

A SEARCH FOR THE STANDARD MODEL HIGGS BOSON DECAYING INTO TWO MUONS  
IN THE VECTOR BOSON ASSOCIATED PRODUCTION MODE AT THE CMS EXPERIMENT

By

XUNWU ZUO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTORATE OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2021

© 2021 Xunwu Zuo

## ACKNOWLEDGEMENTS

Still compiling the list of people to thank.

## TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGEMENTS .....	3
LIST OF TABLES.....	7
LIST OF FIGURES .....	14
ABSTRACT .....	15
 CHAPTER	
1    INTRODUCTION .....	17
1.1  Quantum field theory and gauge symmetry .....	18
1.2  The standard model Lagrangian.....	20
1.3  The Higgs mechanism .....	23
1.4  Parameters of the standard model.....	27
2    THE LHC AND CMS .....	30
2.1  The Large Hadron Collider .....	30
2.1.1  The LHC machine .....	30
2.1.2  Luminosity of collisions .....	34
2.1.3  Proton collision environment .....	36
2.2  The Compact Muon Solenoid experiment .....	37
2.2.1  Solenoid magnet .....	39
2.2.2  Silicon tracker.....	40
2.2.3  Calorimeters .....	42
2.2.4  Muon detectors .....	47
2.2.5  Trigger system.....	52
3    OVERVIEW OF THE SEARCH OF $H \rightarrow \mu\mu$ DECAY AT CMS.....	57
3.1  Data and simulation samples.....	59
3.1.1  The simulation of the signal processes .....	61
3.1.2  The simulation of the background processes.....	62
3.2  Exclusive analyses and their strategies .....	63
4    OBJECT RECONSTRUCTION AND IDENTIFICATION .....	67
4.1  CMS object reconstruction .....	68
4.1.1  Tracks .....	69
4.1.2  Primary vertices .....	70
4.1.3  Beamspot.....	71
4.1.4  Muons.....	71
4.1.5  Electrons .....	73
4.1.6  Photons .....	75
4.1.7  Hadrons .....	77
4.1.8  Jets .....	78
4.1.9  B jets .....	80
4.1.10  Hadronic tau decays .....	81

4.1.11	Missing transverse momentum.....	82
4.2	Object selection in the H to dimuon analysis.....	83
4.2.1	Muon selection.....	83
4.2.2	Electron selection .....	84
4.2.3	Jet selection.....	85
5	MUON MOMENTUM CORRECTION AND CALIBRATION .....	87
5.1	Rochester correction.....	88
5.2	FSR recovery .....	88
5.3	GeoFit correction.....	91
5.3.1	Geometry of the track displacement .....	92
5.3.2	Development of GeoFit.....	96
5.3.3	Performance and validation.....	97
5.3.4	GeoFit vs track refit .....	100
5.4	Muon calibration results.....	101
6	SEARCH FOR H <sub>2</sub> MU TARGETING THE VH PRODUCTION MODE .....	111
6.1	Event selection .....	113
6.2	MVA discrimination .....	115
6.2.1	BDT targeting WH → ℓν + μμ signal .....	116
6.2.2	BDT targeting ZH → ℓℓ + μμ signal .....	119
6.2.3	Validation of the BDTs .....	123
6.3	Event Categorization .....	128
6.4	Signal and background modeling .....	129
6.4.1	Signal modeling.....	130
6.4.2	Background modeling.....	132
6.5	Systematic uncertainties.....	135
6.5.1	Signal shape uncertainties .....	136
6.5.2	Signal rate uncertainties .....	136
6.5.3	Uncertainties from background biases .....	141
6.6	Results of the VH analysis .....	145
7	RESULTS OF THE H <sub>2</sub> MU SEARCH .....	150
	REFERENCES .....	154

## LIST OF TABLES

<u>Tables</u>	<u>page</u>
3-1 Production modes of the Higgs boson in $pp$ collisions at the LHC, their cross sections for $m_H = 125\text{GeV}$ , and the corresponding expected number of $H \rightarrow \mu\mu$ decays in the dataset for this analysis (the cross sections multiplying $\mathcal{B}(H \rightarrow \mu\mu)$ and the integrated luminosity ( $137\text{fb}^{-1}$ )). The leptons ( $\ell$ ) in the table refer to electrons or muons. ....	61
3-2 Summary of the specification for the simulated Higgs signal samples. ....	62
3-3 Summary of the specification for the simulated background samples. ....	63
6-1 Selection criteria on muons and electrons in the VH analysis. ....	113
6-2 Event selections for the WH and ZH categories. ....	115
6-3 List of input variables used to train the signal-background separation BDT in the WH category. In this table, $\mu\mu_H$ is the Higgs candidate, $\ell$ is the lepton from the W decay, $\mu_{OS}$ ( $\mu_{SS}$ ) refers to the muons in the Higgs candidate which OS (SS) to the lepton. ....	117
6-4 List of input variables used to train the signal-background separation BDT in the ZH category. In this table, $\mu\mu_H$ is the Higgs candidate, and $\ell\ell_Z$ is the Z candidate. ....	121
6-5 Normalization uncertainties on the Higgs boson production cross sections for various modes at $\sqrt{s} = 13\text{TeV}$ . ....	137
6-6 Uncertainties on different signal components in the WH and ZH channels related to pileup reweight and L1 prefire reweight. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected. ....	138
6-7 Uncertainties on different signal components in the WH and ZH channels related to the muon energy scale. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected. ....	138
6-8 Uncertainties on different signal components in the WH and ZH channels related to lepMVA scale factor. The lepMVA scale factor is the only scale factor applied to correct for the lepton efficiency modeling. The ID scale factor and Isolation scale factor are covered by the lepMVA scale factors. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected. ....	139
6-9 Uncertainties on different signal components in the WH and ZH channels related to b-jet vetoing. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected. ....	139

6-10Uncertainties on different signal components in the WH Cat1 related to jet energy calibration. JEC uncertainties are in general small for the main signals in the WH and ZH channels. WH Cat1 is shown as an example. Numbers in other categories are similar. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.....	140
6-11Summary of the expect and observed limits and significance in each individual subcategory and the combination. .....	149

## LIST OF FIGURES

<u>Figures</u>	<u>page</u>
1-1 The Standard Model of particles physics: 12 elementary fermions and 5 elementary bosons. Plot taken from [1]. .....	18
1-2 The shape of the Higgs potential with different $\mu^2$ parameter configurations. ....	23
1-3 Summary of the CMS measurements on the Higgs coupling to fermions and bosons. Plot taken from [2]. .....	29
2-1 Photo taken from ... (original source unknown). An aerial view of the LHC landscape. The French-Swiss border is indicated by the yellow line and the LHC tunnels are outlined in white. The triangular building complex just below ATLAS in the picture is the main campus of CERN.....	31
2-2 The schematic layout of the LHC. Each octant contains a insertion point. Points 1, 2, 5, and 8 are the locations of collision experiments. Points 3 and 7 are for beam collimation. Hadrons are injected at Points 1 and 8, accelerated at Point 4, and eventually discarded at Point 6. Plot taken from Ref. [3]. .....	32
2-3 Magnetic field of an ideal quadrupole. Plot taken from Ref. [4].....	33
2-4 The evolutions of the peak instantaneous luminosity (upper plot) and the cumulative integrated luminosity (lower plot) of the LHC Run 2 (2015-2018) $pp$ collisions. Plot taken from Ref. [5].....	35
2-5 An illustration of the interactions in a $pp$ collision. The primary interaction in this example is the Higgs boson production in association with a top quark pair, whose Feynman diagram is shown in 3-2. Plot taken from Ref. [6].....	36
2-6 A cutaway diagram of the CMS detector. The design is specified to a design stage called the Phase-1 detector upgrade [7, 8, 9]. Plot taken from Ref. [10].....	38
2-7 A longitudinal section view of the CMS detector, gridded with the CMS coordinates. Plot taken from Ref. [11].....	39
2-8 The magnetic field in CMS displayed in a longitudinal section, operating with a central field strength of 3.8 T. The field value $ B $ is shown on the left and the field lines are shown on the right. Plot taken from Ref. [12]. .....	40
2-9 Sketch of one quarter of the Phase-1 CMS tracking system in r-z view. The pixel detectors are shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively. Plot taken from Ref. [13]. ....	41
2-10 Schematic view of the CMS ECAL. Plot taken from Ref. [14].....	43
2-11 Schematic view of one quarter of the Phase-1 CMS HCAL in r-z view, showing the locations of the HB, HE, HO, and HF calorimeters. Plot taken from Ref. [8]. .....	45
2-12 A longitudinal section view of the current CMS detector specifying the locations of muon chambers. This plot is different from Figure 2-7 only by the GEM detector installed in 2019/2020. Plot taken from Ref. [15].....	47

2-13Left: sketch of a cell of the drift tube detector showing drift lines and isochrones. Right: schematic view of DT superlayers. Plot taken from Ref. [16].	48
2-14Left: layout of a CSC chamber showing arrangement of anode wires and cathode strips. Right: an illustration of the principle of CSC operation. Plots taken from Ref. [16].	50
2-15Schematic view of a double gap RPC. Plot taken from Ref. [16]	51
2-16Left: schematic view showing the GEM structure and its working principle. Right: exploded view of the mechanical design of GEM. Plot taken from Ref. [15].	52
2-17Diagram of the CMS Level-1 trigger structure during Run 2. Plot taken from Ref. [17].	54
3-1 A conceptual plot for the dimuon mass shapes for the signal and the background. The blue line shows the expected background shape, while the red line shows the expected signal shape on top of it.	58
3-2 Main production modes of the Higgs boson.	59
3-3 Examples of minor Higgs boson production modes.	60
3-4 Scheme of the procedure of assigning events to different categories. All events passing the common baseline selection are divided into four mutually exclusive categories: ggH, VBF, VH (WH and ZH), and t̄tH (leptonic and hadronic).	64
4-1 A transverse view of the CMS detector layout in the barrel region, with illustrations of the interactions between different particles and different detector components. Plot taken from Ref. [18].	67
5-1 Performance of the <i>FSR recovery</i> in the simulated $H \rightarrow \mu\mu$ events. The $m_{\mu\mu}$ before and after the <i>FSR recovery</i> are shown for the events that contain at least one FSR photon (left), and for the inclusive signal events (right). Plot taken from Ref. [19].	90
5-2 Performance of the <i>FSR recovery</i> in the $Z \rightarrow \mu\mu$ events that contain FSR photons, in both data and simulation. A good agreement between simulation and data is observed, both before and after the correction. Plot taken from Ref. [19].	90
5-3 Scheme of the $d_0$ definition in CMS. The $(x_0, y_0)$ is the coordinate of a point near the vertex in the frame where the vertex is at $(0, 0)$ , and the $\phi_0$ is the azimuthal angle of the track at $(x_0, y_0)$ .	93
5-4 Scheme of the track geometry in the transverse plane. The blue lines show the geometry of the reconstructed track, compared to the black lines, which are the geometry of the true track. The difference between the blue track and black track is exaggerated in this scheme. The blue track and the black track must intersect at two points. $l$ is the distance between the two intersections, and $x$ is the distance between the true vertex and the first intersection. $s$ is the distance between the circular centers of the two tracks.	94

5-5 Example plots showing the correlation between $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ and $d_0 \times$ charge. The vertices used for the $d_0$ calculation are the PV (left) and the BS (right). The PV plot shows a modulated dependence from expectation while the BS plot shows a linear shape as expected. Only barrel tracks from 2016 data are shown as examples. Plots of other $ \eta $ regions and other data-taking periods show a similar behavior. Plots credit to Efe Yigitbasi. ....	97
5-6 Plots for the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs $d_0$ correlation in the 2016 DY simulation, and the linear fits to them. Muon tracks are divided into three different $ \eta $ regions: $ \eta  < 0.9$ (left), $0.9 <  \eta  < 1.7$ (middle), and $1.7 <  \eta $ (right). Plots credit to Efe Yigitbasi. ....	97
5-7 Plots for the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs $d_0$ correlation in the 2017 DY simulation, and the linear fits to them. Muon tracks are divided into three different $ \eta $ regions: $ \eta  < 0.9$ (left), $0.9 <  \eta  < 1.7$ (middle), and $1.7 <  \eta $ (right). Plots credit to Efe Yigitbasi. ....	98
5-8 Plots for the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs $d_0$ correlation in the 2018 DY simulation, and the linear fits to them. Muon tracks are divided into three different $ \eta $ regions: $ \eta  < 0.9$ (left), $0.9 <  \eta  < 1.7$ (middle), and $1.7 <  \eta $ (right). Plots credit to Efe Yigitbasi. ....	98
5-9 Plots showing the $p_T$ dependence on the $d_0$ value with different stages of muon correction. The plots compare the $Z \rightarrow \mu\mu$ peak in data and simulation for three years (2016-2018) combined. All positively charged muons are put in the left plot and all negatively charged ones are put in the right plot. The $p_T - d_0$ dependence is reversed for positive and negative muons. ....	99
5-10 Plots showing the <i>GeoFit correction</i> improvement on the four main $H \rightarrow \mu\mu$ signal modes, ggH and VBF plotted on the left, and VH and $t\bar{t}H$ plotted on the right. The plots are made combining the expected signal in all three years of data-taking (2016-2018). The relative improvements on $m_{\mu\mu}$ resolution for ggH, VBF, VH, and $t\bar{t}H$ modes are, respectively, 6.1%, 7.8%, 8.0%, and 9.8%. ....	99
5-11 Plots showing $d_0$ and $p_T$ distributions of the four main $H \rightarrow \mu\mu$ signal modes. The $d_0$ profiles (left plot) of the 4 signals are the same. The $p_T$ profiles (right plot) are different, which explains the different relative improvements that the signals receive from the <i>GeoFit correction</i> . ....	100
5-12 Plots of the $m_{\mu\mu}$ shape of the 2018 ggH simulation sample, comparing different muon correction methods. The left plot shows the $m_{\mu\mu}$ distribution calculated with muon tracks refitted with the additional BS constraint, compared with the particle flow shape (left plot). The <i>Rochester correction</i> is not applied in the left plot for both the red and the blue lines. The right plot shows the $m_{\mu\mu}$ distribution from the refit method, with the <i>Rochester correction</i> applied, compared with the shape from <i>GeoFit correction + Rochester correction</i> (right plot). Plots credit to Pierluigi Bortignon. ....	101

5-13 Muon calibration plots vs $\eta(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. . .	103
5-14 Muon calibration plots vs $\phi(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. . .	104
5-15 Muon calibration plots vs $\phi(\mu^-)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. . .	105
5-16 Muon calibration plots vs $p_T(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. The $p_T$ binning sculpts the shape of the $m_{\mu\mu}$ peak, which leads to a jump at the $p_T = 45$ GeV in the plots.....	106
5-17 Muon calibration plots vs $p_T(\mu\mu)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. . .	107
5-18 Muon calibration plots vs $\eta(\mu\mu)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. . .	108
5-19 Muon calibration plots vs $d_0(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. . .	109
5-20 Muon calibration plots vs $d_0(\mu^-)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. . .	110
6-1 Plots of the performance of the $WH \rightarrow 3\ell$ . On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training sample in red and testing sample in blue. A slight over-training is observed in the region of low signal efficiency, because of the fluctuation in background. As will be shown in Fig. 6-5, the BDT does not sculpt the shape of $m_{\mu\mu}$ .....	119
6-2 Input variables to the $WH \rightarrow 3\ell$ BDT, with signal in blue and background in red. .	120
6-3 Plots of the performance of the $ZH \rightarrow 4\ell$ BDT. On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training in red and testing in blue.....	122
6-4 Input variables to the $ZH \rightarrow 4\ell$ BDT, with signal in blue and background in red. .	122

6-5	For the WH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right). .....	123
6-6	For the ZH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right). .....	124
6-7	The WH BDT output (left) and the ZH BDT output (right) in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ .....	125
6-8	Input variables to the WH BDT in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ .....	126
6-9	Input variables to the ZH BDT in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ . .....	127
6-10	The input variable $M_T(\mu_{SS}, \text{MHT})$ to the WH BDT (left), and the BDT output for signal and background in different $M_T(\mu_{SS}, \text{MHT})$ bins (right). A mild disagreement is seen between the simulation and data in the low bins of $M_T(\mu_{SS}, \text{MHT})$ , while the BDT is not sensitive to the $M_T(\mu_{SS}, \text{MHT})$ values in that region. .....	127
6-11	Scans for the first (left), second (middle) and a potential third (right) BDT cut in the WH channel. The first BDT cut is chosen at 0.3. The second BDT cut is chosen at -0.1. A third BDT cut is not necessary.....	128
6-12	Scans for the BDT cut in the ZH channel. The BDT cut is chosen at -0.1.....	129
6-13	The signal modeling in the $\text{WH} \rightarrow \ell + \mu\mu$ inclusive category. Considered signal modes are WH (top left), qqZH (top middle), ggZH (top right), ttH (bottom left), THQ (bottom middle), and THW (bottom right).....	131
6-14	The signal modeling in the $\text{ZH} \rightarrow \ell\ell + \mu\mu$ inclusive category. Considered signals modes are qqZH (left) and ggZH (right). .....	131
6-15	Schemes on how the bias affects the uncertainty on the signal strength measurement. The left plot is an illustration of the $[-\sigma, \sigma]$ coverage of the biased signal strength measurement. The middle plot shows this coverage becomes less as the bias gets larger. The right plot shows how the bias impacts the best fit signal strength uncertainty $\sigma_{fit}$ . The gray dash lines in the right plot indicates the conventional acceptable range of the bias in this analysis.....	142
6-16	Bias in different BDT-based WH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 0.3] (middle), Cat3 BDT [0.3, 1.0] (right). In the tables, the Power stands for a single "Power" function and the "PowerInt" stands for a single power function plus a constant. .....	144

6-17 Bias in different BDT-based ZH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 1.0] (right). In the tables, the "Power" stands for a single power function and the "PowerInt" stands for a single power function plus a constant.....	144
6-18 Post-fit $m_{\mu\mu}$ distribution of the WH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 0.3] (middle), Cat3 BDT [0.3, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty .....	145
6-19 Post-fit $m_{\mu\mu}$ distribution of the ZH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.....	146
6-20 Post-fit $m_{\mu\mu}$ distribution of the weighted combination of all WH and ZH subcategories. Subcategories are weighted proportionally to $S/(S+B)$ ratio. The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.....	146
6-21 Impacts of the nuisance parameters $\theta$ on the signal strength $\hat{r}$ . Parameter names in black are systematic uncertainties, which are Gaussian-constrained to their pre-fit uncertainties. Parameter names in gray are shape and normalization parameters, which do not have pre-fit constraints. $(\hat{\theta} - \theta_0)/\Delta\theta$ shows the pull of each nuisance parameter, while $\Delta\hat{r}$ is its impact on signal strength. Unconstrained parameters do not have pre-fit assumptions and therefore no pulls. The best-fit values of them are shown in the place of pulls. ....	147
6-22 The 95% CL upper limits on the signal strength in subcategories and combinations. Observed (expected) values are marked in the plot. ....	148
7-1 Expected fraction of signal events per production mode in the different subcategories for $m_H = 125.38$ GeV. The tH contribution is defined as the sum of tHq and tHW processes. ....	150
7-2 Expected $S/(S+B)$ and $S/\sqrt{B}$ in different subcategories, where $S$ and $B$ indicate the number of expect signal and background events, respectively. ....	151

7-3 Signal strength modifiers measured for $m_H = 125.38$ GeV in each production category (black points) are compared to the result of the combined fit (solid red line) and the SM expectation (dashed gray line).....	151
7-4 The local p-values as a function of different $m_H$ hypotheses, for each individual categories as well as the overall combination. The left plot shows the observed p-values, each solid marker indicating a mass point, for which the observed p-values are computed. The right plot shows the expected p-values calculated using the background estimate from the S+B fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$ .....	152
7-5 The observed profile likelihood ratio as a function of $\kappa_\mu$ for $m_H = 125.38$ GeV ...	153
7-6 Left: A duplicate of Figure 1-3, taken from Ref. [2]. Summary of the CMS measurements on the Higgs coupling to fermions and bosons based on data recorded in 2016. Right: The coupling strength measurements updated with this $H \rightarrow \mu\mu$ result. ....	154

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctorate of Philosophy

A SEARCH FOR THE STANDARD MODEL HIGGS BOSON DECAYING INTO TWO MUONS  
IN THE VECTOR BOSON ASSOCIATED PRODUCTION MODE AT THE CMS EXPERIMENT

By

Xunwu Zuo

May 2021

Chair: Darin Acosta

Major: Physics

The Higgs boson was proposed in the 1960s as one of the fundamental particles in the standard model (SM) of particle physics. In 2012, the ATLAS and CMS experiments at the LHC discovered a scalar boson with a mass about 125 GeV that was later confirmed, through dedicated measurements of its properties, to be the Higgs boson of the SM. One of the important properties of the Higgs boson is its coupling to muons, which is an elementary second generation fermion in the SM. Any disagreement between the observed coupling and the SM prediction would indicate new physics phenomena beyond the SM, while an agreement would further validate the SM and help measure parameters in the theory of Higgs physics.

This thesis report on the first evidence for the Higgs boson decay to a pair of muons, based on proton-proton collision data at  $\sqrt{s} = 13$  TeV recorded by the CMS experiment at the LHC, corresponding to an integrated luminosity of  $137 \text{ fb}^{-1}$ . The analysis is performed in four exclusive categories targeting different production modes of the Higgs boson: via gluon fusion, via vector boson fusion, in association with a vector boson, and in association with a top quark pair. This thesis is focused on the exclusive category of the Higgs boson production in association with a vector boson, in which an excess of events over the background expectation is observed in data with a significance of 2.0 standard deviations, while the expected significance for the SM Higgs boson with mass of 125.38 GeV is 0.42 standard deviations.

The combination of the four exclusive categories yields an overall observed significance of 3.0 standard deviations, while the expectation is 2.5. This result is also combined with the that from the data recorded at  $\sqrt{s} = 7$  and 8 TeV, corresponding to integrated luminosities of 5.1 and  $19.7 \text{ fb}^{-1}$ , respectively, whose effect is an increase in both the expected and observed significance by 1%. The measured signal strength, relative to the SM prediction, is  $1.19^{+0.40}_{-0.39}(\text{stat})^{+0.15}_{-0.14}(\text{syst})$ . This result establishes the first evidence for the Higgs boson decay to second generation fermions, with a branching fraction consistent with the SM prediction, and provides the most precise measurement of the Higgs boson coupling to muons to date.

## CHAPTER 1

### INTRODUCTION

The success of modern particle physics has been a great success in describing the basic constituents of matter and the fundamental interactions between them. The Standard Model (SM) of particle physics summarizes the behavior of all observed elementary particles, shown in Figure 1-1. There are 12 types of elementary fermions, 4 types of elementary gauge bosons, and 1 Higgs boson. Elementary fermions are matter particles and are further categorized into quarks and leptons. Both quarks and leptons have 3 generations. In each generation, there is an electric charge =  $+\frac{2}{3}e$  quark (u, c, and t quarks), a charge =  $-\frac{1}{3}e$  quark (d, s, and b quarks), a charge =  $-1e$  lepton ( $e$ ,  $\mu$ , and  $\tau$  leptons), and a neutral lepton ( $\nu_e$ ,  $\nu_\mu$ , and  $\nu_\tau$  neutrinos). Gauge bosons have spin = 1 and are carriers of fundamental interactions: gluons for the strong interaction, photons for the electromagnetic interaction, and W and Z bosons for the weak interaction. Gravity is another fundamental interaction whose quantization has not been observed experimentally, and is not included in the SM. The Higgs boson is the only scalar boson (spin = 0) in the SM. It plays a special role in the SM as it interacts with all massive elementary particles and introduces the mechanism for them to have nonzero masses. Each of the particles has a counterpart called an antiparticle, which has the same mass and spin as the particle itself, but opposite charge and other quantum numbers. The neutral bosons (gluons, photons, Z, and Higgs bosons) are their own antiparticles.

The theory of the SM is formulated in quantum field theory and exhibits rich physics phenomena. The mathematical construction of the SM is given in the following sections: Section 1.1 describes the basic formalism of quantum field theory and the idea of gauge symmetry; Section 1.2 shows the mathematical description of the SM without the Higgs field and its drawback in describing particle masses; Section 1.3 introduces the Higgs field and its interplay with the gauge fields and fermion fields to allow particles to retain masses; and Section 1.4 discusses the free parameters in the SM and the motivation to study the Higgs to muons decay.

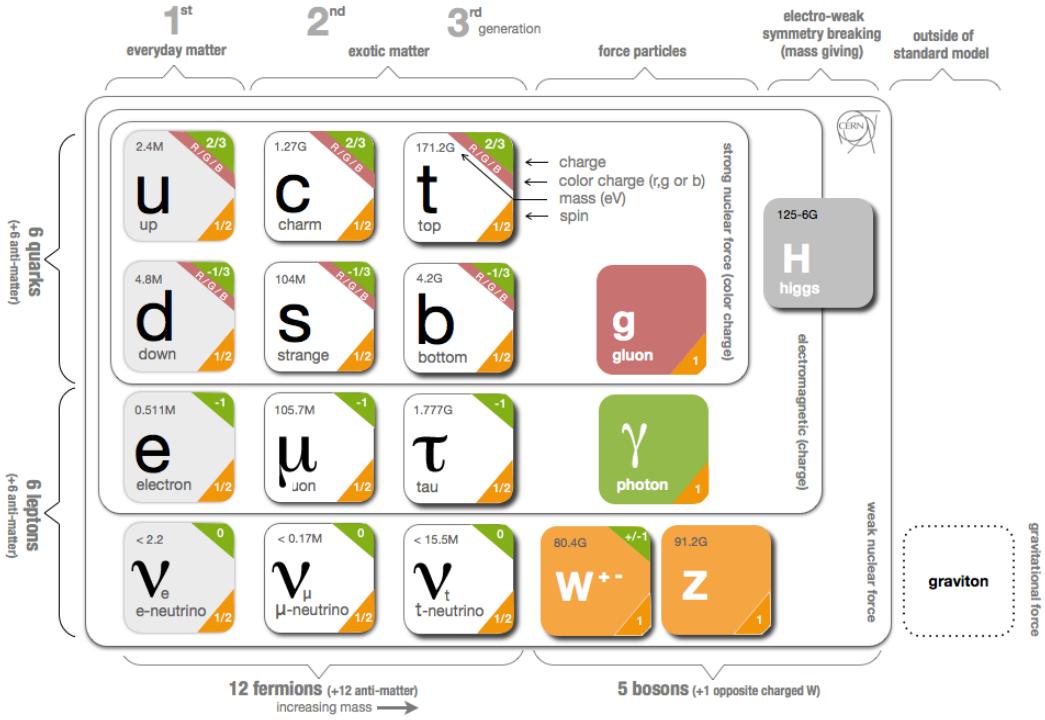


Figure 1-1. The Standard Model of particles physics: 12 elementary fermions and 5 elementary bosons. Plot taken from [1].

## 1.1 Quantum field theory and gauge symmetry

Symmetry is a key concept in physics. A symmetry regarding a certain quantity means that the equations of motion do not depend on this quantity and implies the existence of a conservation law. For example the symmetries under translations in space and time correspond to the conservations of momenta and energy.

In quantum field theory, the Lagrangian for a free Dirac field is

$$\mathcal{L} = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) \quad (1-1)$$

where  $\psi(x)$  is the Dirac field,  $\partial_\mu$  is the regular partial derivative with Lorentz vector index

$\mu$ , and  $m$  is the mass of the Dirac spinor.  $\gamma_\mu$  denotes the Dirac matrices

$$\gamma^0 = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \quad \gamma^k = \begin{pmatrix} 0 & \sigma^k \\ -\sigma^k & 0 \end{pmatrix}, \quad \text{with } k = 1, 2, 3 \quad (1-2)$$

where  $I_2$  is the  $2 \times 2$  identity matrix and  $\sigma^k$  denotes the Pauli matrices.

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1-3)$$

This Lagrangian is invariant under an internal phase translation

$$\psi(x) \rightarrow e^{i\theta_0} \psi(x) \quad (1-4)$$

where  $\theta_0$  is an arbitrary constant. This is a global symmetry as  $\theta_0$  is independent from the spacetime location. A crucial practice in quantum field theory is to promote global symmetries to local symmetries, also known as gauge symmetries, under transformations like

$$\psi(x) \rightarrow e^{i\theta(x)} \psi(x) \quad (1-5)$$

where  $\theta(x)$  is no longer constant but a function of the spacetime coordinates  $x$ . Obviously Equation 1-1 is not invariant under this gauge transformation, as the transformation give rise to a new term  $\partial_\mu \theta(x)$ . The way to achieve gauge invariance is to replace the partial derivative  $\partial_\mu$  by the covariant derivative, through whose construction new interactions are introduced.

The covariant derivative of transformation Equation 1-5 is written as

$$D_\mu = \partial_\mu + ieA_\mu \quad (1-6)$$

where  $e$  is an arbitrary constant and  $A_\mu$  is a new field whose behavior under the same transformation is

$$A_\mu(x) \rightarrow A_\mu(x) - \frac{1}{e}\partial_\mu \theta(x) \quad (1-7)$$

In fact, this exact formula can be used to describe electromagnetism, for which  $A_\mu$  is the vector potential of the electromagnetic field and  $e$  is the electric charge.

Modifying Equation 1-1 with Equation 1-6 and adding the EM field term, we get a gauge invariant Lagrangian describing a Dirac field interacting with an EM field

$$\mathcal{L} = \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1-8)$$

where  $F_{\mu\nu}$  is the EM field tensor

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) \quad (1-9)$$

The transformation Equation 1-5 takes the form of the group U(1) at each point  $x$ . Therefore we say the electromagnetic interaction follows a U(1) gauge symmetry.

## 1.2 The standard model Lagrangian

Following the formalism in Section 1.1 and knowing there are 6 quarks, 6 leptons, and 4 gauge bosons, the SM Lagrangian without considering a Higgs boson and without mass terms can be written as:

$$\mathcal{L} = \bar{\psi}^i(i\gamma^\mu)(D_\mu)_{ij}\psi^j - \frac{1}{4}G_{\mu\nu}^kG^{k\mu\nu} - \frac{1}{4}W_{\mu\nu}^aW^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \quad (1-10)$$

In the equation:

- $\psi$  is the fermion field and  $D_\mu$  is the covariant derivative;
- $G_{\mu\nu}^k$ ,  $W_{\mu\nu}^a$ , and  $B_{\mu\nu}$  are field strength tensors for the interaction fields corresponding to the groups SU(3), an SU(2), and an U(1), respectively;
- $\mu$  and  $\nu$  are the Lorentz vector indices. Index  $k$  and  $a$  are the index of the group generators,  $k$  running from 1 to 8, and  $a$  running from 1 to 3;

This Lagrangian takes the structure of groups  $SU(3)_c \times SU(2)_L \times U(1)_Y$ . The  $SU(3)_c$  represents the strong interaction with the subscript  $c$  standing for color. The  $SU(2)_L \times U(1)_Y$  is the electroweak interaction, where the  $L$  means it only acts on left-handed fermions and the  $Y$  is the hypercharge associated with the U(1) group.<sup>1</sup> Here the  $SU(2)_L \times U(1)_Y$

---

<sup>1</sup>The hypercharge  $Y$  is not the electric charge  $e$  in Equation 1-6. In fact, for all fermions, the electric charge  $Q = T_3 + Y/2$ , where  $T_3$  is an SU(2) group generator.

cannot be disentangled and viewed separately as the weak interaction and electromagnetic interaction.

The gauge transformations on the fermion field are:

$$\begin{aligned} \text{U(1)} : \quad \psi(x) &\rightarrow \exp[i\theta_Y(x)Y]\psi(x) \\ \text{SU(2)} : \quad \psi(x) &\rightarrow \exp[i\theta_L^a(x)T^a]\psi(x) \\ \text{SU(3)} : \quad \psi(x) &\rightarrow \exp[i\theta_c^k(x)t^k]\psi(x) \end{aligned} \quad (1-11)$$

with  $D_\mu$  taking the form of

$$D_\mu = \partial_\mu - ig'B_\mu Y - igW_\mu^a T^a - ig_s G_\mu^k t^k \quad (1-12)$$

where  $g'$ ,  $g$ , and  $g_s$  are field strength coefficients, and  $Y$ ,  $T^a$ , and  $t^k$  are interaction operators, for U(1), SU(2), and SU(3) respectively. In this construction the transformations on the interaction fields are:

$$\begin{aligned} \text{U(1)} : \quad B_\mu(x) &\rightarrow B_\mu(x) + \frac{1}{g'}\partial_\mu\theta_Y(x) \\ \text{SU(2)} : \quad W_\mu^a(x) &\rightarrow W_\mu^a(x) + \frac{1}{g}\partial_\mu\theta_L^a(x) + \epsilon^{abc}W_\mu^b(x)\theta_L^c(x) \\ \text{SU(3)} : \quad G_\mu^k(x) &\rightarrow G_\mu^k(x) + \frac{1}{g_s}\partial_\mu\theta_c^k(x) + f^{klm}G_\mu^l(x)\theta_c^m(x) \end{aligned} \quad (1-13)$$

Furthermore, the field tensors in Equation 1-10 are related to the field vectors as

$$\begin{aligned} \text{U(1)} : \quad B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ \text{SU(2)} : \quad W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon^{abc}W_\mu^b W_\nu^c \\ \text{SU(3)} : \quad G_{\mu\nu}^k &= \partial_\mu G_\nu^k - \partial_\nu G_\mu^k + g_s f^{klm}G_\mu^l G_\nu^m \end{aligned} \quad (1-14)$$

in which the  $\epsilon^{abc}$  and  $f^{klm}$  are structure constants of SU(2) and SU(3) governing the relations between group generators

$$\begin{aligned} [T^a, T^b] &= i\epsilon^{abc}T_c \\ [t^k, t^l] &= if^{klm}t_m \end{aligned} \quad (1-15)$$

Equation 1-10 models the strong and electroweak interactions with matter fields.

However, all particles in this model are massless as there are no mass terms like

$$\mathcal{L}_{\text{mass}} = -m_f \bar{\psi}_f^i \psi_{fi} + \frac{1}{2} m_G^2 G_\mu G^\mu + \frac{1}{2} m_W^2 W_\mu W^\mu + \frac{1}{2} m_B^2 B_\mu B^\mu \quad (1-16)$$

None of these terms is gauge invariant. The gauge boson terms,  $m_G^2 G_\mu G^\mu$ ,  $m_W^2 W_\mu W^\mu$ , and  $m_B^2 B_\mu B^\mu$ , are not gauge invariant because their transformations always introduce new terms depending on the transformations as described in Equation 1-13. The reason why the fermion terms,  $m_f \bar{\psi}_f^i \psi_{fi}$ , are not gauge invariant is not so obvious in this form but becomes clear when written in left and right chiral forms:  $\psi = \psi_R + \psi_L$ , in which

$$\psi_R = \begin{pmatrix} \psi_+ \\ 0 \end{pmatrix}, \text{ and } \psi_L = \begin{pmatrix} 0 \\ \psi_- \end{pmatrix} \quad (1-17)$$

We also introduce the projection operators:

$$P_\pm = \frac{1}{2}(1 \pm \gamma^5), \text{ with } \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 \quad (1-18)$$

who act on  $\psi$  as

$$\begin{aligned} P_+ \psi &= \psi_R, & P_- \psi &= \psi_L, \\ \bar{\psi} P_+ &= \bar{\psi}_L, & \bar{\psi} P_- &= \bar{\psi}_R \end{aligned} \quad (1-19)$$

Expanding the term  $m\bar{\psi}\psi = m(\bar{\psi}_R + \bar{\psi}_L)(\psi_R + \psi_L)$ , the terms  $\bar{\psi}_R\psi_R$  and  $\bar{\psi}_L\psi_L$  vanish.

$$\bar{\psi}_R\psi_R = \bar{\psi} P_- P_+ \psi = \bar{\psi} \frac{1}{4}(1 - (\gamma^5)^2)\psi = 0 \quad (1-20)$$

The surviving terms are  $m\bar{\psi}\psi = m\bar{\psi}_R\psi_L + m\bar{\psi}_L\psi_R$ . The  $SU(2)_L$  transformation only acts on the left chiral component, so its phase only appear in  $\psi_L$  and is not canceled by  $\psi_R$ . The  $U(1)_Y$  hypercharge is also different for the left and right components, leaving the  $U(1)$  phase uncanceled. Therefore the fermion mass terms are not gauge invariant. This is different from the part  $\bar{\psi}(i\gamma^\mu)(D_\mu)\psi$ , whose surviving terms are  $\bar{\psi}_R(i\gamma^\mu)(D_\mu)\psi_R$  and  $\bar{\psi}_L(i\gamma^\mu)(D_\mu)\psi_L$ , in which the  $SU(2)_L$  "phase" term cancels properly, making them gauge invariant.

Therefore, Equation 1-16 cannot be attached to Equation 1-10 and all particles must be massless, which contradicts the experimental findings. The follow section describes a mechanism to allow particles to retain masses through the introduction of a new field component called the Higgs field.

### 1.3 The Higgs mechanism

The Higgs field is a complex scalar field whose potential is:

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \quad (1-21)$$

in which  $\Phi$  is the complex scalar field and  $\mu$ ,  $\lambda$  are coefficients. The shape of the this potential is shown in Figure 1-2. Parameter  $\lambda$  needs to be positive for  $V(\Phi)$  to have a minimum (ground state). If  $\mu^2 \leq 0$ , there is a single minimum of  $V(\Phi)$  at zero when  $|\Phi| = 0$ . If  $\mu^2 > 0$ ,  $V(\Phi)$  can have a degenerate minimum of  $-\frac{\mu^4}{4\lambda}$  when  $\Phi^\dagger \Phi = \frac{\mu^2}{2\lambda}$ .

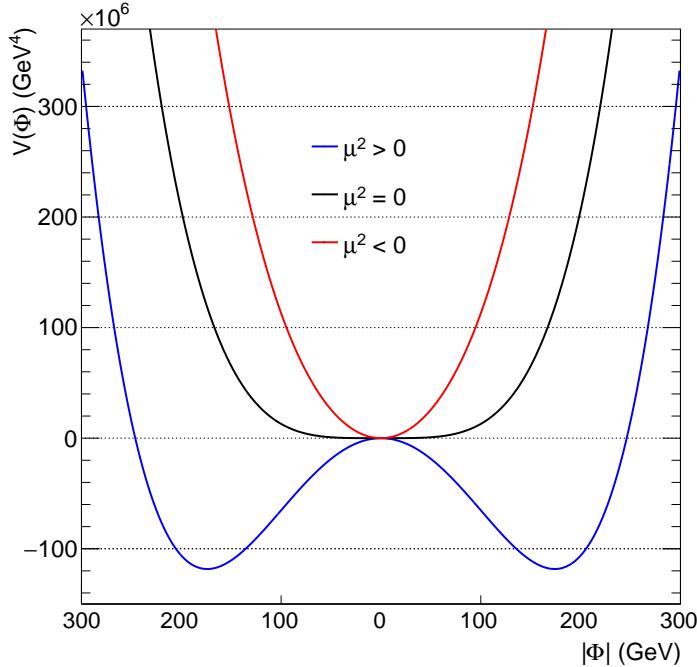


Figure 1-2. The shape of the Higgs potential with different  $\mu^2$  parameter configurations.

The equation  $\Phi^\dagger \Phi = \frac{\mu^2}{2\lambda}$  represents a circle in the complex plane of  $\Phi$ . As  $\mu^2$  increase from below 0 to above 0, the ground state of the field moves from  $\Phi = 0$  in an arbitrary

direction to a point on the circle  $\Phi^\dagger\Phi = \frac{\mu^2}{2\lambda}$ , making an asymmetric solution. This phenomenon is known as the spontaneous symmetry breaking, which is also widely observed in physics systems such as the bending of a cylindrical rod under axial pressure, or the cooling of a ferromagnetic material across its Curie temperature.

Now we look at the SM Lagrangian. Putting the massless form of Equation 1-10 together with the Higgs field, it becomes:

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu} \\ & + |D_\mu\Phi|^2 + \mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2 \\ & + \sum_f [\bar{\psi}_R^f(i\gamma^\mu)(D_\mu)\psi_R^f + \bar{\psi}_L^f(i\gamma^\mu)(D_\mu)\psi_L^f] \\ & - \sum_f g_f [\bar{\psi}_R^f\Phi^\dagger\psi_L^f + \bar{\psi}_L^f\Phi\psi_R^f + \bar{\psi}_R^f\tilde{\Phi}^\dagger\psi_L^f + \bar{\psi}_L^f\tilde{\Phi}\psi_R^f] \end{aligned} \quad (1-22)$$

where the first and second lines are the gauge field and Higgs field, the third line is the fermion kinetics written in the chiral form, and the fourth line is the Yukawa coupling between fermions and the Higgs, also in the chiral form. The index  $f$  in the third and fourth lines just sums all fermions, with  $g_f$  as the coupling strength for each fermion type. The  $\psi_L$  are doublets of the SU(2)

$$\psi_L = \frac{1}{2}(1+\gamma_5) \begin{pmatrix} v^i \\ l^i \end{pmatrix}, \quad \text{or} \quad \frac{1}{2}(1-\gamma_5) \begin{pmatrix} U^i \\ D^i \end{pmatrix} \quad (1-23)$$

and  $\psi_R$  are singlets

$$\psi_R = \frac{1}{2}(1-\gamma_5)l^i, \quad \frac{1}{2}(1-\gamma_5)U^i, \quad \text{or} \quad \frac{1}{2}(1-\gamma_5)D^i \quad (1-24)$$

where  $v^i, l^i, U^i, D^i$  are neutrinos, charged leptons, up-type quarks and down-type quarks of three generations. The existence of right-handed neutrinos is not confirmed by experiments and therefore not assumed in the (simplest) SM. The Higgs field is a doublet with 4

total degrees of freedom

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix}, \quad \tilde{\Phi} = \begin{pmatrix} \Phi^{0\dagger} \\ -\Phi^- \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^3 - i\phi^4 \\ -\phi^1 + i\phi^2 \end{pmatrix} \quad (1-25)$$

By choosing the direction of symmetry breaking to be the real part of  $\Phi^0$ , we can make  $\phi^1 = \phi^2 = \phi^4 = 0$  at the ground state, and rewrite the Higgs doublet as

$$\Phi \rightarrow \frac{1}{\sqrt{2}} \exp[i\theta_L^a(x)T^a] \exp[i\theta_Y(x)Y] \begin{pmatrix} 0 \\ v + \rho \end{pmatrix} \quad (1-26)$$

with  $v = \sqrt{\frac{\mu^2}{\lambda}}$  known as the vacuum expectation value, and  $\rho$  as the deviation from the ground state position.

**The mass of the Higgs boson** comes from the potential term  $\mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2$ , which expands to

$$\mu^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2 = \lambda v^2\rho^2 - \lambda v\rho^3 - \frac{1}{4}\lambda\rho^4 + \frac{1}{4}\lambda v^4 \quad (1-27)$$

The perturbation  $\rho$  acts as the apparent Higgs field. The term  $\lambda v^2\rho^2$  describes the Higgs boson mass

$$m_H = \sqrt{2\lambda}v \quad (1-28)$$

Furthermore, terms  $\rho^3$  and  $\rho^4$  describes important Higgs properties of the triple and quartic self-couplings.

**The masses of gauge bosons** come from the  $|D_\mu\Phi|^2$  term, which expands to

$$\begin{aligned} |D_\mu\Phi|^2 &= \frac{1}{2}|\partial_\mu\rho|^2 + \frac{1}{8}(v+\rho)^2[g^2(W_\mu^1W^{1\mu} + W_\mu^2W^{2\mu}) + (g'B_\mu - gW_\mu^3)^2] \\ &= \frac{1}{8}v^2[2g^2W_\mu^+W^{-\mu} + (g^2 + g'^2)Z_\mu Z^\mu] \\ &\quad + \frac{1}{4}v\rho[2g^2W_\mu^+W^{-\mu} + (g^2 + g'^2)Z_\mu Z^\mu] \\ &\quad + \frac{1}{8}\rho^2[2g^2W_\mu^+W^{-\mu} + (g^2 + g'^2)Z_\mu Z^\mu] + \frac{1}{2}|\partial_\mu\rho|^2 \end{aligned} \quad (1-29)$$

In the second step the mass eigenstates of gauge bosons are defined as

$$\begin{aligned} W_\mu^\pm &= \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}} \\ Z_\mu &= -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3 \\ A_\mu &= \cos\theta_W B_\mu + \sin\theta_W W_\mu^3 \end{aligned} \tag{1-30}$$

with the Weinberg angle  $\tan\theta_W = g'/g$ . The first part in Equation 1-29 becomes the mass terms for the gauge bosons with

$$\begin{aligned} m_W &= \frac{v g}{2} \\ m_Z &= \frac{v \sqrt{g^2 + g'^2}}{2} \\ m_A &= 0 \end{aligned} \tag{1-31}$$

In addition, the second part of Equation 1-29 describes the 3-point couplings between a Higgs boson and a pair of gauge bosons, while the third part describes the 4-point couplings between a Higgs boson pair and a gauge boson pair.

**The masses of fermions** come from the Yukawa coupling term in Equation 1-22. The term  $\bar{\psi}_R^f \Phi^\dagger \psi_L^f + \bar{\psi}_L^f \Phi \psi_R^f$  gives the masses for charged leptons and down-type quarks.

$$\begin{aligned} &-\sum_f g_f [\bar{\psi}_R^f \Phi^\dagger \psi_L^f + \bar{\psi}_L^f \Phi \psi_R^f] \\ &= -\sum_{l^i} g_{l^i} \frac{1}{2\sqrt{2}} \bar{l}^i \begin{pmatrix} 0, & v+\rho \end{pmatrix} \begin{pmatrix} v^i \\ l^i \end{pmatrix} - \sum_{l^i} g_{l^i} \frac{1}{2\sqrt{2}} \begin{pmatrix} \bar{v}^i, & \bar{l}^i \end{pmatrix} \begin{pmatrix} 0 \\ v+\rho \end{pmatrix} l^i \\ &\quad - \sum_{D^i} g_{D^i} \frac{1}{2\sqrt{2}} \bar{D}^i \begin{pmatrix} 0, & v+\rho \end{pmatrix} \begin{pmatrix} U^i \\ D^i \end{pmatrix} - \sum_{D^i} g_{D^i} \frac{1}{2\sqrt{2}} \begin{pmatrix} \bar{U}^i, & \bar{D}^i \end{pmatrix} \begin{pmatrix} 0 \\ v+\rho \end{pmatrix} D^i \\ &= -\frac{1}{\sqrt{2}} \sum_{l^i} g_{l^i} (v+\rho) \bar{l}^i l^i - \frac{1}{\sqrt{2}} \sum_{D^i} g_{D^i} (v+\rho) \bar{D}^i D^i \end{aligned} \tag{1-32}$$

In this equation some calculations are omitted to save space: the SU(2) phase cancels between  $\Phi^\dagger$  and  $\psi_L$ , the SU(1) phase cancels between  $\bar{\psi}_R$ ,  $\Phi^\dagger$  and  $\psi_L$  as their hypercharges add up to 0, and the product of projection operators in  $\bar{\psi}_R$  and  $\psi_L$  is 2 and is absorbed in

the coefficients. Similar to Equation 1-32, the term  $\bar{\psi}_R^f \tilde{\Phi}^\dagger \psi_L^f + \bar{\psi}_L^f \tilde{\Phi} \psi_R^f$  gives the masses for up-type quarks.

$$-\sum_f g_f [\bar{\psi}_R^f \tilde{\Phi}^\dagger \psi_L^f + \bar{\psi}_L^f \tilde{\Phi} \psi_R^f] = -\frac{1}{\sqrt{2}} \sum_i g_{U^i} (\nu + \rho) \bar{U}^i U^i \quad (1-33)$$

There is no term here for neutrinos as they do not have right-handed components. Overall, the Yukawa term is summarized as

$$\mathcal{L}_{\text{Yukawa}} = -\sum_f \frac{1}{\sqrt{2}} g_f (\nu \bar{\psi}^f \psi^f + \rho \bar{\psi}^f \psi^f) \quad (1-34)$$

where the first part is the mass term for fermions, and the second part is the interactions between the apparent Higgs field  $\rho$  and the fermions. The fermion mass  $m_f$  is proportional to its Yukawa coupling strength  $g_f$

$$m_f = \frac{\nu g_f}{\sqrt{2}} \quad (1-35)$$

#### 1.4 Parameters of the standard model

In the Lagrangian we derived in the last Section, there are 14 unconstrained real parameters (assuming massless neutrinos):

- 3 gauge coupling constants  $g$ ,  $g'$ , and  $g_s$ ;
- 2 Higgs potential shape parameters  $\lambda$  and  $\mu^2$ ;
- 9 Yukawa coupling constants  $g_f$  for 3 charged leptons and 6 quarks.

There are 5 other parameters not elaborated here

- 4 parameters in the quark mixing matrix (CKM matrix), which is not elaborated here, including 3 angle parameters and 1 phase parameter;
- A parameter for CP violation in the strong interaction.

These total 19 parameters determine the exact properties of SM particles and the measurements of them are a key topic of high energy experiments. In practice, some other variable sets are frequently used. For example, the parameter set  $[g, g', \lambda, \mu^2, g_f]$  is interchangeable

with the set consisting of the vacuum expectation, and the masses of Higgs, W, Z, and fermions  $[v, m_H, m_W, m_Z, m_f]$ .

The Higgs boson was first observed at the LHC by ATLAS and CMS Collaborations in 2012 [20, 21, 22]. Various measurements on the Higgs properties have been conducted to verify their agreement with the SM prediction and to put constraints on the free SM parameters. The latest and most precise measurement of the Higgs boson mass is  $125.38 \pm 0.14$  GeV, provided by the CMS Collaboration [23]. The Higgs boson decays to gauge bosons and third-generation fermions (except the top quark) have been observed and measured [24, 25, 26, 27, 28, 29, 30, 31, 32]. The Higgs coupling to the top quark has also been measured through the Higgs boson production process in association with a top quark pair [33, 34]. The Higgs coupling measurements from CMS with the LHC proton-proton collision data collected in 2016 [2] is summarized in Figure 1-3, in which, the gauge couplings and Yukawa couplings are expressed by the coupling modifiers,  $\kappa_V$  and  $\kappa_F$ , in the  $\kappa$ -framework [35]. As the Higgs couplings to fermions are proportional to the fermion masses and the Higgs couplings to weak bosons are proportional to the square of the boson masses, the results from  $\kappa$ -framework are expressed in the plot as  $\kappa_F m_F / v$  for fermions, and  $\sqrt{\kappa_V} m_V / v$  for bosons.

Ref. [2] considers 6 free coupling parameters:  $\kappa_W$ ,  $\kappa_Z$ ,  $\kappa_t$ ,  $\kappa_\tau$ ,  $\kappa_b$ , and  $\kappa_\mu$ . The Higgs couplings to other fermions are hard to probe at the LHC, and their coupling modifiers are assumed to be the same as their higher generation counterparts,  $\kappa_u = \kappa_c = \kappa_t$ ,  $\kappa_d = \kappa_s = \kappa_b$ , and  $\kappa_e = \kappa_\mu$ . The study of Higgs to muons decay is of particular importance in this context, as it is the most experimentally sensitive way to probe Higgs couplings to fermions out of the third generation and the only direct measurement of such couplings at the LHC. An analysis on  $H \rightarrow \mu\mu$  decay has been recently conducted based on proton-proton collision data collected by CMS in years 2016-2018 [36], superseding the one considered in Ref. [2].

This thesis presents this recent analysis of the Higgs to muons decay from CMS with a focus on the work in the vector boson associated Higgs boson production channel (VH).

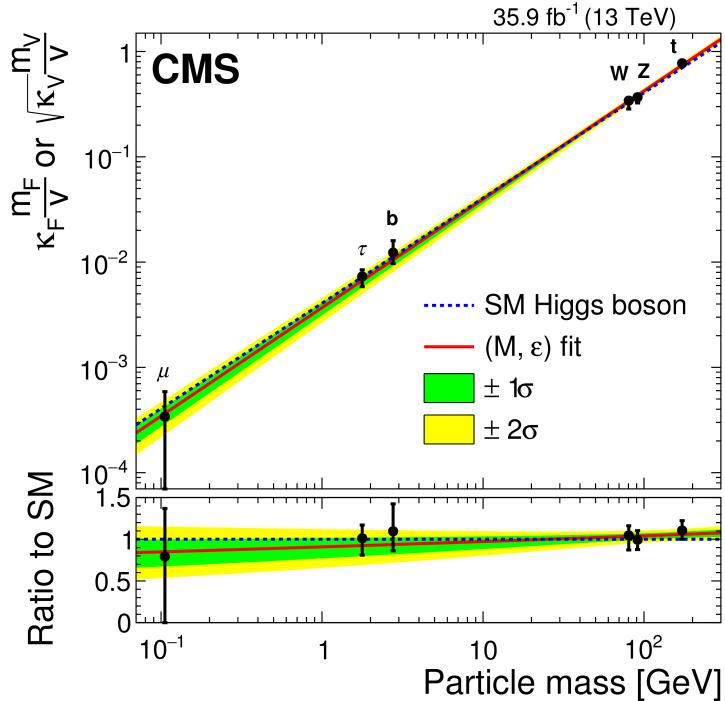


Figure 1-3. Summary of the CMS measurements on the Higgs coupling to fermions and bosons. Plot taken from [2].

The following chapters are organized as follows: Chapter 2 describes the LHC machine and the CMS detector, Chapter 3 gives an overview of the  $H \rightarrow \mu\mu$  analysis, which is conducted in 4 exclusive event categories targeting different Higgs boson production modes, Chapter 4 lists the reconstruction of physics objects and the selections on them in the analysis, Chapter 5 details the correction methods and calibration studies on the muon momentum, which is crucial to the accuracy of this work, Chapter 6 reports the full procedures of the analysis in the VH channel, and Chapter 7 summarizes the results from the VH analysis as well as the combined results of all 4 analysis channels.

## CHAPTER 2

### THE LHC AND CMS

All elementary particles in the Standard Model has been observed. The main focus of particle physics nowadays are to precisely measure SM parameters and to search for physics beyond the Standard Model (BSM). A good way to pursue these goal is to perform high energy collision experiments. The Large Hadron Collider is the most powerful hadron collider that humans have built for this purpose, and the Compact Muon Solenoid (CMS) experiment is one of the experiments at the LHC that study the outcome of the collisions. An overview of the LHC is given in Section 2.1, and the CMS detector is described in Section 2.2.

#### 2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [3] at CERN near Geneva, Switzerland is the world's largest and most powerful machine for particle physics research. It is a double-ring superconducting hadron accelerator and collider installed in a 26.7 km long circular tunnel inherited from its predecessor, the Large Electron-Positron Collider (LEP). The LHC tunnel lies in the rock stratum between 45 m and 170 m underground, and spans across the French-Swiss border from the bank of the Geneva Lake to the base of the Jura mountains. Figure 2-1 shows the geographical location of the LHC. Two series of hadron bunches rotate in opposite directions in the main LHC ring and collide at four interaction points, each hosting a major LHC experiments: Point 1 for ATLAS, Point 2 for ALICE, Point 5 for CMS, and Point 8 for LHCb.

##### 2.1.1 The LHC machine

The LHC tunnel consists of eight arcs and eight straight sections. The arcs make the majority of the LHC circumference, accommodating thousands of the magnet units to bend and tighten the particles' trajectory. The straight sections are approximately 528 m long each, serving as insertions for experiments or utility. The arcs and straight sections are grouped into eight octants, each covering a straight section and two halves of its neighboring arcs, whose geometrical layout is shown in Figure 2-2.

Strong magnets are what guide the high energy hadrons to circulate and collide in

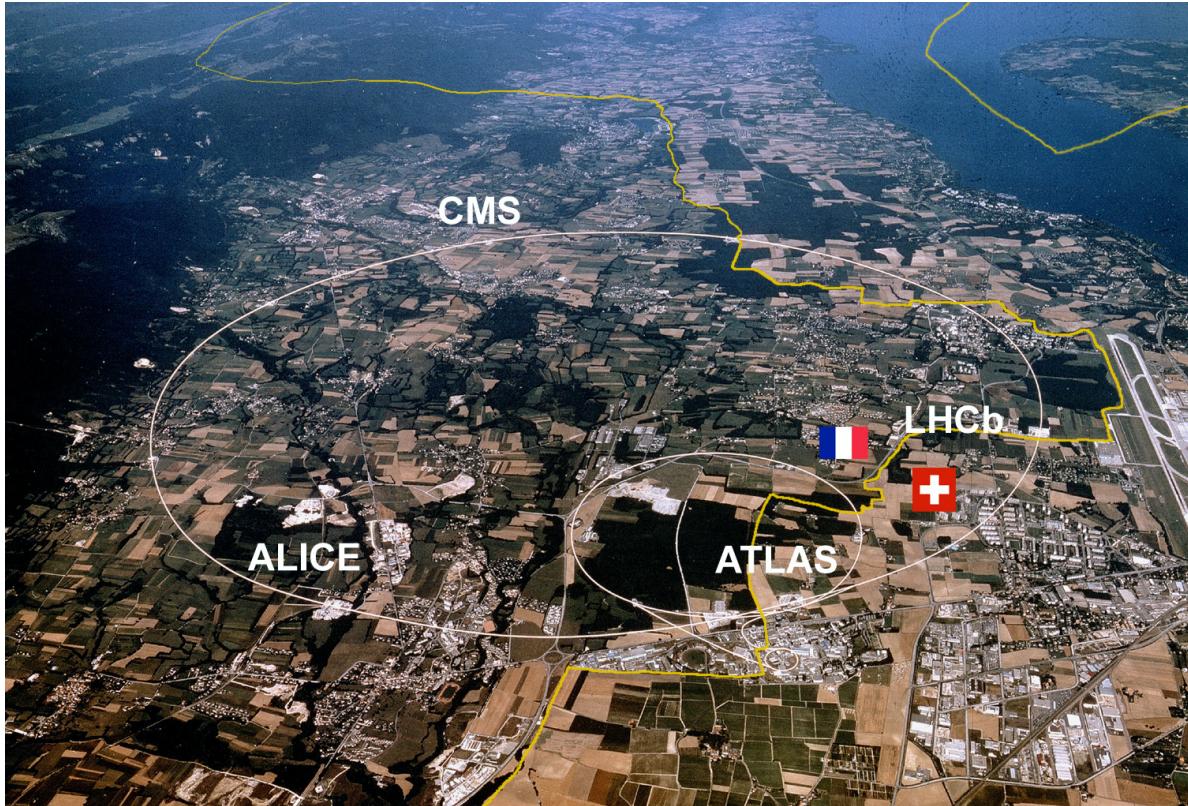


Figure 2-1. Photo taken from ... (original source unknown). An aerial view of the LHC landscape. The French-Swiss border is indicated by the yellow line and the LHC tunnels are outlined in white. The triangular building complex just below ATLAS in the picture is the main campus of CERN.

the LHC. The LHC magnet system is based on Nb-Ti cables, cooled by superfluid helium to a temperature of 1.9 K, where the cables stay superconductive and generate a magnetic field up to their critical field strength. The maximum operable magnetic field of the LHC magnets is 8.33 T, which corresponds to a proton beam energy of 7 TeV, or a heavy ion beam energy of 2.76 TeV per nucleon.

The LHC magnet system consists of 1232 main dipole magnets, about 450 quadrupole magnets, a few thousands multipole corrector magnets, and several types of specialized magnets at the eight insertion points. The dipoles bend the beams so they circulate in the LHC tunnel. Since both beams are positively charged but travel in opposite directions, the magnetic fields for the two beams need to be opposite as well. Given the space limitation in the tunnel and the need to keep the budget down, a "twin-bore" design is adopted, in

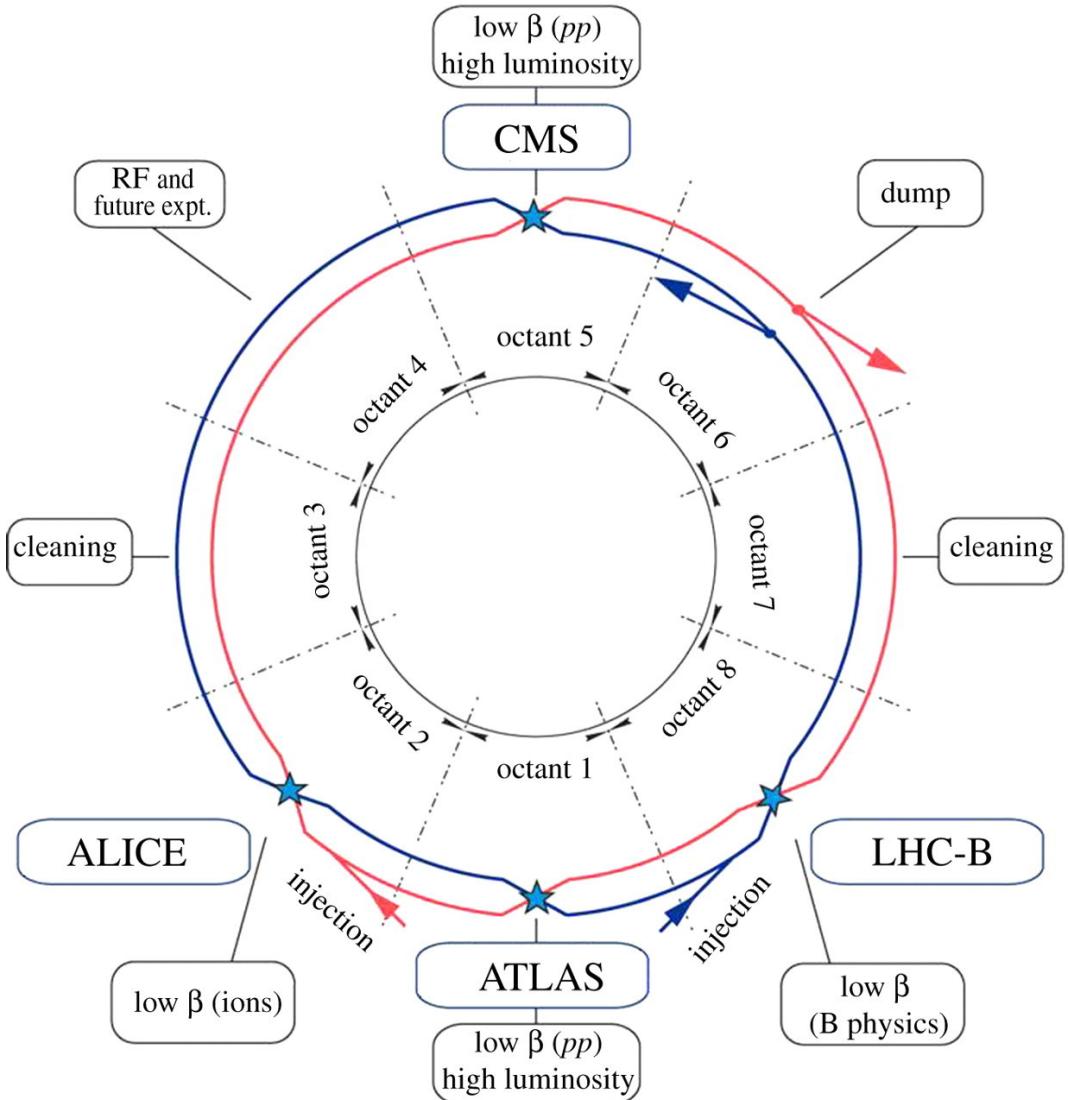


Figure 2-2. The schematic layout of the LHC. Each octant contains a insertion point. Points 1, 2, 5, and 8 are the locations of collision experiments. Points 3 and 7 are for beam collimation. Hadrons are injected at Points 1 and 8, accelerated at Point 4, and eventually discarded at Point 6. Plot taken from Ref. [3].

which the two beam pipes and two sets of magnet coils are installed next to each other in the same piece of mechanical housing, called the cold mass. The cold mass is a cylindrical solid iron structure bored at its center and surrounded by a superfluid helium vessel. Each cold mass has a length of about 15 m, a diameter of about 570 mm, and a mass of about 27.5 t. It provides a stable 1.9 K environment for the magnet coils, and in the meantime serves as their magnetic yoke. The dipoles of LHC are manufactured identically

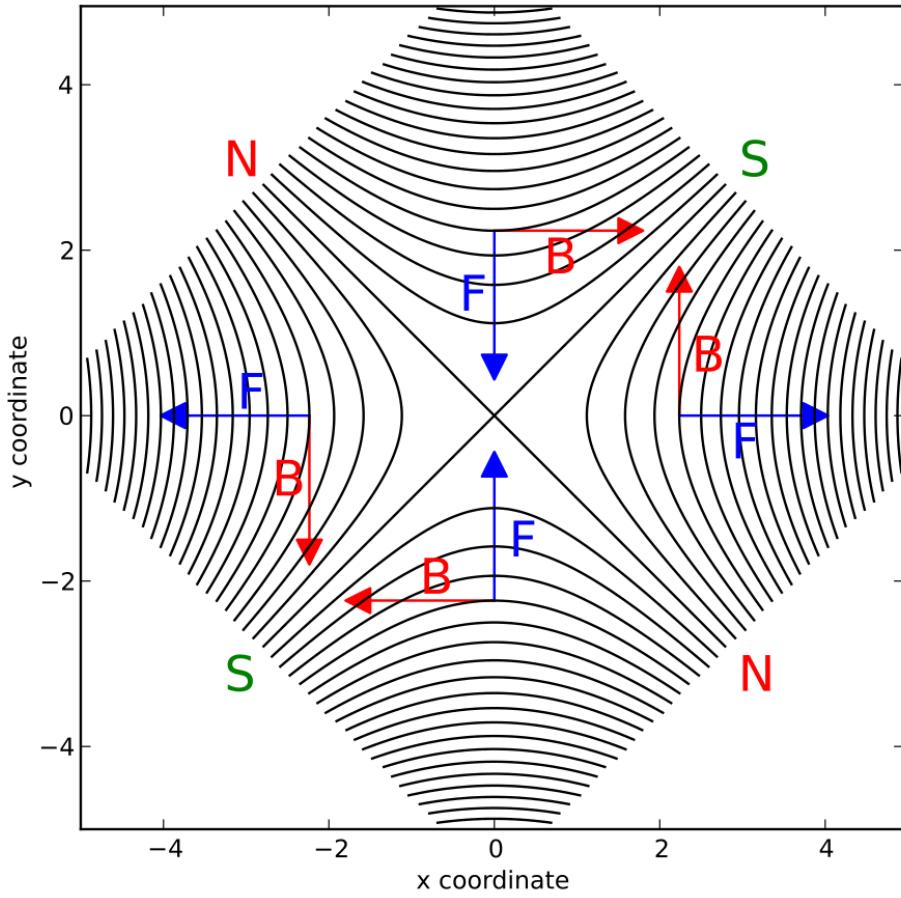


Figure 2-3. Magnetic field of an ideal quadrupole. Plot taken from Ref. [4].

up to a high precision. The relative variation in the magnetic field strength and the field inhomogeneity must not exceed  $10^{-4}$ . The quadrupoles provide gradient fields (shown in Figure 2-3) that squeeze the beam in one direction and disperse it in the other. A few quadrupoles in series, with certain field geometry, can focus or defocus the beams. They keep the beams from dispersing in the beam pipes, focus the beams to high intensity before collisions, and defocus them after collisions. Quadrupoles are also installed in twin-bored cold masses, each about 3.1 m long. Several types of small-scale multipole correctors are installed as components of the main dipoles and quadrupoles, which help to fine-tune the beam parameters. The insertion magnets serve various purposes: to adjust

the beam parameters to the needs of each dedicated experiment, or to abruptly change the direction of the beam for injection or abortion. Most insertion magnets are based on Nb-Ti superconductors, while some, in radiation areas, are built of normal conducting material. The electric currents in these various magnets range from 60 A (for small correctors) to 12 kA (for main dipoles and quadrupoles), while the total energy stored in the magnets is about 10 GJ during full operations of the LHC.

### 2.1.2 Luminosity of collisions

At LHC, protons are clustered into bunches in each beam. The two beams circulate in opposite directions and cross each other at a very high rate. The occurrence of a given physics process in the bunch crossings is evaluated by the product of its cross section  $\sigma$  and the instantaneous luminosity  $L$  of crossings. The cross section takes the unit of Barn ( $b = 10^{-24} \text{ cm}^2$ ) and represents the likelihood for a certain physics process to happen under a certain collision energy, while the instantaneous luminosity takes the unit of  $\text{cm}^{-2} \text{ s}^{-1}$  and represents the intensity of proton crossing. The instantaneous luminosity is parametrized as

$$L = \frac{fN^2}{4\pi\sigma_x\sigma_y} F \quad (2-1)$$

where  $f$  is the bunch crossing frequency with a nominal value of 40 MHz at the LHC, and  $N$  is the number of protons in each bunch, which is of order  $1 \times 10^{11}$  for nominal bunches. The term  $4\pi\sigma_x\sigma_y$  measures the effective area of the beam cross section, in which  $\sigma_x$  and  $\sigma_y$  are the Gaussian widths of the beam with nominal values of order  $10 \mu\text{m}$ . The additional factor  $F$  is close to 1. It makes a geometrical correction as the beams do not collide exactly head-on but with a small crossing angle.

The instantaneous luminosity accumulates to the integrated luminosity over a period of collision experiments. The recent data-taking period at the LHC spanned from 2015 to 2018 and is known as the LHC Run 2. Figure 2-4 shows the instantaneous and integrated luminosity delivered by the LHC during Run 2.

Overall, the LHC delivered  $41.0 \text{ fb}^{-1}$ ,  $49.8 \text{ fb}^{-1}$ , and  $67.9 \text{ fb}^{-1}$  of collision data to CMS

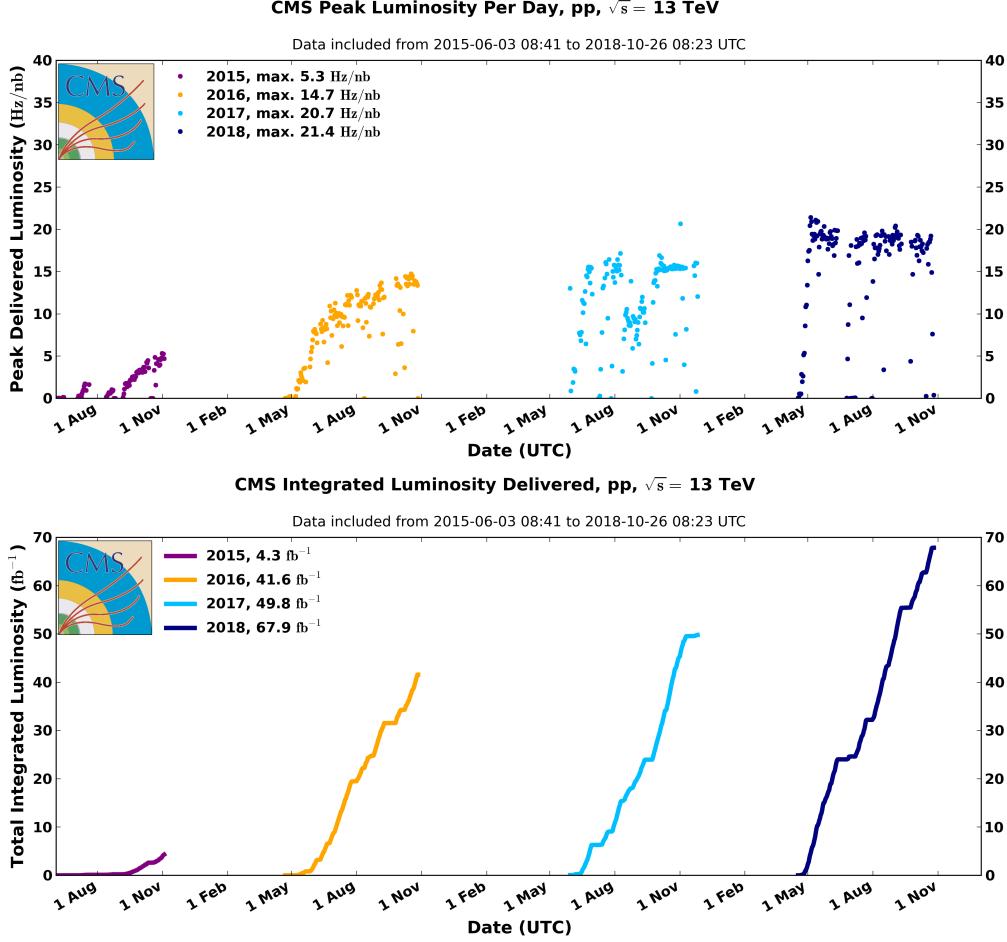


Figure 2-4. The evolutions of the peak instantaneous luminosity (upper plot) and the cumulative integrated luminosity (lower plot) of the LHC Run 2 (2015-2018)  $pp$  collisions. Plot taken from Ref. [5].

during 2016, 2017, and 2018, respectively. A fraction of these data is not recorded by CMS or recorded but not good for physics analyses, because some detector components may not be in working conditions. The datasets good for physics analyses are  $35.9 \text{ fb}^{-1}$ ,  $41.5 \text{ fb}^{-1}$ , and  $59.7 \text{ fb}^{-1}$  for the three years, summing to about  $137 \text{ fb}^{-1}$  in total. The 2015 dataset is much smaller compared to the other years. It adds little to the overall statistics but cost a lot to calibrate and analyze. It is usually not included in the Run 2 data because of this low cost efficiency. The  $H \rightarrow \mu\mu$  analysis described in the follow chapters of this work is based on this  $137 \text{ fb}^{-1}$  of the Run 2 data.

### 2.1.3 Proton collision environment

Proton-proton ( $pp$ ) collisions at the LHC initiate a diverse range of processes over a wide energy spectrum, making an ideal tool for general purpose studies and searches for new phenomena. Quarks and gluons, together called partons, in protons can initiate QCD processes, in which a cascade of quarks and gluons are produced, which in turn hadronize and form various hadrons at high multiplicity. Quarks also participate in the electroweak interaction, producing gauge bosons that may decay to leptons. The Higgs boson can also be produced from quarks and gauge bosons. Proton collision experiments are also a great way to reach for high energy and high luminosity. Compared to other experimental options with leptons or antiprotons, protons are easier to obtain (than antiprotons) and accelerate (than leptons).

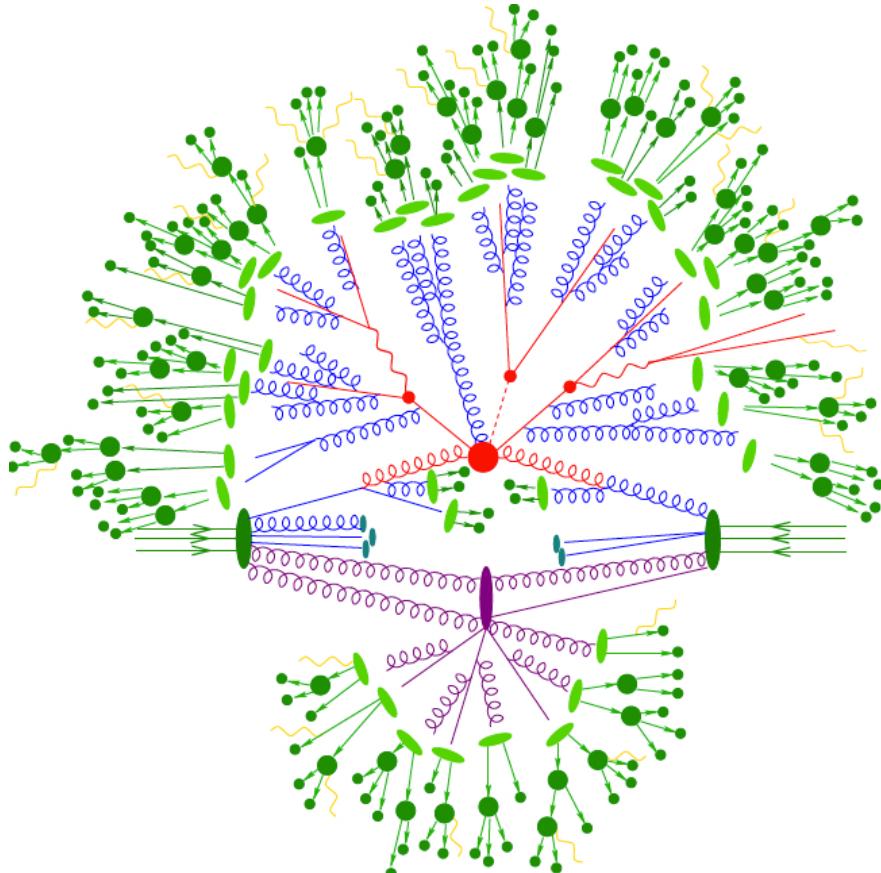


Figure 2-5. An illustration of the interactions in a  $pp$  collision. The primary interaction in this example is the Higgs boson production in association with a top quark pair, whose Feynman diagram is shown in 3-2. Plot taken from Ref. [6].

Meanwhile, proton collisions are a extremely noisy environment. Figure 2-5 demonstrates an example of the interactions initiated in an energetic  $pp$  scatter instance. The primary interaction in this example is the production of the Higgs boson associated with a top quark pair ( $t\bar{t}H$ ), in which the loopy red lines represent the incoming gluons, the big red blob is the vertex of the primary interaction, and the small red blobs are the decay vertices of the Higgs boson (red dashed line) and the top (and antitop) quarks. Additional QCD radiations are indicated by the loopy blue lines, which undergo hadronization (light green blobs) and form hadrons (dark green blobs). The hard interaction in this example is accompanied with a softer secondary interaction (purple blob), which produce a bunch of hadrons through QCD processes. Finally, photons (curvy yellow lines) can be emitted from the final-state hadrons and leptons.

Furthermore, multiple  $pp$  collisions occur in each proton bunch crossing at the LHC, known as pileup. In most cases, all of these simultaneous interactions are QCD interactions, while occasionally one of the collisions is a hard scatter that leads to processes interesting to physicists. In those occasions, the collision containing the hard scatter is considered as the primary interaction, and the other collisions are called pileup interactions. Within the primary interaction, the process of interest, for example the  $t\bar{t}H$  process in Figure 2-5, is called the prompt interaction, while the other QCD-induced byproducts are called the underlying event.

## 2.2 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) [37] is a general purpose detector operating at one of the collision sites of the LHC. It is named after its large-bore superconducting solenoid magnet, which provides a 4 T field at its core and enables precise measurements of the various collision products. The CMS detector has an overall length of 28.7 m, a diameter of 15.0 m and a mass of 14000 t. A cutaway diagram of CMS is shown in Figure 2-6. It consists, from inside to outside, of a silicon-based tracking system (blue slices in the figure), a lead tungstate crystal electromagnetic calorimeter (ECAL) (cyan

blocks), a brass-scintillator hadron calorimeter (HCAL) (yellow blocks), a superconducting solenoid (white blocks), and an iron return yoke (red blocks) interleaved with a multi-layer muon detector (white panels). These detector components are described individually in Section 2.2.1 to 2.2.4. In addition, the data acquisition of CMS relies on a trigger system, which is described in Section 2.2.5.

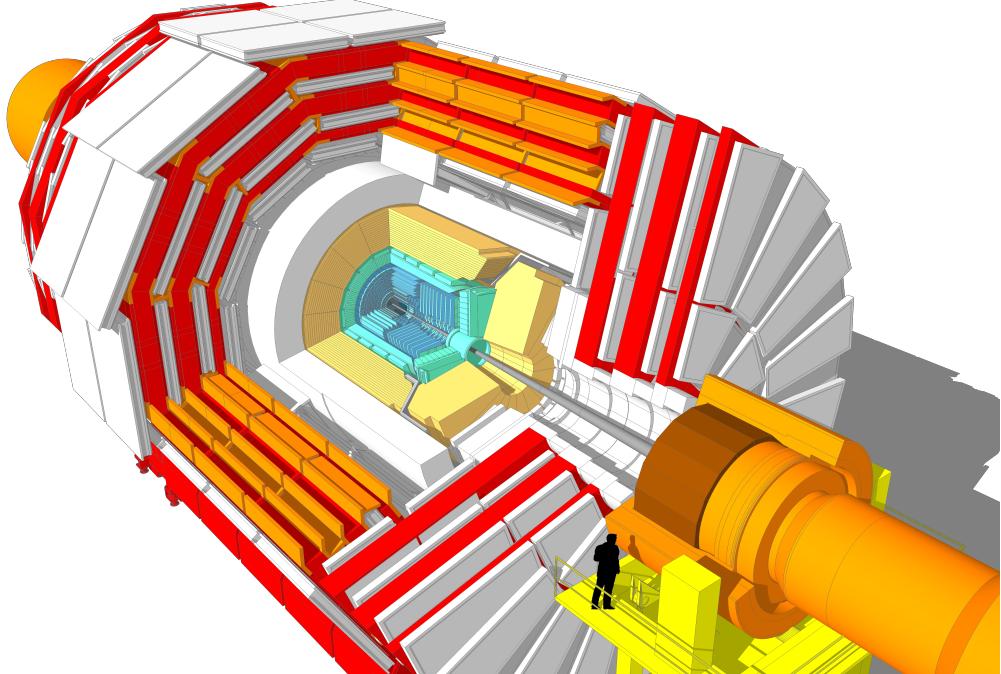


Figure 2-6. A cutaway diagram of the CMS detector. The design is specified to a design stage called the Phase-1 detector upgrade [7, 8, 9]. Plot taken from Ref. [10].

CMS adopts a cylindrical coordinates convention: the origin is positioned at the geometrical center of CMS expected for collisions to happen; the  $z$  axis is along the beam pipe with its positive direction pointing toward the Jura mountain; the  $\phi = 0$  direction (or the  $x$  axis of the Cartesian coordinates) points horizontally toward the LHC ring center; this leaves the  $y$  axis of the Cartesian coordinates pointing upward to the sky. In addition, the polar angle  $\theta$  is in most cases replaced by a variable called pseudorapidity  $\eta$ , defined as  $\eta = -\ln[\tan(\frac{\theta}{2})]$ . The pseudorapidity is a good approximation of the longitudinal rapidity

$y_L$  of particles in the CMS frame in the limit of  $|\mathbf{p}| \gg m$ . The CMS coordinates, along with the  $\eta - \theta$  correspondence, are shown in Figure 2-7.

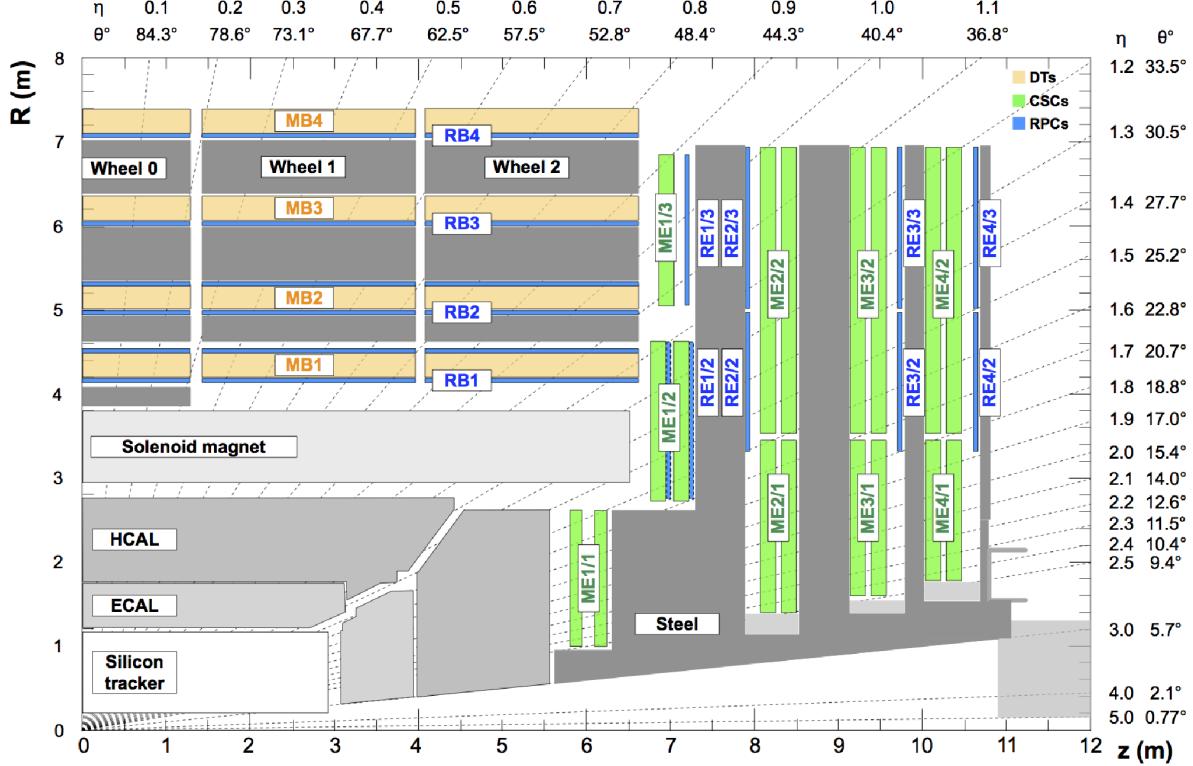


Figure 2-7. A longitudinal section view of the CMS detector, gridded with the CMS coordinates. Plot taken from Ref. [11].

### 2.2.1 Solenoid magnet

The solenoid magnet of CMS is based on the same material as the LHC magnets, operating also at 1.9 K. During its full operation, the superconducting coil carries an electric current of 18500 A and stores an energy of 2.6 GJ. It generates a homogeneous 3.8 T field inside its bore of 6-m diameter and 12.5-m length, providing the functioning environment for the detector components installed inside of it. The superconducting coil is enclosed by its iron yoke, which is the heaviest part of the CMS detector, weighing about 12000 t, almost twice as in the Eiffel Tower. The iron yoke guides the magnetic field outside of the coil, and in the meantime serves as the supporting frame for all other CMS detector components. Figure 2-8 shows the operating magnetic field of CMS.

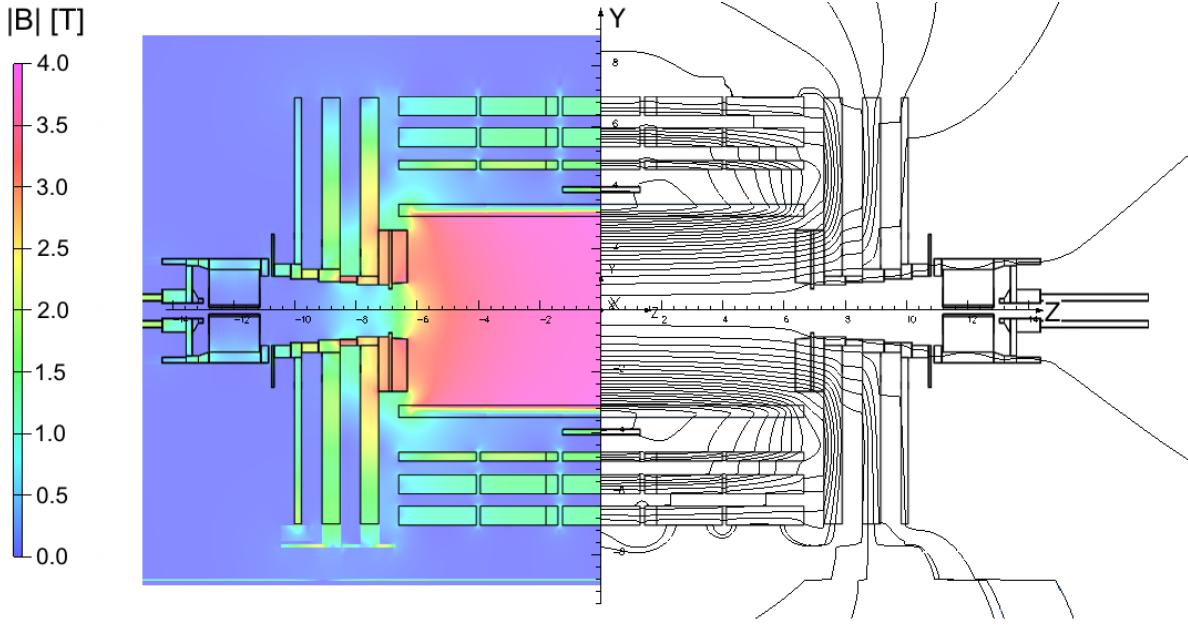


Figure 2-8. The magnetic field in CMS displayed in a longitudinal section, operating with a central field strength of 3.8 T. The field value  $|B|$  is shown on the left and the field lines are shown on the right. Plot taken from Ref. [12].

### 2.2.2 Silicon tracker

The CMS inner tracker is designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions. It is laid out in a cylindrical volume of 5.8-m length and 2.5-m diameter surrounding the beam pipe, shown in Figure 2-9. It is composed of a pixel detector with four barrel layers at radii of 2.9 cm, 6.8 cm, 10.9 cm, and 16.0 cm [13], and a silicon strip detector with 4 + 6 barrel layers extending to a radius of 1.1 m [37]. Each system is completed by endcaps, which consist of three disks in the pixel detector and 3 + 9 disks in the strip detector. The CMS tracker has about 124 million pixel channels and about 9.3 million strip channels in total, and provides its full tracking ability up to a pseudorapidity range of  $|\eta| < 2.5$ .

In the pixel detector, the standard pixel size is  $100 \times 150 \mu\text{m}^2$  in  $r\phi \times z$  plane, with a thickness of  $285 \mu\text{m}$ . At the nominal LHC luminosity, about 1000 charged particles are produced in each bunch crossing, corresponding to a hit occupancy of the order  $10^{-4}$  per pixel per bunch crossing.

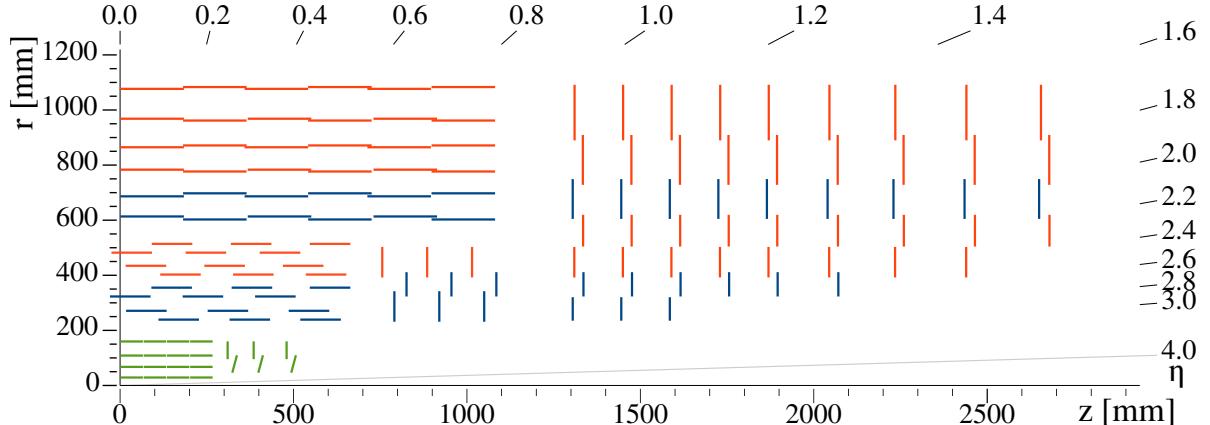


Figure 2-9. Sketch of one quarter of the Phase-1 CMS tracking system in  $r$ - $z$  view. The pixel detectors are shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively. Plot taken from Ref. [13].

At its operation, a charged particle usually generate signals in a few neighboring pixels, known as the charge-sharing. The pixel system reads with analog pulse height readout, which enables an interpolation between the neighboring pixels and achieves a spatial resolution in the range of 15-20  $\mu\text{m}$ .

The strip detector is made up of two subsystems in two regions, the inner region ( $20\text{ cm} < r < 55\text{ cm}$ ) and the outer region ( $55\text{ cm} < r < 110\text{ cm}$ ), both composed of silicon micro-strip detectors. A typical micro-strip cell in the inner region has a thickness of 320  $\mu\text{m}$  and size of  $10\text{ cm} \times 80\text{ }\mu\text{m}$ , leading to an occupancy of up to 2-3% per strip per LHC bunch crossing. Micro-strip cells in the outer region, given the larger radii and reduced particle density, are larger in size: 500  $\mu\text{m}$  in thickness and up to about  $25\text{ cm} \times 180\text{ }\mu\text{m}$  in size, and corresponding to an occupancy of about 1%. The spatial resolution of the strip cells, after the interpolation of charge-sharing, ranges from 23-35  $\mu\text{m}$  in the inner barrel layers, and from 35-53  $\mu\text{m}$  in the outer barrel layers.

All strip cells are placed parallel to the beam pipe in the barrel, and along the radial direction in the endcaps. To provide a measurement of the coordinate along the strip length ( $z$  in the barrel and  $r$  in the endcaps), some layers of the strip detector are constructed with a double-strip design, in which a second micro-strip detector module is

mounted back-to-back to each original strip module with a stereo angle of 100 mrad. This measurement achieves a resolution of 230  $\mu\text{m}$  in the inner barrel and 530  $\mu\text{m}$  in the outer barrel, while the resolution varies in the endcap disks depending on the hit location.

The whole tracking system, consisting of numerous silicon sensors and their readout system, consumes about 60 kW of electric power, which in turn is dissipated as heat in the tracker volume. A cooling system is built to maintain its operation temperature of -10 °C, in which the pixel layers are cooled with aluminum conducting tubes, and the strip layers are cooled with a continuous flow of C<sub>6</sub>F<sub>14</sub> liquid.

### 2.2.3 Calorimeters

The tracker provides high precision measurements on the charged particles but does not directly tell their particle types. It does not acquire information on the numerous neutral particles like photons and neutral hadrons, either. Calorimeters are placed outside of the tracker to absorb most types of particles and provide energy measurements on them. Photons and electrons interact with matter primarily through the electromagnetic interaction, where photons convert into electron-positron pairs while electrons (and positrons) emit bremsstrahlung photons. As a result, a photon or electron incident to thick material turns into a cascade of photons and electrons, called an electromagnetic shower. Similarly, hadrons interact primarily with nuclei in matter through the strong interaction, where multiple secondary hadrons are produced, which in case of dense material form a cascade of hadrons called a hadronic shower. The spatial developments of the electromagnetic showers and hadronic showers are characterized by the radiation length ( $X_0$ ) and the nuclear interaction length ( $\lambda_I$ ), respectively. CMS contains an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL) to initiate and measure these two types of showers, and in turn measure the energies of the corresponding particle types.  $X_0$  is in general much smaller than  $\lambda_I$ , therefore the ECAL is placed on the inside of the HCAL.

The ECAL is a hermetic homogeneous calorimeter made of 75848 lead tungstate (PbWO<sub>4</sub>) crystals, 61200 of which are mounted in the barrel part and 7324 in each of the

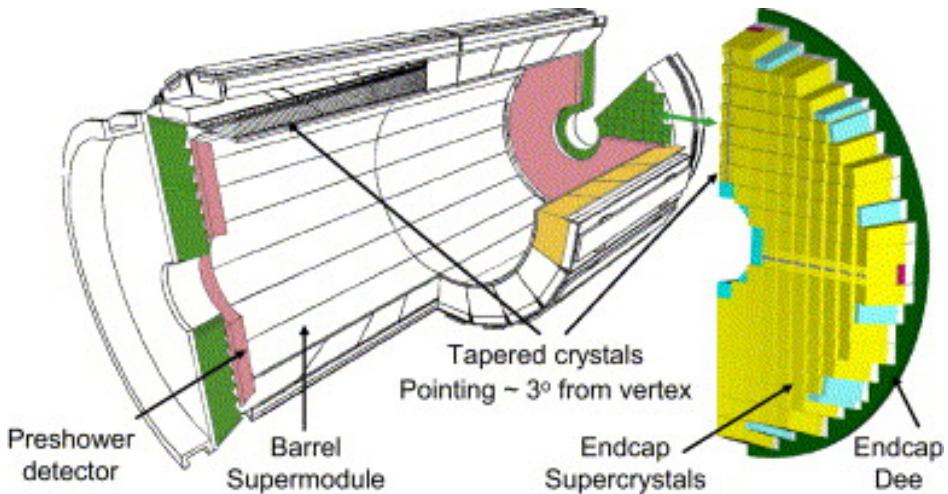


Figure 2-10. Schematic view of the CMS ECAL. Plot taken from Ref. [14].

two endcaps. It is laid out in a cylindrical volume encircling the tracker, as shown in Figure 2-10. The barrel part of the ECAL (EB) covers the pseudorapidity range of  $|\eta| < 1.479$ , with a granularity of 360-fold in  $\phi$  and  $2 \times 85$ -fold in  $\eta$ . They are mounted not exactly toward the collision region but with a small angle of  $3^\circ$  in both the  $\phi$  and  $\eta$  projections. This avoids the intermodule cracks aligning with potential particle trajectories and creating blind regions. The front faces of the crystals are at a radius of 1.29 m. Each barrel crystal has a cross section of  $22 \times 22 \text{ mm}^2$  at the front face and  $26 \times 26 \text{ mm}^2$  at the rear face, and a length of 230 mm ( $25.8X_0$ ). PbWO<sub>4</sub> has a density of  $8.28 \text{ g/cm}^3$ , and overall the barrel crystals weigh 67.4 t. The endcap ECAL (EE) covers the pseudorapidity range of  $1.479 < |\eta| < 3.0$ , with the front faces of crystals 315.4 cm away from the collision region in the longitudinal direction. The endcap crystals have an identical shape, with a cross section of  $28.62 \times 28.62 \text{ mm}^2$  at the front face and  $30 \times 30 \text{ mm}^2$  at the rear face, and a length of 220 mm ( $24.7X_0$ ). The crystals are grouped in mechanical units of  $5 \times 5$  crystals arranged in a rectangular  $x - y$  grid. All crystals point at a focus 1300 mm beyond the collision region, so that the intermodule cracks are  $2^\circ - 8^\circ$  tilted from the directions to the collision region. Each endcap is divided into 2 semicircles called *Dees*, each containing 138 standard  $5 \times 5$  units and 18 partial units on the inner and outer circumferences. The two endcaps together weigh 24.0 t.

$\text{PbWO}_4$  is an optically transparent material that scintillates at a wavelength of 420–430 nm when excited by electrons and photons passing through it. The crystals have a refractive index of  $n = 2.29$  around the scintillation wavelength, and are optically isolated from each other. Photons undergo total internal reflection in the crystal and are collected by the photodetector installed at the rear face of the crystal. The number of scintillation photons emitted by the crystals and the amplification of the photodetectors are both temperature dependent. To maintain a stable ECAL performance, the operation temperature of the ECAL is kept precisely within  $18 \pm 0.05^\circ\text{C}$  by constant  $18^\circ\text{C}$  water flows, which thermally couple to the crystals and electronics via aluminum grids.

The ECAL resolution is affected by several factors: the stochastic fluctuation in the shower containment, in photostatistics, and in the energy absorbed by preceding detector material along the particle trajectory; the noises in electronics and from pileup interactions; and the constant error in the crystal nonuniformity and in detector calibration. Overall, the resolution response can be parametrized as:

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

in which  $\sigma$  is the resolution,  $E$  is the deposit energy,  $S$  is the stochastic term with an empirical value of 2.8%,  $N$  is the noise term with an empirical value of 12%, and  $C$  is the constant term with an empirical value of 0.3%.

In addition to the crystals, a preshower detector is placed in front of each ECAL endcap. It is a sampling calorimeter covering a pseudorapidity range of  $1.653 < |\eta| < 2.6$ , consisting of two layers of lead radiator + silicon strip sensor composite. The lead radiator initiate electromagnetic showers from incoming photons and electrons and the strip sensors measure the shower energy without absorbing them. Each strip is 1.9 mm wide, and the strip orientation in the two layers are orthogonal. This provides a much better spatial resolution than the crystals and significantly improves the identification of single photons from collinear photon pairs produced by the decay of neutral pions.

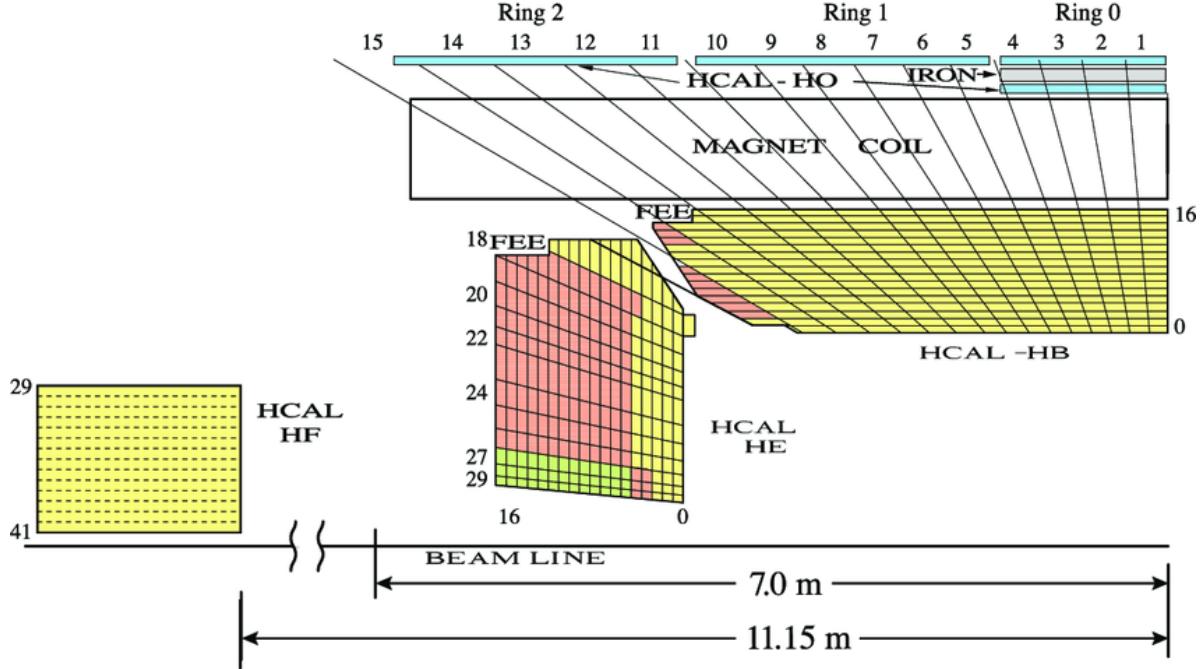


Figure 2-11. Schematic view of one quarter of the Phase-1 CMS HCAL in r-z view, showing the locations of the HB, HE, HO, and HF calorimeters. Plot taken from Ref. [8].

The HCAL is designed to absorb and measure hadrons produced in  $pp$  collisions. It consists of four main parts: a barrel calorimeter (HB) placed in between the outer extent of the EB (at  $R = 1.77\text{ m}$ ) and the inner extent of the magnet coil (at  $R = 2.95\text{ m}$ ), two endcap calorimeters (HE) located in between the rear of the EE (at  $z = \pm 3.9\text{ m}$ ) and the front of the first endcap muon station (at  $z = \pm 5.6\text{ m}$ ), an outer calorimeter (HO) installed outside of the solenoid complementing the HB, and two forward forward calorimeters (HF) outside of the last endcap muon station at  $z = \pm 11.2\text{ m}$  around the beam pipe. The layout of these HCAL components are shown in Figure 2-11. The HB and HO covers the pseudorapidity range of  $|\eta| < 1.3$ , the HE covers  $1.3 < |\eta| < 3.0$ , and the HF covers  $3.0 < |\eta| < 5.2$ .

The HB consists of 36 identical azimuthal wedges, each divided into 4 sectors in  $\phi$  and 16 sectors in  $\eta$ . The wedges are constructed out of 16 flat absorber plates aligned parallel to the beam axis, each has a thickness of 40-75 mm. The front and rear plates are made of stainless steel, while the other 14 plates in between are made of brass. The total absorber thickness ranges from 5.4 to  $10.3\lambda_I$  depending on  $\eta$ , and the ECAL in

front adds about another  $1.1 \lambda_I$  of material. The hadronic energy deposits are sampled by 17 scintillator layers inserted in between the absorber plates (including the front and rear surfaces). The scintillator material is called Kuraray SCSN81, with optical fibers embedded to guide the light. The two outermost scintillator layers are 9 mm thick while the other layers are 3.7 mm. The HE is built with a similar design with 17 interposing 79-mm brass plates and 3.7-mm SCSN81 scintillators, summing to a thickness of about  $10 \lambda_I$ . It consists of 36 azimuthal wedges, each has 5 divisions between  $1.3 < |\eta| < 1.6$  and 9 divisions between  $1.6 < |\eta| < 3.0$ .

As a complement to the limited thickness of HB, the HO sits outside of the solenoid coil, which adds about  $1.4/\sin\theta$  interaction lengths before HO. The HO is geometrically aligned with the muon system and has 5 2.536-m-wide rings along  $z$ -axis. Each ring has a single-layer scintillator at  $r = 4.07$  m. In particular, the HB has the minimal absorber depth near  $\eta = 0$  therefore the central ring has two layers of scintillators at  $r = 3.82$  m and 4.07 m, sandwiching an additional 19.5 cm thick piece of iron. HO rings are divided into 12 identical  $\phi$  sectors and each sector is further divided into 6 slices. Each slice is in turn divided into basic units called tiles along  $\eta$ , with 8 divisions in ring 0, 6 divisions in rings  $\pm 1$ , and 5 divisions in rings  $\pm 2$ .

The HF is overall a cylindrical structure with an outer radius of 130.0 cm and an inner bore radius of 12.5 cm to accommodate the beam pipe. The whole cylindrical structure is housed in a hermetic radiation shielding consisting of composite layers of steel, concrete and polyethylene. Each HF calorimeter is divided into 18 equal wedges in  $\phi$  and 13 units, known as towers, with roughly equal  $\eta$  intervals. Each tower is made of a bundle of quartz fibers that run parallel to the beam line. Quartz is chosen because of its resilience to harsh radiation environment in the forward region: on average 760 GeV of energy per  $pp$  collision is deposited into the HF, while 100 GeV is distributed in the rest of the detector. Signals are detected not as hadronic showers but as the Cherenkov radiation generated by charged particles in the quartz medium, therefore the HF is insensitive to neutral hadrons.

Overall, HCAL has a granularity of  $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$  for  $|\eta| < 1.6$  (HB and HE),  $0.17 \times 0.17$  for  $1.6 < |\eta| < 3.0$  (HE), and  $0.175 \times 0.175$  for  $3.0 < |\eta| < 5.0$  (HF).

#### 2.2.4 Muon detectors

Muon detection at CMS, as implied by its name, is of central importance in its design. CMS uses three types of gaseous particle detectors for muon detection in its original design: drift tubes (DT), cathode strip chambers (CSC), and resistive plate chambers (RPC). Additional gas electron multipliers (GEM) were installed after the LHC Run 2 as a part of the phase-2 upgrade plan of CMS.<sup>1</sup> The locations of different muon detectors are shown in Figure 2-12.

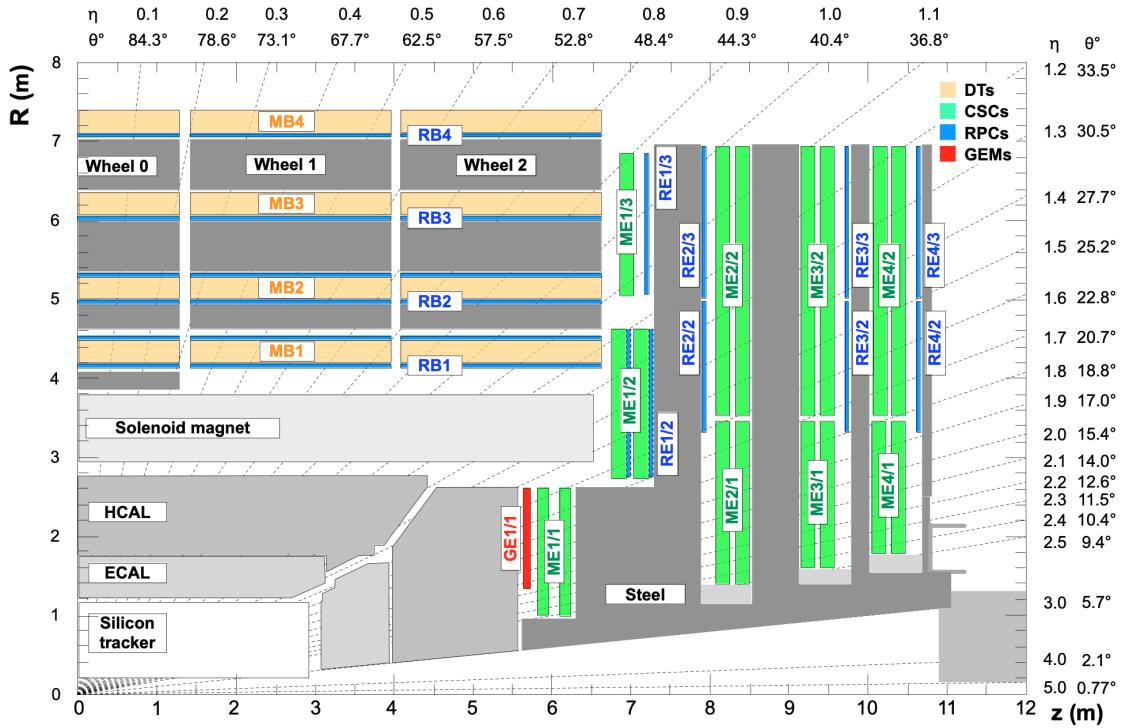


Figure 2-12. A longitudinal section view of the current CMS detector specifying the locations of muon chambers. This plot is different from Figure 2-7 only by the GEM detector installed in 2019/2020. Plot taken from Ref. [15].

The muon system consists of a barrel section and two endcaps, each of them containing four layers of composite detectors, called stations, interleaved with the return yoke of

<sup>1</sup>The physics analysis in this thesis is based on the LHC Run 2 dataset, which does not involve GEM signals.

the solenoid magnet. The barrel of the muon system covers  $|\eta| < 1.2$  with 5 wheels along the  $z$  direction. Each wheel consists of 4 stations along the  $r$  direction and each station is made of 12 sectors. Each sector is a planar structure parallel to  $z$  and perpendicular to  $r$ , and the 12 of them together encircle the barrel. In each sector, there are one DT layer and one or two RPC layers: the DT layers in the first and second stations is sandwiched by two layers of RPC, while the DT layers in the third and fourth stations is only accompanied with an RPC layer on the inner side of it. The muon endcaps cover  $0.9 < |\eta| < 2.4$ , forming an overlap with the barrel to make sure there is no acceptance gap. Each endcap is a disk with 4 stations of muon chambers. Each station is divided into 2 or 3 rings (3 for the first station and 2 for the others), and each ring consists of six  $60^\circ$  sectors. In each sector, there are six  $10^\circ$  CSC chambers except for the inner ring of stations 2-4, which have three  $20^\circ$  degree chambers. The CSC chambers in the inner ring of the first station (ME1/1) are accompanied with GEM chambers whose boundaries are aligned with the CSC chambers. The CSC chambers in the outer rings of station 1-4 are accompanied with RPC chambers aligned with them.

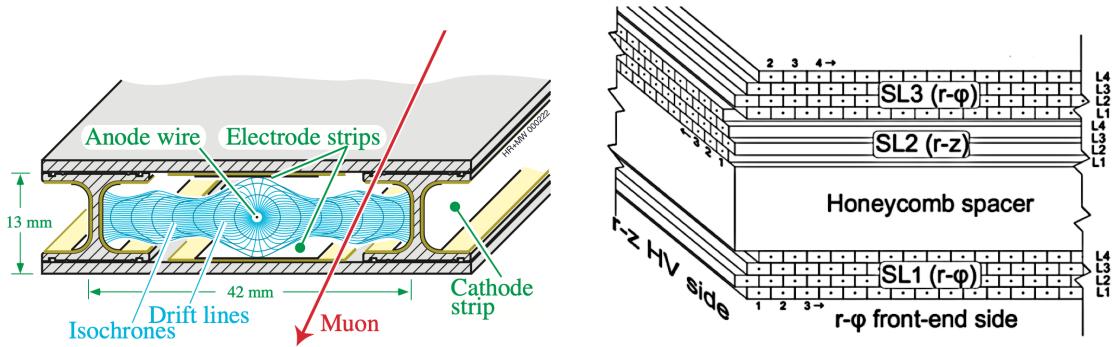


Figure 2-13. Left: sketch of a cell of the drift tube detector showing drift lines and isochrones. Right: schematic view of DT superlayers. Plot taken from Ref. [16].

The DT chambers are arrays of rectangular drift cell, whose design is illustrated in the left plot of Figure 2-13. Each cell has a transverse size of  $42 \times 13\text{ mm}^2$  with a  $50\text{-}\mu\text{m}$ -diameter gold-plated stainless steel wire positioned in the center through its full length. A

DT chamber is made of 3 (or 2 for the fourth station) superlayers, each made of 4 stacked layers of drift cells, staggered by half a cell. The superlayers are aligned alternately along the  $z$  and  $\phi$  directions, shown in the right plot of Figure 2-13. The cell length is 2.5 m for those aligned along  $z$ , and ranges from 1.9 m to 4.1 m for those aligned along  $\phi$ . DT tubes are filled with a gas mixture of 85% Ar and 15% CO<sub>2</sub>. Electric voltages are applied to different parts in the tube to generate a certain potential gradient: the steel wires are at +3600 V as the anodes, the aluminum walls on the 13-mm sides are at -1200 V as the cathodes, and the aluminum walls on the 42-mm sides are at +1800 V to help guide the gradient from anodes to cathodes.

When a muon passes through DT cells, it ionizes the Ar-CO<sub>2</sub> mixture. The resulting free electrons are attracted to the anodes and accelerated in the field, ionizing more gas molecules along their path. This avalanche makes a gain about 10<sup>5</sup> and leads to a detectable current as the signal. The electron avalanche drifts at a roughly constant velocity about 50  $\mu\text{m}/\text{ns}$ , meaning that it takes at maximum close to 400 ns for the electric signal to develop (for incident muons at the farthest corners 21 mm away from the anode wire). This drift time is measured from the signal timing differences between the 4 staggered layers in a superlayer [38]. The determination of this drift time is particularly important as the  $pp$  bunch crossings are only 25 ns apart from each other. Overall, the DT can achieve a time resolution of about 2 ns, which can be translated into a spatial resolution of order 100  $\mu\text{m}$  given the constant drift velocity. However, a caveat is accompanied with this great time and spatial resolution: in a very unlikely case where two or more muons pass through the same drift cell in consecutive bunch crossings, the cell may not recover from the earlier avalanche and become inefficient to the later ones. This is mitigated by the fast-responding RPC detectors, which are described later in this section.

The CSCs, shown as the left plot of Figure 2-14, are multiwire chambers comprised of 6 anode wire planes interleaved among 7 cathode strip panels. The wires are equally spaced of about 3 mm and run along the  $\phi$  direction. Each cathode strip panel contains 80

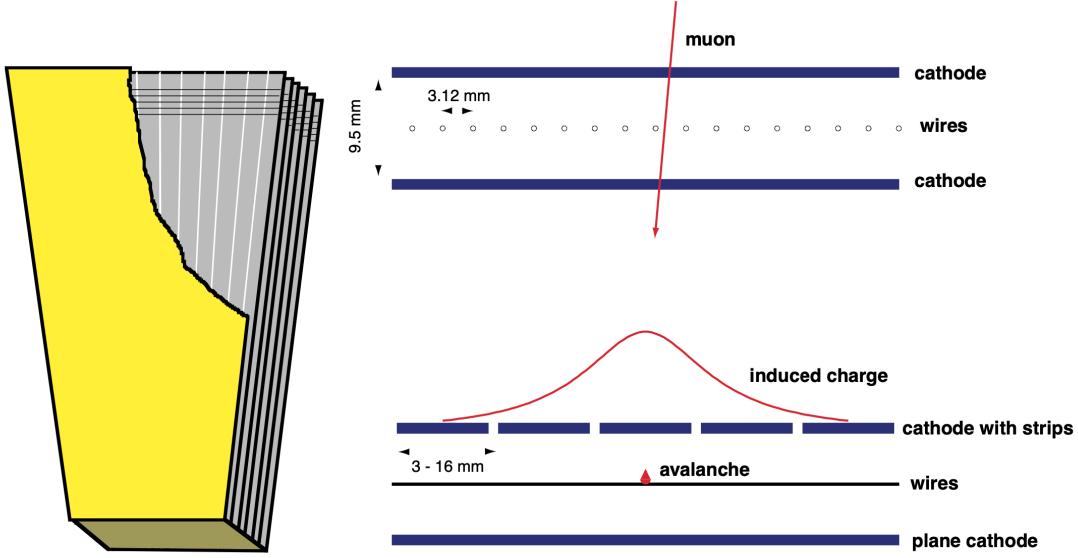


Figure 2-14. Left: layout of a CSC chamber showing arrangement of anode wires and cathode strips. Right: an illustration of the principle of CSC operation. Plots taken from Ref. [16].

strips running along the  $r$  direction, each covering a constant  $\phi$  interval and corresponding to a width between 3 mm to 16 mm. The gap between adjacent cathode panels (with wires in between) is about 1 cm. The overall dimension of CSCs vary for different stations and rings, the largest of which is about  $3.4 \times 1.5 \text{ m}^2$ .

A gas mixture of  $40\% \text{Ar} + 50\% \text{CO}_2 + 10\% \text{CF}_4$  is filled in the CSC chambers and a 3.6 kV voltage difference is applied between the anodes and cathodes. The principle for CSC operation is the same as that described for the DT. The drift time in CSC is within a few bunch crossings as it has a shorter drift length than the DT. After offline calibrations, the CSC can achieve a time resolution of about 3 ns. In particular, when electrons move toward the anode wires, charges are induced in the nearby cathode strips, as illustrated in the right plot of Figure 2-14. By interpolating the distribution of these charges, a spatial resolution of the incident particle can be achieved better than the strip pitch, as good as about 2 mm at the trigger level and around  $100 \mu\text{m}$  in off-line reconstruction.

The RPC chambers are used in both the barrel and endcaps, with their dimensions aligned with the corresponding DT or CSC chambers. A basic double-gap RPC module is

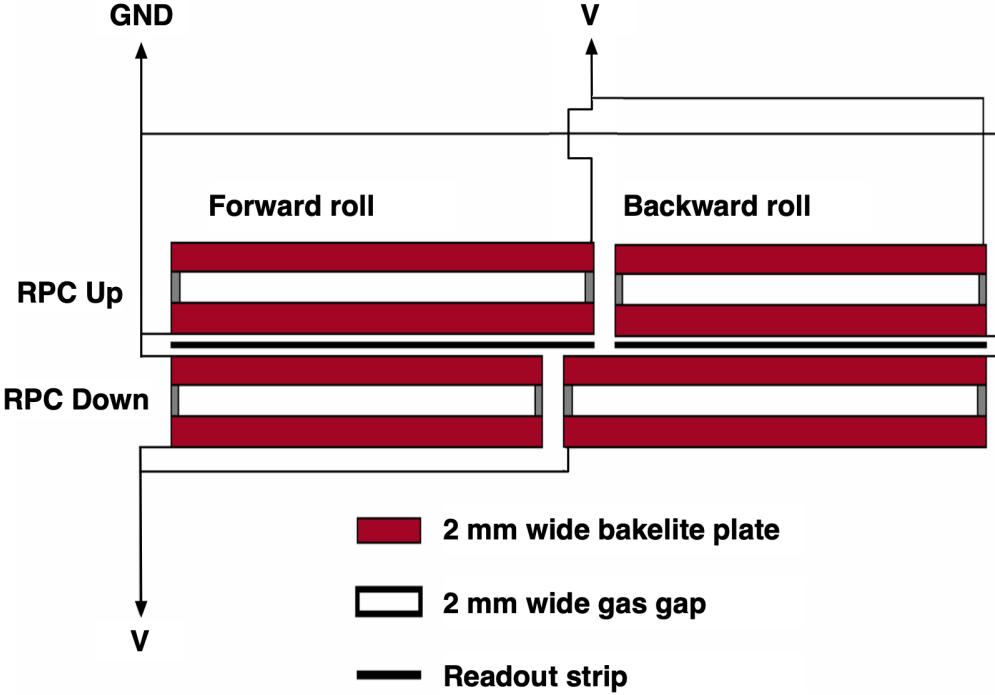


Figure 2-15. Schematic view of a double gap RPC. Plot taken from Ref. [16].

shown in Figure 2-15. Each gas gap consists of resistive plates made of graphite-coated 2-mm bakelite plates, which are separated by 2-mm spacers and a composite gas mixture. A layer of anode strips is inserted between the double gas gaps as the readout. The nominal voltage applied to the outer resistive plates is 9.6 kV.

The RPC features a response rate much shorter than 25 ns, which is crucial in providing bunch crossing assignment and keeping high efficiency of the overall muon system. Its spatial resolution, on the contrary, is worse than the DT and CSC, ranging between 0.8 cm to 1.4 cm.

The GEM detectors are a new component installed in the muon endcaps in front of the CSC inner ring in 2019. At the moment only one GEM station has been installed for each endcap, as shown in Figure 2-12. The basic unit of GEM is a "Triple-GEM detector", shown in Figure 2-16. A triple-GEM chamber features a stack of three GEM foils placed with gaps of a few millimeter immersed in a gas mixture. Each GEM foil is a thin metal-

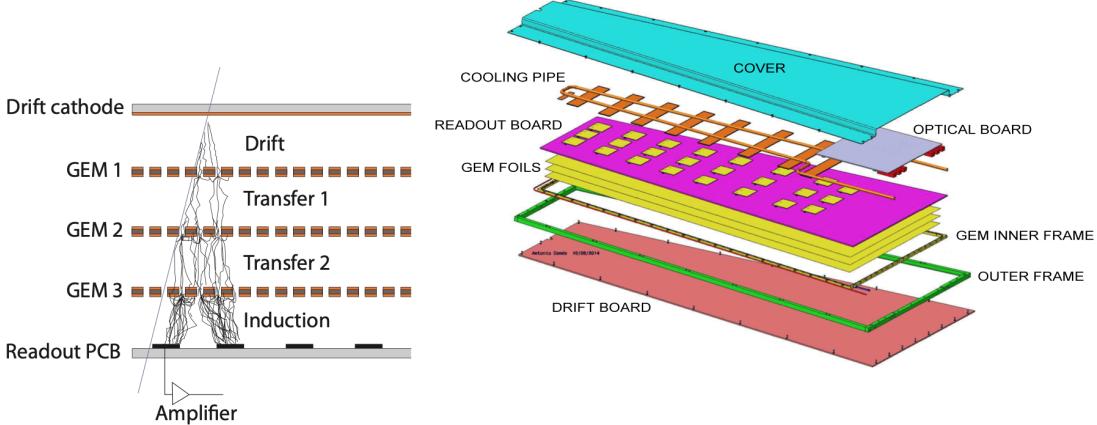


Figure 2-16. Left: schematic view showing the GEM structure and its working principle. Right: exploded view of the mechanical design of GEM. Plot taken from Ref. [15].

clad polymer foil with arrays of holes on it distributed in a hexagonal pattern. GEM foils are encased by a drift board and a readout board, which are connected to a nominal voltage difference of 3200 V. The total distance between the drift board and the readout board is about 7 mm. The holes in GEM foils guide the electric field through them, providing a high avalanche amplification factor while preventing electrical breakdown problems.

The readout board is fine grained with a  $300 \mu\text{rad}$  (or 0.8 mm) precision along the  $\phi$  direction. As the GEM chamber has a short drift distance and a fast drift speed, it is fast-responding with a time resolution better than 10 ns. Two triple-GEM chambers are always stacked to form a superchamber, providing independent measurements. Combination of the two measurements improves the time and spatial resolution and also recover inefficiency, which are expected to be 3% for a single chamber.

### 2.2.5 Trigger system

The LHC produces an immense amount of data colliding protons bunches with a nominal bunch spacing of 25 ns and an instantaneous luminosity of order  $1.0 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . With tens of millions of readout channels in the CMS detector, it is impossible to record all collision data, which correspond to a data flow of about 50 TB/s. Meanwhile, most  $pp$  collisions are soft scatters and do not contain physics processes of primary interests. To

reduce the intense requirement on offline storage and processing, and to filter out uninteresting collisions, a trigger system is implemented in CMS selecting and recording only 0.0025% of the collision events.

The CMS trigger system has two levels. The first level is called the Level-1 Trigger (L1T), which is built from custom electronics specialized in fast parallel processing. It makes decisions based on signals from calorimeters and muon chambers and reduces the event rate from the collision rate of 40 MHz to below 100 kHz. The second level is the High Level Trigger (HLT), a software system operating on a commercial processor farm. It utilizes the readouts from all detectors and reduces the event rate to less than 1 kHz.

The L1T receives coarsely segmented data, known as Trigger Primitives (TP), from the calorimeters and the muon system, while holding the full-precision data in pipelined memory buffers corresponding to a latency of  $4 \mu\text{s}$ . The L1T must analyze every bunch crossing and decide whether to keep the data before they are pushed out of the buffer and lost forever. As electronic signals only travel (at the speed of light) 1200 m in  $4 \mu\text{s}$ , the L1T needs to be physically near the detector for swift communication. Part of L1T electronics is installed on the CMS detector, and the rest is located in the underground control room about 90 m from the experimental cavern. The full-precision data of L1T accepted events are transmitted to the above-ground computing center at a rate of about 100 GB/s and are further analyzed by the HLT. The HLT processes these events among more than 20000 CPU cores with the per-event processing time relaxed to  $O(100 \text{ ms})$ . The HLT runs a full reconstruction algorithm similar to the offline reconstruction described in Section 4.1 and selects events based on their physics signatures. Events passing the HLT selection, corresponding to a data flow of a few GB/s, are distributed in the CERN computing grid for storage and further processing, composing the collision dataset recorded by the CMS experiment.

The structure of the L1T is shown in Figure 2-17, it consists of a muon trigger and a calorimeter trigger. The muon trigger includes three muon track finders (MTF) cover-

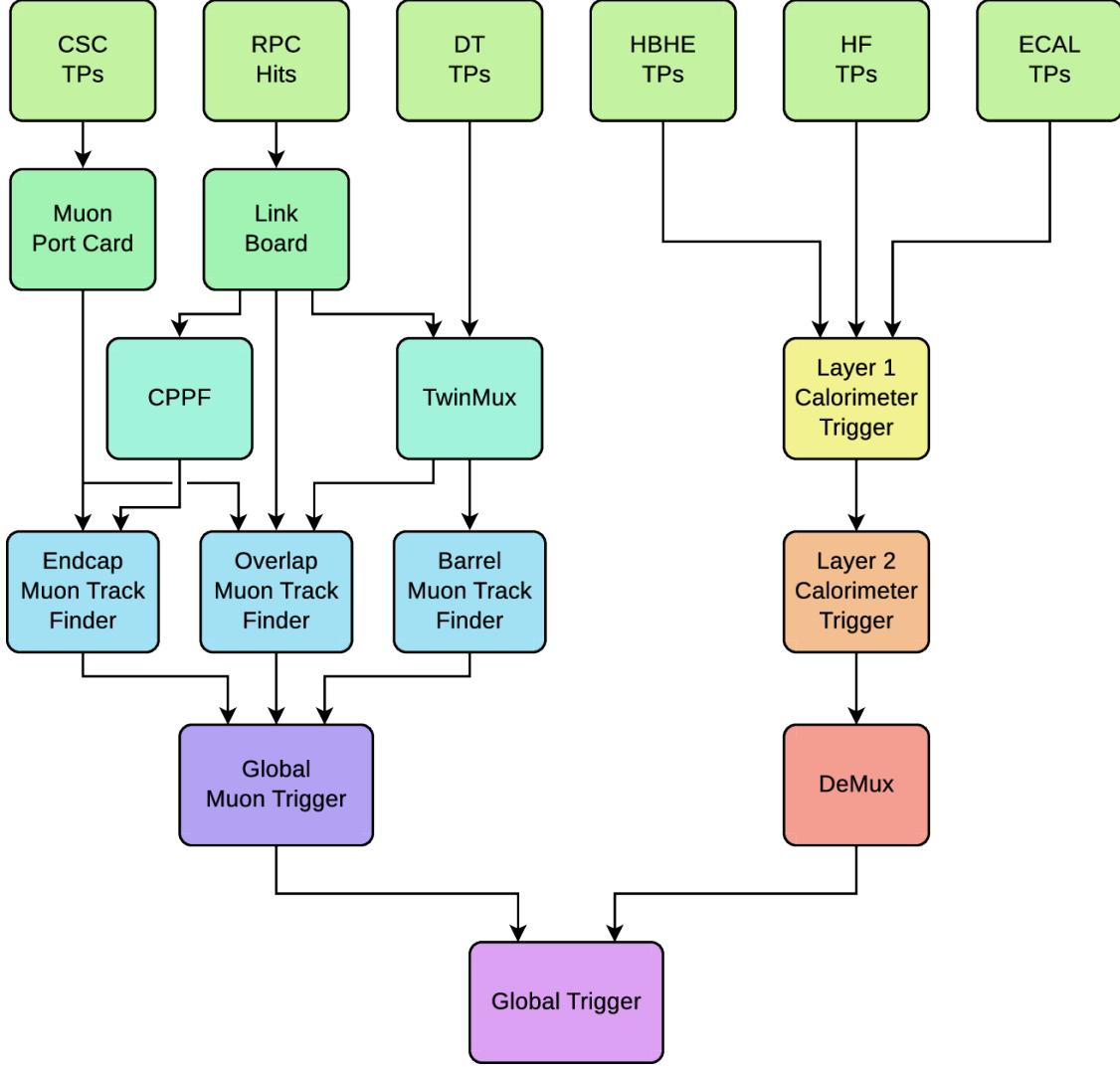


Figure 2-17. Diagram of the CMS Level-1 trigger structure during Run 2. Plot taken from Ref. [17].

ing the barrel (BMTF), overlap (OMTF), and endcap (EMTF) regions, and a global muon trigger for final muon selection. Trigger Primitives (TP) are generated for each detector sector containing information like  $\eta$ ,  $\phi$  positions, timing (bunch crossing), and hit patterns. The CSC TPs are track segments called local charged tracks (LCT) and are transmitted by the muon port cards (MPC). The DT TPs are also track segments. They are combined with nearby RPC hits to form superprimitives in processors called TwinMux [39]. As RPC chambers have less internal layers, the RPC TPs are hits containing only position and timing information. RPC hits in the barrel region are processed by the TwinMux, while hits in

the endcaps are clustered and formatted by the concentration preprocessing and fan-out (CPPF) boards. The BMTF takes inputs from TwinMux, the EMTF takes inputs from both MPC and CPPF, while the OMTF receives inputs within its  $\eta$  coverage from TwinMux, MPC, and the pre-CPPF link boards. The MTFs combine TPs from different muon stations, build muon track candidates, evaluate their momentum, and assign a ranking based on their track qualities. Each MTF can send up to 36 muon candidates to the global muon trigger, which resolves duplicates from different sources, sorts the overall ranking, and sends up to 8 highest ranked muons to the global trigger for final L1T decisions.

The calorimeter trigger consists of two layers. The Layer-1 receives TPs from ECAL and HCAL and calibrates their energy deposits. The Layer-2 combines the calibrated TPs and reconstructs trigger objects such as electrons, photons, tau leptons, jets, and energy sums. (Electrons and photons are indistinguishable at L1T level and are together referred to as  $E/\gamma$  candidates.) The Layer-2 is a time multiplexed trigger [40], in which different bunch crossings are analyzed in parallel by multiple identical boards. In this way each board has a more relaxed processing time and can access the full acceptance and granularity of the ECAL and HCAL. A demultiplexer (DeMux) board summarizes, reorders, and formats the output from multiplexers and transmits the results to the global trigger.

All the mentioned L1T components are built on Xilinx Virtex-7 Field Programmable Gate Array (FPGA) boards. This design greatly boosts the reusability and flexibility of the trigger system, and reduces the workload for development and maintenance. The largely programmable boards also enable the application of complex multivariate algorithms at trigger level in the form of look-up tables.

The global trigger examines all muon and calorimeter objects and makes decisions based on a list of trigger requirements known as the trigger menu. The trigger menu consists of about 400 trigger seeds combined with "or" logic. Each trigger seed is a set of requirements on the  $p_T$ ,  $\eta$ , isolation, and other quantities of certain trigger object(s). For example, a single muon trigger may ask for a muon with  $p_T > 22$  GeV and to be isolated

from other trigger objects, and a single muon + double jets trigger may ask for one muon and two jets in the events that are close together. A detailed list of trigger seeds can be found in Ref. [17]. The selection requirements in trigger seeds are adjusted so that the total trigger rate is contained within a reasonable level. For some cases where there is a need to make loose selections while keeping the trigger rate low, a "prescale" method is applied. A trigger with a prescale of  $N$  means that in every  $N$  events passing this trigger, only one of them is accepted and recorded. In general, the unprescaled triggers are used to collect data for physics analyses, while the prescaled triggers are used for calibrations and trigger performance studies. One extreme example of the prescaled triggers is the zero-bias trigger, which does not require any particular objects from the proton collisions. It fires at a very low rate and makes the baseline dataset for various trigger studies.

The HLT menu is similar to the L1T menu but is based on much better reconstructed physics objects. In particular, the trigger paths taken by the analysis of  $H \rightarrow \mu\mu$  decay are the single isolated muon triggers with a  $p_T$  threshold of 24/27/24 GeV for 2016/2017/2018 datasets. These single muon triggers in general have an efficiency greater than 90% for muons above the trigger threshold. With two muons from the Higgs boson decay, the overall trigger efficiency for signal events is close to 100% with regard to the offline analysis selection detailed in Section 4.2.

## CHAPTER 3

### OVERVIEW OF THE SEARCH OF $H \rightarrow \mu\mu$ DECAY AT CMS

As described in Section 1.4, the Yukawa couplings between the Higgs boson and the elementary fermions are essential parameters to the SM. The Higgs couplings to fermions of the first and second generations are hard to probe because the Higgs boson decay ratios to these light fermions are small. At the LHC, the most experimentally sensitive way to probe such light Yukawa couplings is the study of the  $H \rightarrow \mu\mu$  decay. Prior to this work, searches for the  $H \rightarrow \mu\mu$  decay has been conducted using  $pp$  collision data collected at center-of-mass energies of 7, 8, and 13 TeV by CMS [41, 42] and ATLAS [43, 44, 45], among which the most sensitive result [42] is reported by CMS with an observed (expected in absence of  $H \rightarrow \mu\mu$  decay) upper limit of 2.9 (2.2) times the SM prediction of the Higgs boson production and  $\mathcal{B}(H \rightarrow \mu\mu)$ , at the 95% confidence level (CL). This thesis reports the latest  $H \rightarrow \mu\mu$  analysis based on  $137 \text{ fb}^{-1}$  of data collected by CMS from 2016 to 2018 [36].

The search for the  $H \rightarrow \mu\mu$  decay, in a nutshell, is a struggle to make the signal events stand out from the vast backgrounds with statistical significance. The  $H \rightarrow \mu\mu$  decay has a branching ratio of  $\mathcal{B}(H \rightarrow \mu\mu) = 2.18 \times 10^{-4}$ , which corresponds to an expectation of about 1000 event instances in the data collected by CMS from 2016 to 2018. In contrast, these 1000 signal events are accompanied by millions of events produced through other processes (background events) that mimic their experimental signature. The most important difference between the signal events and the background events lies in the dimuon invariant mass ( $m_{\mu\mu}$ ) distribution, illustrated in Figure 3-1. The SM Higgs boson has a low production cross section and a narrow natural width, making a small sharp peak near 125 GeV. The background events are dominated by the  $Z/\gamma^* \rightarrow \mu\mu$  process (Drell-Yan process), which peaks at 91.2 GeV in the  $m_{\mu\mu}$  spectrum and leaves a smooth falling tail around 125 GeV. The overall signal-to-background ratio ( $S/B$ ) between 120 and 130 GeV is about 1/500 in the 2016-2018 CMS data.

This  $S/B$  can be enhanced based on further kinematic distinctions between various signal and background processes. The Higgs boson is produced via several production

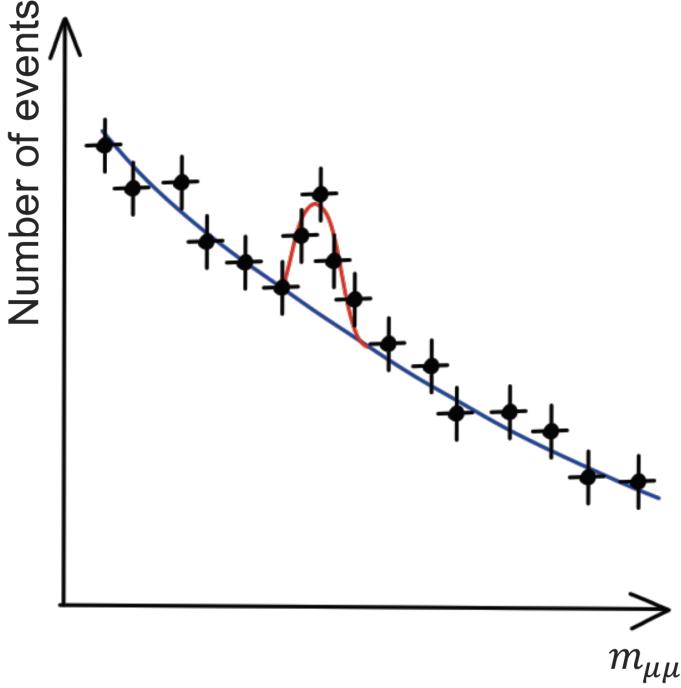


Figure 3-1. A conceptual plot for the dimuon mass shapes for the signal and the background. The blue line shows the expected background shape, while the red line shows the expected signal shape on top of it.

modes. The four main modes, ordered by their cross sections, are gluon fusion ( $ggH$ ), vector boson fusion (VBF or  $qqH$ ), associated production with a weak vector boson ( $VH$ ), and associated production with a pair of top quarks ( $t\bar{t}H$ ). The Feynman diagrams for these main production modes are shown in Figure 3-2. The minor production modes include associated production with a pair of bottom quarks ( $b\bar{b}H$ ), associated production with a Z boson through gluon fusion ( $ggZH$ ), associated production with a top quark and a W boson ( $tHW$ ), and associated production with a top quark and a light quark ( $tHq$ ). The Feynman diagrams for these minor production modes are shown in Figure 3-3. Table 3-1 summarizes the cross sections for all these production modes, along with the expected number of events in the  $137 \text{ fb}^{-1}$  data.

To fully exploit the kinematic profiles of different production modes, the  $H \rightarrow \mu\mu$  analysis is performed in four event categories targeting each of the four main modes, with the analysis procedures optimized separately in each category. No dedicated category is

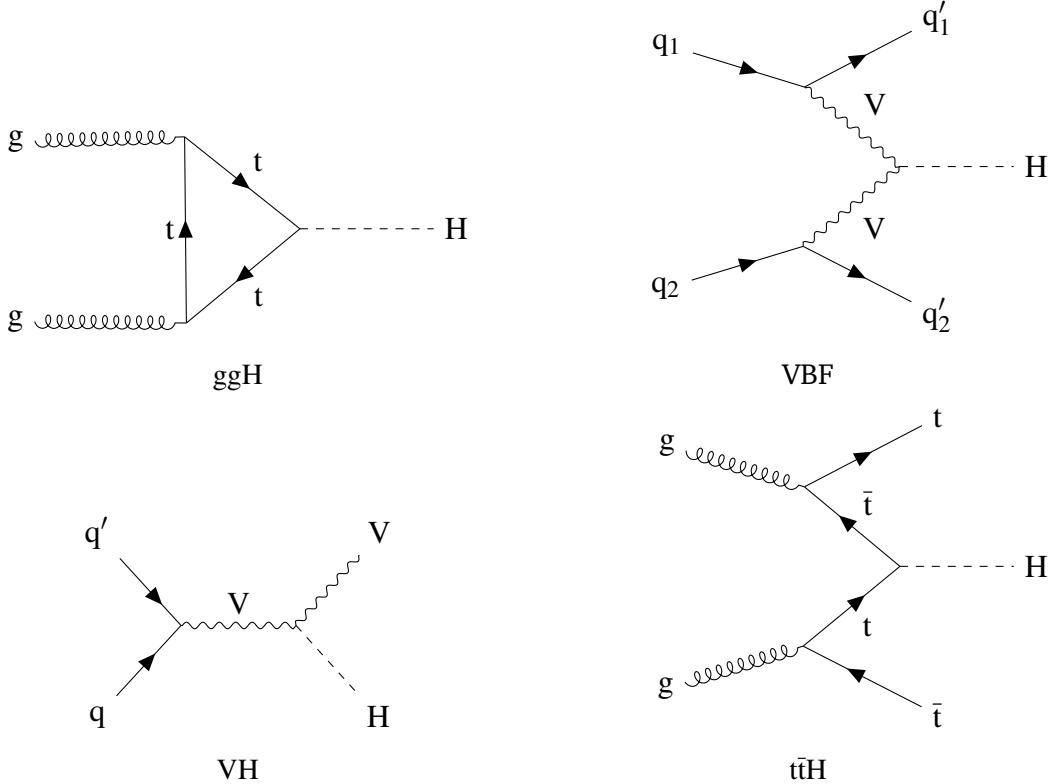


Figure 3-2. Main production modes of the Higgs boson.

made for the minor signal modes, either because they have very similar features to one of the main modes, or because their cross sections are too small to be statistically significant. This chapter gives an overview of the full  $H \rightarrow \mu\mu$  analysis, with Section 3.1 describing the data and simulation samples used in the analysis, and Section 3.2 explaining the analysis strategies in the four exclusive categories.

### 3.1 Data and simulation samples

This analysis uses the  $pp$  collision data collected by CMS from 2016 to 2018, corresponding to an integrated luminosity of  $137\text{fb}^{-1}$ . The triggers used in this analysis are the single muon HLT triggers, which impose some loose isolation requirements and a  $p_T$  threshold on the HLT muon candidates. The  $p_T$  threshold is 24/27/24 GeV for 2016/2017/2018 datasets. The efficiencies of these triggers are above 90% for single muons above the trigger thresholds, and the overall efficiency for events with two muons is close to 100%.

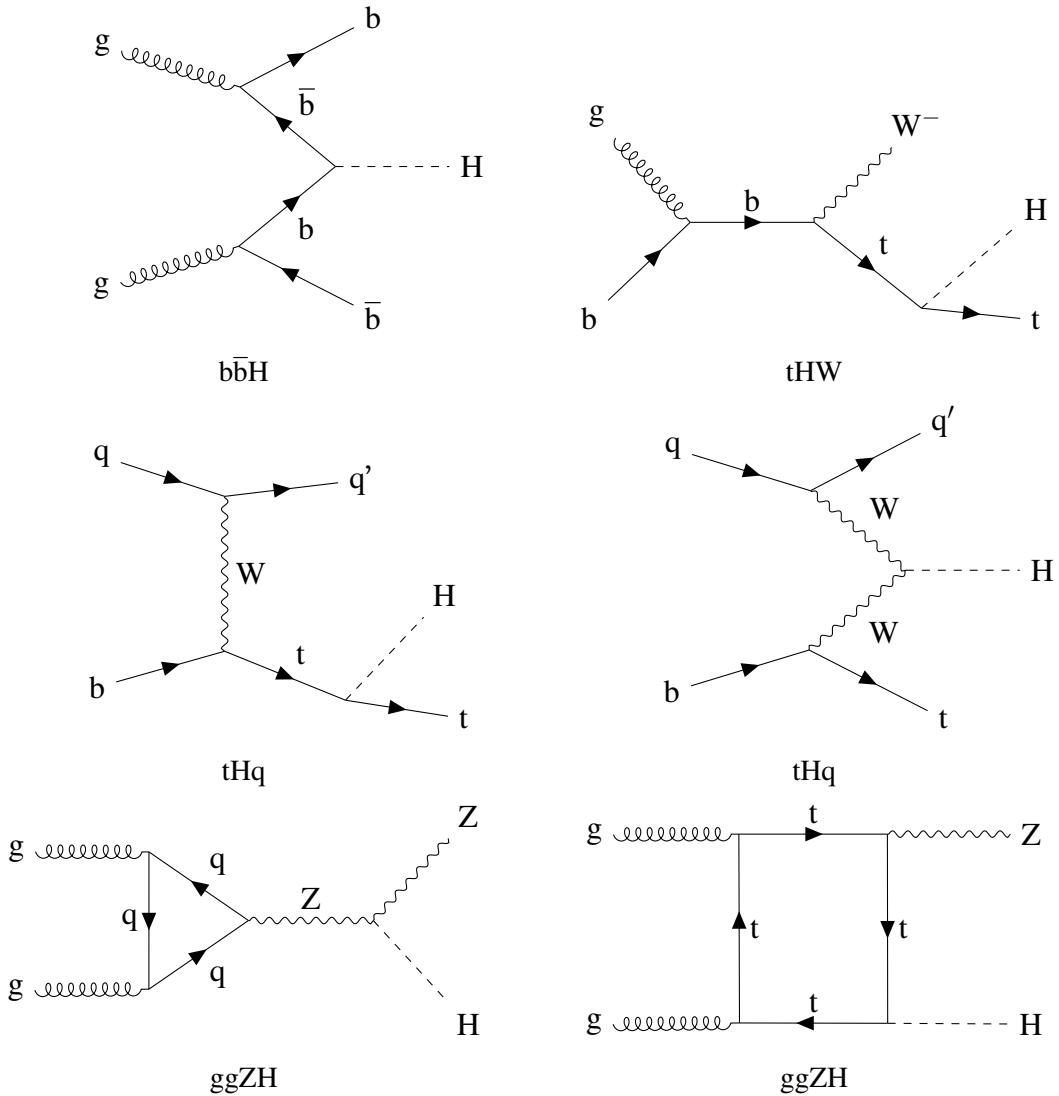


Figure 3-3. Examples of minor Higgs boson production modes.

Simulations of signal and background processes are produced by Monte Carlo (MC) event generators. The generators are listed for different processes in the rest of this section. All simulated samples except the EW  $Z + jj$  samples use PYTHIA 8.2 [46] to model the parton showering (PS), hadronization, and the underlying event (UE), while the EW  $Z + jj$  samples use HERWIG++ and HERWIG7 [47] for the same purpose. The effect of pileup is modeled by overlaying simulated inelastic  $pp$  collisions on the hard-scattering event. The generated events are processed through a simulation of the CMS detector based on GEANT4 [48] and are reconstructed with the same algorithms that are used for data.

Table 3-1. Production modes of the Higgs boson in  $pp$  collisions at the LHC, their cross sections for  $m_H = 125\text{GeV}$ , and the corresponding expected number of  $H \rightarrow \mu\mu$  decays in the dataset for this analysis (the cross sections multiplying  $\mathcal{B}(H \rightarrow \mu\mu)$  and the integrated luminosity ( $137\text{fb}^{-1}$ )). The leptons ( $\ell$ ) in the table refer to electrons or muons.

signal mode	decay mode	Cross section (pb)	Expected $H \rightarrow \mu\mu$ events
ggH	inclusive	48.58	1450
VBF	inclusive	3.782	113
WH	inclusive	1.373	41.0
	$W \rightarrow \ell\nu$	0.293	8.75
qq $\rightarrow$ ZH	inclusive	0.761	22.7
	$Z \rightarrow \ell\ell$	0.051	1.53
ggZH	inclusive	0.123	3.67
	$Z \rightarrow \ell\ell$	0.008	0.25
t $\bar{t}$ H	inclusive	0.507	15.1
	$\geq 1 t \rightarrow \text{leptons}$	0.193	5.76
	Both $t \rightarrow \text{hadrons}$	0.230	6.86
sum of above	inclusive	55.13	1646
bbH	inclusive	0.488	14.6
tHq	inclusive	0.074	2.21
tHW	inclusive	0.015	0.45
sum of all	inclusive	55.70	1663

### 3.1.1 The simulation of the signal processes

The ggH signal process is simulated at next-to-leading order (NLO) accuracy in perturbative QCD, using both the `MADGRAPH5_aMC@NLO v2.4.2` [49] and `POWHEG v2.0` [50, 51, 52, 53] event generators. The  $p_T$  distribution of the Higgs boson in ggH process is then reweighted to match the `POWHEG NNLOPS` prediction [54, 55]. The VBF, WH, qqZH, and t $\bar{t}$ H processes are simulated with `POWHEG v2.0` [56, 57, 58] at NLO precision in QCD. The b $\bar{b}$ H process is simulated at NLO precision in QCD with `POWHEG`. the tHq, and tHW processes are generated at leading order (LO) with the `MADGRAPH5_aMC@NLO` generator. The ggZH process is simulated at LO with the `POWHEG` generator. Simulated signal events are generated, for each production mode, at  $m_H$  values of 120, 125, and 130 GeV. A table summarizing the simulation for signals is shown in Table 3-2.

Expected signal yields are normalized to the production cross sections and  $\mathcal{B}(H \rightarrow \mu\mu)$  values taken from the recommendations of LHC Yellow Report [59]. The ggH

Table 3-2. Summary of the specification for the simulated Higgs signal samples.

Sample	Generator (Perturbative order)	Parton Shower	Cross section	Additional corrections
ggH	MADGRAPH5_amc@NLO (NLO QCD)	PYTHIA	N3LO QCD, NLO EW	$p_T(H)$ from NNLOPS
VBF	POWHEG (NLO QCD)	PYTHIA dipole shower	NNLO QCD, NLO EW	-
qq → VH	POWHEG (NLO QCD)	PYTHIA	NNLO QCD, NLO EW	-
ggZH	POWHEG (LO)	PYTHIA	NNLO QCD, NLO EW	-
t̄tH	POWHEG (NLO QCD)	PYTHIA	NLO QCD, NLO EW	-
b̄bH	POWHEG (NLO QCD)	PYTHIA	NLO QCD	-
tHq	MADGRAPH5_amc@NLO (LO)	PYTHIA	NLO QCD	-
tHW	MADGRAPH5_amc@NLO (LO)	PYTHIA	NLO QCD	-

production cross section is computed at next-to-next-to-NLO (N3LO) precision in QCD, and at NLO in EW theory [60]. The cross section of Higgs boson production in the VBF [61] and qq → VH [62] modes is calculated at next-to-NLO (NNLO) in QCD, including NLO EW corrections, while the t̄tH cross section is computed at NLO in QCD and EW theory [63, 64]. The b̄bH, tHq, and tHW cross sections are computed at NLO in QCD without including higher-order EW corrections [59, 65, 66]. The H →  $\mu\mu$  partial width is computed with HDECAY [67, 68] at NLO in QCD and EW theory.

### 3.1.2 The simulation of the background processes

The background is modeled considering various SM processes, summarized in Table 3-3. The main background in the ggH and VBF categories is the DY process, which is simulated at NLO in QCD using the MADGRAPH5\_amc@NLO generator. The corresponding cross section is calculated with FEWZ v3.1b2 [69] at NNLO in QCD and NLO accuracy in EW theory. The EW production of a Z boson in association with two jets (Z + jj) is an important background in the VBF category. This process is simulated at LO using the MADGRAPH5\_amc@NLO v2.6.5 generator. The WZ, q̄q → ZZ, and WW processes, which constitute the main backgrounds in the VH category, are simulated at NLO in QCD using either the POWHEG or MADGRAPH5\_amc@NLO generators. Their production cross sections are corrected with the NNLO/NLO K factors taken from Refs. [70], [71], and [72]. The gluon-initiated loop-induced ZZ process (ggZZ) is simulated with the MCFM v7.0 generator [73] at LO and the corresponding production cross section is corrected to match higher-order QCD predictions, following the strategy detailed in Ref. [26]. Minor con-

tributions from triboson processes (WWW, WWZ, WZZ, and ZZZ) are also taken into account and are simulated at NLO in QCD using the `MADGRAPH5_AMC@NLO` generator. The main backgrounds in the  $t\bar{t}H$  category involve the production of top quarks. The  $t\bar{t}$  background is simulated with NLO precision in QCD using the `POWHEG` generator, and its cross section is obtained from the `TOP++ v2.0` [74] prediction that includes NNLO corrections in QCD and resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms. The single top quark processes are simulated at NLO in QCD via either `POWHEG` or `MADGRAPH5_AMC@NLO` and their cross sections are computed, at the same order of precision, using `HATHOR` [75]. Finally, contributions from the  $t\bar{t}Z$ ,  $t\bar{t}W$ ,  $t\bar{t}WW$ ,  $t\bar{t}\bar{t}$ , and  $tZq$  processes are also considered and are simulated using the `MADGRAPH5_AMC@NLO` generator at NLO precision in QCD. For the simulated samples corresponding to the 2016 (2017–2018) data-taking periods, the `NNPDF v3.0` (v3.1) NLO (NNLO) parton distribution functions (PDFs) are used [76, 77]. For processes simulated at NLO (LO) in QCD with the `MADGRAPH5_AMC@NLO` generator, events from the matrix element (ME) characterized by different parton multiplicities are merged via the `FxFx` (MLM) prescription [78, 79].

**Table 3-3. Summary of the specification for the simulated background samples.**

Sample	Generator (Perturbative order)	Parton Shower	Cross section	Additional corrections
Drell-Yan	<code>MADGRAPH5_AMC@NLO</code> (NLO QCD)	<code>PYTHIA</code>	NNLO QCD, NLO EW	-
Zjj-EW	<code>MADGRAPH5_AMC@NLO</code> (LO)	<code>HERWIG++/HERWIG7</code>	LO	-
$t\bar{t}$	<code>POWHEG</code> (NLO QCD)	<code>PYTHIA</code>	NNLO QCD	-
Single top quark	<code>POWHEG/MADGRAPH5_AMC@NLO</code> (NLO QCD)	<code>PYTHIA</code>	NLO QCD	-
Diboson (VV)	<code>POWHEG/MADGRAPH5_AMC@NLO</code> (NLO QCD)	<code>PYTHIA</code>	NLO QCD	NNLO/NLO $K$ factors
ggZZ	<code>MCFM</code> (LO)	<code>PYTHIA</code>	LO	NNLO/LO $K$ factors
$t\bar{t}V$ , $t\bar{t}VV$	<code>MADGRAPH5_AMC@NLO</code> (NLO QCD)	<code>PYTHIA</code>	NLO QCD	-
Triboson (VVV)	<code>MADGRAPH5_AMC@NLO</code> (LO)	<code>PYTHIA</code>	NLO QCD	-

### 3.2 Exclusive analyses and their strategies

The  $H \rightarrow \mu\mu$  analysis is conducted independently in four event categories: the ggH, VBF, VH and  $t\bar{t}H$  categories. The workflow to divide events into these categories is shown in Figure 3-4. A prerequisite for all categories, after the trigger selection, is that each events should contain two opposite-charged (or opposite-sign, OS) muons that make the candidate for the Higgs boson decay. Then, as a first step, events containing b-tagged jets (either one medium tag or two loose tag of the DeepCSV [80] working points) are

classified into the  $t\bar{t}H$  category. Events in the  $t\bar{t}H$  category are further divided into the  $t\bar{t}H$  leptonic subcategory if they contain electrons or additional muons, or divided into the  $t\bar{t}H$  hadronic subcategory if they contain at least three jets, or discarded if they contain neither of them. The events without b-tagged jets may fall into the VH category if they contain additional leptons (electrons or muons). Inside the VH category, events are further tagged as WH events if there is one and only one extra lepton in the event, or tagged as ZH events if there are two same-flavor opposite-sign (SFOS) extra leptons. For the events with neither b-tagged jets nor additional leptons, if they have at least two energetic jets composing a jet pair with  $m_{jj} > 400\text{GeV}$  and  $\Delta\eta_{jj} > 2.5$ , they are tagged as the VBF events. Finally, the ggH category collects all events that are not assigned to other categories. The definitions of the different objects used in this categorization is detailed in Section 4.2.

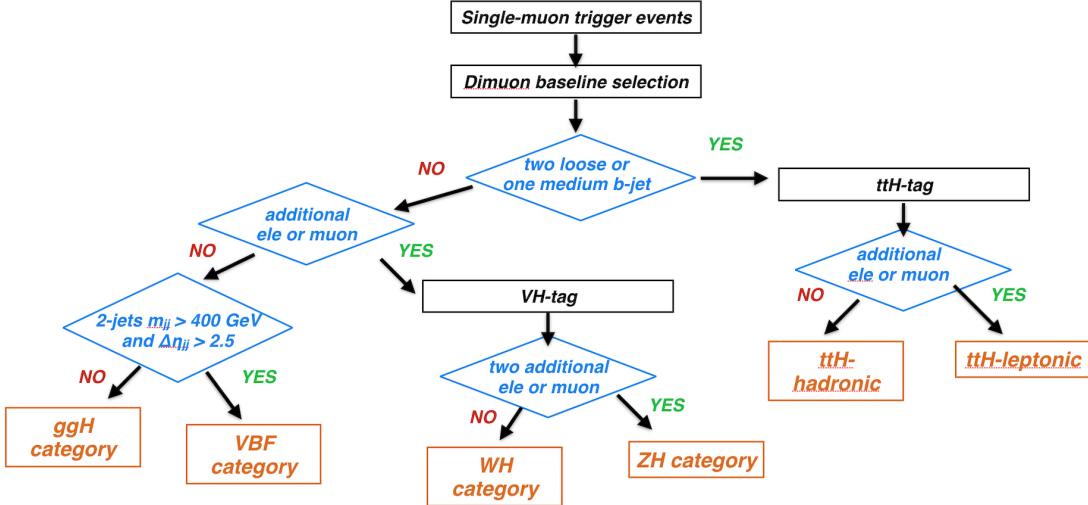


Figure 3-4. Scheme of the procedure of assigning events to different categories. All events passing the common baseline selection are divided into four mutually exclusive categories: ggH, VBF, VH (WH and ZH), and  $t\bar{t}H$  (leptonic and hadronic).

These categories have distinct profiles in expected signal yield, signal purity, and background composition. Therefore the optimal analysis strategies are different. Two strategies are considered for these categories:

- **Data-driven parametric fit to the  $m_{\mu\mu}$  spectrum:** As is done in the previously published analyses on the the data collected prior to 2017 [41, 42], a multivariate anal-

ysis (MVA) method is used to profile the separation between the signal and the background processes. The MVA can be either cut-based as in the Run 1 analysis [41], or machine learning (ML) based as in the analysis on the 2016 data [42]. The MVA considers the kinematic information that is uncorrelated with the  $m_{\mu\mu}$ , and is used to divide the events into several regions with different  $S/B$ , called the MVA-categories. In each MVA-category, the signal strength is evaluated from fits to the  $m_{\mu\mu}$  spectrum in data, in what is called the *signal fit region*, for example  $110 < m_{\mu\mu} < 150 \text{ GeV}$ . Both the signal and the background are modeled by parametric functions that are carefully studied to provide a truthful description of the distributions of physics processes. The total yield of the background is unconstrained in the fit and is determined entirely by the data. The effects of the systematic uncertainties from various sources on either the signal yield or the signal shape are assessed and propagated to the fit result. The systematic uncertainties do not affect the background estimation since it is based on data rather than predictions from simulations.

- **MC-based template fit to the Neural Network discriminator:** This approach is also based on an MVA, for which a ML algorithm, Deep Neural Network (DNN), is taken. The DNN takes all the kinematic variables *including*  $m_{\mu\mu}$ , and profiles the discrimination between the signal and the background. Without making further categories, the binned template of the DNN output in the whole phase space is used for the signal strength evaluation. Since the fit is applied to the DNN output rather than the  $m_{\mu\mu}$  distribution, the *signal fit region* is further divided into two parts: the *signal region*,  $115 < m_{\mu\mu} < 135 \text{ GeV}$ , and the *sideband region*,  $110 < m_{\mu\mu} < 115 \text{ GeV}$  or  $135 < m_{\mu\mu} < 150 \text{ GeV}$ . The data are fit simultaneously in both regions using the DNN templates of the signal and background simulation. The systematic uncertainties affect both the signal and the background prediction, and are employed as variations in either the yield or the shape of the templates. The background yield is estimated from simulation and is allowed to vary within its uncertainty in the fit, in the same manner as the other systematic uncertainties. The signal strength is extracted from the fit in the *signal region*. The *sideband region* does not contain any signal contribution, but is nonetheless used in the fit, to enhance the constraint on the background estimation.

These two strategies should give comparable results in the ideal case, where there are abundant statistics in both data and simulation, and where the data are well described by simulation. However these conditions are usually not met in real analyses, and one strategy becomes preferable. In scenarios where simulations do not model data very well, or where the uncertainties from simulations are not much smaller than the statistical uncertainty in data, it is more advantageous to follow the data-driven approach. In contrast, if an analysis lacks enough statistics in data but has abundant simulations that model data

well, it is more beneficial to perform a MC-based analysis.

The ggH category contains the majority of  $H \rightarrow \mu\mu$  events with a very low  $S/B$ . The statistical uncertainty of data is smaller than the systematic uncertainties of the background prediction from simulations. Therefore it takes the data-driven strategy. The VBF category has a good amount of events, although much less than the ggH category, and a good  $S/B$ . This makes it possible to pick very high  $S/B$  regions with the help of MVA discriminators. The VBF analysis prefers the MC-based strategy as there may be too few events in the high  $S/B$  regions for a data-driven analysis. The VH and  $t\bar{t}H$  categories both have very few events, but high  $S/B$ , which seem like good playgrounds for the MC-based approach. However, the main backgrounds in the VH and  $t\bar{t}H$  categories involve extra lepton(s) from nonprompt sources, which lacks accurate simulation estimates. Moreover, the VH and  $t\bar{t}H$  categories have less sensitivity than the ggH and VBF categories. Adopting the MC-based approach in the VH and  $t\bar{t}H$  categories would take a lot of computation resources and lead to insignificant improvements to the overall result. The data-driven approach is much more cost-effective in the VH and  $t\bar{t}H$  categories. Overall, the ggH, VH, and  $t\bar{t}H$  categories follow the data-driven strategy, while the VBF category takes the MC-based approach.

This thesis is focused on the analysis in the VH category, the procedures of which are detailed in Chapter 6. The summary of all categories is reported in Ref. [36]. Chapter 7 describes the results of the VH analysis and the combined results of all four categories.

## CHAPTER 4

### OBJECT RECONSTRUCTION AND IDENTIFICATION

In CMS, many particles decay very shortly after their production before they reach the detectors. Some particles interact very weakly with detector material and do not leave electronic signals. Particles received by the detector are electrons, muons, photons, and hadrons. These particles leave different signals in multiple detector components, as illustrated in Figure 4-1. As thousands of these particles are produced by  $pp$  collisions in each event, the reconstruction of them requires a holistic processing of the information from all parts of the detector. This is achieved by the particle-flow (PF) algorithm [81], which combines all detector signals per event and finds the optimal identification and reconstruction of these basic particle objects (PF candidates). More complex objects like jets,  $\tau$  leptons, and b hadrons are reconstructed with clusters of PF candidates. Properties of all these objects are calibrated centrally by CMS and provided to the analyzers. The selection of each type of physics object is optimized by the analyzers based on the characteristics of their analyses.

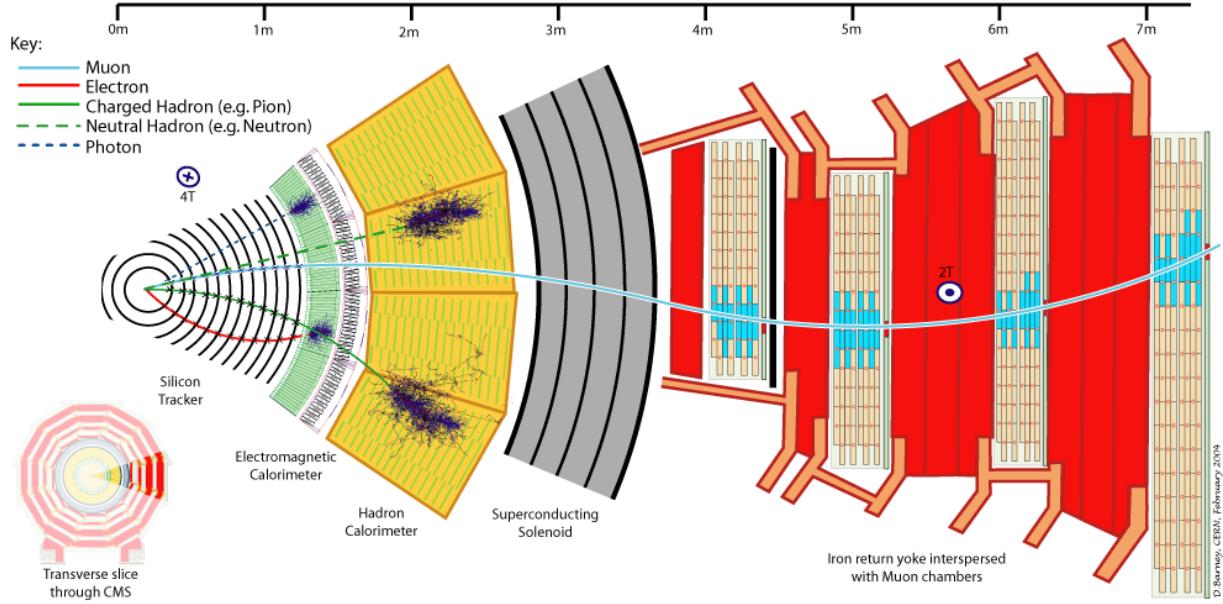


Figure 4-1. A transverse view of the CMS detector layout in the barrel region, with illustrations of the interactions between different particles and different detector components. Plot taken from Ref. [18].

Section 4.1 gives a brief description of the reconstruction sequence for different par-

ticles. Section 4.2 lists the selection criteria for different objects adopted in the  $H \rightarrow \mu\mu$  analysis.

## 4.1 CMS object reconstruction

In  $pp$  collisions in CMS, final-state particles are produced at the beam interaction region. As they travel outward, they first enter the tracker system, in which the trajectories of charged particles (electrons, muons, and charged hadrons) bend in a strong uniform magnetic field and leave hits in the tracker layers. Passing the tracker, electrons and photons are absorbed in the ECAL, which in turn yields the measurements of their position and energy. Hadrons may leave energy deposits in the ECAL as well, but can only be fully absorbed in the HCAL encircling the ECAL. Muons and neutrinos pass the calorimeters and the superconducting magnet with little or no interactions. Muons then bend and leave hits in the muon detectors outside of the magnet, while neutrinos escape all detector layers without leaving any electronic signal. A graphical summary of these interactions is given in Figure 4-1.

The PF algorithm combines all detector signals described above and attains a global description of final-state particles, along with the measurements of their properties. Particles are identified and reconstructed in a sequential manner in PF and are described accordingly in the following sections: tracks of charged particles are first built based on the tracker hits (Section 4.1.1); vertices are reconstructed from groups of compatible tracks (Section 4.1.2); the interaction region is profiled by putting many vertices together (Section 4.1.3); muons and electrons are identified by associating tracks to signals in either the muon chambers or the ECAL (Section 4.1.4 and 4.1.5); hadrons and photons are reconstructed by combining measurements from ECAL, HCAL, with unmatched tracks (Section 4.1.6 and 4.1.7); muons, electrons, hadrons, and photons are all single final-state particles called PF candidates, which can be encapsulated as jets or hadronic decays of  $\tau$  leptons; and finally, the missing transverse momentum ( $E_T^{\text{miss}}$ ) is defined as the negative of the vectorial sum of all PF candidates, indicating the presence of undetected particles

such as neutrinos from the collision (Section 4.1.11).

#### 4.1.1 Tracks

Tracks are the best measured objects in CMS and are the foundation of the reconstruction of various particles. Any charged particle can leave hits in the tracking system and can be reconstructed as a track. Tracks can later be linked to signals in the muon detectors and identified as a muon, or be linked to energy deposits in the ECAL and identified as an electron, or be linked to energy deposits in the HCAL and identified as a charged hadron. An ensemble of tracks can also provide information on the interaction vertices, opening the possibility to identify collision point, converted photons, secondary decays of b quarks and  $\tau$  leptons, and possibly unexpected long-lived particles (LLPs).

In a typical  $pp$  event at CMS during the 2016-2018 data-taking period, where the center-of-mass energy is 13 TeV and the pileup averages about 34, order of a thousand tracks are produced by the collisions, leaving numerous hits in the tracking layers. In order to correctly sew these hits into tracks, a track finder based on a combinatorial Kalman filter (KF) [82] is applied:

- Initial seeds are generated with a few hits compatible with a charged particle trajectory.
- For each seed trajectory, the next layer is surveyed for hits compatible with the seed. A new trajectory candidate is generated for each possible new hit association, with the trajectory quality updated based on the compatibility.
- This pattern recognition is repeated until it reaches the outermost layer. The total number of trajectory candidates is truncated at each layer to avoid an exponential increase.
- The final trajectory candidates are cleaned for duplications, and are evaluated for their properties.

The KF track finder is applied in several successive iterations [83], each targeting a different type of track: first prompt high  $p_T$  tracks, then prompt low  $p_T$  tracks, followed by displaced tracks, and finally tracks with significant detector inefficiencies. After each iteration, all hits associated with the selected tracks are removed from the consideration

of the remaining iterations. In this way, the tracking efficiency is maximized while the misreconstruction rate is kept as low as possible.

Tracks identified by the finder algorithm are refit with a Kalman filter and smoother to evaluate their properties: transverse momentum, direction, and origin. A Kalman filter starts with the innermost hits (typically four) on the track, and builds the covariance matrix [84], which is used to propagate hit uncertainties into the uncertainties of global track properties. This filter proceeds forward through the full list of hits, updating the trajectory estimate and the covariance matrix with each hit. At each step, the hit position uncertainty is reevaluated using the current trajectory estimate as well. As a complement, an additional filter is initialized at each hit with the result of the forward filter but works backward using all hits outside of the current hit. The weighted average of the track parameters from the two filters is taken as the estimate of the trajectory at that hit. This practice is called a smoothing procedure, which ensures the optimal estimate at any hit on the track including, in particular, the innermost and outermost hit. The track is extrapolated to the interaction region and to the calorimeters and muon chambers from the closest hit (the innermost or outermost hit), which are especially important for vertex reconstruction and for track matching to signals in other detectors, respectively.

#### 4.1.2 Primary vertices

In  $pp$  collisions at the LHC, multiple  $pp$  interactions can happen in the same bunch crossing, most of them producing tracks. All the  $pp$  interactions happen in the beam interaction region (called the beamspot) which is focused around the beam axis but spreads along it. It follows a 3D-normal distribution with a standard deviation of order  $10 \mu\text{m}$  in  $x$  and  $y$  directions, and of a few centimeters in  $z$  direction. In order to separate different tracks belonging to different interactions, tracks are extrapolated to the beam axis to find their origins (vertices). The vertex reconstruction is performed with the following procedure [83]:

- Tracks are first selected by the requirement on their compatibility with the center of the beamspot, along with some track quality criteria. This selection rules out tracks

from secondary decays and tracks with low quality, ensuring a high reconstruction efficiency.

- Tracks are then clustered based on their z-coordinates at their point of closest approach to the center of the beamspot. The clustering algorithm starts with a large series of hypothetical vertices along the beam axis, and optimize the global assignment based on the likelihood of associations between each track and each hypothetical vertex. After the optimization, a few ensembles of hypothetical vertices emerge, each considered as a vertex candidate.
- The position of each vertex candidate is fit with the parameters of all the tracks associated to it. The fit is performed adaptively with a weight assigned to each track reflecting the likelihood that it genuinely belongs to the vertex. The weights are updated in each iteration until the sum of weights is maximized.

All the resulting vertices along the beam axis are called the primary vertices (PVs).

The PVs are ordered by the quadratic sum of the  $p_T$  of their tracks,  $\sum p_T^2$ , and the PV with the highest  $\sum p_T^2$  is considered as the hard-scatter vertex, while the other vertices are considered as pileup vertices. In most physics analyses, the hard scatter vertex is the interaction of interest, and is therefore sometimes referred to as the single primary vertex.

#### 4.1.3 Beamspot

The beamspot refers to the 3D-region where  $p_T$  collisions happen, and is determined from the distribution of primary vertices from many events. The position and size of the beamspot are evaluated as the x, y, z coordinates of the beamspot center and their corresponding standard deviations. These values are determined per luminosity section, which is a period of 23 seconds of event collection. If no significant shift of the beamspot center is observed in multiple consecutive luminosity sections, their beamspot position values are merged to extract a final estimate. The beamspot size, however, grows with time as the beams gradually lose focus. Therefore the beamspot size values are kept per luminosity section instead of merging multiple ones.

#### 4.1.4 Muons

The muon detectors in CMS allow muons to be identified with high purity, which is guaranteed by the upstream calorimeters absorbing other particles (except neutrinos). The

muon reconstruction combines information from both the tracker and muon chambers, building three different reconstructed muon types [11]:

- *Standalone muon.* Hits in muon chambers are clustered into track segments with a KF track finding procedure similar to the one applied in track reconstruction in Section 4.1.1. Track segments are built with DT or CSC hits as seeds and are grown into trajectories containing DT, CSC, and RPC hits. The resulting tracks are called the standalone muon tracks.
- *Tracker muon.* Each track built with tracker hits in Section 4.1.1 with a  $p_T$  larger than 0.5 GeV and a total momentum  $p$  larger than 2.5 GeV is extrapolated to the muon system. A comparison is made between the spatial coordinates of the extrapolated tracker track and the nearby DT or CSC muon segments. If the extrapolated track matches at least one muon segment, the tracker track is considered as a tracker muon track.
- *Global muon.* A global muon is built starting with a standalone muon and matching it to a tracker track. The matching is performed by comparing KF parameters of the two tracks. A combined fit is performed using information from both the tracker track and the standalone muon track to determine the final parameters of the global muon track.

Most tracker muons are also global muons, as most muons can be successfully reconstructed as standalone muons. The track muons are more efficient in identifying low  $p_T$  muons, which may only leave hits in the innermost muon station but are not energetic enough to reach the others, failing to be reconstructed as a global muon. On the other hand, hadron shower remnants can sometimes reach the innermost muon station, known as the punch-through, which could fake a tracker muon but not a global one.

The momentum of each global muon is measured fitting all tracker hits plus either zero or one, or multiple hits in muon detectors, depending on their compatibility with the extrapolated tracker track. Since the tracker hits have better spatial resolution than the muon detector hits, the contribution from muon detectors is marginal for muons with  $p_T < 200$  GeV. On the other hand, for muons with  $p_T > 200$  GeV, the information from muon detectors improves the measurement significantly, because the tracks of these high  $p_T$  muons are very straight, and muon detector hits, being far from the tracker hits, provides essential information for the track curvature evaluation.

Isolation is another important property of muon tracks. It helps to distinguish between muons from energetic prompt interactions (prompt muons) and the ones from weak decays within jets (nonprompt muons). It is defined as the ratio of the sum of energy in a geometric cone around the muon to the muon  $p_T$ ,  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ . The summation considers all charged hadrons and neutral particles in PF [81],

$$I_{\text{PF}} = \sum_{h^\pm, \text{HS}} p_T^{h^\pm} + \max(0, \sum_{h^0} p_T^{h^0} + \sum_\gamma p_T^\gamma - \Delta\beta \sum_{h^\pm, \text{PU}} p_T^{h^\pm}) \quad (4-1)$$

where HS means hard scatter and PU means pileup.  $h^0$  and  $h^\pm$  in the equation are neutral and charged hadrons. This calculation aims to mitigate the contamination from pileup and only focus on the hard scatter. The factor  $\Delta\beta$  is set to be 0.5, which corresponds approximately to the ratio of neutral particle to charged hadron production in inelastic  $pp$  collisions, as an estimate to remove the particle energy from pileup interactions.

In addition to the kinematic variables, a set of variables is defined to reflect the quality of the muon reconstruction, such as the  $\chi^2$  of the track fit, the number of hits per track (either in the tracker or the muon detectors, or both), and for global muons specifically, the degree of matching between the tracker track and the standalone muon track. There are other variables evaluating whether the muon track is compatible with the primary vertex, and whether a kink can be found in the muon track. Based on these variables, CMS offers several sets of official muon identification criteria for physics analyses. The specific muon selection criteria adopted in the  $H \rightarrow \mu\mu$  analysis is detailed in Section 4.2.1.

#### 4.1.5 Electrons

Electrons produce hits in the tracker layers and are eventually absorbed by the ECAL. Both the tracker and ECAL provide information on electron identification but both suffer from significant inefficiencies: most electrons emit bremsstrahlung photons as they travel through the tracker layers and lose a sizeable fraction of their energy, which degrades the quality of the reconstructed tracks and biases the momentum measurements; while most of the electron energy including the bremsstrahlung photons is captured by a series of

neighboring ECAL units, this energy deposit is very often overlapped with deposits from other photons and hadrons and is very hard to disentangle. The electron identification relies on the combination of tracker and ECAL information, starting with two types of seed [85]:

- *ECAL-based seed.* Energy deposits in several ECAL channels are grouped into clusters. If the energy deposit in a cluster exceeds 1 GeV, it is considered as a potential incident particle and the cluster is called a seed cluster. Since a final-state electron may reach the ECAL as several electrons and/or photons because of bremsstrahlung radiation and photon conversion, which scatter in multiple clusters tangent to the electron track, clusters near the seed cluster are further grouped into a supercluster (SC). A SC is defined with a small window in  $\eta$  and an extended window in  $\phi$  in order to follow the azimuthal bending of electron tracks in the magnetic field. If a SC can be loosely matched to a track seed (as described in Section 4.1.1), and if there is no significant energy deposit in the HCAL towers behind the SC, this SC along with the track seed(s) linked to it are selected as an ECAL-based electron seed.
- *Track-based seed.* An electron seed can also be built from a generic track if its  $p_T$  exceeds 2 GeV. When the energy radiated by the electron is small, the reconstructed track has good quality and can be extrapolated to the ECAL surface. If the ratio of the closest ECAL cluster energy to the track momentum is compatible with unity, The track along with the ECAL cluster is selected as a track-based seed. For electrons that lose considerable energy via photon emission, their tracks may have kinks and generally have unideal qualities. These tracks are selected with some loose quality criteria and refit with a Gaussian-sum filter (GSF) [86], which is more CPU intensive than the KF used for general tracking, but more adapted to sudden and substantial energy losses along the trajectory. The GSF refit track is extrapolated to the ECAL surface, and a multivariate discriminator is evaluated combining the track and ECAL information. The track + ECAL seeds are selected based on the discriminator output.

Electron seeds obtained from these two approaches are merged and submitted to a more exhaustive GSF refit. In addition, a survey is performed among generic tracks near the seeds, to pick up any track pair with a common displaced vertex. Such track pairs are likely to originate from photon conversions (more details in Section 4.1.6). Conversion tracks may lead to a refinement of the SC shape and an update (or removal) of the GSF tracks. An electron is reconstructed only if the GSF track can be built, and if the HCAL energy deposit behind the ECAL SC does not exceed 10% of the SC energy. Energy collected by the ECAL SC is subject to losses for several reasons: material in the tracker, electromag-

netic shower leakages, and intermodule gaps in ECAL. The total energy in the ECAL SC is corrected with a multivariate regression algorithm to account for these energy losses [87]. The final energy assignment for electrons is based on a combination of the corrected SC energy and the momentum of the GSF track.

The electron identification is based on several aspects of the reconstruction. The isolation of electrons follows the definition in Equation 4-1 and is the main handle to distinguish prompt electrons from nonprompt ones. The electromagnetic shower shape also helps to reject fake electrons from hadrons and is described by a few variables: the ratio between the SC energy and the HCAL energy behind it, the variable  $\sigma_{i\eta i\eta}$  indicating the second moment of the ECAL energy in a  $5 \times 5$  crystal array, and the variable  $R_9$  defined as the most energetic  $3 \times 3$  crystal array divided by the total SC energy. A few other variables evaluate the compatibility between the ECAL SC and the track: the distance in  $\eta$  between the seed cluster and the track, the distance in  $\phi$  between the energy-weighted SC position and the track, the difference between the inverse of the SC energy and the inverse of the track momentum. Furthermore, some other variables are also included accounting for how many missing hits are in the track and whether the electron is likely to come from the conversion of a photon.

CMS offers several sets of official electron identification criteria based on these variables, either in a cut-based fashion or as a multivariate discriminator. The specific electron selection criteria adopted in the  $H \rightarrow \mu\mu$  analysis is detailed in Section 4.2.2.

#### 4.1.6 Photons

Photons can be produced by various processes in CMS: from prompt interactions, from bremsstrahlung emission of electrons, and from secondary decay of hadrons (mostly  $\pi^0$ ). Although photons themselves do not produce hits in trackers, they have a significant probability to convert into an electron-positron pair in the tracker material. Photons, if not converted, are all fully absorbed by the ECAL, whose signals can very often be mixed with the signals from nearby electrons and hadrons if there are any. As a result, photon iden-

tification is entangled with the reconstruction of electrons and hadrons. In CMS, photons can be defined in three cases at different stages of the PF reconstruction: the converted photons based on displaced track pairs and their associated ECAL clusters, the isolated unconverted photons based on ECAL deposits with little HCAL energy deposit behind it, and the nonisolated unconverted photons based on ECAL deposits with significant HCAL energy deposit behind it.

The reconstruction of converted photons involves tracks and ECAL clusters, and is performed along with the reconstruction of electrons and isolated unconverted photons. After the electron seeding stage, a conversion finding algorithm [88] examines all tracks near the ECAL SC and looks for tracks significantly displaced from the primary vertex. If displaced tracks of different charges are found, they are paired with some proximity requirement and fit to a common vertex with a kinematic vertex fit. These track-pair candidates are defined as converted photons if they satisfy thresholds on the quality of kinematic fit, on the total  $p_T$ , and on the compatibility between the converted tracks and the associated ECAL clusters.

The reconstruction of isolated photons shares the same procedure as the electron reconstruction described in Section 4.1.5, and is only different at the last stage: an electron is defined if a GSF track can be built from the seed, while a photon is defined if no GSF track can be built associated to the ECAL SC and if the  $E_T$  is greater than 10 GeV. These photon candidates from ECAL seeding are further retained if they are isolated from other tracks and ECAL clusters and if there are no significant HCAL energy deposits around. The corrected ECAL SC energy is used for the final photon energy assignment.

More often, photons in CMS are produced by decays of hadrons and are not isolated from other calorimeter deposits. The reconstruction of nonisolated photons is performed along with the reconstruction of hadrons, which is detailed in Section 4.1.7. Once muons, electrons, isolated photons, and charged hadrons are reconstructed by PF, the tracks and calorimeter deposits associated to them are masked, leaving only unassigned ECAL and

HCAL clusters. In general, photons carry about 25% of the total jet energy, all absorbed by the ECAL, while neutral hadrons leave only 3% of the total jet energy in the ECAL. Therefore, within the tracker acceptance ( $|\eta| < 2.5$ ), as charged hadrons are already identified, all the unassigned ECAL deposits are considered as photons and all remaining HCAL deposits are considered as neutral hadrons. However, out of the tracker acceptance, charged hadrons cannot be distinguished from neutral ones. Charged and neutral hadrons together leaves about 25% of the total jet energy in the ECAL, which is at the same level of photon energy. ECAL deposits can no longer be assumed as photons. In this case, hadrons are reconstructed with HCAL clusters, and a fraction of energy is removed from the associated ECAL clusters. The remaining net ECAL clusters are reconstructed as photons. The energy estimate of these nonisolated photons takes the uncorrected ECAL cluster energy.

The identification of photons is based on the isolation and electromagnetic shower shape variables as described for electron identification (Section 4.1.5). CMS offers identification criteria either as cut-based selections or a multivariate discriminator. These identifications target only isolated photons, while the nonisolated ones are encapsulated in jets (Section 4.1.7). Photons are not used in the  $H \rightarrow \mu\mu$  analysis as primary physics objects, but are only used for the recovery of the energy losses in final-state radiation (FSR). The selection for FSR photons is described in Section 5.2.

#### 4.1.7 Hadrons

Hadrons are produced in plenty by  $pp$  collisions in CMS. Charged hadrons leave hits in the tracker and all hadrons are absorbed by the calorimeters. As described in Section 4.1.8, hadrons are initiated by gluons or quarks in QCD processes, and are always produced as cascades, known as jets. As a result, their detector signals, especially in calorimeters, always overlap with one another. As the detectors only provide information on the trajectory (for charged hadrons only) and energy of the hadrons, CMS reconstruct hadrons as charged and neutral ones without further distinguishment of their exact composition.

Energy deposits in HCAL are grouped into clusters in a similar fashion as that in the ECAL seeding. Cluster seeds are first identified as HCAL cells that have an energy over a threshold and larger than their neighboring cells. Seeds are grown into topological clusters by iteratively including adjacent cells with an energy above another threshold. A topological cluster may cover many seeds if they are connected by cells with significant energy deposits. Each topological cluster is fit with a Gaussian-mixture model, in which the energy distribution in the cells is assumed to be the sum of  $N$  Gaussian energy distributions, where  $N$  is the number of seeds in the topological cluster. Each resulting Gaussian component is considered a a cluster.

Within the tracker acceptance, HCAL clusters linked to tracks are reconstructed as charged hadrons. Nonisolated photons and neutral hadrons are reconstructed from the remaining ECAL and HCAL clusters respectively. Beyond the tracker acceptance, hadrons are reconstructed from the HCAL clusters without differentiating their charges, and photons are reconstructed with the remaining ECAL clusters.

All hadrons (and nonisolated photons) are considered as PF candidates but are not used as individual physics objects for analyses. PF candidates are clustered into more complex objects like jets and the hadronic decays of  $\tau$  leptons, which are used in physics analyses indicating outgoing partons or  $\tau$  leptons from the collisions. Reconstruction of jets and hadronic  $\tau$  decays are detailed in Section 4.1.8 and 4.1.10.

#### 4.1.8 Jets

All partons carry color charges. They cannot exist as free particles because QCD forbids isolated color charges, known as color confinement. When partons are emitted from collisions, the strong interaction potential grows rapidly between the color charges as they move away from each other. Once the potential reaches a threshold known as Hagedorn temperature, corresponding to an energy of roughly 150 MeV, the excessive energy is converted into the production of a quark-antiquark pair. If the original outgoing partons are energetic enough, this process is repeated, leading to a cascade of quarks and

gluons. These quarks and gluons bind with each other and form hadrons, which are color-neutral, known as the hadronization process. Some of the hadrons have short lifetime and decay into photons, leptons, or lighter hadrons. As a result, a parton produced in collisions is always detected as a cluster of mixed particles that traverse with roughly collinear directions, known as a jet. In  $pp$  collisions at the LHC, on average, 65% of the jet energy is carried by charged hadrons, 25% by photons, and 10% by neutral hadrons.

In particle-flow, jets are reconstructed with the anti- $k_T$  algorithm [89, 90], which clusters particles based on a parametrized distance between them. A jet can be reconstructed considering all PF candidates (PF jets), or with PF candidate except charged hadrons from pileup vertices via charged hadron subtraction (CHS). The jet momentum is calculated as the vector sum of all of its constituent PF candidates. In the same fashion, an invariant mass of the jet is also calculated. PF jets are reconstructed down to a  $p_T$  of 15 GeV.

The energies of PF jets are systematically underestimated for several reasons. First of all, very low  $p_T$  charged hadrons loop in the tracker and do not reach the calorimeter. It is usually difficult to cluster these tracks with other particles and many of them end up unassigned. In addition, neutral hadrons with low energy ( $< 10\text{GeV}$ ) may not be successfully identified from the HCAL noise, and are sometimes not reconstructed. Furthermore, neutral hadrons usually also leave energy deposits in the ECAL, which are systematically identified as (nonisolated) photons, following the reconstruction procedure described in Section 4.1.6. These ECAL deposits are calibrated following the photon hypothesis, which is not suitable for hadrons and leads to an systematic underestimate of energy.

To mitigate this systematic bias in the jet energy estimate, a jet energy correction (JEC) procedure [91] is applied, which brings the jet response, defined as the mean ratio between the reconstructed jet energy to the true jet energy, to unity across different  $p_T$  and  $\eta$  regions. The jet energy resolution is mostly determined by the resolution of the HCAL, therefore the relative resolution is better for more energetic jets. After the JEC, in the barrel region, the relative energy resolution is typically 15% for 20 GeV jets, 10% for

100 GeV jets, and 5% for 1 TeV jets.

The  $H \rightarrow \mu\mu$  analysis uses PF+CHS jets, with JEC applied. Jets are particularly important in the VBF category, which is tagged with two energetic forward jets. In the VH category, on the other hand, jets are not used explicitly but only considered for the calculation of the missing transverse energy. The selection criteria of jets are detailed in Section 4.2.3.

#### 4.1.9 B jets

Jets originating from bottom or charm quarks contain heavy-flavor hadrons, and are called heavy-flavor jets. These heavy-flavor hadrons have lifetimes of the order of 1 ps, and their decays give rise to secondary vertices (SV) a few mm away from the corresponding primary vertex. Therefore the heavy-flavor jets can be distinguished from the light-flavor ones (originating from light-flavor quarks or gluons). The identification of jets initiated by b quarks, called b jets, are particularly important to both SM measurements and BSM searches. CMS provides several algorithms to evaluate the probability of a jet to be a b jet [80], among which the DeepCSV tagger is adopted in the  $H \rightarrow \mu\mu$  analysis.

The DeepCSV considers many discriminating variables of a jet: the track multiplicity, the kinematic properties of the tracks in the jet, the energy ratio and spatial separation between different tracks, the position and the invariant mass of secondary vertices, and the presence/absence of semileptonic decays of hadrons in the jet. These variables are processed with a deep neural network, whose outputs are probabilities of the jet to contain one b hadron ( $P(b)$ ), or two b hadrons ( $P(bb)$ ), or one c hadron ( $P(c)$ ), or two c hadrons ( $P(cc)$ ), or none of them ( $P(udsg)$ ). The result  $P(b) + P(bb)$  is used as the figure-of-merit to set standard b-tagging working points in CMS. Three working points are officially provided: the loose, medium, and tight tags, which are chosen with a misidentification rate of 10%, 1%, and 0.1%, and correspond to a tagging efficiency of about 85-90%, 70-75%, and 50-55%, respectively. B tagging relies heavily on the tracking information and is therefore only developed and applied inside the tracker acceptance region  $|\eta| < 2.5$ .

B tagging are used in the  $H \rightarrow \mu\mu$  analysis to identify the b jets from the top quark decays in  $t\bar{t}H$  events, or to veto such decays in other event categories. The exact criteria for b-tagged jet selection in the  $H \rightarrow \mu\mu$  analysis is detailed in Section 4.2.3.

#### 4.1.10 Hadronic tau decays

The  $\tau$  lepton has a mean lifetime of 0.3 ps and decays to either an electron or a muon plus two neutrinos, or a few hadrons plus one neutrino. The reconstruction of leptonic  $\tau$  decays is unrealistic because they are only received by the detector as single electrons or muons, while the neutrinos from the decays cannot be separated from other sources of missing energy. On the other hand, the reconstruction of hadronic  $\tau$  decays, denoted as  $\tau_h$ , can be achieved based on the track multiplicity, collimation, and isolation of the decay products.

The main hadronic  $\tau$  decay modes contain either one or three charged hadrons, from zero to two neutral hadrons ( $\pi_0$ ), and a neutrino. Their detector signature resembles that of low-multiplicity jets. Therefore in PF,  $\tau_h$  candidates are reconstructed starting with PF jets as seeds. The reconstruction algorithm is called the hadrons-plus-strips (HPS) algorithm [92, 93], which picks charged hadrons from the seeding jet, and clusters photon and electron constituents of the seeding jet into "strips" in the  $\eta - \phi$  plane. Each strip is considered as a  $\pi_0 \rightarrow \gamma\gamma$  candidate inside the  $\tau_h$  decay. Different combinations of one or three charged hadron candidates plus up to two strips are tested for several  $\tau_h$  hypotheses:  $\tau^\pm \rightarrow h^\pm h^\mp h^\pm \nu_\tau$ ,  $\tau^\pm \rightarrow h^\pm \pi_0 \pi_0 \nu_\tau$ ,  $\tau^\pm \rightarrow h^\pm \pi_0 \nu_\tau$ , and  $\tau^\pm \rightarrow h^\pm \nu_\tau$ . Hypotheses are rejected if they fail requirements on the invariant mass, the sum of charge, the isolation, and the track vertexing in cases of multiple tracks. If multiple combinations inside a jet are compatible with the  $\tau_h$  hypotheses, the one with the largest  $p_T$  is retained and all others are discarded.

The identification of  $\tau_h$  candidates is performed separately against jets, electrons, and muons. The ID discriminants can either be cut-based or MVA-based, on various kinematic properties of the  $\tau_h$  candidates. Several working points are provided for each discriminant, whose efficiencies range from 30% to 70% (vs jets), from 70% to 90% (vs electrons), and

from 98% to 99% (vs muons), all with a misidentification rate below 1%.

Hadronic  $\tau$  candidates are not considered in the  $H \rightarrow \mu\mu$  analysis. In particular, the hadronic *tau* leptons can be used to tag W and Z decays in VH signal events, But it adds negligible statistical significance to the analysis and is therefore not included. More explanation are given in Chapter 6.

#### 4.1.11 Missing transverse momentum

Neutral particles produced in  $pp$  collisions that interact weakly with regular material can traverse the CMS detector undetected. There is no way to determine their number, type, or exact direction. However, when these particles are produced along with detectable particles: muons, electrons, photons, hadrons etc., their presence can be inferred from the detected particles as an imbalance in the total momentum perpendicular to the beam axis. This imbalance is referred to as the missing transverse momentum ( $\vec{p}_T^{\text{miss}}$ ) or the missing transverse energy ( $E_T^{\text{miss}}$ ). In practice, no distinction is made between  $E_T^{\text{miss}}$  and the magnitude of the missing transverse momentum ( $p_T^{\text{miss}}$ ) as there is no way to infer the invariant mass of the missing particle(s).

$E_T^{\text{miss}}$  is crucial to both SM measurements and searches for new weakly-interactive particles beyond the SM. CMS offers two ways to reconstruct the  $E_T^{\text{miss}}$  [94]. The PF  $E_T^{\text{miss}}$  [95] is defined as the negative of the vector sum of the  $p_T$  of all PF candidates in the event. The other reconstruction method is called the "pileup per particle identification" (PUPPI) [96], which sums all PF candidates while rescaling the energy of all hadrons based on their likelihood of originating from the PV. The PF  $E_T^{\text{miss}}$  is used in the  $H \rightarrow \mu\mu$  analysis.

In addition, another missing energy estimate based on high level physics objects,  $H_T^{\text{miss}}$ , is included in many analysis. It sums only identified leptons and jets, instead of all PF candidates, in order to focus on the expected products of the prompt interaction. In the  $H \rightarrow \mu\mu$  analysis, the  $H_T^{\text{miss}}$  definition only considers leptons and jets that pass their selection criteria specific in 4.2.

By its nature, the resolution of  $E_T^{\text{miss}}$  (or  $H_T^{\text{miss}}$ ) is much worse than other physics

objects. A mismeasurement on  $E_T^{\text{miss}}$  can originate from proton debris falling out of the detector acceptance, misreconstruction of objects and misassociation between pileup and primary vertices, and in general experimental resolution in all detector components. Particular to the  $H \rightarrow \mu\mu$  analysis,  $E_T^{\text{miss}}$  can potentially be used to tag the  $Z \rightarrow vv$  decay in  $ZH$  signal events. This selection is very inefficient because of the  $E_T^{\text{miss}}$  resolution and is not considered in the  $H \rightarrow \mu\mu$  analysis.

## 4.2 Object selection in the $H$ to dimuon analysis

### 4.2.1 Muon selection

CMS provides official muon identification (ID) criteria based on several kinematic variables: the number of hits in the muon track; the fit quality of the muon track; the compatibility between the tracker track and the standalone muon for global muons; and the compatibility between the muon track and the primary vertex. The global track fit  $\chi^2$  and a kink-finder  $\chi^2$  are used as indicators of the fit quality of the global muon track. The compatibility between the tracker track and the standalone muon is evaluated with the  $\chi^2$  of the position match, and a variable called the segment compatibility. The compatibility between the track and the primary vertex is evaluated with their impact parameters (as defined in Section 5.3.1) along with a variable (called the SIP) reflecting the significance of the impact parameters relative to its uncertainty.

The official muon identification is provided for different types of muons [11]: ID for generic muons, dedicated ID for low- $p_T$  muons ( $p_T < 20$  GeV), and dedicated ID for high- $p_T$  muons ( $p_T > 200$  GeV). The muons of moderate- $p_T$  are used in most studies on electroweak and Higgs physics, and can be selected with three levels (known as working points) of generic identifications:

- *Loose muon ID* selects PF muons that is either a tracker or a global muon without further requirements.
- *Medium muon ID* poses requirements on the number of hits in the tracker track and on the track fit quality, on top of the loose ID. The number of tracker hits must be more than 80% of the number of tracker layers the muon traverses. The global fit  $\chi^2$  must be less than 3, the kink-finder  $\chi^2$  must be less than 20, and the position

match  $\chi^2$  must be less than 12. The segment compatibility is required to be greater than 0.303 for global muons, or greater than 0.451 for tracker-only muon. It rejects badly-reconstructed muons while keeping a high efficiency for the well-reconstructed ones.

- *Tight muon ID* suppresses muons from decay in flight and from hadronic punch-through. The muon must be a loose muon, as well as a global muon with its track fit  $\chi^2 < 10$ . The muon track must include at least one hit in the muon chamber and at least six layers of the inner tracker, at least one of them been pixel hits. It must also be compatible with the primary vertex, with  $d_{xy}^{PV} < 0.2\text{ cm}$ ,  $d_z^{PV} < 0.5\text{ cm}$ .

Several levels of selection criteria on the PF isolation is also centrally provided, in which the Loose Isolation requires the PF isolation of the muon in a cone of  $\Delta R < 0.4$  to be less than 25% of the muon  $p_T$ . It cleans muons from hadronic activities and keeps a selection efficiency of about 99% regarding the Medium ID selection.

The  $H \rightarrow \mu\mu$  analysis adopts the Medium muon ID and Loose Isolation as a baseline selection. In addition, muons are required to have  $p_T > 20\text{ GeV}$  and  $|\eta| < 2.4$ , and should be compatible with the primary vertex with impact parameters  $d_{xy}^{PV} < 0.05\text{ cm}$ ,  $d_z^{PV} < 0.1\text{ cm}$  and the SIP  $< 8.0$ .

Finally, to further reject nonprompt muons, a multivariate identification method, called the LeptonMVA, is applied. Several analyses in CMS have used the LeptonMVA approach, and the version adopted by the  $H \rightarrow \mu\mu$  analysis is developed in the context of the search for tZq production [97]. This LeptonMVA combines the information of the muon isolation, the vertex compatibility, and the relative position and relative energy between the muon and its closest jet. The selection requirement on the LeptonMVA is chosen to be  $\text{LeptonMVA} > 0.4$ , which corresponds to an efficiency of about 95% and a fake rate of about 3-4%.

#### 4.2.2 Electron selection

As described in Section 4.1.5, the electron identification in CMS is based on the electron track quality and the properties of the ECAL supercluster:  $\sigma_{i\eta i\eta}$ ,  $|\Delta\eta_{in}^{seed}|$ ,  $|\Delta\phi_{in}|$ ,  $H/E$ ,  $|1/E - 1/p|$ , the number of missing hits, and the photon-conversion indicator. The official identification criteria is provided either as a series of selection criteria based on

these variables, or as a multivariate discriminator summarizing these variables [85].

The cut-based ID includes four standard working points: Veto ID is used for vetoing electrons, corresponding to about 95% efficiency; Loose ID is tolerant for fake electrons, with an efficiency of about 90% on real electrons; Medium ID is balanced between the fake rate and the signal efficiency, which is about 80%; and the Tight ID is about 70% efficient for real electrons but also has a low fake rate. The multivariate discriminator is developed combining the aforementioned variables, plus the PF isolation of the electron, the fraction of the track momentum at the outermost tracker layer relative to that at the innermost tracker layer, and a variable evaluating the track-cluster match. Two working points are provided for the MVA-based ID, corresponds to 80% and 90% efficiencies.

The  $H \rightarrow \mu\mu$  analysis adopts the 90%-efficiency working point of the MVA-based ID as a baseline selection. It also requires electrons to have  $p_T > 20\text{GeV}$  and  $|\eta| < 2.5$ , and to be compatible with the primary vertex with impact parameters  $d_{xy}^{PV} < 0.05\text{cm}$ ,  $d_z^{PV} < 0.1\text{cm}$  and the  $\text{SIP} < 8.0$ . The electron must also have less than two missing hits in its track, and must not be linked to a photon conversion, these two requirements being used in the cut-based ID but not included in the MVA-ID. The ECAL is less efficient in the region of  $1.444 < \eta < 1.566$ , where there are gaps between ECAL modules. Electrons in this region are not considered for the analysis.

Finally, the LeptonMVA discriminator, similar to that for muons, is also developed for electrons in Ref. [97]. The selection requirement of  $\text{LeptonMVA} > 0.4$  is also applied to electrons, corresponding to an efficiency of about 93% and a fake rate of about 4%.

#### 4.2.3 Jet selection

CMS provides two types of identification criteria on jets [98]: the jet ID, which distinguishes real jets from fake ones originating from detector noise, and the pileup jet ID, which tells jets originating from the primary interaction apart from jets produced in pileup interactions. The jet ID is determined with a series of selection criteria on the jet constituent variables, which quantifies the number of neutral hadrons, charged hadrons,

photon, and leptons in the jet, as well as the fractions of jet energy carried by them. There is only one working point for the jet ID, corresponding to an efficiency of 98-99%, referred to as the standard jet ID, or historically as the tight jet ID. The pileup jet ID is an MVA-based discriminator summarizing the jet properties such as the number of objects in the jet, the vertexing of tracks in the jet (if inside the tracker coverage), and the shape of the jet cluster. The pileup ID is only trained with (and applied to) jets with  $p_T < 50\text{GeV}$ , where it is most likely for pileup jets to appear. Three working points are offered for the pileup ID, loose, medium, and tight, which correspond to 99(95)% , 90(90)% , and 80(80)% efficiencies in  $|\eta| < 2.5$  ( $|\eta| > 2.5$ ) regions, respectively.

The  $H \rightarrow \mu\mu$  analysis considers PF+CHS jets with  $p_T > 25\text{GeV}$  and  $|\eta| < 4.7$  after the JEC. Jets are required to pass the jet ID and the loose pileup ID in general. To mitigate the exceptional ECAL endcap noise in 2017, jets in 2017 datasets are specifically required to pass the tight pileup ID if they lie in  $2.6 < |\eta| < 3.0$  region. Finally, jets are rejected if they are found near a selected muon or electron with the geometrical separation of  $\Delta R < 0.4$ .

B-tagged jets are also considered in the  $H \rightarrow \mu\mu$  analysis, mainly for the determination of the  $t\bar{t}H$  event category. B-tagged jets should meet the selection requirements on generic jets, have  $|\eta| < 2.5$ , and are tagged with the DeepCSV tagger described in Section 4.1.9. Events containing at least one medium b-tagged jet or at least two loose b-tagged jets are classified into the  $t\bar{t}H$  category.

## CHAPTER 5

### MUON MOMENTUM CORRECTION AND CALIBRATION

This analysis aims to find a sharp signal peak on top of a smooth background in the  $m_{\mu\mu}$  distribution. It is of crucial importance to correct any mismeasurement in the muon momentum scale and to remove any momentum dependence on variables like  $\eta$  and  $\phi$  of muons. It is also crucial to minimize the momentum and resolution differences between data and simulation, so that there is no significant bias in the signal modeling.

Three sets of corrections are applied in this analysis: the *Rochester correction* [99], the recovery of the final-state radiation (FSR) photons, and the *GeoFit correction*. The *Rochester correction* is a centrally provided correction (by CMS), which corrects the biases in the muon momentum resulted from the mismodeling of detector alignment and magnetic field. A brief description of the *Rochester correction* is given in Section 5.1, while the technical details can be found in the Ref. [99]. The *FSR recovery* is a common practice in many CMS analyses, which corrects the muon energy loss via FSR. The recovery scheme in this analysis is optimized specifically for the  $H \rightarrow \mu\mu$  decay, which is described in Section 5.2 and in more detail in Ref. [19]. The *GeoFit correction* is developed by the author in the context of the  $H \rightarrow \mu\mu$  analysis and approved by the CMS collaboration. It uses information of muon vertexing to correct the biases in muon momentum of the reconstructed muon tracks. The development of the *GeoFit correction* is described in details in Section 5.3. The effects of the three corrections are orthogonal. In practice, the *Rochester correction* is applied to all muons, then each muon is surveyed for FSR photons. If an FSR photon is found associated to the muon, the *FSR recovery* is applied, if not, the *GeoFit correction* is applied.<sup>1</sup>

These three corrections are applied to both data and simulation. The performance of the corrections is examined with the study on the  $Z \rightarrow \mu\mu$  peak, which is listed in details in Section 5.4. These corrections fix all the known biases in muon measurement, and ensure a per-mille-level agreement between data and simulation.

---

<sup>1</sup>The performance of *GeoFit correction* on muons with FSR is not validated and maybe suboptimal. Furthermore, as FSR muons are a small fraction, the difference whether to apply *GeoFit correction* to them is negligible.

## 5.1 Rochester correction

In reality, the CMS detector can have various imperfections, such as the misalignment of the detector components, and the uncertainties in the magnetic field. Sometimes these imperfections are not correctly emulated in reconstruction softwares, and as a result, the reconstructed muons can be inaccurate. These inaccuracies are reflected as the dependences of the muon momentum on its  $\eta$ ,  $\phi$  coordinates, and its charge.

On the other hand, in the simulation of CMS events, none of the imperfections are assumed, which leads to slightly different detector responses from those in data, and in turn over-optimistic modelings in the muon measurement. Therefore the correction for simulation and for data have to be different. The generated muon information, smeared with some functional forms to match the experimental resolution, is used as the reference for both data and simulation.

The well-understood  $Z \rightarrow \mu\mu$  events are used to develop the *Rochester correction*. The idea of the correction is briefly summarized as follows:

- For data, reconstructed simulation (reco-sim), and the reference simulation (ref-sim), muons are divided into different  $\eta$  and  $\phi$  bins, separately for  $\mu^+$  and  $\mu^-$ . In each bin, the  $1/p_T$  distributions of data and reco-sim are corrected so that the mean value of the distribution becomes the same as that in the ref-sim.
- The  $1/p_T$  distribution in reco-sim is usually narrower than that in data. A smearing is applied to the reco-sim  $1/p_T$  distribution so that it matches the resolution in data.
- After the steps above, the  $m_{\mu\mu}$  in each bin may still be off from the expected distribution by some small amounts. The ratio between this offset and the nominal  $Z$  mass is applied to the muon  $p_T$  as a correction factor iteratively, until the offset is minimized.

The *Rochester correction* removes the dependences of  $m_{\mu\mu}$  on muon  $\eta$ ,  $\phi$ , and charge, as well as the  $m_{\mu\mu}$  resolution differences between data and the simulation. Details of the performance of the *Rochester correction* can be found in Section 5.4.

## 5.2 FSR recovery

In CMS, muons produced in  $pp$  collisions may radiate photons and lose energy, which is referred to as the final-state radiation (FSR). The radiation may carry substantial energy

and lead to an underestimation of the original muon momentum. This leads to two effects in this analysis: a loss of event acceptance, and a smearing of the  $m_{\mu\mu}$  resolution. This can be mitigated by identifying some of the FSR photons adding their energy back to the muon energy, called the *FSR recovery*.

The selection for the FSR photons is modified on top of the strategy developed in the CMS H → ZZ analyses [26, 100]. The selection criteria is summarized as follows:

- Photons with transverse energy  $E_T^\gamma > 2$  GeV and  $|\eta| < 1.4$ ,  $1.6 < |\eta| < 2.4$  are considered as FSR candidates.
- The photon is required to be within the cone of  $\Delta R < 0.5$  around its closest muon that satisfies  $p_T > 20$  GeV and  $|\eta| < 2.4$ .
- The photon is not identified as a bremsstrahlung photon associated with a reconstructed electron.
- The PF isolation of the photon in a cone of  $\Delta R < 0.3$  should be less than 1.8, i.e.  $\sum_i p_T^i(\Delta R(\gamma, i) < 0.3)/p_T(\gamma) < 1.8$ , where  $i$  iterates the PF objects around the photon other than the candidate muon.
- The separation between the photon and the muon satisfies  $\Delta R(\mu, \gamma)/p_T^2(\gamma) < 0.012$ .
- In order to suppress energetic photons from the H → Z $\gamma$  →  $\mu\mu\gamma$  process, the  $p_T$  ratio between the photon and the muon is required to be less than 0.4, i.e.  $p_T(\gamma)/p_T(\mu) < 0.4$ .
- If multiple FSR photons are associated to the same muon, only the photon with the smallest  $\Delta R(\mu, \gamma)/p_T^2(\gamma)$  is taken.

With this set of selection, about 3% of signal events are tagged with FSR photons. The selection is also very effective in reducing the H → Z $\gamma$  events, whose final yield is about 0.1% of the overall H →  $\mu\mu$  signal and can be neglected. The momentum of the FSR photons are added to the muon momentum, while the photons themselves are removed from the calculation of the muon isolation. The *FSR recovery* significantly improves the  $m_{\mu\mu}$  reconstruction in the FSR tagged events, as shown in the left plot of Figure 5-1. The overall effect on the inclusive signal, as shown in the right plot of Figure 5-1, is a 3% improvement on the  $m_{\mu\mu}$  resolution and a 1.7% increase in the total signal yield. The *FSR recovery* is applied in all categories of the H →  $\mu\mu$  analysis and resulting improvement on

the combined significance is about 3%. The performance of *FSR recovery* is also validated with the  $Z \rightarrow \mu\mu$  events, shown in Figure 5-2, where a good agreement is kept between simulation and data with or without the *FSR recovery*. The *FSR recovery* is expected to perform the same way on data as on simulation, and no bias is introduced by the application of the *FSR recovery*.

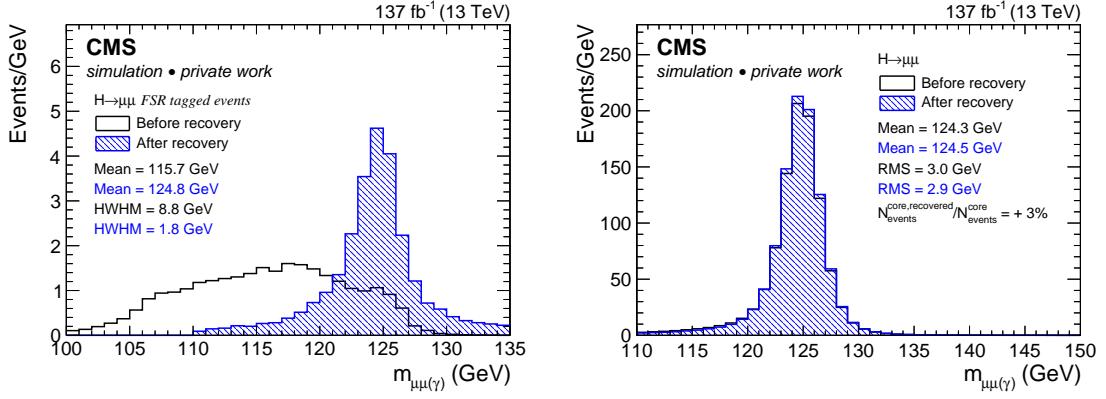


Figure 5-1. Performance of the *FSR recovery* in the simulated  $H \rightarrow \mu\mu$  events. The  $m_{\mu\mu}$  before and after the *FSR recovery* are shown for the events that contain at least one FSR photon (left), and for the inclusive signal events (right). Plot taken from Ref. [19].

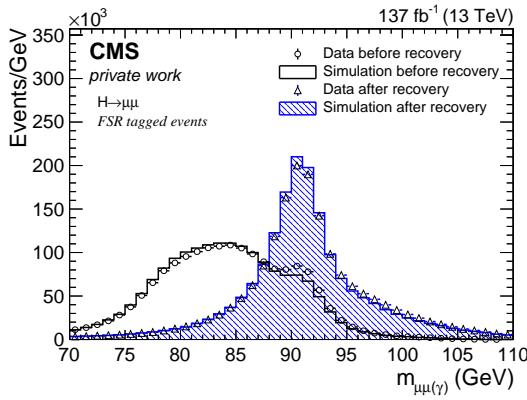


Figure 5-2. Performance of the *FSR recovery* in the  $Z \rightarrow \mu\mu$  events that contain FSR photons, in both data and simulation. A good agreement between simulation and data is observed, both before and after the correction. Plot taken from Ref. [19].

### 5.3 GeoFit correction

As described in Section 4.1.1, tracks are built from hits in silicon trackers and extrapolated to the collision region without any assumption on their vertices. Reconstructed tracks may have nonzero displacements from their true origins because of uncertainties in track fits. If there is a way to locate the true origin of a track, it can be used as a constraint on the track to improve the track momentum measurement. Prompt tracks originate from  $pp$  interaction points, which are estimated by primary vertices (PV, Section 4.1.2) or the beamspot (BS, Section 4.1.3). This section reports the development of a momentum correction for prompt muons (the *GeoFit correction*) by taking the PV or BS as their true origins.

This correction is based on a geometrical correlation between the  $p_T$  mismeasurement and the displacement from the reconstructed track to its true vertex. The correction is applied as a simple analytic function whose parameters are determined from fits to simulations. It is therefore named the *GeoFit correction*. Section 5.3.1 explains the geometry of the track displacement and the correlation between different variables. Section 5.3.2 describes the studies on simulated samples to find the best fit parameters in that correlation. The *GeoFit correction* is developed using muon tracks from the  $Z \rightarrow \mu\mu$  process. It removes the dependence of  $m_{\mu\mu}$  on track displacement, which leads to an improvement on the  $m_{\mu\mu}$  resolution of the combined signal ranging from 3% to 10%, depending on the data-taking period. Details of the *GeoFit correction* performance, along with validation studies are shown in Section 5.3.3. In addition, an alternative way to correct this  $p_T$  bias is to redo the track fit including the colliding vertex as an additional hit in the track, which should achieve a more fundamental correction at the cost of more computational resources. A preliminary study comparing the *GeoFit correction* with the track refit shows the two methods give almost equivalent results, detailed in Section 5.3.4.

### 5.3.1 Geometry of the track displacement

The displacement of a track from a vertex is usually measured as the impact parameters,  $d_{xy}$  and  $d_z$ , which are the signed distances between the vertex and its point of closest approach (PCA) to the track, in the transverse and longitudinal directions. In CMS, because most studies only care about the transverse impact parameter, the PCA is defined as the point on the 2D-projection of the track in the transverse plane that is the closest to the vertex. Note that the PCA of the 2D-track is not necessarily the 2D-projection of the PCA in the 3D-space. The  $d_z$  is calculated at the 3D-point corresponding to the 2D-PCA, rather than the 3D-PCA. As the  $d_z$  is not used in our studies, the term "impact parameter", if not otherwise stated, refers specifically to the transverse impact parameter  $d_{xy}$ , also denoted as  $d_0$ . The definition of  $d_0$  can be expressed as

$$d_0 = -x_0 \cdot \sin(\phi_0) + y_0 \cdot \cos(\phi_0) \quad (5-1)$$

where  $(x_0, y_0)$  is the coordinate of a point near the vertex in the frame, in which the vertex is at  $(0, 0)$ , and  $\phi_0$  is the azimuthal angle of the track at  $(x_0, y_0)$ . A scheme for this definition is shown in Figure 5-3.

In the track geometry, illustrated in Figure 5-4, the reconstructed track is very close to, but slightly deviated from, the true track, which leads to a small  $d_0$  between the reco track and the true vertex, as well as a small distance between the circular centers of the reco track and the true track. The circular centers of the two tracks are labeled as  $O$  and  $O'$  for the true track and the reconstructed track, and  $s$  is the distance between  $O$  and  $O'$ . The radii of the two tracks are  $r$  and  $r'$ , with  $\Delta r = r' - r$ . The two circles must intersect at two points, labeled as point  $M$  and  $N$ , with the distance between  $M$  and  $N$  denoted as  $l$ .  $\beta$  is half of the central angle spanned by the chord  $l$  in the true track, while  $\alpha$  is the angle  $\angle O'MO$ . The true vertex is denoted as  $V$ , with  $d_0$  as the impact parameter of the reco track to it, while the PCA on the reco track is denoted as  $P$ . The distance between  $M$  and  $V$  is marked as  $x$ . As this scheme represents typical muon tracks in CMS, the radii of the tracks

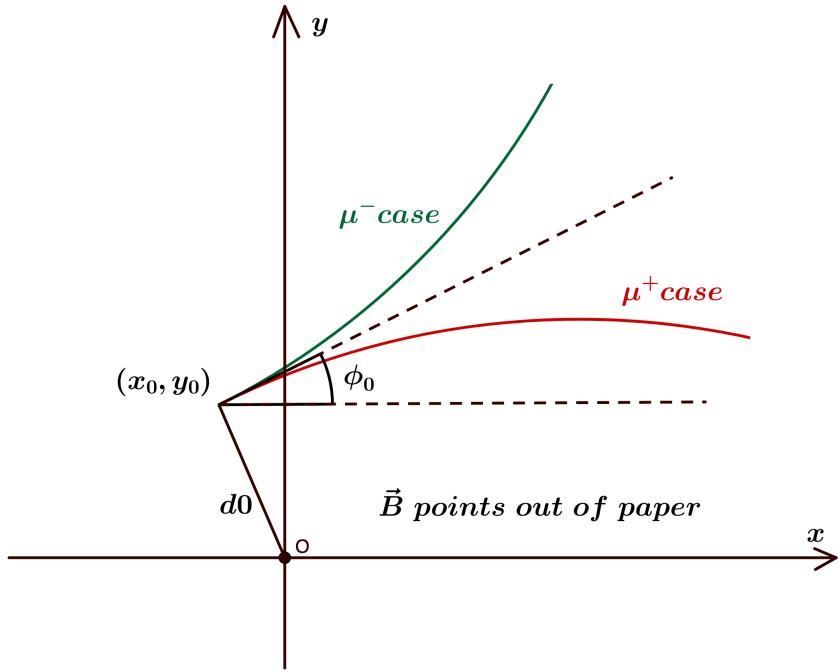


Figure 5-3. Scheme of the  $d_0$  definition in CMS. The  $(x_0, y_0)$  is the coordinate of a point near the vertex in the frame where the vertex is at  $(0, 0)$ , and the  $\phi_0$  is the azimuthal angle of the track at  $(x_0, y_0)$ .

under study are at the scale of several tens of meters, and the  $\Delta r$  is expected to be much smaller than  $r$ . Points  $M$  and  $N$  are expected to be around the coverage of the CMS tracker system, which is about a meter. Therefore  $x$  and  $l$  are expected to be much smaller than  $r$  as well. Finally, the  $d_0$  scale of the tracks under study is about ten microns, which is much smaller than  $x$ ,  $l$ , and  $r$ .

In this setup, a few geometrical relationships can be found between different variables, listed as follows: Since  $x \ll r$ , arc  $\widehat{VM}$  and  $\widehat{PM}$  can be viewed as line segments which are respectively perpendicular to  $OM$  and  $O'M$ . Therefore in triangle  $\triangle VMP$ ,

$$d_0 = x \cdot \sin\alpha \quad (5-2)$$

In triangle  $\triangle O'MO$ , the sine law gives

$$\frac{s}{\sin\alpha} = \frac{r + \Delta r}{\sin\beta} \quad (5-3)$$

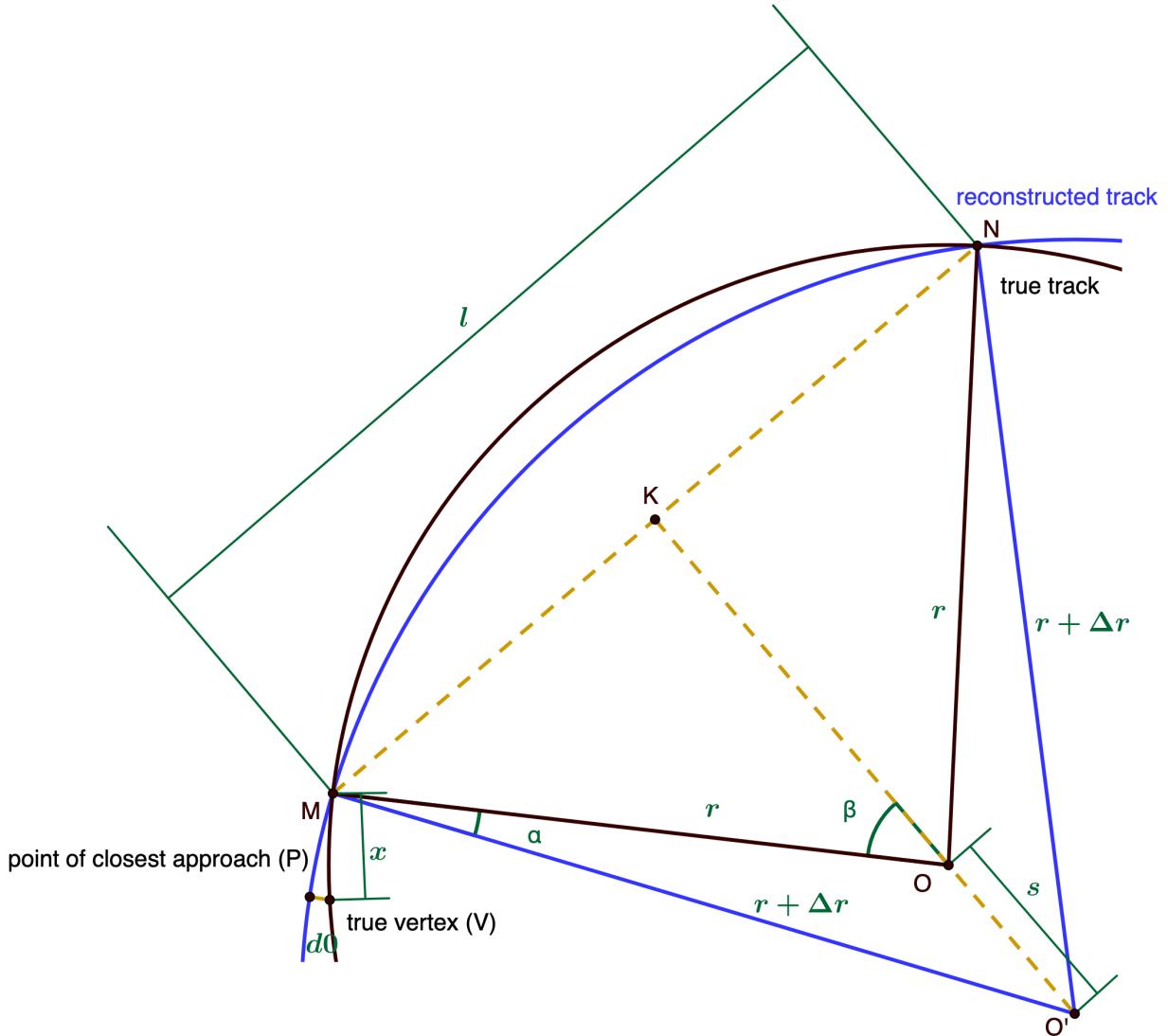


Figure 5-4. Scheme of the track geometry in the transverse plane. The blue lines show the geometry of the reconstructed track, compared to the black lines, which are the geometry of the true track. The difference between the blue track and black track is exaggerated in this scheme. The blue track and the black track must intersect at two points.  $l$  is the distance between the two intersections, and  $x$  is the distance between the true vertex and the first intersection.  $s$  is the distance between the circular centers of the two tracks.

And in triangle  $\triangle OMK$ ,

$$\sin \beta = \frac{l/2}{r} \quad (5-4)$$

Then, using the Pythagorean theorem in both triangle  $\triangle O'MK$  and triangle  $\triangle OMK$ , there

is

$$s = \sqrt{(r + \Delta r)^2 - (l/2)^2} - \sqrt{r^2 - (l/2)^2} \quad (5-5)$$

Combining Equation 5-2 to 5-5 and assuming  $r \gg l$ , one can get

$$d_0 = \frac{xl}{2} \cdot \frac{\Delta r}{r^2} \quad (5-6)$$

Note that in CMS, under the 3.8T magnetic field, tracks follow

$$p_T \text{ (in GeV)} = 1.14 \cdot r \text{ (in meter)} \quad (5-7)$$

We reach

$$d_0 \propto \frac{\Delta p_T}{p_T^2} \quad (5-8)$$

Now a quantitative relationship is extracted between  $d_0$  and  $p_T$ , but with one caveat: the variables  $x$  and  $l$  in the scheme above may vary track by track, and are impossible to measure in real data, meaning that the coefficient in the proportionality is not a constant for different tracks, and Equation 5-8 can be smeared. Therefore, to validate this proportionality, studies are performed on simulated samples comparing the reconstructed  $p_T$  and the generated  $p_T$  of muon tracks. Plots of  $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$  vs  $d_0$  are made, to see whether the proportionality can be observed after the smearing, which is the topic of Section 5.3.2.

Another remark needs to be made, that in Figure 5-4 the  $p_T$  mismeasurement is related to the relative position of the true vertex to the reconstructed track. To be more specific, if the true vertex is inside of the reco track, the  $p_T$  is overestimated, while if the true vertex is outside of the reco track, the  $p_T$  is underestimated. However, in the CMS definition of  $d_0$  shown in Figure 5-3, the sign of  $d_0$  corresponds to an opposite relative position between the vertex and the track for the positively charged muons and the negatively charged muons. A positive  $d_0$  value means the true vertex is inside of the reco track if the muon is positive, but outside of the reco track if the muon is negative. Therefore in CMS convention the  $d_0 - p_T$  correlation is expected be reversed for different muon charges, and

in Section 5.3.2 studies are always performed evaluating  $d_0 \times$  charge rather than just  $d_0$ .

### 5.3.2 Development of GeoFit

The  $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$  vs  $d_0$  plots are made with the following steps: The values  $p_T^{reco}$ ,  $p_T^{gen}$ , and  $d_0$  are extracted for each track in simulated samples. The distribution of  $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$  is made for tracks in different  $d_0 \times$  charge bins. The maximum position and the corresponding full-width-half-maximum (FWHM) is found for each fine-binned  $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$  distribution and set as the value and the uncertainty of one data point in the plots in Figure 5-5. The plots are then fit with analytic functions, which are considered as the experimental realization of Equation 5-8.

In CMS, the colliding vertex is measured by two physics objects, the primary vertex (PV) and the beamspot (BS), as described in Sections 4.1.2 and 4.1.3. Both vertex types are tested and compared in the development of *GeoFit correction*. Examples of the corresponding  $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$  vs  $d_0$  dependences are shown in Figure 5-5. The dependence in the PV plot is not linear as the reconstructed PV is pulled towards the energetic muon tracks, while the dependence in the BS plot follows a linear trend as predicted in Equation 5-1. This is understood as the PV position, reconstructed with a limited number of tracks, can be biased toward the few energetic tracks associated to it, while the BS, averaging numerous tracks from many events, is less affected by individual tracks. Therefore the BS is considered as the position of the true vertex in the rest of the study.

The  $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$  vs  $d_0$  correlation is found to be different in different  $|\eta|$  regions and data-taking periods: different  $|\eta|$  regions are covered by different detector components, and there have been upgrades on the detector and the reconstruction algorithm between different data-taking periods. Overall, the  $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$  vs  $d_0$  correlation is evaluated by three years (2016, 2017, 2018) and three  $|\eta|$  regions (barrel, overlap, endcap), shown in Figure 5-6 for 2016, 5-7 for 2017, and 5-8 for 2018. Each of the plots is fit with a linear function, whose best fit parameters are also shown in the plot.

These fit results are applied as the analytic correction to muon  $p_T$ , based on the  $d_0$ ,

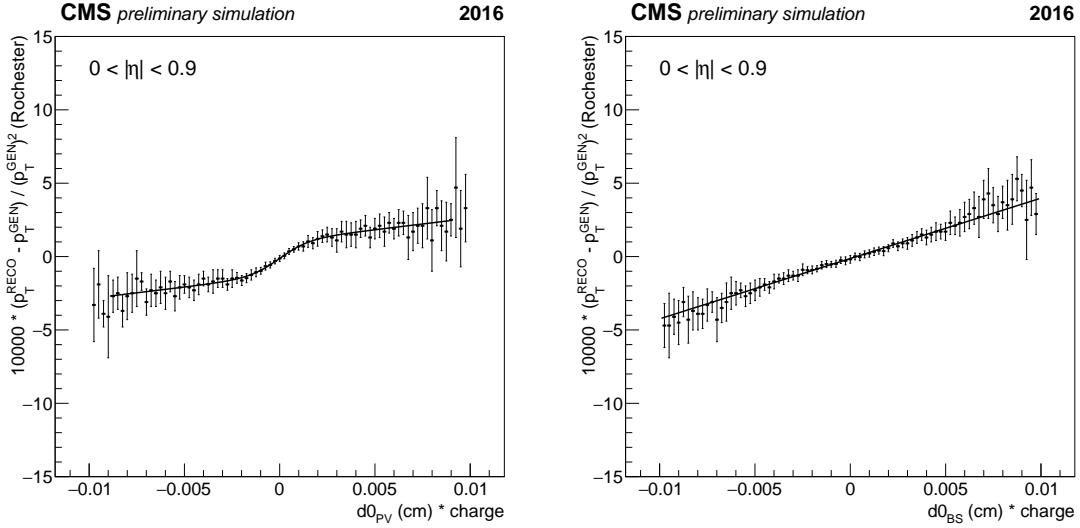


Figure 5-5. Example plots showing the correlation between  $(p_T^{\text{reco}} - p_T^{\text{gen}}) / (p_T^{\text{gen}})^2$  and  $d_0 \times \text{charge}$ . The vertices used for the  $d_0$  calculation are the PV (left) and the BS (right). The PV plot shows a modulated dependence from expectation while the BS plot shows a linear shape as expected. Only barrel tracks from 2016 data are shown as examples. Plots of other  $|\eta|$  regions and other data-taking periods show a similar behavior. Plots credit to Efe Yigitbasi.

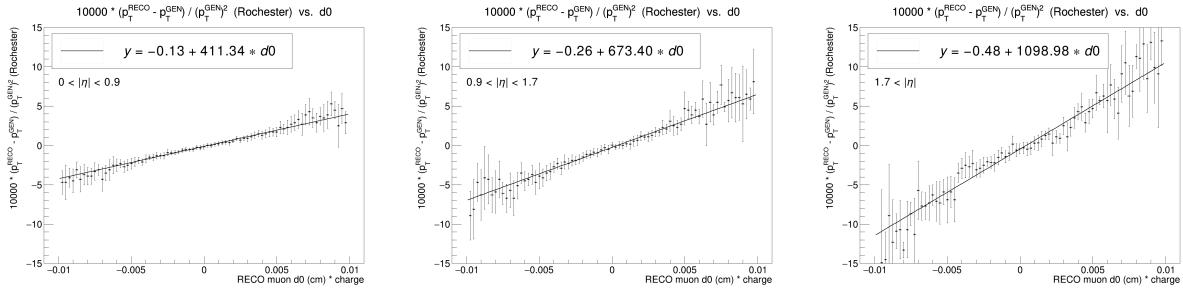


Figure 5-6. Plots for the  $(p_T^{\text{reco}} - p_T^{\text{gen}}) / (p_T^{\text{gen}})^2$  vs  $d_0$  correlation in the 2016 DY simulation, and the linear fits to them. Muon tracks are divided into three different  $|\eta|$  regions:  $|\eta| < 0.9$  (left),  $0.9 < |\eta| < 1.7$  (middle), and  $1.7 < |\eta|$  (right). Plots credit to Efe Yigitbasi.

$p_T$ ,  $|\eta|$ , and charge of the muon. The correction is applied to all muons in data and simulation in all categories in the  $H \rightarrow \mu\mu$  analysis, unless the muon is tagged for *FSR recovery*. The performance of this correction is detailed in Section 5.3.3.

### 5.3.3 Performance and validation

The *GeoFit correction* removes the  $p_T$  dependence on  $d_0$ , whose overall effect on the  $Z \rightarrow \mu\mu$  peak is illustrated in Figure 5-9. A clear trend in the  $m_{\mu\mu}$  is seen regarding  $d_0$

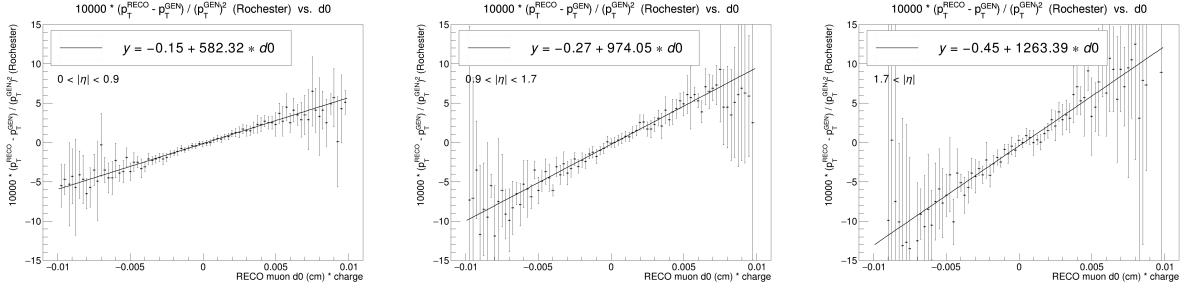


Figure 5-7. Plots for the  $(p_T^{reco} - p_T^{gen}) / (p_T^{gen})^2$  vs  $d_0$  correlation in the 2017 DY simulation, and the linear fits to them. Muon tracks are divided into three different  $|\eta|$  regions:  $|\eta| < 0.9$  (left),  $0.9 < |\eta| < 1.7$  (middle), and  $1.7 < |\eta|$  (right). Plots credit to Efe Yigitbasi.

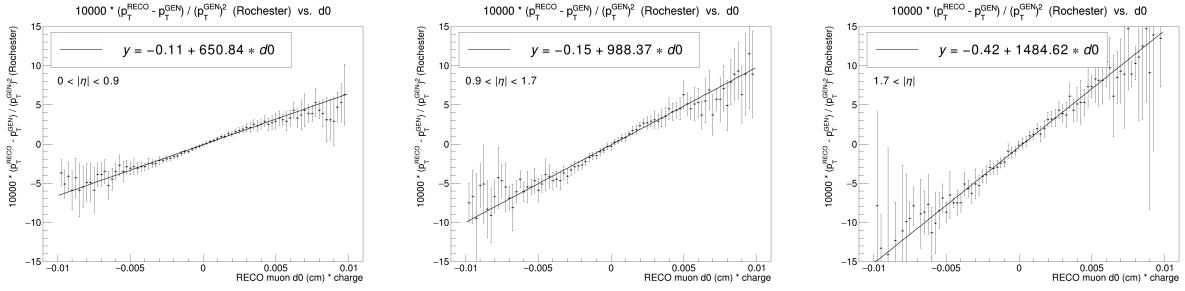


Figure 5-8. Plots for the  $(p_T^{reco} - p_T^{gen}) / (p_T^{gen})^2$  vs  $d_0$  correlation in the 2018 DY simulation, and the linear fits to them. Muon tracks are divided into three different  $|\eta|$  regions:  $|\eta| < 0.9$  (left),  $0.9 < |\eta| < 1.7$  (middle), and  $1.7 < |\eta|$  (right). Plots credit to Efe Yigitbasi.

before the *GeoFit correction*, while no significant dependence remains after the correction. As a side remark, the  $m_{\mu\mu}$  mismeasurement in Figure 5-9 can be as large as 1.5 GeV for extreme  $d_0$  values, but in data and simulation the distribution of the muon  $d_0$  is roughly a Gaussian shape with a standard deviation around 15  $\mu\text{m}$ . So most of the events are near the center of the plots, and the size of the correction is not as exaggerated as the values at the tails.

Overall, the removal of the  $p_T - d_0$  dependence leads to an improvement on the inclusive  $m_{\mu\mu}$  resolution. This improvement is different for different processes depending on their kinematic profiles in  $p_T$  and  $|\eta|$ . Figure 5-10 shows the improvement on  $m_{\mu\mu}$  resolution in the four main expected signal modes, ggH, VBF, VH, and  $t\bar{t}H$ . The relative improvements on  $m_{\mu\mu}$  resolution for ggH, VBF, VH, and  $t\bar{t}H$  modes are, respectively, 6.1%,

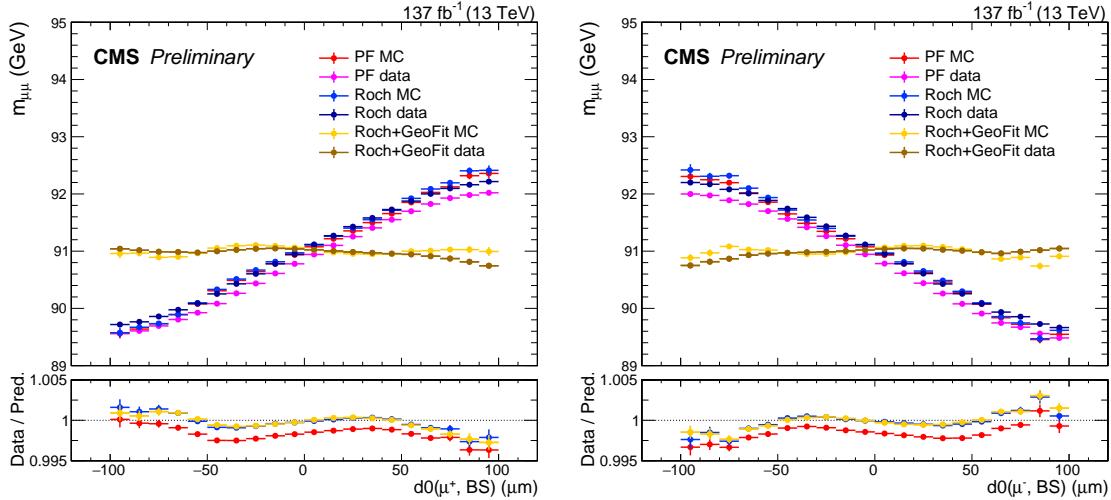


Figure 5-9. Plots showing the  $p_T$  dependence on the  $d_0$  value with different stages of muon correction. The plots compare the  $Z \rightarrow \mu\mu$  peak in data and simulation for three years (2016-2018) combined. All positively charged muons are put in the left plot and all negatively charged ones are put in the right plot. The  $p_T - d_0$  dependence is reversed for positive and negative muons.

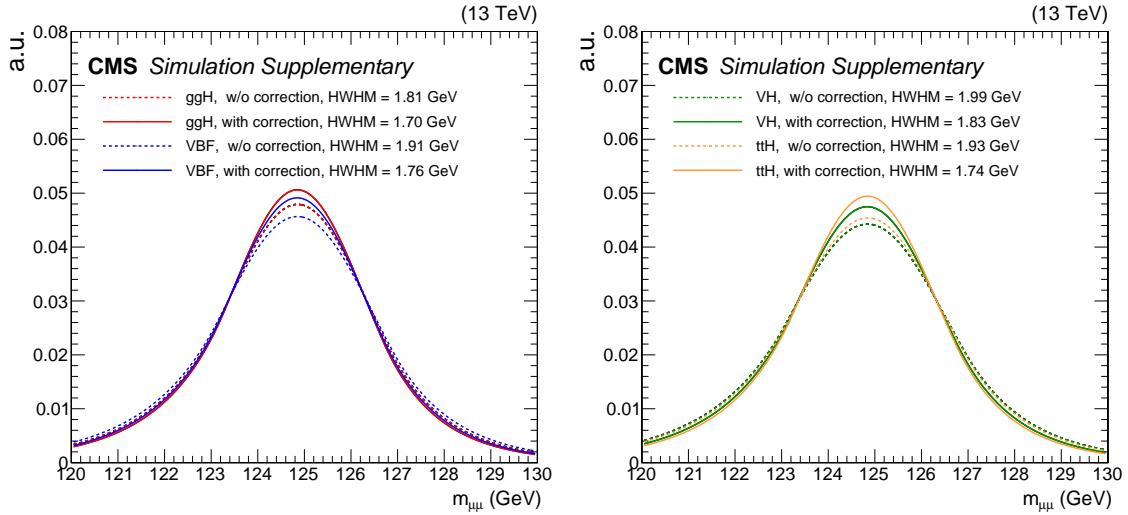


Figure 5-10. Plots showing the *GeoFit correction* improvement on the four main  $H \rightarrow \mu\mu$  signal modes, ggH and VBF plotted on the left, and VH and  $t\bar{t}H$  plotted on the right. The plots are made combining the expected signal in all three years of data-taking (2016-2018). The relative improvements on  $m_{\mu\mu}$  resolution for ggH, VBF, VH, and  $t\bar{t}H$  modes are, respectively, 6.1%, 7.8%, 8.0%, and 9.8%.

7.8%, 8.0%, and 9.8%. This improvement on signal resolution translates to about 5% improvement on the significance of the inclusive  $H \rightarrow \mu\mu$  analysis.

The different improvements in different signal modes comes from their  $p_T$  profiles. As the correction is proportional to  $d_0$  and  $p_T^2$ , the relative improvement  $\Delta p_T/p_T$  is proportional to  $d_0$  and  $p_T$ . Figure 5-11 shows the  $d_0$  (left) and  $p_T$  (right) distributions of different signal modes. The  $d_0$  profiles of the 4 signals are exactly the same, while the  $p_T$  profiles are different.  $t\bar{t}H$  signal has more high  $p_T$  muons than other signal modes, while  $ggH$  signal has more muons on the low  $p_T$  side. Therefore they gain the largest and smallest improvements from the *GeoFit correction*, respectively. VBF and VH signals have very similar  $p_T$  profiles, and their improvements from the *GeoFit correction* are about the same.

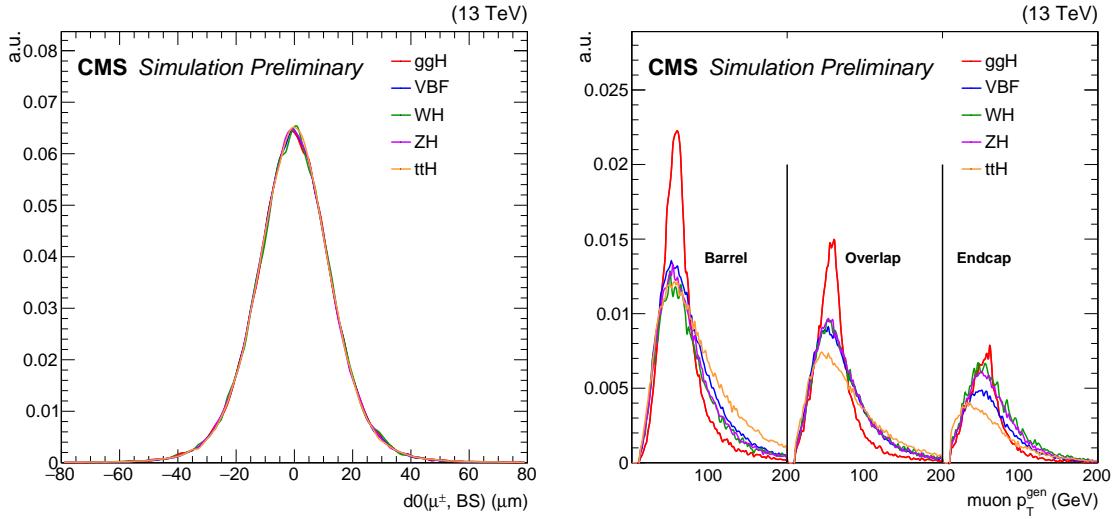


Figure 5-11. Plots showing  $d_0$  and  $p_T$  distributions of the four main  $H \rightarrow \mu\mu$  signal modes. The  $d_0$  profiles (left plot) of the 4 signals are the same. The  $p_T$  profiles (right plot) are different, which explains the different relative improvements that the signals receive from the *GeoFit correction*.

### 5.3.4 GeoFit vs track refit

The *GeoFit correction* provides a simple method to correct the  $p_T$  dependence on  $d_0$  based on high level physics variables. Since the origin of this  $p_T$  dependence is well-understood, it is also possible to derive a more fundamental correction by refitting each muon track including the BS position as an additional constraint to the track. This method requires lower-level information of muon reconstruction and is computationally more expensive, but is in principle more precise. To compare the performance of the *GeoFit correction*

tion and the refit method, a preliminary study is made on the 2018 ggH signal simulation. The  $m_{\mu\mu}$  shape of the inclusive signal is plotted applying the track refit method vs applying the *GeoFit correction*, shown in Figure 5-12. This comparison shows that the  $m_{\mu\mu}$  shapes from the two methods are almost equivalent. The *GeoFit correction*, although an approximation method, captures most of the effect and provides about the same improvement in  $m_{\mu\mu}$  resolution as the refit method. The *GeoFit correction* is therefore chosen in the  $H \rightarrow \mu\mu$  analysis to speed up the workflow.

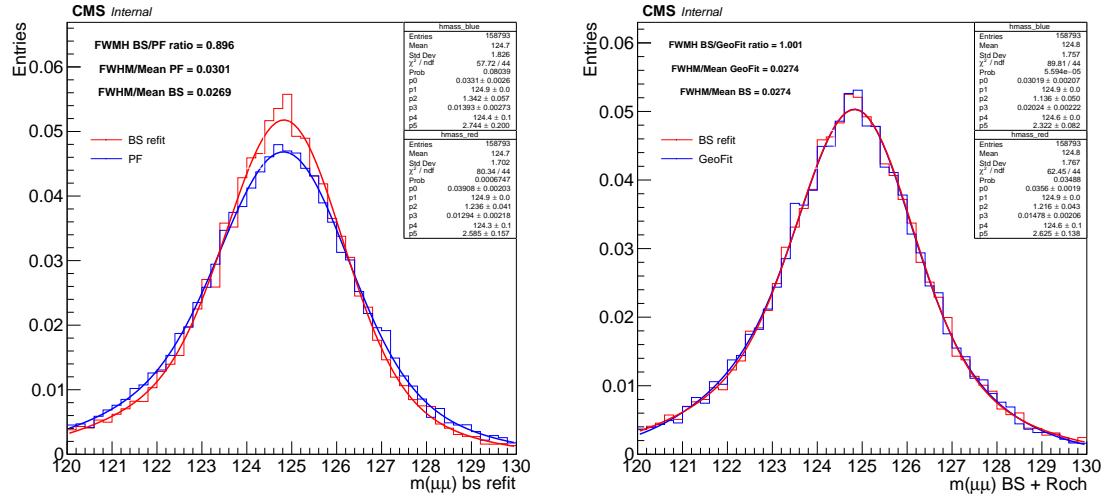


Figure 5-12. Plots of the  $m_{\mu\mu}$  shape of the 2018 ggH simulation sample, comparing different muon correction methods. The left plot shows the  $m_{\mu\mu}$  distribution calculated with muon tracks refitted with the additional BS constraint, compared with the particle flow shape (left plot). The *Rochester correction* is not applied in the left plot for both the red and the blue lines. The right plot shows the  $m_{\mu\mu}$  distribution from the refit method, with the *Rochester correction* applied, compared with the shape from *GeoFit correction + Rochester correction* (right plot). Plots credit to Pierluigi Bortignon.

## 5.4 Muon calibration results

The  $Z \rightarrow \mu\mu$  is a well-understood process with a mass scale not far from the Higgs boson and with a much larger number of events at the LHC. It is therefore used as a candle to monitor the performance of the *Rochester correction* and the *GeoFit correction*, and validate that these corrections do not introduce new biases. In this study, the distribution of the  $m_{\mu\mu}$  is plotted in different bins of various dimuon kinematic variables. The  $m_{\mu\mu}$  distributions

are fit with a Voigtian + Exponential function, in which the Voigtian part is a convolution of a Breit-Wigner function and a Gaussian function. The parameter "mean mass" from the Breit-Wigner part and standard deviation from the Gaussian part are taken as the mean value and the experimental resolution of the  $m_{\mu\mu}$  distribution. They are plotted against the dimuon kinematic variable of interest to check for potential trends.

The calibration plots are made by year as the corrections are provided by year. Events containing *FSR recovery* are removed from this study as it is a separate effect. Different variables are tested in the Figures listed: Figure 5-13 for the  $\eta$  of the positive muon, Figure 5-14 for the  $\phi$  of the positive muon, Figure 5-15 for the  $\phi$  of the negative muon, Figure 5-16 for the  $p_T$  of the positive muon, Figure 5-17 for the  $p_T$  of the dimuon system, Figure 5-18 for the  $\eta$  of the dimuon system, Figure 5-19 for the  $d_0$  of the positive muon, and Figure 5-20 for the  $d_0$  of the negative muon.

From these plots, it can be concluded that all the known biases in muon  $p_T$  are removed and no new bias has been introduced. The *Rochester correction* and *GeoFit correction* correct orthogonal effects, and do not interfere with the performance of each other. After the corrections, a per-mille level agreement is achieved between data and simulation in the  $m_{\mu\mu}$  value, while the agreement in  $m_{\mu\mu}$  resolution is about a few percent.

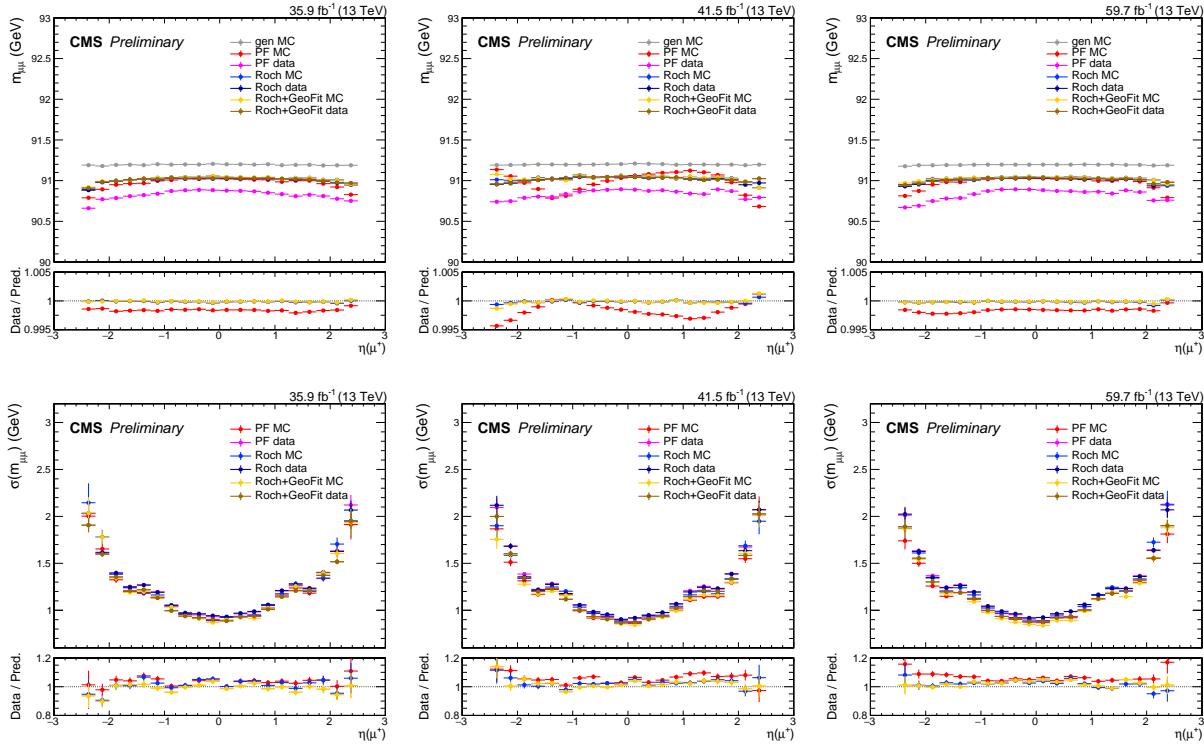


Figure 5-13. Muon calibration plots vs  $\eta(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution.

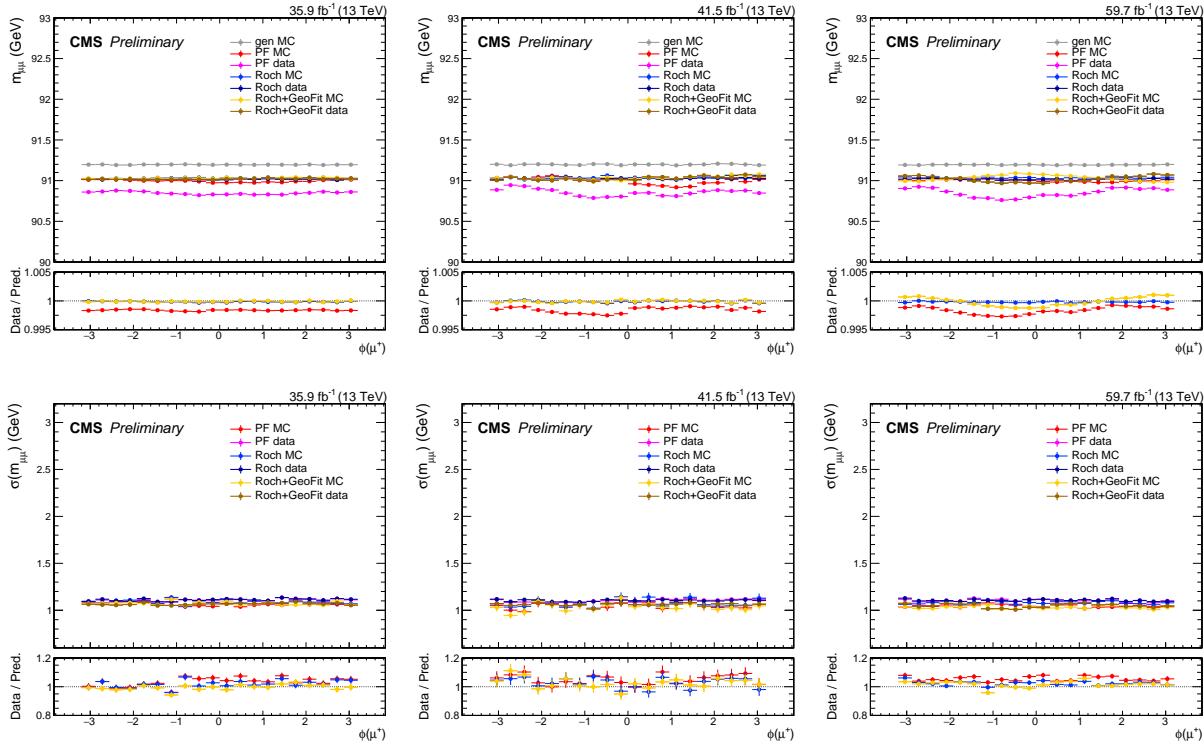


Figure 5-14. Muon calibration plots vs  $\phi(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution.

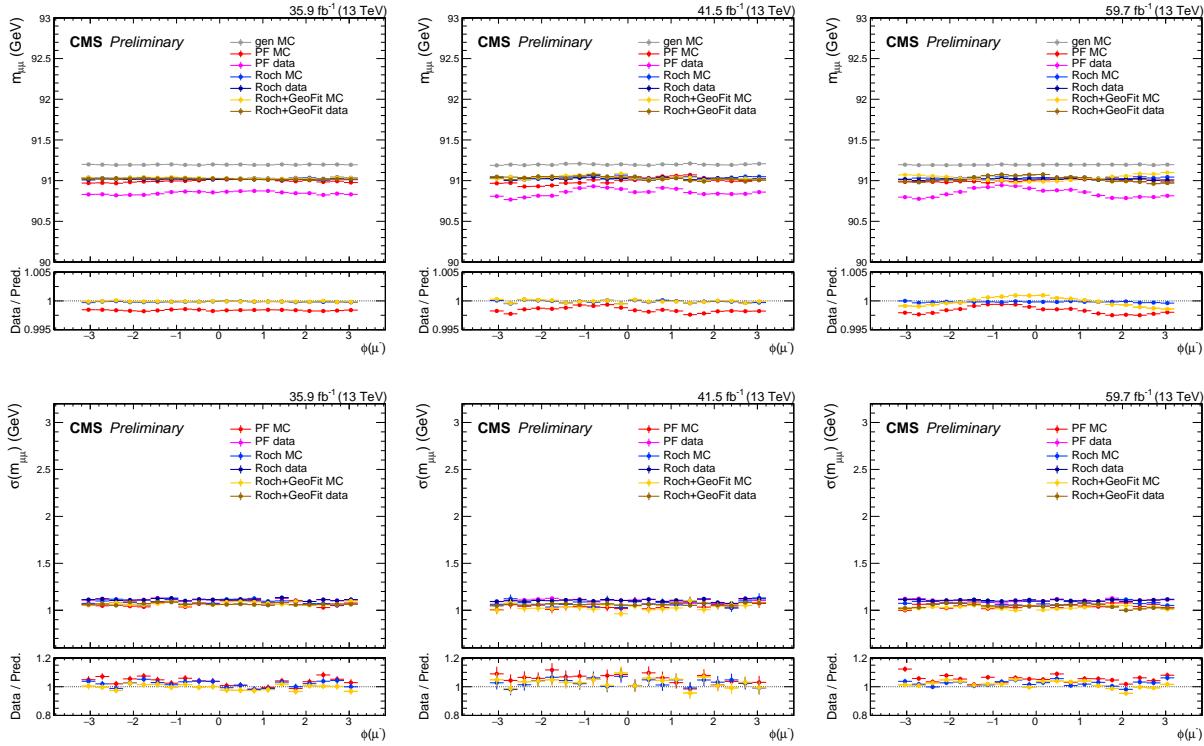


Figure 5-15. Muon calibration plots vs  $\phi(\mu^-)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution.

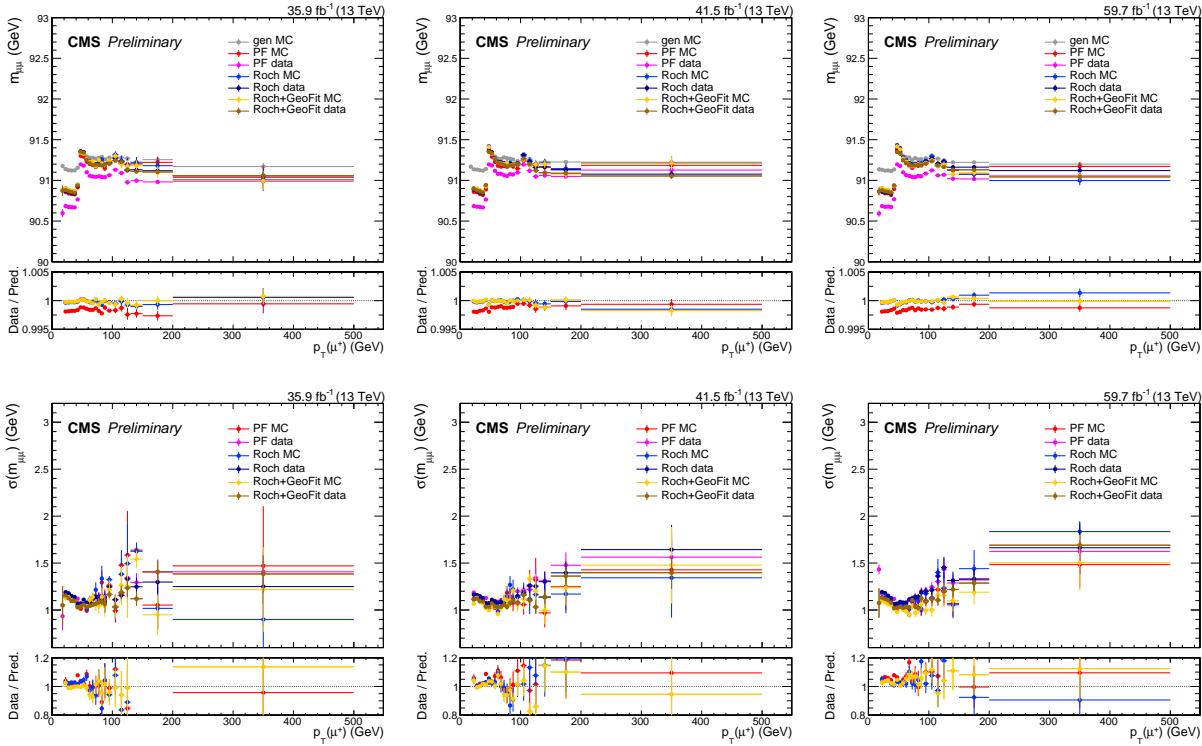


Figure 5-16. Muon calibration plots vs  $p_T(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution. The  $p_T$  binning sculpts the shape of the  $m_{\mu\mu}$  peak, which leads to a jump at the  $p_T = 45$  GeV in the plots.

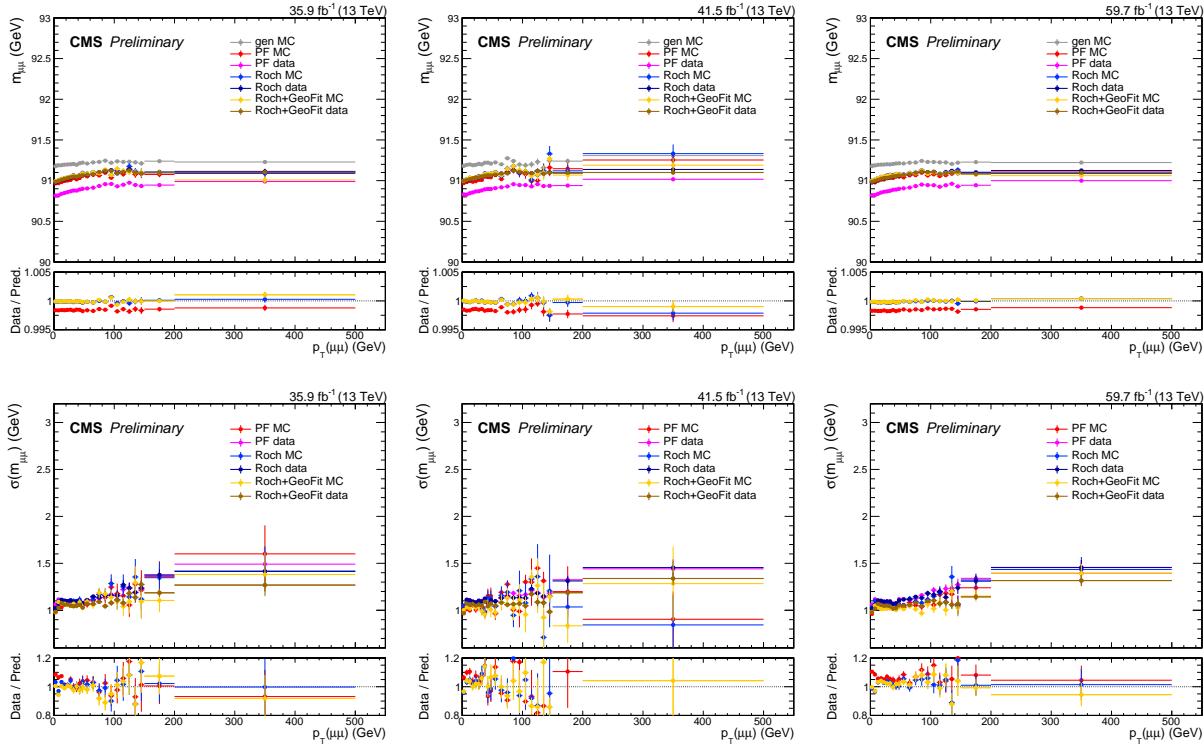


Figure 5-17. Muon calibration plots vs  $p_T(\mu\mu)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution.

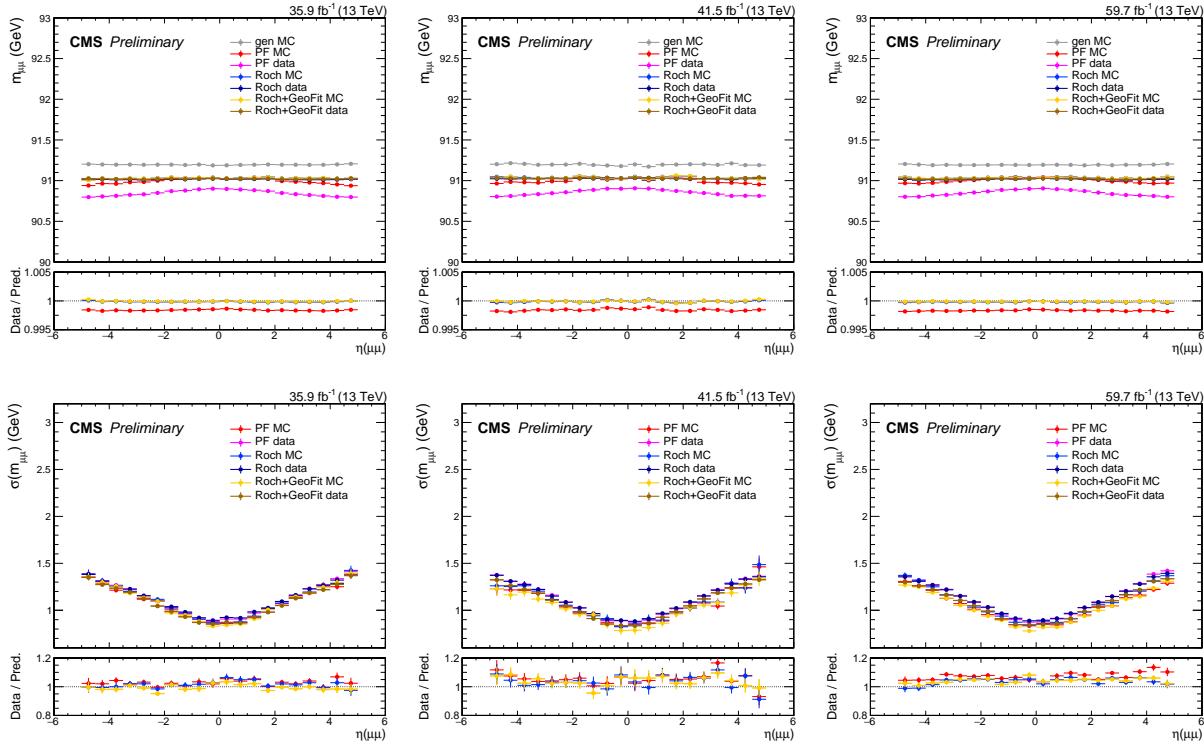


Figure 5-18. Muon calibration plots vs  $\eta(\mu\mu)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution.

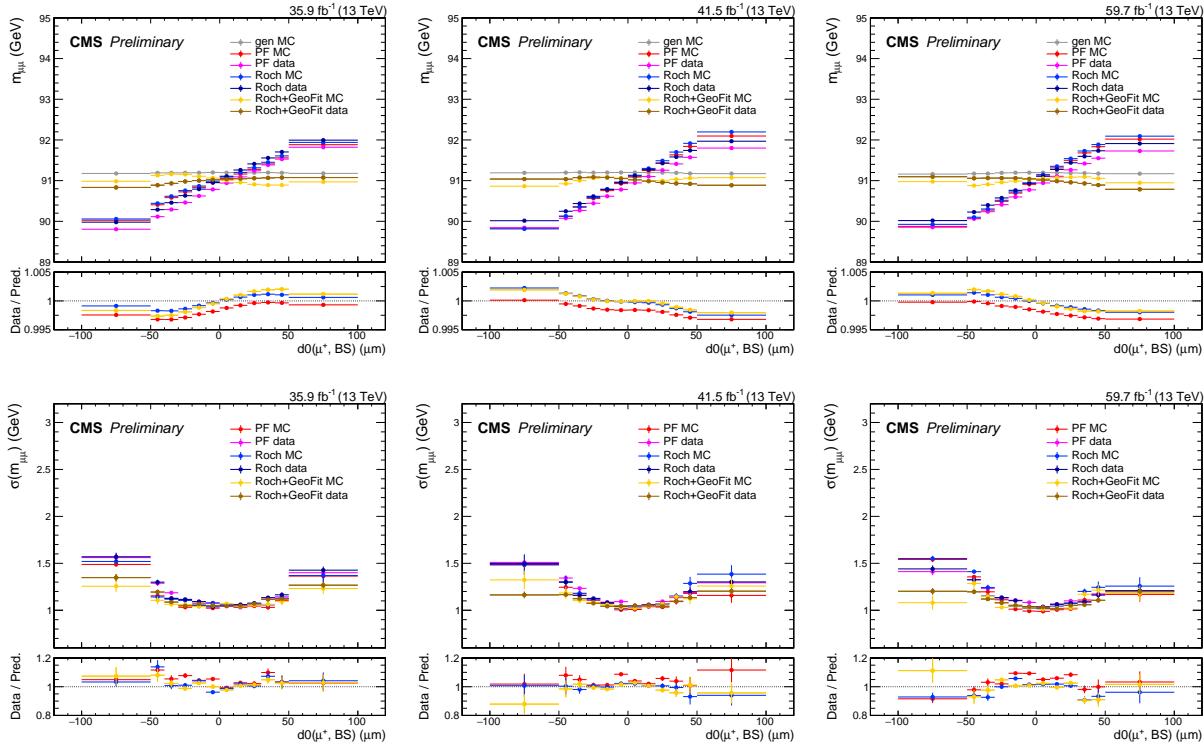


Figure 5-19. Muon calibration plots vs  $d_0(\mu^+)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution.

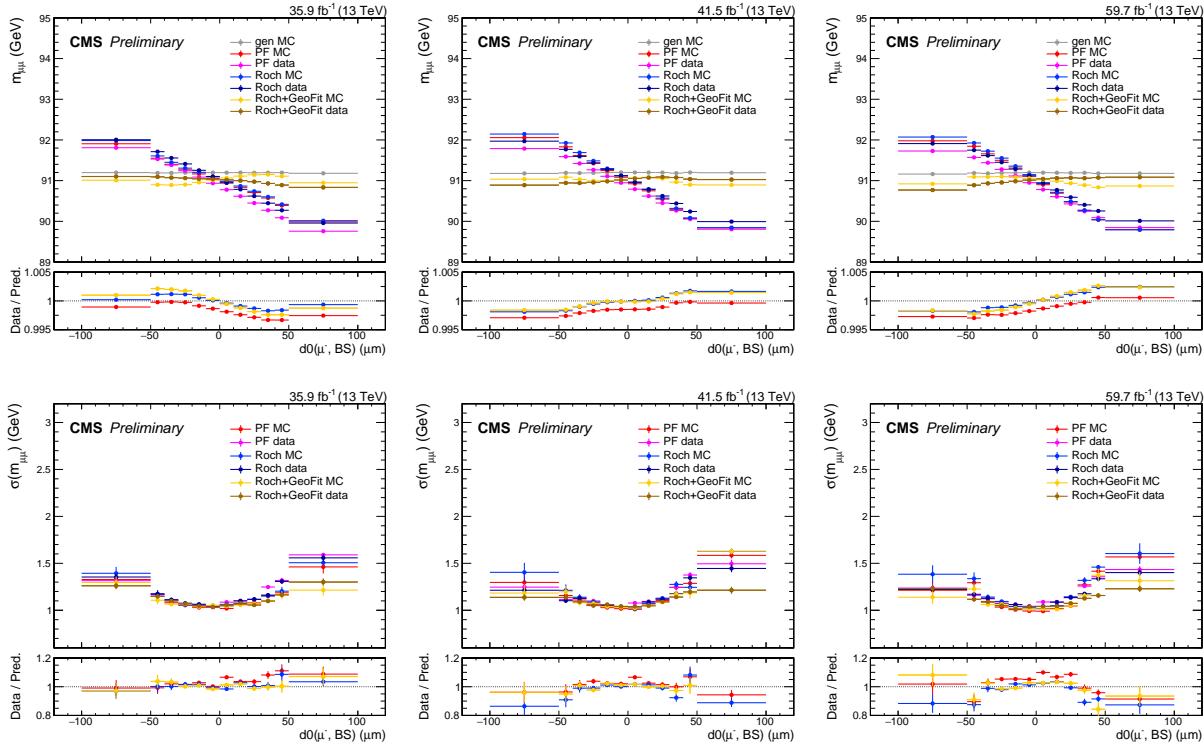


Figure 5-20. Muon calibration plots vs  $d_0(\mu^-)$ , for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the  $m_{\mu\mu}$  distribution, while the bottom row shows its experimental resolution.

## CHAPTER 6

### SEARCH FOR H2MU TARGETING THE VH PRODUCTION MODE

As described in Section 3.2, the analysis in the VH category targets the VH production modes of the Higgs boson, and is established as two independent parts, the WH category and the ZH category. This chapter provides a full description of the procedures and the results in these two categories.

The VH analysis focuses on the leptonic ( $e$  or  $\mu$ ) decay modes of vector bosons ( $V = W$  or  $Z$ ), which lead to distinct final states involving extra well-reconstructed charged lepton(s) in addition to the two muons from the Higgs boson decay. By requiring these additional leptons, the main backgrounds in the generic  $\mu\mu$  phase-space, the DY and the  $t\bar{t}$  processes, are greatly suppressed, leading to a high  $S/B$  in the VH category, and ensuring a good expected significance with only a handful of signal events. The other decay modes of the vector boson are disregarded for reasons as described below.

The vector bosons decay to tau leptons ( $W \rightarrow \tau + \nu_\tau$ , or  $Z \rightarrow \tau\bar{\tau}$ ) at the same rate as they decay to electrons or muons. The reconstruction of  $\tau$ , as described in Section 4.1.10, targets the hadronic decays of taus and suffers from a sizable fake rate. Tagging final states with hadronic tau decays gives a low  $S/B$  phase-space and yields negligible contribution to the overall sensitivity. Tau leptons may also decay into leptons ( $e/\mu$  + neutrinos). The resulting  $e$  or  $\mu$  only carries about one third of the  $\tau$  momentum, and has a displacement of a few millimeters from the primary vertex. The majority of them fail electron or muon selection criteria of the analysis.

The hadronic decays of the  $W$  or the  $Z$  boson lead to two jets in the event, making an invariant mass near the mean mass of the boson. Although amounting to a larger branching ratio than the leptonic decay modes, these hadronic decays turned out not as helpful in enriching the VH events. The dijet selection is subject to immense QCD backgrounds and has a much worse mass resolution than for leptons. It gives a phase-space dominated by ggH+jets signal vs DY+jets background. Therefore the hadronic VH events are considered as minor signal contributions in the ggH category, rather than a dedicated VH hadronic tag.

The Z boson can decay to a pair of neutrinos at a branching ratio of 20%, which leave no electronic signal in the CMS detector and appear as missing transverse momentum ( $E_T^{\text{miss}}$ ) in the event. This  $E_T^{\text{miss}}$  equals to the  $p_T$  of the Z boson and can provide discrimination against some backgrounds. However, as described in Section 4.1.11, the  $E_T^{\text{miss}}$  resolution is much worse than those of leptons and jets. The purity of the  $ZH \rightarrow vv + \mu\mu$  signal can only be enhanced if a very tight cut on the  $E_T^{\text{miss}}$  (for example  $E_T^{\text{miss}} > 100$  GeV) is applied. Only a small fraction of ZH signal would pass this selection while some large irreducible backgrounds still remain, like the  $t\bar{t}$  and diboson processes. Overall, a tag for the  $ZH \rightarrow vv + \mu\mu$  would be much less sensitive than the existing  $ZH \rightarrow \ell\ell + \mu\mu$  tag, and is not deployed. The  $ZH \rightarrow vv + \mu\mu$  signal events are considered as minor signal contributions in the ggH category.

After requiring additional lepton(s) in the event, the main background in the resulting VH phase-spaces becomes the WZ and ZZ processes, for the WH and ZH categories respectively. For both WZ and ZZ backgrounds, if more than two muons are present in the event, there are different possibilities of the association between the muons and their parent particles. In fact, the wrong pairing of muons, called the combinatorial background, yields the majority of the WZ and ZZ background. For example, in an on-shell WZ event, the muon from the W decay can be falsely paired with the oppositely signed (OS) muon from the Z decay, making an invariant mass near the Higgs boson mass value. A set of cut-based event selection is optimized to reduce the combinatorial background, further improving the  $S/B$ . More details are given in Section 6.1.

The minor backgrounds may include the triboson processes, the DY process accompanied with additional nonprompt leptons, or the t quark associated processes, for example the  $t\bar{t}$ ,  $tW$ , and  $t\bar{t}V$  processes, where the b quarks from the top decays either fall out of the acceptance of the detector or fails the b-tagging. All these minor backgrounds have different kinematic profiles from the signal and can be reduced to different extents. Multivariate discriminators are trained in both the WH and the ZH categories to account for the

differences between the signal and the inclusive background as much as possible.

Events are further divided in to several subcategories with different  $S/B$  based on the Multivariate discriminators. In each subcategory, the  $m_{\mu\mu}$  spectrum is plotted, and analytic functions are used to model the shapes of signal and background. The strength of the  $H \rightarrow \mu\mu$  signal is evaluated by fitting the signal and the background functions to the  $m_{\mu\mu}$  spectrum of data.

The following sections of this chapter are organized as follows: Section 6.1 describes the event selection. Section 6.2 discusses the details of the training of the MVA discriminators, and the determination of subcategories based on them. Section 6.4 shows the performances of different analytic functions on the signal and background modeling. Section 6.5 lists different sources of systematic uncertainties considered in this analysis. And Section 6.6 gives the statistical interpretation of the results.

## 6.1 Event selection

The VH analysis takes standard CMS physics objects described in Chapter 4, whose selections in this analysis are detailed in Section 4.2. The selection criteria for electrons and muons, which are the most important for this analysis, are also summarized in Table 6-1.

Table 6-1. Selection criteria on muons and electrons in the VH analysis.

Variable	Muon	Electron
$p_T$	$> 20\text{GeV}$	$> 20\text{GeV}$
$ \eta $	$< 2.4$	$< 2.5$
ID and Iso	Medium ID + Loose Iso	MVA wp90
ECal gap veto	-	(1.444, 1.566)
$d_{xy}(\text{PV})$	$< 0.05\text{ cm}$	$< 0.05\text{ cm}$
$d_z(\text{PV})$	$< 0.10\text{ cm}$	$< 0.10\text{ cm}$
SIP	$< 8.0$	$< 8.0$
Conversion Veto	-	✓
Number of Missing Hits	-	$< 2$
lepMVA	$> 0.4$	$> 0.4$

The event selection targets the leptonic decays of the W or the Z bosons in the VH signals. The selection steps are devised to suppress different background processes and

optimize the  $S/B$  in the WH and ZH categories respectively. The event selection in the  $\text{WH} \rightarrow \ell\nu + \mu\mu$  category is described as follows:

- At least one muon must have  $p_T > 26\text{GeV}/29\text{GeV}/26\text{GeV}$  for years 2016/2017/2018 respectively, which is matched to a single-muon trigger object
- All SFOS lepton pairs must have an invariant mass  $> 12 \text{ GeV}$
- The charge of the three leptons must add up to  $\pm 1$
- At least one  $\mu^+\mu^-$  pair must have an invariant mass between 110 and 150 GeV
- If two  $\mu^+\mu^-$  pairs fall in the 110 - 150 GeV mass window, the pair with the higher  $p_T$  is chosen as the Higgs candidate (denoted  $\mu\mu_H$ )
- The event must contain exactly 0 medium b-tagged jet and less than 2 loose b-tagged jets
- In  $3\mu$  events, the non-Higgs-candidate  $\mu^+\mu^-$  pair ( $\mu\mu_{OS}$ ) must not have an invariant mass between 81 and 101 GeV, to suppress WZ and Z+jets backgrounds

The event selection in the  $\text{ZH} \rightarrow \ell\ell + \mu\mu$  category is described as follows:

- At least one muon must have  $p_T > 26\text{GeV}/29\text{GeV}/26\text{GeV}$  for years 2016/2017/2018 respectively, and the event must contain an unprescaled single-muon trigger object
- The charge of the four leptons must add up to 0.
- All SFOS lepton pairs must have an invariant mass  $> 12 \text{ GeV}$ .
- In  $\mu\mu ee$  events, the  $e^+e^-$  pair must have invariant mass between 70 and 110 GeV, and the  $\mu^+\mu^-$  pair must have invariant mass between 110 and 150 GeV.
- In  $4\mu$  events, if it is possible to form two distinct  $\mu^+\mu^-$  pairs each with a mass between 81 and 101 GeV, the event is discarded.
- In  $4\mu$  events, one muon pair must have mass between 110 and 150 GeV, and the other muon pair must have mass between 81 and 101 GeV.
- In  $4\mu$  events, if both combinations have a muon pair in the Z-mass window and a muon pair in the signal-mass window, the combination in which the mass of the Z candidate is closer to 91 GeV is chosen.
- The event must contain exactly 0 medium b-tagged jet and less than 2 loose b-tagged jets.

These selection criteria are also summarized in Table 6-2.

Table 6-2. Event selections for the WH and ZH categories.

Criterion	WH category	ZH category
Muon trig match	✓	✓
b-jets veto, 0 medium and < 2 loose	✓	✓
$\mu^+ \mu^-$ pair with $110 < m_{\mu\mu} < 150$ GeV	✓	✓
Additional lepton(s)	1	1 SFOS pair
Low-mass resonance veto $m_{\ell\ell} > 12$ GeV	✓	✓
Number of $ m_{\mu\mu} - m_Z  < 10$ GeV or $ m_{ee} - m_Z  < 20$ GeV	= 0	= 1
Choice of muon combination	Highest $p_T(\mu\mu)$ as $\mu\mu_H$	Smallest $ m_{\mu\mu} - m_Z $ as $\mu\mu_Z$

## 6.2 MVA discrimination

After the event selection of the WH or the ZH categories, the remaining background processes resemble the kinematic signatures of the signals and cannot be decisively reduced by simple selection cuts. MVA discriminators are trained to further enhance the  $S/B$ . The algorithm used in this analysis is called the Boosted decision tree (BDT), which is a supervised machine learning algorithm that classifies data by iteratively generating and averaging (boosting) a series of decision trees.

Different variables can be effective in separating different background processes. For example, the muon pair from t quark associated processes usually have more  $p_T$  than those from the WH signal process, while the  $E_T^{\text{miss}}$  in the DY process is likely to be smaller than that in the WH events. The main background, WZ or ZZ in the WH and ZH categories respectively, in which the Higgs candidate  $\mu\mu$  pair comes from an off-shell Z boson, is a more complicated story as it looks almost kinematically identical to the signal. The key in discriminating them lies in the spin difference between the Z and the H bosons, which is measured as a difference in the helicity angle  $\theta^*$  between the decay products of the Z (H) boson.

The helicity angle is defined in a decay system in the frame in which the parent particle is at rest, as the angle between the direction of the decay and the boost direction of the parent particle. The distribution of this angle is determined by the spins of the parent particle and the decay products. In the case of  $W \rightarrow WZ$  ( $Z \rightarrow ZZ$ ) vs  $W \rightarrow WH$  ( $Z \rightarrow ZH$ ), the helicity angle between the W ( $Z_1$ ) and the  $Z_2$  follows the distribution  $1 + \cos^2 \theta^*$  while the the helicity angle between the W ( $Z$ ) and the H follows a flat distribution. Similarly,

for  $Z \rightarrow \mu\mu$  vs  $H \rightarrow \mu\mu$ , the muons from a  $Z$  decay tend to align with the polarization of the  $Z$  boson, while the muons from a  $H$  decay do not have a preference in direction.

All these helicity angles,  $\theta_{WH}^*$ ,  $\theta_{ZH}^*$ , and  $\theta_{\mu\mu}^*$ , can in principle provide significant discrimination between the signal and the background. However, in practice,  $\theta_{WH}^*$  cannot be reconstructed because of the lack of information of the neutrino from the  $W$  decay, and the distribution of  $\theta_{\mu\mu}^*$  is severely sculpted by the acceptance of the CMS detector, rendering a similar shape for the  $Z \rightarrow \mu\mu$  and  $H \rightarrow \mu\mu$  processes. On the other hand, the spin information can be partially captured in other variables, for example the helicity angle between the  $W$ -lepton and the  $\mu\mu_H$  system, and some other angular correlations like  $\Delta\phi$ ,  $\Delta\eta$ . All these variables are tested as inputs to the BDT for the performance. Some of them turned out insignificant and are later trimmed off from the input collection. To make sure the BDTs do not sculpt the  $m_{\mu\mu}$  shape, which will be used for the signal extraction, variables that are strongly correlated with  $m_{\mu\mu}$ , for example the  $p_T$  of the muons, are not used in the BDTs.

### 6.2.1 BDT targeting $WH \rightarrow \ell\nu + \mu\mu$ signal

The  $WH$  BDT takes variables of three different kinds, the kinematic variables of leptons that are uncorrelated with  $m_{\mu\mu}$ , the angular correlations between different leptons as discussed above, and the variables reflecting the missing energy in the event. In total, there are 16 input variables to the BDT, which are listed in Table 6-3. The variables related to the missing energy includes the missing energy itself, the transverse mass ( $M_T$ ) between a lepton and the missing energy, and the angular separation between a lepton and the missing energy. Two types of missing energy are tested,  $E_T^{\text{miss}}$ , which is the negative of the vector sum of the  $p_T$  of all PF candidates, and  $H_T^{\text{miss}}$ , which only considers well-defined jets, photons, and leptons in the similar calculation. The  $E_T^{\text{miss}}$ -related and  $H_T^{\text{miss}}$ -related variables are expected to play interchangeable roles in the BDT. The  $H_T^{\text{miss}}$ -related variables turned out to be slightly better performing and are kept in the final BDT, while the  $E_T^{\text{miss}}$ -related variables are trimmed.

Another important feature in the WH category is that about 40% of the WZ background are from the wrong combination of muons in  $3\mu$  events. The duplication of some variables calculated with the alternative combination of the muons in each event may also help with the discrimination. For example, apart from the transverse mass between the nominal W-lepton and the  $H_T^{\text{miss}}$ , another transverse mass is also considered, using the  $H_T^{\text{miss}}$  and the same-sign muon from the nominal Higgs candidate, which turns out to be effective.

Table 6-3. List of input variables used to train the signal-background separation BDT in the WH category. In this table,  $\mu\mu_H$  is the Higgs candidate,  $\ell$  is the lepton from the W decay,  $\mu_{OS}$  ( $\mu_{SS}$ ) refers to the muons in the Higgs candidate which OS (SS) to the lepton.

Variable	Description
$p_T(\mu\mu_H)$	$p_T$ of the Higgs candidate
$ \eta(\mu_1) $	$\eta$ of the leading muon in the Higgs candidate
$ \eta(\mu_2) $	$\eta$ of the trailing muon in the Higgs candidate
$\Delta R(\mu_{SS}, \mu_{OS})$	$\Delta R$ between the two muons in the Higgs candidate
$p_T(\ell)$	$p_T$ of the extra lepton in the event
Number of electrons	Number of electrons in the event
$\Delta R(\ell, \mu\mu_H)$	$\Delta R$ between the extra lepton and the Higgs candidate
$\Delta\eta(\ell, \mu\mu_H)$	$\Delta\eta$ between the extra lepton and the Higgs candidate
$\Delta\eta(\ell, \mu_{SS})$	$\Delta\eta$ between the extra lepton and the SS muon
$\cos\theta^*(\ell, \mu_{SS})$	$\cos\theta^*$ between the extra lepton and the SS muon
$\Delta R(\ell, \mu_{OS})$	$\Delta R$ between the extra lepton and the OS muon
$\Delta\eta(\ell, \mu_{OS})$	$\Delta\eta$ between the extra lepton and the OS muon
$\cos\theta^*(\ell, \mu_{OS})$	$\cos\theta^*$ between the extra lepton and the OS muon
$M_T(\mu_{SS}, MHT)$	transverse mass of the $H_T^{\text{miss}}$ and the SS muon
$M_T(\ell, MHT)$	transverse mass of the $H_T^{\text{miss}}$ and the extra lepton
$ \Delta\phi(\ell, MHT) $	$ \Delta\phi $ between the $H_T^{\text{miss}}$ and the extra lepton

The sensitivity of this analysis depends largely on the resolution of the  $m_{\mu\mu}$  peak, which is determined by muon momentum resolution and in turn primarily depends on  $\eta$  of the muons as the detector condition differs in different  $\eta$  regions in CMS. It is important to divide events with different resolution into different categories, which leads to an enhancement to the overall  $S/B$ . In the previous  $H \rightarrow \mu\mu$  analysis [42, 101], the categorization is achieved by dividing events based on both the BDT output and the  $\eta$  value of

the muons. While in this work, the resolution information is incorporated into the BDT, not as an input variable, but by weighting the signal events by  $1/\sigma(m_{\mu\mu})$ , which is the per-event experimental dimuon mass resolution, calculated from the  $p_T$  uncertainty of the muon tracks. In this way, the BDT output encapsulates both the kinematic information and the resolution information in its output, allowing for a categorization based on a single variable, achieving a better significance with fewer categories. The resolution is not used as an direct input to the BDT because its distribution is not very different between signal and background. The weights are only applied in the training on signal events, and not applied in the evaluation of the BDT score.

The BDT is trained with a collection of simulated samples from all eras in Run 2. The training is performed in the mass window of  $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ . To make sure the BDT is not sensitive to the  $m_{\mu\mu}$  value, signal samples with different Higgs boson mass assumptions,  $m_H = 120, 125, 130 \text{ GeV}$ , are all used as signals in the training. Signal events are only used if the candidate  $\mu\mu$  pair truly originates from the Higgs boson decay, so that the BDT only picks the true kinematic signatures of the signal. No parent matching is required for backgrounds. To benefit from the maximal statistics in simulation while keeping sensitive to all kinematic features, events with  $e + \mu\mu$  and  $\mu + \mu\mu$  are used together in the training, but can be distinguished by the "number of electrons" as one of the input variables to the BDT. To increase the statistics of the nonprompt backgrounds in the simulated events, a relaxed lepMVA cut is used for the training collection than in the actual analysis, and the nonprompt yields are scaled by a factor of 0.5 to account for the increased nonprompt lepton efficiency. For both the signal and background samples, half of the events is used for the training, while the other half is used for the testing.

A receiver operating characteristic (ROC) curve illustrates the performance of binary classifiers by plotting the true positive rate against the false positive rate. The BDT output and its ROC curve are shown in Figure 6-1, in which the BDT performs the same on training and testing samples, indicating no over-training. Distributions of the BDT input variables

are shown in Figures 6-2.

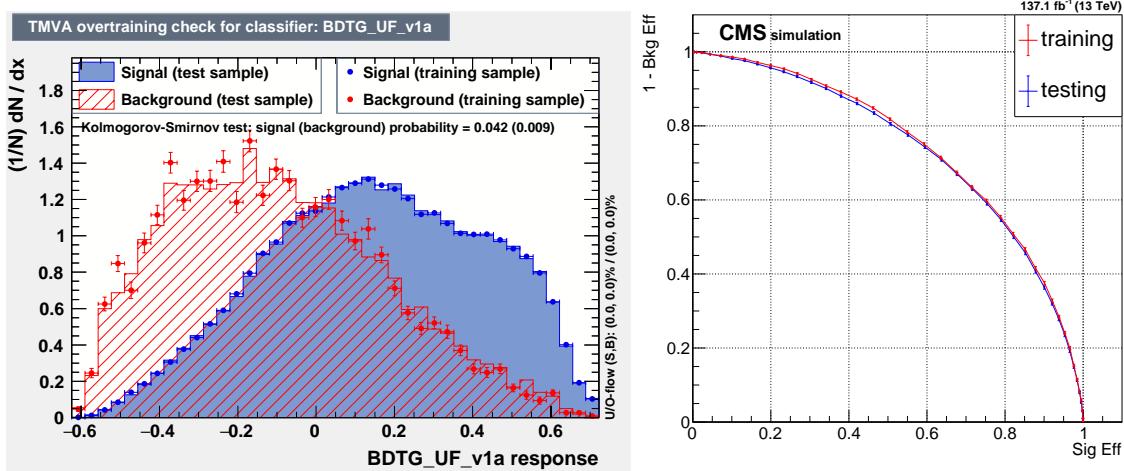


Figure 6-1. Plots of the performance of the  $\text{WH} \rightarrow 3\ell$ . On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training sample in red and testing sample in blue. A slight over-training is observed in the region of low signal efficiency, because of the fluctuation in background. As will be shown in Fig. 6-5, the BDT does not sculpt the shape of  $m_{\mu\mu}$ .

### 6.2.2 BDT targeting $\text{ZH} \rightarrow \ell\ell + \mu\mu$ signal

After the event selection of the ZH category, the background is almost purely composed of  $ZZ \rightarrow 4\ell$  and  $ggZZ \rightarrow 4\ell$  processes. Other backgrounds, prompt or nonprompt, have negligible contribution in this channel. Both ZZ and ggZZ processes have the identical final states as the ZH signal. Apart from the dimuon mass, which is used in the last stage for signal extraction, the most distinct discrimination between the signal and the background lies in the helicity angles, between the leptons from the H (Z) decay, and between the H ( $Z_1$ ) and the Z ( $Z_2$ ) bosons.

The input variables to the BDT are listed in Table 6-4 and shown in Figure 6-4, among which,  $\cos \theta^*(\mu\mu_H, \ell\ell_Z)$ , the helicity angle between the Higgs candidate and the Z candidate, is one of the most discriminating. In the ZZ background process, a propagator Z boson ( $Z_0$ ) couples to two Z bosons ( $Z_1$  and  $Z_2$ ), which in turn decay to lepton pairs. Since Z bosons are spin-1 particles, in the  $Z_0 \rightarrow Z_1 Z_2$  process, the direction of the decay

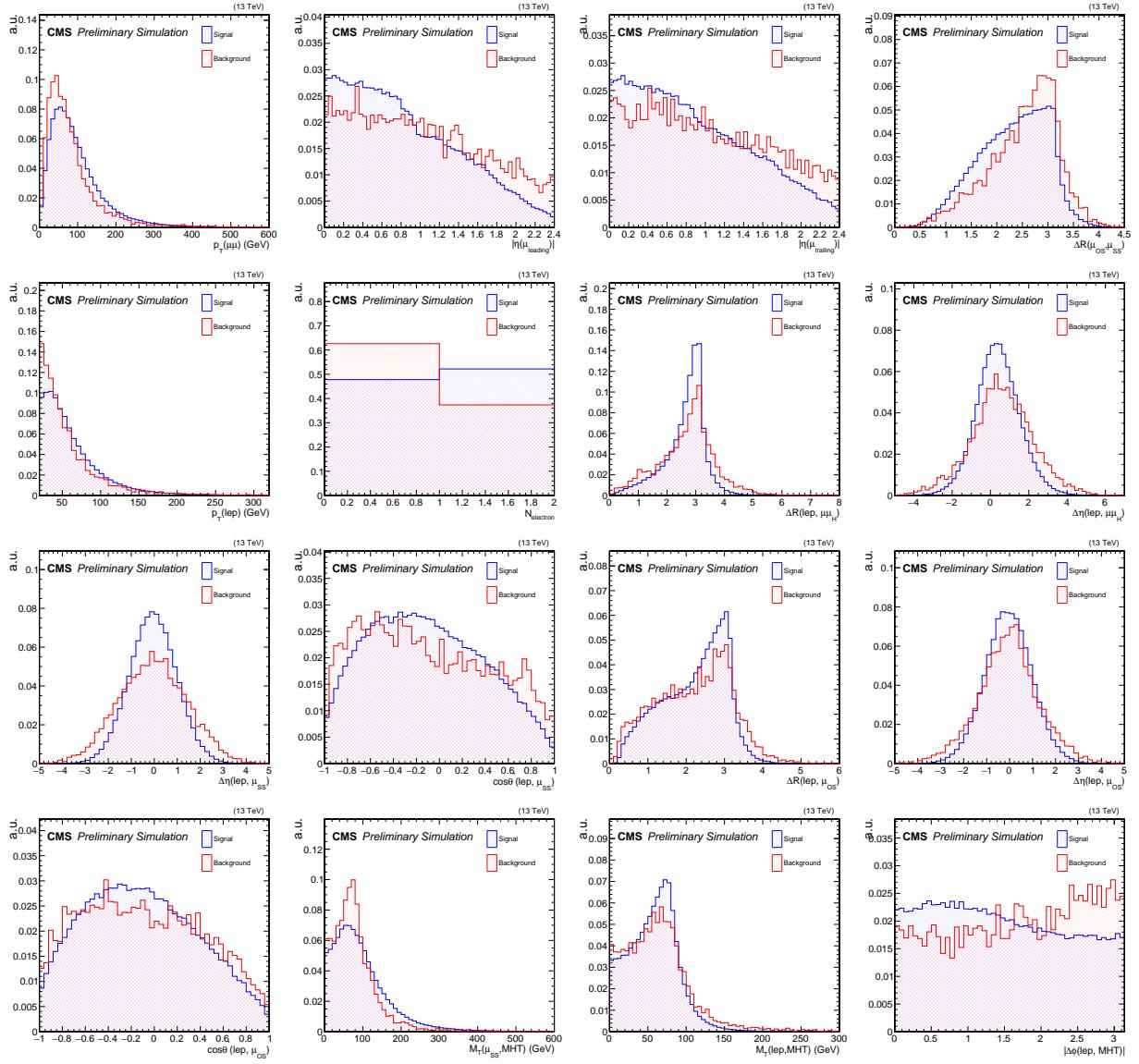


Figure 6-2. Input variables to the  $\text{WH} \rightarrow 3\ell$  BDT, with signal in blue and background in red.

is more likely to align with the direction of the momentum of  $Z_0$ . Whereas in the  $\text{ZH}$  events, since the Higgs bosons are spin-0 particles, there is no preferred direction for the the  $Z_0 \rightarrow Z_1 H$  decay. A similar kinematic discrimination is also present in the helicity angle  $\cos \theta^*(\mu_1, \mu_2)$ , between the  $Z \rightarrow \mu\mu$  decay and the  $H \rightarrow \mu\mu$  decay, where in the  $Z$  decay the muons prefer to align with the momentum of their parent and in the  $H$  decay they follow a flat distribution in  $\cos \theta^*$ . However, in this analysis, because of the acceptance of the

CMS detector, the distribution of  $\cos \theta^*(\mu_1, \mu_2)$  is sculpted and turns out not very different between signal and background. This variable is included in the initial training and later discarded during the variable trimming process.

Similar to the WH BDT training, as described in Section 6.2.1, the training is performed with simulated samples from all eras in Run2. The training is performed in the mass window of  $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ . This training was performed prior to the production of ggZH signal samples, so only qqZH samples are used as signal events. Signal samples with different Higgs boson mass assumptions,  $m_H = 120, 125, 130 \text{ GeV}$ , are used. Signal events are only used if the candidate  $\mu\mu$  pair truly originates from the Higgs boson decay. Signal events are weighted by  $1/\sigma(m_{\mu\mu}^H)$ . Events with  $ee + \mu\mu$  and  $\mu\mu + \mu\mu$  are used together in the training, but can be distinguished with the "lepton flavor" as one of the input variables. To increase the statistics of training events, a relaxed lepMVA cut is used in the training. Even so, there is no nonprompt background component passing the loosened selection.

The BDT output and the ROC curve are shown in Figure 6-3, in which the BDT performs the same on training and testing samples, indicating no over-training. Distributions of the BDT input variables are shown in Figure 6-4.

Table 6-4. List of input variables used to train the signal-background separation BDT in the ZH category. In this table,  $\mu\mu_H$  is the Higgs candidate, and  $\ell\ell_Z$  is the Z candidate.

Variable	Description
$p_T(\mu\mu_H)$	$p_T$ of the Higgs candidate
$ \eta(\mu\mu_H) $	$ \eta $ of the Higgs candidate
$ \Delta\phi(\mu\mu_H) $	$ \Delta\phi $ between the muons in the Higgs candidate
$M(\ell\ell_Z)$	invariant mass of the Z candidate
$p_T(\ell\ell_Z)$	$p_T$ of the Z candidate
$ \eta(\ell\ell_Z) $	$ \eta $ of the Z candidate
$\Delta R(\ell\ell_Z)$	$\Delta R$ between the leptons in the Z candidate
lepton flavor	flavor of the Z candidate lepton pair
$\cos \theta^*(\mu\mu_H, \ell\ell_Z)$	cosine helicity angle between the Higgs and the Z candidates
$\Delta\eta(\mu\mu_H, \ell\ell_Z)$	$\Delta\eta$ between the Higgs and the Z candidates

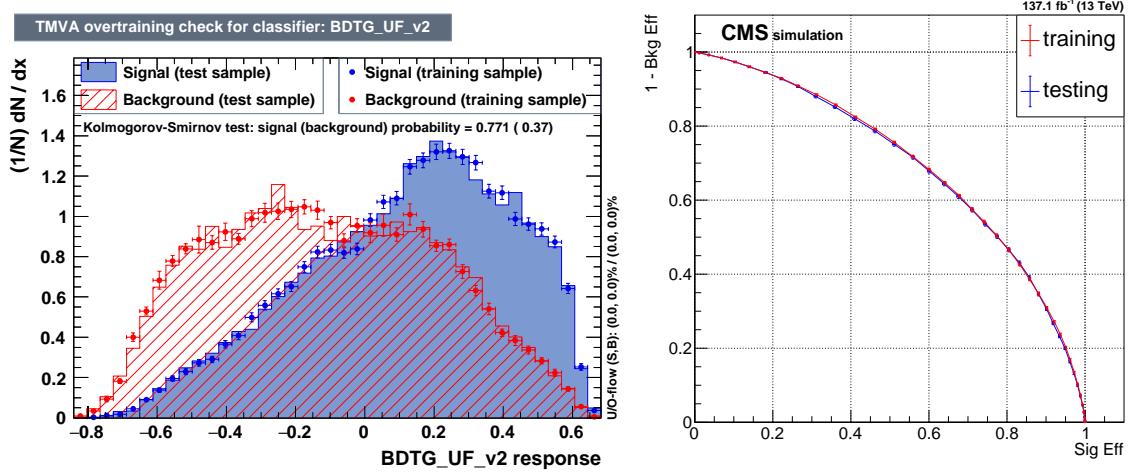


Figure 6-3. Plots of the performance of the  $ZH \rightarrow 4\ell$  BDT. On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training in red and testing in blue.

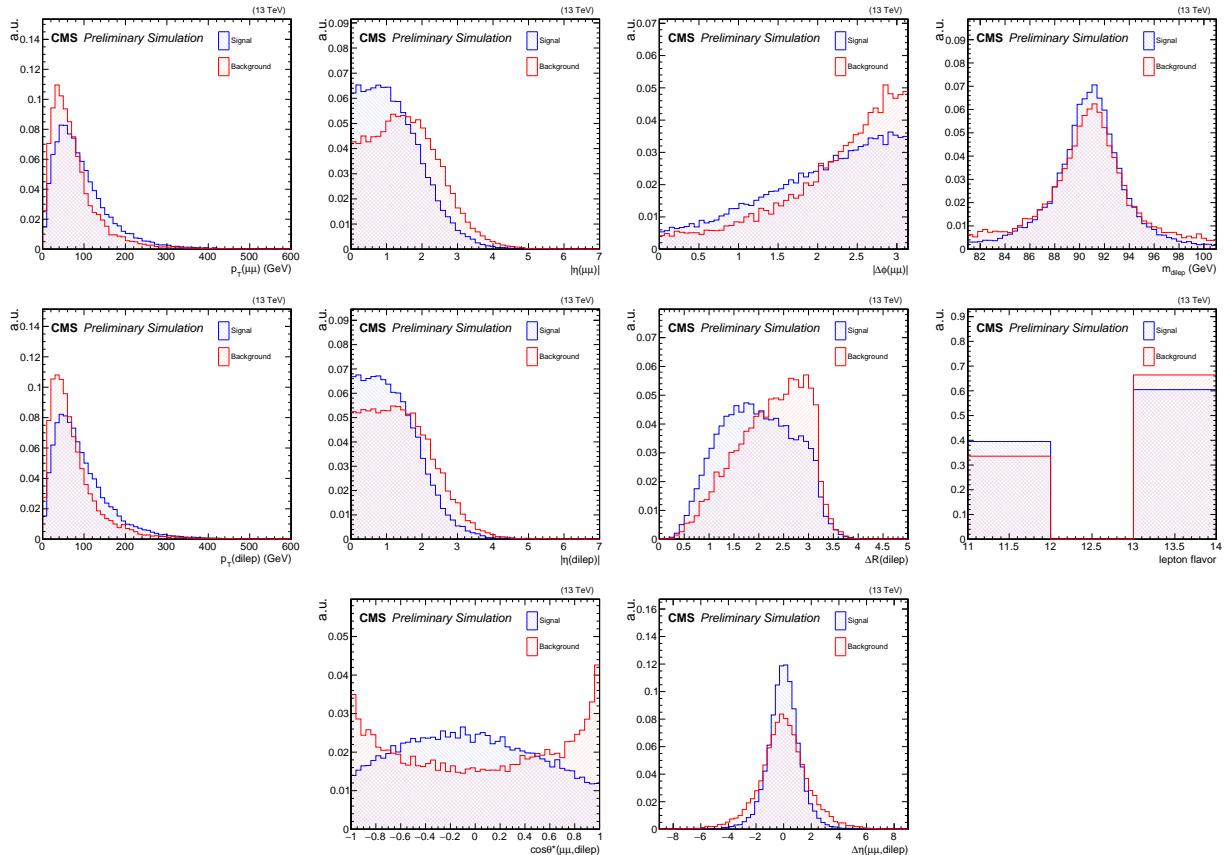


Figure 6-4. Input variables to the  $ZH \rightarrow 4\ell$  BDT, with signal in blue and background in red.

### 6.2.3 Validation of the BDTs

As discussed in 3.2, the strategy of this analysis is to divide events into subcategories with different  $S/B$ , and consequently maximize the overall sensitivity. The signal extraction is performed by fitting the  $m_{\mu\mu}$  spectrum, therefore it is crucial that any selection cut applied to the BDT score should not sculpt the  $m_{\mu\mu}$  shape. Two checks are performed for this purpose:

- The  $m_{\mu\mu}$  shape of the background is compared between events in different BDT quantiles, shown as the left plots of Figures 6-5 and 6-6.
- The BDT output is compared between several signal samples with different  $m_H$  assumptions, shown as the right plots of Figures 6-5 and 6-6.

The  $m_{\mu\mu}$  shape of the background in all BDT quantiles are smooth falling. The left plot of Figure 6-5 shows a mild dependence of slope on the BDT quantiles. But no spurious peak-like structure is seen in any of the distributions. This slope difference should not affect the signal evaluation. The BDT distributions in different signal samples agree very well, so the BDTs are not biased toward any particular  $m_{\mu\mu}$  range. Overall, the WH and ZH BDTs do not sculpt of the  $m_{\mu\mu}$  shape and can be used for categorization.

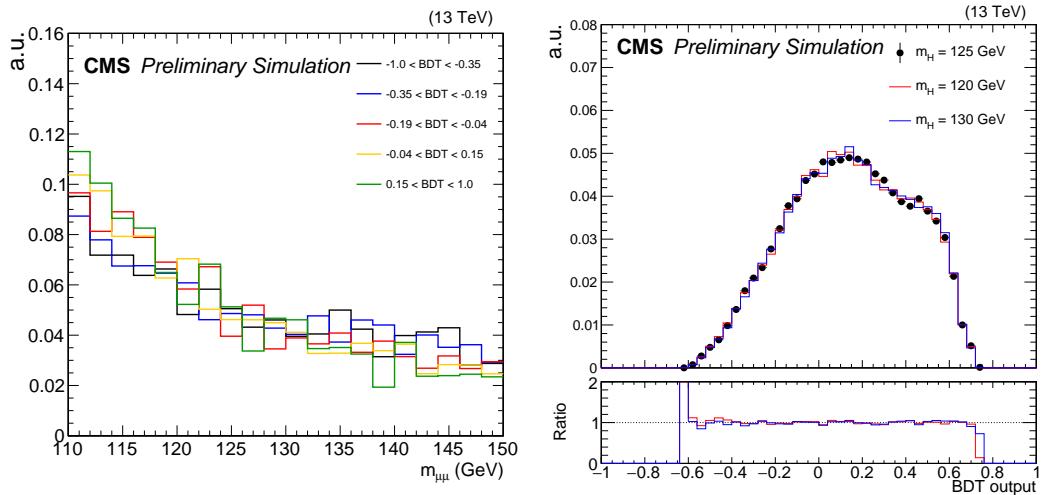


Figure 6-5. For the WH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right).

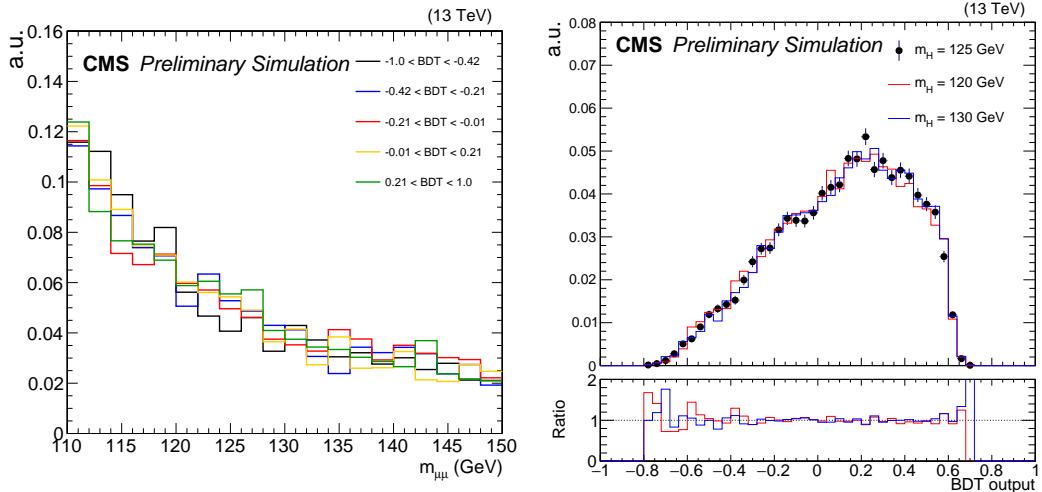


Figure 6-6. For the ZH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right).

Furthermore, it is also important to make sure the BDT would perform the same way on data as it does on simulation. In order to do this, the inputs and output of the BDTs are plotted comparing between data and the simulation. Figure 6-7 shows the output of the WH BDT and the ZH BDT, and Figure 6-8 and 6-9 show the input variables to the WH and ZH BDTs respectively. Overall, data and simulation agree with each other within the uncertainties for the BDT outputs and most of the inputs. Some fluctuations are seen in data, especially in the ZH category, as the total number of events is small. These fluctuations are expected within the statistical uncertainty and do not indicate any systematic disagreement.

An arguable disagreement is seen in the leftmost region in one of the inputs to the WH BDT,  $M_T(\mu_{SS}, \text{MHT})$ , shown as the second plot in the bottom row of Figure 6-8, also put separately in Figure 6-10. To understand if this disagreement would translate into a mismodeling of the WH BDT, the BDT output is plotted for both signal and background in different  $M_T(\mu_{SS}, \text{MHT})$  bins, shown as the right plot in Figure 6-10. In this plot, for both signal and background, the BDT profile is almost the same for events with  $M_T(\mu_{SS}, \text{MHT}) < 40$  GeV and events with  $40 < M_T(\mu_{SS}, \text{MHT}) < 80$  GeV, while it is different between

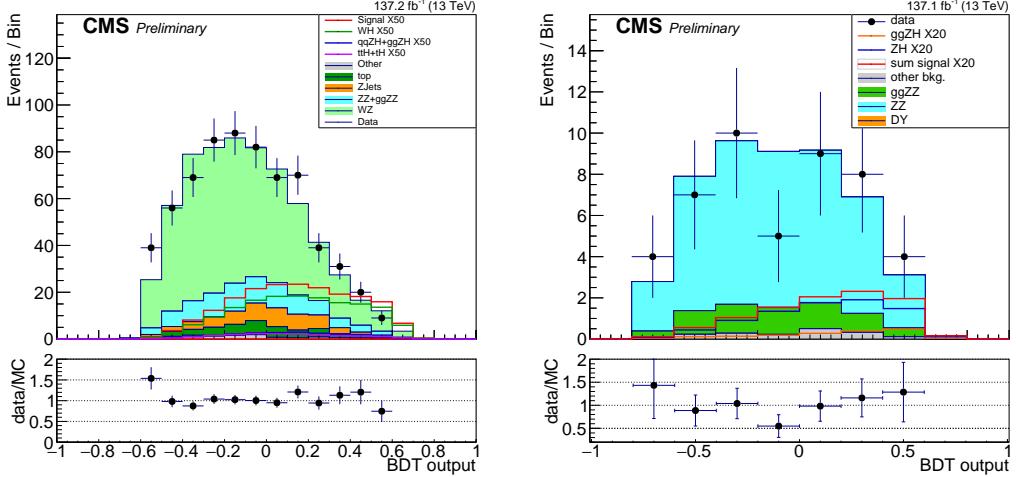


Figure 6-7. The WH BDT output (left) and the ZH BDT output (right) in full Run 2 in the signal region  $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ .

events with  $M_T(\mu_{SS}, \text{MHT}) < 80 \text{ GeV}$  from events with  $M_T(\mu_{SS}, \text{MHT}) > 80 \text{ GeV}$ . The BDT is sensitive to whether the  $M_T(\mu_{SS}, \text{MHT})$  is greater or smaller than 80 GeV, but does not further distinguish events if the  $M_T(\mu_{SS}, \text{MHT})$  is less than 80 GeV. Once the bins below 80 GeV are merged in the left plot of Figure 6-10, there is no significant disagreement, therefore it should not cause any mismodeling of the BDT.

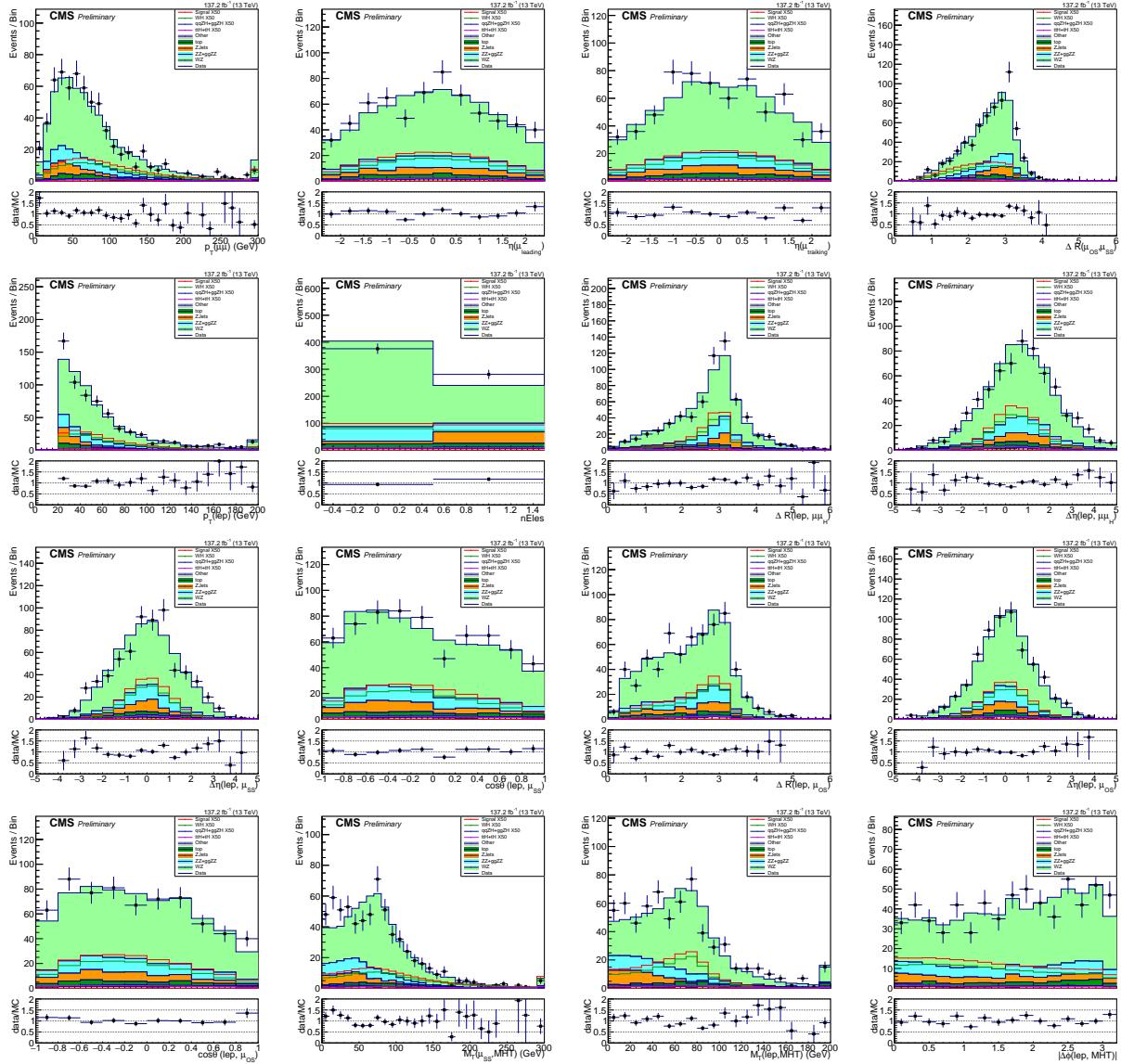


Figure 6-8. Input variables to the WH BDT in full Run 2 in the signal region  $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ .

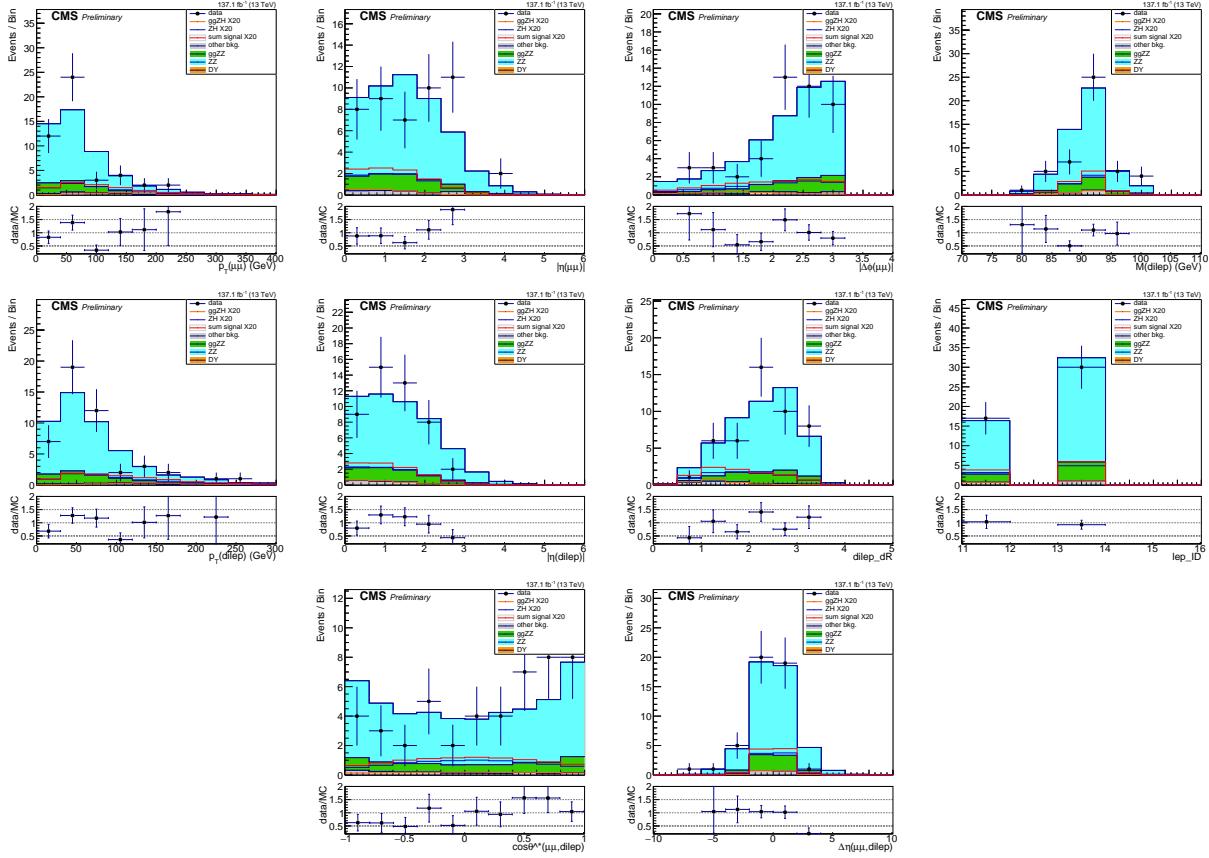


Figure 6-9. Input variables to the ZH BDT in full Run 2 in the signal region  $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ .

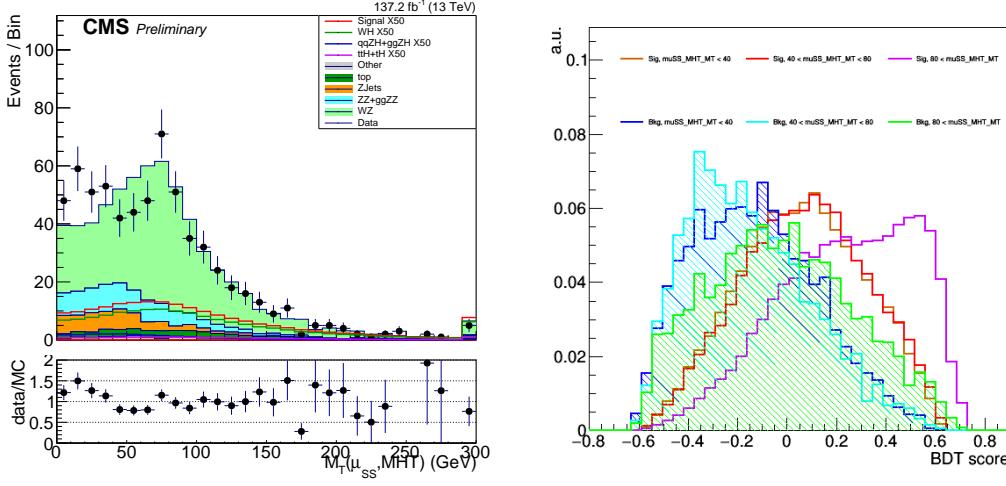


Figure 6-10. The input variable  $M_T(\mu_{SS}, MHT)$  to the WH BDT (left), and the BDT output for signal and background in different  $M_T(\mu_{SS}, MHT)$  bins (right). A mild disagreement is seen between the simulation and data in the low bins of  $M_T(\mu_{SS}, MHT)$ , while the BDT is not sensitive to the  $M_T(\mu_{SS}, MHT)$  values in that region.

### 6.3 Event Categorization

To optimize the overall sensitivity of the VH analyses, the WH and the ZH phase-spaces are divided into several subcategories with different S/B ratios, based on the BDT discriminants described in Section 6.2.

To achieve the maximal sensitivity with a reasonable number of subcategories, an iterative procedure is taken. In each iteration, a cut is scanned at a step of 0.01 of the BDT value and the sum of the significance of the resulting subcategories is calculated as the figure of merit. The figure of merit is defined as the  $S/\sqrt{B}$  in each subcategory summed in quadrature, where S and B represent the expected signal and background yields within the full width at half maximum (FWHM) of the signal peak in each subcategory. In addition, to ensure that there are enough events in each subcategory to perform a shape analysis, all subcategories have to meet a minimal total event yield requirement during the BDT scanning process. The choice of this minimal requirement is somewhat arbitrary, and is determined based on the experience with the subsequent "bias studies" described in Section 6.5.3. This number is chosen to be 30 in the WH BDT scans, and 16 in the ZH BDT scans, as the total number of events in the ZH category is less than 50.

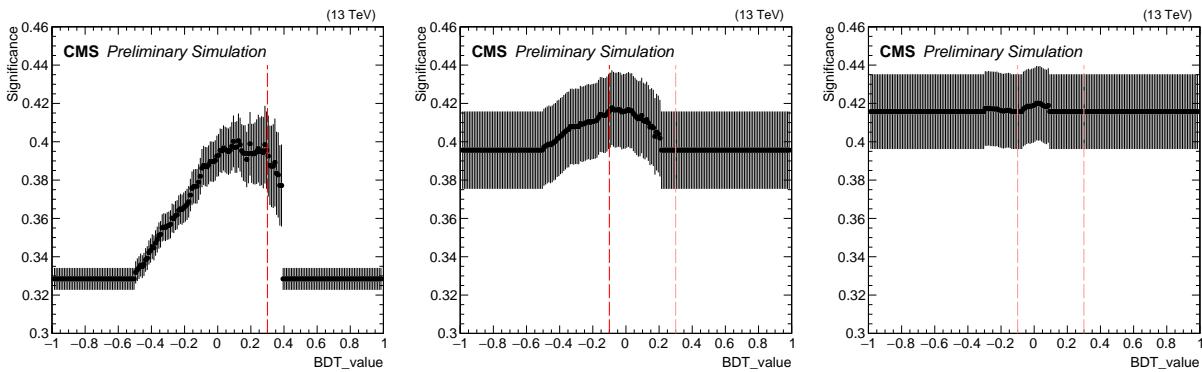


Figure 6-11. Scans for the first (left), second (middle) and a potential third (right) BDT cut in the WH channel. The first BDT cut is chosen at 0.3. The second BDT cut is chosen at -0.1. A third BDT cut is not necessary.

Figure 6-11 shows the iterations performed on the WH BDT. In the first scan, the overall significance maximizes around  $0 \sim 0.3$ , and the cut is chosen at 0.3 so that there are

enough events on its left side for a second cut. In the second scan, the overall significance maximizes around  $-0.1 \sim 0.05$ , and the position of the second cut can be any value in this range. To help decide the second cut, several third scans are performed under different assumptions of the second cut, all showing negligible changes of the overall significance (similar to the right plot in Figure 6-11). Therefore, there is no need for a third cut, and the choice for the second cut can be somewhat arbitrary. The second cut is decided at -0.1 so that there are a good number of events in the middle subcategory to ensure a stable shape analysis. As a result, two BDT boundaries are set, dividing the WH phase-space into 3 subcategories, BDT within  $[-1.0, -0.1]$ , BDT within  $[-0.1, 0.3]$  and BDT within  $[0.3, 1.0]$ .

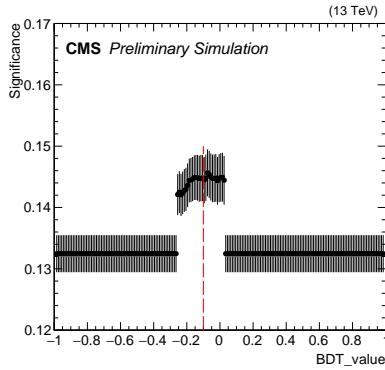


Figure 6-12. Scans for the BDT cut in the ZH channel. The BDT cut is chosen at -0.1.

Similarly, Figure 6-12 shows the scan performed on the ZH BDT. In the BDT scan, the overall significance maximizes around  $-0.15 \sim 0.05$ . The BDT cut is chosen at -0.1, dividing the ZH events into two roughly equal halves. A second cut is not needed as the number of events is not enough for a further division. The resulting 2 ZH subcategories are, BDT within  $[-1.0, -0.1]$ , and BDT within  $[-0.1, 1.0]$ .

#### 6.4 Signal and background modeling

The extraction of signal is performed by fitting analytic functions to the  $m_{\mu\mu}$  spectrum in each subcategory. Different functions are used to model the expected signal and background shapes: a sharp signal peak near 125 GeV, and a smooth falling background shape in  $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$ . Functions are first tested on the simulated samples to

make sure the perform well in describing the shapes. The parameters of the signal function are constrained, with systematic uncertainties described in Section 6.5.1, to the best fit values to simulation as the expectation of the SM signal. The parameters of the background function are allowed to float freely, so that no prior assumption on the background is imposed, and the background prediction relies completely on data. The final evaluation of the signal strength is achieved by fitting the signal + background functions to data, where the normalizations of both the signal function and the background function are allowed to float freely. The normalization of the signal, in particular, is called the signal strength modifier and represents the signal strength relative to the SM prediction.

Function modeling of signal and background are described in Sections 6.4.1 and 6.4.2 respectively.

#### 6.4.1 Signal modeling

In all subcategories, signals are modeled independently by different production modes, with the contributions from three years (2016, 2017, 2018) summed together. In particular, qqZH and ggZH signals are modeled separately as there are no other signal component in the ZH category. Each of the components is modeled with a Double-sided Crystal Ball function (DCB), as described in Equation 6-1. In all DCB functions, the parameters  $n_L$  and  $n_R$  are fixed to 2.0, since they only affect the shape in tails and can take values in a large range without changing the quality of the fit by much. Other parameters are allowed to float freely.

$$\text{DCB}(m_{\mu\mu}) = \begin{cases} e^{-(m_{\mu\mu}-s)^2/(2\sigma^2)} & -\alpha_L < (m_{\mu\mu}-s)/\sigma < \alpha_R \\ (\frac{n_L}{|\alpha_L|})^{n_L} \times e^{-\alpha_L^2/2} \times (\frac{n_L}{|\alpha_L|} - |\alpha_L| - (m_{\mu\mu}-s)/\sigma)^{-n_L} & (m_{\mu\mu}-s)/\sigma \leq -\alpha_L \\ (\frac{n_R}{|\alpha_R|})^{n_R} \times e^{-\alpha_R^2/2} \times (\frac{n_R}{|\alpha_R|} - |\alpha_R| + (m_{\mu\mu}-s)/\sigma)^{-n_R} & (m_{\mu\mu}-s)/\sigma \geq \alpha_R \end{cases} \quad (6-1)$$

Examples of signal modeling are shown in Figures 6-13 and 6-14. Please note that the plots shown are the signals in the inclusive WH and ZH categories. The actual models

used in each subcategory are slightly different. ggH, VBF and bbH have negligible contributions to the WH category and are not considered. Similarly, in the ZH category only qqZH and ggZH are considered since all other contributions are negligible.

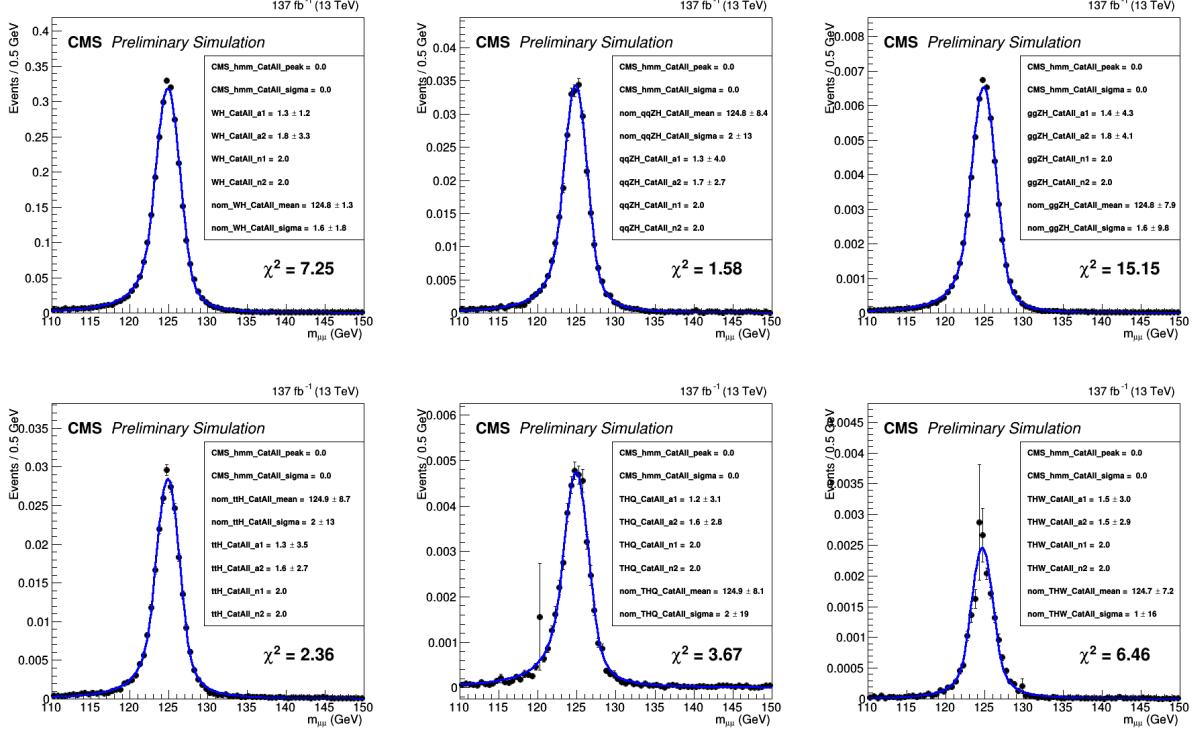


Figure 6-13. The signal modeling in the  $WH \rightarrow \ell + \mu\mu$  inclusive category. Considered signal modes are WH (top left), qqZH (top middle), ggZH (top right), ttH (bottom left), THQ (bottom middle), and THW (bottom right).

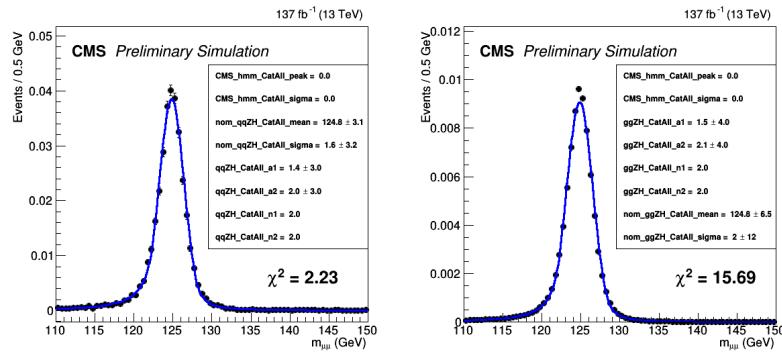


Figure 6-14. The signal modeling in the  $ZH \rightarrow \ell\ell + \mu\mu$  inclusive category. Considered signals modes are qqZH (left) and ggZH (right).

### 6.4.2 Background modeling

As discussed in Section 6.1, the main background in WH (ZH) category is the WZ (ZZ) process, a fraction of which consists wrong pairing of the muons. The  $m_{\mu\mu}$  spectrum of the correctly paired WZ (ZZ) events follows a Breit-Wigner tail of the Z boson, while the spectrum of the wrongly paired WZ (ZZ) events is rather flat. Overall the background shapes in the WH and ZH categories are smoothly falling, and Breit-Wigner like to some extent. Therefore, a group of different functional forms are considered as candidates for the background modeling. Some of them are physics-inspired, meaning that they are modified from the form of the Breit-Wigner function, the others are agnostic that take the form of some general functional bases. The physics-inspired function candidates include the *BWZ* function (Equation 6-2), which is a Breit-Wigner core times an exponential term , the *BWZRedux* (Equation 6-3), which is a Breit-Wigner core times a exponential term with more degrees of freedom (DOFs), the *BWZGamma* (Equation 6-4), which is a linear combination of the *BWZ* function and an exponential function, and the *BWZ × Bernstein* (Equation 6-5), which is the *BWZ* function times a Bernstein polynomial.

$$\text{BWZ}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \quad (6-2)$$

$$\text{BWZRedux}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu} + b \cdot m_{\mu\mu}^2}}{(m_{\mu\mu} - m_Z)^c + (\Gamma_Z/2)^c} \quad (6-3)$$

$$\text{BWZGamma}(m_{\mu\mu}) = f \cdot \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} + (1 - f) \cdot \frac{e^{a \cdot m_{\mu\mu}}}{m_{\mu\mu}^2} \quad (6-4)$$

$$\text{BWZ} \times \text{Bernstein}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \times \text{Bern}_n(m_{\mu\mu}) \quad (6-5)$$

The agnostic function candidates include the Bernstein polynomials (Equation 6-6), a series of exponential functions (Equation 6-7), and a series of power functions (Equation 6-8). In the actual fits, given the low statistics in the VH subcategories, the sum

of exponential or power functions are usually reduced to a single exponential or power function plus a constant (Equation 6-9 and 6-10).

$$\text{Bernstein}(m_{\mu\mu}) = \sum_i^n a_i \cdot \binom{n}{i} m_{\mu\mu}^i (1 - m_{\mu\mu})^{n-i} \quad (6-6)$$

$$\text{S-exponential}(m_{\mu\mu}) = \sum_i^n a_i \cdot e^{b_i \cdot m_{\mu\mu}} \quad (6-7)$$

$$\text{S-power-law}(m_{\mu\mu}) = \sum_i^n a_i \cdot m_{\mu\mu}^{b_i} \quad (6-8)$$

$$\text{Exponential+constant}(m_{\mu\mu}) = f + (1 - f) \times e^{a \cdot m_{\mu\mu}} \quad (6-9)$$

$$\text{Power-law+constant}(m_{\mu\mu}) = f + (1 - f) \times m_{\mu\mu}^a \quad (6-10)$$

In each subcategory, the function candidates are fit to the  $m_{\mu\mu}$  shape in the range of  $110 < m_{\mu\mu} < 150$  GeV, with events blinded in the signal region  $120 < m_{\mu\mu} < 130$  GeV, so the functions are not aware of the existence of any signal. Because the limited number of events in VH subcategories, the distribution of data is subject to large fluctuations. The  $m_{\mu\mu}$  shape of data reflects both the underlying physics shape, as well as the specific features from the fluctuation of this particular dataset. It is important to make sure the modeling of background does not over-fit these specific features. On the other hand, as shown in Section 6.1 and 6.2, the simulated samples are known to provide a good modeling of data, and the  $m_{\mu\mu}$  shape from the simulation can be assumed to be a good representation of the true physics shape. Therefore, the simulation can be used to study the performance of the background function candidates to learn how they would model generic expected physics shapes, and the data is treated as a particular realization of these physics distributions. The fit to simulation takes the  $m_{\mu\mu}$  shape of the simulation, but assumes the statistical error in each bin as the Poisson error of the expected number of events in that bin rather than

the number of simulated events in the sample. If a function candidate provides a good fit to the simulation, it is then tested on the real dataset, to make sure the fit does not break down because of the fluctuation. If the function gives consistently good fit performances on the simulation and data, it is considered as a good candidate. It is worth a remark that the fit to data does not assume any parameter information from the fit to simulation, so the simulation is only used to study the performance of the background functions, but not used to constrain the specific shapes.

All the functional forms listed above can be used with different DOFs. For the physics-inspired functions, parameters  $m_Z$  and  $\Gamma_Z$  can either be fixed at the nominal value for the Z boson or allowed to float freely, while for the agnostic functions, the order of the series can be adjusted. To find the right DOFs, each functional form is tested with different setups, and the optimal DOFs is determined following the idea of the likelihood ratio test. A standard likelihood ratio test compares the likelihood ratio between the fits with  $n$  and  $n+1$  DOFs, usually calculated as  $2(\text{Log}\mathcal{L}_{n+1} - \text{Log}\mathcal{L}_n)$ , where the  $\text{Log}\mathcal{L}_n$  is the likelihood of the fit with  $n$  DOFs. This quantity should follow the  $\chi^2_1$  distribution, the chi-square distribution with one degree of freedom, whose p-value is then used to decide whether adding one more DOF in the fit leads to a significantly better fit quality.

In the practice of background fitting in the VH subcategories, which all have low expected number of events, it turns out in most cases two DOFs are enough, one for overall normalization and one for shape variation. In some subcategories with very low statistics, even functions without any shape DOF give good performances, namely, a function with all its shape parameters fixed at the best fit values to the simulation can be a good fit to data. These fixed shapes are included as some candidates along side with their freely floating versions. On the other hand, when the shape parameters are allowed to float, agnostic functions with low DOFs do not always fit well as they lack enough flexibility. For example, an order-1 Bernstein polynomial (2 shape DOFs) is just a straight line and is obviously not the true  $m_{\mu\mu}$  shape, and the fits with a single exponential or power function are not stable

and sometimes do not converge. To mitigate these behaviors, a BWZ  $\times$  order-1 Bernstein (2 shape DOF in total), in which the BWZ part is fixed with the nominal Z boson shape, is used instead of the plain Bernstein, and a single exponential (power) function plus a free constant (2 shape DOF in total) is used instead of the plain exponential (power) function. Overall, the good function candidates include fixed forms (1 normalization DOF + 0 shape DOF) and floating forms (1 normalization DOF + 1  $\sim$ 2 shape DOFs).

The final choices of the background function in each subcategory is decided based on the bias it may have against other possibilities, described in detail in Section 6.5.3. If several functions pass the bias requirement, the function with the fewest DOF is chosen, as it leads to the highest significance in statistical analysis.

## 6.5 Systematic uncertainties

A crucial task in the statistical analysis is to evaluate all the systematic uncertainties that affect the signal and background estimation. In this analysis, both the signal and background are described by analytic functions, and the statistical analysis is performed based on the fits of them. All sources of systematic uncertainties are therefore translated into the variations of the parameters of signal and background functions.

Several sources of signal systematic uncertainties are considered, divided into two types, the *shape* and the *rate* uncertainties. The *shape* uncertainties, described in Section 6.5.1, account for the factors affecting the expected shape of the signal peak, while the *rate* uncertainties, described in Section 6.5.2, are those affecting the expected signal yield.

A different approach is taken to evaluate the systematic uncertainty in background. The background estimation always takes the best fit to data and does not rely on simulation, therefore none of the theoretical or experimental uncertainties considered for the signal needs to be considered for the background. However, by fitting the background shape with an analytical function, a potential bias could be introduced between the chosen background model and the underlying real distribution. A bias between the background

estimation and the true background appearing at the position of the signal is essentially a spurious signal. This bias has to be small so that it does not impact the validity of the signal strength evaluation. The study to evaluate this potential bias is described in details in Section 6.5.3.

### 6.5.1 Signal shape uncertainties

For all Higgs boson production modes, the expected  $m_{\mu\mu}$  signal shape is primarily affected by the uncertainties in muon energy scale and resolution, in other words the mean and sigma values in the DCB fits of the signal peak. As described in Chapter 5, the *Rochester correction* is implemented to correct for differences in both scale and resolution between data and simulated events, while the *FSR recovery* and *GeoFit correction* are not expected to introduce new differences between data and simulation. In the meantime, Section 5.4 shows that the simulation of the DY peak agrees with the data up to a per-mille level in scale and a percent level in resolution. The shape uncertainties can be estimated accordingly.

The muon energy scale shape uncertainty is estimated to be 0.1% of the mean value of the  $m_{\mu\mu}$  peak, and the muon energy resolution uncertainty is conservatively estimated to be 10% of the resolution of the  $m_{\mu\mu}$  peak. The effect of the scale uncertainty is an overall shift of the  $m_{\mu\mu}$  peak to higher or lower mass value, while the effect of the resolution uncertainty is a stretching or squeezing of the width of the  $m_{\mu\mu}$  peak. Both uncertainties are modeled as a Gaussian constrained nuisance parameter that is correlated across different production modes but uncorrelated between different subcategories.

### 6.5.2 Signal rate uncertainties

The rate uncertainties are the ones that affect the signal yield in each subcategory, and may come from various sources. Some of them affect the overall prediction of the signal and act as a factor on the overall normalization of the signal. The normalization uncertainties include the theoretical uncertainties on the cross sections of signal productions, and theoretical uncertainties on the  $\mathcal{B}(H \rightarrow \mu\mu)$ , as well as the uncertainties on

the CMS luminosity measurement. Other uncertainties tweak the event kinematics and affect the acceptance of signals in each subcategory. The acceptance uncertainties include the uncertainties on all the event weights in the simulation, the uncertainties from all the efficiency scale factors applied in the analysis, and the uncertainties from all the physics object calibrations and corrections.

The impacts from theoretical uncertainties are shown in Table 6-5. The uncertainty on the  $\mathcal{B}(H \rightarrow \mu\mu)$  is  $\pm 1.23\%$ , independent from the production modes. The luminosity uncertainty for each year is set following the official recommendation of CMS, which is 2.5%, 2.3% and 2.5% for 2016, 2017 and 2018 respectively. Since the signals are modeled summing all years, the luminosity uncertainty in each year reflected in the overall signal yield is 0.7%, 0.7% and 1.1%, for 2016, 2017, and 2018, respectively.

Table 6-5. Normalization uncertainties on the Higgs boson production cross sections for various modes at  $\sqrt{s} = 13\text{TeV}$ .

Process	Perturbative Order	+QCD scale unc. (%)	-QCD scale unc. (%)	+ (PDF + $\alpha_s$ ) unc. (%)	- (PDF + $\alpha_s$ ) unc. (%)
WH	NNLO (QCD)	+0.5	-0.7	+1.9	-1.9
	NLO (EWK)				
qqZH	NNLO (QCD)	+0.5	-0.6	+1.9	-1.9
	NLO (EWK)				
ggZH	NLO (QCD)	+25.1	-18.9	+2.4	-2.4
$t\bar{t}H$	NLO (QCD)	+5.8	-9.2	+3.6	-3.6
	NLO (EWK)				
tHq	NLO (QCD)	+6.5	-14.9	+3.7	-3.7
tHW	NLO (QCD)	+4.9	-6.7	+6.3	-6.3

The impacts from the pileup reweight and ECAL L1 trigger prefiring reweight are shown in Table 6-6. The acceptance impacts from the muon energy scale corrections are shown in Table 6-7. The impacts from the muon and electron ID scale factors (the lepMVA scale factor) are shown in Table 6-8. The impacts from the b-jet ID scale factors (for the b-jet vetoing) are shown in Table 6-9. The impacts from the jet energy calibrations are shown in Table 6-10.

Table 6-6. Uncertainties on different signal components in the WH and ZH channels related to pileup reweight and L1 prefiring reweight. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty	Category	WH	qqZH	ggZH	ttH	THQ	THW
pileup 2016 (%)	WH cat1	+0.8/-0.7	+0.8/-0.7	+0.6/-0.5	+0.9/-0.7	+1.0/-0.9	+0.4/-0.4
	WH cat2	+0.8/-0.7	+0.6/-0.5	+0.6/-0.5	+0.6/-0.5	+1.0/-0.9	+0.7/-0.6
	WH cat3	+0.6/-0.5	+0.4/-0.3	+0.6/-0.4	+0.7/-0.6	+0.2/-0.3	+0.5/-0.4
	ZH cat1	-	+0.7/-0.7	+0.8/-0.7	-	-	-
	ZH cat2	-	+0.8/-0.7	+0.7/-0.6	-	-	-
pileup 2017 (%)	WH cat1	+0.6/-0.5	+0.2/-0.2	+0.3/-0.3	+0.3/-0.2	+0.3/-0.5	+0.6/-0.7
	WH cat2	+0.4/-0.4	+0.4/-0.4	+0.3/-0.3	+0.3/-0.3	+0.4/-0.5	+0.3/-0.4
	WH cat3	+0.5/-0.5	+0.5/-0.3	+0.3/-0.3	+0.6/-0.6	+0.5/-0.5	+0.4/-0.2
	ZH cat1	-	+0.4/-0.4	+0.4/-0.5	-	-	-
	ZH cat2	-	+0.3/-0.4	+0.4/-0.4	-	-	-
pileup 2018 (%)	WH cat1	+0.6/-0.6	+0.4/-0.4	+0.5/-0.5	+0.6/-0.6	+0.5/-0.5	+0.5/-0.5
	WH cat2	+0.5/-0.5	+0.3/-0.3	+0.4/-0.4	+0.5/-0.5	+0.4/-0.4	+0.8/-0.8
	WH cat3	+0.4/-0.4	+0.2/-0.3	+0.4/-0.4	+0.4/-0.3	+0.5/-0.5	+0.8/-0.8
	ZH cat1	-	+0.6/-0.6	+0.5/-0.5	-	-	-
	ZH cat2	-	+0.6/-0.6	+0.5/-0.5	-	-	-
prefire 2016 (%)	WH cat1	+0.1/-0.1	+0.1/-0.1	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2
	WH cat2	+0.1/-0.1	+0.1/-0.1	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.1/-0.1
	WH cat3	-	+0.1/-0.1	+0.1/-0.1	+0.1/-0.1	+0.2/-0.2	+0.1/-0.1
	ZH cat1	-	+0.1/-0.1	+0.1/-0.1	-	-	-
	ZH cat2	-	+0.1/-0.1	+0.1/-0.1	-	-	-
prefire 2017 (%)	WH cat1	+0.2/-0.2	+0.3/-0.3	+0.3/-0.3	+0.4/-0.4	+0.3/-0.3	+0.2/-0.3
	WH cat2	+0.1/-0.1	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3	+0.3/-0.4	+0.2/-0.2
	WH cat3	+0.1/-0.1	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.1/-0.1
	ZH cat1	-	+0.2/-0.2	+0.3/-0.3	-	-	-
	ZH cat2	-	+0.1/-0.1	+0.2/-0.2	-	-	-

Table 6-7. Uncertainties on different signal components in the WH and ZH channels related to the muon energy scale. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty (%)	WH	qqZH	ggZH	ttH	THQ	THW
WH cat1	-	+0.0/+0.1	-	-0.1/+0.3	+0.0/-0.1	+0.0/+0.3
WH cat2	-	-0.1/-0.0	+0.1/-0.1	+0.1/-0.0	-0.1/+0.2	-0.1/-0.2
WH cat3	+0.0/+0.1	-0.2/-0.0	-0.2/+0.1	-0.2/-0.0	+0.0/+0.4	+0.1/-0.0
ZH cat1	-	+0.0/+0.1	+0.0/+0.1	-	-	-
ZH cat2	-	-0.1/-0	-	-	-	-

Table 6-8. Uncertainties on different signal components in the WH and ZH channels related to lepMVA scale factor. The lepMVA scale factor is the only scale factor applied to correct for the lepton efficiency modeling. The ID scale factor and Isolation scale factor are covered by the lepMVA scale factors. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty	Category	WH	qqZH	ggZH	ttH	THQ	THW
muon SF (%)	WH cat1	-1.8/+1.8	-1.7/+1.8	-2.3/+2.3	-2.3/+2.3	-2.1/+2.1	-2.5/+2.5
	WH cat2	-1.7/+1.8	-1.7/+1.7	-2.4/+2.4	-2.0/+2.0	-1.9/+2.0	-2.3/+2.3
	WH cat3	-2.3/+2.3	-1.9/+2.0	-2.5/+2.5	-2.5/+2.5	-2.3/+2.4	-2.8/+2.9
	ZH cat1	-	-1.9/+1.9	-2.5/+2.6	-	-	-
	ZH cat2	-	-2.5/+2.6	-3.3/+3.4	-	-	-
electron SF (%)	WH cat1	-0.3/+0.3	-0.4/+0.4	-0.4/+0.4	-0.3/+0.3	-0.4/+0.4	-0.2/+0.2
	WH cat2	-0.5/+0.5	-0.6/+0.6	-0.5/+0.5	-0.5/+0.5	-0.6/+0.6	-0.5/+0.5
	WH cat3	-0.5/+0.5	-0.6/+0.6	-0.6/+0.6	-0.5/+0.5	-0.6/+0.6	-0.5/+0.5
	ZH cat1	-	-0.6/+0.6	-0.6/+0.6	-	-	-
	ZH cat2	-	-0.8/+0.8	-0.6/+0.6	-	-	-

Table 6-9. Uncertainties on different signal components in the WH and ZH channels related to b-jet vetoing. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty (%)	WH	qqZH	ggZH	ttH	THQ	THW
WH cat1	+0.1/-0.1	+0.1/-0.1	-0.8/+0.8	+5.5/-5.3	-0.9/+0.9	-1.4/+1.4
WH cat2	+0.1/-0.1	+0.1/-0.1	-0.9/+0.9	+5.7/-5.5	-0.8/+0.8	-1.3/+1.3
WH cat3	+0.1/-0.1	+0.0/-0.1	-1.0/+1.0	+5.3/-5.1	-0.9/+0.9	-1.4/+1.4
ZH cat1	-	+0.1/-0.1	-0.6/+0.6	-	-	-
ZH cat2	-	+0.2/-0.2	-0.5/+0.5	-	-	-

Table 6-10. Uncertainties on different signal components in the WH Cat1 related to jet energy calibration. JEC uncertainties are in general small for the main signals in the WH and ZH channels. WH Cat1 is shown as an example. Numbers in other categories are similar. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty (%)	WH	qqZH	ggZH	ttH	THQ	THW
flavorQCD	+0.1/-0.0	-0.4/-0.0	+0.1/-0.0	+0.6/-0.9	+0.2/-0.5	+1.5/-0.8
relativeBal	+0.1/-0.0	-0.2/-0.0	+0.1/-0.0	+0.7/-0.5	+0.0/-0.2	+0.7/-0.5
absolute	-	+0.0/+0.1	-	-	+0.0/+0.1	+0.1/+0.4
BBEC1	+0.1/-0.0	-0.4/-0.0	+0.1/-0.1	+1.5/-1.1	+0.2/-0.7	+1.2/-0.7
EC2	-	-	-	-0.1/-0.0	-0.1/-0.3	+0.2/+0.1
HF	-0.1/+0.1	-0.2/-0.0	+0.1/+0.1	+0.5/-0.6	+0.0/-0.5	+0.5/+0.3
relativeSample_2016	-	-0.1/-0.0	-	+0.1/-0.3	-	+0.1/-0.3
absolute_2016	-	+0.0/-0.1	-	+0.1/-0.1	+0.0/+0.1	-
BBEC1_2016	-	+0.0/-0.1	-	+0.0/-0.1	+0.0/+0.1	+0.1/-0.0
EC2_2016	-	-	-	-	+0.0/+0.1	+0.1/-0.0
HF_2016	-	-	-	-	-	+0.1/-0.0
relativeSample_2017	-	-0.1/+0.2	+0.1/-0.0	+0.1/-0.2	-0.2/+0.1	+0.3/+0.3
absolute_2017	-	-0.1/+0.2	-	+0.2/-0.1	-	+0.3/+0.2
BBEC1_2017	-	-0.1/+0.1	-	+0.2/+0.0	-0.1/+0.1	+0.2/+0.0
EC2_2017	-	+0.1/-0.0	-	-	+0.1/+0.3	+0.1/+0.3
HF_2017	-	-	-	-	-0.1/-0.2	-
relativeSample_2018	-0.1/+0.1	+0.1/+0.2	+0.1/-0.0	+0.4/-0.5	+0.1/-0.4	+0.7/-0.7
absolute_2018	+0.0/+0.1	-0.2/-0.0	-	+0.0/-0.1	+0.2/-0.4	+0.3/-0.6
BBEC1_2018	+0.0/+0.1	-0.2/-0.0	-	+0.0/-0.1	-	+0.1/-0.1
EC2_2018	+0.0/+0.1	-0.1/+0.1	-	+0.0/+0.1	+0.0/+0.1	+0.0/+0.1
HF_2018	-	+0.0/-0.1	-	+0.0/+0.1	+0.0/-0.1	-

### 6.5.3 Uncertainties from background biases

As described in Section 6.4.2, the background modeling follows a data-driven approach, and is not affected by any systematic uncertainty in the simulation. Instead of evaluating the impacts of uncertainties as done for the signal modeling, the main task for background is to make sure it is robust against spurious signals.

A spurious signal is produced by the bias between the analytic background function and the true background shape at the position of the expected signal. For an analysis with finite statistics, there is a statistical uncertainty on the signal strength,  $\sigma_{stat}$ , resulting from the statistical fluctuation of background events. The best fit signal strength  $\mu_{fit}$  under the null hypothesis, which is the hypothesis without the existence of a true signal, should follow  $\mathcal{N}(0, \sigma_{stat})$ , the normal distribution with a mean of 0 and a standard deviation of  $\sigma_{stat}$ . The  $1\sigma$  or  $2\sigma$  range of this distribution gives the 68.3% or 95.4% confidence intervals for the exclusion of this null hypothesis. If there is a systematic spurious signal  $\hat{\mu}_{SS}$ , namely a bias, the probability distribution for the  $\mu_{fit}$  becomes  $\mathcal{N}(\hat{\mu}_{SS}, \sigma_{stat})$ , and the coverage of the  $1\sigma_{stat}$  range becomes equation 6-11, which is not 68.3%.

$$\int_{-\sigma_{stat}}^{\sigma_{stat}} \mathcal{N}(\hat{\mu}_{SS}, \sigma_{stat}) = \frac{1}{2} [erf(\frac{\sigma_{stat} + \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}}) + erf(\frac{\sigma_{stat} - \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}})] \quad (6-11)$$

The left plot of Figure 6-15 illustrates this difference in coverage. The middle plot of Figure 6-15 shows how the coverage changes as the bias gets larger. As a result, to achieve the 68.3% confidence level, the signal uncertainty from the fit,  $\sigma_{fit}$ , needs to satisfy equation 6-12 and becomes larger than the  $\sigma_{stat}$ .

$$\frac{1}{2} [erf(\frac{\sigma_{fit} + \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}}) + erf(\frac{\sigma_{fit} - \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}})] = 68.3\% \quad (6-12)$$

In this way, the bias in the background modeling, even if it is not strictly an uncertainty, adds to the overall uncertainty of the signal strength measurement. The relationship between  $\sigma_{fit}$  and  $\hat{\mu}_{SS}$  is shown in the right plot of Figure 6-15. As a convention in the  $H \rightarrow \mu\mu$  analysis, biases below 20% of the statistical uncertainty of the signal are considered ac-

ceptable, which corresponds to less than 2% inflation of the signal uncertainty.

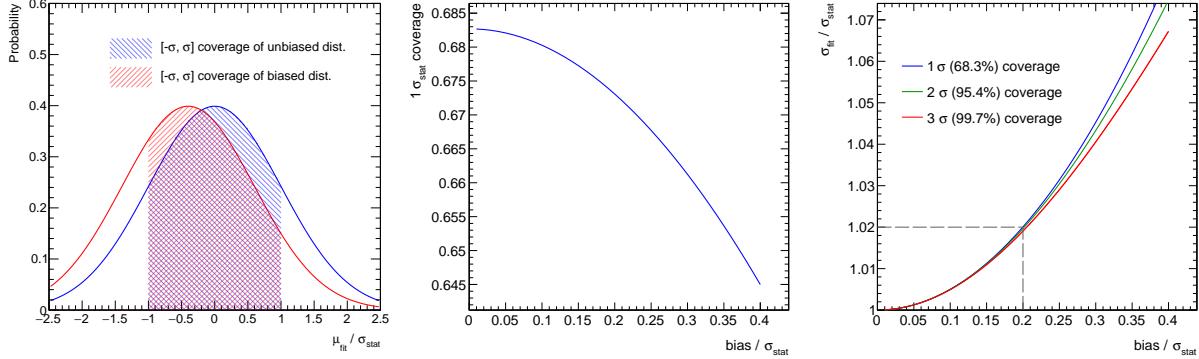


Figure 6-15. Schemes on how the bias affects the uncertainty on the signal strength measurement. The left plot is an illustration of the  $[-\sigma, \sigma]$  coverage of the biased signal strength measurement. The middle plot shows this coverage becomes less as the bias gets larger. The right plot shows how the bias impacts the best fit signal strength uncertainty  $\sigma_{fit}$ . The gray dash lines in the right plot indicates the conventional acceptable range of the bias in this analysis.

In this analysis, the bias is evaluated between different function candidates via studies on groups of pseudo-experiments (toys). The true background shape is of course unknown, but is believed to be covered by the flexibility of the collective set of functional forms. Procedures for the bias evaluation are as follows:

### 1. Toy generation

- One function candidate  $f(m_{\mu\mu})$  is fit to the background shape of the simulation to find the best fit parameters.
- The best fit shape of  $f(m_{\mu\mu})$  is used as the Probability Density Function (PDF) to generate toy datasets. In each toy, the number of events in each bin is taken sampling the Poisson distribution of the expected number of events given by  $f(m_{\mu\mu})$ .
- For each selected function  $f(m_{\mu\mu})$ , 3000 toys are generated.

### 2. Signal injection

- For each background toy, an artificial signal is also generated following the Poisson distribution of a given signal strength  $\hat{\mu}_{inj}$ . In this set of study, two sets of artificial signal strength are tested, which are zero or the expected SM signal strength.

- The artificial signal toys are added to the background toys and completes the signal + background toys.

### 3. Signal extraction

- For each signal + background toy of function  $f(m_{\mu\mu})$ , the shape analysis is performed using another function  $g(m_{\mu\mu})$ . In these toy analyses, systematic uncertainties on the signal modeling are not included, as they are unrelated to the bias estimation.
- From these fits, the best fit signal strength  $\mu_{fit}$  and its standard deviation  $\sigma_{fit}$  are extracted.

### 4. Bias evaluation

- The spurious signal between function  $f(m_{\mu\mu})$  and  $g(m_{\mu\mu})$  in each toy is defined as:

$$\mu_{ss}(f,g) = \frac{\mu_{fit} - \hat{\mu}_{inj}}{\sigma_{fit}} \quad (6-13)$$

- The distribution (of 3000 toys) of this spurious signal is fit with a gaussian function. As stated above, this spurious signal should follow the Gaussian distribution  $\mathcal{N}(\hat{b}, 1)$ , where  $\hat{b} = \hat{\mu}_{ss}/\sigma_{stat}$ . The mean value from the Gaussian fit is the bias between function  $f(m_{\mu\mu})$  and  $g(m_{\mu\mu})$ .

Such bias is evaluated in each subcategory between each combination of  $f(m_{\mu\mu})$  and  $g(m_{\mu\mu})$ , and the results are summarized in Figure 6-16 and 6-17. Most of the functions have good bias response against other functions can be used for the analysis. In the final analysis, the BWZGamma function (2 shape DOF) is chosen as the background model in the WH [-1.0, -0.1] category, and the BWZ function (1 shape DOF) is chosen in all other categories.

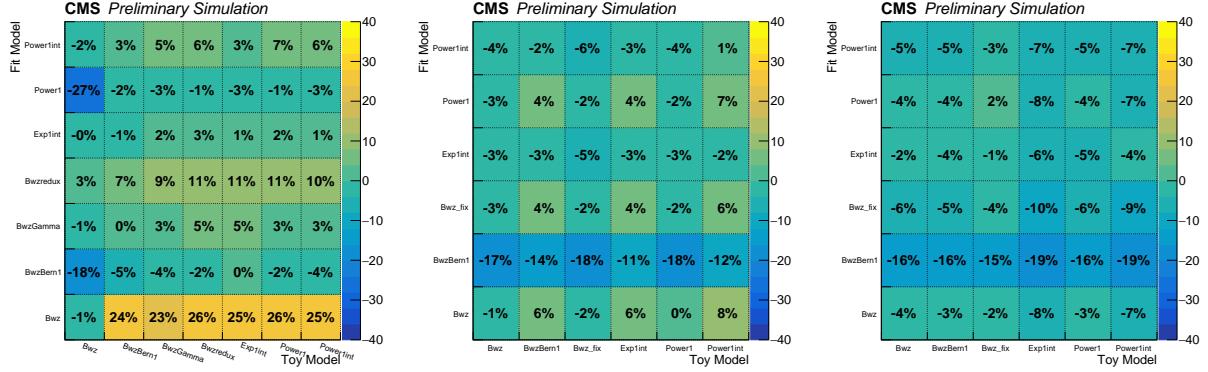


Figure 6-16. Bias in different BDT-based WH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 0.3] (middle), Cat3 BDT [0.3, 1.0] (right). In the tables, the Power stands for a single "Power" function and the "PowerInt" stands for a single power function plus a constant.

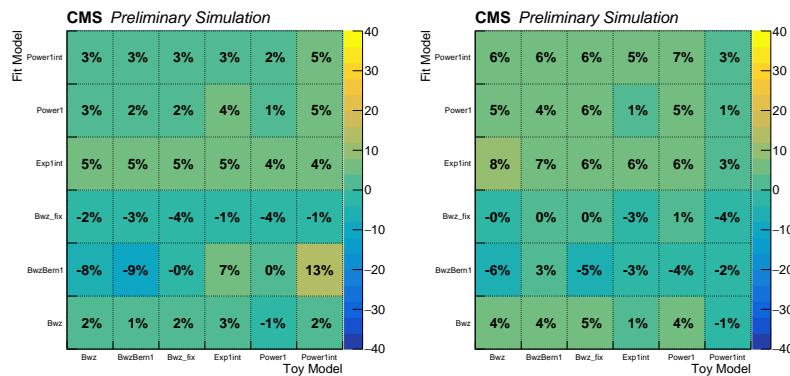


Figure 6-17. Bias in different BDT-based ZH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 1.0] (right). In the tables, the "Power" stands for a single power function and the "PowerInt" stands for a single power function plus a constant.

## 6.6 Results of the VH analysis

The final results are extracted by performing a binned maximum-likelihood fit in each VH subcategory. The fit is performed on the observed  $m_{\mu\mu}$  distribution in the mass range of  $110 < m_{\mu\mu} < 150$  GeV. The SM Higgs mass is assumed at  $m_H = 125$  GeV. The parameter of interest (POI) of this fit is the signal strength modifier, which is defined as the ratio between the observed signal rate to the SM expectation,  $\mu = (\sigma \mathcal{B}(H \rightarrow \mu\mu))_{\text{obs}} / (\sigma \mathcal{B}(H \rightarrow \mu\mu))_{\text{SM}}$ . All the different signal modes in different subcategories share a common signal strength modifier.

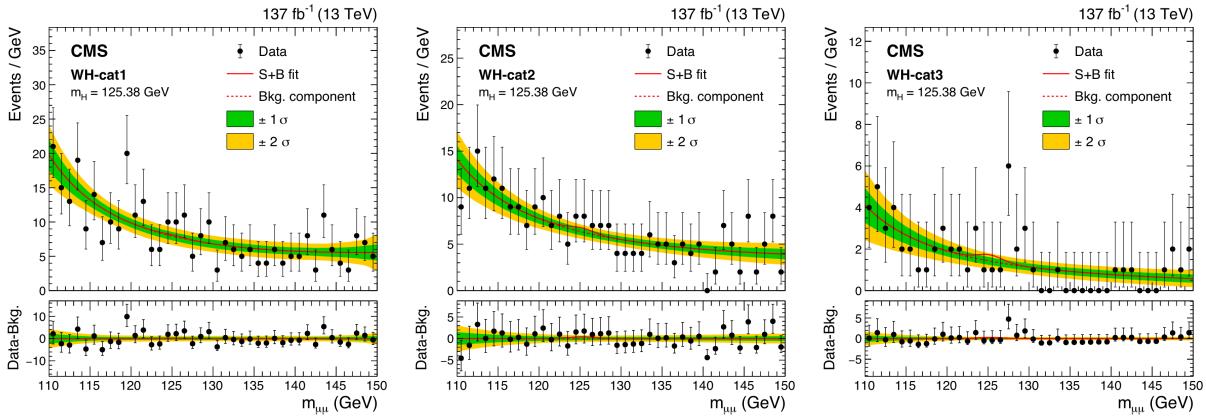


Figure 6-18. Post-fit  $m_{\mu\mu}$  distribution of the WH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 0.3] (middle), Cat3 BDT [0.3, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.

Figure 6-18 and 6-19 show the post-fit results of the signal-plus-background fits in individual WH and ZH subcategories, while Figure 6-20 shows the summary of all subcategories weighted by  $S/(S+B)$ , where  $S$  and  $B$  are number of expected signal and background events within the FWHM range of the expected signal peak.

The estimate of POI can be affected by many other parameters, such as the systematic uncertainties and the free background shape parameters. These parameters are called nuisance parameters and are allowed to float freely in the fit mentioned above. The fit

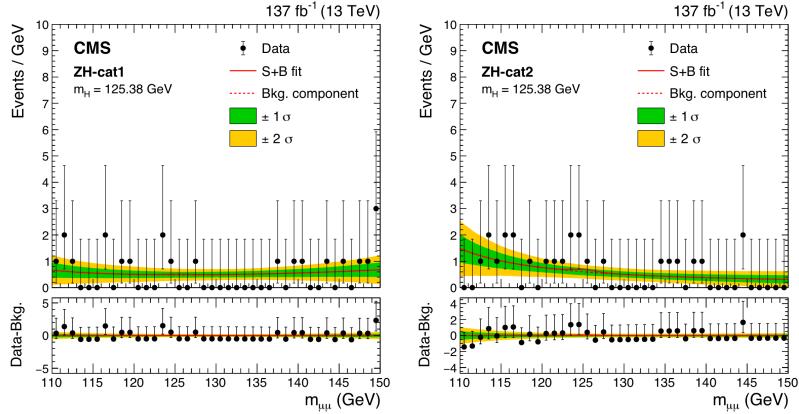


Figure 6-19. Post-fit  $m_{\mu\mu}$  distribution of the ZH subcategories. The subcategories are: Cat1 BDT [-1.0, -0.1] (left), Cat2 BDT [-0.1, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.

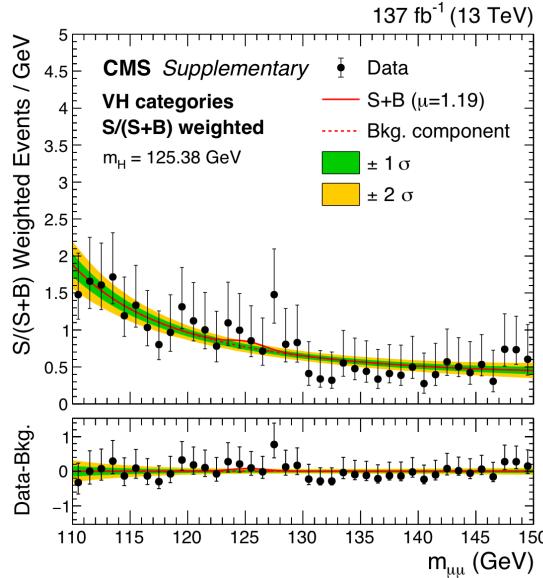


Figure 6-20. Post-fit  $m_{\mu\mu}$  distribution of the weighted combination of all WH and ZH subcategories. Subcategories are weighted proportionally to  $S/(S+B)$  ratio. The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.

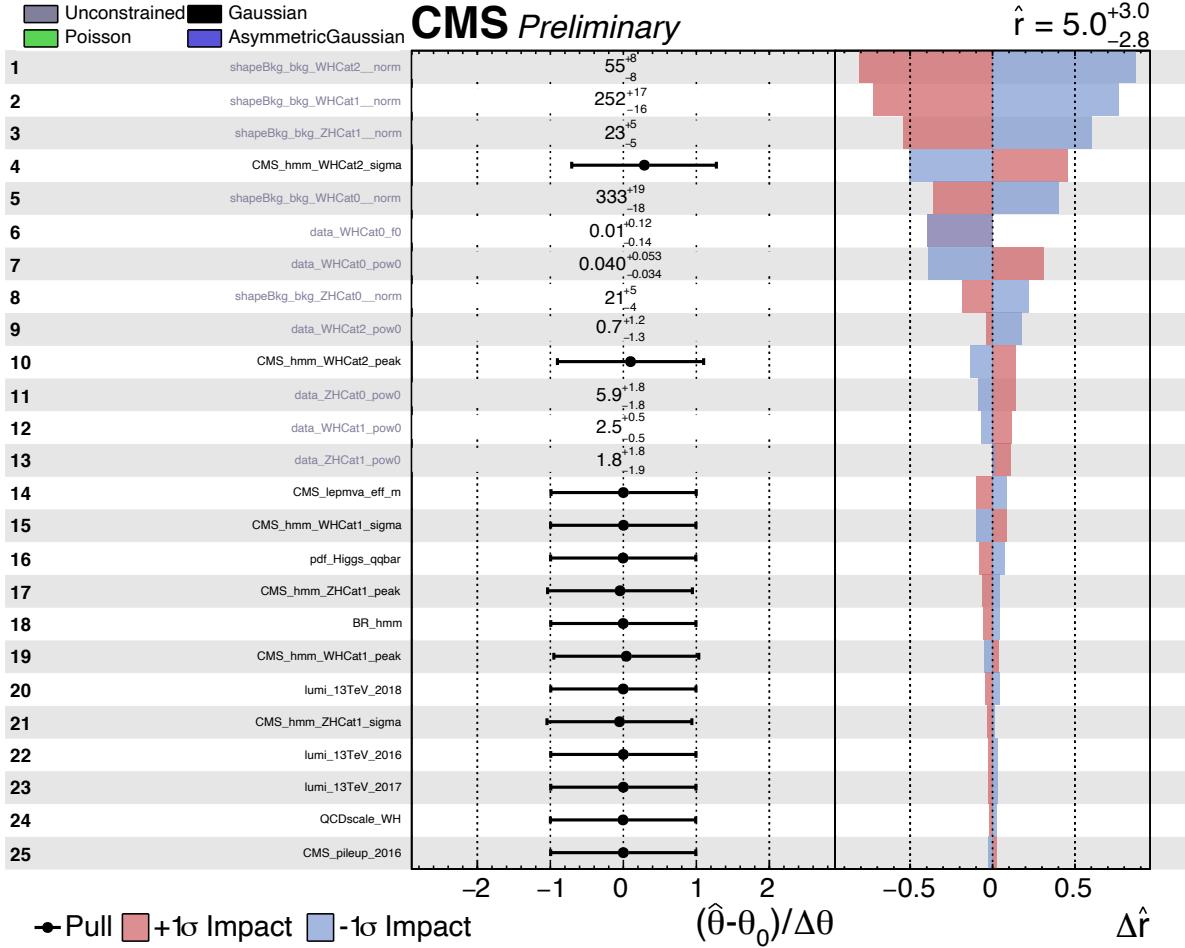


Figure 6-21. Impacts of the nuisance parameters  $\theta$  on the signal strength  $\hat{r}$ . Parameter names in black are systematic uncertainties, which are Gaussian-constrained to their pre-fit uncertainties. Parameter names in gray are shape and normalization parameters, which do not have pre-fit constraints.  $(\hat{\theta} - \theta_0)/\Delta\theta$  shows the pull of each nuisance parameter, while  $\Delta\hat{r}$  is its impact on signal strength. Unconstrained parameters do not have pre-fit assumptions and therefore no pulls. The best-fit values of them are shown in the place of pulls.

also returns the best fit value of all nuisance parameters and their post-fit uncertainties. To evaluate the impact of each nuisance parameter, a set of variation fits are performed by independently shifting each nuisance parameter by its post-fit uncertainty (up and down). The difference in POI between the initial fit and each variation fit measures the impact of the corresponding nuisance parameter. Figure 6-21 summarizes the results of the impact study, showing 25 nuisance parameters of the highest ranked impacts. In the plot, the

pull between the pre-fit and post-fit values of each parameter is defined as the difference between the pre-fit and post-fit values  $\hat{\theta} - \theta$  divided by the pre-fit uncertainty  $\Delta\theta$ . Pulls of most parameters are centered near 0 and have uncertainties close to 1, meaning their post-fit values change very little from pre-fit values. The best fit signal strength is denoted as  $\hat{r}$  in the plot, with the impact defined as the variation  $\Delta\hat{r}$  induced by the shifting one nuisance parameter. The largest impacts come from background normalization parameters. The largest impact among systematic uncertainties is around 0.5, while the overall uncertainty of the signal strength is about 3. The sensitivity of this analysis is mainly limited by statistical uncertainty.

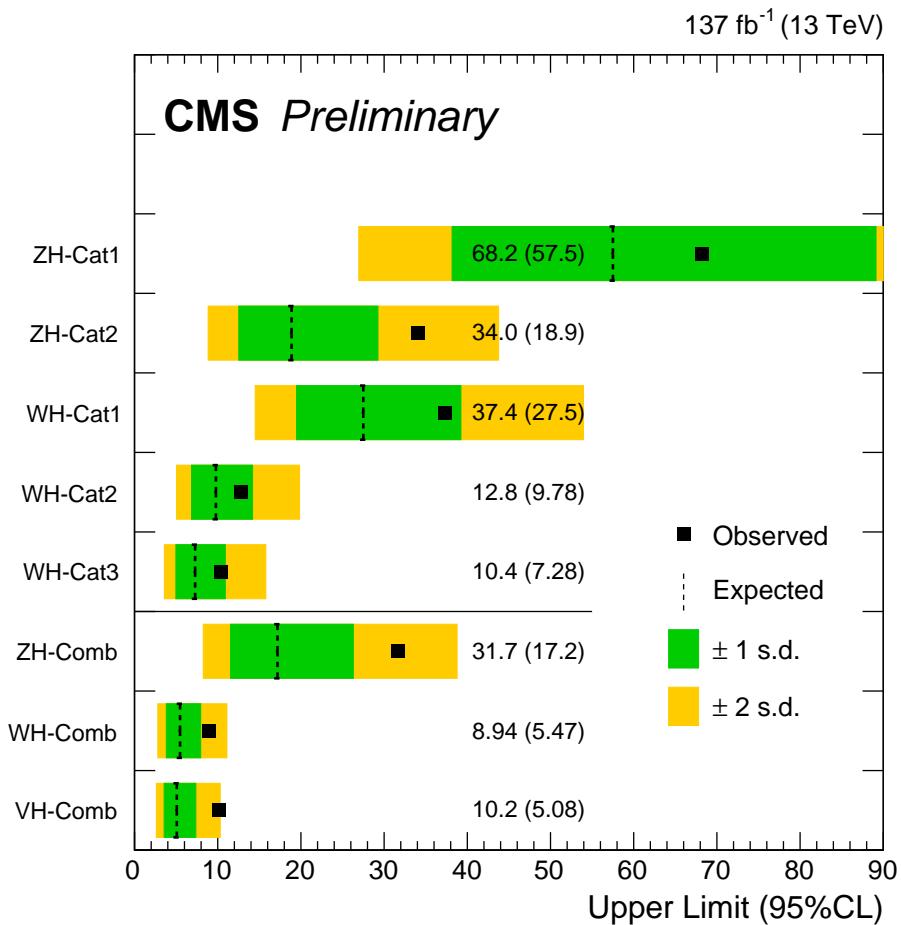


Figure 6-22. The 95% CL upper limits on the signal strength in subcategories and combinations. Observed (expected) values are marked in the plot.

Figure 6-22 shows the 95% confidence level (CL) upper limits on the signal strength

in each VH subcategory as well as the combinations. The limits are calculated with an asymptotic approximation [102] of the  $CL_s$  method [103]. The expected limits and uncertainties are based on the background-only hypothesis. The expected and observed significance for the VH subcategories and combinations are summarized in Table 6-11. The expected significance is based on the SM signal strength hypothesis.

Table 6-11. Summary of the expect and observed limits and significance in each individual subcategory and the combination.

Category	Expected Signif.	Observed Signif.
ZH $\rightarrow 4\ell$ Cat1	0.05	0.65
ZH $\rightarrow 4\ell$ Cat2	0.14	1.81
WH $\rightarrow 3\ell$ Cat1	0.08	0.89
WH $\rightarrow 3\ell$ Cat2	0.22	0.79
WH $\rightarrow 3\ell$ Cat3	0.33	0.64
ZH combine	0.15	1.92
WH combine	0.40	1.15
WH and ZH combine	0.43	1.86

These results from the VH analysis are combined with other categories (ggH, VBF, and  $t\bar{t}H$ ) to make the inclusive  $H \rightarrow \mu\mu$  analysis. Studies and results of the combination are covered in Chapter 7.

## CHAPTER 7

### RESULTS OF THE H<sub>2</sub>MU SEARCH

The analysis in the VH category is combined with those in the ggH, VBF, and tt>H categories [36]. All results and plots in this chapter are published in Ref. [36]. In this combined analysis, the mass of the Higgs boson is expected at  $m_H = 125.38$  GeV [23], which is the most precise measurement of the Higgs boson mass up to date. Figure 7-1 summarizes the expected signal composition in all subcategories of the  $H \rightarrow \mu\mu$  analysis. Figure 7-2 summarizes the expected  $S/(S+B)$  and  $S/\sqrt{B}$  in all subcategories, in which the signal and background yields are calculated by integrating the expectations within the FWHM range of the signal peak for the ggH, VH, and tt>H subcategories, while for the VBF category the considered mass range is  $115 \text{ GeV} < m_{\mu\mu} < 135 \text{ GeV}$ .

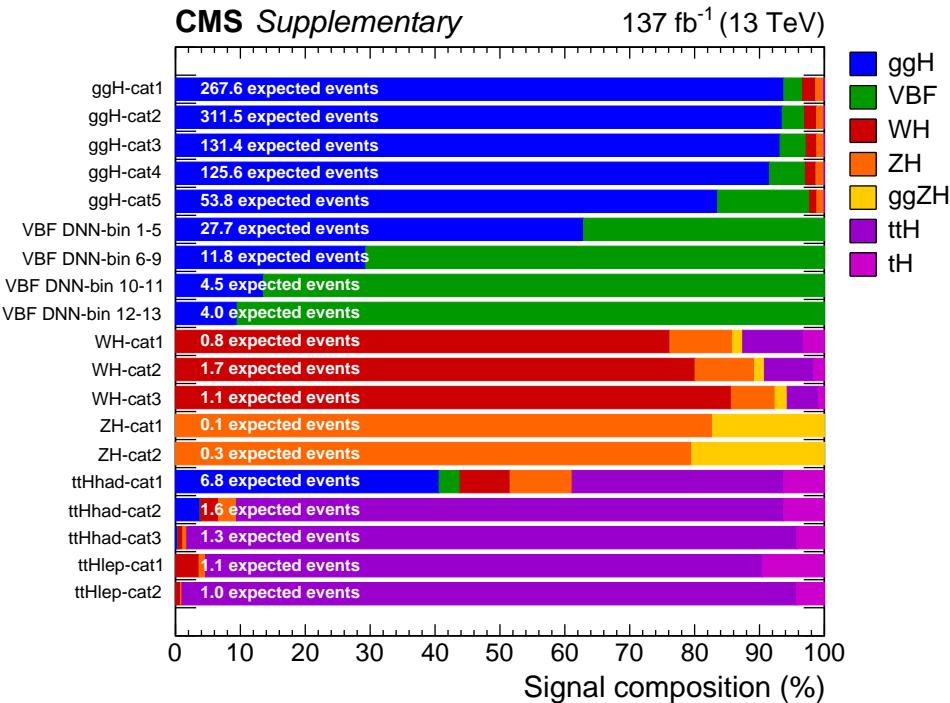


Figure 7-1. Expected fraction of signal events per production mode in the different subcategories for  $m_H = 125.38$  GeV. The tH contribution is defined as the sum of tHq and tHW processes.

The individual signal strength modifier from each category is summarized in Figure 7-3, with the Higgs signal expected at  $m_H = 125.38$  GeV. A combined fit is performed across all categories with one common signal strength modifier, whose best fit value is  $\hat{\mu} = 1.19^{+0.41}_{-0.40}(\text{stat})^{+0.17}_{-0.16}(\text{syst})$ . The statistical uncertainty is determined by performing a

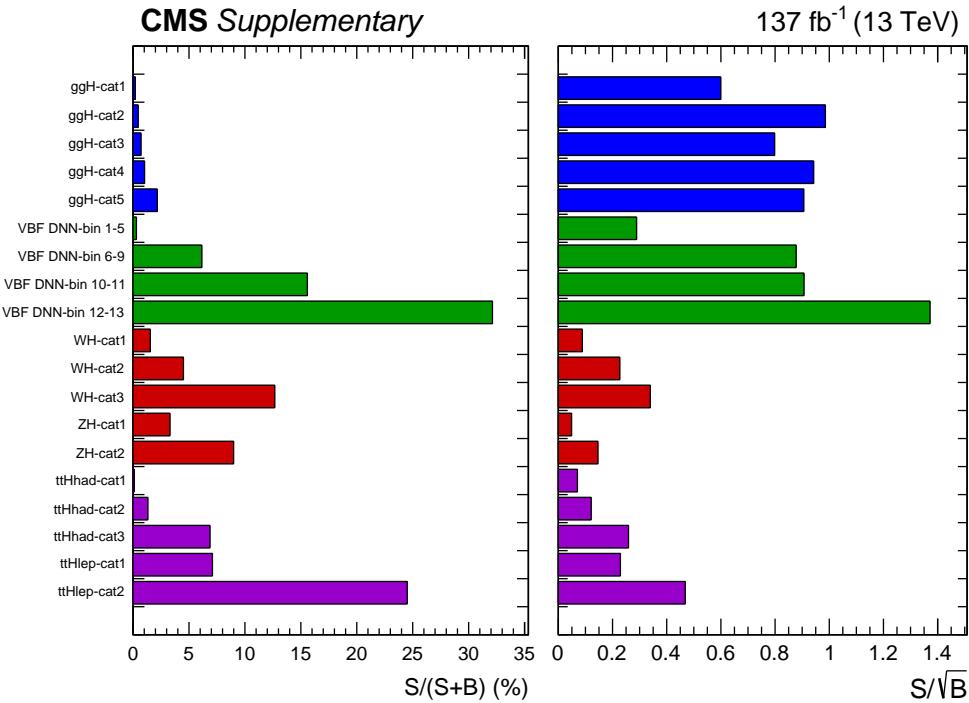


Figure 7-2. Expected  $S/(S+B)$  and  $S/\sqrt{B}$  in different subcategories, where  $S$  and  $B$  indicate the number of expect signal and background events, respectively.

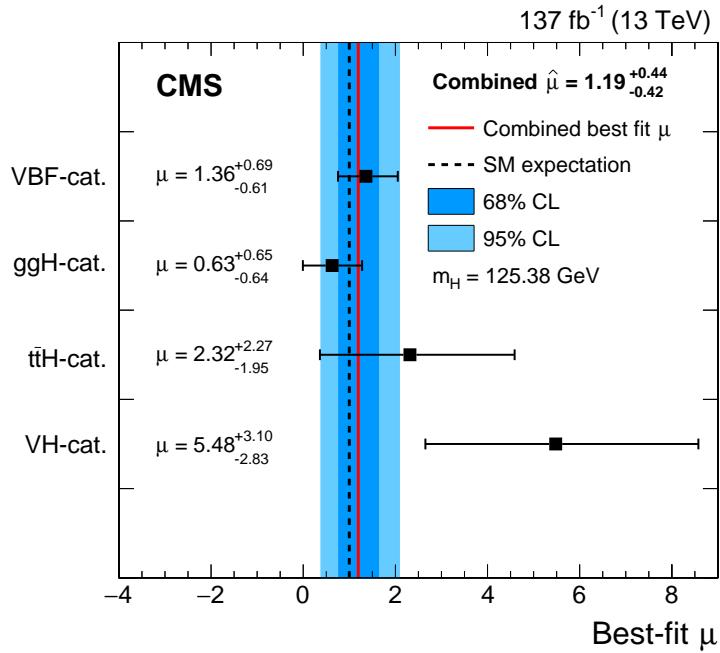


Figure 7-3. Signal strength modifiers measured for  $m_H = 125.38$  GeV in each production category (black points) are compared to the result of the combined fit (solid red line) and the SM expectation (dashed gray line).

likelihood scan of  $\mu$  with all nuisance parameters fixed at their best fit values. And the systematic uncertainty is calculated as the difference in quadrature between the total uncertainty and the statistical component. This result makes the most precise measurement of the  $H \rightarrow \mu\mu$  decay rate up to date. Assuming SM production cross sections for all Higgs production modes, the  $H \rightarrow \mu\mu$  branching ratio is constrained at 95% CL to be within  $0.8 \times 10^{-4} < \mathcal{B}(H \rightarrow \mu\mu) < 4.5 \times 10^{-4}$ .

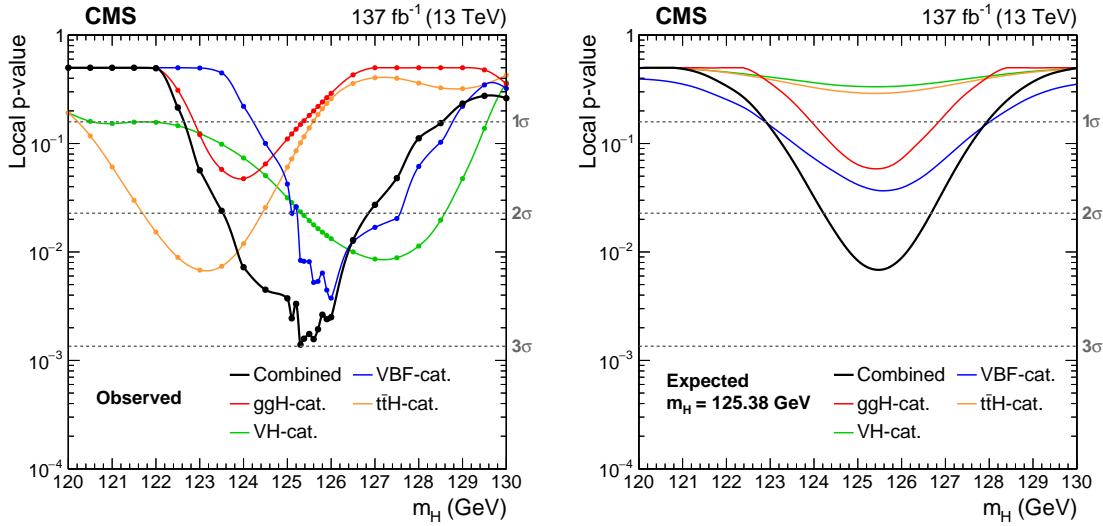


Figure 7-4. The local  $p$ -values as a function of different  $m_H$  hypotheses, for each individual categories as well as the overall combination. The left plot shows the observed  $p$ -values, each solid marker indicating a mass point, for which the observed  $p$ -values are computed. The right plot shows the expected  $p$ -values calculated using the background estimate from the S+B fit and injecting a signal with  $m_H = 125.38$  GeV and  $\mu = 1$ .

The statistical significance of a signal presence is tested against the null hypothesis, in which no signal is assumed and the observed distribution is purely raised by statistical fluctuation in background. The local  $p$ -value quantifies the probability for the background to produce such a fluctuation larger than the apparent signal observed in the search region. A scan of  $p$ -value is performed across the mass range of  $120 \text{ GeV} < m_H < 130 \text{ GeV}$ . Figure 7-4 shows the observed and expected  $p$ -values for individual categories as well as the overall combination. The expected  $p$ -values are calculated with a SM signal at 125.38 GeV injected on top of the background expectation. The observed (expected for  $\mu = 1$ ) signifi-

cance at  $m_H = 125.38$  GeV of the incompatibility with the background-only hypothesis is 3.0 (2.5) standard deviations, while the observed (expected for  $\mu = 0$ ) upper limit of the signal strength at the 95% confidence level is 1.9 (0.8) times the SM expectation. This result establishes the first evidence of the Higgs boson decay to fermions of the second generation.

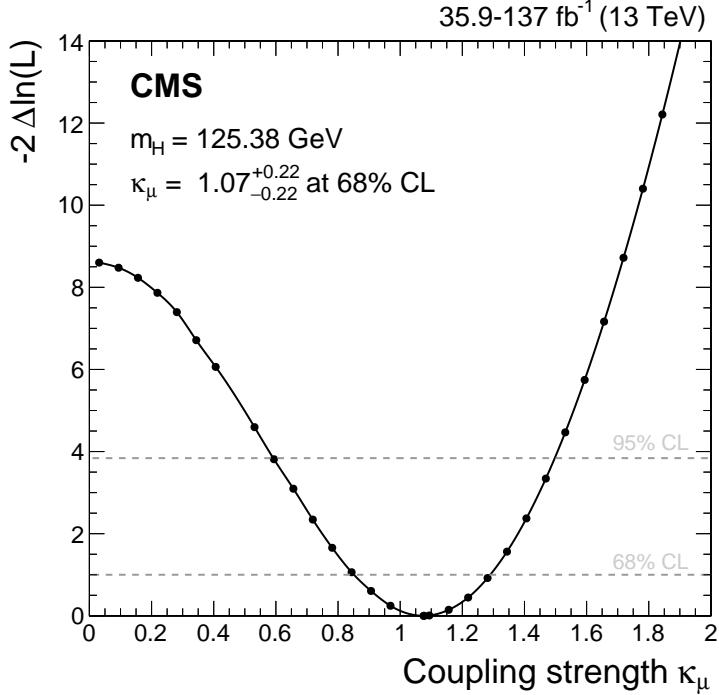


Figure 7-5. The observed profile likelihood ratio as a function of  $\kappa_\mu$  for  $m_H = 125.38$  GeV.

Finally, as discussed in Section 1.4, the coupling strength between the Higgs boson and the muon is an essential parameter to SM. This work also provides the best constraint on the Higgs-muon coupling up to date. The signal strength of  $H \rightarrow \mu\mu$  decay cannot be translated directly into a measurement of the Higgs boson coupling to muons, as the  $H \rightarrow \mu\mu$  decay ratio is determined by all possible Higgs decays. Assuming the Higgs boson does not decay to unknown particles, the Higgs coupling strengths in SM are evaluated in the  $\kappa$ -framework [35]. The measurement of Higgs coupling strengths is performed by combining many Higgs analyses, and the latest measurement by CMS [2] prior to this work is based on collision data at  $\sqrt{s} = 13$  TeV recorded in 2016, corresponding to an integrated luminosity

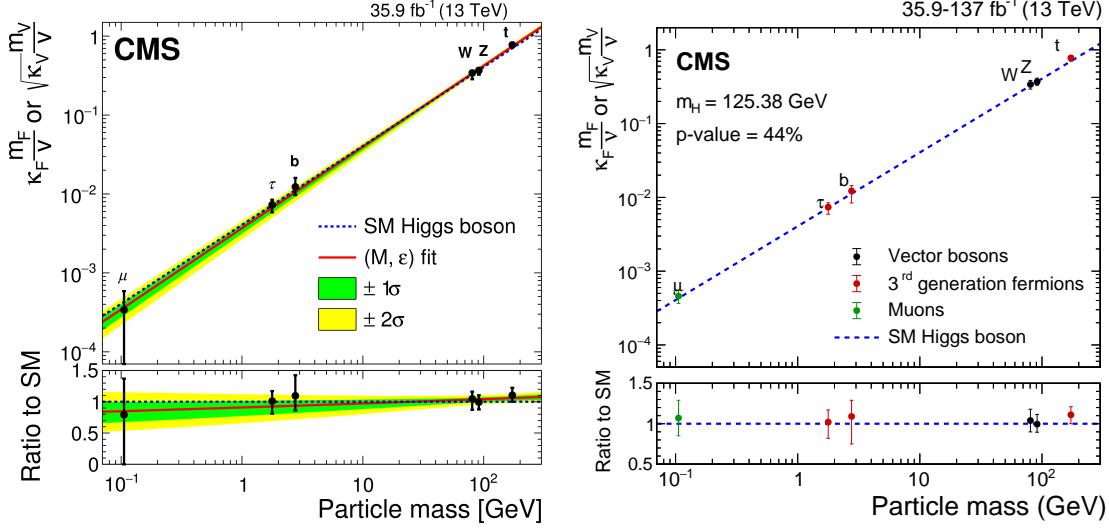


Figure 7-6. Left: A duplicate of Figure 1-3, taken from Ref. [2]. Summary of the CMS measurements on the Higgs coupling to fermions and bosons based on data recorded in 2016. Right: The coupling strength measurements updated with this  $H \rightarrow \mu\mu$  result.

of  $35.9 \text{ fb}^{-1}$ . This  $H \rightarrow \mu\mu$  analysis is combined with Ref. [2], updating the coupling strengths measurement. The likelihood scan of  $\kappa_\mu$  is shown in Figure 7-5. The best fit value for  $\kappa_\mu$  is 1.07 and the corresponding observed 68% confidence interval is  $0.85 < \kappa_\mu < 1.29$ . The global result of all coupling modifiers is shown in Figure 7-6, updating Figure 1-3.

In summary, an analysis on the  $H \rightarrow \mu\mu$  decay is performed with  $pp$  collision data at  $\sqrt{s} = 13 \text{ TeV}$  collected by CMS, corresponding to an integrated luminosity of  $137 \text{ fb}^{-1}$ . An excess of events over background-only expectation is observed (expected) with a significance of 3.0 (2.5) standard deviations for the SM Higgs boson at  $125.38 \text{ GeV}$ . The measured signal strength, relative to the SM expectation, is  $1.19^{+0.41}_{-0.40}(\text{stat})^{+0.17}_{-0.16}(\text{syst})$ . This result makes the first evidence for the decay of the Higgs boson to second generation fermions and provides the most precise measurement of the Higgs boson coupling to muons up to date.

## REFERENCES

- [1] Go on a particle quest at the first cern webfest. URL <https://cds.cern.ch/journal/CERNBulletin/2012/35/News%20Articles/1473657>.
- [2] Combined measurements of Higgs boson couplings in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Eur. Phys. J. C*, 79(arXiv:1809.10733. CMS-HIG-17-031-005):421. 67 p, Sep 2018. doi: 10.1140/epjc/s10052-019-6909-y. URL <https://cds.cern.ch/record/2640611>. Submitted to Eur.Phys.J.
- [3] Lyndon Evans and Philip Bryant. LHC machine. *Journal of Instrumentation*, 3(08):S08001–S08001, aug 2008. doi: 10.1088/1748-0221/3/08/s08001. URL <https://doi.org/10.1088/1748-0221/3/08/s08001>.
- [4] Magnetic field of an idealized quadrupole. URL [https://en.wikipedia.org/wiki/File:Magnetic\\_field\\_of\\_an\\_idealized\\_quadrupole\\_with\\_forces.svg](https://en.wikipedia.org/wiki/File:Magnetic_field_of_an_idealized_quadrupole_with_forces.svg).
- [5] Cms luminosity - public results. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [6] T Gleisberg, S Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007–007, feb 2009. doi: 10.1088/1126-6708/2009/02/007. URL <https://doi.org/10.1088/1126-6708/2009/02/007>.
- [7] The CMS Phase-1 Pixel Detector Upgrade. Technical Report arXiv:2012.14304, Dec 2020. URL <https://cds.cern.ch/record/2748381>.
- [8] J Mans, J Anderson, B Dahmes, P de Barbaro, J Freeman, T Grassi, E Hazen, J Mans, R Ruchti, I Schimdt, T Shaw, C Tully, J Whitmore, and T Yetkin. CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter. Technical Report CERN-LHCC-2012-015. CMS-TDR-10, Sep 2012. URL <https://cds.cern.ch/record/1481837>.
- [9] A Tapper and Darin Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. Technical Report CERN-LHCC-2013-011. CMS-TDR-12, Jun 2013. URL <https://cds.cern.ch/record/1556311>.
- [10] Tai Sakuma. Cutaway diagrams of CMS detector. May 2019. URL <https://cds.cern.ch/record/2665537>.
- [11] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Journal of Instrumentation*, 13(06):P06015–P06015, jun 2018. doi: 10.1088/1748-0221/13/06/p06015. URL <https://doi.org/10.1088/1748-0221/13/06/p06015>.
- [12] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03):T03021–T03021, mar 2010. doi: 10.1088/1748-0221/5/03/t03021. URL <https://doi.org/10.1088/1748-0221/5/03/t03021>.

- [13] Sketches of the cms tracker detector. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.
- [14] R.M. Brown. The cms electromagnetic calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 572(1):29–32, 2007. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2006.10.240>. URL <https://www.sciencedirect.com/science/article/pii/S016890020601970X>. Frontier Detectors for Frontier Physics.
- [15] A Colaleo, A Safonov, A Sharma, and M Tytgat. CMS Technical Design Report for the Muon Endcap GEM Upgrade. Technical Report CERN-LHCC-2015-012. CMS-TDR-013, Jun 2015. URL <https://cds.cern.ch/record/2021453>.
- [16] CMS Collaboration. The performance of the CMS muon detector in proton-proton collisions at  $\sqrt{s} = 7$  TeV at the LHC. *Journal of Instrumentation*, 8(11):P11002–P11002, nov 2013. doi: 10.1088/1748-0221/8/11/p11002. URL <https://doi.org/10.1088/1748-0221/8/11/p11002>.
- [17] CMS Collaboration. Performance of the CMS level-1 trigger in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Journal of Instrumentation*, 15(10):P10017–P10017, oct 2020. doi: 10.1088/1748-0221/15/10/p10017. URL <https://doi.org/10.1088/1748-0221/15/10/p10017>.
- [18] Siona Ruth Davis. Interactive Slice of the CMS detector. Aug 2016. URL <https://cds.cern.ch/record/2205172>.
- [19] Oliver Rieger. *Evidence for Higgs boson decays to muons with the CMS experiment at the LHC*. PhD thesis, University of Hamburg, 2021.
- [20] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012. doi: 10.1016/j.physletb.2012.08.020.
- [21] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012. doi: 10.1016/j.physletb.2012.08.021.
- [22] Serguei Chatrchyan et al. Observation of a new boson with mass near 125GeV in pp collisions at  $\sqrt{s} = 7$  and 8TeV. *JHEP*, 06:081, 2013. doi: 10.1007/JHEP06(2013)081.
- [23] A.M. Sirunyan et al. A measurement of the Higgs boson mass in the diphoton decay channel. *Physics Letters B*, 805:135425, 2020. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2020.135425>. URL <https://www.sciencedirect.com/science/article/pii/S037026932030229X>.
- [24] A.M. Sirunyan et al. Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP*, 11(CMS-HIG-16-040).

CMS-HIG-16-040-003):185. 56 p, Apr 2018. doi: 10.1007/JHEP11(2018)185. URL <https://cds.cern.ch/record/2312121>. Submitted to JHEP. All figures and tables can be found at <http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-16-040> (CMS Public Pages).

- [25] A.M. Sirunyan et al. Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at  $\sqrt{s} = 13\text{TeV}$ . *Physics Letters B*, 791:96–129, 2019. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2018.12.073>. URL <https://www.sciencedirect.com/science/article/pii/S0370269319301169>.
- [26] Albert M Sirunyan et al. Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at  $\sqrt{s} = 13\text{TeV}$ . *JHEP*, 11:047, 2017. doi: 10.1007/JHEP11(2017)047.
- [27] A. M. Sirunyan et al. Observation of Higgs Boson Decay to Bottom Quarks. *Phys. Rev. Lett.*, 121:121801, Sep 2018. doi: 10.1103/PhysRevLett.121.121801. URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.121801>.
- [28] A.M. Sirunyan et al. Observation of the Higgs boson decay to a pair of  $\tau$  leptons with the CMS detector. *Physics Letters B*, 779:283–316, 2018. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2018.02.004>. URL <https://www.sciencedirect.com/science/article/pii/S0370269318301035>.
- [29] M. Aaboud et al. Cross-section measurements of the Higgs boson decaying into a pair of  $\tau$ -leptons in proton-proton collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector. *Phys. Rev. D*, 99:072001, Apr 2019. doi: 10.1103/PhysRevD.99.072001. URL <https://link.aps.org/doi/10.1103/PhysRevD.99.072001>.
- [30] M. Aaboud et al. Observation of  $H \rightarrow b\bar{b}$  decays and VH production with the ATLAS detector. *Physics Letters B*, 786:59–86, 2018. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2018.09.013>. URL <https://www.sciencedirect.com/science/article/pii/S0370269318307056>.
- [31] M. Aaboud et al. Measurements of gluongluon fusion and vector-boson fusion Higgs boson production cross-sections in the  $H \rightarrow WW^* \rightarrow e\nu\mu\nu$  decay channel in pp collisions at  $s=13\text{ TeV}$  with the ATLAS detector. *Physics Letters B*, 789:508–529, 2019. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2018.11.064>. URL <https://www.sciencedirect.com/science/article/pii/S0370269318309936>.
- [32] M. Aaboud and et al. Measurement of the Higgs boson coupling properties in the  $H \rightarrow ZZ^* \rightarrow 4l$  decay channel at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector. *Journal of High Energy Physics*, 2018(3), Mar 2018. ISSN 1029-8479. doi: 10.1007/jhep03(2018)095. URL [http://dx.doi.org/10.1007/JHEP03\(2018\)095](http://dx.doi.org/10.1007/JHEP03(2018)095).
- [33] A. M. Sirunyan et al. Observation of  $t\bar{t}H$  Production. *Phys. Rev. Lett.*, 120:231801, Jun 2018. doi: 10.1103/PhysRevLett.120.231801. URL <https://link.aps.org/doi/10.1103/PhysRevLett.120.231801>.

- [34] M. Aaboud et al. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Physics Letters B*, 784:173–191, 2018. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2018.07.035>. URL <https://www.sciencedirect.com/science/article/pii/S0370269318305732>.
- [35] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 7 2013. doi: 10.5170/CERN-2013-004.
- [36] A. M. Sirunyan and et al. Evidence for Higgs boson decay to a pair of muons. *Journal of High Energy Physics*, 2021(1), Jan 2021. ISSN 1029-8479. doi: 10.1007/jhep01(2021)148. URL [http://dx.doi.org/10.1007/JHEP01\(2021\)148](http://dx.doi.org/10.1007/JHEP01(2021)148).
- [37] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004–S08004, aug 2008. doi: 10.1088/1748-0221/3/08/s08004. URL <https://doi.org/10.1088/1748-0221/3/08/s08004>.
- [38] CMS Collaboration. Bunched beam test of the cms drift tubes local muon trigger. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 534(3):441–485, 2004. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2004.06.169>. URL <https://www.sciencedirect.com/science/article/pii/S0168900204016559>.
- [39] Performance of the CMS TwinMux Algorithm in late 2016 pp collision runs. Dec 2016. URL <https://cds.cern.ch/record/2239285>.
- [40] R Frazier, S Fayer, G Hall, C Hunt, G Iles, D Newbold, and A Rose. A demonstration of a time multiplexed trigger for the CMS experiment. *Journal of Instrumentation*, 7(01):C01060–C01060, jan 2012. doi: 10.1088/1748-0221/7/01/c01060. URL <https://doi.org/10.1088/1748-0221/7/01/c01060>.
- [41] V. Khachatryan et al. Search for a standard model-like Higgs boson in the  $\mu^+\mu^-$  and  $e^+e^-$  decay channels at the LHC. *Physics Letters B*, 744:184 – 207, 2015. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2015.03.048>. URL <http://www.sciencedirect.com/science/article/pii/S0370269315002117>.
- [42] A. M. Sirunyan et al. Search for the Higgs Boson Decaying to Two Muons in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV. *Phys. Rev. Lett.*, 122:021801, Jan 2019. doi: 10.1103/PhysRevLett.122.021801. URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.021801>.
- [43] G. Aad et al. Search for the Standard Model Higgs boson decay to  $\mu^+\mu^-$  with the ATLAS detector. *Physics Letters B*, 738:68 – 86, 2014. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2014.09.008>. URL <http://www.sciencedirect.com/science/article/pii/S0370269314006583>.
- [44] M. Aaboud et al. Search for the Dimuon Decay of the Higgs Boson in  $pp$  Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector. *Phys. Rev. Lett.*, 119:051802, Aug 2017. doi: 10.1103/PhysRevLett.119.051802. URL <https://link.aps.org/doi/10.1103/PhysRevLett.119.051802>.

- [45] Georges Aad et al. A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector. Submitted to *Phys. Lett. B*, 2020.
- [46] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159 – 177, 2015. ISSN 0010-4655. doi: <https://doi.org/10.1016/j.cpc.2015.01.024>. URL <http://www.sciencedirect.com/science/article/pii/S0010465515000442>.
- [47] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76: 196, 2016. doi: 10.1140/epjc/s10052-016-4018-8.
- [48] S. Agostinelli et al. Geant4a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. ISSN 0168-9002. doi: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [49] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.
- [50] Paolo Nason. A new method for combining NLO QCD with shower monte carlo algorithms. *Journal of High Energy Physics*, 2004(11):040–040, nov 2004. doi: 10.1088/1126-6708/2004/11/040. URL <https://doi.org/10.1088%2F1126-6708%2F2004%2F11%2F040>.
- [51] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007(11):070–070, nov 2007. doi: 10.1088/1126-6708/2007/11/070. URL <https://doi.org/10.1088%2F1126-6708%2F2007%2F11%2F070>.
- [52] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. doi: 10.1007/JHEP06(2010)043.
- [53] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini. Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM. *JHEP*, 02:088, 2012. doi: 10.1007/JHEP02(2012)088.
- [54] Keith Hamilton, Paolo Nason, Emanuele Re, and Giulia Zanderighi. NNLOPS simulation of Higgs boson production. *JHEP*, 10:222, 2013. doi: 10.1007/JHEP10(2013)222.
- [55] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. Finite quark-mass effects in the NNLOPS POWHEG+MinLO Higgs generator. *JHEP*, 05:140, 2015. doi: 10.1007/JHEP05(2015)140.

- [56] Paolo Nason and Carlo Oleari. NLO Higgs boson production via vector-boson-fusion matched with shower in POWHEG. *JHEP*, 02:037, 2010. doi: 10.1007/JHEP02(2010)037.
- [57] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano.  $\text{HW}^\pm/\text{HZ} + 0$  and 1 jet at NLO with the POWHEG BOX interfaced to gosam and their merging within MiNLO. *JHEP*, 10:083, 2013. doi: 10.1007/JHEP10(2013)083.
- [58] Heribertus B. Hartanto, Barbara Jager, Laura Reina, and Doreen Wackerlo. Higgs boson production in association with top quarks in the POWHEG BOX. *Phys. Rev. D*, 91:094003, 2015. doi: 10.1103/PhysRevD.91.094003.
- [59] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 10 2016.
- [60] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Elisabetta Furlan, Thomas Gehrmann, Franz Herzog, Achilleas Lazopoulos, and Bernhard Mistlberger. High precision determination of the gluon fusion Higgs boson cross-section at the LHC. *JHEP*, 05:058, 2016. doi: 10.1007/JHEP05(2016)058.
- [61] Matteo Cacciari, Frédéric A. Dreyer, Alexander Karlberg, Gavin P. Salam, and Giulia Zanderighi. Fully differential vector-boson-fusion Higgs production at next-to-next-to-leading order. *Phys. Rev. Lett.*, 115:082002, 2015. doi: 10.1103/PhysRevLett.115.082002. [Erratum: DOI10.1103/PhysRevLett.120.139901].
- [62] Oliver Brein, Abdelhak Djouadi, and Robert Harlander. NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders. *Phys. Lett. B*, 579:149, 2004. doi: 10.1016/j.physletb.2003.10.112.
- [63] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackerlo. Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections. *Phys. Rev. D*, 68:034022, 2003. doi: 10.1103/PhysRevD.68.034022.
- [64] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. Weak corrections to Higgs hadroproduction in association with a top-quark pair. *JHEP*, 09:065, 2014. doi: 10.1007/JHEP09(2014)065.
- [65] Federico Demartin, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro. Higgs production in association with a single top quark at the LHC. *Eur. Phys. J. C*, 75: 267, 2015. doi: 10.1140/epjc/s10052-015-3475-9.
- [66] Federico Demartin, Benedikt Maier, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro. tWH associated production at the LHC. *Eur. Phys. J. C*, 77:34, 2017. doi: 10.1140/epjc/s10052-017-4601-7.
- [67] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension. *Comput. Phys. Commun.*, 108:56, 1998. doi: 10.1016/S0010-4655(97)00123-9.

- [68] Michael Spira. QCD effects in Higgs physics. *Fortsch. Phys.*, 46:203, 1998. doi: 10.1002/(SICI)1521-3978(199804)46:3<203::AID-PROP203>3.0.CO;2-4.
- [69] Ye Li and Frank Petriello. Combining QCD and electroweak corrections to dilepton production in FEWZ. *Phys. Rev. D*, 86:094034, 2012. doi: 10.1103/PhysRevD.86.094034.
- [70] Massimiliano Grazzini, Stefan Kallweit, Dirk Rathlev, and Marius Wiesemann.  $W^\pm Z$  production at the LHC: fiducial cross sections and distributions in NNLO QCD. *JHEP*, 05:139, 2017. doi: 10.1007/JHEP05(2017)139.
- [71] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev. ZZ production at the LHC: fiducial cross sections and distributions in NNLO QCD. *Phys. Lett. B*, 750:407, 2015. doi: 10.1016/j.physletb.2015.09.055.
- [72] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi.  $W^+W^-$  production at hadron colliders in next-to-next-to-leading order QCD. *Phys. Rev. Lett.*, 113:212001, 2014. doi: 10.1103/PhysRevLett.113.212001.
- [73] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 07:018, 2011. doi: 10.1007/JHEP07(2011)018.
- [74] Michal Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185:2930, 2014. doi: 10.1016/j.cpc.2014.06.021.
- [75] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer. HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. *Comput. Phys. Commun.*, 191:74, 2015. doi: 10.1016/j.cpc.2015.02.001.
- [76] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. doi: 10.1007/JHEP04(2015)040.
- [77] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77:663, 2017. doi: 10.1140/epjc/s10052-017-5199-5.
- [78] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473, 2008. doi: 10.1140/epjc/s10052-007-0490-5.
- [79] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012. doi: 10.1007/JHEP12(2012)061.
- [80] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13TeV. *JINST*, 13:P05011, 2018. doi: 10.1088/1748-0221/13/05/P05011.

- [81] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, oct 2017. doi: 10.1088/1748-0221/12/10/p10003. URL [https://doi.org/10.1088%2F1748-0221%2F12%2F10%2Fp10003](https://doi.org/10.1088/2F1748-0221%2F12%2F10%2Fp10003).
- [82] Wolfgang Adam, Boris Mangano, Thomas Speer, and Teddy Todorov. Track Reconstruction in the CMS tracker. Technical Report CMS-NOTE-2006-041, CERN, Geneva, Dec 2006. URL <http://cds.cern.ch/record/934067>.
- [83] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009–P10009, oct 2014. doi: 10.1088/1748-0221/9/10/p10009. URL <https://doi.org/10.1088/1748-0221/9/10/p10009>.
- [84] Are Strandlie and W Wittek. Propagation of Covariance Matrices of Track Parameters in Homogeneous Magnetic Fields in CMS. Technical Report CMS-NOTE-2006-001, CERN, Geneva, Jan 2006. URL <http://cds.cern.ch/record/927379>.
- [85] CMS Collaboration. Electron and photon reconstruction and identification with the cms experiment at the cern lhc, 2020.
- [86] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9–N20, jul 2005. doi: 10.1088/0954-3899/31/9/n01. URL <https://doi.org/10.1088/0954-3899/31/9/n01>.
- [87] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, M. Friedl, and et al. Observation of the diphoton decay of the higgs boson and measurement of its properties. *The European Physical Journal C*, 74(10), Oct 2014. ISSN 1434-6052. doi: 10.1140/epjc/s10052-014-3076-z. URL <http://dx.doi.org/10.1140/epjc/s10052-014-3076-z>.
- [88] CMS Collaboration. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Journal of Instrumentation*, 10(08):P08010–P08010, aug 2015. doi: 10.1088/1748-0221/10/08/p08010. URL <https://doi.org/10.1088/1748-0221/10/08/p08010>.
- [89] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-ktjet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008. doi: 10.1088/1126-6708/2008/04/063. URL <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [90] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. Fastjet user manual. *The European Physical Journal C*, 72(3), 2012. ISSN 1434-6052. doi: 10.1140/epjc/s10052-012-1896-2. URL <https://doi.org/10.1140/epjc/s10052-012-1896-2>.
- [91] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, feb

2017. doi: 10.1088/1748-0221/12/02/p02014. URL <https://doi.org/10.1088/1748-0221/12/02/p02014>.
- [92] Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu\tau$  at CMS. *Journal of Instrumentation*, 11(01):P01019–P01019, jan 2016. doi: 10.1088/1748-0221/11/01/p01019. URL <https://doi.org/10.1088/1748-0221/11/01/p01019>.
  - [93] CMS Collaboration. Performance of reconstruction and identification of  $\tau$  leptons decaying to hadrons and  $\nu\tau$  in pp collisions at  $\sqrt{s}=13$  TeV. *Journal of Instrumentation*, 13(10):P10005–P10005, oct 2018. doi: 10.1088/1748-0221/13/10/p10005. URL <https://doi.org/10.1088/1748-0221/13/10/p10005>.
  - [94] CMS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV using the CMS detector. *Journal of Instrumentation*, 14(07):P07004–P07004, jul 2019. doi: 10.1088/1748-0221/14/07/p07004. URL <https://doi.org/10.1088/1748-0221/14/07/p07004>.
  - [95] CMS collaboration. Performance of the CMS missing transverse momentum reconstruction in pp data at  $\sqrt{s} = 8$  TeV. *Journal of Instrumentation*, 10(02):P02006–P02006, feb 2015. doi: 10.1088/1748-0221/10/02/p02006. URL <https://doi.org/10.1088/1748-0221/10/02/p02006>.
  - [96] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup per particle identification. *Journal of High Energy Physics*, 2014(10), 2014. ISSN 1029-8479. doi: 10.1007/JHEP10(2014)059. URL [https://doi.org/10.1007/JHEP10\(2014\)059](https://doi.org/10.1007/JHEP10(2014)059).
  - [97] A. M. Sirunyan et al. Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV. *Phys. Rev. Lett.*, 122: 132003, Apr 2019. doi: 10.1103/PhysRevLett.122.132003. URL <https://link.aps.org/doi/10.1103/PhysRevLett.122.132003>.
  - [98] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018–P09018, sep 2020. doi: 10.1088/1748-0221/15/09/p09018. URL <https://doi.org/10.1088/1748-0221/15/09/p09018>.
  - [99] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov. Extracting muon momentum scale corrections for hadron collider experiments. *Eur. Phys. J. C*, 72: 2194, 2012. doi: 10.1140/epjc/s10052-012-2194-8.
  - [100] Albert M Sirunyan et al. Search for a new scalar resonance decaying to a pair of Z bosons in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JHEP*, 06:127, 2018. doi: 10.1007/JHEP06(2018)127, 10.1007/JHEP03(2019)128. [Erratum: JHEP03,128(2019)].
  - [101] Andrew M. Carnes. *A Search for the Standard Model Higgs Boson Decaying to Two Muons at the CMS Experiment*. PhD thesis, University of Florida, 2018.

- [102] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. doi: 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. [Erratum: Eur. Phys. J.C73,2501(2013)].
- [103] A L Read. Presentation of search results: the CLs technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693–2704, sep 2002. doi: 10.1088/0954-3899/28/10/313.