Section 3.3, under the assumption that the correct vertex, i.e. $|z_{\text{Higgs vtx.}} - z_{\text{selected vtx.}}| < 1$ cm, has been selected for the diphoton mass reconstruction, only the photon energy resolution contributes the diphoton mass resolution. The per-event relative mass resolution estimate in this case is computed from the propagation of the photon energy resolution estimates as

$$\sigma_{rv} = \frac{1}{2} \sqrt{\left(\frac{\sigma_{E1}}{E_1}\right)^2 + \left(\frac{\sigma_{E2}}{E_2}\right)^2} \quad (4.19)$$

where σ_E is the energy resolution of photons given by the photon energy regression, described in Section 3.3.2. However, for the wrong vertex assumption, i.e. $|z_{\text{Higgs vtx.}} - z_{\text{selected vtx.}}| > 1$ cm, an additional angle resolution arising from the wrong diphoton vertex choice significantly contributes the diphoton mass resolution. The per-event relative mass resolution estimate in the wrong vertex case is thus computed as

$$\sigma_{wv} = \frac{1}{2} \sqrt{\left(\frac{\sigma_{E1}}{E_1}\right)^2 + \left(\frac{\sigma_{E2}}{E_2}\right)^2 + \left(\frac{\sigma_\alpha}{\tan(\alpha/2)}\right)^2} \quad (4.20)$$

where σ_α is analytically estimated by the opening angle difference from the Higgs boson interaction vertex to the selected vertex, whose z coordinate distance follows a Gaussian distribution with a width of $\sqrt{2}\sigma_Z^{\text{beamspot}}$ ¹⁰.

During the diphoton BDT training, the QCD multi-jets sample is down-weighted by a factor of 25 to prevent overtraining on specific values of the input variables¹¹. The weight w_{sig} given by Eq. (4.21) is applied to the signal events in order to correct the BDT score which is inversely proportional to the mass resolution.

$$w_{\text{sig}} = \frac{p_{vtx}}{\sigma_{rv}} + \frac{1 - p_{vtx}}{\sigma_{wv}} \quad (4.21)$$

The weights of the signal events are also scaled such that the sum of weights of all signal events is equal to the sum of weights of all background events in order to avoid having imbalanced classes. To perform the overtraining test, the training splits the samples to 70% for training and 30% for test ones. A range of hyperparameters of the BDT are also studied in order to maximize the performance. It's observed that the learning rate and maximum tree depth, and L1 and L2 regularisation parameters significantly change the performance and Table 4.6 shows the results of the optimized hyperparameters without overtraining. Finally, the performance of the training gives a ROC curve with an AUC value of 87% which is the same as that of the test, shown in Figure 4.31 (right). After applying the diphoton BDT to overall diphoton events, Figure 4.31 (left) shows the comparison between data and simulation/data-driven samples is agreement, removing the $M_{\gamma\gamma} \in [120, 130]$ GeV.

¹⁰The assumption is based on the $Z \rightarrow \mu\mu$ validation.

¹¹Since these events tend to occupy the background-like part of the output distribution, this does not have any significant impact on the training performance.

	Trees	Max. Depth	Learning rate	L1 regularization	L2 regularization
Diphoton BDT	600	6	0.3	0	1

Table 4.6: Optimized hyperparameters for the diphoton BDT classifiers.

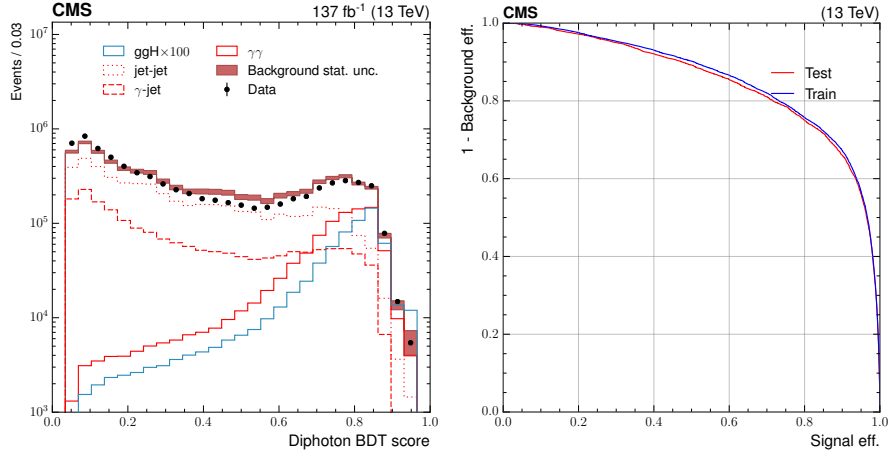


Figure 4.31: The left is the comparison between data and simulation/data-driven samples of the diphoton BDT distribution with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV. The non-resonance $\gamma\gamma$ and (non-prompt) data-driven components are expressed by two overlapping histograms with the red solid and dashed. Total background and its statistical uncertainty combining these two background components are described with a brown solid line and box, respectively. The histogram with the blue solid line represent the samples simulating the ggH process with $m_H = 125$ GeV, respectively, rescaled 100 times than the expected yield for visibility. The right shows the ROC curve between training and test samples. It's observed that there is no significant overtraining issue and the corresponding AUC is 87%.

VH Hadronic BDT

The VH Hadronic BDT is a dedicated classifier which focuses on the VH Hadronic category. Through the jet and VH kinematic variables, it can further rejects the prompt photon background and the ggH with two additional jets which has the same final states as the VH production with the W and Z boson hadronic decay. The samples simulating the VH production with the W and Z boson hadronic decay and $m_{\gamma\gamma} = 125$ GeV by the Powheg generator are as signal, while the non-resonance $\gamma\gamma$ sample and data-driven sample for replacing the QCD multi-jets and $\gamma+$ jets samples are considered as background. Additionally, the ggH production sample with $m_H = 125$ GeV simulated by the Powheg generator is also added into background. Only one inclusive classifier by combining all 2016-2018 simulation samples is prepared. All samples have to pass the selections in the VH Hadronic category.

All input variables used to train the VH Hadronic BDT are listed in the following :

- the transverse momenta of the two leading jets ($p_T^{\text{jet1},(\text{jet2})}$);
- the pseudo-rapidity of the two leading jets ($\eta^{\text{jet1},(\text{jet2})}$);

- the di-jet invariant mass (m_{jj});
- the difference in pseudo-rapidity of the two leading jets (η_{jj});
- the transverse momenta for both photons, rescaled for the diphoton mass ($p_T^{\gamma^{1,(\gamma^2)}}/m_{\gamma\gamma}$);
- the cosine of the helicity angle between di-jet and diphoton systems ($\cos\theta^*$).

The θ^* quantity is the helicity angle defined by the angle between the motion of the diphoton system in the diphoton-dijet center-of-mass frame, and the motion of the diphoton-dijet system in the lab frame, described by Figure 4.32 (left). The distribution of this variable is rather uniform for the VH production, while it is strongly peaked at 1 for the common photon backgrounds and ggH production, shown in Figure 4.32 (right).

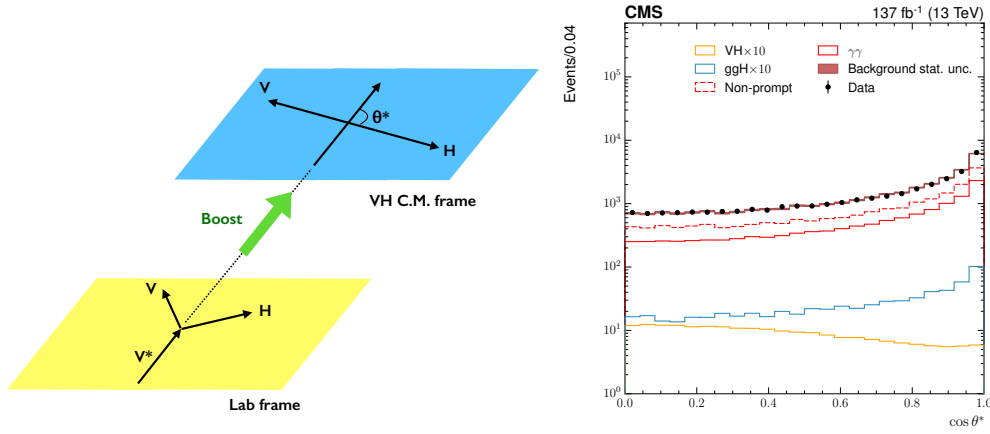


Figure 4.32: The left is the diagram for the VH production in lab frame (yellow plane) boosted to VH (or V^*) C.M. frame (blue plane) with θ^* as a helicity angle. The right is the comparison between data and simulation/data-driven samples of the cosine of the helicity angle with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV. The non-resonance $\gamma\gamma$ and (non-prompt) data-driven components are expressed by two overlapping histograms with the red solid and dashed. Total background and its statistical uncertainty combining these two background components are described with a brown solid line and box, respectively. Two overlapping histograms with the yellow and blue solid lines represent the samples simulating $V(\text{had.})H$ and ggH processes with $m_H = 125$ GeV, respectively, rescaled 10 times than the expected yield for visibility.

The training treatments like the event weights and training/test splitting and hyperparameter optimization in the VH Hadronic BDT training are the same as those in the diphoton BDT training. After the training task, the performance of the training gives a ROC curve with an AUC value of 87% which is the same as that of the test, shown in Figure 4.33 (right). Applying the VH Hadronic BDT to the VH Hadronic category, Figure 4.33 (left) shows the comparison between data and simulation/data-driven samples matches well, removing the $M_{\gamma\gamma} \in [120, 130]$ GeV.

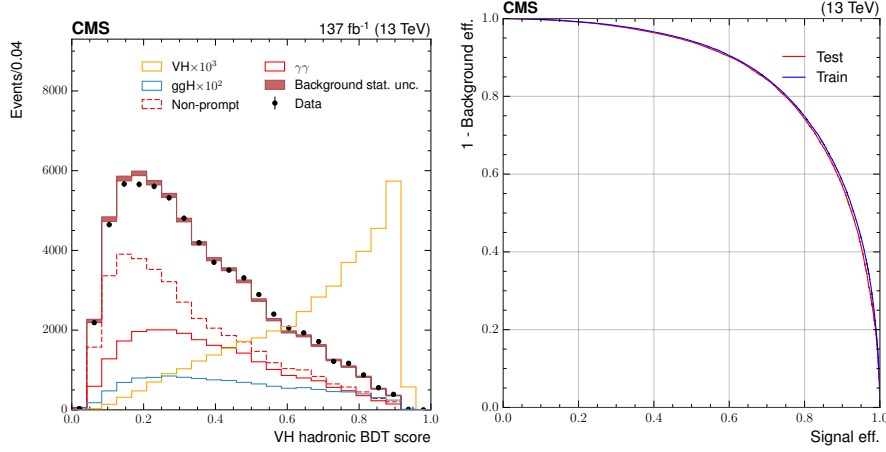


Figure 4.33: The left is the comparison between data and simulation/data-driven samples of the VH Hadronic BDT distribution with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV. The non-resonance $\gamma\gamma$ and (Non-prompt) data-driven components are expressed by two overlapping histograms with the red solid and dashed. Total background and its statistical uncertainty combining these two background components are described with a brown solid line and box, respectively. Two overlapping histograms with the yellow and blue solid lines represent the samples simulating $V(\text{had.})H$ and ggH processes with $m_H = 125$ GeV, rescaled 100-1000 times than the expected yield for visibility. The right shows the ROC curve between training and test samples. It's observed that there is no significant over-training issue and the corresponding AUC is 87%.

4.3 Categorization for the VH STXS stage 1.2

In Section 4.1, the events have been divided into four analysis categories by the final states of the W and Z bosons for enriching three VH STXS stage 0 bins such as $q\bar{q}(gg) \rightarrow Z(\text{lep.})H$, $q\bar{q}' \rightarrow W(\text{lep.})H$ and $q\bar{q} \rightarrow V(\text{had.})H$. In Section 4.2, the BDT classifiers are also prepared in each analysis category to exclude non-resonance background events and non-target STXS bins falling into the VH analysis categories for better sensitivity and less theoretical dependence. However, some non-target STXS bins could only contaminate the VH production with lower energy scale. Some targeted STXS bins could also have larger variation originating from their experimental uncertainties in some specific kinematic regions. These cause that the analysis needs to extrapolate these effects in specific phase spaces to a larger phase space. To maximally reduce these effects, the STXS bins at the stage 0 is further split to those at the stage 1.2 with finer granularity by the transverse momenta of the associated W and Z boson, p_T^V , and the jet multiplicity. Through the stage 1.2, these effects can be isolated to some specific analysis categories for easier extrapolation. Such STXS bin splitting even allows measuring the kinematic structure of the associated VH production.

In the STXS stage 1.2 of the VH production with the W and Z leptonic decay, each bin in the stage 0 is further split by $p_T^V = 75, 150$ and 250 GeV according to the recommendation of the LHC Higgs Cross Section Working Group Report [6]. In particular, the $150 \text{ GeV} < p_T^V <$

250 GeV bin is also separated into the 0-jet and ≥ 1 -jet cases. Figure 4.34 shows the schematic diagram for the VH leptonic bins at the STXS stage 1.2, corresponding cross sections shown in Table 4.7.

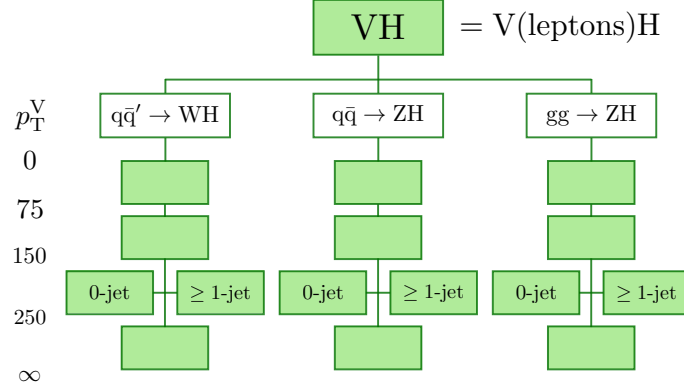


Figure 4.34: Schematic of the VH leptonic STXS stage 1.2 bins from the VH leptonic STXS stage 0 bins split by the kinematics such as the p_T^V and jet multiplicity.

On the other hand, the VH hadronic STXS bin at the stage 1.2 keeps the same as that at the stage 0. Since the VBF and ggH productions in the electroweak qqH and ggH STXS bins also move on the stage 1.2 to reduce their correlation and variation, these of which are considered as contamination by the VH hadronic STXS bin indirectly reduce the theoretical dependence of the VH measurements in the hadronic part.

To enrich the VH STXS stage 1.2 bins, the event categorization moves on the step 2 and first split the analysis categories in the step 1 by the reconstructed p_T^V . Afterwards, each analysis category will be divided into several analysis sub-categories by the corresponding BDT distribution to improve more sensitivity of the STXS measurements. Eventually, all optimized analysis sub-categories are considered as the final analysis categories in the VH analysis and propagated to the step of the statistical analysis for fitting, described in Chapter 6. However, there are two issues occurring when the VH analysis moves on the STXS stage 1.2. One is the event migration arising from that the reconstructed p_T^V is inconsistent with the particle-level one. This is discussed in Section 4.3.1. The other involves in limited statistics in the VH analysis. For some productions or decay modes of the VH processes with limited statistics, analysis categories with finer granularity might reduce the sensitivity even lower than those in the Stage 0. For this reason, some analysis categories for the STXS stage 1.2 bins need to be merged and so are the STXS bins at the stage 1.2. The bin merging and sub-category construction are shown in Section 4.3.2.

VH leptonic STXS stage 0					
Production mode	Order of σ calc	$\sigma \cdot \text{BR}$ [fb] 13 TeV, $m_H = 125$ GeV			
$qq \rightarrow W(\ell\nu)H$	NNLO(QCD)+NLO(EW)	1.016			
$qq \rightarrow Z(\ell\ell/\nu\nu)H$	NNLO(QCD)+NLO(EW)	0.520			
$gg \rightarrow Z(\ell\ell/\nu\nu)H$	NNLO(QCD)+NLO(EW)	0.084			
STXS region	Definition units of p_T^V in GeV	Fraction of total			$\sigma \cdot \text{BR}$ [fb]
		$qq \rightarrow W(\ell\nu)H$	$qq \rightarrow Z(\ell\ell/\nu\nu)H$	$gg \rightarrow Z(\ell\ell/\nu\nu)H$	
WH lep forward	$ Y_H > 2.5$	12.13%	-	-	0.123
ZH lep forward		-	11.21%	-	0.058
ggZH lep forward		-	-	2.71%	0.002
$ Y_H < 2.5$					
WH lep $p_T^V < 75$	No jet requirements, $p_T^V < 75$	46.55%	-	-	0.473
WH lep $75 < p_T^V < 150$	No jet requirements, $75 < p_T^V < 150$	29.30%	-	-	0.298
WH lep 0J $150 < p_T^V < 250$	Exactly 0 jets, $150 < p_T^V < 250$	5.10%	-	-	0.052
WH lep $\geq 1J$ $150 < p_T^V < 250$	At least 1 jet, $150 < p_T^V < 250$	3.97%	-	-	0.040
WH lep $p_T^V > 250$	No jet requirements, $p_T^V > 250$	2.95%	-	-	0.030
ZH lep $p_T^V < 75$	No jet requirements, $p_T^V < 75$	-	45.65%	-	0.237
ZH lep $75 < p_T^V < 150$	No jet requirements, $75 < p_T^V < 150$	-	30.70%	-	0.160
ZH lep 0J $150 < p_T^V < 250$	Exactly 0 jets, $150 < p_T^V < 250$	-	5.16%	-	0.027
ZH lep $\geq 1J$ $150 < p_T^V < 250$	At least 1 jet, $150 < p_T^V < 250$	-	4.27%	-	0.022
ZH lep $p_T^V > 250$	No jet requirements, $p_T^V > 250$	-	3.01%	-	0.016
ggZH lep $p_T^V < 75$	No jet requirements, $p_T^V < 75$	-	-	15.96%	0.013
ggZH lep $75 < p_T^V < 150$	No jet requirements, $75 < p_T^V < 150$	-	-	43.32%	0.036
ggZH lep 0J $150 < p_T^V < 250$	Exactly 0 jets, $150 < p_T^V < 250$	-	-	9.08%	0.008
ggZH lep $\geq 1J$ $150 < p_T^V < 250$	At least 1 jet, $150 < p_T^V < 250$	-	-	20.49%	0.017
ggZH lep $p_T^V > 250$	No jet requirements, $p_T^V > 250$	-	-	8.45%	0.007

Table 4.7: The cross section times branching ratios for the VH leptonic STXS bins, evaluated at a centre-of-mass energy of 13 TeV and $m_H = 125$ GeV. The total cross section times branching ratio is provided for each production mode, along with details on the order of the calculation. Also provided are the percentage breakdown of the production modes into the STXS bins.

4.3.1 Migration of the STXS stage 1.2 bins in the VH Leptonic categories

As the stage 1.2 has described, the VH leptonic STXS bins are split not only by production processes but also according to the transverse momentum of the vector boson with the boundaries at $p_T^V = 75, 150, 250$ GeV. Three VH Leptonic categories in the step 1 should be also split by p_T^V which is reconstructed from corresponding lepton final states, i.e. charged leptons and neutrinos. Owing to the limitation of the CMS detector, neutrinos can not be detected and reconstructed as particles. The splitting only relies on the missing transverse in each event, described in Section 3.5. However, the significant uncertainty for the missing energy originating from experimental acceptance causes that the STXS stage 1.2 bins migrate between analysis categories with the p_T^V boundaries, especially, the WH Leptonic and VH MET categories which treat the missing energy as neutrinos. To reduce the effect of the STXS bin migration, the VH

analysis uses the feature of the VH kinematic topology in the CMS detector.

For the VH kinematic topology, the vector boson and Higgs boson in the associated VH production without any additional hard parton are expected to tend to a back-to-back propagation from the interaction point in the transverse plane of the CMS detector. Therefore, the transverse momentum of the Higgs boson and W/Z boson should have similar scales but opposite directions. From this reason, it's possible to use the transverse momentum of the diphoton system, $p_T^{\gamma\gamma}$, as the boundaries of the VH Leptonic categories in the step 2. Figure 4.35 describes the comparison between $p_T^{\gamma\gamma}$ and reconstructed W or Z boson p_T quantity, i.e. p_T^{miss} for the VH MET category, $p_T^{\ell\nu}$ for the WH Leptonic category and $p_T^{\ell\ell}$ for the ZH Leptonic category, with respect to the particle-level p_T^V . It turns out that the $p_T^{\gamma\gamma}$ in the WH Leptonic category and the VH MET category is more close to the p_T^V under particle level. As for the ZH Leptonic category, $p_T^{\ell\ell}$ is expected to have better performance due to two precise charged tracks.

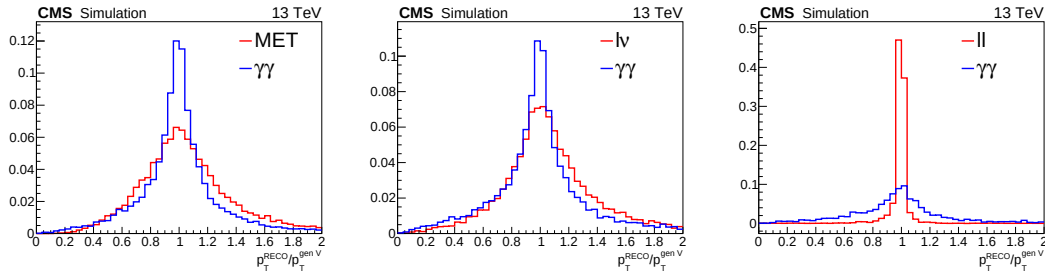


Figure 4.35: Comparison between the ratio of the reconstructed and particle-level W/Z p_T (red) and of the reconstructed diphoton and particle-level W/Z p_T (blue) in the VH MET (left) and WH Leptonic (middle) and ZH Leptonic (right) categories. All histograms are normalized to 1.

Moreover, an event migration test is also shown in Figure 4.36 for three VH Leptonic categories. Less migration for the $p_T^{\gamma\gamma}$ boundary in the VH MET category and for the $p_T^{\ell\ell}$ boundary in the ZH Leptonic category are expected, while the $p_T^{\ell\nu}$ boundary, in fact, gives slightly better splitting than the $p_T^{\gamma\gamma}$ boundary in the WH Leptonic category. However, considering the missing energy with large uncertainty with respect to $p_T^{\gamma\gamma}$, the WH Leptonic category still uses $p_T^{\gamma\gamma}$ for splitting. In the result, for the splitting of the VH analysis categories in the step 2, the WH Leptonic and VH MET categories use the transverse momentum of diphoton system, $p_T^{\gamma\gamma}$, as the boundaries, and the ZH Leptonic category is the transverse momentum of two charged and same flavour leptons, $p_T^{\ell\ell}$.

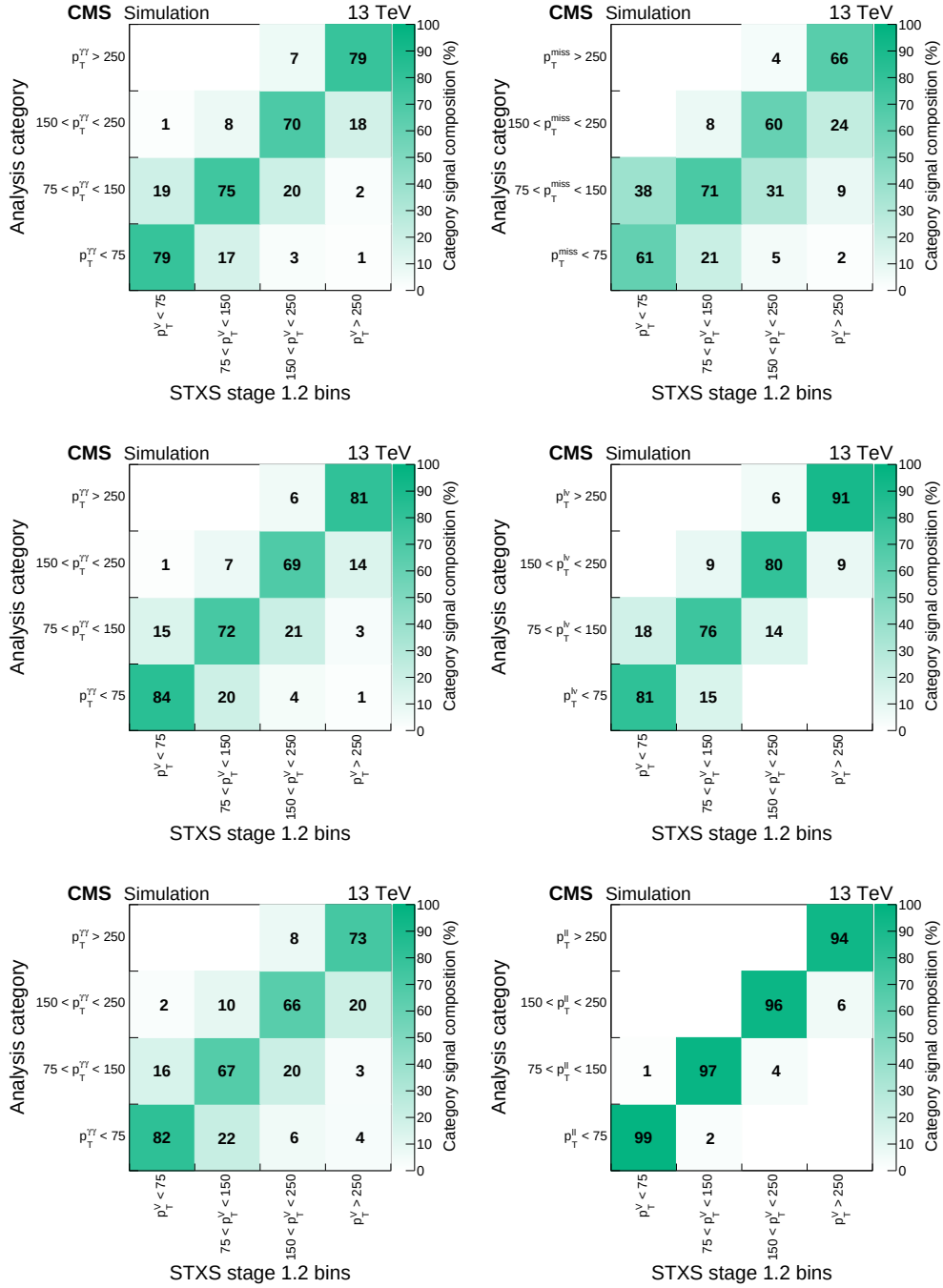


Figure 4.36: Competition of the STXS stage 1.2 bins with $m_H = 125$ GeV with the p_T^{miss} , $p_T^{\ell\nu}$, $p_T^{\ell\ell}$ versus the particle-level p_T^V in the VH MET, WH Leptonic and ZH Leptonic categories (top, middle and bottom plots in the right column), respectively, and with the $p_T^{\gamma\gamma}$ versus the particle-level p_T^V in these three analysis categories (top, middle and bottom plots in the left column).

4.3.2 Analysis categories at the stage 1.2

Although the fine granularity at the STXS stage 1.2 can reduce the theoretical dependence of the STXS measurements, the splitting at some boundaries leads to some analysis categories with less statistics which could lose sensitivity. It is quite important to decide which analysis categories and corresponding STXS bins at the stage 1.2 should be merged. The analysis strategy for merging is to mutually compare the sensitivity of all possible p_T^V splittings and get the largest one as the best choice. However, the sensitivity of this best choice of the p_T^V splitting must be compatible or larger than one of no splitting case which is the stage 0, otherwise, the analysis category will keep the step 1 categorization.

Among the boundary determination, to gain more sensitivity for each analysis category in the step 2, the analysis includes those prepared BDT classifiers during the sensitivity comparison. Instead of applying a single cut to a given BDT distribution to remove the background-like region in the corresponding category, each BDT distribution is divided into several region corresponding to several analysis sub-categories with different sensitivity by performing a dedicated procedure, where the most background-like region is removed, illustrated by Figure 4.37.

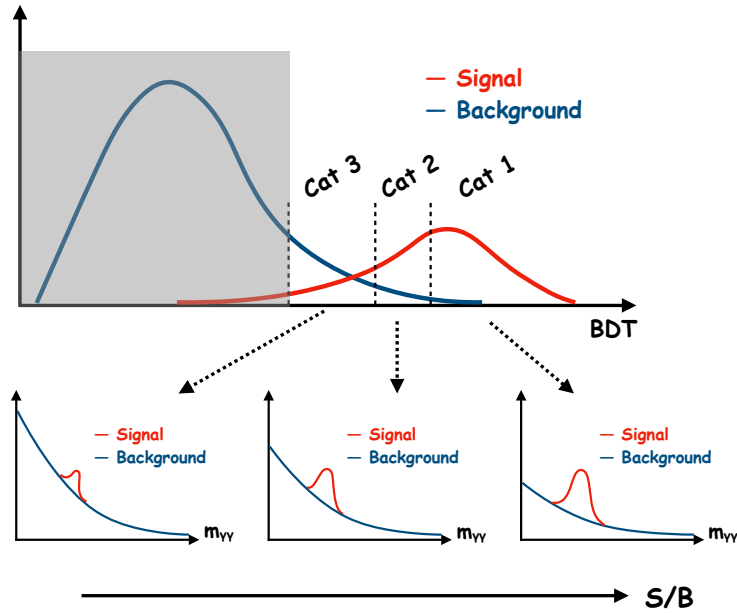


Figure 4.37: Schematic of the analysis sub-categories divided by applying the dedicated boundary optimization procedure to the corresponding BDT distribution. There are four sub-categories with different signal to background ratios shown by the $m_{\gamma\gamma}$ distributions. The fourth category is the most background-like blocked by a gray area and removed from the analysis.

Through this boundary optimization procedure, the combined sensitivity of these analysis sub-categories is expected to become larger than that of a single category. However, this expectation is actually driven by the statistics of given analysis categories as the analysis categories with

lower statistics still use a single cut to the BDT distribution for example. A complete procedure of this boundary optimization is shown in the following:

1. The splitting starts from a given category where is an complete region of the BDT distribution $\in [-1, 1]$
2. Make 99 splitting points within $[-1, 1]$ and find the best splitting point, x , which make two sub-categories, i.e. BDT score $\in [-1, x]$ and $\in [x, 1]$, give the maximal expected significance.
 - In each sub-category, the procedure builds the signal model by a double-Gaussian function to fit to the targeted STXS bins; builds the resonance background model by another double-Gaussian function to fit to the non-target STXS bins; builds the non-resonance background model using the Laurent series function with the 2nd order by fitting to total background samples, where all signal and background simulated samples are normalized to their expected yields.
 - Build a signal-plus-background model and generate a signal-plus-background and a background only Asimov datasets by this model. Then, Compute the expected significance by taking the square root of a likelihood ratio given by

$$q = -2\ln \frac{L(b)}{L(s+b)} \quad (4.22)$$

where s is the signal and b is the resonance and non-resonance backgrounds.

- Calculate the significance with two sub-categories by square sum of significance values in these two sub-categories.
 - Choose the best splitting point which gives the maximal expected significance.
3. The iteration starts finding new splitting points until the improvement of combined significance from a given additional splitting is less than 2%. For each iteration,
 - For each new sub-category from the previous iteration, repeat the second step ¹² to find out a new splitting point candidate.
 - Collect these new splitting point candidates and find out which splitting candidate gives the maximal combined significance as the final splitting point for this iteration.
 - the iteration is terminated if the additional splitting from this iteration has an improvement of less than 2 % with respect to the previous iteration.
 4. The category with the minimal expected significance which is the most background-like region is dropped from this analysis category.

Due to limited statistics, the p_T^V splitting is not applied to the ZH Leptonic category but the boundary optimization procedure still divides this category into two analysis sub-categories as

¹²Note that -1 and 1 now should be replaced by the boundary of this category.

the best choice. Although the VH MET category has sufficient events for the p_T^V splitting, it's checked that there is no significant improvement after splitting. The VH MET category keep one category with three analysis sub-categories given by the boundary optimization procedure. For the WH Leptonic category, there are three possible splitting cases with the corresponding boundaries as $p_T^V = 75$ GeV, $p_T^V = 150$ GeV and $p_T^V = 75, 150$ GeV¹³. Each analysis category in each scenario is also performed to use the boundary optimization procedure to gain optimal sensitivity. Besides, no p_T^V splitting case with the boundary optimization is also treated as a case. Eventually, the combined significance for four cases present at most 5 % difference. Considering the reduction of the theoretical dependence and the understanding of the kinematic structure in the WH production, the boundaries $p_T^V = 75, 150$ GeV with the corresponding sub-categories are used in the WH Leptonic category. On the other hand, since only one STXS bin at stage 1.2 is derived in the $q\bar{q} \rightarrow V(\text{had.})H$ process, the VH Hadronic category directly uses the boundary optimization procedure that gives two analysis sub-categories as the best choice.

In summary, Table 4.8 shows the details of the analysis categories in the step 2 with the corresponding sub-categories and boundaries which can be visualized by Figure 4.38 and 4.39. Total 12 analysis sub-categories are considered as the final analysis categories corresponding to the 5 STXS stage 1.2 bins in the VH analysis. These final analysis categories are expected to give the optimal sensitivity of the VH STXS measurements at the stage 0 and stage 1.2 under an integrated luminosity of 137 fb^{-1} .

Category (step 1)	Category (step 2)	$N_{\text{sub-cat}}$	BDT boundary
ZH Leptonic	-	2	[0.2307, 0.342, 1]
WH Leptonic	$p_T^V < 75 \text{ GeV}$	2	[0.18438, 0.402, 1]
	$75 < p_T^V < 150 \text{ GeV}$	2	[0.27914, 0.482, 1]
	$p_T^V > 150 \text{ GeV}$	1	[0.454, 1]
VH MET	-	3	[0.64612, 0.766, 0.85024, 1]
VH Hadronic	-	2	BDT _{VH Had.} $\in [0.835, 1]$, BDT _{$\gamma\gamma$} $\in [0.81, 1]$
			BDT _{VH Had.} $\in [0.534, 0.835]$, BDT _{$\gamma\gamma$} $\in [0.83, 1]$

Table 4.8: List of the VH analysis categories at the step 1 and 2 with the corresponding number of sub-categories and their boundaries determined by applying a dedicated boundary optimization procedure to the corresponding BDT distributions. The analysis categories without any p_T^V splitting are not shown in the column "Category (step 2)".

¹³Here, each p_T^V , in fact, represents $p_T^{\gamma\gamma}$ in the WH Leptonic category described in Section 4.3.1.

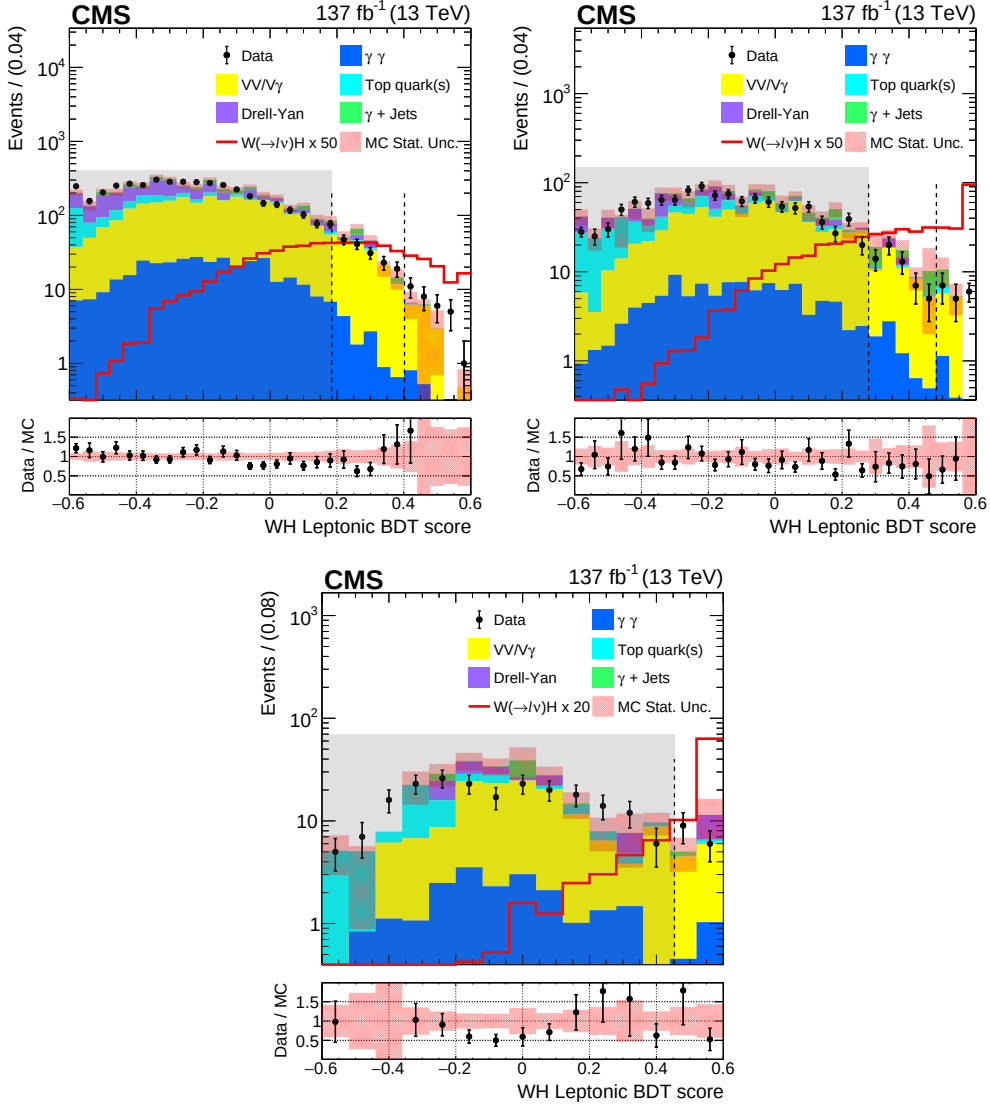


Figure 4.38: Data-and-simulation comparison of the WH Leptonic BDT distributions in three WH Leptonic categories for the stage 1.2, $p_T^{\gamma\gamma} < 75$ GeV (left), $75 < p_T^{\gamma\gamma} < 150$ GeV (middle) and $p_T^{\gamma\gamma} > 150$ GeV (right) with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV. The different background components are stacked with different color and solid histograms. The corresponding statistical uncertainty given by background samples is also shown with pink boxes. The histograms with red lines in these three plots represent the samples simulating the $W(\rightarrow \ell\nu)H$ process with $m_H = 125$ GeV, rescaled 20-50 times than the expected yields for visibility. The gray areas in three plots are the most background-like regions dropped from corresponding categories and the dashed lines are the optimized boundaries of the analysis sub-categories. The bottom panel in each figure is the data-to-simulation ratio.

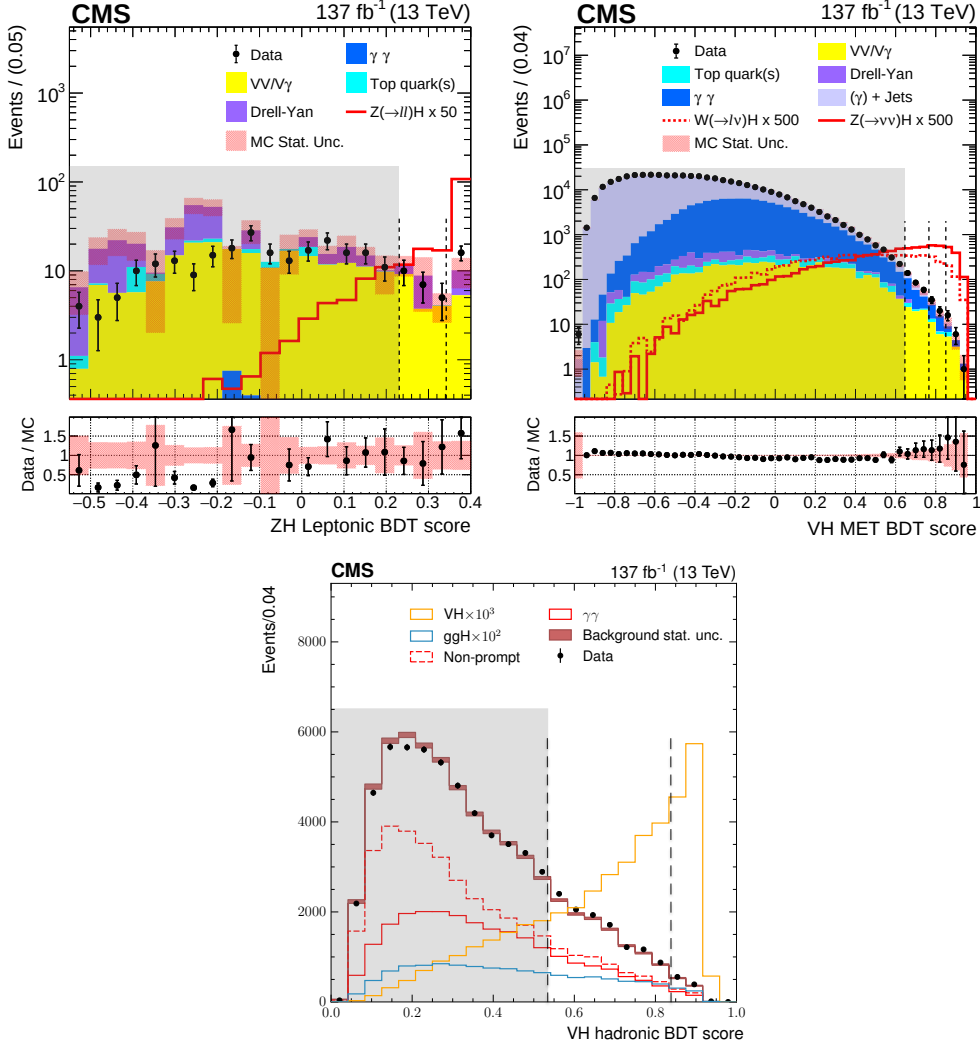


Figure 4.39: Data-and-simulation comparison of the corresponding BDT distributions in three VH categories without p_T^V splitting, the ZH Leptonic category (left), VH MET category (middle) and VH Hadronic category (right) with 2016-2018 year merged without a region of $m_{\gamma\gamma} \in [120, 130]$ GeV. The different background components are stacked with different color and solid histograms. The corresponding statistical uncertainty given by background samples is also shown with pink boxes. For visibility, the histograms with red solid lines in (left, middle) and dashed line in (middle) are the sample simulating the $q\bar{q}(gg) \rightarrow Z(\rightarrow \ell\ell)H$ and $Z(\rightarrow \nu\nu)H$, and $W(\rightarrow \ell\nu)H$ processes rescaled 50-500 times than the expected yield. The histograms with yellow and blue lines are in (right) are the sample simulating $V(\rightarrow \text{had.})H$ and the ggH processes with $m_H = 125$ GeV, respectively, rescaled 100-1000 times than the expected yield. The gray areas in three plots are the most background-like regions dropped from corresponding categories and the dashed lines are the optimized boundaries of the analysis sub-categories. The bottom panel in each figure is the data-to-simulation ratio.

Analysis categories	Total	Target STXS bin(s)	(qqH)			(VH lep)		Top	S/(S+B)
			VBF	VH had	WH lep	ZH lep	ggZH lep		
ZH lep Tag0	2.4	99.6%	—	—	—	82.0%	17.7%	0.4%	0.57
ZH lep Tag1	0.9	97.5%	0.1%	—	0.2%	80.7%	16.9%	2.2%	0.32
WH lep $p_T^V < 75$ Tag0	2.0	81.1%	—	0.2%	95.0%	3.3%	0.2%	1.3%	0.43
WH lep $p_T^V < 75$ Tag1	4.5	75.7%	2.6%	0.5%	87.2%	7.0%	0.3%	2.4%	0.19
WH lep $75 < p_T^V < 150$ Tag0	3.0	77.7%	0.7%	0.3%	93.2%	3.4%	0.8%	1.6%	0.56
WH lep $75 < p_T^V < 150$ Tag1	3.3	60.8%	1.7%	1.4%	83.1%	7.7%	1.6%	4.4%	0.33
WH lep $p_T^V > 150$ Tag0	3.5	79.9%	0.5%	0.4%	91.5%	3.6%	1.1%	2.8%	0.77
VH MET Tag0	2.2	97.9%	0.4%	0.9%	23.5%	56.9%	17.6%	0.8%	0.48
VH MET Tag1	3.6	90.5%	4.6%	3.1%	28.8%	46.0%	15.7%	1.9%	0.34
VH MET Tag2	6.6	72.2%	15.5%	8.8%	27.7%	33.5%	11.0%	3.5%	0.18
qqH VH-like Tag0	16.3	55.8%	36.5%	2.8%	55.0%	1.4%	—	4.2%	0.25
qqH VH-like Tag1	47.1	26.8%	64.9%	4.7%	26.4%	1.2%	—	2.9%	0.13

Table 4.9: The expected number of signal events for $m_H = 125$ GeV in analysis categories targeting VH production with the W or Z leptonic or hadronic decays, shown for an integrated luminosity of 137 fb^{-1} . The fraction of the total number of events arising from each production mode in each analysis category is provided, as is the fraction of events originating from the targeted STXS bin or bins. Entries with values less than 0.05% are not shown. Here ggH includes contributions from the $gg \rightarrow Z(q\bar{q})H$ with the Z hadronic decay and bbH production modes, qqH incorporates both VBF and VH production with hadronic vector boson decays, and “Top” represents both tH and tH production together. The final column shows the expected ratio of signal to signal-plus-background, $S/(S+B)$, where S and B are the numbers of expected signal and background events in a $\pm 1\sigma_{\text{eff}}$ window centred on m_H .

Chapter 5

Signal and background modelling

The final result extraction in the VH STXS measurements is performed by a maximum likelihood fit of the signal and background models to the $m_{\gamma\gamma}$ distribution in data. The signal models for the STXS stage 1.2 bins and background models in all analysis categories thus have to be constructed first. The strategy of the model construction follows the previous publication of the $H \rightarrow \gamma\gamma$ analysis [5] using parametrization functions to describe signal and background shapes under the $m_{\gamma\gamma}$ distribution. Parametric functions provide more flexible description and smoother model adjustment against a variety of the $m_{\gamma\gamma}$ distribution in data to obtain better fit quality. This is very useful for precision measurements such as the STXS measurements where the slight tuning under the $m_{\gamma\gamma}$ distribution could significantly affect the final results. However, too flexible functions could lead to unstable fit quality to create overfitting issue when containing too many parameters, to balance the flexibility and stability is quite important for estimating correct fit results. Hence, the signal and background modelling including this balance will be discussed in Section 5.1 and 5.2, respectively.

On the other hand, although the rare decay of the $H \rightarrow \gamma\gamma$ causes that statistical uncertainties dominate final results, systematic uncertainties originate from imperfect knowledge of theory, physical objects and detector situation during data taking could also affect final results. These systematic uncertainties give penalties to fitting, change model shapes and quote additional uncertainties for overall normalization to the final results. Therefore, some sources resulting in systematic uncertainties in the measurements of the VH production are also described in Section 5.3.

5.1 Signal modelling

For the $H \rightarrow \gamma\gamma$ analysis at the STXS measurement, the signal shape for the $m_{\gamma\gamma}$ distribution relies on the simulated samples generated by the Madgraph5_aMC@NLO, which gives more precise kinematics under particle-level. These simulated samples need to apply all photon energy correction, resolution smearing and scale factors from the corresponding selections or identifications in order to correct the $m_{\gamma\gamma}$ distribution and expected yields of all production

processes.

The signal shape is independently constructed for each analysis category, each STXS stage 1.2 bin corresponding to each production process, and each year due to different CMS situations. Since the shape of the $m_{\gamma\gamma}$ distribution also depends on the diphoton vertex hypothesis, that is, if the diphoton vertex is correctly identified within 1 cm with respect to the Higgs interaction vertex, the construction is further separated to the correct vertex (within 1 cm) and wrong vertex (beyond 1 cm) scenarios. Therefore, each vertex scenario, a parametric signal shape is constructed by using the sum of at most five Gaussian functions to fit to the $m_{\gamma\gamma}$ distribution from corresponding simulated samples. To avoid using too many parameters to build a function which is unstable and could lead to the overfitting issue, an $\mathcal{F}$ -test [81] is performed determining the optimal number of Gaussian functions for balance between flexibility and stability. This $\mathcal{F}$ -test takes different signal shapes with the sum of different number of Gaussian functions to fit to given simulated events and picks up a number of Gaussian functions with the minimal $\chi^2/\text{d.o.f.}$ as the best choice for the signal modelling.

Moreover, since the mass of the Higgs boson (m_H) is not exactly known, the signal model is further parametrized by m_H to allow m_H floating. This parametrization constructs a functional form where the mean and width of each Gaussian function in the origin signal shape are encoded by the polynomials of m_H , described by Eq. (5.1).

$$f(x) = \sum_{i=1}^{\max} a_i \text{Gaus.}_{(i)}(x, \mu_i(m_H), \sigma_i(m_H)) \quad (5.1)$$

To decide the order of all polynomials and their coefficients in this functional form, the model construction performs a simultaneous signal fitting (SSF) using this functional form to fit to three kinds of simulated events with $m_H = 120, 125, 130$ GeV hypotheses¹. It's observed that the order of all polynomials beyond one has no any improvement for fitting quality. Therefore, the polynomials are actually the linear functions of m_H . Besides, systematic uncertainties corresponding to the energy scale and resolution of the individual photon, the material corrections and ECAL crystal light yields are also incorporated into the signal shape as its nuisance parameters. Finally, combining two functional forms corresponding to two vertex scenarios, a functional form of the signal shape generalized to all m_H points is eventually constructed for each STXS stage 1.2 bin in a given analysis category per year. Figure 5.1 and 5.2 show the results with m_H fixed to 125 GeV after signal modelling in the VH analysis categories.

5.2 Background modelling

For an analysis with a small signal falling on a large background, a slight change for the background shape could significantly affect the measured signal yield or peak position. The STXS measurements of the $H \rightarrow \gamma\gamma$ are particularly sensitive to the floating of background shapes. A way to extract background shapes are to rely on samples simulating background events. In the $H \rightarrow \gamma\gamma$ analysis, the main background candidates for most of the analysis categories

¹The optimal number of Gaussian functions for $m_H = 120, 130$ GeV hypotheses are assigned to be the same as one for $m_H = 125$ GeV hypothesis.

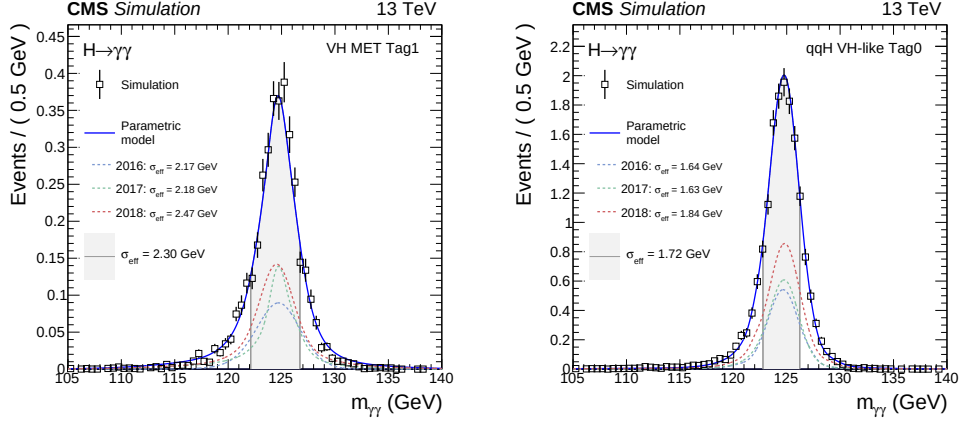


Figure 5.1: Signal models and corresponding simulated events with $m_H = 125$ GeV in the $m_{\gamma\gamma}$ distribution combining 2016-2018 years and all STXS stage 1.2 bins together in one of the VH MET (left) and VH hadronic (right) categories at the stage 1.2. The corresponding effective widths, σ_{eff} , for each year and inclusive years where there are 68% events, are shown.

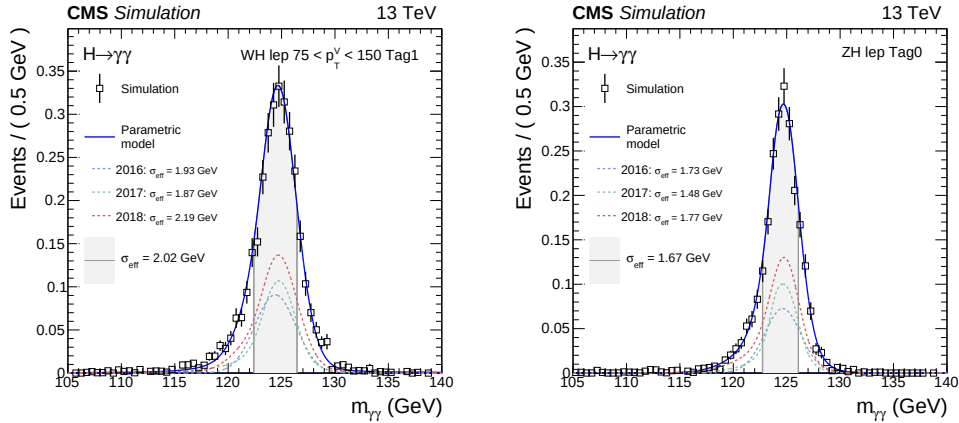


Figure 5.2: Signal models and corresponding simulated events with $m_H = 125$ GeV in the $m_{\gamma\gamma}$ distribution combining 2016-2018 years and all STXS stage 1.2 bins together in one of the WH Leptonic (left) and ZH Leptonic (right) categories at the stage 1.2. The corresponding effective widths, σ_{eff} , for each year and inclusive years where there are 68% events, are shown.

are the QCD multi-jets and $\gamma + \text{jets}$ processes, but the corresponding simulated samples have poor description to data. Although in Section 4.1 we have performed the data-driven methods to replace these samples for the construction of the multivariate classifiers, it's not confirmed that the data-driven samples precisely reflect the background shapes in data. Moreover, the background samples simulating leptonic final states are the limited size with large uncertainty which is also unreliable. Even unknown background candidates can contribute in the analysis. For these reasons, the STXS analysis thus adopts a data-driven fitting method, taking a parametric background shape to directly fit to the $m_{\gamma\gamma}$ distribution in data. The background model is independently built for each analysis category where data is combined in three years to gain

statistics for constraining the background shape. To avoid the background shape significantly biased in the signal region, i.e. $m_{\gamma\gamma} \in (120, 130)$ GeV, the fitting region is extended to a $m_{\gamma\gamma}$ range between 100 and 180 GeV so that a wide range of background events in data can dominantly constrain the background shape. That is why the diphoton selections choose a wide diphoton mass region between 100 to 180 GeV.

Although the data-driven method for the background modelling can directly reflect the $m_{\gamma\gamma}$ distribution in data, choosing different background shapes is sensitive to the signal measurement, demonstrated in Figure 5.3. Therefore, the background shape choice in the $m_{\gamma\gamma}$ distribution should be considered as a kind of systematic uncertainty.

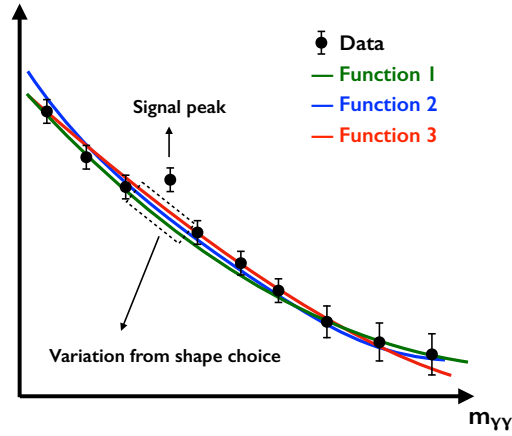


Figure 5.3: Schematic of the diphoton mass distribution to which there are three parametric functions fitted. The normalization variation in the signal region originating from the choice of three background shapes is sensitive to the estimation of the Higgs boson signal yield.

To construct a background model with a systematic uncertainty originating from the background shape choice, a *discrete profiling method* [82] is applied to each analysis category in the STXS measurement. In general, when a shape with several continuous nuisance parameters fits to a given dataset, the best fit value for some parameter of interest x is determined by minimizing the corresponding twice negative log-likelihood function (2NLL) denoted by Λ . Furthermore, by minimizing Λ with respect to the nuisance parameters at each x near the best fit value, the minimal Λ at each x form a profile likelihood curve, where the minimal point gives the best fit value. The uncertainty interval of the best fit value can be defined by the boundaries of x whose Λ on the profile likelihood curve are 1 unit larger than that for the best fit. However, for another point of view, the likelihood curve for the best fit, where all nuisance parameters are fixed to optimized values, is narrower than the profile likelihood curve as those nuisance parameters without floating give x with a smaller error. Likewise, the likelihood curves with all nuisance parameters fixed to arbitrary values are also narrower and totally above the profile likelihood curve. Hence, an overall minimal Λ as a function of x given by the likelihood curves with nuisance parameters fixed to different values thus forms a so-called *envelope*. This envelope curve can actually approximate to the profile likelihood one, described by Figure 5.4,

which implies the uncertainty from the envelope curve is similar to that from the profile curve. Now, if continuous nuisance parameters are now transferred to discrete ones, this concept becomes the discrete profiling method.

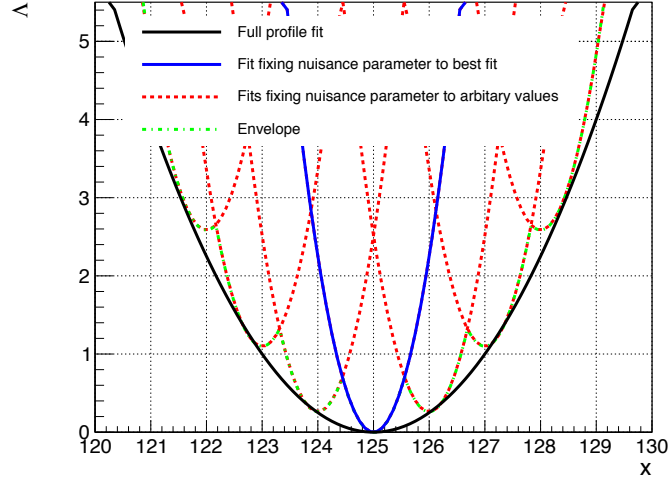


Figure 5.4: Diagram of the profile likelihood curve (black solid curve) and the likelihood curves with nuisance parameters fixed to the best fit (blue solid curve) and arbitrary values (red dashed line). The envelope curve (green curve) is built by tracing the overall minimal Λ from these likelihood curves with the blue solid and red dashed curves [82].

In the background modelling, the discrete profiling method collects several appropriate background functions and labels each background function an integer number. These functions are then encoded to a functional form where the integer is regarded as a discrete nuisance parameter. This discrete nuisance parameter reflects the uncertainty coming from the choice of background shapes, that is, switching the integer will transfer the background shapes from one to another. Moreover, the method adds the complexity of each background function as a penalty to correct Λ by

$$\Lambda_{\text{corr.}} = \Lambda + N \quad (5.2)$$

where N is the number of parameters. Therefore, when profiling this functional form with respect to this integer nuisance parameter, the discrete profiling method scans the x space to obtain the overall minimal $\Lambda_{\text{corr.}}$ from profile likelihood curves, each of which comes from a possible integer corresponding to a background function and has been profiled with respect to its continuous nuisance parameters. The overall minimal $\Lambda_{\text{corr.}}$ forms an envelope curve, described in Figure 5.5. The envelope curve is eventually considered as a discrete profile curve where the uncertainty from the background shape choice is automatically decoded.

In practice, the STXS analysis in the background modelling considers four families of analytic functions :

- Sums of exponentials :

$$f_N(x) = \sum_{i=1}^N p_{2i} e^{p_{2i+1}x},$$

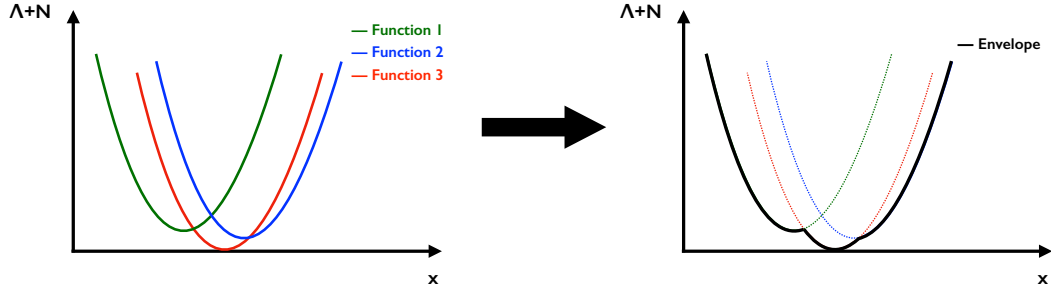


Figure 5.5: Schematic of the discrete profile method for some background functional form containing three functions with corresponding integers (1, 2, 3). The left plot contains three profile likelihood curves with corrected Λ which stand for three different background functions corresponding to three different integers (1, 2, 3). The right is the envelope curve given by picking up the overall minimal Λ_{corr} from three profile curves.

- Sums of polynomials (in the Bernstein basis) :

$$f_N(x) = \sum_{i=0}^N p_i b_{(i,N)}, \text{ where } b_{(i,N)} := \binom{N}{i} x^i (1-x)^{N-i},$$

- Laurent series :

$$f_N(x) = \sum_{i=1}^N p_i x^{-4+\sum_{j=1}^i (-1)^j (j-1)},$$

- Sums of power-law functions :

$$f_N(x) = \sum_{i=1}^N p_{2i} x^{-p_{2i+1}},$$

The function for each order in each family is fitted to the $m_{\gamma\gamma}$ distribution in data to constrain its parameters and then is labeled by an integer number for putting in a background functional form. To avoid overfitting due to too many parameters once each family grows up very high order, especially analysis categories with limited statistics, an $\mathcal{F}$ -test [81] is performed testing the highest order that each family can achieve. This $\mathcal{F}$ -test uses a likelihood ratio test for the lower order and higher order functions with nuisance parameters $\vec{\theta} \in \Theta_0$ and $\vec{\theta}' \in \Theta_1$, respectively. Due to $\Theta_0 \subseteq \Theta_1$ for the same family, according to Wilks' theorem [83], twice the negative log-likelihood ratio function is allowed asymptotically approximating to a χ^2 distribution with M degree of freedom given by the difference between the number of free parameters from the lower to higher order functions :

$$\chi^2(M) \approx -2 \ln \frac{L(x, \vec{\theta})|_{\vec{\theta} \in \Theta_0}}{L(x, \vec{\theta}')|_{\vec{\theta}' \in \Theta_1}} \quad (5.3)$$

The corresponding p -value is calculated by the area under the χ^2 distribution with M d.o.f and above the χ^2 value given by Eq. (5.3). Hence, for each family, the order is increased until the p -value between $N + 1$ and N orders is larger than 0.05 as no significant improvement for the $N + 1$ order with respect of the N order. Finally, for each family, only the functions with the order below $N + 1$ are put in the list of the background functional form. Table 5.1 shows the complete information of the collected functions in all VH analysis categories by applying the $\mathcal{F}$ -test to four families of the analytic functions, some of which are also revealed with Figure 5.6 and 5.7. Once all functions are collected from these four families and the background functional form is made, the discrete profile work will be delivered to the combination fit for the final STXS extrapolation.

Analysis category	$N_{\text{Bernstein}}$	$N_{\text{Exponential}}$	$N_{\text{Power-law}}$	N_{Laurent}
VH MET Tag 0	1 (Best)	1	1	1
VH MET Tag 1	1, 2 (Best)	1	1	1
VH MET Tag 3	1, 2 (Best)	1	1	1
WH Leptonic $0 < p_T^V < 75$ Tag 0	1, 2	1	1 (Best)	1
WH Leptonic $0 < p_T^V < 75$ Tag 1	1, 2 (Best)	1	1	1
WH Leptonic $75 < p_T^V < 150$ Tag 0	1, 2 (Best)	1	1	1
WH Leptonic $75 < p_T^V < 150$ Tag 1	1	1 (Best)	1	1
WH Leptonic $p_T^V > 150$ Tag 0	1	1, 3 (Best)	1	1
ZH Leptonic Tag 0	1 (Best)	1	1	1
ZH Leptonic Tag 1	1, 2	1	1	1 (Best)
VH Hadronic Tag 0	1, 2	1 (Best)	1	1, 2
VH Hadronic Tag 1	2, 3	1 (Best)	1	1, 2

Table 5.1: Collected functions in each VH analysis category for the STXS stage 1.2 by applying the $\mathcal{F}$ -test to four families of the analytic functions such as Bernstein, Exponential, Power-law and Laurent functions. N represents number of parameters for corresponding functions and "Best" means that the functions give a global minimal $\Delta_{\text{corr.}}$, that is, the best fit to data in the given VH analysis categories.

5.3 Systematic uncertainties

This analysis derives many dedicated methods to optimize the performance and sensitivity such as the energy correction and identification in the physical objects. These methods have to rely on Monte Carlo simulations including physical processes and detector configurations. However, in reality, data collected from the CMS detector could be affected by some unexpected detector status or lack of understanding of particles interacting with matters. Besides, those predicted cross sections and kinematics of the physical processes could be also imperfect due to missing higher order terms in perturbative regions or limited knowledge to deal with non-perturbative regions like parton kinematics inside protons, which are only described by a set of parton density functions. These cause that the simulated events can't perfectly describe data.

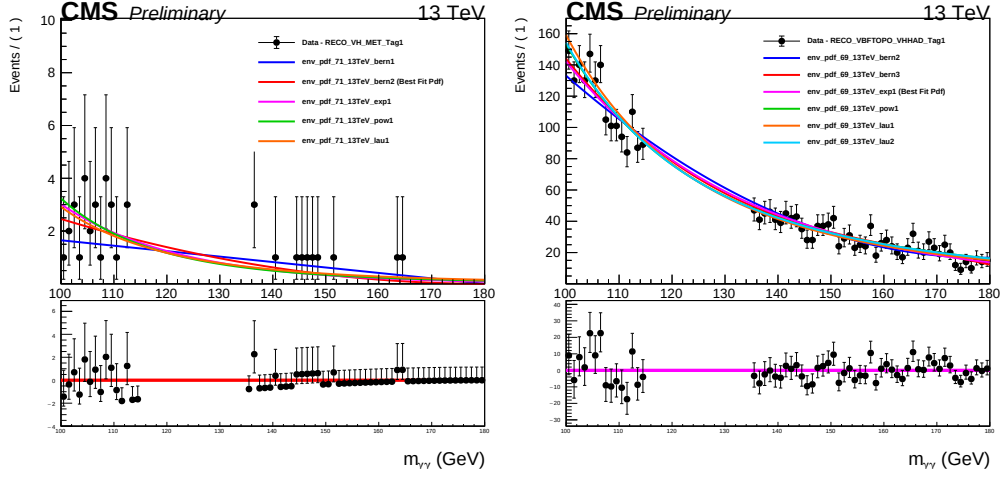


Figure 5.6: Background models from four families after $\mathcal{F}$ -test and corresponding data combining 2016-2018 years together in the VH MET and VH Hadronic categories for the stage 1.2. Data located on a signal region of $115 < m_{\gamma\gamma} < 135$ GeV is blinded.

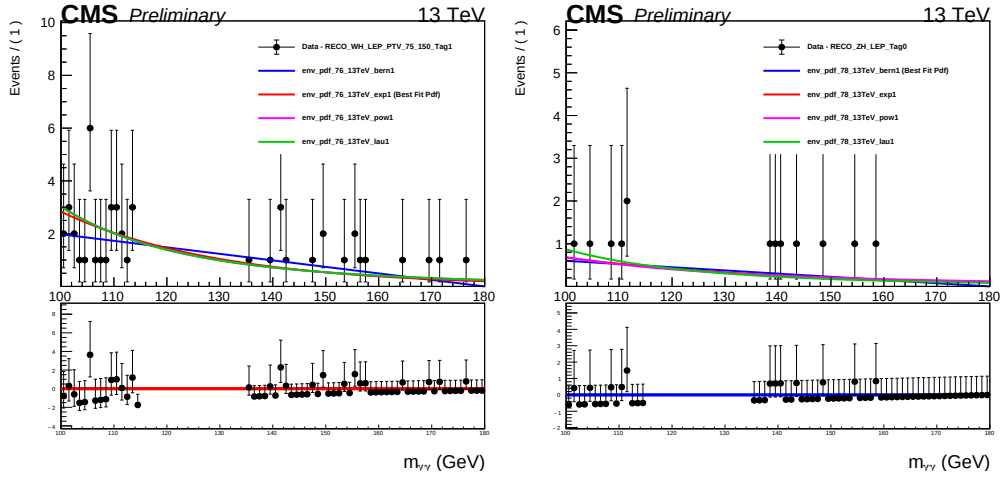


Figure 5.7: Background models from four families after $\mathcal{F}$ -test and corresponding data combining 2016-2018 years together in the WH Leptonic and ZH Leptonic categories for the stage 1.2. Data located on a signal region of $115 < m_{\gamma\gamma} < 135$ GeV is blinded.

Although the analysis also tries to reduce such inconsistency through some dedicated methods, for example, the shower shape correction, described in Section 3.3.4, the residue differences can't be completely eliminated. To account for these discrepancies, the analysis introduces systematic uncertainties to cover them.

In the cross section measurements of the VH production in the $H \rightarrow \gamma\gamma$ channel, systematic uncertainties are given by signal and background modelling. The systematic uncertainty associated with the data-driven background estimation is handled using the discrete profiling method, described in Section 5.2, while there are a lot of systematic uncertainties that affect the signal model. According to the property of the uncertainties, the uncertainties can be handled

through two ways. The uncertainties that change the shape of the $m_{\gamma\gamma}$ distribution are encoded into the signal model as nuisance parameters. These uncertainties are typically the experimental uncertainties related to the energy of two photons. In contrast, the uncertainties which do not affect the shape of the $m_{\gamma\gamma}$ distribution are treated as log-normal variations in the event yield as external constraints. These uncertainties are composed of the theory and experimental parts such as those affecting the VH BDTs used to category events. The magnitude of each uncertainty's impact is determined individually for each STXS bin in each analysis category. The following sections will individually introduce the theoretical and experimental uncertainties which affect the signal modelling.

5.3.1 Theoretical uncertainties

Theoretical uncertainties affect both the overall cross section prediction for a given STXS bin, and the distribution of kinematic variables used in the event selection and categorization. However, the uncertainties affecting on the overall cross section prediction are not included in the STXS measurements since the STXS is directly measured from data. The sources of the theoretical uncertainties considered in this analysis are as follows :

- *Renormalization and factorization scale uncertainties* : The uncertainty arises from variations of the renormalization and factorization scales used when computing the expected SM cross section and event kinematic properties. These account for the missing higher order terms in perturbative calculations. The recommendations provided in Ref [40] are followed. The uncertainty in the overall normalization for the VH production process is estimated using three sources: varying the renormalization scale by a factor of two, varying the factorization scale by a factor of two, and varying both in the same direction simultaneously. Depending on the production process, the size of the uncertainty in the overall normalization varies about 1–2% for the VH production.

To estimate the uncertainty in the event kinematics, the distribution of events falling into each analysis category is recalculated when varying both the renormalization and factorization scales by a factor of two in the same direction simultaneously and the overall cross section for a given STXS bin is kept constant. These uncertainties, representing migrations between analysis categories, are decorrelated for different VH production modes and different regions of the Higgs boson phase space. The migration uncertainties are in general around 1%.

- *Parton density function uncertainties* : These account for the uncertainty due to imperfect knowledge of the composition of the proton, which affects the partons that are most likely to initiate high energy events. The overall normalization uncertainty for the VH production process also include the uncertainty in the value of the strong force coupling constant α_s , and are taken from Ref [40]. Uncertainties in the event kinematic properties are calculated following the PDF4LHC.100 prescription [40, 84, 85] using the MC2hessian procedure [86]. As with the Renormalization and factorization scale uncertainties, the

normalization for a given STXS bin is kept constant when calculating the migrations between analysis categories. The overall normalization uncertainties are around 1.4%, with the migrations significantly smaller, usually less than 1%.

- *Uncertainty in the strong force coupling constant* : The uncertainty in the value of the strong force coupling constant α_s is included in the treatment of the PDF uncertainties, following the PDF4LHC prescription. The impact on the overall normalization for the VH production is 0.8%. An additional source is included to account for changes in the event kinematic properties due to the uncertainty in α_s . This is calculated using a similar procedure to the renormalization and factorization scale migration uncertainties, but instead varying the value of α_s , and corresponds to uncertainties that are in general less than 1%.
- *Uncertainty in the $H \rightarrow \gamma\gamma$ branching fraction* : The probability of the Higgs boson decaying to two photons is required to calculate the SM expected cross section, but this branching fraction is not exactly known due to missing higher order corrections. The uncertainty is currently estimated to be 2% [40].
- *Underlying event and parton shower uncertainties* : These uncertainties are obtained using dedicated simulated samples where the specific tune of the event generator have been modified. They are split into two components for those affecting the overall normalization of the STXS bins and those representing migrations of events between analysis categories. The normalization uncertainties are in general around 1.8% for the VH process.
- *Uncertainties in the VH leptonic STXS fractions* : For VH leptonic production, additional sources are introduced to account for the uncertainty in the modelling of the p_T^V distributions, and the number of jets in the event. Four independent sources are defined to reflect the migrations of events across the p_T^V boundaries at 75, 150, and 250 GeV, in addition to the migration between the zero and greater than one-jet bins for events with p_T^V of 150-250 GeV. These sources are defined separately for the WH leptonic, ZH leptonic, and $ggZH$ leptonic production modes, leading to 12 independent nuisance parameters. In each case, the uncertainty is computed by varying the renormalization and factorization scales and recalculating the fractional breakdown of VH leptonic STXS stage 1.2 cross sections. The total magnitude varying between bins is at most 5% for the dominant WH and ZH leptonic production modes. These are considered as uncertainties in the SM predictions when performing cross section measurements.

As Chapter 4 described, when measuring cross sections, merging certain STXS bins is necessary to avoid sensitivity reduction due to lack statistics in corresponding analysis categories or very high correlations between STXS bins. If two bins are measured individually, the uncertainty related to event migration between two bins is not included since both cross sections are being fitted. However, when performing the bin merging across a boundary, the measurement is sensitive to the relative fraction of the two bins. The uncertainty accounting for migrations across the merged boundaries must be included to the measurements.

5.3.2 Experimental uncertainties

The uncertainties that affect the shape of the signal $m_{\gamma\gamma}$ distribution are listed below.

- *Photon energy scale and resolution* : The uncertainties associated with the residual corrections applied to the photon energy scale in data and the resolution in simulation, described in Section 3.3.2, are evaluated using $Z \rightarrow ee$ events. The estimate is computed by varying the regression training scheme, the distribution of R_9 , and the electron selection criteria. For the majority of photons the resulting uncertainty in the energy scale is between 0.05–0.015%, although for those with very high p_T the effect can be 0.5–3%.
- *Nonlinearity of the photon energy scale* : An additional uncertainty covers possible residual differences in the linearity of the photon energy scale between data and simulation. The uncertainty is estimated using boosted $Z \rightarrow ee$ events. In this analysis, an uncertainty of 0.2% on the photon energy scale is assigned, which accounts for the nonlinearity across the full range of photon p_T values.
- *Shower shape corrections* : An uncertainty accounts for the effect of the shower shape corrections, described in Section 3.3.4, on the photon energy regression, described in Section 3.3.2. The impact is estimated by the difference of the ratio of the energy scale before and after the shower shape corrections in simulated $Z \rightarrow ee$ and $H \rightarrow \gamma\gamma$ events. The magnitude of the uncertainty in the energy scale ranges from 0.01–0.15%, depending on the photon $|\eta|$ and R_9 values.
- *Non-uniformity of the light collection* : The shower depth maximum for photons is deeper than that of electrons by approximately one radiation length, which is 0.89 cm for the ECAL crystal. Hence the differences in the light collection efficiency along the length of the ECAL crystals will lead to a difference in the ECAL response to electrons and photons [87]. Since the photon energy scale is derived from $Z \rightarrow ee$ events, an additional systematic uncertainty is assigned to the photon energy scale to account for this difference. The uncertainty is between 0.16–0.25% for photons with $R_9 > 0.96$, whilst the magnitude for low R_9 photons is below 0.07%.
- *Modelling of the material budget in front of the ECAL* : The material that photons pass through before the ECAL could affect the behavior of the electromagnetic showers, and this effect may not be perfectly modeled in simulation. Dedicated samples where the material budget is uniformly varied by $\pm 5\%$ are used to estimate the impact on the photon energy scale by the difference of the energy scale in simulated $Z \rightarrow ee$ and $H \rightarrow \gamma\gamma$ events, with $|\eta|$ and R_9 bins. The magnitude of the resulting uncertainty ranges from 0.02–0.05% for the most central photons, increasing to as much as 0.24% for those in the endcap.

The set of the uncertainty sources that only modifies the event yield is described below.

- *Integrated luminosity* : The uncertainties of 2.5%, 2.3% and 2.5% are determined by the CMS luminosity monitoring for the 2016, 2017, and 2018 data sets [88–90], respectively, in order to account for common sources in the luminosity measurement schemes.

- *Photon identification BDT score* : The uncertainty arises from the residual discrepancies between data and simulation in the photon identification BDT score after shower shape corrections, described in Section 3.3.4. It is estimated by varying the set of events used to train the quantile regression corrections. The uncertainty in the signal yields is estimated by propagating this uncertainty through the full category selection procedure. The impact in the most sensitive analysis categories is around 3%.
- *Jet energy scale and smearing corrections* : The uncertainty in the jet energy scale is a few percent and depends on p_T and η . The impact of jet energy scale uncertainties in event yields is evaluated by varying the jet energy corrections within their uncertainties and propagating the effect to the final result. Correlations between years are introduced for the different jet energy scale uncertainty sources, ranging between 0 and 100%. The impact on the VH Hadronic category yields can be as high as 22% for the scale uncertainties, but is less than around 8% for the resolution.
- *Per-photon energy resolution estimate* : The uncertainty in the per-photon energy resolution is parametrized as a rescaling of the resolution by $\pm 5\%$ about its nominal value. This is designed to cover all differences between data and simulation in the distribution, which is an output of the energy regression. For most analysis categories, the impact is below the percent level.
- *Trigger efficiency* : The efficiency of the trigger selection is measured with $Z \rightarrow ee$ events using the tag-and-probe technique. The size of its uncertainty is less than 1%. An additional uncertainty is also introduced to account for a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region at $|\eta| > 2.0$, which caused a specific trigger inefficiency during 2016 and 2017 data taking [62].
- *Photon preselection* : The uncertainty in the photon preselection efficiency is computed by the tag-and-probe technique in $Z \rightarrow ee$ events and varying the tag-and-probe conditions to fluctuate the ratio of the efficiency measured in data and simulation. Additionally, the R_9 and $|\eta|$ distributions for the probe electron are weighted to those for photons in $H \rightarrow \gamma\gamma$ events. Its magnitude is less than 1%.
- *Missing transverse momentum* : This uncertainty is computed by shifting the reconstructed p_T of the particle candidates entering the missing transverse momentum computation, within the momentum scale and resolution uncertainties appropriate to each type of reconstructed object, as described in Section 3.5. In this analysis, the impact on the category yields is never larger than 5%, even for analysis categories that explicitly use the missing transverse momentum in their definition.
- *Pileup jet identification* : The uncertainty in the pileup jet classification output score accounts for the inconsistency in data and simulation under $Z + \text{jets}$ events. It is estimated by the score differences of the simulated $Z + \text{jets}$ events with parton shower modeled by PYTHIA and HERWIG++. The difference is of the order 1%.

- *Lepton isolation and identification* : This uncertainty affecting electrons and muons is computed by using the tag-and-probe technique in $Z \rightarrow ee, \mu\mu$ events and varying the tag-and-probe conditions to fluctuate the ratio of the efficiency measured in data and simulation. The resulting impact on the analysis categories selecting leptons is up to around 1%.

Most of the experimental uncertainties are left uncorrelated among the different years. The exceptions are the partial correlations introduced for the luminosity and jet energy correction uncertainties.

Chapter 6

Results

In Chapter 4, all analysis categories of the VH productions with optimal sensitivity for cross section measurements have been defined according to the STXS bins at the stage 1.2 and the boundaries of the BDT classifiers. The signal model for each STXS bin in each analysis category is also constructed and so is the background model in each analysis category through the Chapter 5. In this chapter, the cross section of the STXS bins for the VH production and the other Higgs boson productions is estimated by performing a combination fit. The combination fit uses the state-of-the-art simultaneous fit to all analysis categories in the $H \rightarrow \gamma\gamma$ channel with cross sections like parameters of interest, including all systematic uncertainties as constraints of these cross sections. This enables the same STXS bin in different categories to correct and share the same cross section during fitting.

For the VH analysis, although the BDT classifiers reject most of the contributions from non-target STXS bins in all VH analysis categories, unstable normalization between targeted and non-target bins lead to huge correlation arising from the indistinguishable diphoton mass shape in all STXS bins during fitting. Table 4.9 shows very pure WH and ZH Leptonic categories but the “ qqH VH -like” analysis category has compatible contamination from the STXS bins of the ggH production. Nevertheless, through this simultaneous fit to overall STXS analysis, these compatible contamination from the ggH STXS bins in the “ qqH VH -like” analysis category can be constrained by those ggH analysis categories, described in Figure 6.1.

The same concept also occur in the other categories. Therefore, through the combination fit, the cross section of all STXS bins can be well-measured. That is the reason why the STXS framework is, in particular, appropriate to the combination fit and vice versa. This state-of-the-art simultaneous fit is introduced in Section 6.1.

Besides revealing the observed cross section for the VH production under the STXS framework, this chapter also provides the measurements of the signal strength modifier defined by the ratio of the observed cross section and that from the Standard Model prediction as

$$\mu = \frac{\sigma_{\text{obs}}}{\sigma_{\text{SM}}}$$

where include the theoretical uncertainties in terms of overall normalization. Therefore, two

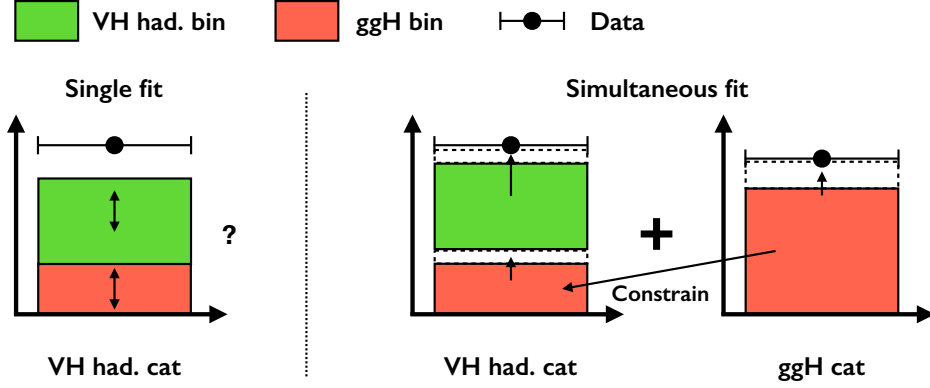


Figure 6.1: Schematic of the composition between a single fit (left) in the VH Hadronic category and a simultaneous fit combining the VH Hadronic and ggH categories (right). A single fit makes the normalization of the VH had. bin (green) and ggH bin (orange) become ambiguous and create large correlation between both bins. However, by adding an additional ggH analysis category to do a simultaneous fit, the ggH bin in the VH Hadronic category is constrained by the ggH category. The VH had. bin can be correctly measured.

kinds of final results by combining 2016-2018 year scenarios with a luminosity of 137 fb^{-1} are showed in Section 6.2 for the signal strength modifier and Section 6.3 for the cross sections of the VH STXS bins including other Higgs boson productions. These results are totally the same as the latest $H \rightarrow \gamma\gamma$ STXS paper of Ref [72] where I'm one of the main contributions.

6.1 Statistical methodology

In general, an unbinned maximum-likelihood fit provides a complete information from the $m_{\gamma\gamma}$ distribution in data to extract the most precise results. Since the overall STXS measurements use a lot of data, model parameters and constraints from systematic uncertainties, using an unbinned fit can consume much time and CPU usage for calculation. Instead, it's observed that to perform a binned fit to the $m_{\gamma\gamma}$ distribution with a bin width of 250 MeV, equivalent to 320 bins in the range of $100 < m_{\gamma\gamma} < 180 \text{ GeV}$, not only speeds up the calculation but also keep compatible precision. Hence, the STXS extraction performs a simultaneous binned maximum-likelihood fit using all signal-plus-background models to fit to the $m_{\gamma\gamma}$ distribution over the range $100 < m_{\gamma\gamma} < 180 \text{ GeV}$ for data in all analysis categories.

In the STXS measurements, the simultaneous fit originally establishes a likelihood function with signal, background models and data in each analysis category, defined by

$$\mathcal{L}_c(\text{data} | \vec{\sigma}, m_H, \vec{\theta}_s, \vec{\theta}_b, \vec{\theta}) = \prod_{i=1}^{N_{\text{bin}}} \text{Poisson}\left(d_i | s_i(\vec{\sigma}, m_H, \vec{\theta}_s) + b_i(\vec{\theta}_b)\right) \cdot p(\vec{\sigma} | \vec{\theta}) \quad (6.1)$$

where $\vec{\sigma}$ is a set of parameters of interest (POIs), which is one or more cross section parameters in this analysis; $\vec{\theta}_s$, $\vec{\theta}_b$ and $\vec{\theta}$ are the nuisance parameters from the signal shapes, background

shapes including shape indices and those which affect the measurements but are not of interest, respectively; N_{bin} is the number of bins for the $m_{\gamma\gamma}$ distribution which is 320; "Poisson" represents a Poisson distribution calculated with

$$\text{Poisson}(d_i | s_i + b_i) = \frac{(s_i + b_i)^{d_i}}{d_i!} e^{-(s_i + b_i)} \quad (6.2)$$

where d_i , s_i and b_i are the observed number of events, signal and background expectation at the i -th bin, respectively; $p(\vec{\sigma}|\vec{\theta})$ is the probability density functions (usually log-normal distributions) of all nuisance parameters to give penalties as external constraints to the POIs. Finally, a complete likelihood function is a product over likelihood functions for all analysis categories given by

$$\mathcal{L}(\text{data} | \vec{\sigma}, m_H, \vec{\theta}_s, \vec{\theta}_b, \vec{\theta}) = \prod_{c=1}^{N_c} \mathcal{L}_c(\text{data} | \vec{\sigma}, m_H, \vec{\theta}_s, \vec{\theta}_b, \vec{\theta}) \quad (6.3)$$

where N_c is the number of analysis categories.

For the estimation of the POIs, that is, the cross sections, the simultaneous fit then performs a maximum likelihood estimation to minimize twice the negative log-likelihood (2NLL) from Eq. (6.3), including the discrete profile method, described in Section 5.2 :

$$-2 \ln \tilde{\mathcal{L}}(\text{data} | \vec{\sigma}, m_H, \vec{\Theta}) = -2 \ln \mathcal{L}(\text{data} | \vec{\sigma}, m_H, \vec{\theta}_s, \vec{\theta}_b, \vec{\theta}) + \sum_j^{N_c} N_{b(j)} \quad (6.4)$$

where $\vec{\Theta} = \vec{\theta}_s + \vec{\theta}_b + \vec{\theta}$ and $N_{b(j)}$ is the number of parameters for the background shape in the j -th category. The free parameters in the fit are the $\vec{\sigma}$, m_H and nuisance parameters of signal models and background models including their indices which freely switch one to another corresponding to changing background shapes for minimizing the 2NLL. The values of all parameters after the 2NLL minimization are treated as the best fit values. On the other hand, the confidence intervals for the POIs, $\vec{\sigma}$, following the frequentist method are estimated by performing a $\vec{\sigma}$ scan near the best fit values to a profiled likelihood-ratio test statistic :

$$q(\vec{\sigma}) = -2 \ln \frac{\tilde{\mathcal{L}}(\text{data} | \vec{\sigma}, \hat{m}_H, \hat{\vec{\Theta}})}{\tilde{\mathcal{L}}(\text{data} | \hat{\vec{\sigma}}, \hat{m}_H, \hat{\vec{\Theta}})}, \quad \text{with a constraint : } \vec{\sigma} > \vec{0} \quad (6.5)$$

where $\hat{\vec{\sigma}}, \hat{m}_H, \hat{\vec{\Theta}}$ are the best fit values of the POIs and nuisance parameters; $\hat{m}_H$ and $\hat{\vec{\Theta}}$ are evaluated by performing maximum likelihood estimations to Eq. (6.4) for $\vec{\sigma}$ fixed to some value. According to Wilks' theorem, the test statistic Eq. (6.5) can asymptotically approximate to a χ^2 distribution with its degree of freedom given by the number of measured parameters of interest. For an one-dimensional scan, such as each STXS measurement, where other STXSs are considered as nuisance parameters, the scan result, $q(\sigma)$, can be illustrated as the blue curve in Figure 6.2 with the lowest point $q(\hat{\sigma}) = 0$ as the best fit, and the 68 % confidence interval can be defined by the interval of $q(\sigma) < 1.0$. This value originates from the fact that the area under the χ^2 distribution with one degree of freedom below $q(\sigma) = 1.0$ is given by an occupancy rate of 68 % with respect to the whole area under the χ^2 distribution. In other words, suppose the

parameter of interest, σ , is measured from an amount of pseudo-experiments, and there are 68 % results that fall in this region.

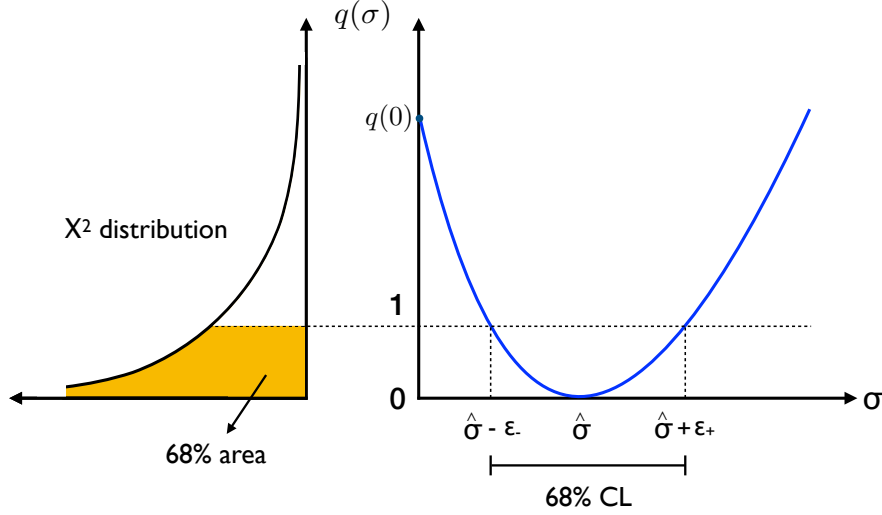


Figure 6.2: Profile likelihood-ratio curve (blue curve) for the σ scan with the lowest point as the best fit corresponding to a best fit value of the POI $\hat{\sigma}$. The 68 % confidence interval of the POI is defined by two boundaries from two cross points of the profile curve and $q(\sigma) = 1.0$ line projected to the σ axis, where ε_- and ε_+ are regarded as the uncertainties of the POI. This definition originates from that the area (yellow area) under the χ^2 distribution with one degree of freedom below $q(\sigma) = 1.0$ occupies 68 % with respect to a complete area under this χ^2 distribution. $q(0)$ is the test statistic under the background-only hypothesis.

Therefore, the lower and upper boundaries for the 68 % confidence interval can be determined by the cross points between the $q(\sigma)$ curve and $q(\sigma) = 1$ straight line projecting to the σ axis, described in Figure 6.2. The ε_- and ε_+ are regarded as the uncertainties of the best value $\hat{\sigma}$.

Besides, a signal excess level under the background-only hypothesis can be quantified by a p -value defined by

$$p\text{-value} = 1 - P(q(\sigma) < q(0)), \quad q(0) = -2 \ln \frac{\tilde{\mathcal{L}}(\text{data} | \sigma = 0, \hat{m}_H, \hat{\hat{\theta}})}{\tilde{\mathcal{L}}(\text{data} | \hat{\sigma}, \hat{m}_H, \hat{\hat{\theta}})}, \quad (6.6)$$

that is, area under the χ^2 distribution above $q(0)$, where $q(0)$ is the profiled likelihood-ratio test statistic under the background-only hypothesis. This statistical quantity can be interpreted as a probability of background producing a fluctuation as large or larger than the observed one assuming the presence of only background. Since the square of one random number from the standard normal Gaussian distribution (mean = 0, sigma = 1) satisfies the χ^2 distribution with one degree of freedom, the p -value in this χ^2 distribution is equal to twice the p -value in the plus side of this Gaussian distribution above Z_0 , described in Figure 6.3. Hence, Z_0 can express

the significance of an signal excess, estimated by

$$Z_0 = \sqrt{q(0)}. \quad (6.7)$$

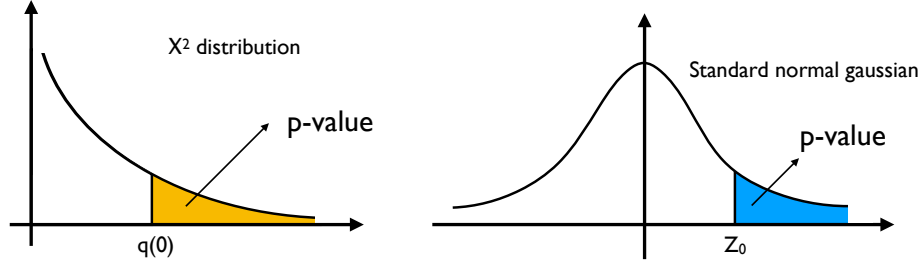


Figure 6.3: Square of the one random number from the standard normal Gaussian distribution is equivalent to the χ^2 distribution with one degree of freedom. Therefore, the area under the χ^2 curve above $q(0)$ (yellow) as a p-value is equal to twice that under standard normal Gaussian curve at the plus side above Z_0 (blue), where "twice" comes from the minus side which is a unphysical region.

6.2 Signal strength modifier

In this section, the observed signal strength modifier for the inclusive $H \rightarrow \gamma\gamma$ production and four exclusive productions such as the ggH , VBF , VH and top-Higgs productions are shown, including a 68 % confidence interval as their uncertainties and combining 2016-2018 year scenarios with a luminosity of 137 fb^{-1} . In the inclusive measurement, all STXS bins are grouped together and scaled by an inclusive signal strength modifier as a parameter of interest to fit to all analysis categories. Since the denominator of the signal strength modifier is donated by the cross section given by the SM prediction, not only the systematic uncertainties related to the event kinematics but those involving in the overall cross section normalization are included during the fitting. The observed result for the inclusive signal strength modifier is shown as

$$\mu = 1.12^{+0.09}_{-0.09} = 1.12^{+0.06}_{-0.06}(\text{theo})^{+0.03}_{-0.03}(\text{syst})^{+0.07}_{-0.06}(\text{stat})$$

where the uncertainty is decoupled into the theoretical, experimental and statistical components. The p -value in this fit with respect to the SM prediction is about 17 % which is compatible to the SM prediction. Figure 6.4 shows the best fit of the signal-plus-background model with data combining all analysis categories and years. For visibility of the signal with respect to the background expectation, the events and signal-plus-background model in each analysis category are weighted by $S/(S+B)$, where S and B are the expected signal and background events within a $\pm\sigma_{\text{eff}}$ window centered on $m_H = 125.38 \text{ GeV}$, where the background events are given by the signal-plus-background fit. These weighted signal, weighted background models

and weighted data are normalized such that the integral of the weighted signal model matches the number of signal events obtained from the best fit.

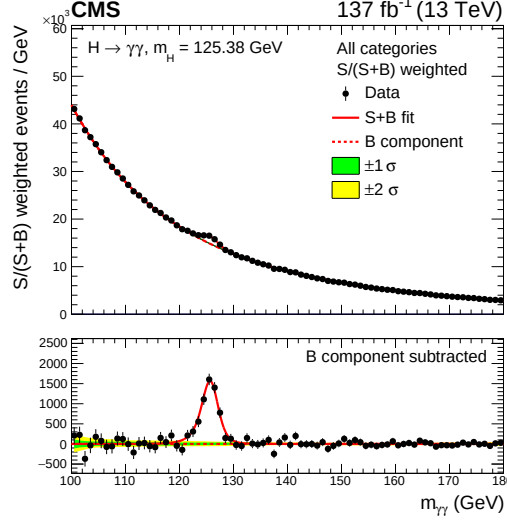


Figure 6.4: The inclusive best fit of the signal-plus-background model (red curve) to the diphoton invariant mass distribution in data (black dot) by combining all STXS bins, all analysis categories and three years. For visibility of signal, the events and signal-plus-background model in each analysis category are weighted by $S/(S+B)$, where S and B are the expected signal and background events within a $\pm\sigma_{\text{eff}}$ window centered on $m_H = 125.38$ GeV, where background events are given by the signal-plus-background fit. Finally, the plot is normalized such that the integral of the weighted signal model matches the number of signal events obtained from the best fit. The green and yellow bands are the 1σ and 2σ variations coming from the background choice. The lower panel shows the residuals after background subtraction.

For the exclusive measurements, the STXS bins of the ggH and $b\bar{b}H$ processes are considered as the same process scaled with the same signal strength modifier, μ_{ggH} ; all VBF STXS bins share the same modifier, μ_{VBF} , for scaling in the fit; the VH hadronic and leptonic bins including all $ggZH$ processes are bundled together and scaled by μ_{VH} ; μ_{Top} equally controls to $t\bar{t}H$, tHq and tHW production modes at the same time. Total four free parameters are estimated by the simultaneous fit to all analysis categories. The observed values and corresponding uncertainties for four signal strength modifiers are showed in Figure 6.5.

The compatibility of this fit with respect to the SM prediction expressed by p -value are approximately 50 %. It is worth noting that the μ_{ggH} starts to be limited by its systematic uncertainties under the current luminosity. Focusing on the result for the VH production, the observed value and corresponding uncertainties for the VH signal strength modifier is given by

$$\mu_{VH} = 1.34^{+0.36}_{-0.33} = 1.34^{+0.07}_{-0.06}(\text{theo})^{+0.05}_{-0.04}(\text{syst})^{+0.35}_{-0.32}(\text{stat})$$

where the uncertainty is decoupled into the theoretical, experimental and statistical components. The corresponding p -values with respect to the background-only hypothesis and SM prediction are about $2.6 \times 10^{-6} \%$ and 32 %, corresponding to the significance of 4.7σ and 1σ ,

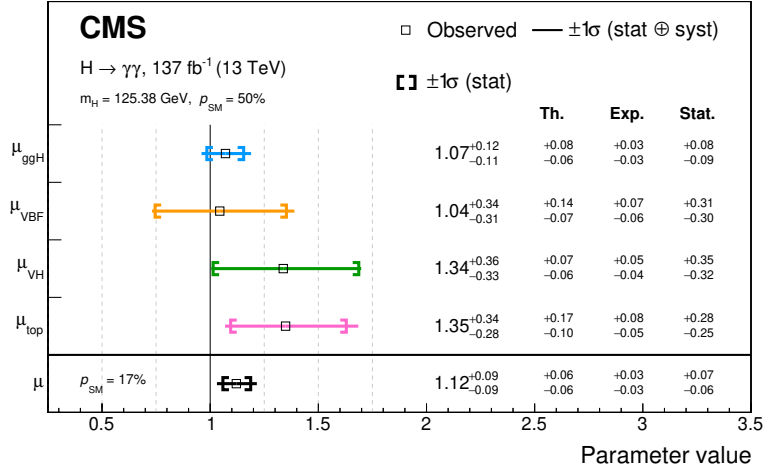


Figure 6.5: Observed results of the signal strength modifiers of four principal Higgs boson production modes. Each result contains the best fit value and its uncertainties with the total version and decoupled versions including the theoretical, experimental and statistical uncertainties. The color scheme follows Figure 4.1, where the tH production is merged into the ttH production. The signal compatibility for these four observed signal strength modifiers with respect to the Standard Model prediction is expressed by a p -value of 50 %. The observed inclusive signal strength modifier is also shown by the black color scheme, with the p -value w.r.t. the SM prediction of 17 %.

respectively, which is compatible to the SM prediction. As expected, the result shows that the statistical uncertainty is dominant in the total uncertainty. On the other hand, in the VH analysis, the systematic uncertainties for the experimental sources mainly come from integrated luminosity, photon energy corrections (scale and smearing) and photon identification, each of which gives a value of approximately 2 % deviation with respect to the observed μ_{VH} , while each systematic uncertainty from the theoretical sources are approximately 2 % after fitting, shown in Figure 6.6. More importantly, the measurement of the μ_{VH} signal strength modifier gives essential improvement with respect to Ref [5] if the improvement from luminosity raising is not considered. This originates from the application of multivariate techniques and sub-categories in this analysis, described in Chapter 4. Figure 6.7 shows the best fit of signal-plus-background model and data in this fit by including all VH analysis categories weighted by $S/(S+B)$ as Figure 6.7.

6.3 Simplified template cross sections

Under the STXS framework, the finer STXS bins are able to not only reduce the theoretical dependence for the STXS results but also decode more information such as kinematic structures of the Higgs boson productions and even explore the potential of physics beyond the Standard Model from the STXS bins with high energy scale. However, such fine granularity could lose sensitivity due to limited size of data. As Chapter 4 have shown, the limitation of the STXS

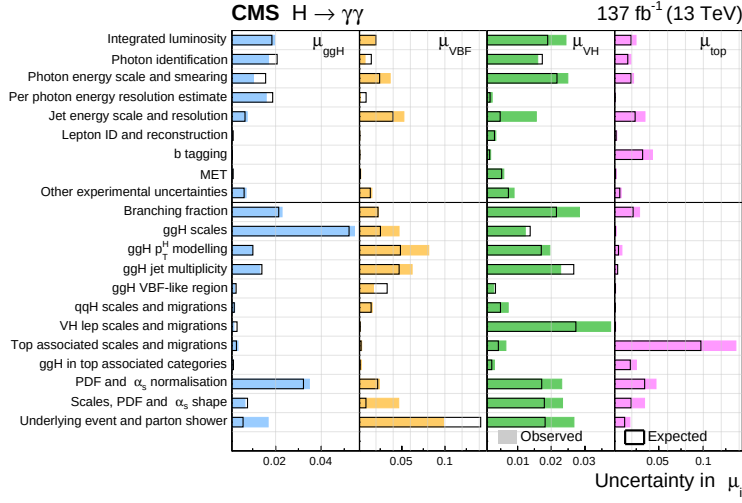


Figure 6.6: Main systematic uncertainties of the signal strength modifiers for four principal Higgs boson production modes. The observed (expected) uncertainties are the best fit values by performing a combination fit to data (the SM prediction), shown by histograms with color filled (black edges). The color scheme follows Figure 4.1, where the tH production is merged into the ttH production.

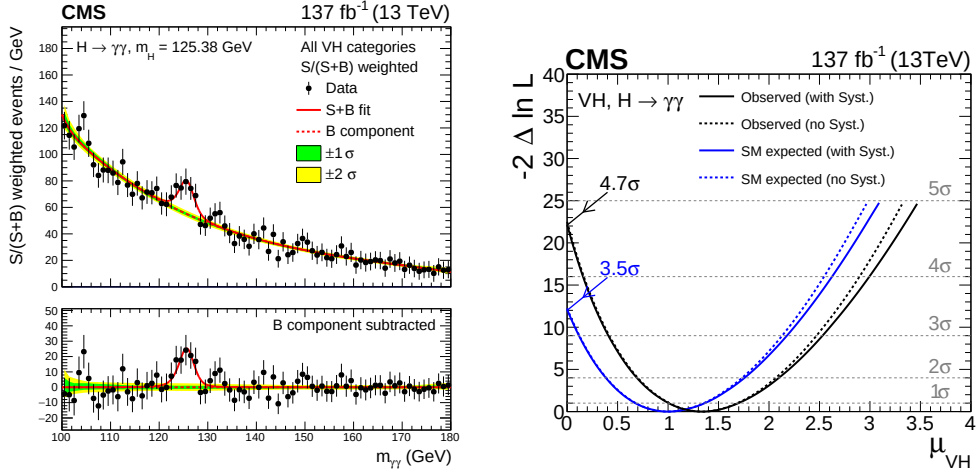


Figure 6.7: VH best fit of the signal-plus-background model (red curve) to the diphoton invariant mass distribution in data (black dot) by combining all STXS bins in all VH analysis categories. For visibility of signal, the events and signal-plus-background model in each analysis category are weighted by $S/(S+B)$, where S and B are the expected signal and background events within a $\pm\sigma_{\text{eff}}$ window centered on $m_H = 125.38$ GeV, where background events are given by the signal-plus-background fit. Finally, the plot is normalized such that the integral of the weighted signal model matches the number of signal events obtained from the best fit. The green and yellow band are the 1σ and 2σ variations coming from the background choice. The lower panel shows the residuals after background subtraction.

analysis for the $H \rightarrow \gamma\gamma$ channel under the current luminosity of 137 fb^{-1} is the stage 1.2. The corresponding analysis categories have been optimized and used to measure the signal strength modifiers in Section 6.2. Therefore, in this section, the observed cross sections under the STXS stage 1.2 bins are revealed by combining 2016-2018 year scenarios with a luminosity of 137 fb^{-1} . Since the STXS measurements no longer rely on the cross sections given by the Standard Model prediction as the definition of the signal strength modifier, the theoretical uncertainties related to overall normalization are not included in the fit and the results thus have less theoretical dependence. Besides the minimal merging scheme with the maximal granularity, i.e. the STXS stage 1.2 bins, in Section 6.3.2, this section also provides the maximal merging scheme which is the STXS stage 0 bins to compare with the previous $H \rightarrow \gamma\gamma$ analysis in 2016, described in Section 6.3.1.

6.3.1 STXS at Stage 0

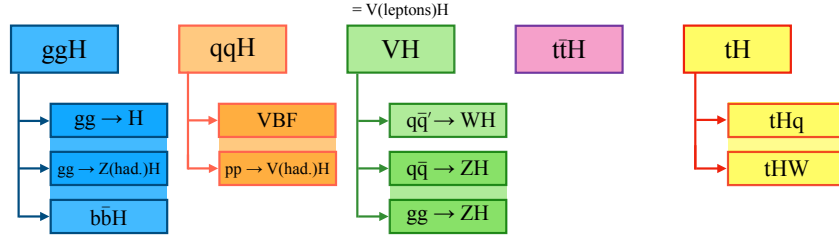


Figure 6.8: Schematic diagram for the STXS stage 0 bins. Some bins are merged by placing similar but light color areas behind these STXS bins.

At this stage, the results with the least granularity in the cross section measurements under the STXS framework are shown. In this fit, all ggH , $b\bar{b}H$ and $ggZH$ with the Z boson decaying hadronically are bundled and scaled together with a single ggH bin. The hadronic WH and ZH processes are grouped with VBF production and the leptonic WH process and leptonic ZH process including $ggZH$ are grouped individually. The tHq and tHW are merged together to form a tH bin which is split from the $t\bar{t}H$ bin. Total 6 parameters of interest are fitted to data in 80 analysis categories. The Figure 6.9 shows the observed results of the cross section measurement under the STXS stage 0 with the complete results put in Appendix A. The corresponding p -value with respect to the SM gives 35 % which is in agreement with the SM prediction. In particular, the details of the observed VH leptonic STXS results, also shown with Table 6.1, gain significant improvement, not considering luminosity raising, with respect to the old one in 2016 [5], shown in Figure 6.10, where the VH analysis used the cut-based strategy. These new results are more close to the SM prediction with less uncertainties.

6.3.2 SXTS at Stage 1.2

In this section, the cross section measurements with the extraordinary granularity for a single Higgs production is shown with the diphoton decay channel with respect to other Higgs decay

Parameters	$\sigma\mathcal{B}$ (fb)				$\sigma\mathcal{B}/(\sigma\mathcal{B})_{\text{SM}}$
	SM prediction	Observed (Expected)			Observed (Expected)
	($m_{\text{H}} = 125.38$ GeV)	Best fit	Stat. unc.	Syst. unc.	Best fit
WH lep	$0.88^{+0.03}_{-0.03}$	$1.19^{+0.49}_{-0.43} \left(\begin{smallmatrix} +0.51 \\ -0.42 \end{smallmatrix} \right)$	$+0.48 \left(\begin{smallmatrix} +0.50 \\ -0.43 \end{smallmatrix} \right)$	$+0.07 \left(\begin{smallmatrix} +0.07 \\ -0.04 \end{smallmatrix} \right)$	$1.35^{+0.55}_{-0.49} \left(\begin{smallmatrix} +0.57 \\ -0.47 \end{smallmatrix} \right)$
ZH lep	$0.54^{+0.03}_{-0.02}$	$0.71^{+0.41}_{-0.36} \left(\begin{smallmatrix} +0.42 \\ -0.35 \end{smallmatrix} \right)$	$+0.40 \left(\begin{smallmatrix} +0.41 \\ -0.35 \end{smallmatrix} \right)$	$+0.07 \left(\begin{smallmatrix} +0.06 \\ -0.02 \end{smallmatrix} \right)$	$1.32^{+0.76}_{-0.66} \left(\begin{smallmatrix} +0.77 \\ -0.65 \end{smallmatrix} \right)$

Table 6.1: Results of the STXS stage 0 [72]. The best fit cross sections are shown together with the respective 68% CL intervals. The uncertainty is decomposed into the systematic and statistical components. The expected uncertainties on the fitted parameters are given in brackets. The SM predictions for the cross sections and the theoretical uncertainty in those predictions are listed.

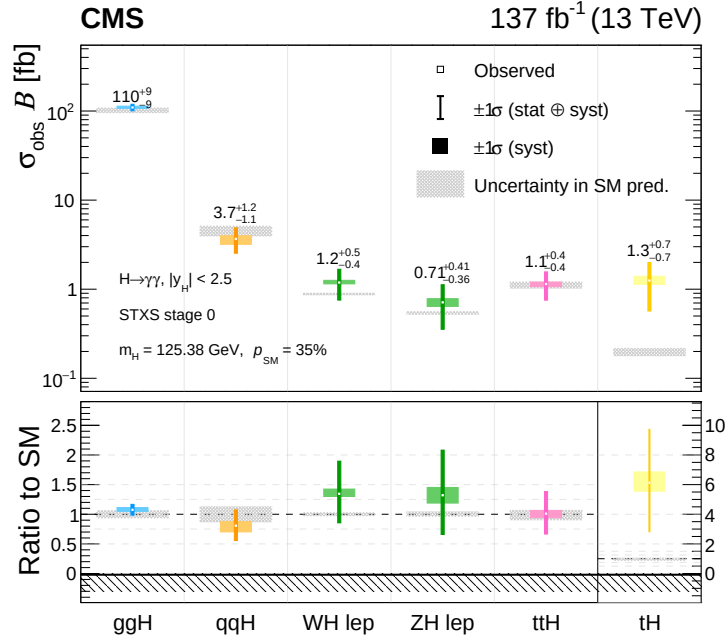


Figure 6.9: Observed results of the STXS at the stage 0 in the $H \rightarrow \gamma\gamma$ channel with $m_H = 125.38$ GeV and without $|y_H| > 2.5$ region by combining 2016-2018 year scenarios with a luminosity of 137 fb^{-1} [72]. Each result contains the best fit value and its systematic-only and statistical-plus-systematic uncertainties, with a colorful block and bar, respectively, where the color scheme follows Figure 6.8. The uncertainties from the Standard Model prediction are shown by the gray blocks. The signal compatibility for the STXS at the stage 0 with respect to the Standard Model prediction is expressed by a p -value of 35 %. The ratio of the observed STXS with respect to the Standard Model prediction is also shown in the bottom panel.

channels. All STXS bins in this stage are visualized in Figure 6.11, where some bins are merged by placing similar but light color areas behind these STXS bins. For the VH leptonic STXS bins, all $q\bar{q} \rightarrow ZH$ and $gg \rightarrow ZH$ are merged together, while the WH leptonic bins with $p_T^V > 150$ GeV are grouped. This is to avoid low statistics in corresponding analysis categories which leads to sensitivity loss. On the other hand, some STXS bins are adjusted according to different situations during the fitting. For example, it's observed that the significant anti-correlation

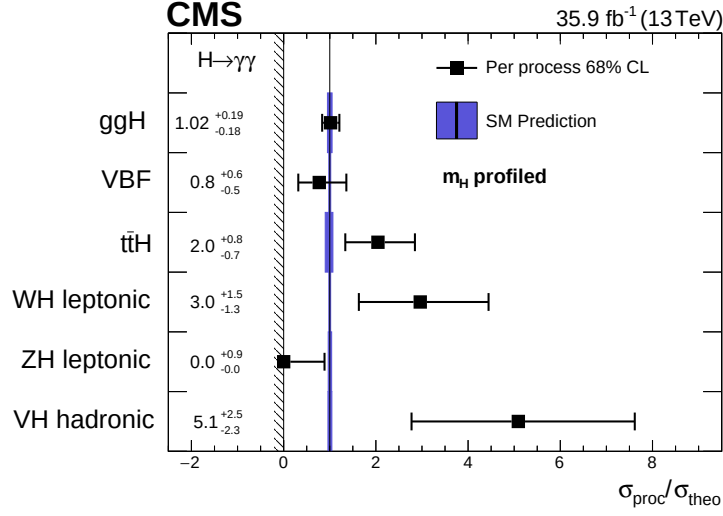


Figure 6.10: Observed results of the STXS divided by the STXS from the Standard Model prediction at the stage 0 in the $H \rightarrow \gamma\gamma$ channel with profiled m_H and without $|y_H| > 2.5$ region for only 2016 year scenario [5]. Each result contains the best fit value and its statistical-plus-systematic uncertainties, with a black bar. The uncertainties from the Standard Model prediction are shown by the purple blocks.

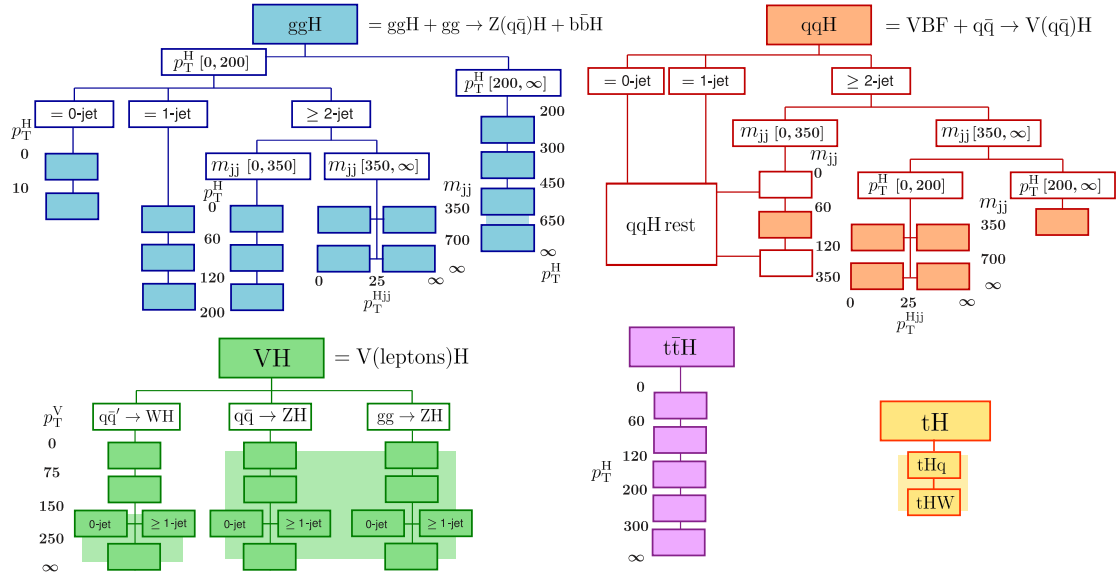


Figure 6.11: Schematic diagram for the STXS stage 1.2 bins [72]. Some bins are merged by placing similar but light color areas behind these STXS bins.

between the ggH and qqH bins occurs at the $m_{jj} > 350$ GeV plus $0 < p_T^H < 200$ GeV. These ggH and qqH bins with the same kinematics are merged one-by-one as the ggH and qqH bins with $350 < m_{jj} < 700$ GeV plus $0 < p_T^{Hjj} < 25$ GeV are merged and so on. Moreover, since there is no analysis category targeting the qqH bins with either $N_{jet} \leq 1$ or $0 < m_{jj} < 60$ GeV and $120 < m_{jj} < 350$ GeV, described as “ qqH -rest” in Figure 6.11, these bins are fixed to

the Standard Model prediction. Therefore, total 27 parameters of interest are fitted to all 80 analysis categories in data.

The observed results are revealed in Figure 6.13, where the complete results are put in Appendix A and the corresponding p -value with respect to the SM is up to 70 % which is compatible to the SM prediction. The observed cross sections of five VH STXS bins in this stage are also extracted in Table 6.2. Since the ZH leptonic bin at this stage is the same as that at the stage 0, the observed cross sections for both are almost the same as its corresponding analysis category is quite pure. Besides, the observed STXS for the VH hadronic bin, i.e. qqH VH -like bin, is first extracted from the qqH bins at this stage and also gives significant improvement with respect to that shown in Figure 6.10. More importantly, among the WH leptonic STXS bins, although the observed STXS in the high- p_T region which is BSM sensitive, i.e. the WH leptonic bin ($p_T^V > 150$ GeV), has no any excess observed, these WH leptonic STXS results are the most granularity for the WH leptonic STXS analysis at the LHC currently. In summary, by decoupling the kinematic structure of the WH leptonic process through the STXS framework, the SM prediction is successful again.

Parameters	SM prediction ($m_H = 125.38$ GeV)	$\sigma\mathcal{B}$ (fb)			$\sigma\mathcal{B}/(\sigma\mathcal{B})_{SM}$	
		Observed (Expected)		Syst. unc.	Observed (Expected)	
		Best fit	Stat. unc.		Best fit	
qqH VH-like	$1.22^{+0.05}_{-0.05}$	$1.53^{+1.20}_{-1.21} \left(\begin{smallmatrix} +1.13 \\ -1.27 \end{smallmatrix} \right)$	$\begin{smallmatrix} +1.20 & (+1.12) \\ -1.20 & (-1.27) \end{smallmatrix}$	$\begin{smallmatrix} +0.11 & (+0.05) \\ -0.19 & (-0.08) \end{smallmatrix}$	$1.25^{+0.98}_{-0.99} \left(\begin{smallmatrix} +0.92 \\ -1.04 \end{smallmatrix} \right)$	
WH lep $p_T^V < 75$	$0.47^{+0.02}_{-0.02}$	$0.71^{+0.68}_{-0.54} \left(\begin{smallmatrix} +0.75 \\ -0.47 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.68 & (+0.75) \\ -0.54 & (-0.47) \end{smallmatrix}$	$\begin{smallmatrix} +0.05 & (+0.05) \\ -0.05 & (-0.05) \end{smallmatrix}$	$1.51^{+1.45}_{-1.15} \left(\begin{smallmatrix} +1.60 \\ -1.00 \end{smallmatrix} \right)$	
WH lep $75 < p_T^V < 150$	$0.29^{+0.02}_{-0.02}$	$0.44^{+0.37}_{-0.32} \left(\begin{smallmatrix} +0.37 \\ -0.29 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.37 & (+0.37) \\ -0.32 & (-0.29) \end{smallmatrix}$	$\begin{smallmatrix} +0.03 & (+0.02) \\ -0.02 & (-0.02) \end{smallmatrix}$	$1.49^{+1.26}_{-1.08} \left(\begin{smallmatrix} +1.25 \\ -1.00 \end{smallmatrix} \right)$	
WH lep $p_T^V > 150$	$0.12^{+0.01}_{-0.01}$	$0.12^{+0.12}_{-0.09} \left(\begin{smallmatrix} +0.13 \\ -0.10 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.12 & (+0.13) \\ -0.09 & (-0.10) \end{smallmatrix}$	$\begin{smallmatrix} +0.01 & (+0.01) \\ -0.00 & (-0.01) \end{smallmatrix}$	$0.98^{+0.98}_{-0.76} \left(\begin{smallmatrix} +1.05 \\ -0.79 \end{smallmatrix} \right)$	
ZH lep	$0.54^{+0.03}_{-0.02}$	$0.71^{+0.41}_{-0.35} \left(\begin{smallmatrix} +0.42 \\ -0.35 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.40 & (+0.41) \\ -0.35 & (-0.35) \end{smallmatrix}$	$\begin{smallmatrix} +0.07 & (+0.04) \\ -0.03 & (-0.04) \end{smallmatrix}$	$1.32^{+0.76}_{-0.65} \left(\begin{smallmatrix} +0.77 \\ -0.65 \end{smallmatrix} \right)$	

Table 6.2: Results of the STXS stage 1.2 [72]. The best fit cross sections are shown together with the respective 68% CL intervals. The uncertainty is decomposed into the systematic and statistical components. The expected uncertainties on the fitted parameters are given in brackets. The SM predictions for the cross sections and the theoretical uncertainty in those predictions are also listed.

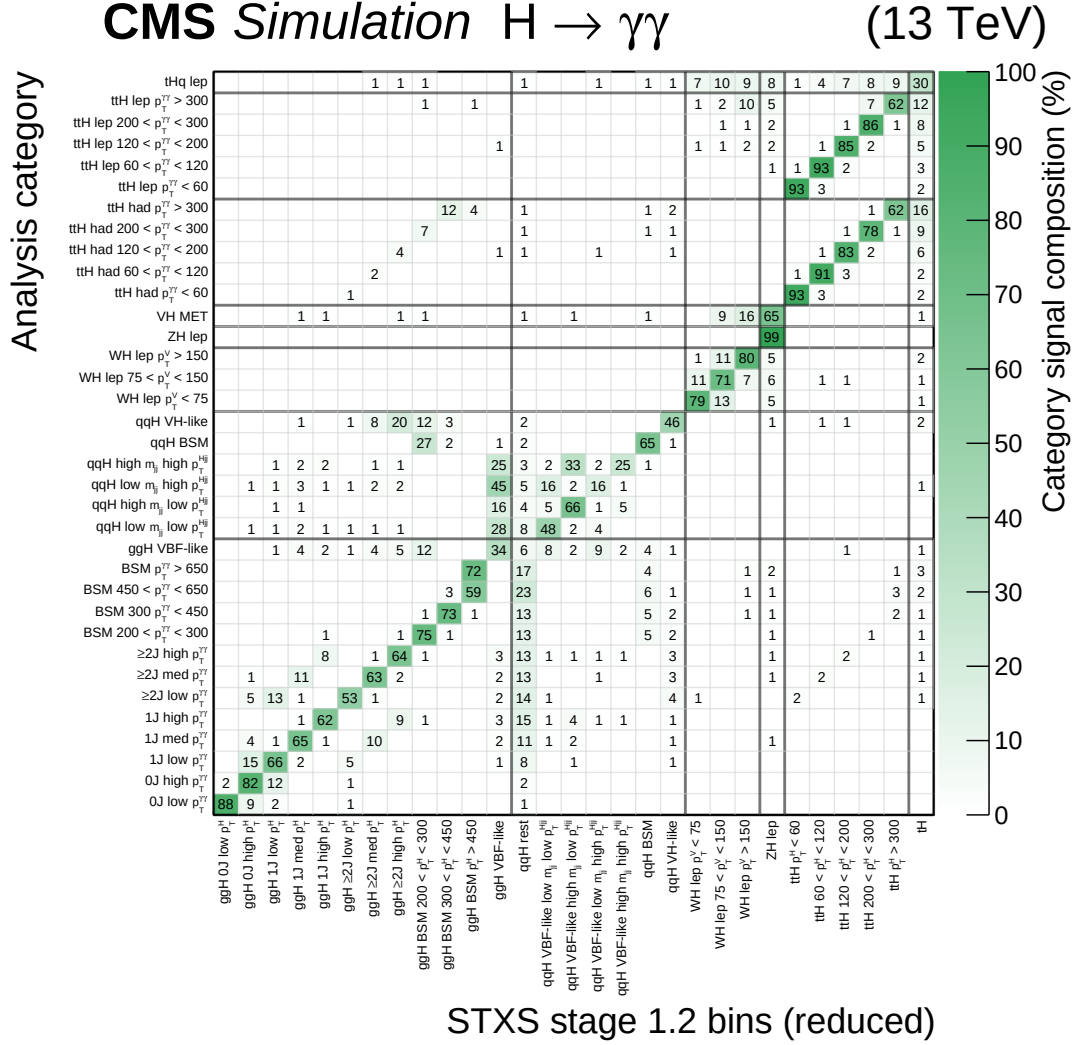


Figure 6.12: The composition of the analysis categories in terms of a merged set of STXS bins is shown in [72]. The granularity of the STXS bin merging corresponds to the finest granularity used for the cross section measurements in this analysis. Analysis categories targeting a common STXS region are summed, where the signal compositions of the individual categories are weighted in the sum by the expected ratio of signal to signal-plus-background events. The color scale corresponds to the fractional yield in each analysis category group (rows) accounted for by each STXS process (columns). Each row therefore sums to 100%. Entries with values less than 0.5% are not shown. Simulated events for each year in the period from 2016 to 2018 are combined with appropriate weights corresponding to their relative integrated luminosity in data. The column labelled as qqH -rest includes contributions from the qqH 0J, qqH 1J, qqH $m_{jj} < 60$ GeV and qqH $120 < m_{jj} < 350$ GeV STXS bins.

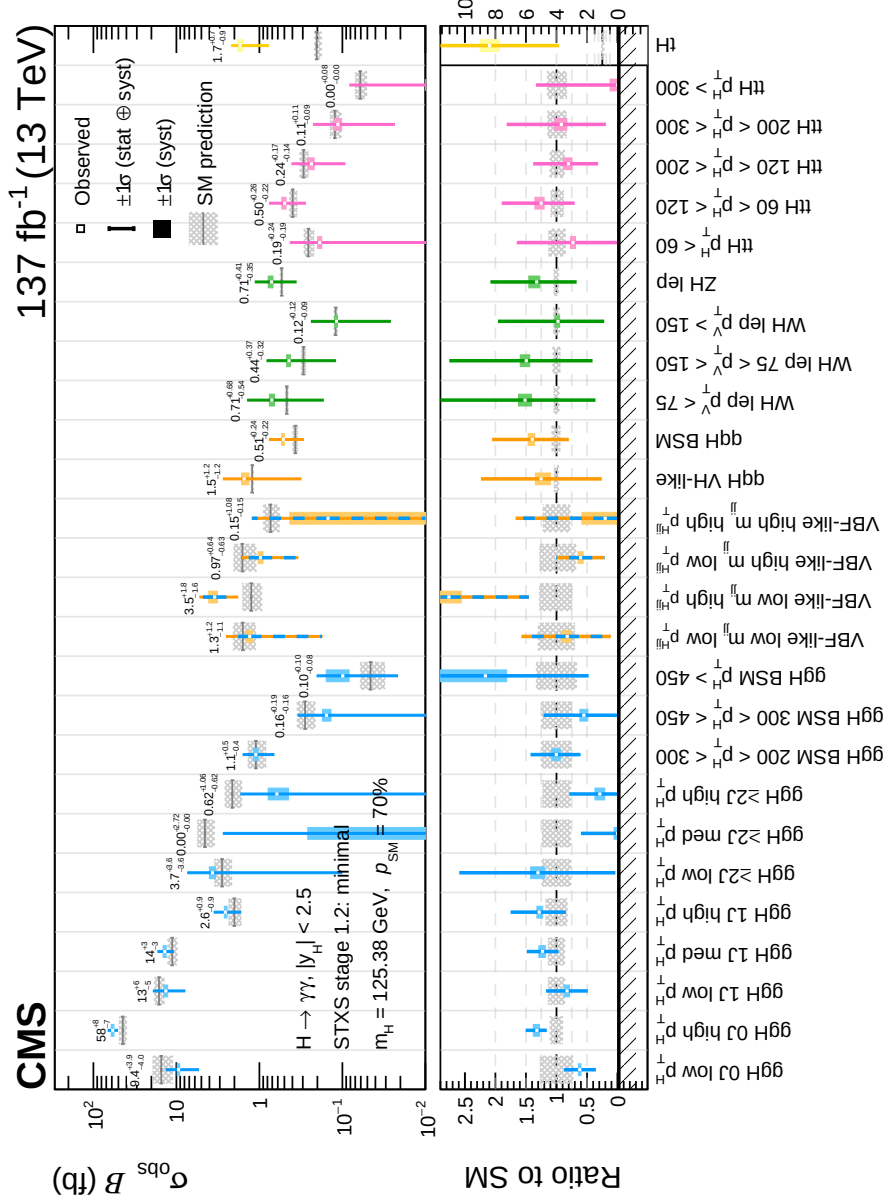


Figure 6.13: Observed results of the STXS at the stage 1.2 in the $H \rightarrow \gamma\gamma$ channel with $m_H = 125.38$ GeV and without $|\eta_H| > 2.5$ region [72]. Each result contains the best fit value and its systematic-only and statistical-plus-systematic uncertainties, with a colorful block and bar, respectively, where the color scheme follows Figure 6.11. The uncertainties from the Standard Model prediction are shown by the gray blocks. The signal compatibility for the STXS at the stage 1.2 with respect to the Standard Model prediction is expressed by a p -value of 70 %. The ratio of the observed STXS with respect to the Standard Model prediction is also shown in the bottom panel.

Conclusion

The cross section measurements of the associated VH production, with the $H \rightarrow \gamma\gamma$ channel are shown in this thesis. The measurements use events with two isolated photons from the proton-proton collisions at a center-of-mass energy of 13 TeV collected by the CMS detector during 2016-2018 corresponding to an integrated luminosity of 137 fb^{-1} .

The analysis strategy performs the simplified template cross section framework to optimize the sensitivity of the associated VH production and reduce the theoretical dependence from other Higgs processes. In the STXS framework, the VH production cross sections are split into different STXS bins by different productions and the transverse momenta of the W and Z bosons under the particle level. The analysis splits events into different analysis categories according to the final states of the W and Z bosons and their transverse momenta under the reconstruction level. The Boost Decision Tree (BDT) classifiers are applied to these analysis categories to define finer and more signal enriched analysis categories. Such application not only reduces non-resonance backgrounds but excludes non-target STXS bins for maximizing the sensitivity of the VH process. Finally, the state-of-the-art combination fit is performed to extract the observed cross sections of the associated VH production.

For the final results, the combination fit including four relevant Higgs boson productions gives an inclusive signal strength modifier of the $H \rightarrow \gamma\gamma$ production of

$$\mu = 1.12^{+0.09}_{-0.09} = 1.12^{+0.06}_{-0.06}(\text{theo})^{+0.03}_{-0.03}(\text{syst})^{+0.07}_{-0.06}(\text{stat})$$

where the total uncertainty is decoupled to different sources. This result satisfies the Standard model prediction and its uncertainty starts to be limited by its systematic uncertainty. The signal strength modifier of four relevant Higgs boson production is also revealed in Figure 6.5, especially, that of the associated VH production given as

$$\mu_{VH} = 1.34^{+0.36}_{-0.33} = 1.34^{+0.07}_{-0.06}(\text{theo})^{+0.05}_{-0.04}(\text{syst})^{+0.35}_{-0.32}(\text{stat})$$

where the observed significance is up to a standard deviation of 4.7. This result almost announces that the Higgs-Strahlung process is observed under the $H \rightarrow \gamma\gamma$ channel. On the other hand, the observed STXS with the minimum merging scheme and the scheme with the most granularity at the LHC collaborations are also shown in Figure 6.9 and 6.13, respectively, revealing the kinematic structure of the associated VH production. The compatibility of both schemes with respect to the Standard Model is given by two p -values of 35 and 70 %, respec-

tively. In particular, from the most granularity scheme, it's observed that there is no excess in the highest energy scale bin of the WH production, " $WH \text{ lep } p_T^V > 150$ " bin, where physics beyond the Standard Model (BSM) is sensitive. Therefore, all results are compatible to the Standard Model prediction and the Standard Model is successful again.

One reason that BSM effect can't be observed in this analysis could be the limited branching ratio ($\approx 0.23\%$) of the diphoton channel in the Higgs boson decay. This reason leads to large statistic uncertainties in the STXS measurements with the most granularity. These imprecise results due to low statistics could dilute the relatively small effects from BSM. For a certainty, two isolated photons give pretty clean event topology against the QCD events. The associated VH production with the W and Z decaying leptonically further provides higher purity and excludes more backgrounds. Hence, once this analysis moves on more statistics periods as the Run III with expected twice luminosity with respect to Run II or even the HL-LHC era with interpreted luminosity up to 3000 fb^{-1} , more than 20 times, this channel will become more important. More statistics not only enables the measured cross sections to be more precise and granularity but provides more sensitivity to the effective field theory in order to indirectly explore BSM hidden in the HVV coupling.

Another reason is that BSM effect in high energy region makes the Higgs boson become boosted such as heavy vector bosons decaying to the VH final states. Two photons from the boosted Higgs boson could be close to each other and even collinear ($\Delta R < 0.3$). Unfortunately, this kind of events is killed by the selections and trigger requirements which aim at two isolated photons ($\Delta R > 0.3$). Hence, in future, the analysis strategy for the VH part needs be modified to extend additional phase space with such event topology. In any case, the associated VH production with $H \rightarrow \gamma\gamma$ channel will be one of the best candidates for understanding the HVV coupling and BSM effect in the long term.

Appendix A

STXS results

The complete results of the STXS stage 0 and 1.2 in the $H \rightarrow \gamma\gamma$ channel are shown in Table A.1 and Table A.2, respectively. Among these results, the best fit cross sections are shown together with the respective 68% CL intervals. The uncertainty is decomposed into the systematic and statistical components. The expected uncertainties on the fitted parameters, computed assuming the SM predicted cross section values, are given in brackets. Also listed are the SM predictions for the cross sections and the theoretical uncertainty in those predictions.

Parameters	SM prediction ($m_H = 125.38$ GeV)	$\sigma\mathcal{B}$ (fb)			$\sigma\mathcal{B}/(\sigma\mathcal{B})_{\text{SM}}$	
		Observed (Expected)		Syst. unc.	Observed (Expected)	
		Best fit	Stat. unc.		Best fit	
ggH	$102.02^{+6.80}_{-6.80}$	$109.65^{+9.41}_{-9.17} \left(\begin{smallmatrix} +10.02 \\ -9.72 \end{smallmatrix} \right)$	$\begin{smallmatrix} +8.59 & (+9.34) \\ -8.76 & (-9.20) \end{smallmatrix}$	$\begin{smallmatrix} +3.84 & (+3.64) \\ -2.73 & (-3.14) \end{smallmatrix}$	$1.07^{+0.09}_{-0.09} \left(\begin{smallmatrix} +0.10 \\ -0.10 \end{smallmatrix} \right)$	
qqH	$4.55^{+0.60}_{-0.61}$	$3.67^{+1.21}_{-1.14} \left(\begin{smallmatrix} +1.22 \\ -1.19 \end{smallmatrix} \right)$	$\begin{smallmatrix} +1.17 & (+1.14) \\ -1.03 & (-1.12) \end{smallmatrix}$	$\begin{smallmatrix} +0.31 & (+0.43) \\ -0.48 & (-0.40) \end{smallmatrix}$	$0.81^{+0.27}_{-0.25} \left(\begin{smallmatrix} +0.27 \\ -0.26 \end{smallmatrix} \right)$	
WH lep	$0.88^{+0.03}_{-0.03}$	$1.19^{+0.49}_{-0.43} \left(\begin{smallmatrix} +0.51 \\ -0.42 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.48 & (+0.50) \\ -0.43 & (-0.42) \end{smallmatrix}$	$\begin{smallmatrix} +0.07 & (+0.07) \\ -0.04 & (-0.04) \end{smallmatrix}$	$1.35^{+0.55}_{-0.49} \left(\begin{smallmatrix} +0.57 \\ -0.47 \end{smallmatrix} \right)$	
ZH lep	$0.54^{+0.03}_{-0.02}$	$0.71^{+0.41}_{-0.36} \left(\begin{smallmatrix} +0.42 \\ -0.35 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.40 & (+0.41) \\ -0.35 & (-0.35) \end{smallmatrix}$	$\begin{smallmatrix} +0.07 & (+0.06) \\ -0.07 & (-0.02) \end{smallmatrix}$	$1.32^{+0.76}_{-0.66} \left(\begin{smallmatrix} +0.77 \\ -0.65 \end{smallmatrix} \right)$	
ttH	$1.13^{+0.08}_{-0.11}$	$1.14^{+0.42}_{-0.39} \left(\begin{smallmatrix} +0.42 \\ -0.39 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.42 & (+0.41) \\ -0.38 & (-0.39) \end{smallmatrix}$	$\begin{smallmatrix} +0.06 & (+0.09) \\ -0.08 & (-0.04) \end{smallmatrix}$	$1.01^{+0.37}_{-0.34} \left(\begin{smallmatrix} +0.37 \\ -0.35 \end{smallmatrix} \right)$	
tH	$0.20^{+0.01}_{-0.03}$	$1.25^{+0.74}_{-0.68} \left(\begin{smallmatrix} +0.71 \\ -0.20 \end{smallmatrix} \right)$	$\begin{smallmatrix} +0.73 & (+0.71) \\ -0.67 & (-0.20) \end{smallmatrix}$	$\begin{smallmatrix} +0.15 & (+0.09) \\ -0.12 & (-0.09) \end{smallmatrix}$	$6.13^{+3.63}_{-3.34} \left(\begin{smallmatrix} +3.49 \\ -1.00 \end{smallmatrix} \right)$	

Table A.1: Results of the STXS stage 0 [72].

Parameters	SM prediction ($m_H = 125.38$ GeV)	$\sigma\mathcal{B}$ (fb)			$\sigma\mathcal{B}/(\sigma\mathcal{B})_{\text{SM}}$		
		Observed (Expected)		Syst. unc.	Observed (Expected)		Best fit
		Best fit	Stat. unc.		Best fit	Stat. unc.	
ggH 0J low p_T^H	$15.21^{+4.14}_{-4.18}$	$9.41^{+3.91}_{-4.00}$ ($+4.19$ -4.06)	$+3.90$ ($+4.16$ -3.99)	$+0.37$ ($+0.50$ -0.30)	$0.62^{+0.26}_{-0.26}$ ($+0.28$ -0.27)		
ggH 0J high p_T^H	$44.25^{+4.84}_{-4.61}$	$58.46^{+8.12}_{-7.17}$ ($+7.87$ -7.78)	$+7.69$ ($+7.66$ -6.91)	$+2.60$ ($+1.78$ -1.94)	$1.32^{+0.18}_{-0.16}$ ($+0.18$ -0.18)		
ggH 1J low p_T^H	$16.20^{+2.25}_{-2.27}$	$13.40^{+5.59}_{-5.50}$ ($+5.70$ -5.58)	$+5.53$ ($+5.64$ -5.46)	$+0.79$ ($+0.77$ -0.67)	$0.83^{+0.34}_{-0.34}$ ($+0.35$ -0.34)		
ggH 1J med p_T^H	$11.23^{+1.56}_{-1.55}$	$13.80^{+2.90}_{-2.94}$ ($+3.14$ -3.41)	$+2.82$ ($+3.08$ -2.90)	$+0.68$ ($+0.59$ -0.51)	$1.23^{+0.26}_{-0.26}$ ($+0.28$ -0.30)		
ggH 1J high p_T^H	$2.00^{+0.36}_{-0.36}$	$2.57^{+0.94}_{-0.88}$ ($+0.92$ -0.90)	$+0.94$ ($+0.91$ -0.87)	$+0.08$ ($+0.13$ -0.12)	$1.28^{+0.47}_{-0.44}$ ($+0.46$ -0.45)		
ggH ≥ 2 J low p_T^H	$2.82^{+0.68}_{-0.68}$	$3.67^{+3.63}_{-3.57}$ ($+3.74$ -2.82)	$+3.62$ ($+3.71$ -3.56)	$+0.34$ ($+0.49$ -0.30)	$1.30^{+1.29}_{-1.27}$ ($+1.33$ -1.00)		
ggH ≥ 2 J med p_T^H	$4.53^{+1.07}_{-1.07}$	$0.00^{+2.72}_{-0.00}$ ($+2.90$ -2.80)	$+2.71$ ($+2.86$ -0.00)	$+0.26$ ($+0.45$ -0.00)	$0.00^{+0.60}_{-0.00}$ ($+0.64$ -0.62)		
ggH ≥ 2 J high p_T^H	$2.12^{+0.49}_{-0.50}$	$0.62^{+1.06}_{-0.62}$ ($+1.15$ -1.10)	$+1.04$ ($+1.11$ -0.62)	$+0.17$ ($+0.30$ -0.17)	$0.29^{+0.50}_{-0.29}$ ($+0.54$ -0.52)		
ggH BSM $200 < p_T^H < 300$	$1.10^{+0.28}_{-0.27}$	$1.11^{+0.47}_{-0.44}$ ($+0.56$ -0.45)	$+0.46$ ($+0.56$ -0.43)	$+0.08$ ($+0.05$ -0.07)	$1.00^{+0.42}_{-0.40}$ ($+0.51$ -0.41)		
ggH BSM $300 < p_T^H < 450$	$0.28^{+0.07}_{-0.07}$	$0.16^{+0.19}_{-0.16}$ ($+0.20$ -0.18)	$+0.18$ ($+0.19$ -0.16)	$+0.02$ ($+0.03$ -0.02)	$0.55^{+0.66}_{-0.55}$ ($+0.69$ -0.65)		
ggH BSM $p_T^H > 450$	$0.05^{+0.02}_{-0.02}$	$0.10^{+0.10}_{-0.08}$ ($+0.10$ -0.05)	$+0.09$ ($+0.09$ -0.08)	$+0.06$ ($+0.04$ -0.02)	$2.16^{+2.25}_{-1.69}$ ($+2.19$ -1.00)		
VBF-like low m_{ij} low p_T^{Hij}	$1.59^{+0.49}_{-0.48}$	$1.31^{+1.19}_{-1.13}$ ($+1.22$ -1.16)	$+1.18$ ($+1.21$ -1.13)	$+0.14$ ($+0.13$ -0.09)	$0.82^{+0.75}_{-0.71}$ ($+0.77$ -0.73)		
VBF-like low m_{ij} high p_T^{Hij}	$1.25^{+0.35}_{-0.32}$	$3.46^{+1.76}_{-1.64}$ ($+1.79$ -1.25)	$+1.65$ ($+1.76$ -1.62)	$+0.61$ ($+0.32$ -0.25)	$2.76^{+1.40}_{-1.31}$ ($+1.43$ -1.00)		
VBF-like high m_{ij} low p_T^{Hij}	$1.60^{+0.45}_{-0.51}$	$0.97^{+0.64}_{-0.63}$ ($+0.72$ -0.63)	$+0.63$ ($+0.71$ -0.62)	$+0.07$ ($+0.11$ -0.07)	$0.61^{+0.40}_{-0.39}$ ($+0.45$ -0.39)		
VBF-like high m_{ij} high p_T^{Hij}	$0.73^{+0.16}_{-0.16}$	$0.15^{+1.08}_{-0.15}$ ($+0.93$ -0.73)	$+1.04$ ($+0.92$ -0.15)	$+0.28$ ($+0.14$ -0.15)	$0.20^{+1.47}_{-0.20}$ ($+1.26$ -1.00)		
qqH VH-like	$1.22^{+0.05}_{-0.05}$	$1.53^{+1.20}_{-1.21}$ ($+1.13$ -1.27)	$+1.20$ ($+1.12$ -1.20)	$+0.11$ ($+0.05$ -0.19)	$1.25^{+0.98}_{-0.99}$ ($+0.92$ -1.04)		
qqH BSM	$0.37^{+0.03}_{-0.02}$	$0.51^{+0.24}_{-0.22}$ ($+0.24$ -0.24)	$+0.24$ ($+0.24$ -0.22)	$+0.03$ ($+0.03$ -0.01)	$1.40^{+0.66}_{-0.60}$ ($+0.66$ -0.65)		
WH lep $p_T^V < 75$	$0.47^{+0.02}_{-0.02}$	$0.71^{+0.68}_{-0.54}$ ($+0.75$ -0.47)	$+0.68$ ($+0.75$ -0.54)	$+0.05$ ($+0.05$ -0.05)	$1.51^{+1.45}_{-1.15}$ ($+1.60$ -1.00)		
WH lep $75 < p_T^V < 150$	$0.29^{+0.02}_{-0.02}$	$0.44^{+0.37}_{-0.32}$ ($+0.37$ -0.29)	$+0.37$ ($+0.37$ -0.32)	$+0.03$ ($+0.02$ -0.02)	$1.49^{+1.26}_{-1.08}$ ($+1.25$ -1.00)		
WH lep $p_T^V > 150$	$0.12^{+0.01}_{-0.01}$	$0.12^{+0.12}_{-0.09}$ ($+0.13$ -0.10)	$+0.12$ ($+0.13$ -0.09)	$+0.01$ ($+0.01$ -0.00)	$0.98^{+0.98}_{-0.76}$ ($+1.05$ -0.79)		
ZH lep	$0.54^{+0.03}_{-0.02}$	$0.71^{+0.41}_{-0.35}$ ($+0.42$ -0.35)	$+0.40$ ($+0.41$ -0.35)	$+0.07$ ($+0.04$ -0.03)	$1.32^{+0.76}_{-0.65}$ ($+0.77$ -0.65)		
$t\bar{t}H$ $p_T^H < 60$	$0.26^{+0.03}_{-0.04}$	$0.19^{+0.24}_{-0.19}$ ($+0.23$ -0.19)	$+0.24$ ($+0.23$ -0.19)	$+0.01$ ($+0.03$ -0.01)	$0.73^{+0.92}_{-0.73}$ ($+0.89$ -0.75)		
$t\bar{t}H$ $60 < p_T^H < 120$	$0.40^{+0.04}_{-0.05}$	$0.50^{+0.26}_{-0.22}$ ($+0.28$ -0.22)	$+0.25$ ($+0.28$ -0.22)	$+0.04$ ($+0.03$ -0.02)	$1.25^{+0.65}_{-0.55}$ ($+0.72$ -0.55)		
$t\bar{t}H$ $120 < p_T^H < 200$	$0.29^{+0.03}_{-0.04}$	$0.24^{+0.17}_{-0.14}$ ($+0.17$ -0.16)	$+0.17$ ($+0.17$ -0.14)	$+0.02$ ($+0.02$ -0.02)	$0.80^{+0.58}_{-0.49}$ ($+0.58$ -0.53)		
$t\bar{t}H$ $200 < p_T^H < 300$	$0.12^{+0.02}_{-0.02}$	$0.11^{+0.11}_{-0.09}$ ($+0.10$ -0.09)	$+0.11$ ($+0.10$ -0.09)	$+0.01$ ($+0.01$ -0.01)	$0.92^{+0.89}_{-0.73}$ ($+0.81$ -0.75)		
$t\bar{t}H$ $p_T^H > 300$	$0.06^{+0.01}_{-0.01}$	$0.00^{+0.08}_{-0.00}$ ($+0.07$ -0.06)	$+0.08$ ($+0.07$ -0.00)	$+0.01$ ($+0.02$ -0.00)	$0.00^{+1.34}_{-0.00}$ ($+1.21$ -1.00)		
tH	$0.20^{+0.01}_{-0.03}$	$1.71^{+0.71}_{-0.93}$ ($+1.01$ -0.20)	$+0.70$ ($+1.00$ -0.92)	$+0.13$ ($+0.11$ -0.13)	$8.38^{+3.48}_{-4.55}$ ($+4.93$ -1.00)		

Table A.2: Results of the STXS stage 1.2 [72].

Bibliography

- [1] CMS Collaboration, “A New Boson with a Mass of 125 GeV Observed with the CMS Experiment at the Large Hadron Collider”, *Science* 338 (2012) 1569–1575.
- [2] ATLAS Collaboration, “A particle consistent with the Higgs Boson observed with the ATLAS Detector at the Large Hadron Collider”, *Science* 338 (2012) 1576–1582.
- [3] Particle Data Group, “Review of Particle Physics”, *PTEP* 2020.8 (2020) 083C01.
- [4] CMS Collaboration, “Observation of the Diphoton Decay of the Higgs Boson and Measurement of Its Properties”, *Eur. Phys. J. C* 74.10 (2014) 3076.
- [5] CMS Collaboration, “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* 11 (2018) 185.
- [6] Nicolas Berger et al., “Simplified Template Cross Sections - Stage 1.1” (June 2019).
- [7] Mark Thomson, *Modern Particle Physics*, Cambridge University Press, 2013.
- [8] Matthew D. Schwartz, *Quantum Field Theory and the Standard Model*, Cambridge University Press, 2013.
- [9] Benjamin W. Lee, C. Quigg, and H. B. Thacker, “The Strength of Weak Interactions at Very High-Energies and the Higgs Boson Mass”, *Phys. Rev. Lett.* 38 (1977) 883–885.
- [10] LEP Working Group for Higgs boson searches, “Search for the standard model Higgs boson at LEP”, *Phys. Lett. B* 565 (2003) 61–75.
- [11] “Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data” (July 2011).
- [12] M. Baak et al., “Updated Status of the Global Electroweak Fit and Constraints on New Physics”, *Eur. Phys. J. C* 72 (2012) 2003.
- [13] CMS Collaboration ATLAS, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* 114 (2015) 191803.
- [14] CMS Collaboration, “Measurement of Higgs Boson Production and Properties in the WW Decay Channel with Leptonic Final States”, *JHEP* 01 (2014) 096.
- [15] ATLAS Collaboration, “Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector”, *Phys. Rev. D* 92.1 (2015) 012006.

-
- [16] CMS Collaboration, “Measurement of Higgs boson production in association with a W or Z boson in the $H \rightarrow WW$ decay channel”, *CMS Physics Analysis Summary CMS-PAS-HIG-19-017* (2021).
 - [17] ATLAS Collaboration, “Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector”, *Phys. Rev. D* **91.1** (2015) 012006.
 - [18] CMS Collaboration, “Observation of Higgs boson decay to bottom quarks”, *Phys. Rev. Lett.* **121.12** (2018) 121801.
 - [19] ATLAS Collaboration, “Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector”, *Phys. Lett. B* **786** (2018) 59–86.
 - [20] CMS Collaboration, “Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector”, *Phys. Lett. B* **779** (2018) 283–316.
 - [21] ATLAS Collaboration, “Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Rev. D* **99** (2019) 072001.
 - [22] CMS Collaboration, “Observation of $t\bar{t}H$ production”, *Phys. Rev. Lett.* **120.23** (2018) 231801.
 - [23] ATLAS Collaboration, “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector”, *Phys. Lett. B* **784** (2018) 173–191.
 - [24] CMS Collaboration, “Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel”, *Phys. Rev. Lett.* **125.6** (2020) 061801.
 - [25] CMS Collaboration, “A search for the standard model Higgs boson decaying to charm quarks”, *JHEP* **03** (2020) 131.
 - [26] ATLAS Collaboration, “Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment”, *Phys. Rev. Lett.* **120.21** (2018) 211802.
 - [27] CMS Collaboration, “Evidence for Higgs boson decay to a pair of muons”, *JHEP* **01** (2021) 148.
 - [28] ATLAS Collaboration, “CP Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \rightarrow \gamma\gamma$ with the ATLAS Detector”, *Phys. Rev. Lett.* **125.6** (2020) 061802.
 - [29] CMS Collaboration, “Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to τ lepton pairs”, *Phys. Rev. D* **100.11** (2019) 112002.
 - [30] CMS Collaboration, “Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state”, *Phys. Rev. D* **99.11** (2019) 112003.
 - [31] CMS Collaboration, “Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV” (2020).

- [32] CMS Collaboration, “Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state” ([Apr. 2021](#)).
- [33] CMS Collaboration, “Constraints on Higgs boson properties using $WW^*(\rightarrow e\nu\mu\nu)jj$ production in 36.1fb^{-1} of $\sqrt{s} = 13\text{TeV}$ pp collisions with the ATLAS detector” ([Nov. 2020](#)).
- [34] CMS Collaboration, “Test of CP invariance in vector-boson fusion production of the Higgs boson in the $H \rightarrow \tau\tau$ channel in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector”, *Phys. Lett. B* 805 (2020) 135426.
- [35] CMS Collaboration, “Combination of searches for Higgs boson pair production in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$ ”, *Phys. Rev. Lett.* 122.12 (2019) 121803.
- [36] ATLAS Collaboration, “Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector”, *Phys. Lett. B* 800 (2020) 135103.
- [37] CMS Collaboration, “Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$ ”, *JHEP* 03 (2021) 257.
- [38] ATLAS Collaboration, “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 $\sqrt{s} = 13\text{ TeV}$ ”, *Phys. Lett. B* 801 (2020) 135145.
- [39] ATLAS Collaboration, “Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton–proton collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector”, *JHEP* 07 (2020) 108.
- [40] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”, 2/2017 ([Oct. 2016](#)).
- [41] Duccio Pappadopulo et al., “Heavy Vector Triplets: Bridging Theory and Data”, *JHEP* 09 (2014) 060.
- [42] You-Ying Li, Rosy Nicolaïdou, and Stathes Paganis, “Exclusion of heavy, broad resonances from precise measurements of WZ and VH final states at the LHC”, *Eur. Phys. J. C* 79.4 (2019) 348.
- [43] Xing-Fu Su et al., “High P_T Higgs excess as a signal of non-local QFT at the LHC”, *Eur. Phys. J. C* 81 (2021) 796.
- [44] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”, *CERN* (2016).
- [45] ATLAS Collaboration, “Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector”, *Eur. Phys. J. C* 81.2 (2021) 178.
- [46] ATLAS Collaboration, “Measurement of the associated production of a Higgs boson decaying into b -quarks with a vector boson at high transverse momentum in pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector”, *Phys. Lett. B* 816 (2021) 136204.
- [47] Lyndon Evans and Philip Bryant, “LHC Machine”, *JINST* 3 (2008) S08001.

-
- [48] R. Bruce et al., “Reaching record-low β^* at the CERN Large Hadron Collider using a novel scheme of collimator settings and optics”, *Nucl. Instrum. Meth. A* 848 (2017) 19–30.
 - [49] Jorg Wenninger, “Operation and Configuration of the LHC in Run 2” (Mar. 2019).
 - [50] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *JINST* 15.09 (2020) P09018.
 - [51] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* 3 (2008) S08004.
 - [52] CMS Tracker Group, “The CMS Phase-1 Pixel Detector Upgrade”, *JINST* 16.02 (2021) P02027.
 - [53] “CMS Technical Design Report for the Pixel Detector Upgrade” (Sept. 2012), ed. by David Aaron Matzner Dominguez et al.
 - [54] CMS Collaboration, “The Phase-2 Upgrade of the CMS Tracker” (June 2017), ed. by K. Klein.
 - [55] CMS Collaboration, “The CMS ECAL performance with examples”, *JINST* 9 (2014), ed. by Pietro Govoni et al. C02008.
 - [56] P. Adzic et al., “Energy resolution of the barrel of the CMS electromagnetic calorimeter”, *JINST* 2 (2007) P04004.
 - [57] CMS Collaboration, “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV”, *JINST* 15.05 (2020) P05002.
 - [58] CMS ECAL/HCAL Collaborations, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c”, *Eur. Phys. J. C* 60 (2009) 359–373.
 - [59] CMS Collaboration, “Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays”, *JINST* 5 (2010) T03021.
 - [60] CMS Collaboration, “The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC”, *JINST* 8 (2013) P11002.
 - [61] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* 13.06 (2018) P06015.
 - [62] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* 15.10 (2020) P10017.
 - [63] CMS Collaboration, “The CMS trigger system”, *JINST* 12.01 (2017) P01020.
 - [64] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* 12 (2017) P10003.
 - [65] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth. A* 262 (1987) 444–450.
 - [66] Serguei Chatrchyan et al., “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* 9.10 (2014) P10009.
 - [67] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* 86.11 (1998) 2210–2239.
 - [68] “Muon Reconstruction and Identification Performance with Run-2 data” (July 2020).

-
- [69] CMS Collaboration, “Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV”, *Phys. Lett. B* 713 (2012) 68–90.
- [70] CMS Collaboration, “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”, *JINST* 16.05 (2021) P05014.
- [71] CMS Collaboration, “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV”, *JINST* 8 (2013) P09009.
- [72] CMS Collaboration, “Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV”, *JHEP* 07 (2021) 027.
- [73] Eleftherios Spyromitros-Xioufis et al., “Multi-target regression via input space expansion: treating targets as inputs”, *Machine Learning* 104.1 (2016) 55–98.
- [74] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* 04 (2008) 063.
- [75] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, “FastJet User Manual”, *Eur. Phys. J. C* 72 (2012) 1896.
- [76] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* 13.05 (2018) P05011.
- [77] CMS Collaboration, “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector”, *JINST* 14.07 (2019) P07004.
- [78] Y. Freund, “Boosting a Weak Learning Algorithm by Majority”, *Information and Computation* 121.2 (1995) 256–285, ISSN: 0890-5401.
- [79] P. Speckmayer et al., “The toolkit for multivariate data analysis, TMVA 4”, *J. Phys. Conf. Ser.* 219 (2010), ed. by Jan Gruntorad and Milos Lokajicek 032057.
- [80] Tianqi Chen and Carlos Guestrin, “XGBoost”, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Aug. 2016).
- [81] R. A. Fisher, “On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P”, *J. Royal Stat. Soc.* 85 (1922) 87, ISSN: 09528385.
- [82] P. D. Dauncey et al., “Handling uncertainties in background shapes: the discrete profiling method”, *JINST* 10 (2015) P04015.
- [83] Samuel S Wilks, “The large-sample distribution of the likelihood ratio for testing composite hypotheses”, *The annals of mathematical statistics* 9.1 (1938) 60–62.
- [84] Jon Butterworth et al., “PDF4LHC recommendations for LHC Run II”, *J. Phys. G* 43 (2016) 023001.
- [85] Richard D. Ball et al., “Parton distributions for the LHC Run II”, *JHEP* 04 (2015) 040.
- [86] S. Carrazza et al., “An Unbiased Hessian Representation for Monte Carlo PDFs”, *Eur. Phys. J. C* 75 (2015) 369.
- [87] Albert M Sirunyan et al., “A measurement of the Higgs boson mass in the diphoton decay channel”, *Phys. Lett. B* 805 (2020) 135425.

-
- [88] CMS Collaboration, *CMS luminosity measurements for the 2016 data taking period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.
 - [89] CMS Collaboration, *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.
 - [90] CMS Collaboration, *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.