Measurement of the properties of the Higgs boson in CMS experiment at the LHC

A Thesis

Submitted to the

Tata Institute of Fundamental Research, Mumbai

for the degree of Doctor of Philosophy

in Subject Board of Physics

by
Soumya Mukherjee

Department of High Energy Physics

Tata Institute of Fundamental Research

Mumbai

Final Version Submitted in May, 2023

Abstract

Since the groundbreaking discovery of the Higgs boson (H) in 2012, the primary focus of the experiments conducted at the Large Hadron Collider (LHC) has been to accurately measure its properties. This essential physics program is driven by two key objectives: firstly, to validate the description of generation of mass as in the standard, and secondly, to utilize the measured properties of the H to explore the new physics phenomena, beyond the standard model physics. Towards this end this thesis describes essentially two analyses of current interest and high importance, carried out using the Run-2 p-p collision data collected by the CMS experiment at the LHC. The first study concerns the search for the non-resonant H pair produced through Vector Boson Fusion (VBF) mode in two photons and two b-jets final state. The constraints on the anomalous coupling of a pair of H with a pair of gauge bosons was determined for the first time by the CMS collaboration. Further, a projection study of the feasibility of the non-resonant H pair production in the context of the High luminosity LHC is also presented. In the second study the measurement of H through VBF production and subsequent decays to two b-quarks is carried out for the first time using Run-2 data from CMS. Though this study has used CMS datasets of 2016 and 2018, the derived signal strength is comparable with the analysis carried out by the ATLAS collaboration, using full dataset of Run-2. This CMS analysis also reported the first observation of the VBF production of followed by decay to b-quark pair, in a hadron collider.

Finally, two detector performance related studies are reported regarding a) the performance check of the missing transverse energy filters, and, b) the estimation of the resolution and the response of the missing transverse energy in the events containing photons along with the jets.

DECLARATION

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

The work was done under the guidance of Professor Kajari Mazumdar, at the Tata Institute of Fundamental Research, Mumbai.

Soumya Mukherjee

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Prof. Kajari Mazumdar

Date: 29. 5. 2023

Dedicated to my parents and my uncle

Acknowledgment

I am grateful to TIFR, the one of the leading scientific research institutes in India, for providing me with the very relevant and useful coursework programs and world class research opportunity to successfully grow my research career. I would like to thank all the TIFR members who went above and beyond to help me from every aspect of my stay at TIFR campus for the last 5 years. I would like to give my special thanks to some of them for their crucial role in my PhD days.

I am starting with giving my sincere thanks to my PhD supervisor Prof. Kajari Mazumdar for her constant support, guidance and patience for the last 4 years, which helped me to successfully finish my PhD thesis work in time. The research projects which she gave to me for my thesis are outstanding in the context of current research interest in the LHC physics programs. She always encouraged me to learn new things and to implement them in my research work. She has devoted her valuable time for me to discuss physics and to monitor my works regularly.

Next, I thank Prof. Gobinda Majumder, who is my PhD co-supervisor. I have learned many things about high energy physics from his course on experimental particle physics, which I audited and also from the discussions we had in our TIFR CMS group meetings. I am always grateful to him for the thorough checks of my thesis chapters and the important comments and suggestions. I would also like to thank Prof. Monoranjan Guchait, for his advice both in academic and nonacademic purposes. I am also thankful to other faculty members of the CMS group at TIFR, Prof. Gagan Mohanty, Prof. Sashi Dugad, Dr. Shilpi Jain and especially to Dr. Rajdeep Mohan Chatterjee. Rajdeep da was CMS Higgs to diphoton group Level-3 convener, when I started my first project, and later he was the analysis review committee member for my second PhD project and currently he is a faculty at TIFR; I have learned and am still learning many things from him. I thank our TIFR IndiaCMS Tier-2 grid computing administrators Brij Jashal and Punnet Patel for maintaining the system smoothly, without this system it is not possible for me at all to finish the physics analyses on time. Let me thanks to our scientific officer Mandakini Patil, who taught me the hardware aspects of the CMS Trigger systems many times. I also thank our Experimental High Energy Physics group secretary Minal Rane for all kinds of official support. I also acknowledge Dr. Tuhin S. Roy from the Department of Theoretical Physics group for monitoring my thesis work for the last four years and for the many useful discussions which ease the understanding of the subject.

I would like to thank those persons here, without whom it is not possible to start work in CMS,

I am always thankful to my girlfriend Rubia, for supporting me everyday. She is the only person who was always with me from morning to deep night over the phone during my working days, so that I would not be bored. It is not possible for me to express in words the impact of the mental support I got from her in the last five years.

Finally I thank my father, mother and my uncle. They did a lot of sacrifices and struggles to build this position where I am currently now. Only for their support and constant motivation; I came to do PhD. I also thank my other family members for their constant support.

Publications and public documents

1. Search for non-resonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s}$= 13 TeV.
   CMS Collaboration
   arXiv:2011.12373, JHEP 03 (2021) 257

2. Prospects for non-resonant Higgs boson pair production measurement in $bb\gamma\gamma$ final states in proton-proton collisions at $\sqrt{s}$=14 TeV at the High-Luminosity LHC.
   CMS Collaboration
   CMS-PAS-FTR-21-004

3. Measurement of the signal strength of the Higgs boson decaying to bottom quarks and produced from Vector Boson Fusion using proton-proton collision data at $\sqrt{s}$=13 TeV.
   CMS Collaboration
   CMS-PAS-HIG-22-009, Target Journal: JHEP

4. Mitigation of anomalous missing transverse momentum measurements in data collected by CMS at $\sqrt{s}$= 13 TeV during the LHC Run 2.
   CMS Collaboration
   CMS-DP-2020/018

5. Performance of missing transverse momentum reconstruction in events containing a photon and jets collected by CMS during proton-proton collisions at $\sqrt{s}$= 13 TeV in 2018
   CMS Collaboration
   CMS DP-2020/031

6. HL-LHC Experiment Data Quality Working Group Summary Report
   CMS Collaboration
   CERN-BE-2022-001

7. The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger Techinical Design Report, CERN-LHCC-2021-007, CMS-TDR-022.

8. Prospects for nonresonant HH production measurement in bb$\gamma\gamma$ final states with CMS at the HL-LHC. [sec: 2.4.1, Snowmass ATLAS-CMS white paper, CMS-PAS-FTR-22-001]

CMS internal documents

1. Search for non resonant H(bb)H($\gamma\gamma$) for Run 2.
   CMS analysis note: **CMS AN-17-286**
   Contributors: N. Chernyavskaya, M. Donega, S. Pigazzini, K. Mazumdar, **S. Mukherjee** and others

2. Update on HHbbgg studies for EDQ.
   CMS analysis note: **AN-20-153**
   Contributors: K. Mazumdar, **S. Mukherjee**

3. HL-LHC projection for the Higgs boson self coupling: non-resonant diHiggs production in bb$\gamma\gamma$ final state.
   CMS analysis note: **AN-21-223**
   Contributors: K. Mazumdar, **S. Mukherjee**

4. Measurement of the Higgs boson production via Vector Boson Fusion process with subsequent decay of the Higgs boson into a pair of bottom quarks.
   CMS analysis note: **AN-21-045**
   Contributors: K. Mazumder, **S. Mukherjee**, A. Raspereza, A. Tumasyan

5. Performance of missing transverse momentum in pp collision at 13 TeV with the full Run2 dataset. CMS analysis note: **AN-21-167**
   Contributors: K. Mazumder, **S. Mukherjee** and others

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Extensive experimental measurements during the last several decades have established the Standard Model (SM) of particle physics to be the most complete and consistent description of the subatomic world. The Higgs boson was the last fundamental particle of the standard model, to be discovered by the ATLAS and the CMS experiments in 2012 at the CERN LHC. Over a decade after the discovery, both of the experimental collaborations are still continuing the exploration to characterize of the Higgs boson ever more precisely. The Higgs sector occupies the centre stage in the mathematical structure of the SM, being responsible for the generation of mass of all the elementary particles, including the Higgs boson itself, via spontaneous electroweak symmetry breaking mechanism [28, 29]. It envisages mass-dependent interactions of the particles with the Higgs field (ϕ). Though the SM predicts the nature of the fundamental interactions of the elementary particles, each mass value is to be determined experimentally. The mass, in turn, defines the strength of the interaction. Explicitly, the coupling of the Higgs boson with a massive gauge boson, ($V = W/Z$), with mass M_V, is given as $g_{HVV} = C_V = \frac{2M_V^2}{v}$; for a fermion with mass m_f, it is the Yukawa coupling, $g_{Hf\bar{f}} = y_f = \frac{\sqrt{2}m_f}{v}$. Here $v = 246$ GeV is the vacuum expectation value of the Higgs field after the electroweak symmetry is broken. The self coupling of the Higgs boson follows a different structure and is given by $\lambda = m_H^2/2v$, for Higgs boson mass at 125 GeV, the value of λ is expected to be 0.13.

In this thesis work, we present essentially two studies, using the LHC Run 2 proton-proton (p-p) collision data collected by the CMS experiment during 2016-2018, towards better understanding the properties of the Higgs boson: (i) the self-coupling, and, (ii) the Yukawa coupling of Higgs boson to the bottom quark. In both cases ample use of machine learning tools, as

applicable, have been used. Further, the signal sensitivity is increased via event categorization based on the purity.

The thesis starts with the description of the standard model physics in Chapter: 2 by recapitulating salient features of the fundamental interactions, including Higgs mechanism. Chapter 3 describes the early searches for the Higgs boson in the Large Electron-Positron collider the Tevatron, proton-antiproton collider at Fermilab. This chapter also highlights the post-discovery studies towards precision measurements of its properties and finally the current focus for the diHiggs searches.

The next two Chapters. 4 and 5, present briefly of the experimental apparatus: the LHC machine and the CMS detector system. The CMS experiment is a highly complex system with various subdetectors meant for measuring energy and momenta of various stable particles, like electron, photon, muon, hadronic jets etc. The reconstruction of the different physics level objects from the detector level hits or energy deposition is described in Chapter. 6.

The non-resonant process of inclusive di-Higgs boson production, followed by the decays to a pair of photons and a pair of b-quarks, is studied for the measurement of the self-coupling (λ) as well as the coupling of the pair of Higgs boson to a pair of gauge boson ($g_{HHVV} = C_{2V}$) through the vector boson fusion production mode. The strategy is to identify the signal-like events containing possible production of the Higgs boson pair with the candidates decaying to a pair of photons (H $\rightarrow \gamma\gamma$) and to a pair of bottom quarks (H $\rightarrow b\bar{b}$). The individual invariant mass distributions are accounted for the signal as well as the background contributions, and fitted simultaneously to extract the signal strength and eventually interpret the values of the couplings. As an extension, a projection study for the future high luminosity operation of LHC (HL-LHC) was also carried out to estimate the physics potential of the CMS experiment. They are described in Chapter 7 and 8 respectively.

The other major part of the thesis concerns the measurement of single Higgs boson production through vector boson fusion (VBF) mode and decaying to a bottom quark-antiquark pair to extract the value of the bottom quark Yukawa coupling; the analysis is presented in chapter. 9. Here also similar strategy, as in case of di-Higgs analysis, has been followed for extracting the relevant parameters of interest.

The studies related to the performance of the detector, which are crucial for event reconstruction

and physics analyses are also part of the thesis as well, summarized in Chapter. 10.

Chapter 2

The standard model of particle physics

2.1 Introduction

The standard model of particle physics (SM) is based on quantum field theory which can describe all the known elementary particles and the fundamental interactions they corroborate to. The SM does not encompass the gravitational interaction since masses involved are extremely tiny. The mathematical structure of the SM is based on a four dimensional space-time Lagrangian describing the dynamics and the interactions of all the fields representing various particles. This must be invariant under the unitary transformation of the guage group of $SU(3)_C \times SU(2)_L \times U(1)_Y$ corresponding to the strong, weak and electromagnetic interactions respectively. Each of the gauge transformations describes a specific force among the matter particles, spin-$\frac{1}{2}$ fermions, which is mediated by the corresponding carrier(s) spin-1 boson(s); eg., the massless photon (γ) for $U(1)_Y$, the massive $W^{\pm}$ and Z for $SU(2)_L$, and the massless gluons for $SU(3)_C$. Each particle has a set of quantum numbers. For a given type of interaction, certain types of quantum numbers must be conserved. They are: (i) the electric charge ($Q \cdot e$) for the electromagnetic (EM) interaction described by the theory of quantum electrodynamics (QED), (ii) the weak isospin (I_3) for the weak interaction which can be explained by the electroweak (EW) theory, and, (iii) the color charge (c) for the strong interaction described by the theory of quantum-chromodynamics (QCD). There is an additional, different type of interaction caused by the presence of the Higgs spin-0 scalar field which lead to the mass generation of the elementary particles according to the strength of the interaction.

At the most fundamental level known so far, the matter consists of twelve fermions whose main

properties have already been determined by various experiments. These can be classified into two families based on the nature of their interactions the quarks and the leptons. The quarks possess color and electric charge (either $+\frac{2}{3}$ for up type or $-\frac{1}{3}$ down type) as well as the weak isospin; hence they take part in all three type of interactions. The leptons, on the other hand, have the electric charge of -1 (electrons, muons and taus) and 0 (neutrinos), and, the weak isospin; they can take part only in the electromagnetic and the weak interactions. Each family is further subdivided into 3 generations, with the masses increasing gradually. Thus only the particles belonging to the first generation are stable. Each generation of each family has a pair of particles belonging to a doublet structure with the difference in the electric charges being one unit. Figure2.1 represents the complete particle content of the SM with the basic classifications described above.
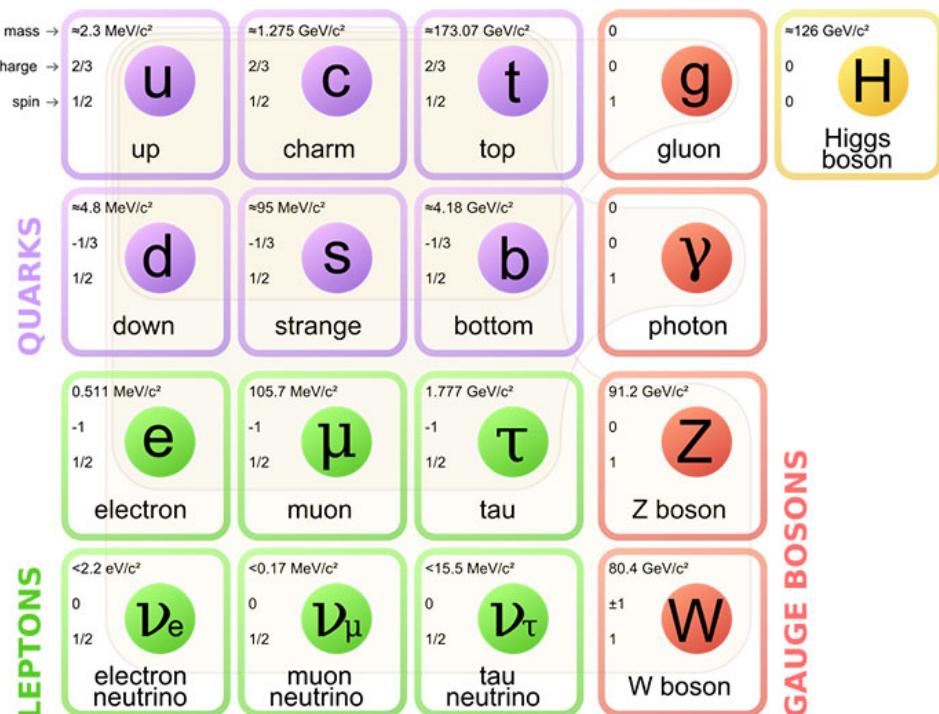
Figure 2.1: The standard model of particle physics. Source: *Wikipedia*

We present below the basic description of the interactions referred to above.

2.1.1 Quantum electrodynamics (QED)

The Lagrangian density for a free fermion field (or the 4-component spinor representing the wavefunction) ψ of mass m can be written, in natural units ($\hbar = c = 1$), as

$$\mathscr{L}_{QED}^{free} = i\bar{\psi}\gamma_\mu\partial^\mu\psi - m\bar{\psi}\psi + \text{h.c.} \tag{2.1}$$

We have included here explicitly the Hermitian conjugate as h.c., though we shall keep it implicit from now on. Here γ_μ are the 4×4 Pauli matrices. A local U(1)$_Q$ gauge transformation on the ψ, with Q as a parameter, can be expressed as

$$\psi(x) \rightarrow e^{iQ\alpha(x)}\psi(x), \tag{2.2}$$

To make the Lagrangian invariant under the above transformation, a vector gauge field $A_\mu(x)$ must be introduced, which should transform as

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu\alpha(x) \tag{2.3}$$

in addition to the required replacement of the ordinary derivative ∂_μ by a covariant derivative in the Lagrangian:

$$D^\mu = \partial^\mu - iQA^\mu \tag{2.4}$$

The Lagrangian in Eq. 2.1 describing only a free charged fermion now should be modified to include the effect of the newly introduced vector field A^μ as well. Including the kinetic term for A^μ in the Lagrangian,

$$\mathscr{L}_{QED} = i\bar{\psi}\gamma_\mu D^\mu\psi - m\bar{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.5}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor. By replacing D^μ in terms of the ordinary derivative ∂^μ we get the explicit form of the interaction of the fermion with the vector field. We call it vector current and the field A_μ is identified with the photon. The arbitrary parameter Q now represents strength of the coupling of ψ with A_μ.

$$\mathscr{L}_{QED} = i\bar{\psi}\gamma_\mu\partial^\mu\psi - m\bar{\psi}\psi - iQ\bar{\psi}\gamma_\mu A^\mu\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.6}$$

This expression describes the Lagrangian of the relativistic quantum theory of light with matter particle, or the QED.

2.1.2 Electroweak Theory

Unlike EM interaction, the mediators of the charged current weak interaction (eg., beta decay) electrically charged and hence they couple to the photon. Thus the theory of weak interaction actually encompasses both weak and EM interactions via the gauge group $SU(2)_L \times U(1)_Y$. Experimental observations established that only the left handed fermions participate in the weak interaction, and equally importantly, the charge-parity (CP) conservation is not respected. Under the $SU(2)_L$ transformation left and right handed fermions transform differently: as left chiral doublets (also called weak-isospin doublets) and a right chiral singlets (weak-isospin singlets) respectively. We can write,

$$
\begin{aligned}
\psi &= \psi_L + \psi_R \\
\psi_L &= \frac{1 - \gamma_5}{2} \psi \\
\psi'_R &= \frac{1 + \gamma_5}{2} \psi'
\end{aligned}
\tag{2.7}
$$

where $\frac{1-\gamma_5}{2}$ and $\frac{1+\gamma_5}{2}$ are the left- and right-handed projection operators respectively. For example, the representation of the first generation of leptons can be written as:

$$
\psi_L = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}, \ \psi_R = e_R, \nu_{eR}
\tag{2.8}
$$

In the electroweak theory, the gauge invariance of $SU(2)_L \times U(1)_Y$ transformation conserves two quantities: the weak isospin I (and its third component I_3) and the weak hypercharge Y. For the up-type quarks and the neutrinos, $I_3 = 1/2$ while for the down type quarks and the charged leptons (e, μ, τ) $I_3 = -1/2$. The Gellmann-Nishijama formula [30–32] provides the relation between the electric charge (Q), third isospin component (I_3) and the hypercharge (Y).

$$
Q = I_3 + Y
\tag{2.9}
$$

The mathematical treatment of weak interaction is similar to QED. The gauge invariance of the Lagrangian under $SU(2)_L$ transformation results in the generation of 3 gauge fields W_μ^i, $(i = 1, 2, 3)$. The interaction of the left-handed fermion doublet with the local $U(1)_Y$ group results in the single gauge field B_μ, which, however, interacts with both the left- and right-handed fermions. Thus the Lagrangian density of the electroweak theory can be presented as:

$$
\mathcal{L}_{EWk} = i\bar{\psi}_L \gamma_\mu D^\mu \psi_L + i\bar{\psi}_R \gamma_\mu D^\mu \psi_R + i\bar{\psi}'_R \gamma_\mu D^\mu \psi'_R
\tag{2.10}
$$

Here the covariant derivative D_μ suitable for making the Lagrangian guage invariant must be defined differently for the left- and the right-handed fermions:

$$D_\mu(\text{left handed}) = \partial_\mu - ig \sum_{j=1}^{3} W_\mu^j T_j - ig' \frac{Y}{2} B_\mu \tag{2.11}$$

$$D_\mu(\text{right handed}) = \partial_\mu - ig' Y B_\mu \tag{2.12}$$

where for the left-handed fermion ψ_L, the generators for the $SU(2)_L$ group are $T_j = \frac{\sigma_j}{2}$ with $\sigma_j, j = 1, 2, 3$ being the three 2×2 Pauli matrices; for the right-handed fermion ψ_R, the generator is $T_i = 0$. Here g and g' are the coupling strength for the charged and the neutral current interactions respectively. To keep the Lagrangian invariant, the 4 gauge fields are needed to be transformed as follows.

$$W_\mu \to e^{-\frac{i}{2}\sigma_j\alpha_j} W_\mu e^{\frac{i}{2}\sigma_j\alpha_j} - \frac{i}{g}(\partial_\mu e^{\frac{i}{2}\sigma_j\alpha_j}) e^{-\frac{i}{2}\sigma_j\alpha_j} \tag{2.13}$$

$$B_\mu \to B_\mu + \frac{i}{g'}\partial_\mu\beta \tag{2.14}$$

Using the definition of the covariant derivative, and the transformed forms of the gauge fields, Eq. 2.10 can be written as:

$$\mathscr{L}_{EWk} = \underbrace{[i\bar{\psi}_L\gamma_\mu\partial^\mu\psi_L + i\bar{\psi}_R\gamma_\mu\partial^\mu\psi_R + i\bar{\psi}'_R\gamma_\mu\partial^\mu\psi'_R]}_{\mathscr{L}_{kin}} +$$

$$\underbrace{\frac{g}{\sqrt{2}}[W^+\bar{\psi}_L\gamma^\mu\psi'_L + W^-\bar{\psi}'_L\gamma^\mu\psi_L]}_{\mathscr{L}_{charged\ current}} +$$

$$\underbrace{\frac{g}{\sqrt{2}}W_\mu^3[\bar{\psi}_L\gamma^\mu\psi_L - \bar{\psi}'_L\gamma^\mu\psi'_L] + \frac{g'}{\sqrt{2}}B_\mu[Y_L(\bar{\psi}_L\gamma^\mu\psi_L + \bar{\psi}'_L\gamma^\mu\psi'_L) + Y_R\bar{\psi}'_R\gamma^\mu\psi'_R + Y'_R\bar{\psi}'_R\gamma^\mu\psi'_R]}_{\mathscr{L}_{neutral\ current}}$$

$$\tag{2.15}$$

In the Eq. 2.15 the first, second and the third terms correspond to the kinetic part of the left- and the right-handed fermions, the charged and the neutral currents respectively. Here $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \pm iW_\mu^2)$ and referred to as W$^+$ and W$^-$ bosons respectively. Neutral currents are described by the linear combinations of the two neutral fields W^3 and B_μ.

$$\begin{bmatrix} Z_\mu \\ A_\mu \end{bmatrix} = \begin{bmatrix} cos\theta_W & -sin\theta_W \\ sin\theta_W & cos\theta_W \end{bmatrix} \begin{bmatrix} W_\mu^3 \\ B_\mu \end{bmatrix} \tag{2.16}$$

Here Z_μ and A_μ are identified to be carriers of the weak neutral current, the Z boson, and the photon field of QED respectively; θ_W refers to the mixing between the fields W_μ^3 and B_μ, called

"weak mixing angle" or the "Weinberg angle". Further more, the transformation of A_μ can describe the EM interaction provided

$$g\sin\theta_W = g'\cos\theta_W = e \tag{2.17}$$

where e is electric charge of the electron. Now the electroweak Lagrangian can be expressed as:

$$\begin{aligned}
\mathscr{L}_{EWk} = i\bar{\psi}_L\gamma_\mu D^\mu\psi_L + i\bar{\psi}_R\gamma_\mu D^\mu\psi_R + i\bar{\psi}'_R\gamma_\mu D^\mu\psi'_R \\
-\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}\sum_j W_i^{\mu\nu}W_{\mu\nu}^i
\end{aligned} \tag{2.18}$$

Here $B_{\mu\nu}$ and $W_{\mu\nu}$ are the two field strength tensors defined as

$$\begin{aligned}
B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \\
W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk}W_\mu^j W_\nu^k
\end{aligned} \tag{2.19}$$

The first three terms are correspond to the free fermion Dirac Lagrangian for the left and the right handed fermions the next one is the kinetic term of the gauge boson and, importantly, the last one defines the self-interactions of the gauge bosons representative Feynman disarms of which are shown in Fig. 2.2.
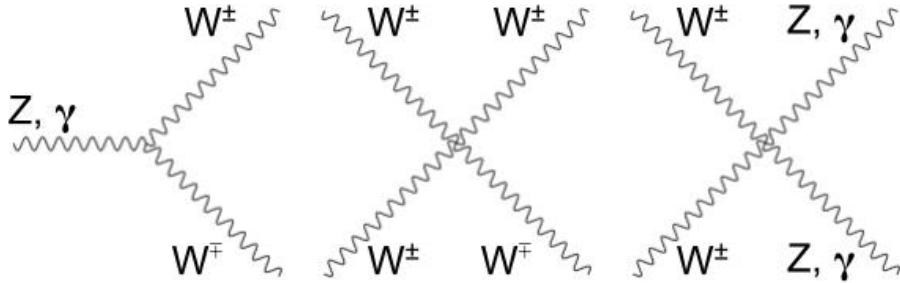
Figure 2.2: Representative Feynman diagrams of the EWK processes consisting the gauge bosons, $W^\pm$, Z and photon.

It is to be noted that the mass terms of the fermions are not present in the Eq. 2.15. The mass terms for the fermions can be written in the Lagragian as: $\mathscr{L}_{mass} = m\bar{\psi}\psi = m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$. Since the fields for the left and the right handed fermions ψ_L and ψ_R transform differently under $SU(2)_L$ group, the mass term violates the gauge invariance, and hence can not be present in the electroweak Lagrangian. Therefore, all the fermions and gauge bosons in electroweak Lagrangian are considered massless so far. The simplest method for the generation of the masses of the fundamental particles in the theory will be dicsussed shortly.

2.1.3   Quantum Chromodynamics

The strong force is described by the theory of quantum chromodynamics (QCD) based on the principle of the local gauge invariance of the $SU(3)_c$ group, where a new quantum numbers called the *color* charge, and denoted by c, is the conserved quantity under the corresponding transformation. Having color charges, only the quarks (and antiquarks) can participate in the strong interactions which is mediated by massless spin-1 gauge bosons, called the *gluons*. The free-field Lagrangian density of the fermions can be presented as:

$$\mathscr{L} = \psi(\bar{x})i\gamma^\mu \partial_\mu \psi(x) \tag{2.20}$$

where $\psi(x)$ is now an $SU(3)_C$ triplet of the quark fields. The local gauge incvariance can be written as:

$$\psi(x) \rightarrow e^{ig_s\beta_a\theta_a(x)}\psi(x) \tag{2.21}$$

where β_a are the eight $(3^2 - 1 = 8)$ Gell-Mann matrices which are generated according to the $SU(3)_C$ group and it also produces the 8 vector Gluon fields G_μ. Here g_s is related to the strong coupling constant $\alpha_s = \frac{g_s^2}{4\pi}$. A similar treatment applied to QED, the normal derivative needs to be replaced by the covariant derivative defined by,

$$D_\mu = \partial_\mu - ig_s t^a G_a^\mu \tag{2.22}$$

and the field strength tensor can be expressed as,

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_{b\mu} G_{c\nu} \tag{2.23}$$

The QCD Lagrangian under the $SU(3)_C$ gauge group can now be written as:

$$\begin{aligned}
\mathscr{L}_{QCD} &= \bar{\psi}(i\gamma_\mu D^\mu - m)\psi - \frac{1}{4}G^{\mu\nu}G_{\mu\nu} \\
&= \underbrace{i\bar{\psi}\gamma_\mu\partial^\mu\psi}_{\mathscr{L}_q^{kin}} - \underbrace{m\bar{\psi}\psi}_{\mathscr{L}_q^m} - \underbrace{g_s\bar{\psi}\gamma_\mu t^a G_a^\mu\psi}_{\mathscr{L}_q^{int}g} - \underbrace{\frac{1}{4}(\partial_\mu G_\mu^a)(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)}_{\mathscr{L}_g^{kin}} \\
&\quad - \underbrace{\frac{g_s}{2}f^{abc}(\partial_\mu G_{\nu a} - \partial_\nu G_{\mu a})G_b^\mu G_c^\nu}_{\mathscr{L}_{3g}^{int}} - \underbrace{\frac{g_s^2}{4}f^{abc}f^{ade}G_{b\mu}G_{c\nu}G_d^\mu G_e^\nu}_{\mathscr{L}_{4g}^{int}}
\end{aligned} \tag{2.24}$$

under the transformation of gluon fields as:

$$G_\mu^a \rightarrow G_\mu^a - \frac{1}{g_s}\partial_\mu\alpha^a - f^{abc}\alpha_b G_{c\mu} \tag{2.25}$$

The terms $\mathscr{L}_q^{kin}$ and $\mathscr{L}_g^{kin}$ are the propagators corresponding to the quarks and gluons while $\mathscr{L}_{qg}^{int}$ represents the interactions between the gluons and quarks; a representative Feynman diagram of the quark-gluon interaction is shown in Fig. 2.3 (left). The terms $\mathscr{L}_{3g}^{int}$ and $\mathscr{L}_{4g}^{int}$ describe the interactions among three and four gluons respectively; the corresponding Feynman diagrams shown in the middle and the right of Fig. 2.3.
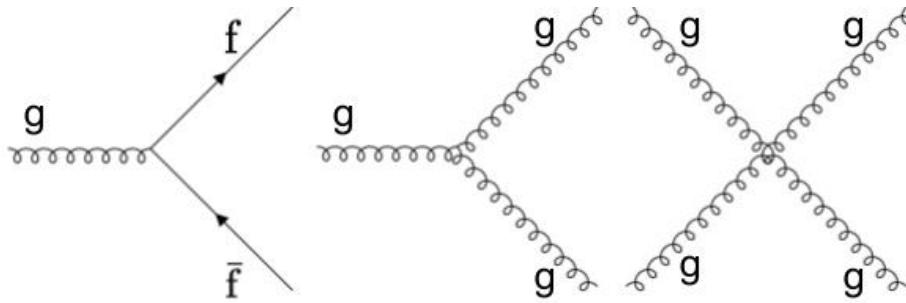
Figure 2.3: Feynman diagrams for the interaction between QCD fields : quark-gluon interaction vertex (left), three-gluon vertex (middle), four-gluon vertex (right).

2.1.4   Higgs mechanism and generation of mass

As mentioned already, the Lagrangian describing the electroweak and the strong interactions consider the fermions and the gauge bosons as massless particles and the introduction of the mass term in the Lagrangian would break the guage invariance. However, the experimental measurements reveal non-zero, unique masses of the fermions, as well as the W and Z bosons; this contradicts the theoretical assumptions discussed so far. The solution of this problem was suggested by three independent groups based on the ideas of Anderson and Nambu to explain the phenomenon of *spontaneous symmetry breaking* in context of condensed matter physics. Notably, the mechanism proposed by Englert and Brout, and Higgs [28, 29], referred to as BEH mechanism, is based on the spontaneous electroweak symmetry breaking (EWSB) of a local gauge field.

The BEH mechanism is responsible for the generation of masses of the gauge bosons and the fermions, where the massive gauge bosons (W $^\pm$ and Z) have 3 degrees of freedom and the massless boson (γ) has 2. In order to explain the origin of masses of fermions and bosons, a complex scalar $\Phi(x)$ (popularly known as the Higgs field) is introduced to the Lagrangian; it is

a doublet of complex scalar fields under $SU(2)_L$ transformation given by

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1(x) + i\phi_2(x) \\ \phi_3(x) + i\phi_4(x) \end{bmatrix} = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix} \tag{2.26}$$

where $\phi_1, \phi_2, \phi_3, \phi_4$ are normalized real scalar fields. The corresponding additional terms in the Lagrangian involving Φ can be written as:

$$\mathscr{L}_\Phi = (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi) + \mathscr{L}_{Yukawa} \tag{2.27}$$

Here the first term contains the kinetic and gauge invariance terms via the covariant derivative, the second describes a potential while the third one defines the Yukawa interaction of the scalar fields with a pair of fermions. In the most general approach, the structure of the potential $V(\Phi)$ is given as

$$V(\Phi) = -\mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \tag{2.28}$$

The shape of this potential depends on the choice of the parameters μ^2 and λ:

1. If $\lambda < 0$, then V is unbound without any stable state.

2. If $-\mu^2 > 0$ and $\lambda > 0$ the potential is symmetric around $|\Phi|=0$

3. If $-\mu^2 < 0$ and $\lambda > 0$, the symmetry of the potential is broken and the minimum lies away from $|\Phi|=0$. Here the vacuum or minimum energy state is not invariant under $SU(2)_L \times U(1)_Y$ guage transformations, which *spontaneously* breaks the symmetry of the vacuum.

Considering the last case among the three possibilities mentioned above, the miminum of the potential is found at:

$$\Phi^\dagger\Phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) \tag{2.29}$$

Minimization of the potential in Eq. 2.28 gives the values of:

$$|\Phi|^2 = \Phi^\dagger\Phi = \frac{\mu^2}{2\lambda} = \frac{v^2}{2} \tag{2.30}$$

Where $v = \sqrt{\frac{\mu^2}{\lambda}}$ is a vacuum expectation value (vev) of the scalar potential $V(\Phi)$. Now the shape of the potential in Eq. 2.28 can be written in terms of the four real scalars as

$$V = -\frac{\mu^2}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) + \frac{\lambda}{4}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2)^2 \tag{2.31}$$

In a compact 2-dimensional plane the shape is represented in Fig. 2.4
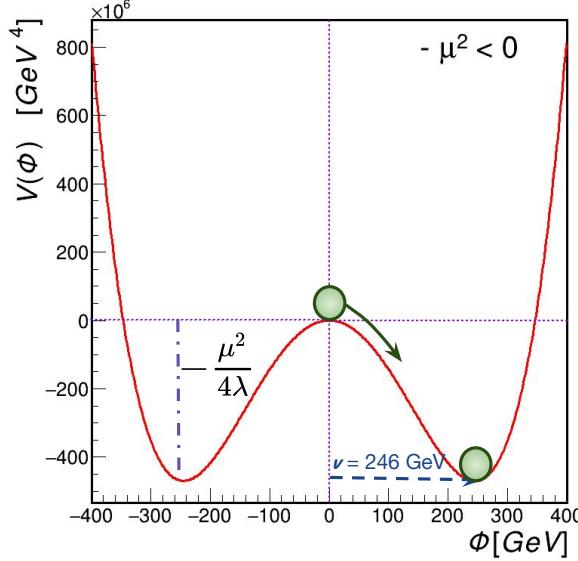
Figure 2.4: The Higgs potential $V(\Phi)$ as a function of Φ, for the condition of $-\mu^2 < 0$ and $\lambda > 0$.

By specifically choosing the basis of the states such that the expectation values are

$$< \phi_3 > = v = \sqrt{\frac{\mu^2}{\lambda}},$$

$$< \phi_1 > = < \phi_2 > = < \phi_4 > = 0$$

(2.32)

we then introduce a new scalar field h as a small perturbation around v, with $< h > = 0$. Thus

$$\phi_3 = h + v$$

(2.33)

Therefore, the Higgs field can be written as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 + i\phi_2 \\ v + h + i\phi_4 \end{bmatrix}$$

(2.34)

and the potential becomes

$$V(\Phi) = -\frac{\mu^2}{2}(\phi_1^2 + \phi_2^2 + (h+v)^2 + \phi_4^2) + \frac{\lambda}{4}(\phi_1^2 + \phi_2^2 + (h+v)^2 + \phi_4^2)^2$$

(2.35)

Using the relation $\mu^2 = \lambda v^2$, the above equation can be rewritten as

$$V(\Phi) = \text{constant} + 0 \cdot \phi_1^2 + 0 \cdot \phi_2^2 + \lambda v^2 h^2 + 0 \cdot \phi_4^2 + O(h^3) + O(h^4)$$

(2.36)

The qudratic terms in the above expression correspond to the mass terms for the real scalar components of the Higgs field. Thus $\phi_1, \phi_2, \& \phi_4$ are massless while the field h gets a non-zero mass given by $m_h = \sqrt{2\lambda v^2}$. In the above expansion of the perturbative field, the terms

$O(h^3)$ and $O(h^4)$ signify the tri-linear and quartic self-interactions which of the Higgs; their importance is explained in Sec. 3.5 in the context of this thesis. After the symmetry breaking the complex doublet field Φ can be written in a convenient form,

$$\Phi = \frac{1}{\sqrt{2}} exp\left(\frac{i\sigma_i\theta^i}{v}\right)\begin{bmatrix} 0 \\ v + h \end{bmatrix} \tag{2.37}$$

Here the three arbitrary rotation degrees of freedom θ^i correspond to the three massless Goldstone bosons which are generated by the spontaneous symmetry breaking. Since $SU(2)_L$ symmetry allows for a rotation, any dependence on the θ^i can be removed by a specific choice of the gauge transformation:

$$\Phi \rightarrow \Phi' = exp\left(\frac{-i\sigma_i\theta^i}{v}\right)\Phi = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ v + h \end{bmatrix} \tag{2.38}$$

Here only one real scalar massive field is left which is called the Higgs field and the the three degrees of freedom, corresponding to the Goldstone bosons, were absorbed by the W and Z bosons transverse **of the longitudinal** polarizations.

Generation of gauge boson masses

The effect of the covariant derivatives on Φ is given by

$$D_\mu\Phi = \frac{1}{\sqrt{2}}\begin{bmatrix} -\frac{i}{2}(W_\mu^1 - iW_\mu^2(v + h)) \\ \partial_\mu h + \frac{i}{2}(gW_\mu^3 - g'B_\mu)(v + h) \end{bmatrix} \tag{2.39}$$

Accordingly, the Lagrangian density can be now written as follows

$$\mathcal{L} = (D_\mu\Phi)^\dagger(D^\mu\Phi) = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g^2(v+h)^2(W_\mu^1 - iW_\mu^2)(W^{1\mu} + iW^{2\mu}) + \frac{1}{2}(v+h)^2(-g'B_\mu + gW_\mu^3)^2 \tag{2.40}$$

The first term is the kinetic term for the real scalar field h while the second term can be elaborated as

$$\begin{aligned} \mathcal{L} &= \frac{1}{8}g^2(v + h)^2(W_\mu^1 - iW_\mu^2)(W^{1\mu} + iW^{2\mu}) \\ &= \frac{1}{4}g^2(v + h)^2 W_\mu^+ W^{-\mu} \\ &= \frac{g^2v^2}{4}W_\mu^+ W^{-\mu} + \frac{g^2v}{2}hW_\mu^+ W^{-\mu} + \frac{g^2}{4}hhW_\mu^+ W^{-\mu} \end{aligned} \tag{2.41}$$

The first term of Eq. 2.41 is identified as the mass term for the W boson, $m_W^2 = \frac{g^2v^2}{4}$. From the direct measurements of m_W and indirect estimation of g from the measurement of muon

lifetime, we can estimate the value of "vev" v to be 246 GeV. The second and the third terms signify the interactions of one and two Higgs bosons respectively with a pair of W bosons, eg., W^+W^-. The form of these interactions can be uniquely predicted in terms of M_W and v.

$$
\begin{aligned}
hW^+W^- &: \quad i\frac{g^2 v}{2}g_{\mu\nu} = 2i\frac{M_W^2}{v}g_{\mu\nu} \\
hhW^+W^- &: \quad i\frac{g^2}{2}g_{\mu\nu} = 2i\frac{M_W^2}{v^2}g_{\mu\nu}
\end{aligned}
\tag{2.42}
$$

The third term of Eq. 2.40 can be written as:

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{8}(v+h)^2(gW_\mu^3 - g'B_\mu)^2 \\
&= \frac{1}{8}(g^2 + g'^2)(v+h)^2(\frac{g}{\sqrt{g^2+g'^2}}W_\mu^3 - \frac{g'}{\sqrt{g^2+g'^2}}B_\mu)^2 \\
&= \frac{1}{8}(g^2 + g'^2)(v+h)^2 Z^\mu Z_\mu \\
&= \frac{(g^2 + g'^2)v^2}{8}Z^\mu Z_\mu + \frac{(g^2 + g'^2)v}{4}hZ^\mu Z_\mu + \frac{(g^2 + g'^2)}{8}hhZ^\mu Z_\mu
\end{aligned}
\tag{2.43}
$$

The first term of the Eq. 2.43 gives the mass of the Z bosons as $M_Z = \sqrt{\frac{(g^2+g'^2)v^2}{4}}$, and the other terms generate the coupling of one or two Higgs bosons with a pair Z bosons. The corresponding couplings are given as

$$
\begin{aligned}
hZ_\mu Z_\nu &: \quad i\frac{(g^2 + g'^2)v}{2}g_{\mu\nu} = 2i\frac{M_Z^2}{v}g_{\mu\nu} \\
hhZ_\mu Z_\nu &: \quad i\frac{(g^2 + g'^2)}{2}g_{\mu\nu} = 2i\frac{M_Z^2}{v^2}g_{\mu\nu}
\end{aligned}
\tag{2.44}
$$

Finally the coupling strength ($C_V = HVV$, $V : W^\pm, Z$) of the interactions of a pair weak gauge boson $W^\pm$ and Z with a single Higgs boson is proportional to $\frac{m_V^2}{v}$, and the coupling strength of a pair of weak gauge bosons with a pair of Higgs boson ($C_{2V} = HHVV$) is proportional to $\frac{m_V^2}{v^2}$. These two types of the couplings appearing are shown in Fig. 2.5 through Feynman diagrams.

Generation of the lepton masses

The Lagrangian describing the Yukawa interaction of the Higgs doublets with the lepton fields has the structure in terms of left- and right-handed components; eg., for electron the Yukawa term is written as:

$$
\mathcal{L}_{\text{Yukawa}} = -\left(y_e \bar{e}_R \Phi^\dagger e_L + y_e^* \bar{e}_L \Phi e_R\right)
\tag{2.45}
$$
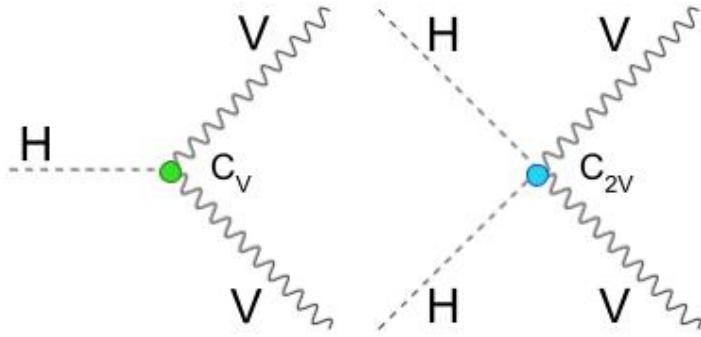
Figure 2.5: The coupling of the pair of vector bosons with single Higgs (left) and double Higgs (right), consisting C_V and C_{2V} couplings.

In the above Lagrangian the second term is the complex conjugate of the first term. The effect of $\Phi^\dagger$ on e_L can be written as:

$$\Phi^\dagger e_L = \begin{bmatrix} 0 & \frac{v+h}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} v_{eL} \\ e_L \end{bmatrix} = \frac{v+h}{\sqrt{2}} e_L \tag{2.46}$$

Hence we can rewrite Eq. 2.45 as

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}} &= -y_e \frac{1}{\sqrt{2}} [(v+h)\bar{e}_R e_L + (v+h)\bar{e}_L e_R] \\ &= -\frac{y_e}{\sqrt{2}} (v+h)\bar{e}e \\ &= -\frac{y_e v}{\sqrt{2}} \bar{e}e - \frac{y_e}{\sqrt{2}} h\bar{e}e \end{aligned} \tag{2.47}$$

The first term of Eq. 2.47 gives the mass of the electron as

$$m_e = \frac{y_e v}{\sqrt{2}} \tag{2.48}$$

The second term corresponds to the coupling of the Higgs bosons with the leptons, with the coupling strength (y_e) being directly proportional to the mass as:

$$h\bar{e}e : -\frac{iy_e}{\sqrt{2}} = -\frac{im_e}{v} \tag{2.49}$$

The above procedure is the same for higher generation charged leptons, ie., muon and tau. It is evident that higher the mass of the particle, larger is the coupling or the interaction strength with the Higgs field.

Generation of the quark masses

The structure of the Lagrangian for the interaction of quarks with the Higgs fields is the same as described above for the case of leptons, and for the first generation of quark doublet it is

written as

$$
\begin{aligned}
\mathcal{L}_{\text{Yukawa}} &= -[y_d \bar{d}_R \Phi^\dagger Q_L + y_d^* \bar{Q}_L \Phi d_R] \\
&= -\left(\frac{y_d v}{\sqrt{2}}\right) \bar{d}d - \frac{y_d}{\sqrt{2}} h\bar{d}d.
\end{aligned}
\tag{2.50}
$$

Again, the first term corresponds to the mass for the down quark, $m_d = \frac{y_d v}{\sqrt{2}}$ and the second term gives the coupling of the Higgs boson with the down quark.

$$
h\bar{d}d : -\frac{iy_d}{\sqrt{2}} = -\frac{im_d}{v}
\tag{2.51}
$$

To generate the mass of the up type quark we need to utilize the properties of the $SU(2)$ group. The conjugate of the Higgs doublet can be written as:

$$
\tilde{\Phi} = i\sigma^2 \Phi^* = i
\begin{bmatrix}
0 & -i \\
i & 0
\end{bmatrix}
\begin{bmatrix}
0 \\
\frac{v+h}{\sqrt{2}}
\end{bmatrix}
\tag{2.52}
$$

We use $\tilde{\Phi}$ in the gauge invariant Lagrangian as

$$
\mathcal{L} = -[y_u \bar{u}_R \tilde{\Phi}^\dagger Q_L + y_u^* \bar{Q}_L \tilde{\Phi} d_R]
\tag{2.53}
$$

where

$$
\tilde{\Phi}^\dagger Q_L =
\begin{bmatrix}
\frac{v+h}{\sqrt{2}} & 0
\end{bmatrix}
\begin{bmatrix}
u_L \\
d_L
\end{bmatrix}
= \frac{v+h}{\sqrt{2}} u_L
\tag{2.54}
$$

Therefore,

$$
\mathcal{L} = -\left(\frac{y_u v}{\sqrt{2}}\right) \bar{u}u - \frac{y_u}{\sqrt{2}} h\bar{u}u.
\tag{2.55}
$$

We identify the first term as the mass for the up quark, $m_y = \frac{y_u v}{\sqrt{2}}$ and the second term providing the interaction of the up quark with the Higgs field via the coupling.

$$
h\bar{u}u : -\frac{iy_u}{\sqrt{2}} = -\frac{im_u}{v}
\tag{2.56}
$$

As anticipated, as in case of leptons and down quark, the coupling strength of the Higgs with the up quark is proportional to its mass. This pattern repeats across generation. Hence for the most massive particle, the top quark, the Yukawa coupling is $y_t = \frac{\sqrt{2}m_t}{v} \sim 1$ corresponding to the measured mass of the top quark $m_t \sim 173$ GeV.

2.1.5 Conclusion

We have discussed the basic features of the standard model (SM) of particle physics in terms of electroweak theory, QCD, and the BEH mechanism. The Lagrangian describing the fundamental interactions among the all the known elementary particles and which is invariant under

$SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge transformation is presented as

$$
\begin{aligned}
\mathcal{L}_{SM} = &-\frac{1}{4}G_a^{\mu\nu}G_{a\mu\nu} - \frac{1}{4}W_i^{\mu\nu}W_{i\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \\
&+ (D^\mu\Phi)^\dagger(D_\mu\Phi) + i\bar{Q}_{Li}\gamma^\mu D_\mu Q_{Li} + i\bar{U}_{Ri}\gamma^\mu D_\mu U_{Ri} + i\bar{D}_{Ri}\gamma^\mu D_\mu D_{Ri} \\
&+ i\bar{L}_{Li}\gamma^\mu D_\mu L_{Li} + i\bar{E}_{Ri}\gamma^\mu D_\mu E_{Ri} \\
&+ Y_{ij}^u \bar{Q}_{Li} U_{Rij}\tilde{\Phi} + Y_{ij}^d \bar{Q}_{Li} D_{Rij}\Phi + Y_{ij}^e \bar{L}_{Li} E_{Rij}\Phi + h.c \\
&- V(\Phi)
\end{aligned}
\tag{2.57}
$$

Here the first row contains the three kinetic terms of the gauge bosons which convey various gauge interactions. Q_{Li} and L_{Li} are the three $SU(2)_L$ left-handed doublets for the quarks and the leptons respectively which take part in weak interaction. U_{Ri}, D_{Ri} and E_{Ri} are referred to as the $SU(2)_L$ singlets for the up and down type and the right handed charged leptons respectively. The terms consisting of Y_{ij} refer to the *Yukawa* couplings between Higgs and the fermions (both leptons and quarks) uniquely giving rise their masses. The main properties and various types of interactions of the all the fermions and the gauge bosons are summarized in Tab. 2.1.

Table 2.1: Properties of the known fermions and the gauge bosons in the standard model [23]. The electric charge, weak-isospin, weak-hypercharge are represented by Q, I_3 and Y respectively, where $Q = I_3 + Y/2$.

Particles	Mass	Q, I_3, Y	Generation	Interaction
Quarks				
up(u)	$2.16^{+0.49}_{-0.26}$ MeV	$+\frac{2}{3}, +\frac{1}{2}, +\frac{1}{3}$	I	Strong, EM, Weak
down(d)	$4.67^{+0.48}_{-0.17}$ MeV	$-\frac{1}{3}, -\frac{1}{2}, +\frac{1}{3}$		
charm(c)	1.27 ± 0.02 GeV	$+\frac{2}{3}, +\frac{1}{2}, +\frac{1}{3}$	II	
strange(s)	$93.4^{+8.6}_{-3.4}$ MeV	$-\frac{1}{3}, -\frac{1}{2}, +\frac{1}{3}$		
top(t)	172.69 ± 0.30 GeV	$+\frac{2}{3}, +\frac{1}{2}, +\frac{1}{3}$	III	
bottom(b)	$4.18^{+0.03}_{-0.02}$ GeV	$-\frac{1}{3}, -\frac{1}{2}, +\frac{1}{3}$		
Leptons				
electron(e)	0.511 MeV	$-1, -\frac{1}{2}, -1$	I	EM, Weak
ele neutrino(ν_e)	< 2.05 eV	$0, +\frac{1}{2}, -1$		Weak
muon(μ)	105.658 MeV	$-1, -\frac{1}{2}, -1$	II	EM, Weak
mu neutrino(ν_μ)	< 0.23 eV	$0, +\frac{1}{2}, -1$		Weak
tau(τ)	1.776 GeV	$-1, -\frac{1}{2}, -1$	III	EM, Weak
tau neutrino(ν_τ)	< .23 eV	$0, +\frac{1}{2}, -1$		Weak
Gauge Bosons				
W^+	80.377 ± 0.012 GeV	$+1, +1, 0$		EM,Weak
W^-	80.377 ± 0.012 GeV	$-1, -1, 0$		EM,Weak
Z	91.188 ± 0.002 GeV	0,0,0		Weak
gluon	0	0,0,0		Strong
photon (g)	0	0,0,0		EM
Higgs (H)	125.25 ± 0.17 GeV	$0, -\frac{1}{2}, \frac{1}{2}$		Weak

2.2 Physics at the hadron colliders

2.2.1 Parton density function (PDF)

For collisions between hadrons, the interacting partons carry fractional energies of the colliding protons. The exact energy fraction of the parton is not known but it can be interpreted in terms

of a probability distribution and is called the parton density function (PDF). Thus the PDFs are essential to calculate the physical (measurable) cross section of a given partonic interaction as discussed below. The PDF of a particular parton depends on its energy or momentum fraction (x) and the energy scale of the interaction (Q). Here x is also called Bjorken parameter, defined as $x = \frac{p_z}{E_b}$, where p_z is the longitudinal momentum of the parton and E_b is the energy of the parent hadron, the beam energy. The PDFs corresponding to different partons are extracted mainly from various measurements performed at deep inelastic scattering (DIS) as well as other hadron collider experiments at typically lower energy scales and then extrapolated to the higher energy scale. The PDF set of a given hadron satisfies the relation,

$$\sum_{i \epsilon \text{parton}} \int x f_i(x, Q) dx = 1 \tag{2.58}$$

This signifies that the total momentum of a colliding proton at the LHC is just the integration of the energy fraction of all the partons according to their PDFs. The cross section of a particular parton interaction can be calculated using the matrix element (ME) for the given transition which includes the momentum transfer and the couplings of the initial and final states which takes place. Figure 2.6 shows an example of a partonic interaction ($ab \rightarrow cd$), where a and b are the initial partons, carry x_1 and x_2 energy fraction of the two colliding hadrons h_1 and h_2 respectively, with corresponding PDFs $f_{a/h_1}(x_1)$ and $f_{b/h_2}(x_2)$ and c and d are the final state particles. The inclusive cross section at the total centre-of-mass energy ($\sqrt{s}$) can be expressed as:

$$\sigma_{\text{total}}^{ab \rightarrow cd}(s, Q) = \int_{x_1=0}^{1} \int_{x_2=0}^{1} f_1(x_1, Q) f_2(x_2, Q) \sigma_{\text{partonic}}^{ab \rightarrow cd}(x_1, x_2, s, Q) dx_1 dx_2 \tag{2.59}$$

here, the partonic cross section ($\sigma_{\text{partonic}}^{ab \rightarrow cd}(x_1, x_2, s, Q)$) depends on the sub-process energy $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$ and includes all the possible Feynman diagrams for the particular process. In the current state-of-the-art monte-carlo (MC) event generators, used for the LHC, the most precise PDF is Neural Network PDF(NNPDF) [33]. If the partonic interaction occurs with a large momentum transfer, the inelastic process produces high energy particles; this is called hard-scatter. The other partons which are not participating into the hard-scatter process, are the remnants. They produce soft interactions at lower energy scales, referred to as the underlying event (UE). Thus a complete description of a proton-proton collision includes the hard scatter as well as the UE followed by the parton shower, hadronization and the fragmentation process. This is depicted via the cartoon presented in Fig. 2.7 below.
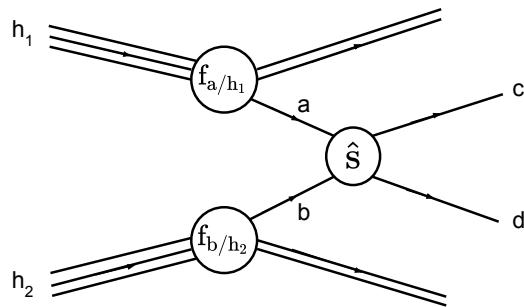
Figure 2.6: Feynman diagram of a hard-scattering process of two partons a & b coming from two hadrons h_1 and h_2 producing a final states with c and d, where f_{a/h_1} & $f_{b/h2}$ are the PDFs of the two hadrons.
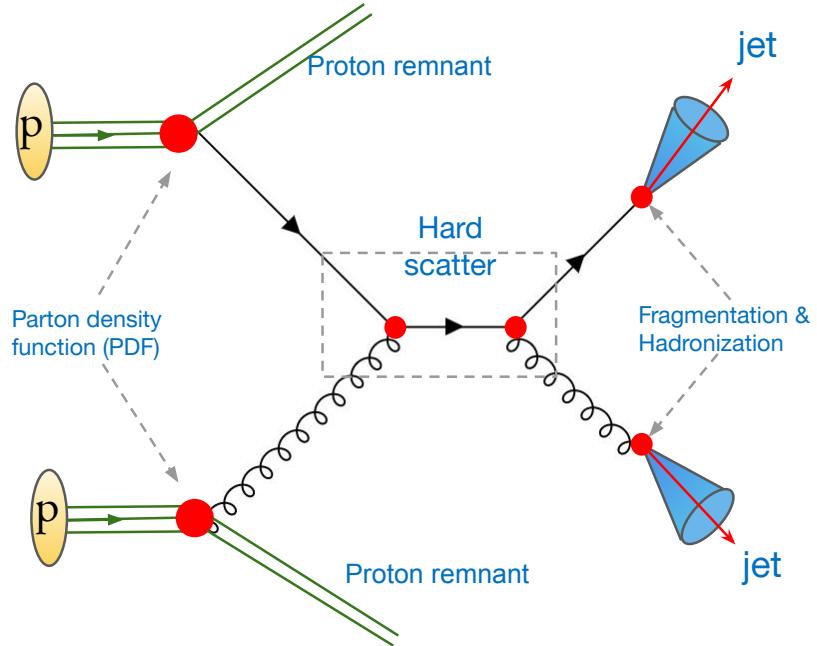
Figure 2.7: A cartoon pf a p-p collision event at the LHC from parton level to observational level at the experiment. Here the hard-scatter involves quark-gluon interaction at the sub-process level leading to two hadronic jets. The proton remnants are responsible for the under lying events (UE).

Chapter 3

The Higgs boson search

3.1 Early searches of the Higgs boson in the pre-LHC era

Various experiments in the 1970s and 1980s confirmed the general structure of the standard model and, broadly, the predictions concerning the gauge sector. However, the scalar sector, signifying the generation of mass via Higgs mechanism, remained to be established experimentally. In other words, the existence of a new type of fundamental scalar particle, the Higgs boson remained questionable. This affirmation requiring discovery of the particle continued to be elusive in next few decades making the SM to be an incomplete description. Like all other particles, the mass of the Higgs boson, M_H, is not predicted from the theory, through the nature of the interaction and other relevant aspects are. Various considerations allowed a wide possible range of M_H, up to about 750 GeV [34].

3.1.1 Search for the Higgs boson in the LEP experiments

The Higgs boson was searched extensively in the experiments at the large-electron-positron (LEP) collider at CERN, near Geneva, Switzerland using different production processes depending on the centre-of-mass energy ($\sqrt{s}$). In LEP1 era production of Z boson with subsequent decay through $e^- e^+ \rightarrow Z \rightarrow q\bar{q}$ channel was the main target. In LEP2 with higher $\sqrt{s}$ the Higgsstrahlung process opened up $e^- e^+ \rightarrow Z^* \rightarrow ZH$ and H was searched in the $b\bar{b}$ final state due to the largest decay branching ratio. This was combined with leptonic and hadronic decay modes of Z, with the second type providing good event statistics,

For the highest value of $\sqrt{s} = 206$ GeV at LEP2, the kinematic considerations allowed a

maximum value for M_H to be about 115 GeV. In spite of some experimental hints about possible production of H, the final conclusion from LEP by the turn of the century (2000) was a lower limit on the mass: $M_H > 114.4$ GeV at 95% confidence level (CL). Figure. 3.1 shows the exclusion of standard model Higgs boson hypothesis as a function of probed M_H.
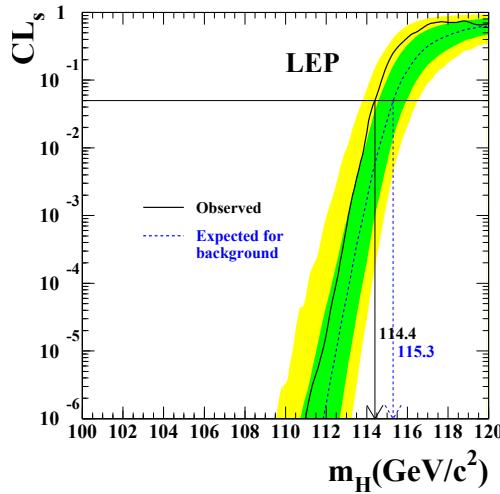
Figure 3.1: The exclusion plot of the standard model Higgs boson mass from the LEP experiment. The figure is taken from Ref. [1].

3.1.2 Search for the Higgs boson at the Tevatron

The H search continued extensively in Tevatron proton-antiproton collider at Fermilab, though being a hadron machine it was more difficult. Using the associated production or Higgsstrahlung, but now initiated via quark-antiquark pair the experimentally sensitive mass region was limited essentially between 140 to 180 GeV.

Just before the physics analysis started at the LHC, the CDF and D0 experiments at the Tevatron excluded the mass range of the H 162-166 GeV at the 95% CL [35, 36], which is shown in Fig. 3.2. Subsequently, a wider range was excluded further: $149 < M_H < 182$ GeV [37]. By the time H was discovered at the LHC in 2012, Tevatorn data also hinted at an excess of 3 standard deviations in the mass range of 115-140 GeV.
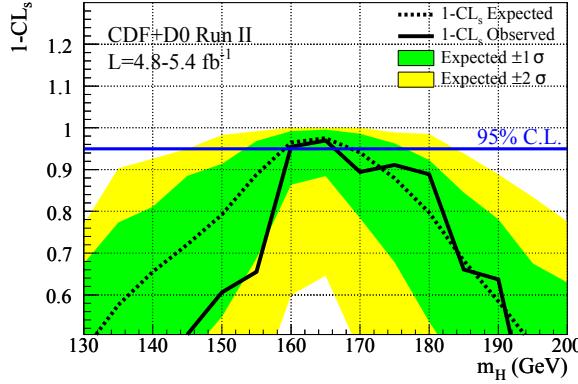
Figure 3.2: Tevatron results for the upper limit on the Higgs production cross section, relative to the SM prediction, as a function of the Higgs mass.

3.2 Search for the Higgs Boson at the LHC

By early 1990s search capability for the Higgs boson became a major benchmark for the experiments. Since the LHC was capable of producing the Higgs boson of any mass, the mandate was to hunt out the particle or resolve the issue of electroweak symmetry breaking. Accordingly, the search strategy and hence the detector design were focused for specific final states for different mass regions.

3.2.1 Production of the Higgs Boson

At the LHC, H can be produced by several mechanisms over the whole possible mass range, due to the availability of highly energetic partons. The major production modes are briefly discussed below.

1. **Gluon-Gluon Fusion (ggH)**: at the LHC energies, this is the most dominant production mode due to its much larger density in the proton compared to that of the quarks. It contributes almost 88 (85)% of the total Higgs cross section at the LHC CM energy of $\sqrt{}$=7, (13) TeV. In the SM, the direct coupling between H and gluons are not allowed, so the production of H proceeds via virtual quark loops. Due to much larger coupling of the H with top quark, the leading contribution is from top quark followed by the bottom quark. This process thus indirectly gives access to the top quark Yukawa coupling (y_t) from the virtual loop. A representative Feynman diagram of ggH process at the leading order (LO) is shown in Fig. 3.3 (left). The QCD corrections at higher orders of the

perturbation theory often leads to additional jets from radiations, but importantly they collectively lead to the increase in the inclusive ggH production cross section almost by a factor of two.
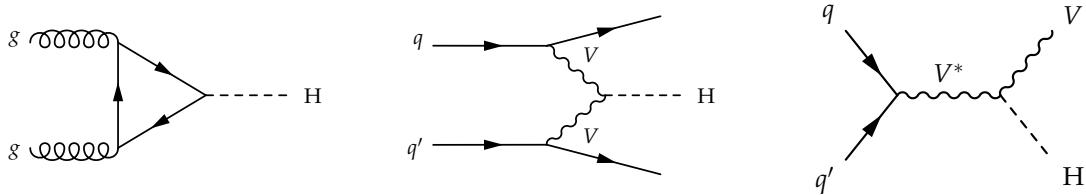
Figure 3.3: Left: Leading order Feynman diagram of ggH process, middle: VBF Higgs production and right: Higgsstrahlung (VH) process of Higgs boson with association with a Vector boson.

2. **Vector Boson Fusion (VBFH)**: This is the second largest production mode at the LHC with a reduced cross section by about a factor of ten than the ggH production process. As the name suggests and as shown in Fig. 3.3 (middle) this process gives direct access to the $HVV = C_V$ coupling. The colliding parton pair simultaneously radiate two vector bosons which are fused to produce the H in the central region. The outgoing quarks continue almost along the original direction. The distinctive topology of the event and the kinematics of the final state makes this process very unique.

3. **Higgsstrahlung (VH):** In this case the H is produced association with a weak interaction gauge boson (V), ie., either $W^{\pm}$ or Z as shown in Fig. 3.3 (right). As in case of VBF, the VH production mode also provides the access to C_V. Tagging the leptonic decay of V, the search for the hadronic decay of H is possible due to reduction of the QCD induced multijet backgrounds.

4. **In associate with top pair (t$\bar{\text{t}}$H)**: The tree-level production of the Higgs boson along a pair of top quarks has the unique property that it allows to measure directly the Higgs-top Yukawa coupling, y_t, in contrast with the y_t measurement from the virtual top quark loop. While top quark mass much more than $m_H/2$, kinematically H cannot decay to top quark pair, and hence this coupling can not be measured from the decay of H to top pair. Representative Feynman diagrams of t$\bar{\text{t}}$H process are shown in Fig. 3.4.

5. **In association with single top (tH)**: In the tH production mode the Higgs boson can be
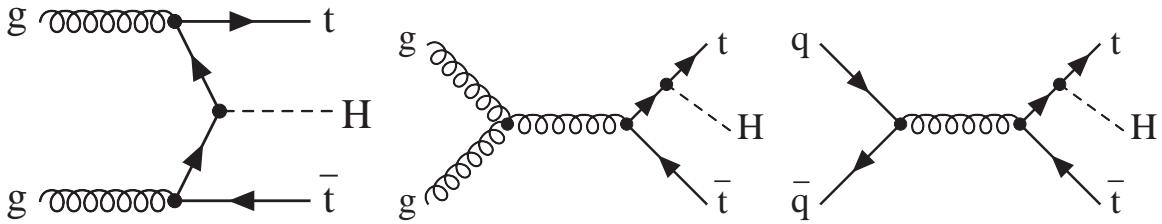
Figure 3.4: Representative tree-level Feynman diagrams of the $t\bar{t}$H production .

radiated either from the exchanged W boson or from the top quark in the two dominant leading order as shown in Fig. 3.5. The relative sign between the two Higgs boson couplings, y_t and C_V, decides the sign of the interference terms of the two diagrams; hence it can be determined by measuring the inclusive cross section. It is to be noted that this process has not yet been observed experimentally at the LHC. The current signal strength of the tH has observed from the multilepton final state study by the CMS experiment to be 5.7±2.7(stat)±3.0(syst) [38] while the SM predicted value of 0.0724 pb.
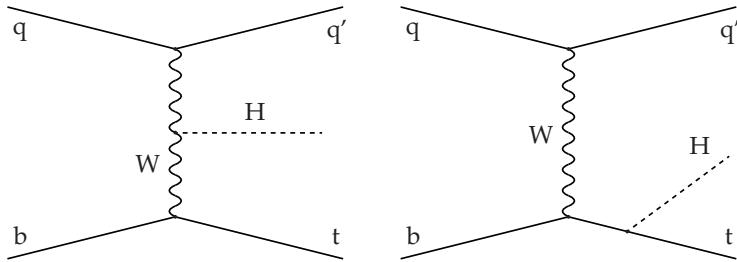
Figure 3.5: Representative tree-level Feynman diagrams of the tH process, first diagram contains a C_V coupling and the second diagram has the y_t coupling.

The cross sections of these above mentioned production modes of the SM H for a definite value of M_H = 125 GeV, as a function of the $\sqrt{s}$, at the LHC is shown in Fig. 3.6.

3.2.2 Decay of Higgs boson

After the production of H, it subsequently decays within a short lifetime of $O(10^{-22})$ s [39]. Driven mainly by the value of M_H and the coupling constants of H to vector bosons and fermions, various decay final states are possible. Interestingly, the measured mass of about M_H allows, fortunately, a large variety of possible decay channels, most of which can be detected experimentally.
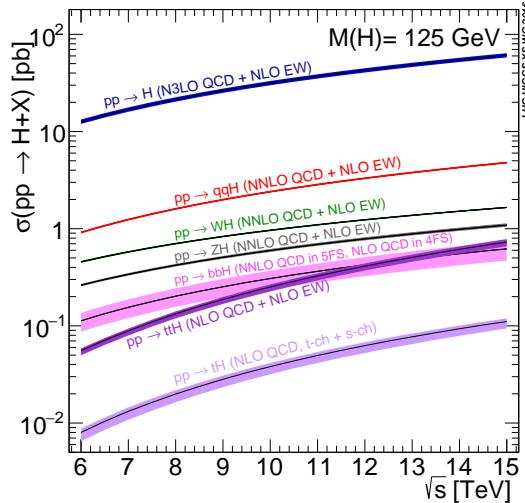
Figure 3.6: Production cross section of a SM Higgs boson of mass $M_H = 125$ GeV as a function of center-of-mass energy $\sqrt{s}$. The lines with different colors correspond to the different production modes with a certain order of accuracy, while the band across each lines give the uncertainty of the calculation.

Of course, at a hadron collider like the LHC, the experimental challenges for measurements are more for the hadronic final states. The general purpose detectors at the LHC are designed to be maximally sensitive to non-hadronic final states involving photons, electrons and muons, such that the discovery milestone could be reached even with limited data.

The branching ratio ($\mathcal{B}r$) to a particular decay final state ($H \rightarrow xx$) is defined as the ratio of decay width of H to that particular decay mode to the total decay width.

$$\mathcal{B}r(xx) = \frac{\Gamma_{xx}}{\Gamma_{total}} \tag{3.1}$$

Fig. 3.7 shows the $\mathcal{B}r$ of H in different final states for $M_H = 125$ GeV. The dependence of $\mathcal{B}r$ on the $\sqrt{s}$ is negligible. The importance to study the H in different final states has been described in the later sections of this thesis.

3.2.3   Discovery of Higgs boson at the LHC

The Higgs boson was discovered by the ATLAS [40] and the CMS [2] Collaborations at the LHC independently and at the mass of $M_H \sim 125$ GeV in 2012. Humongous efforts from a large community consisting of accelerator engineers, theoretical and experimental physicists matched by computing experts made it possible; this discovery has been truly termed as a *big leap for human kind*. This hallmark result established the last part of the SM particle spectrum
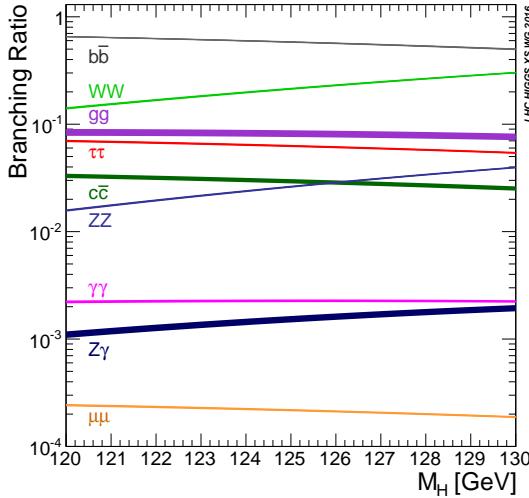
Figure 3.7: Higgs boson branching fractions to various possible decay modes as a function of mass of it .

which was missing for several decades and resolved the mystery about the mass generation of the weak gauge bosons and the fermions. The ATLAS Collaboration reported the existence of a neutral scalar boson with a mass measured at 126.0 ± 0.4 (stat) ± 0.4(syst) GeV with a signal significance of 5.9 standard deviations corresponding to a background fluctuation probability of 1.7×10^{-9} [40]. The analysis was based on the accumulated data of 4.8 fb^{-1} collected at $\sqrt{s} = 7$ TeV in 2011 and 5.8 fb^{-1} at $\sqrt{s} = 8$ TeV in 2012. In parallel, the CMS Collaboration also established an excess of events corresponding to a neutral resonance production at the mass of 125.3 ± 0.4 (stat) ± 0.5(syts) GeV with a signal significance of 5.9 standard deviations based on analysis of 5.1 fb^{-1} and 5.7 fb^{-1} of data collected at $\sqrt{s} = 7$ and 8 TeV [2]. For both the experiments the analysis sensitivities were mostly driven by the di-photon (H $\rightarrow \gamma\gamma$) and four-leptons (H $\rightarrow$ ZZ* $\rightarrow$ 4ℓ) final states with excellent, high resolution ($m/\Delta m \sim 1$ - 2 %) measurements. The discovery plots from CMS Collaboration are presented in Fig. 3.8. As the newly discovered particle decays to a pair of photons, it ensures that its intrinsic spin cannot be one unit and it belongs to the bosonic family.

As mentioned already, the intrinsic mass of H is a free parameter and for the measured value of the M_H, SM can predict almost all the important properties of H. At the same time, for scenarios beyond the SM, the Higgs field structure is extended and thereby existence of multiple physical Higgs bosons are predicted. In some models, eg. minimal supersymmetric extension of SM (MSSM), the lightest member of the Higgs boson family resemble the SM particle, albeit with slight differences in some of the properties, like couplings to various particles. Hence,
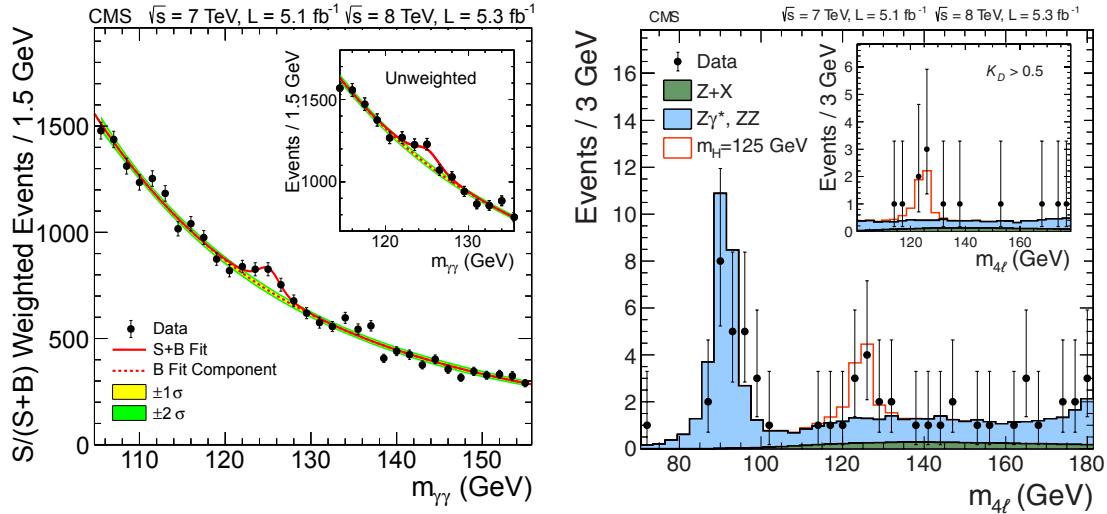
Figure 3.8: Left: The diphoton invariant mass distribution weighted by the S/(S + B) value of its category, the peak around 125 GeV from the SM H contribution on top of the continuum diphoton background is shown with red solid line. Right: Distribution of the four-lepton invariant mass for the ZZ → 4ℓ analysis, H with a mass of M_H 125 GeV has been shown in red slid line sitting on the background. The figure is adapted from Ref. [2]

even after observing the existence of H at $M_H \sim 125$ GeV, the dilemma was whether the discovered resonance belongs to the SM or not. To resolve this, all the properties of H must be measured thoroughly and compared with the attributes in the SM. Both the ATLAS and the CMS experiments have been studying painstakingly various properties of H utilizing both the Run 1 and Run 2 data which is described in Sec. 3.4.

The wisdom gained during last one decade is extremely rich and unexpected to a good extent. The exemplary works of the collider physics community has made many interesting measurements possible. All the measurements are compatible with the predictions of SM. However the current level of uncertainties still allows the particle to belong to certain physics scenarios beyond SM, although the specific nature of them cannot be judged.

3.3 Framework for interpretation of data

3.3.1 Estimating the signal strength μ

This is the first and the most simplified theoretical framework developed to interpret the LHC data pertaining to the Higgs boson. It has been used extensively during the initial years after

the discovery. For a particular mass hypothesis of H, the expected number of signal events in a particular decay mode can be written as:

$$s(M_{\mathrm{H}}) = \sigma_{SM}(M_{\mathrm{H}}).\mathcal{B}r.\mathcal{L}.\epsilon.A \tag{3.2}$$

Here $\sigma_{SM}(M_{\mathrm{H}})$ is the production cross section, $\mathcal{B}r$ is the branching ratio of the particular H decay mode, $\mathcal{L}$ is the integrated luminosity of the data being used and ϵ and A are the efficiency and the geometrical acceptance (detector response) of the experiment. In a total of n number of observed events with b number of background events predicted from SM, $\sigma_{SM}(M_{\mathrm{H}})$ can be fitted like:

$$n = \mu s(M_{\mathrm{H}}) + b \tag{3.3}$$

where μ is called signal strength which is defined as the ratio of the observed value of $\sigma.\mathcal{B}r$ to its expectation predicted from the SM.

$$\mu = \frac{(\sigma.\mathcal{B}r)_{obs}}{(\sigma.\mathcal{B}r)_{SM}} \tag{3.4}$$

Obviously, a measured value of $\mu = 1$ corresponds to the SM prediction, while a deviation indicates the effects of the beyond the SM physics (BSM). Albeit, there are uncertainties in the measurements which have both the statistical and the systematic components. A lot of effort goes into improving the precision.

3.3.2   kappa-framework (κ) for measurement of couplings

In the κ-framework [24], the couplings of H to the massive SM particles (neutrinos are ignored) are taken into account both at the production as well as at the decay vertices. The production cross section, the total width and the decay branching ratio to a particular mode are scaled separately by the relevant scale factors called κ. Considering a process $ii \to \mathrm{H} \to ff$, where H is produced with a cross section of σ_{ii} followed by decay $\mathrm{H} \to ff$ with total decay width of Γ_H and partial width of Γ_{ff}, under narrow-width approximation one can write

$$
\begin{aligned}
(\sigma \cdot BR)(ii \to \mathrm{H} \to ff) &= \sigma(ii \to H).BR(\mathrm{H} \to ff) \\
&= \sigma(ii \to H) \cdot \frac{\Gamma_{ff}}{\Gamma_H} \\
&= \sigma^{SM}(ii \to H) \cdot \frac{\Gamma_{ff}^{SM}}{\Gamma_H^{SM}} \cdot \frac{\kappa_i^2 \cdot \kappa_f^2}{\kappa_H^2}
\end{aligned}
\tag{3.5}
$$

The κs are called the coupling modifiers. κ_i appears due to the couplings at the production side, while κ_f is the coupling modifier for the coupling between the H and its decay products and,

finally, κ_H is the coupling modifier for the total decay width, since some of the couplings are yet to be established.

It is to be noted that experimentally we are only estimating the deviations of the couplings wrt SM via κ measurements; we are not directly measuring the individual couplings. In various measurements there is also an implicit assumptions that the couplings do not "run" or vary across different datasets collected at different energies.

In some H production modes and also in the decay modes via loop contributions (H $\rightarrow \gamma\gamma$), where more than one H couplings contribute either constructively or destructively, the resultant κ comes with non-trivial relations taking into account the weight factors. These couplings are: $\kappa_V(\kappa_W, \kappa_Z)$, $\kappa_g(\kappa_b, \kappa_t)$, $\kappa_\gamma(\kappa_b, \kappa_t, \kappa_\tau, \kappa_W)$ and κ_H where the couplings to lighter particles are ignored due to their smallness. The examples of some of them are described below in details:

1. **H $\rightarrow$ gg** κ_g^2 is the scaling term which appears in the loop-induced process appearing in the ggH production mode since H does not directly couple to the gluons. In the loop contribution, the two heavy quarks (t, b) mostly contributes due to their larger Yukawa couplings (y_f). So the κ_g^2 depends on κ_t and κ_b and it is expressed as:

$$\kappa_g^2(\kappa_t, \kappa_b) = \frac{\kappa_t^2 \cdot \sigma_{ggH}^{tt} + \kappa_b^2 \cdot \sigma_{ggH}^{bb} + \kappa_t\kappa_b \cdot \sigma_{ggH}^{tb}}{\sigma_{ggH}^{tt} + \sigma_{ggH}^{bb} + \sigma_{ggH}^{tb}} \tag{3.6}$$

Here σ_{ggH}^{tt}, σ_{ggH}^{bb} and σ_{ggH}^{tb} are the ggF process cross section due to only top and bottom quark contributions and from top-bottom interference respectively.

2. **κ in VBFH production:** H in VBF production can be produced through either W or Z pair fusion with cross sections σ_{WF} and σ_{ZF} respectively. Hence κ_{VBF} can be formulated as:

$$\kappa_{VBF}^2(\kappa_W, \kappa_Z) = \frac{\kappa_W^2 \cdot \sigma_{WF} + \kappa_Z^2 \cdot \sigma_{ZF}}{\sigma_{WF} + \sigma_{ZF}} \tag{3.7}$$

It is to be noted that the above relations are valid for leading order feynman diagrams. A summary of the scale factors for the production cross sections and the partial decay widths are presented in Tab. 3.1.

3.4 Important milestones for Higgs boson characterization

After the first observation of the Higgs boson at around 125 GeV both the ATLAS and CMS Collaborations initiated extensively, in consultation with theoretical experts, multiple studies

Table 3.1: Leading Order coupling scale factor (κ) relations from both Higgs boson cross sections and its partial decay widths relative to the SM prediction. The table is taken from Ref. [24].

<div style="display:flex; justify-content:space-between;">

Production modes

$$\frac{\sigma_{ggF}}{\sigma_{ggF}^{\text{SM}}} = \begin{cases} \kappa_g^2(\kappa_b, \kappa_t) \\ \kappa_g^2 \end{cases}$$

$$\frac{\sigma_{\text{VBF}}}{\sigma_{\text{VBF}}^{\text{SM}}} = \kappa_{\text{VBF}}^2(\kappa_W, \kappa_Z)$$

$$\frac{\sigma_{\text{WH}}}{\sigma_{\text{WH}}^{\text{SM}}} = \kappa_W^2$$

$$\frac{\sigma_{\text{ZH}}}{\sigma_{\text{ZH}}^{\text{SM}}} = \kappa_Z^2$$

$$\frac{\sigma_{t\bar{t}H}}{\sigma_{t\bar{t}H}^{\text{SM}}} = \kappa_t^2$$

Detectable decay modes

$$\frac{\Gamma_{WW^{(*)}}}{\Gamma_{WW^{(*)}}^{\text{SM}}} = \kappa_W^2$$

$$\frac{\Gamma_{ZZ^{(*)}}}{\Gamma_{ZZ^{(*)}}^{\text{SM}}} = \kappa_Z^2$$

$$\frac{\Gamma_{b\bar{b}}}{\Gamma_{b\bar{b}}^{\text{SM}}} = \kappa_b^2$$

$$\frac{\Gamma_{\tau^-\tau^+}}{\Gamma_{\tau^-\tau^+}^{\text{SM}}} = \kappa_\tau^2$$

$$\frac{\Gamma_{\gamma\gamma}}{\Gamma_{\gamma\gamma}^{\text{SM}}} = \begin{cases} \kappa_\gamma^2(\kappa_b, \kappa_t, \kappa_\tau, \kappa_W) \\ \kappa_\gamma^2 \end{cases}$$

$$\frac{\Gamma_{t\bar{t}}}{\Gamma_{t\bar{t}}^{\text{SM}}} = \kappa_t^2$$

Currently undetectable decay modes

$$\frac{\Gamma_{gg}}{\Gamma_{gg}^{\text{SM}}}$$

$$\frac{\Gamma_{c\bar{c}}}{\Gamma_{c\bar{c}}^{\text{SM}}} = \kappa_t^2$$

$$\frac{\Gamma_{s\bar{s}}}{\Gamma_{s\bar{s}}^{\text{SM}}} = \kappa_b^2$$

$$\frac{\Gamma_{\mu^-\mu^+}}{\Gamma_{\mu^-\mu^+}^{\text{SM}}} = \kappa_\tau^2$$

$$\frac{\Gamma_{Z\gamma}}{\Gamma_{Z\gamma}^{\text{SM}}} = \begin{cases} \kappa_{(Z\gamma)}^2(\kappa_b, \kappa_t, \kappa_\tau, \kappa_W) \\ \kappa_{(Z\gamma)}^2 \end{cases}$$

Total width

$$\frac{\Gamma_H}{\Gamma_H^{\text{SM}}} = \begin{cases} \kappa_H^2(\kappa_i, M_H) \\ \kappa_H^2 \end{cases}$$

</div>

to establish the main features of the H by measuring the individual couplings with various SM particles. From the Run-1 data of LHC, mainly the interaction of H to the gauge bosons were established via decays to pairs of photons [41], Z [42] and W [43].

The fact that H decays to two spin 1 particles (say H $\rightarrow$ $\gamma\gamma$), which indicates that it cannot be a spin 1 boson. There were two possibilities: spin 0 or 2. For an odd parity particle, the kinematics of the final state particles would be different from that of even parity cases. By construction H is a boson class of models considered to be scalar ie. of spin 0. Experimental measurements at the LHC rules out various possibilities for the value of spin and parity. A dedicated measurement to determine the spin and parity of H has been performed by CMS in the leptonic decay of H $\rightarrow$ W$^+$W$^-$ channel [43]. The data is found to favor the SM hypothesis of $J^P = 0^+$, J: spin and P: parity of the H, by excluding the other alternative hypothesis at 99.8% CL.

The natural width of the H boson (Γ_H) is beyond the reach of the direct measurement, but CMS has performed a width measurement using Run 1 data [44] in H $\rightarrow$ ZZ channel from the relative on-shell and off-shell production decay rates, where both the two Z bosons are decaying leptonically. At that time, the upper limit on Γ_H is 22 MeV at a 95% confidence level. Very recently CMS has published Run 2 result [39] in the same final state, the width of the H boson appears at $\Gamma_H = 3.2^{+2.4}_{-1.7}$ MeV, which agrees with the SM predictions (4.1 MeV) within the uncertainties.

After a few years, CMS established the H decay to a pair of τ leptons with a signal significance of 5.9 standard deviations from the Run 1 and the first year of Run 2 (35.9 fb^{-1}) data [45]. This is the first observation of Higgs coupling to a fermion (belonging to the 3rd generation lepton). Subsequently CMS discovered the H coupling to a top-antitop pair by studying t$\bar{\text{t}}$H production process [46], where H decays into pairs of W bosons, Z bosons, photons, τ leptons, and b-jets with a combined significance of 5.2 standard deviations over the background-only hypothesis. This established the first coupling of Higgs bosons with the quark sector.

After accumulating one more year of Run 2 data collected in 2017 amounting to $\mathcal{L} \sim 41.5$ fb^{-1} and combining with the datasets of 2016 and Run 1, CMS observed the H coupling to the bb pair, for the first ever at the LHC, and yielding an observed significance of 5.6 standard deviations [47]. The sensitivity of this analysis was significantly driven by the VH production mode with V decays to either leptons or neutrinos; ie, W $\rightarrow$ $\ell\nu$ and Z $\rightarrow$ $\ell\ell$, $\nu\nu$.

Very recently CMS has extracted the first evidence of H coupling to the 2nd generation of leptons with a significance of 3 standard deviations from background only hypothesis by measuring $H \rightarrow \mu\mu$ process by using the total Run 2 data of 137 fb^{-1}, collected during 2016-2018 [48].

All the above measurements are used to extract the coupling strengths within the κ framework to check the consistency with the SM predictions. Any deviation of a certain κ from the value of 1 may indicate the possible existence of the BSM physics. Figure 3.9 (left) shows the current status of the measurements of H coupling modifiers with the fermions and bosons (κ_f and $\sqrt{\kappa_V}$ scaled with $\frac{m_f}{v}$ and $\frac{m_V}{v}$) respectively [3]. The measured values match with the SM predictions within the uncertainties, indicating that H behaves as a SM particle. The actual values are shown in Fig. 3.9 (right).
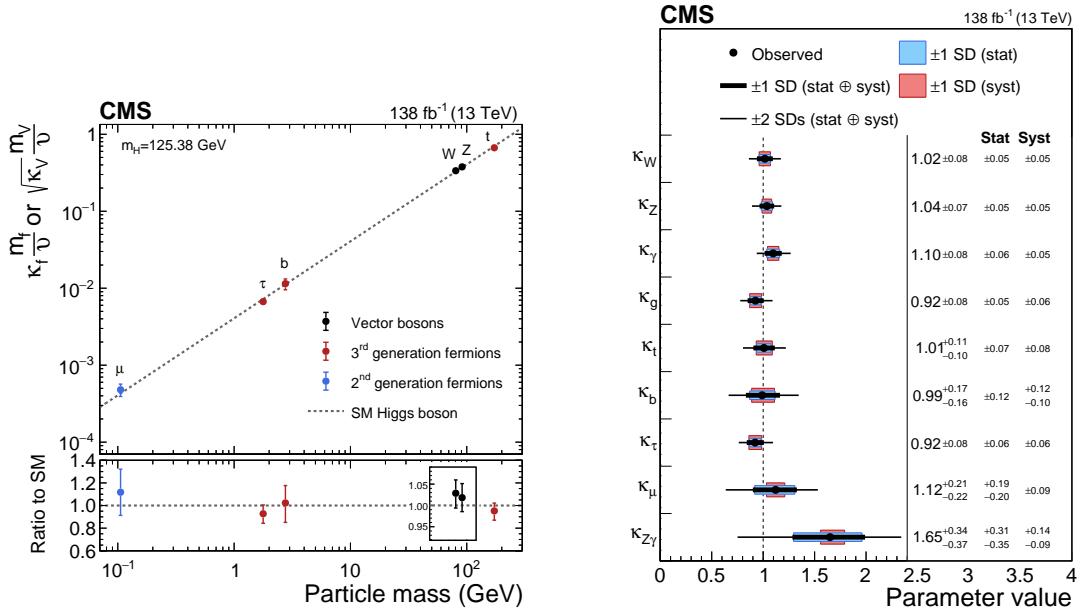
Figure 3.9: Left: The diphoton invariant mass distribution weighted by the S/(S + B) value of its category, the peak around 125 GeV from the SM H contribution on top of the continuum diphoton background is shown with red solid line. Right: Distribution of the four-lepton invariant mass for the $ZZ \rightarrow 4\ell$ analysis, H with a mass of M_H 125 GeV has been shown in red slid line sitting on the background. Figure is adapted from Ref. [3].

3.5 Higgs self-coupling and HH production

The shape of the Higgs potential is directly dependent on the Higgs self-coupling according to the expression

$$V(h) = \frac{m_h^2}{2} h^2 + \lambda_{HHH} v h^3 + \frac{1}{4} \lambda_{HHHH} h^4 \tag{3.8}$$

Here λ_{HHH} and λ_{HHHH} are the trilinear and the quartic coupling strengths and they can be defined in terms of m_H and v. In SM

$$\lambda_{HHH} = \lambda_{HHHH} = \lambda = \frac{m_H^2}{2v^2}$$

However the experimental measurement of λ is essential to validate the above assumption about the Higgs potential in SM. Importantly, the shape of the Higgs potential determines the evolution of the early universe at the epoch after the EWSB, whether there are multiple valleys corresponding to metastable state or stable state.

The expected value of λ from SM is 0.13 for $M_H = 125$ GeV and v = 246 GeV. Thus the experimental measurement of the λ is a very crucial aspect of H characterization and as of today it remains one of the major Higgs properties which is yet to be established. The coupling λ can be directly accessed at the LHC from the non-resonant productions of Higgs pair (HH) and triple Higgs(HHH). Due to very low cross section (~ 80 ab at $\sqrt{s} = 13$ TeV) the search for HHH production is currently beyond the reach of the LHC, although the cross section of the HH is small also, but using the accumulated data so far, it can be probed to constrain the allowed range of the coupling modifier κ_λ. It is assumed that an anomalous value of κ_λ, typically results in higher cross section leading to larger event statistics compared to that expected from SM, which is suitable for analyses with limited data volume.

At the LHC the main production mode of the diHiggs boson is through gluon gluon fusion which produces almost 95% of the HH events. Similar to single-H production there are other subdominant modes of HH production which can also be probed at the LHC by utilizing special properties of the concerned processes.

- **Gluon Gluon Fusion (ggHH)**: The most dominating production mode of the HH at the LHC is via gluon pair fusion with a cross section of about 31.05 fb at N^2LO accuracy in QCD [49–51] at a centre-of-mass energy of 13 TeV; it is about 1000 times smaller than the single-H production. Figure 3.10 shows the leading order diagrams of the ggHH process, where both of them is produced via a heavy quark loop and contain the t-quark
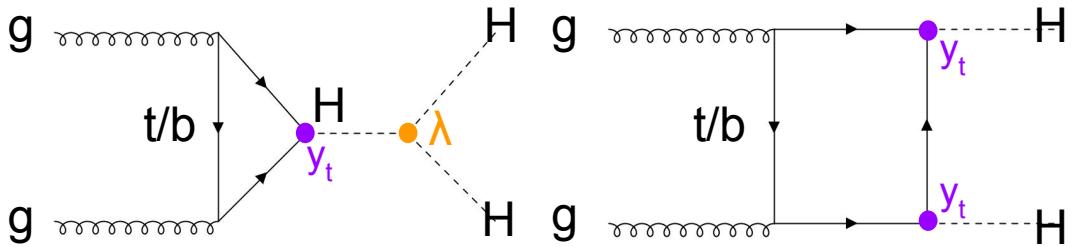
Figure 3.10: LO Feynman diagrams of the gluon-gluon fusion process of Higgs pair (ggHH) production. The left one, called the triangle diagram, contains both y_t and λ and the right one refers to the box diagram involving only y_t.
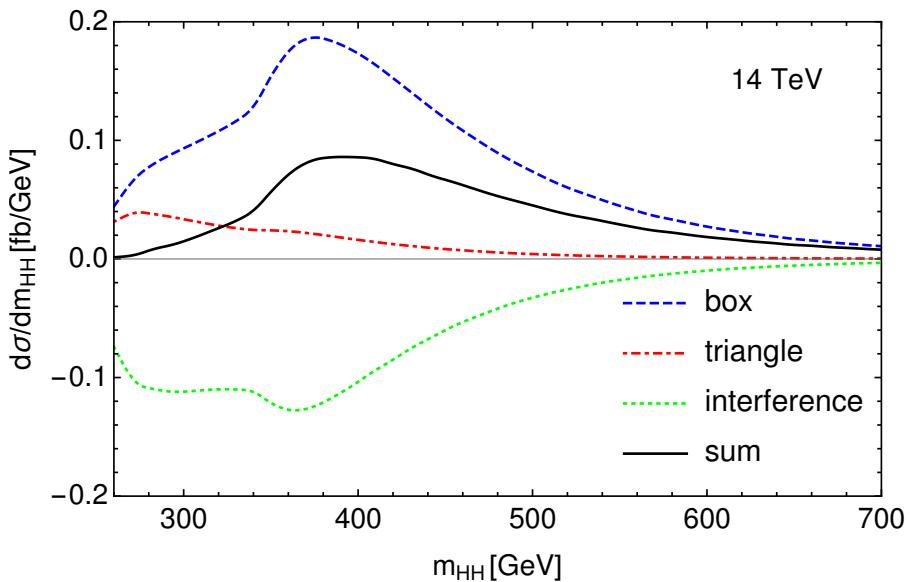
Figure 3.11: The HH invariant mass distribution from the individual ggHH processes (box & triangular diagrams) and their interference terms at LO. The figure is taken from Ref. [4].

Yukawa coupling. The first diagram called triangle diagram contains λ, while the second box diagram doesn't. The cross section and kinematics of the ggHH process depends on the λ and y_t. Two diagrams have opposite signs leading to destructive interference, and hence the total cross section is very small. The contribution of the individual diagrams and the interference term is shown in Fig. 3.11 as a function of HH invariant mass.

- **Vector Boson Fusion (VBFHH)** : HH production via VBF process is the subleading mode, where a soft emission of two massive vector bosons (V = W,Z) from the colliding partons (quarks) is followed by their fusion leading to the hard scattering VV → HH. At $\sqrt{s} = 13$ TeV the cross section of VBFHH process is 1.73 fb at N^{3}LO QCD accuracy [52, 53]. The representative feynman diagrams for the VBFHH process at LO are shown in
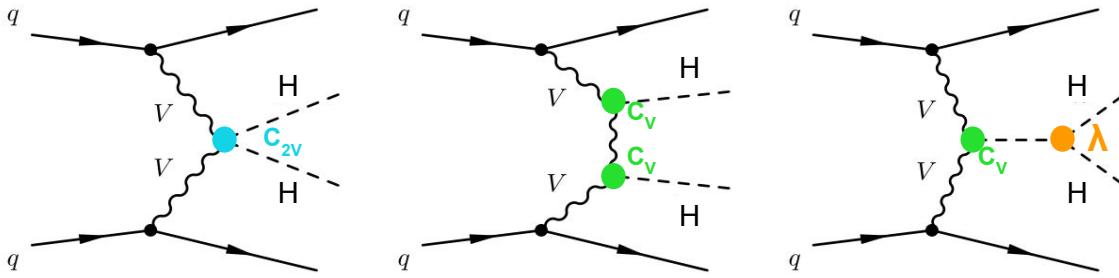
Figure 3.12: Feynman diagrams that contribute to the production of Higgs boson pairs via VBF at LO. On the left the diagram involving the C_{2V} coupling, in the middle the diagram with two C_V couplings, and on the right the diagram with the tri-linear Higgs self-coupling λ.

Fig .3.12. The most interesting fact about VBFHH process is the unique and direct access of the coupling of a pair of Higgs bosons with a pair of vector bosons, $C_{2V} \sim HHVV$, whereas the other two diagrams contains the self-coupling λ and the coupling of a single H with a pair of vector bosons $C_V = HVV$. Despite having a very small cross section, a very small change in the couplings can induce a striking increase of the cross section as a function of the M_{HH}; this enhanced sensitivity potentially adds extra impact on the measurement and the constraints on λ and C_{2V}.

- **Higgs boson pair in association with a vector boson (VHH)**: The production of HH in association with a vector boson contributes to the stduy of inclusive Higgs pair. There are three different Higgs boson couplings involved which are the same as in VBFHH production process: single Higgs boson coupling with a pair of vector bosons C_V, self-coupling λ, and the quartic coupling to vector bosons C_{2V}. The leading order Feynman diagrams for VHH process are shown in Fig. 3.13. In the SM, the cross sections of VHH processes at the LHC are much smaller compared to the ggF and VBF production modes of HH, 0.50 fb for WHH (W$^+$HH: 0.329 fb, and W$^-$HH: 0.173fb) and 0.36 fb for ZHH at $\sqrt{s}$= 13 TeV [54, 55], computed at N^{2}LO accuracy in QCD.

- **Higgs boson pair in association with top pair (t$\bar{t}$HH)**: Similar to t$\bar{t}$H production, t$\bar{t}$HH process has also a very small cross section, and the corresponding Feynman diagrams are shown in Fig 3.14. At the present centre-of-mass energy regime of LHC, t$\bar{t}$HH does not play a very important role but in high energy future colliders it has two-fold relevance: (i) the interference of the contributing terms is constructive unlike the ggHH process and (ii) as t$\bar{t}$HH prodcuction requires a high energy transfer process, at the higher centre of
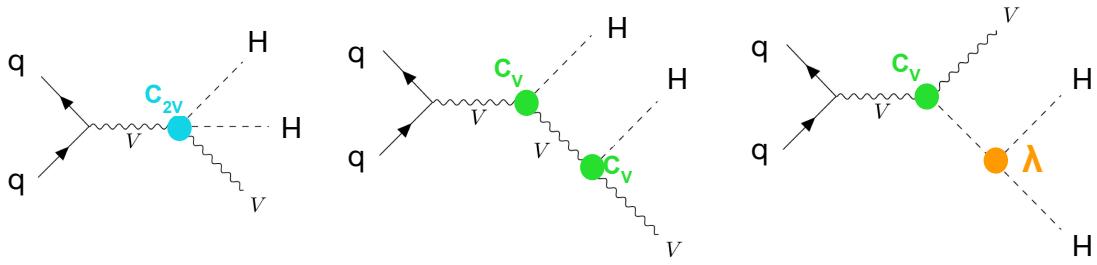
Figure 3.13: Feynman diagrams that contribute to the production of Higgs boson pair in association with a vector boson (VHH) at LO. On the left the diagram involves the C_{2V} coupling, in the middle the diagram has two C_V couplings, and the right diagram involves the tri-linear Higgs self-coupling λ.

mass energy the cross section difference between ggHH and $t\bar{t}$HH is less (the values of $\sigma_{ggHH}/\sigma_{t\bar{t}HH}$ at $\sqrt{s}$ =13 TeV is $\sim$ 40 whereas 14 at $\sqrt{s}$ = 100 TeV collider [56, 57]).
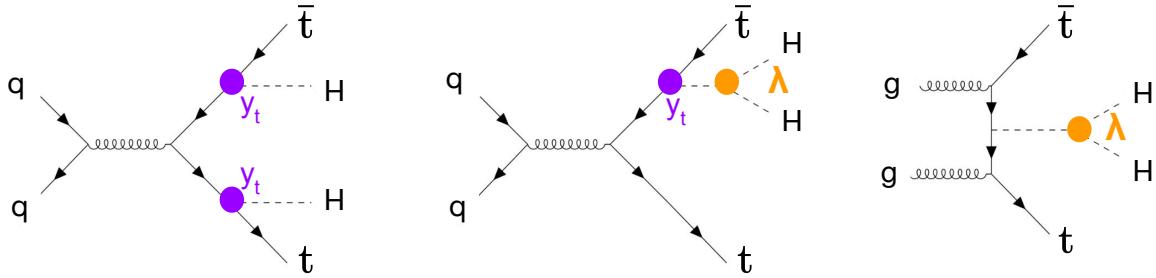
Figure 3.14: Feynman diagrams that contribute to the production of Higgs boson pairs in association with a pair of top quarks at LO. The left two diagrams contain the tri-linear Higgs self coupling (λ), all the diagrams contain the top-quark Yukawa coupling y_t.

The production cross sections of HH process via different production modes mentioned above are presented in Fig. 3.15 as a function of centre-of-mass energy and also summarized in Tab. 3.2.

3.5.1   Current status of HH searches at the LHC

At the LHC both CMS and ATLAS Collaborations have performed the search of HH production in both Run 1 and Run 2 data [58–62,62–68]. Due to it's very small rate of production, the number of expected HH events is extremely low. A rough estimation of the expected events produced in the LHC is given in Tab. 3.3 without considering the detector acceptance and the
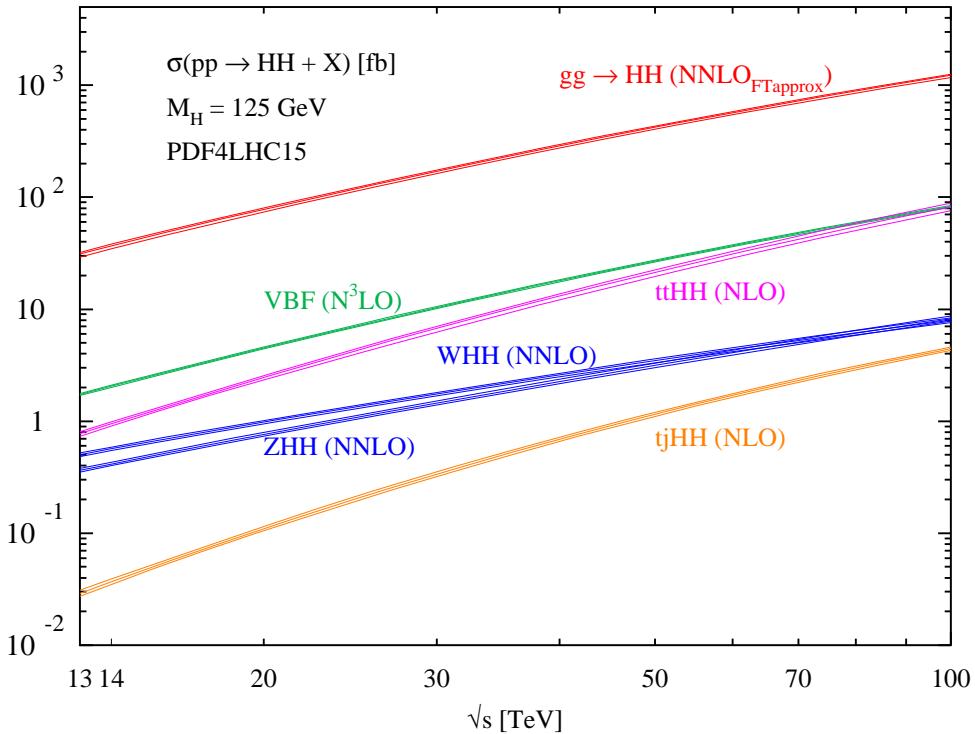
Figure 3.15: Variation of cross section for Higgs boson pair production as a function of centre-of-mass energy via the major modes ggHH, VBFHH, VH (WHH & ZHH) and t$\bar{\text{t}}$HH in different colors. The bands correspond to the associated uncertainties of the cross section computation at certain order. The figure is taken from Ref. [4].

Table 3.2: Summary of different modes of inclusive HH production cross sections (in fb) as a function of centre-of-mass energy. The associated uncertainties (in %) to the cross sections due to QCD scale & PDF+α_S are mentioned along with the values.

Prod. mode	13 TeV	14 TeV	27 TeV	100 TeV
ggHH	$31.05^{+6\%}_{-23\%} \pm 3.0\%$	$36.69^{+6\%}_{-23\%} \pm 3.0\%$	$139.9^{+5\%}_{-22\%} \pm 2.5\%$	$1224^{+4\%}_{-21\%} \pm 2.4\%$
VBFHH	$1.726^{+0.03\%}_{-0.04\%} \pm 2.1\%$	$2.005^{+0.03\%}_{-0.04\%} \pm 2.1\%$	$8.404^{+0.11\%}_{-0.04\%} \pm 2.0\%$	$82.84^{+0.13\%}_{-0.04\%} \pm 2.1\%$
ZHH	$0.363^{+3.4\%}_{-2.7\%} \pm 1.9\%$	$0.415^{+3.5\%}_{-2.7\%} \pm 1.8\%$	$1.23^{+4.1\%}_{-3.3\%} \pm 1.5\%$	$8.23^{+5.9\%}_{-4.6\%} \pm 1.7\%$
W$^+$HH	$0.329^{+0.32\%}_{-0.41\%} \pm 2.2\%$	$0.369^{+0.33\%}_{-0.39\%} \pm 2.1\%$	$0.941^{+0.52\%}_{-0.53\%} \pm 2.1\%$	$4.70^{+0.90\%}_{-0.96\%} \pm 1.8\%$
W$^-$HH	$0.173^{+1.2\%}_{-1.3\%} \pm 2.8\%$	$0.198^{+1.2\%}_{-1.3\%} \pm 2.7\%$	$0.568^{+1.9\%}_{-2.0\%} \pm 2.1\%$	$3.30^{+3.5\%}_{-4.3\%} \pm 1.9\%$
t$\bar{\text{t}}$HH	$0.775^{+1.5\%}_{-4.3\%} \pm 3.2\%$	$0.949^{+1.7\%}_{-4.5\%} \pm 3.1\%$	$5.24^{+2.9\%}_{-6.4\%} \pm 2.5\%$	$82.1^{+7.9\%}_{-7.4\%} \pm 1.6\%$

event reconstruction efficiency. It is evident that the observation of diHiggs events is only possible at the end of HL-LHC with a large enough integrated luminosity of 3000 ab^{-1}. But any existence of new physics or deviation of the relevant couplings of the Higgs boson from the SM expectation can enhance the cross section with possible modifications in event kinematics; this will also modify the event acceptance as well. Hence the current thrust area of the HH searches is to exclude the anomalous value of the HH coupling parameters with as stringent allowed range as possible and to determine the upper limit on the production cross sections. The

Table 3.3: Rough estimation of expected number of HH event production at the different Run of the LHC. Combination of ATLAS and CMS will double the number of events.

LHC era	$\sqrt{s}$ (TeV)	Int. luminosity (fb^{-1})	Expected events
Run 1	7 - 8	~30	~300
Run 2	13	~150	~5000
Run 3	13.6	~300 (projected)	~12000
HL-LHC	14	~3000 (projected)	~120000

number of events finally used for any analysis is typically much lower compared the number actually produced in the experiment, due to various factors. As the production cross section of HH is very low, so it is important to reconstruct such events as precisely as possible from its decay products. The decay branching fraction of the total HH system directly depends on the individual H decay rates since the two decays are independent of each other. Figure 3.16 represents a 2-dimensional representation of the HH decay combinations in different final states. In Run 1 and early Run 2 (only with 2016 data) all searches were confined in the final states along the first row of Fig. 3.16, where at least one of the H candidate decays to a pair of b-jets due to it's maximum branching fraction. COnsideration of different types of final states typically lead to different proportions of the background contaminations. Hence the choice of the final state in HH search is an interplay between the net HH branching fraction and the signal-to-background ratio. The four major HH search channels along with their key features are described below. .

- **HH → bb̄bb̄**: For both of the H candidates decaying through bb̄ mode has the maximum value of the net branching fraction of about 33.7%. Larger branching fraction yields large statistics event sample, but the overwhelming background from QCD multijet production with cross section many orders higher makes the experimental identification of signal highly challenging. Also it reduces the signal purity in the selected event sam-
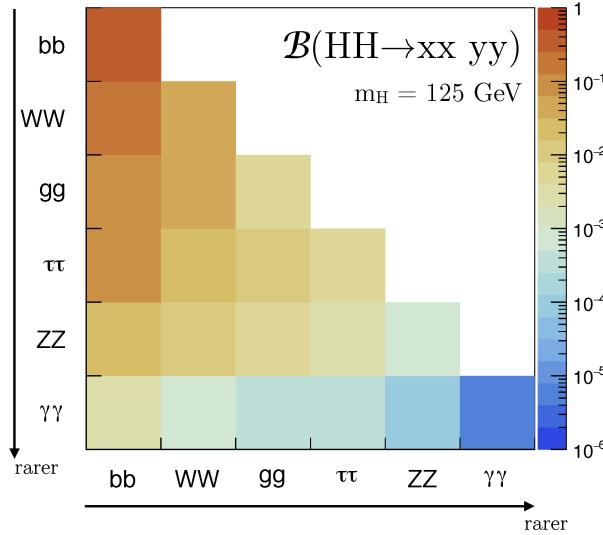
Figure 3.16: Branching fraction of HH decay into different final states taking H decay to SM particles at M_H =125 GeV. The figure is taken from Ref. [5].

ple. However considering a boosted HH system the signal discrimination against the background can be improved. The CMS Collaboration has recently utilised this strategy to search for a boosted VBFHH system and established that the Run 2 data excludes $C_{2V} = 0$ value at 95% CL [69].

- **HH → $\tau\tau$ b$\bar{\text{b}}$**: A branching fraction of 7.3% and comparatively lower background contamination makes this channel feasible to search for HH production with reasonable sensitivity. The only problem here is to reconstruct the kinematics of the HH candidate fully due to the neutrinos from τ decays which lead to a net missing transverse energy in the event.

- **HH → VV b$\bar{\text{b}}$**: Here both Z and W are considered with the final state leading to a substantial branching fraction for this decay mode. Main drawback of this analysis is the irreducible t$\bar{\text{t}}$ background.

- **HH → $\gamma\gamma$ b$\bar{\text{b}}$**: Despite having very small net branching fraction only about 0.26%, it is considered to be a golden channel in the context of HH search due to the excellent resolution of the invariant mass of the two photons coming from H → $\gamma\gamma$ candidate due to the precise measurement of the kinematics in the electromagnetic caloriemeter while the other H → b$\bar{\text{b}}$ candidate helps to acquire reasonable event statistics to peruse the analysis.

After the end of Run 2 with a large amount of dataset, the final states appearing across the second row of Fig. 3.16, where one of the H decays to a pair of W are targeted. HH searches in multilepton final states, ie, utilizing the decay modes WWWW, WWττ and ττττ) [70] and WWγγ [71] adds to the combined HH signal significance. A summary of the results of non-resonant HH searches in the major decay channels using full Run 2 data from both ATLAS [72–75] and CMS [70, 76–79] experiments are presented in Tab: 3.4. The combined results of HH search from the most dominating decay channel analyses in CMS experiment is published in Ref. [3]. The current reach of the HH search in Run 2 data from individual channels and after combining is shown in Fig. 3.17 (left). Figure. 3.17 (right) shows the HH production cross section as functions of the κ_λ, the Higgs boson self-interaction coupling modifier of λ. The allowed range of κ_λ is found to be -1.24 to 6.49 at 95% CL. The recent HH results from ATLAS is shown in Fig. 3.18.

Table 3.4: The expected (exp.) and observed (obs.) upper limit at 95% CL on the production cross section of HH process and branching ratio from different decay channels from ATLAS and CMS experiment in Run 2.

Final states	ATLAS		CMS	
	exp.	obs.	exp.	obs.
$b\bar{b}b\bar{b}$	8.4	5.1	4.0	6.4
$b\bar{b}\,\tau\,\tau$	3.9	4.7	5.2	3.3
$b\bar{b}\,\gamma\,\gamma$	5.7	4.2	5.5	8.4
multilepton	-	-	19	21
$b\bar{b}\,Z\,Z$	-	-	40	32
$b\bar{b}\,W\,W$	40	29	-	-
combined	-	-	2.5	3.4
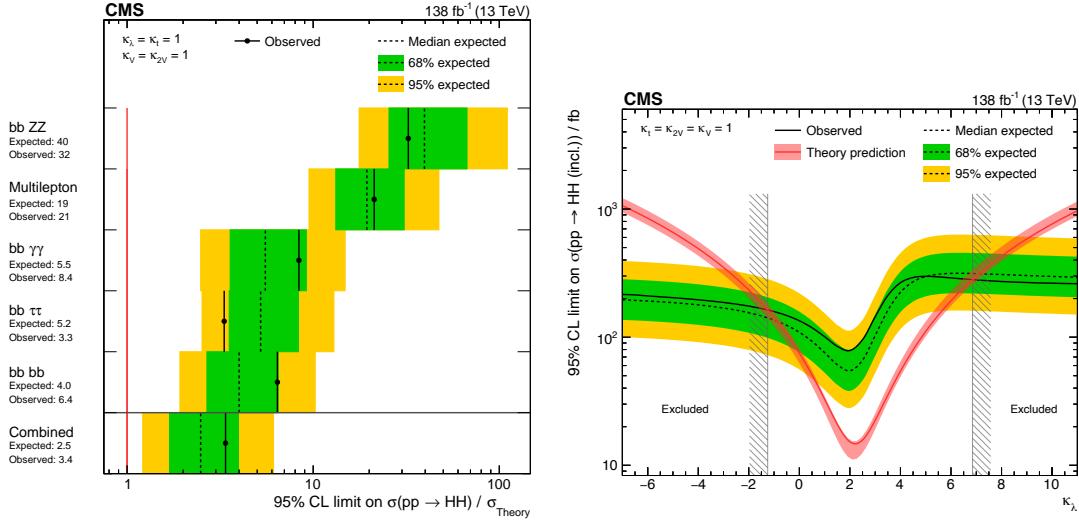
Figure 3.17: Left: 95% CL upper limit on the standard model HH cross section times branching fraction of the different HH final state channels and the combined one. Right: The scan of 95% CL upper limit on the cross section for different values of κ_λ from the combination of searches targeting different HH final states. Ref [3].
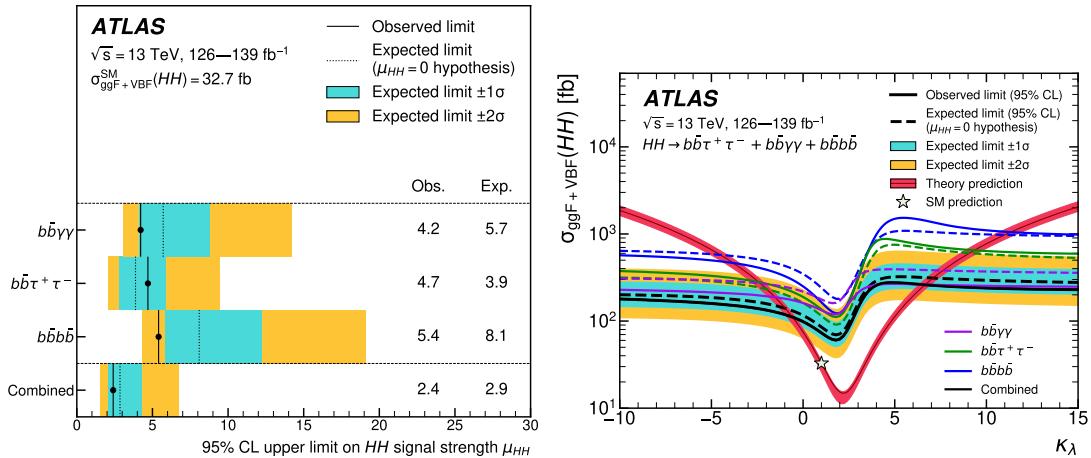
Figure 3.18: The descriptions of the plots are same as in Fig. 3.17. Ref [80]

.

Chapter 4

The Large Hadron Collider

4.1 Particle Colliders

During the last several decades the nature of the fundamental particles, has been explored and investigated in various particle physics experiments, that helped to develop a consistent mathematical structure which we call the standard model (SM) of particle physics. With the advancement of technologies, the latter half of this learning era has been based on mostly experiments at various types of high energies colliding beam (of oppositely moving particles) facilities rather than fixed-target set up where a significant fraction of the beam energy cannot be utilized for achieving high centre-of-mass energy.

Depending on the nature of the colliding particles, the particle accelerators can be classified into mainly two types: (a) collision of leptons, like electron and positron (e^+e^-), as in the case of the large Electron Positron (LEP) collider at CERN, Geneva which operated during 1989-2000, (b) collision of hadrons, like, the proton-antiproton collider, Tevatron at Fermilab, Chicago which was operational during 1983-2011. The currently on-going large hadron collider (LHC) at CERN is capable of colliding protons as well as heavy ions. The LHC machine is discussed in a bit more detail in Sec. 4. In the first type of the collider, the colliding particles are fundamental; hence the collision energy is fixed by the momentum of the initial particles; such facilities can be used for the precision measurements with quite a clean experimental environment. On the other hand, for the second type, the colliding particles are the composite ones, like protons or lead ion etc.. Despite of the fixed beam energy of the colliding hadrons, actually the constituents of these composite particles take part in the basic interactions carrying only a fraction of the

beam energy. This results in a wide range of the probed energy scale at the subprocess level. Hence the hadron machines can be used more efficiently for the discovery or search for the new particles of unknown masses, or effects of new interactions of higher energy scale. Typically at the colliders, the energy of the colliding beams are either the same or different motivated by the physics being addressed. The e^+e^- collider at Japan has asymmetric beam energies. The aspects of the LHC has been described below.

4.2   The LHC

The Large Hadron Collider [81] at CERN (European Organization for Nuclear Research) is the largest and the highest energy particle collider machine built so far. The LHC ring is located inside a tunnel of circumference 26.7 km long, the same tunnel which was used for having the Large Electron Positron Collider (LEP)) and 50-170 meter underground near the France-Switzerland border close to Geneva city, Switzerland. The main motivation of the LHC project is to discover the Higgs boson and resolve the issue of electroweak symmetry breaking as well as to probe physics at TeV energy scale. To study the new particles at high energy regime the centre-of-mass energy of the collision ($\sqrt{s}$) is high enough to make it easy to probe. Not only the centre-of-mass energy, the collision rate of the LHC is also very high which helps to search for rare processes.

The ring has eight straight portions and eight circular sectors with an average internal diameter of 3.7 meter. Inside the accelerator two high energy particle beams collide at four interaction points from opposite directions. As the particle beams travel in the opposite directions, two different tubes are kept inside the same mechanical structure at ultrahigh vacuum with a pressure of 10^{-10} to 10^{-11} mbar. They are guided inside the ring by a strong magnetic field which is maintained by superconducting electromagnets. To achieve this superconductivity a very cold temperature of 1.8 K, even colder than the outer space, is maintained by supplying the liquid helium constantly in the magnetic systems.

The initial plan of the LHC was to collide only the proton-proton (p-p) collision, later heavy ion collisions (i.e ion-ion or ion-proton) were also included in the LHC agenda to produce quark-gluon plasma (QGP) for studying the collective behavior of quarks and gluons (heavy-ion physics). For the LHC heavy ion physics program mostly lead (Pb) ions are used, though Xenon (Xn) has also been used. This thesis work is based on the data from the p-p collision
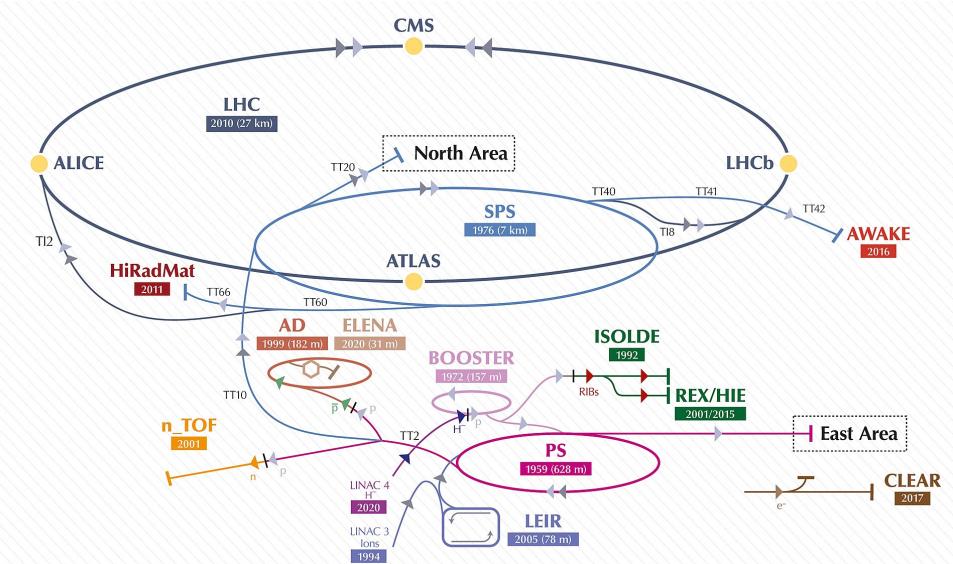
Figure 4.1: A schematic view of LHC accelerator.

during Run 2 of LHC machine (2016-2018). There are four crossing points of the tubes inside the LHC ring, where the particles can collide. Each collision point is surrounded by a detector: A Toroidal LHC Apparatus (ATLAS) [82], A Large Ion Collider Experiment (ALICE) [83], Compact Muon Solenoid (CMS) [84] and Large Hadron Collider beauty (LHCb) [85]. The ATLAS and CMS are two general purpose detectors which discovered the Higgs boson in 2012 independently, ALICE is dedicated to heavy-ion physics and LHCb is to study the physics of beauty quarks (b-physics).

Before collision at the experimental sites, the protons undergo through different stages of the acceleration entering the main LHC ring in the final stages as shown in Fig. 4.1.

Protons of 50 keV are produced first from Hydrogen plasma in Duoplasmon source and passed to radio frequency (RF) quadruples which accelerate them to 750 keV before injecting into the linear accelerator (LINAC 2). LINAC 2 uses the radio-frequency cavities to charge cylindrical conductors, which are charged alternately positive or negative. The conductors behind them push the particles and the conductors ahead of them pull, causing the particles to accelerate. Small quadrupole magnets ensure that the protons remain in a tight beam. At this step protons had reached the energy of 50 MeV. LINAC 2 was turned off on 12 November 2018 after 40 years of constant operation and it was replaced in 2020 during CERNs Long Shutdown 2 by a more powerful LINAC 4 with an acceleration power upto 160 MeV. Proton are then injected to the proton synchrotron booster (PSB). It is the first and the smallest circular proton accelera-

tor step in the LHC proton acceleration chain. It contains four superimposed synchrotron rings with a radius of 25 meters and accelerates the protons upto 1.4 GeV and produces proton bunch before pushing them to the next step of proton synchrotron (PS). The PS is the CERN's oldest synchrotron machine which consists a ring of circumference about 628 meters with 277 conventional (room-temperature) electromagnets, including 100 dipoles to bend the beams round the PS ring. It accelerates proton bunches up to 26 GeV. The PS also served as the pre-accelerator machine for the Intersecting Storage Rings (ISR) experiment [86] in 1960. Proton bunches are then injected to the super-proton-synchrotron (SPS) to gain the acceleration upto 450 GeV before entering the main LHC ring. The SPS has 1317 conventional (room-temperature) electromagnets, including 744 dipoles to bend the beams round the ring of 7 km circumference. The SPS has itself a major role in the Nobel-prize winning experiment for the W and Z boson discovery in 1983.

Inside the LHC ring proton bunches are accelerated to 6500 GeV and two bunches traveling opposite directions collide with each other at centre-of-mass energy of $\sqrt{s} = 13$ TeV during Run 2 operation. The maximum permissible $\sqrt{s}$ of the LHC machine is 14 TeV. The proton bunches are accelerated by RF cavities oscillating at 400 MHz with 16 MV electric field along the straight portions of the ring. 1232 dipole magnets consists of Niobium-Titanium coils produces 8.3 T magnetic field to bend the protons inside the LHC ring and 392 quadruple magnets, each 57 meters long are used to focus the beams of proton bunches by squeezing them at the collision points. The specification of the proton bunches used in the Run 2 LHC period is mentioned in Tab. 4.1.

The rate of the collisions per second depends on the instantaneous luminosity (L). The time integral of L, is called integrated luminosity ($\mathcal{L}$). For a given process with cross section σ, the expected number of events produced at the LHC with luminosity $\mathcal{L}$ is given by,

$$\begin{aligned}
N_{evt} &= \sigma \int L dt \\
&= \sigma \mathcal{L}
\end{aligned} \tag{4.1}$$

The instantaneous luminosity, L is defined as:

$$L = \frac{N_b^2 n_b f \gamma}{4\pi \sigma_x \sigma_y} F = \frac{N_b^2 n_b f \gamma}{4\epsilon_n \beta^*} F \tag{4.2}$$

Some of the values of the different parameters of the above equation are given in Tab. 4.1. In

addition, ϵ_n & β^* are the transverse emittance and amplitude function of the beam respectively, which can be defined in terms of transverse bunch dimension $\epsilon_n \beta^* = \pi \sigma_x \sigma_y$. The transverse emittance (ϵ_n) is a beam quality concept, by keeping the emittance parameter small; one can enhance the rate of the collision i.e achieve higher instantaneous luminosity. The amplitude function, β^*, is determined by the accelerator magnet configuration and power; β small means the beam is narrower and squeezed. γ is the relativistic Lorentz factor and F is a geometric factor which accounts for the luminosity dependence due to the crossing angle θ_c between the beams at the interaction point.

$$F = \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right]^{-1/2} \tag{4.3}$$

where σ^* is the effective dimension of the beam along the transverse direction. Taking the values of the parameters from Tab. 4.1, the instantaneous luminosity is found to be of the order of $10^{34} cm^{-2} s^{-1}$. Pileup (PU) is the average number of simultaneous interactions due to the

Table 4.1: Specification of proton bunches at LHC in Run 2.

Bunch dimension (longitudinal) (σ_z)	~ 5.5 cm
Bunch dimension (transverse) ($\sigma_{x/y}$)	~ 15 μm
Amplitude function (β^*)	40 in 2016, 40 $\rightarrow$ 30 in 2017, 30 $\rightarrow$ 27 $\rightarrow$ 25 in 2018
Beam emittance (ϵ_n)	1.8 - 2.0 mm-μrad
No. of protons per bunch (N_b)	~ 1.25 (1.1) $\times 10^{11}$ in 2016-17 (18)
Average bunch crossing rate (R_b)	29 MHz
No. of bunches in ring (n_b)	2220 in 2016, 2556-1868 in 2017 & 2556 in 2018
No of turns / sec (f)	11245
Half crossing angle (θ_c)	140 - 185 μrad in 2016, 120 - 150 μrad in 2017
	& 130-160 μrad in 2018
Peak instantaneous luminosity (L)	~1.5 (2.1) $\times 10^{34}$ cm^{-2}s^{-1} in 2016 (17 & 18)
No of collisions / sec	~1.0 (1.4) $\times 10^9$ in 2016 (17 & 18)
No of average pileup $< PU >$	~23 (33) in 2016 (17 & 18)

in-time and out-of-time bunch crossings, and define as

$$< PU >= \frac{L\sigma_{in}(pp)}{R_b} \tag{4.4}$$

$\sigma_{in}(pp)$ is the inelastic cross section of p-p collision, roughly $\sigma_{in}(pp)$ ~80 mb and R_b is the average bunch crossing rate. The year-wise distribution of the number of pileup events from
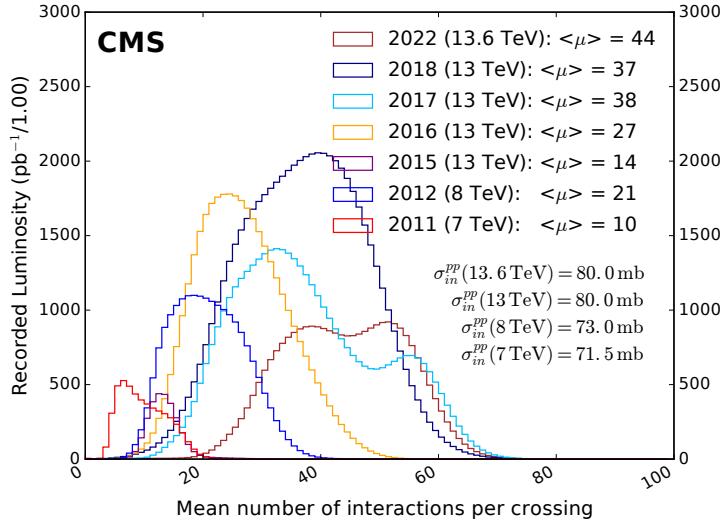
Figure 4.2: The distribution of the average number of pileup interactions in p-p collisions in the data-taking year of 2011 (red), 2012 (blue), 2015 (purple), 2016 (orange), 2017 (light blue), 2018 (navy blue), and 2022 (brown).

CMS experiment at the LHC is shown in Fig. 4.2.

A summary of the past, present and future run status of the LHC machine has been summarized in Fig. 4.3. The operation of LHC started in 2009, and is being continued in three different run periods Run 1 (2010-2012) , Run 2 (2015-2018) & Run 3 (2022-until 2025) with centre of mass energy 7-8 TeV, 13 TeV & 13.6 TeV respectively. These run periods are interleaved with long shutdown (LS) periods when both the machine and the experiments are upgraded in various ways. This pattern of the LHC operations will see LS3 during 2026-2028. Afterwords the LHC will have a major upgrade program towards High Luminosity LHC (HL-LHC) and will be continued its operation till 2040 with higher $\sqrt{s} = 14$ TeV and enormously high instantaneous luminosity (5 - 7.5 $\times 10^{34}$cm^{-2}s^{-1}), expected to collect 3 ab^{-1} data (10 times than the current volume) at the end of HL-LHC data taking period.
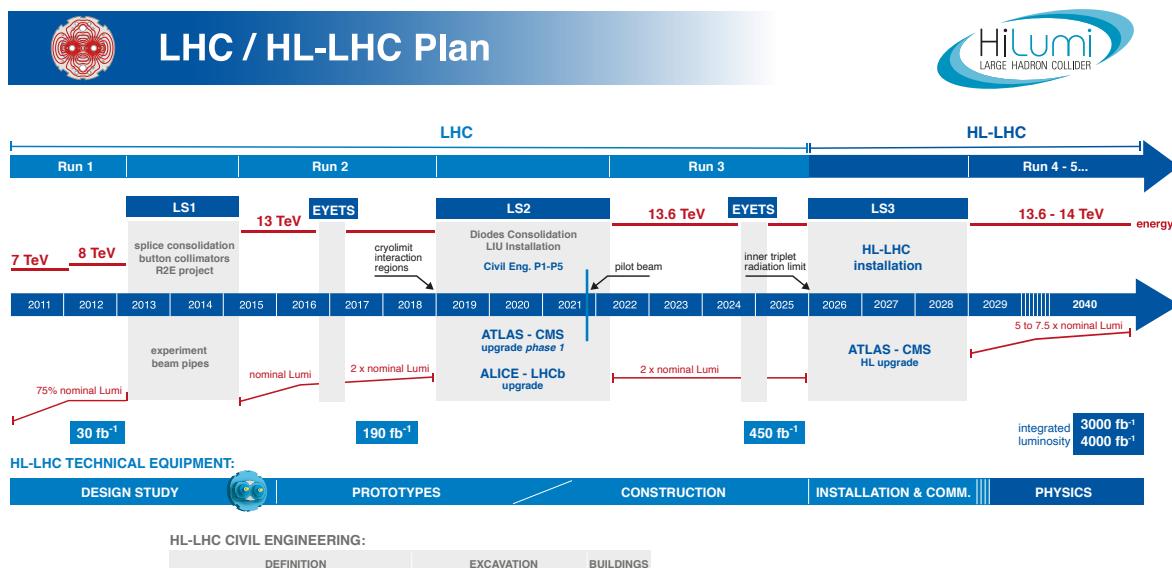
Figure 4.3: Summary of the past, present & planned future operation of the LHC as of 2022.

Chapter 5

Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of the two general purpose experiments located at 90 meters underground at the interaction point 5 (P5) of the LHC ring at the French Village Cessy near Geneva, Switzerland. The main goal of the CMS experiment was to discover the Higgs boson and to resolve the electroweak symmetry breaking mechanism. Besides the discovery of the Higgs boson, CMS has broad area to explore: study of the electroweak and strong interaction (QCD physics) at high energy scale, precision measurements of W and Z bosons and search for new particles predicted by the new physics models. To study the physics of different energy regimes, CMS detector is designed to be sensitive for the particles with a mass ranging from few GeV to few TeV. The importance of having the different sub-detectors of the CMS experiment has been described below.

1. Before the start of the LHC, the LEP experiment already gave a lower bound of H mass to be $M_{\text{H}} > 114.4$ GeV, for the range of M_{H}: 115 - 130 GeV, the best way to detect the Higgs boson is through the H $\rightarrow \gamma\gamma$ channel. Hence the requirement for high resolution electromagnetic calorimeter (ECAL) to enable the very precise measurements of the photon energy with good resolution.

2. If H would be heavier, like, $M_{\text{H}} > 150$ GeV, then it would be the best way to study through either H $\rightarrow$ W$^+$W$^-$ or H $\rightarrow$ ZZ channel, when W or Z decaying to electrons and muons. The leptonic final states are relatively easier to explore than the hadronic one in terms of suppression of the backgrounds. Good ECAL can be used to serve for precise electron detection just like as photon, but to detect muon and to measure its momentum

a muon chamber is needed with a charge identification criteria.

3. Many supersymmetry particles predicted from the BSM theory, or even SM predicted particles like W or Z bosons can decay to quarks, subsequently produce jets from hadronization and fragmentation. Hadron calorimeter (HCAL) is an important sub-detector to measure the energy of the hadrons.

4. To measure the momentum of the charged particles (electrons, muons or charged hadrons) from the bending of the trajectory inside a magnetic field is very important. To bend the track and to measure the bending a powerful magnet and silicon trackers are used in CMS.

5. Reconstruction of the secondary vertices originate from the decay of short lived mesons (B or D meson), a high granularity pixel detector is needed at the very close to the interaction point, which can work in the very large number of tracks with high precision.

6. Finally the detector should be hermetic with a 4π coverage, so that it can measure all the particles produced in one collision, so that the energy carried by an undetectable neutrino can be measured from the momentum imbalance in the event. This is also very important for new physics searches like dark matter candidate, where the interaction of the new particles with the detector material is not known.

5.1   CMS coordinate system

The CMS experiment uses a right handed coordinate system, where the center of it is at the interaction point inside the detector. The x-axis points towards the center of the LHC accelerator ring, y-axis is along the perpendicular upward direction from the plane of the LHC ring and z-axis points towards the anti-clockwise direction of the beam. The longitudinal and the transverse directions are along the z-axis and x-y plane respectively. The coordinate system is shown in Fig. 5.1.

As the LHC collides protons which are not fundamental particles, and the constituents of the protons, the partons collide with each other with unknown fraction of the total beam energy. Hence the net longitudinal momentum in the initial state is not known. However the transverse momentum is defined as p_T ($p_\mathrm{T} = \sqrt{p_x^2 + p_y^2}$), in the initial state is zero and it should also be nil in the final state due to momentum conservation. The azimuthal angle ϕ is measured the x-axis
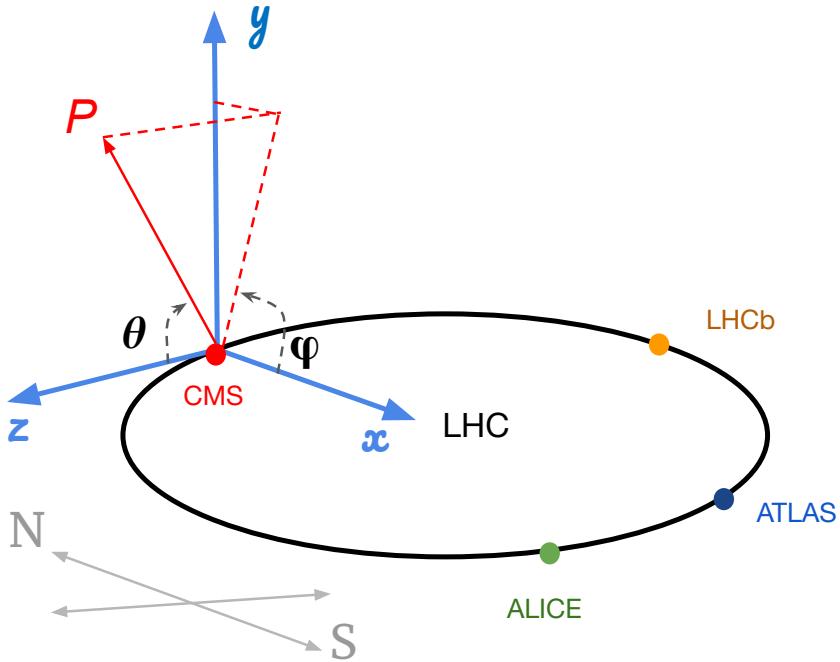
Figure 5.1: The CMS coordinate system.

in the x-y plane, $\phi = \tan^{-1}\frac{p_y}{p_x}$ and the radial coordinate in this plane is denoted by a radial vector r. The polar angle θ is measured wrt the z-axis as illustrate in Fig. 5.1.

An useful kinematic variable, which is often used in the LHC physics is rapidity (y), where the rapidity difference of two particles ($\Delta y = y_1 - y_2$) is a Lorentz invariant quantity, y is defined as:

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z} \tag{5.1}$$

For a massless particle or for a particle with relatively small mass compared to its momentum ($E \approx |\vec{p}| \gg m$), the definition of the rapidity can be approximated by a quantity, called pseudo-rapidity (η), the difference of pseudorapidity of two particles ($\Delta\eta = \eta_1 - \eta_2$) is also a Lorentz invariant quantity. Pseudorapidity can be defined in terms of the polar angle θ as,

$$\eta = -\ln\tan\frac{\theta}{2} \tag{5.2}$$

The value of η is zero at the central part of the detector for polar angle $= 90°$, and $\eta = \pm\infty$ along the beam direction, $\theta = 0, \pi$. The variation of η with respect to the polar angle θ, has been shown in Fig. 5.2. CMS detector covers a region up to $\eta \sim \pm 5$, which corresponds to an angle of $0.8°$ wrt beam line.
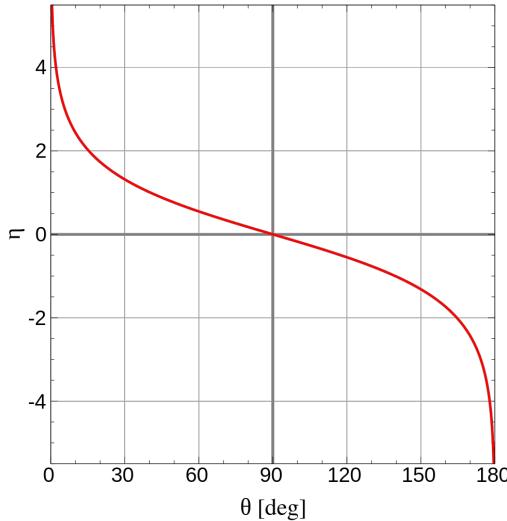
Figure 5.2: Pseudorapidity η as a function of the polar angle θ.

5.2 Superconducting solenoid magnet

An uniform magnetic field in the central region of the detector is provided by a superconducting solenoid of lengths 12.6 m and a diameter of 6 m and weight of 200 ton. The superconducting magnet consists of four layers Niobium-Titanium (NbTi) coils with 542 turns in each of the layers. It is kept inside a liquid Helium cryostat at an operational temperature of -268.65°C to reach at the superconductivity state. It provides an uniform magnetic field of 3.8 Tesla. The pixel tracker, silicon strip detector, ECAL and HCAL are placed inside the solenoid. A steel return yoke surrounds the NbTi coils and produces the bulk of the detector mass. The return yoke acts as an additional absorber to guide and confine the magnetic flux to the detector volume and houses the four muon stations outside the solenoid magnet. The magnetic coils and the return yoke weigh around 12500 tons and provide most of the structural support to hold the full experiment. The main goal of the CMS magnetic system is to measure the momentum of the charged particle from the bending of its trajectory ("sagita") inside the magnetic system, the relationship between the sagita (s), the lever arm (L) and the magnetic field (B) for a particle with transverse momentum p_T is defined as:

$$s \text{ (m)} \simeq \frac{0.3B(T)L^2(m^2)}{8p_T(\text{GeV})} \tag{5.3}$$

The resolution of the momentum of charged particles $\frac{\sigma(p_T)}{p_T}$ depends on the magnetic field, and can be expressed as,

$$\frac{\sigma(p_T)}{p_T} = \frac{\sigma(s)}{s} \propto \frac{p_T}{BL^2} \tag{5.4}$$

Equation 5.4 shows that for the measurement of the momentum with high resolution, high magnetic field is required.

5.3   Tracker

The CMS tracker [87] sub-detector consists of a cylinder of 5.8 m in length and 2.6 m in diameter, placed at the inner most part at the detector. The main goal of the tracking system is to reconstruct the tracks of the charged particles as precise as possible and to reconstruct the position of the secondary vertex which is crucial for the study of the long lived particles (life time $\tau > 1$ ps) and to tag the quark-flavor of the jets. The tracking system was built under several basic requirements.

- In the p-p collision mode of the LHC the number of pileup events are very high, so tracking system needs to reconstruct the tracks from the high particle flux and then needs to associate them to the correct vertices at the origin.

- As the tracker system is closest to the detector interaction point, it experiences an enormously high radiation dose. So it is desirable that the tracking detector should be radiation hard.

- For tracking, the measurement is non destructive. Hence the material budget is expected to be minimal to avoid the energy losses and multiple scattering of particles inside the tracker material before reaching to the calorimeter sectors.

CMS exploits two different types of tracking sub-system: pixel detector and silicon strip detectors. Figure. 5.3 shows the longitudinal view of the CMS tracker system with both types of CMS tracker subsystems.

5.3.1   Pixel tracker

Pixel detector provides three dimensional position measurement of the charged particles at very close to the beam pipe. Originally the pixel detector [84] had three concentric layers made of silicon pixels at the barrel region (BPIX) with radii 44, 73 and 103 mm and two circular disks (FPIX), each side of the interaction vertex at the distances of ±345 and ±465 mm from the interaction point, this is called Phase-0 CMS tracking system and was used in Run 1 and up to 2016 of Run 2 LHC data taking periods. After that, there was a major upgrade
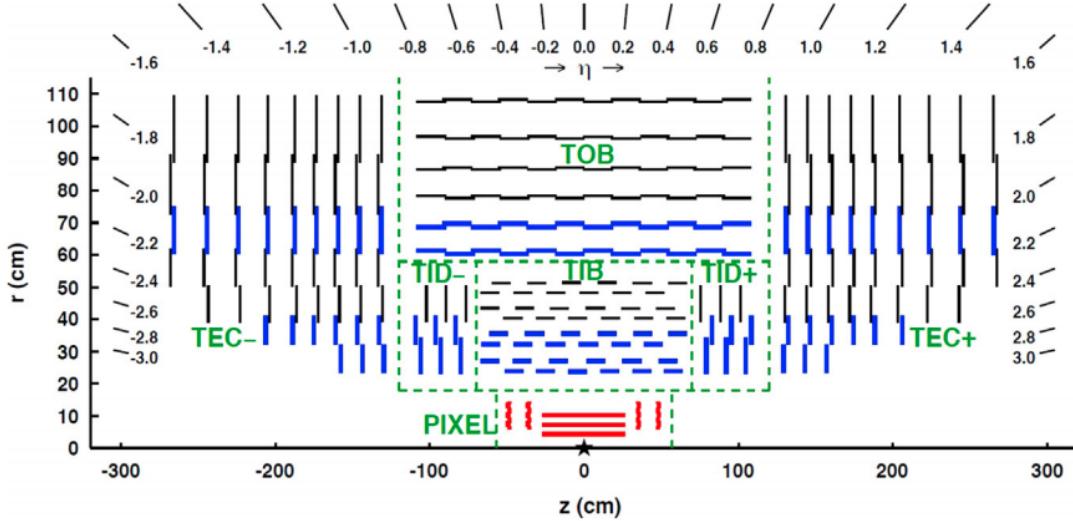
Figure 5.3: Longitudinal view of CMS tracking system operated till 2016 in the r-z plane. Pixel detector layers are shown in red, single and double sided strips are presented by black and blue thick lines respectively. The figure is taken from Ref. [6]
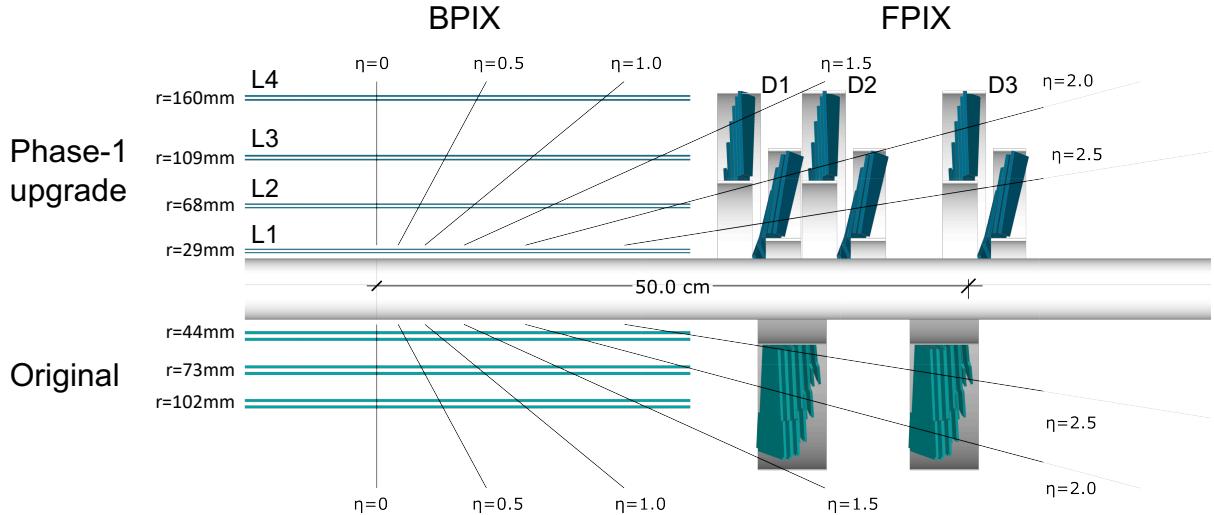
Figure 5.4: Longitudinal layout Phase-1 CMS pixel tracker compared to the original detector.

program during the shutdown period of 2016-17, and its upgraded version is called Phase-1 pixel detector [88, 89] system, which was used in the rest of the Run 2 data taking period starting from 2017. The Phase-1 upgraded pixel detector is shown in Fig. 5.4 on top of it's Phase-0 layout.

The Phase-1 pixel detector consists of four barrel layers (L1-L4) at radial distances of 29, 68, 109, 160 mm and 3 forward circular disks at each side (D1-D3) at distances 291, 396 & 516 mm

from the interaction point. Each FPIX "D" has two parts called inner ring and outer ring. Both in BPIX and FPIX, one additional layers are added at Phase-1 upgrade program, while the inner most layer of the BPIX comes close more to the beam line. The geometric position of the each BPIX and FPIX layers are presented in Tab. 5.1. The pixel detector gives position resolution of 10 μm in transverse (x-y) plane and 20-40 μm in the longitudinal (z) direction. The pixel detector has total 1856 segmented silicon sensor modules covering an area of 1 m^2, while 1194 at BPIX and 672 at FPIX. Each tracker module consists of 160×416 pixels connected to 16 readout chips (ROCs). The total system is operated inside a temperature of -10°C to reduce the leakage current due to the radiation damage.

Table 5.1: Summary of the positions and number of modules of the pixel detectors used in BPIX & FPIX.

Layer	Radius (mm)	z position (mm)	Number of modules
BPIX			
L1	29		96
L2	68	-270 to 270	224
L3	109		352
L4	160		512
FPIX			
D1 inner ring	45 -110	± 338	88
D1 outer ring	96 -161	± 309	136
D2 inner ring	45 -110	± 413	88
D2 outer ring	96 -161	± 384	136
D3 inner ring	45 -110	± 508	88
D3 outer ring	96 -161	± 479	136

5.3.2 Silicon strip detector

The silicon strip detectors are the outer part of the tracking system with a coarse resolution in position measurements than the pixel detector. The particle flux decreases as with increasing of the radial distance from the interaction point. It follows $1/r^n$ rule, where r is the radial length of the detector layer and n $\geq$ 2. Figure. 5.3 shows the longitudinal structure of the silicon strip chambers. The inner most part of the strip detector consists of 4 concentric cylindrical layers in

barrel (TIB) and 2 disks at the side (TID). The position resolution of these inner layers is about 13-38 μm in $r - \phi$ plane. Next layer is the outer silicon strip detector, divided into two parts (i) in barrel six silicon strip layers (TOB) provide resolution of 18-47 μm in $r - \phi$ direction and (ii) nine layers at the endcap (TEC), each containing up to seven concentric rings of silicon strips. Additionally, in the modules of the first two layers (rings) of TIB (TID) and TOB and first, second and fifth rings of the TEC, a second micro-strip detector module is mounted back-to-back to the first with a stereo angle of 100 mrad, called double sided module. The hits from the two back to back layers which provides the measurements both in $r - \phi$ as well as z coordinate. The resolutions from these double sided module of TIB and TOB are 230 μm and 530 μm in $r - \phi$ plane, respectively. Figure. 5.5 shows the material budget of the CMS tracker in units of radiation length(X_0) as a function of pseudorapidity (η). At the very central part of the detector, where η =0, the material budget is about 0.4 X_0; it increases with η, and at $|\eta|$ =1.4 it is about 1.8X_0, and then it falls about 1 X_0 at $|\eta|$ =2.5.
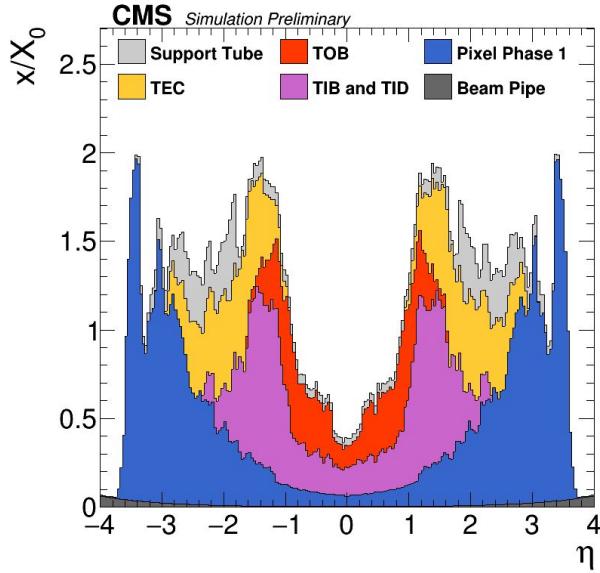
Figure 5.5: The material budget of CMS tracking system in terms of radiation length (X_0) as a function of pseudorapidity (η).

5.3.3 Performance of the CMS tracking system

The performance of the CMS tracking system is expected to depend on the number of pileup (PU) events, as it is directly related to the number of reconstructed track, coming from the primary vertex as well as the PU vertices. Hence the performance has been checked with
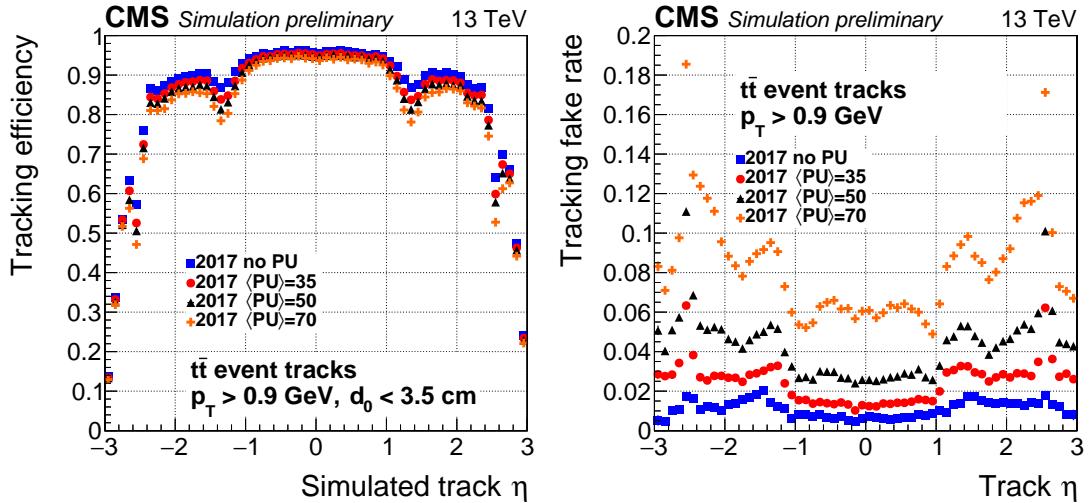
Figure 5.6: Track reconstruction efficiency (left) and fake rate (right) as a function of simulated track η for the CMS tracker at different pileup conditions.

simulated events for inclusive production of top pairs ($t\bar{t}$ +X) configured with full CMS detector geometric condition using GEANT-4 at $\sqrt{s}$ =13 TeV with different PU conditions 0, 35, 50 & 75. Both the efficiency and fake rate of the reconstructed tracks is shown in Fig. 5.6 in different η regions.

It is evident and also understandable that the track reconstruction efficiency is decreasing with the increase of PU. The relative resolution of the track's transverse momentum as a function of p_{T} and η of the tracks has been shown in Fig. 5.7. The comparison of Phase-1 tracker performance with respect to the original (Phase-0) tracker scenario is also shown for comparison.

5.4   Electromagnetic calorimeter

As already mentioned the main goal of the CMS experiment was to discover the Higgs boson through the decay H $\rightarrow \gamma\gamma$ and H $\rightarrow$ W$^+$W$^-$/H $\rightarrow$ ZZ via the leptonic decays of W, Z bosons. This demands a precise reconstruction of photon and electron, with high energy-momentum resolutions as well as position. To achieve the best results the CMS experiments uses an electromagnetic calorimeter [90] made of roughly 76,000 Lead Tungstate (PbWO$_4$) crystals doped with Niobium to increase radiation hardness. The main reasons behind the choice PbWO$_4$ are (i) small radiation length X_0 (0.89 cm) which helps to make the ECAL to be compact (ii) small moliere radius (2.2 cm), which helps to confine the shower in one crystal, increases spatial
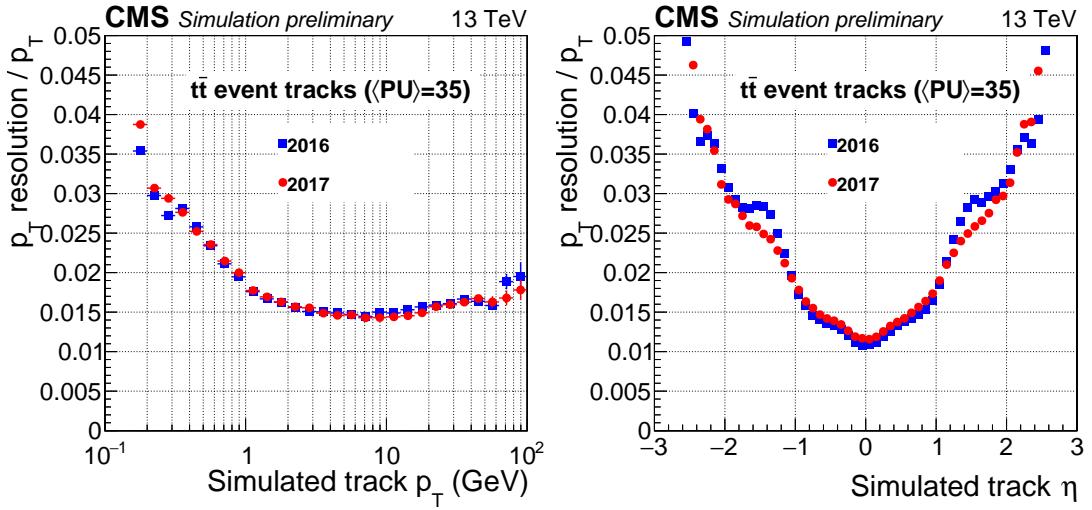
Figure 5.7: The relative resolution of track reconstruction with respect to the p_T (let) and η (right). The performance of Phase-0 (Run 1 and Run 2 till 2016) and Phase-1 (from 2017) tracker subsystems are imposed here for comparison.

resolution in lateral direction (iii) the fast response with decay time of 15 ns, which keeps the ECAL crystal to be ready for the next collision. It consists of two parts, ECAL barrel (EB) mostly at the central part of the detector (coverage up to $|\eta| < 1.479$) and the ECAL endcap (EE) at the higher pseudorapidity region ($1.5 < |\eta| < 3.0$). The layout of the CMS electromagnetic calorimeter, describing the crystal barrel and endcap detectors, as well as the silicon preshower detectors (described below) in front of EE, is shown in Fig. 5.8.

The EB consists of 61200 crystals. The crystals are aligned with a small angle of $3°$ with respect to the interaction point to avoid any trajectory of γ passing through the detector gaps. Each crystal size is 22×22 mm^2 in the font face and 26×26 mm^2 in the rear face and covers an area of 0.0174×0.0174 in $\eta - \phi$ plane. The length of each crystal is 230 mm corresponding to 25.8 X_0.

The EE is placed at a distance of 315.4 cm from the interaction point. On the either side along the z direction, each EE is divided into two Dees, which holds 3662 crystals, each crystal has a 28.62×28.62 mm^2 in the font face and 30×30 mm^2 at the rear face with a length of 220 mm (24.7 X_0).

The light yields of the scintillator material is very low, therefore an amplification is done by applying high gain photodetectors, Avalanche Photo Diode (APD) used in the EB and Vacuum Photo-Triodes (VPT) in the EE. The number of scintillation photons, emitted by the ECAL
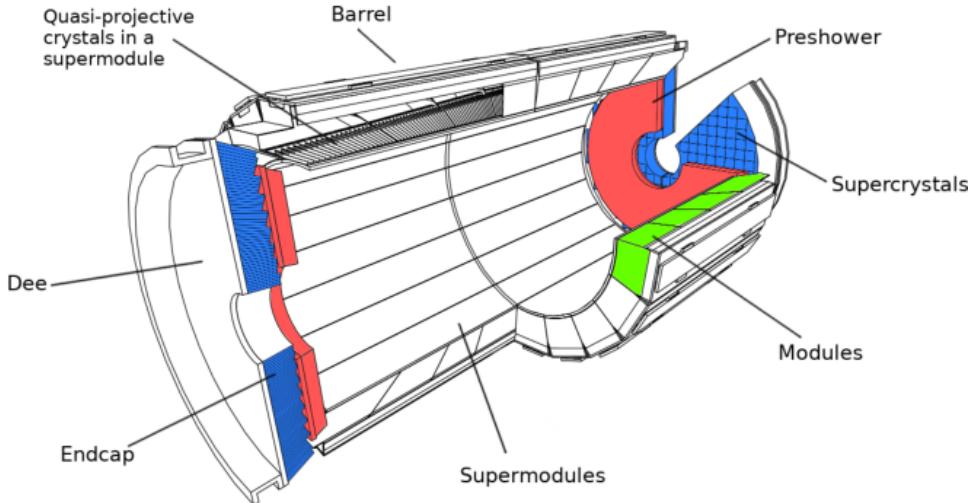
Figure 5.8: Layout of the CMS electromagnetic calorimeter. The figure is taken from Ref. [7].

crystals and the amplification at the photodetectors are both as temperature dependent with negative temperature gradient. Therefor the temperature of ECAL is maintained at a constant value very precisely, requiring a cooling system to reduce the heat dissipated from the read-out electronics. The nominal temperature of the ECAL system is 19° C.

In addition to that a sampling calorimeter, called pre-shower detector is placed, in front of the EE with a coverage of $1.653 < |\eta| < 2.6$. It consists of two silicon micro strips and two lead absorbers. The main purpose of the pre-shower detector is to differentiate the prompt photons and two closely separated photons coming from the decay of π^0.

Finally, the ECAL resolution can be expressed by the following formula,

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \tag{5.5}$$

Where E is the energy of the electromagnetic object and S: stochastic term arising from the statistical fluctuations in the light efficiency, N: noise term due to the electronic noise and pileup and C: constant terms for the detector inhomogeneities and calibration. The terms in the ECAL resolution are derived from the dedicated test beam data analysis, where it is found that S= 2.8%, N = 12%, and C = 0.30%. For the low energy photons, the stochastic and the noise terms dominated over the constant term whereas for very high energetic photons the constant term dominates.

The performance of the ECAL has been measured in collision data from the di-electron mass resolution of the Z $\rightarrow$ e$^+$e$^-$ peak which is found to be 1.2% and 2.9% in EB and EE as shown

in Fig. 5.10. The relative resolution of momentum measurement of a single electron is about 1.7% and 4.1% for EB and EE respectively [8].
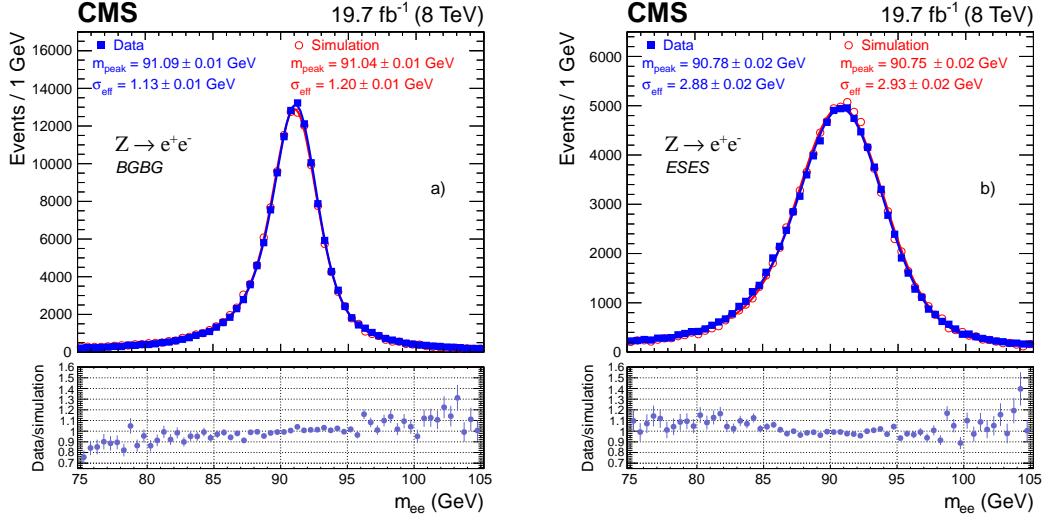
Figure 5.9: The invraint mass distribution of $Z \rightarrow e^{+}e^{-}$ events in data and MC, where the both the two electrons are either in EB (left) or EE (right). Figure is taken from Ref. [8]

The relative resolution of the energy measured for the unconverted or late converted photon is around 1.5% upto $|\eta| = 1.0$ and 2-3 % in $1.0 < |\eta| < 1.5$ and 3.5% at the EE [9], whereas for the converted photons resolution is slightly degraded. The variables R_9 is used to quantify the possibility of a photon getting converted or not, by measuring its deposition of energy fraction within a 3×3 matrix of ECAL crystals around the maximum energy deposited crystal (called seed crystal), which is defined in Sec. 9.4.
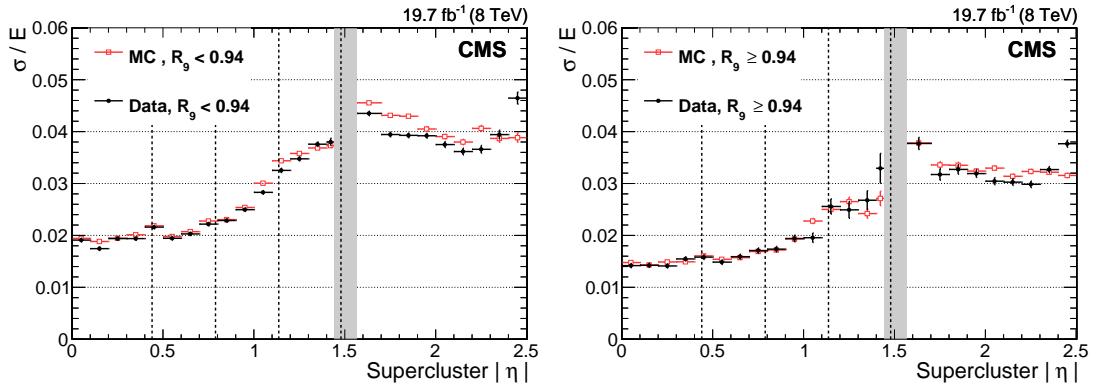
Figure 5.10: The energy resolution of the photons in different η regions for high and low R_9 region. Ref. [9]

5.5 Hadronic calorimeter

Hadron calorimeter (HCAL) serves to measure the energy of jets which is a collimated flux of particles, consisting predominantly of charged hadrons (65%), electromagnetic component (25%) and neutral hadrons (10%). Jet is a very important object to use in search for super symmetric particles as well as in precision studies for various SM physics. The HCAL measures the energy for the charged and neutral hadrons and also provides information part in the reconstruction and isolation of the photons, electrons and muons by comparing the fraction of their energy deposition in the HCAL. These are used in the particle flow (PF) algorithm, discussed in Sec. 6.

The structure of the HCAL system from the longitudinal direction is shown in Fig. 5.11. It consists of four sub-systems (i) HCAL barrel (HB) covers upto $|\eta| < 1.4$, (ii) HCAL endcap (HE) in $1.4 < |\eta| < 3.0$, (iii) HCAL forward (HF) in $2.85 < |\eta| < 5.0$ and finally (iv) one additional layer outside the solenoid magnet (HO) in the barrel region $|\eta| < 1.4$. As the longitudinal spread of the hadronic shower is longer than the electromagnetic shower, it is practically impossible to use a homogeneous calorimeter for the hadrons to confine its full shower. So HCAL system in CMS is a sampling type caloirmeter.
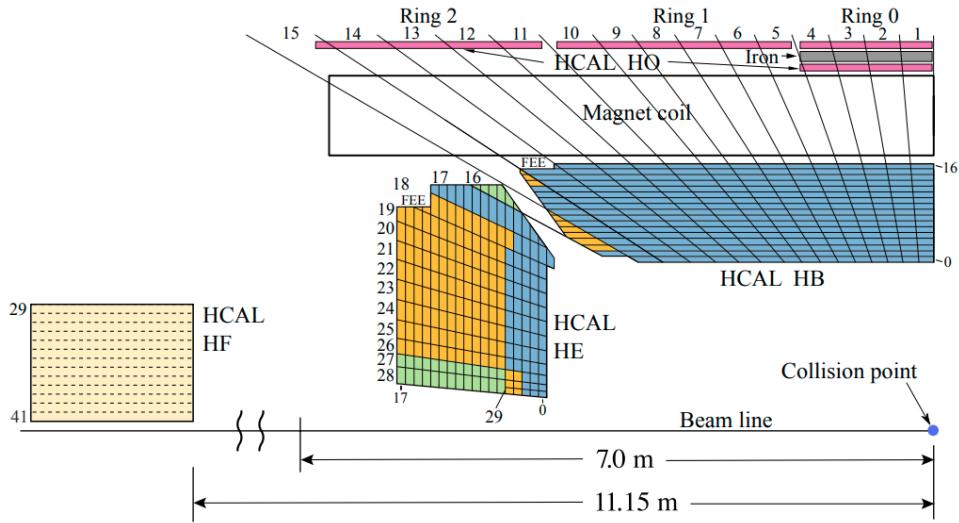
Figure 5.11: Layout of the CMS hadronic calorimeter. Ref. [10].

HCAL barrel (HB) is divided into two component (HB+ and HB-) across the negative and positive sides along the z-axis, divided into 18 wedges in ϕ and built up with 14 brass (an alloy with 90% copper and 10% zinc) plates as absorber. There are eight brass plates of thickness

50.5 mm and six plates of 56.5-mm to initiate the hadronic shower. The plastic scintillator plates are used to measure the produced hadronic shower placed in between the two absorber plates. The plastic scintillator is divided into 16 η sectors with a 0.085×0.085 segmentation in $\eta - \phi$ direction. The front (40 mm thick) and the back (70 mm thick) layers are made of steel to give additional support to the structure. The effective thickness of the HB is polar angle dependent ($\propto 1/\sin\theta$), it is 5.92 radiation length (λ_0) at $\eta = 0$ and 9.6 λ_0 at $|\eta| \sim 1.4$. Addition of ECAL material budget along with HCAL increase the effective radiation length to about 1.1 λ_0.

The HE is placed at the end of the solenoid magnet, so the abosorber material should be a non-magnetic material. It covers the largest solid angle of 13.2%, segmented in 18 wedges. The granularity of the HE calorimeter is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ up to $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for $|\eta| \geq 1.6$. The effected length of HE including ECAL endcap, is about 10 interaction lengths (λ_0).

Due to comparatively small material budget at the central part of the detector, it is difficult to confine the hadronic shower within HB. Hence an additional layer of HCAL has been placed just after the solenoid magnet and before the front portion of the CMS muon chamber, called outer hadron calorimeter (HO) to catch the tail portion of the hadron shower. The HO covers the region of $|\eta| < 1.26$. It consists of five cylindrical rings along the z direction, one at the central, named Ring-0 and two others Ring-1 and Ring-2, on the both side. The material budget of HB at $\eta = 0$ is minimum, so the central Ring-0 of HO consists of two layers of scintallators placed at radial distance of 3.82 m and 4.07 m, on both sides of a 19.5 cm thick iron. The total effective length after inclusion of HO, is thus increased to 11.8 λ_0.

The forward hadron caloimeter (HF) is tasked to function in the region of unprecedented particle fluxes. The detector consists of a steel absorber of 5 mm thick plates and quartz fibers as activation materials, which are placed parallel to the beam to collect the Cherenkov lights. The choice of the quartz fiber is to have the sustainability in exceptionally high radiation field in forward region. Half of the fibers run through the full depth of the absorber (165 cm), they are called long fibers, and another set runs after 22 cm from the front end absorber (short fibers). As the electromagnetic shower mostly confined in the early part of the HF, so long fibers are able to catch both electromagnetic and hadronic showers, whereas the short fibers measures only the hadronic component.

The energy resolution of the jets clustered from the energy depositions in the HCAL and the ECAL combination, can be written as:

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}\text{(GeV)}} \oplus 8.0\% \tag{5.6}$$

The measurements of energy deposition in the HCAL is significantly worse compared to the ECAL. The constituent particles of a jet is dominated by the charged hadrons. The momentum of the charged hadrons can be measured with high precision from the tracker. So, inclusion of tracking information along with the HCAL and the ECAL information resulted in the reduction in the first term of the above equation to 70%. This has very significantly effect on the physics performance of the CMS experiment.

5.6 Muon system

The precise determination of muon momentum in CMS is the key feature for the golden channel for the discovery of the Higgs boson, H $\rightarrow$ ZZ* $\rightarrow$ 4µ, as well as the SM precision studies of W and Z boson properties. In general, the muon system has 3 functions: identification, momentum and charge determination and finally triggering the muon events within a large range of muon momentum from few GeV to few TeV. Muons are almost 200 times heavier than the electrons and thus deposit very small fraction of their energy to the other subdetectors (via multiple scattering) while passing through, so it is not possible to measure the energy of muons from the calorimetric information. The muon system is a gaseous particle detector, meant to measure the momentum of the muons from bending muon tracks under magnetic field. It consists of one cylindrical barrel part and two endcap disks on either side of the detector. Figure. 5.12 shows a schematic view of CMS muon system in r-z plane.

The barrel muon system covers up to $|\eta| < 1.2$ consisting of 4 concentric cylindrical drift tube (DT) chambers (MB1, MB2, MB3 and MB4), each made of 3 independent super layers (SL), excepts MB4, which has 2 SLs. Each SL consists of 4 layers of rectangular drift cells, stacked by a distance of half of their cell size. The wires of the 2 outer SL are parallel to the beam direction and able to provide the measurement the track in $r - \phi$ plane, while the wires of the inner SL are orthogonal to the beam direction and able to measure the z position. A single DT cell is shown in Fig. 5.13, the anode wire is a 50 μm gold-plated stainless steel wire, and cathods are on the either side of anode wire made of aluminum plate of 50 μm thick and 11.5
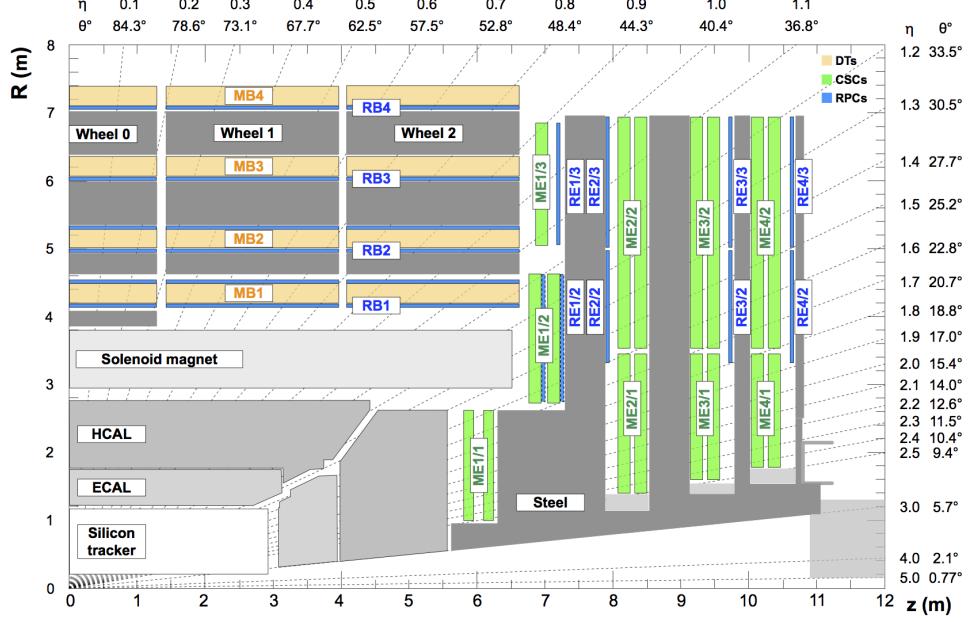
Figure 5.12: Layout of the CMS muon system. The figure is taken from Ref. [11].

mm wide, an high voltage (HV) of 1.75 to 1.85 kV is maintaining in between them. A mixture of 80% Ar and 20% CO_2 gas is used inside the DT cells. The spatial resolution of these DT chambers is 100 μm in the $r - \phi$ plane and 150 μm in the $r - z$ plane.
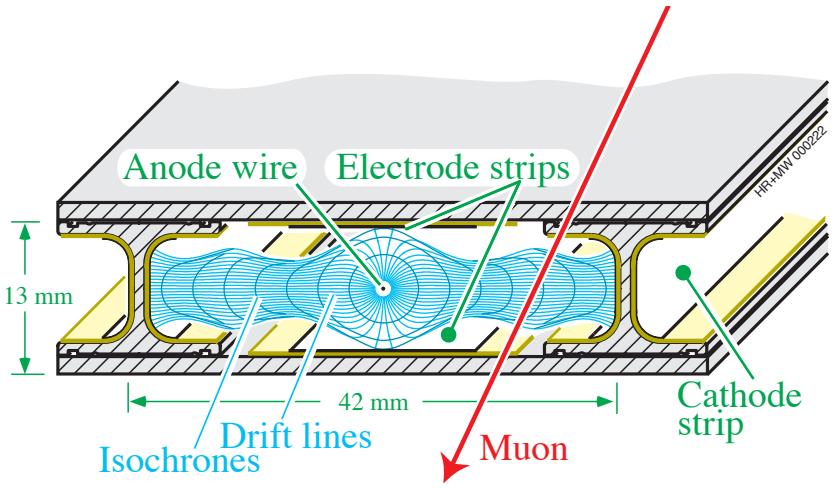
Figure 5.13: Schematic layout of a single Drift Tube cell. Ref. [12].

The endcap of the muon system is placed at $1.2 < |\eta| < 2.4$, in the high and non-uniform magnetic field region. Due to the large Lorentz angle of the DT muons, the position resolution in the endcap deteriorates; so DT is not a good option to use at endcap. The main ingredient of the endcap muon system is Cathode strip chamber (CSC). The endcap consists of four stations of CSC labeled as ME1, ME2, ME3 & ME4. Each stations are divided into two concentric

rings expect for the ME1 which has three. Each concentric rings of ME1 are divided into 36 chambers in ϕ direction and covers $10°$ each, whereas the innermost layers of ME2-ME4 host 18 chambers in ϕ of $20°$ each. Each CSC chamber has seven trapezoidal shaped CSC panels and cathode strips staggered radially; the anode wires are placed perpendicular to the strips. A schematic view of a trapezoidal shaped CSC panel is shown in Fig. 5.14 (left). Whenever a charged particle is traversing one CSC panel, it causes ionization of the gas and subsequent electron avalanche, which produces a charge collection on anode wire and corresponding image charge on cathode strips. A cross sectional view of a operational CSC module is presented in Fig. 5.14 (right). Thus a CSC chamber provides $r-\phi-z$ coordinate measurement by combining six layers, with the spatial resolution of 150-250 μm in $r-\phi$ plane and 1 mm in z directions.
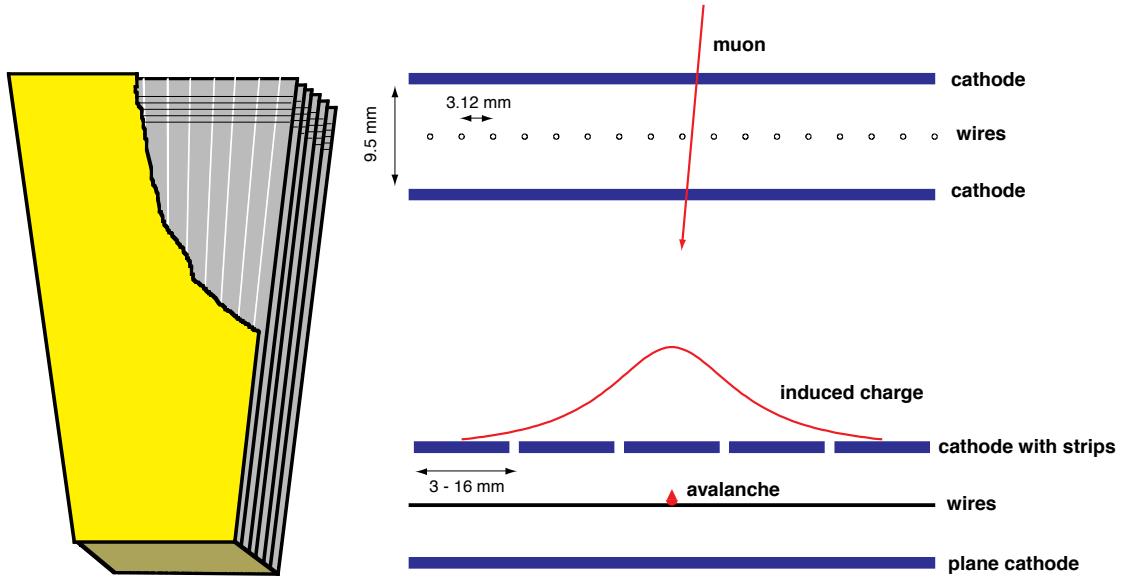
Figure 5.14: Left: A diagram of a trapezoidal shaped 6 layers CSC panel and the orientation of some of the cathode strips and anode wires. Right: Cross-sectional views of the operation of a CSC module, while a muon is traversing.

Both in barrel and endcap regions, resistiive plate chambers (RPC) are used along with DT and CSC to trigger the events with muons, which covers up to $|\eta| < 1.9$. The RPCs are made of Bakelite anode and cathode plates and a mixture of 96% $C_2H_2F_4$, 3.5% Iso-C_4H_{10}, 0.3% SF_6 gases are used. In barrel region, two RPCs are used on the either side of the two innermost DT chambers (MB1 and MB2) and only one RPC each for the last two outermost DT chamber (MB3 and MB4). In the endcap, RPCs are placed in all four muon stations. RPCs are designed to have good time resolution (~ 1 ns), and capable to identify the muons for a particular bunch

crossing with 25 ns interval. Hence RPCs are used to trigger an event with one or more muons.

The muon tracks can be reconstructed only from the hits in the muon chambers. The relative momentum resolution of the reconstructed muon using only information from muon the system is 8-15% at 10 GeV and 20-40% at 1 TeV. The precision of momentum measurement for the muon can be enhanced by combining the tracker information, the relative p_T resolution is 1-1.5% at 10 GeV and 6-17% at 1 TeV. The probability for wrong assignment of electric charge to a muon is less than 0.1% for p_T upto 100 GeV.

5.7   CMS Trigger system

At the LHC two proton bunches collide at an interval of 25 ns. For an instantaneous luminosity of $10^{34} cm^{-2} s^{-1}$, in a second there are about one billion p-p collisions in a second. In general, the digital size of one event is 1-2 Mega Byte (MB), which produces about Petabyte (PB) of raw data every second. This is beyond the reach of the storage capacity due to the limited disk quota. Most of the p-p collision events happen with low energy transfer producing low energy hadrons, which is not of the prime interest of the CMS physics program. So the main job of the trigger system is to take a fast judicious decision whether an event needs to be stored or not for the further offline analysis. For this purpose CMS uses a two level trigger system as described below.

The first level of trigger system (L1 trigger) is made of highly sophisticated customized hardware, which operates synchronously with the LHC collision. The rate of the L1 accept is 100 kHz. The upper threshold of L1 rate is limited by the bandwidth of the data transfer between the front end electronics to the computer storage farm. The time latency is about 3.8 μs, the L1 trigger system consists of two stages, regional and global triggering for both the calorimetric and the muonic system. The Regional Calorimetric Trigger (RCT): RCT receives information about the transverse energy (E_T) from the calorimetric energy depositions in the ECAL and HCAL towers and sends them as potentially interesting candidates along with the sum of E_T of the 4×4 towers to the next level, called Global Calorimetric Triiger (GCT). The GCT first sorts the e/γ candidates objects and construct jets (central, forward and tau candidates) using E_T sums and finally derive the event level objects p_T^{miss} and H_T (sum of p_T of all jets in a event). In the muon system, Regional Track Finder, RTF, uses the muon hits from the three muon sub-systems, drift tube (DT), cathode strip chamber (CSC) and resistive plate chambers (RPC).

The reconstructed tracks are then sent to the global muon trigger (GMT) for merging the muon candidates identified in different sub-systems and to sort them according to the p_T. Finally the GMT is linked with GCT to form the global event by combining the muon tracks with the calorimetric information. A flowchart of the CMS L1 trigger system is shown in Fig. 5.15. Finally a set of selections are imposed on these reconstructed objects at the L1 level to pick good events.
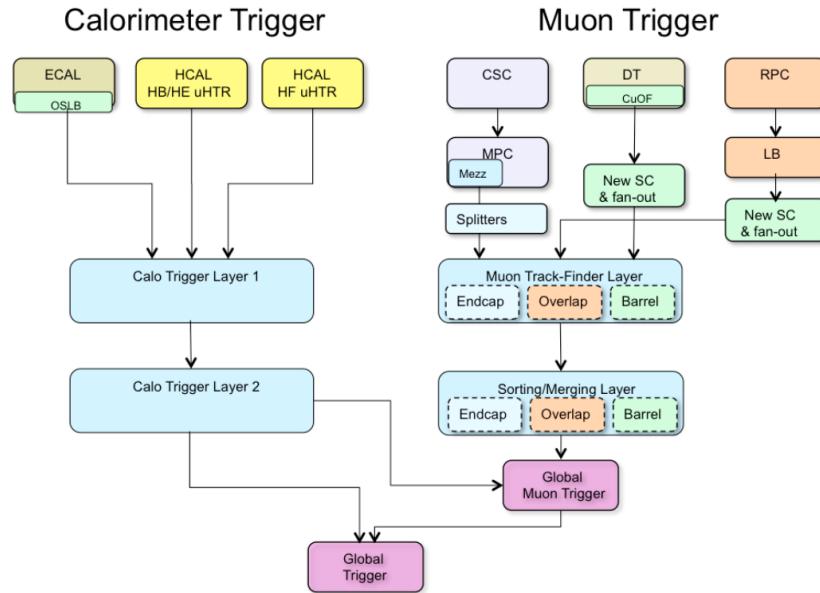
Figure 5.15: Schematic layout of L1 trigger system of the CMS experiment. The figure is taken from Ref. [13].

The second tier of the CMS trigger system is based on a computer farm receiving offline software reconstruction and called the high level trigger (HLT) which utilizes the information from all the sub-detectors including the tracker. The acceptance rate of the HLT system is about 1 kHz, only 1% of events that passed through L1 trigger. The optimization of the HLT algorithms is an interplay between the physics goal having particular final state objects with definite event signature and the processing time. There are several hundreds of HLT paths or criteria. One event can be triggered by multiple HLT paths. The events are stored for offline analysis only if they have passed at least one HLT condition. The output bandwidth of the HLT decision is about 2 GB/s. In the SM precision measurements, sometimes it is required to use particles of very low energy, but the production rate of these particles are extremely high in the LHC. So to control the rate of the HLT system, it stores a random event out of many passing the HLT conditions. This type of HLT paths are called prescale-HLT. Typically the prescale factor is a

function of the instantaneous luminosity foreseen and it is fixed in advance.

Apart from the normal HLT trigger system CMS has comprised two more strategy to accumulate a larger amount of data (i) to search for new physics in a difficult phase space or objects with non-standard p_T regimes, and (ii) to search B physics in a large samples of B hadrons. The first method is called Data scouting, while the second one is Data Parking. It is already mentioned that the HLT rate is about 1 kHz and stores only a fraction of the total p-p collisions due to the limitation of the bandwidth of the data transfer and also of the limited storage system, these constraints become more aggressive over time as LHC luminosity rises. In the scouting data, the loose selections on the HLT-level PF objects have been implemented by discarding the raw event information, which brings down the per event size to a few kB from 1MB. This allows to store larger data volume than the normal HLT triggers. The only limitation of the data scouting is, as there is no raw event information, so the offline level object reconstruction and calibration is not possible, one has to perform analysis only based the HLT-level objects. The term data parking means to the technique of selecting events at the HLT and moving them immediately to the tape storage, skipping the step of the prompt reconstruction. Events selected by this format remain on tape until there are sufficient free computing resources to reconstruct them offline. Data parking allows more than the standard 1 kHz of physics events to be recorded, because there is no constrains on the limited capacity of the prompt reconstruction system. Th parking of data only depends on the bandwidth of the CMS Data system and the amount of tape storage space available. Parking allows to keep the raw event information which discards the disadvantage of the scouting system of offline reconstruction. In CMS there is a significant number of physics analysis are being performed based on the scouting and parking data.

Chapter 6

Physics object reconstruction and identification

6.1 Particle flow algorithm

The Particle flow (PF) algorithm [16] is used widely in the CMS experiment to reconstruct and identify all the final state particles utilizing the full CMS detector information, which are further used to reconstruct composite objects like jet, tau and missing transverse energy. The PF algorithm maintains a hierarchy of the reconstruction of each type of object, it minimizes the confusion of assigning a detector response to a particular type of object. A brief description of different types of object reconstruction and identification is described below.

Muon identification proceeds via a set of selections for the global and tracker muons. Global muons are first selected by requiring the additional information of tracks and calorimetric energy deposits within a cone of radius 0.3 in $\eta - \phi$ plane around the muon's direction. The sum of p_T of all other tracks and the total E_T of the energy deposits inside that cone is required to be less than 10% of the muon p_T, for it to be identified as an isolated object as expected from say W or Z boson. This condition is sufficient to reject the muons inside the jets (nonisolated muons). Similarly there are subsequent requirements for the selection of non-isolated muons mostly muons inside the jets. These identified muons are then masked from the detector signatures for the further processing through PF algorithm to reconstruct the next set of particles.

The reconstruction of electrons and photons rely on the information from the tracker and the

energy deposition in the ECAL. Due to the large material budget of the tracker material, electrons often emit bremsstrahlung photons, and photons converts to e^+e^- , which further emit bremsstrahlung photons. So the tracking and the energy deposition patterns for the electron and the photon candidates are very much similar. The reconstruction algorithm first starts to form clusters of ECAL crystal above a certain threshold of the electronic noise of ECAL crsytal (80 MeV for EB and 300 MeV for EE) around the most energetic crystal (called the seed crystal) and the most energetic cluster above 1 GeV is called the seed cluster. The grouping of clusters around a seed cluster in a specific geometric area forms a supercluster (SC) which includes the photon conversions and electron bremsstrahlung loss. The sum of energies of the HCAL cells inside the radius of 0.15 in $\eta - \phi$ plane of the ECAL SC must not exceed 10% of the SC energy. All tracks and calorimetric clusters are used to form electrons and photons are masked form the PF algorithms for further processing.

After the selection of muons, electrons and photons, the remaining particles which are needed to be reconstructed are the charged hadrons ($\pi^\pm$, $K^\pm$, or protons), neutral hadrons (K^0_L, neutrons). Each of the ECAL and the HCAL clusters inside $|\eta| < 2.5$, which are not selected as isolated electrons or photons in the previous step and also not linked with any tracks, are determined as non-isolated photons and neutral hadrons respectively, the precedence over the photons are given as inside a jet 25% of its energy is coming from the electromagnetic candidates. Beyond $|\eta| > 2.5$, where tracking information is not available the charged and neutral hadrons are treated similarly. Each of the remaining HCAL clusters linked with one or several tracks with compatible momentum (not linked with the other HCAL cells) are treated as charged hadrons.

After reconstructing different type of particles in the detector, it is now important to form the composite objects like jets, hadronic decay of τ leptons, missing transverse energy ($p_\mathrm{T}^\mathrm{miss}$). A detailed description of jet clustering algorithm, different corrections applied to the jets are described below.

6.2   Photon reconstruction and identification

The photon candidates are reconstructed from the energy deposition in the ECAL crystals, that are not linked with the charged tracks coming from the inner most layer of the tracker. But due to the presence of the tracker material before ECAL, almost 60% of the photons start to convert into a pair of electron-positron through the pair-production mechanism. Furthermore,

the produced electron or positron experiences bremsstrahlung photon radiation. So the energy measurement of a photon from a particular single crystal deposition is not possible. Hence, the photon-electron reconstruction algorithm starts by grouping the crystals into 3×3 cluster in η-ϕ space around the most energetic one (called seed crystal). In the presence of the CMS magnetic field the trajectories of the electrons and positrons are bent leading to photon radiations spread radially over ϕ direction. To catch the corresponding energy deposits multiple ECAL clusters are needed to be combined to produce a super cluster (SC). In CMS this step is done by two different algorithm (i) mustache algorithm, which is particularly useful to properly measure the low energy deposits, and (ii) refined algorithm, it uses the tracking information of the extrapolated bremsstrahlung tangents and the conversion tracks of the electron-positron pair to match with the SC position to combine into a single candidate.

Due to the leakage of the energy through the spread of the electromagnetic shower or the loss of the radiation inside the detector material, the accumulated energy from the SC is always less compared to the original (true) energy of the electron/photon (e/γ). To correct this, a multivariate regression technique has been deployed. The target variable (y) of the regression is the ratio of the true energy of the e/γ object to the reconstructed energy; the regression provides a correction factor and an estimate of the resolution (σ_E/E) for each e/γ object. The input features of this regression technique are the different shower shape variables (the definitions of these variables along with the distributions are presented in Sec. 7.5), the position coordinates of the SC and the median energy density (ρ) of the event to take into account the PU impacts. The regression correction factor has been derived from MC simulation, which is surely not perfect and then applied on data; so the regression is not exactly the same in data as in MC simulation. The performance of the regression is shown in Fig. 6.1, by reconstructing Z boson from e^+e^- events before and after the correction. To minimize the residual discrepancies between data and MC, additional correction factor for the energy scale has been derived from the $Z \rightarrow e^+e^-$ events, and an extra smearing term is applied on the MC to match the resolution of the $Z \rightarrow e^+e^-$ spectrum with data. The derived correction factors are then validated in the $Z \rightarrow \mu^+\mu^-\gamma$ process.

After the reconstruction of the photons, there is always a finite probability where a jet can be faked as a photon. To distinguish the prompt photon from a fake photon CMS has developed two different identification (i) cut based : applying threshold values on different shower shape
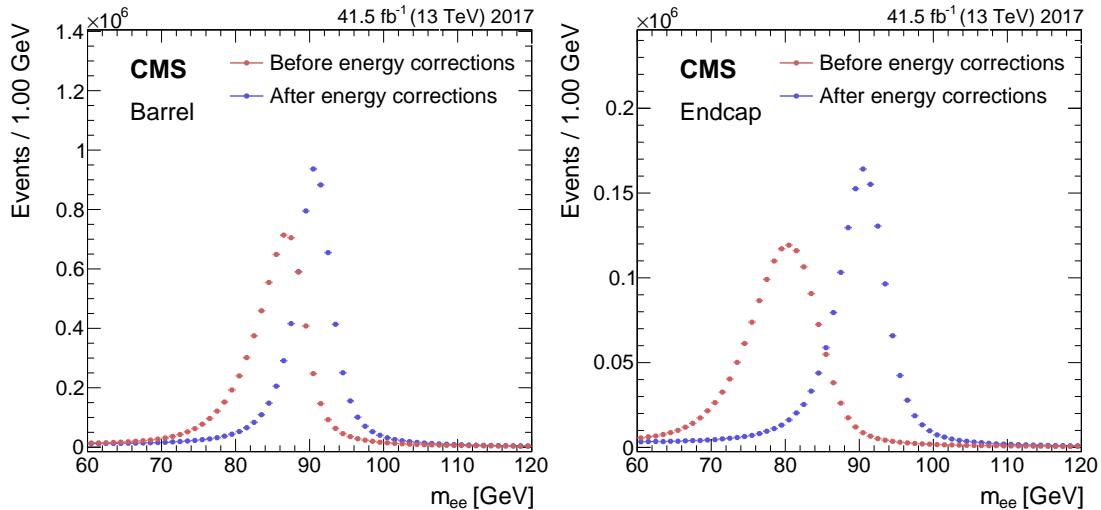
Figure 6.1: Reconstructed $Z \rightarrow e^+e^-$ invariant mass distribution before and after the energy regression correction for the barrel (left) and endcap (right). Ref. [14].

and isolation variables (ii) MVA based: an multivariate based discriminator for the prompt photon against the fake ones using the shower shape and isolation variables. The MVA based photon identification has clearly better performance compared to the cut-based identification; which is shown in Fig. 6.2. The imperfect MC simulation modeling of the input variables is corrected to match the data using another regression method based on $Z \rightarrow e^+e^-$ events. The corrections are then applied to the simulated photons such that the photon shower shape variables and the photon ID correctly model the distributions observed in data.

The clustering algorithms allow to achieve of about 95% of photon energy deposits, including the converted photon. Photon candidates are created within the ECAL and tracker acceptance region ($|\eta| < 2.5$), excluding the ECAL barrel-endcap transition region ($1.44 < |\eta| < 1.57$) as photon reconstruction in this region is not optimal.

6.3   Jet reconstruction

In a parton shower all the constituents partons are recombined in a color less state to produce hadrons in a short time and distance, such that only colorless hadrons are visible at the detector. If in the final state a well-energetic parton is produced, all the hadrons produced after the hadronization from the shower of that parton will be collimated along the direction of it, which is called jet. The jet produced from a particular parton actually carries the features of the original parton and one of the important object in the experiment.
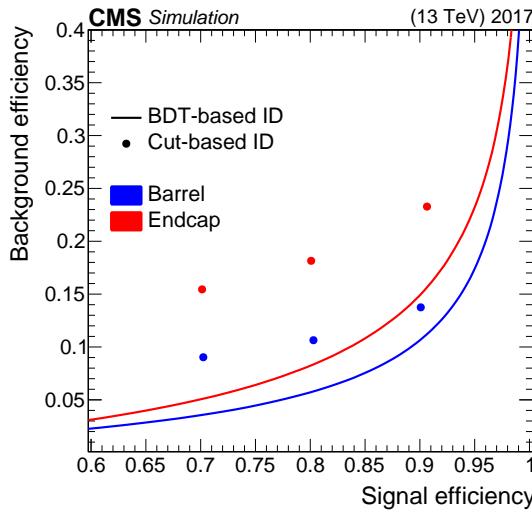
Figure 6.2: Performance of the photon identification based on cut-based and MVA based approach. The three points for the cut-based method refer to the three different working point, loose, medium and tight. Ref. [14].

A brief description of the jet reconstruction algorithm is presented below. In the hadron collider the jet clustering algorithms are mostly based on the sequential recombination technique, where the algorithm starts from the collection of all the particles and sequentially recombine them to produce a "pesdo-particle" and a new set of particle collection will be produced.

Starting from the N number of particles, one needs to compute the angular distance ΔR between each pair of particles.

$$d_{ij} = \min(p_{\mathrm{Ti}}^{2a}, p_{\mathrm{Ti}}^{2a})\frac{\Delta R_{ij}^2}{R^2} \tag{6.1}$$

Where d_{ij} is the angular distance between the two particles i and j in the rapidity-azimuthal (y-ϕ) plane, where $\Delta R = \sqrt{\Delta y^2 + \Delta \phi^2}$, R is called a radius parameter and a is a real parameter. For each particle another distance parameter d_{iB} is calculated, defined as:

$$d_{iB} = p_{\mathrm{Ti}}^{2a} \tag{6.2}$$

We shortly define the parameter a. If the minimum of all combination of d_{ij} is less than d_{iB}, then the particles i and j are merged together and a new particle k is produced with $p^i + p^j = p^k$, where p is the 4-momentum of the particle, otherwise particle i is taken out from the list of particles and considered it as a jet. The jet clustering algorithm is continued until the particle list become empty.

Different values of the above mentioned parameter a correspond to the different algorithms. There are three commonly used algorithms at the LHC.

1. a = 1 → k_T algorithm [91]: This starts clustering with two closely spaced low energetic particles. So the jet is formed through this algorithm is mostly governed by the soft radiation and may have many fake hits due to pileup.

2. a = 0 → Cambridge- Aachen (C/A) algorithm [92]: which starts the clustering depending on their spatial separation. It is commonly used to study the substructure analysis of a jet where a large radius jet is produced by the two closely energetic jets.

3. a = -1 → anti-k_T algorithm [15]: This starts to cluster from highly energetic particles, and continuing clustering the soft particles surrounded by it. The main advantage of this algorithm is that as the algorithm is governed by a highly energetic particles and if there is no other highly energetic particles within a certain distance, then the angular shape of the jet is circular.
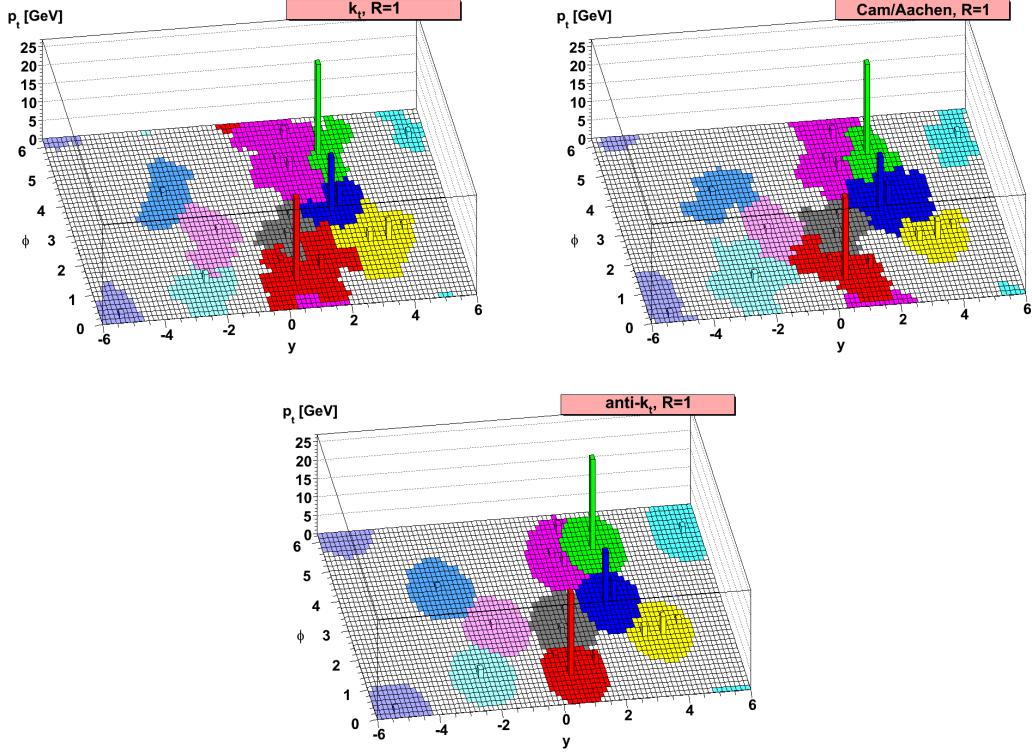
Figure 6.3: Jet clustering using different algorithms (a) k_T (top-left) (b) Cambridge-Aachen (top righ) and (c) anti-k_T (bottom). Ref. [15].

Fugure 6.3 shows the behavior of the jets reconstructed by k_T, C/A and anti-k_T algorithm clustered from the same set of particles and it is also observed that the shape of the jets using anti-k_T algorithms are circular which is easy to calibrate and easy to reduce the fraction of energy clus-

tered inside the jets coming from the pileup particles. In the LHC experiments anti-k_T jets are widely used.

For the analysis of the Run 2 data, CMS is using the anti-k_T 0.4 radius jets built from either all PF reconstructed particles, called PF jets or from the ECAL and HCAL energy depositions in the calorimeter towers, referred as Calo jets. The performance of PF jets in terms of resolution and response is twice good compared to the Calo jets, due to the precise determination of charged hadrons momenta using the track information. Figure. 6.4 shows the relative composition of different type of particles clustered inside a jet, it is found that on average 65% of jet energy carried out by charged hadrons, 25% by electromagnetic objects (mostly photons) and 10% by neutral hadrons. To measure the energy and momentum of the charged hadrons, PF algorithm relies on the highly grained tracker information along with the calorimetric energy depositions, which increase the momentum resolution of the charged hadrons and as well as the reconstructed jet itself.
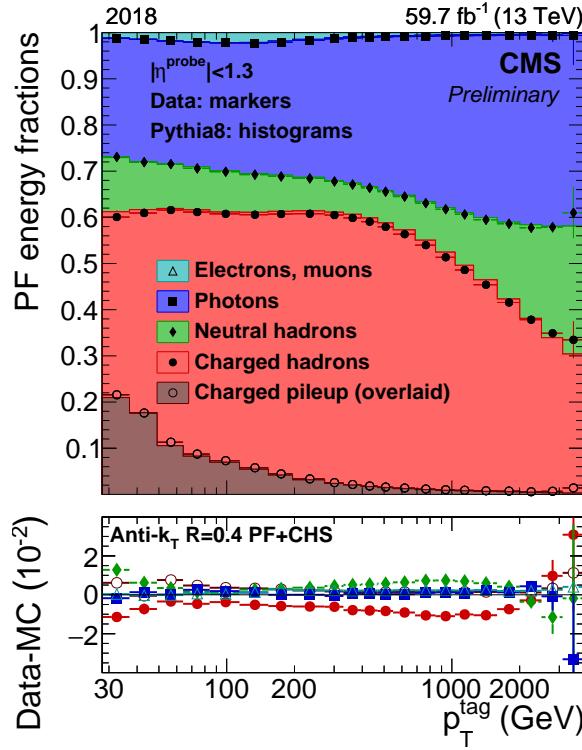
Figure 6.4: Composition of different types of particles clustered inside a jet. Lower panel shows the discripancy between the data and simulation for each type of objects.

The response of a reconstructed jet has been derived with respect to a geometrically matched generator level jet clustered from all final state stable generator level particles and shown in

Fig. 6.5 (left), where the response of the PF jet is almost stable across the η range and close to unity. Figure. 6.5(right) shows the jet energy resolution of PF and Calo jets, and at the lower p_T the resolution of PF jet is much higher than the Calo jets. But at the L1-trigger system as tracker information is not available, it uses Calo jets.
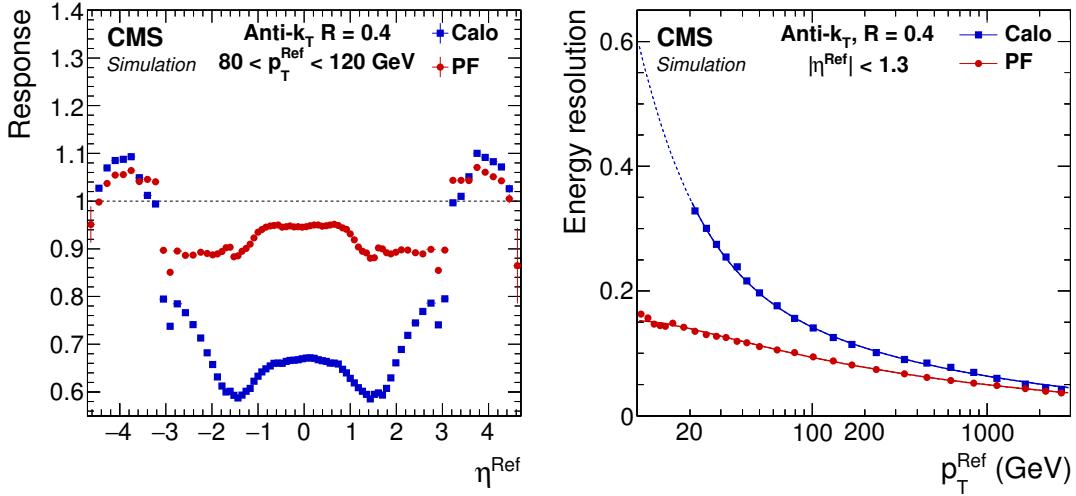
Figure 6.5: Left: Comparison of responses of PF and Calo jets with respect to the jet η. Right: The jet resolution as a function of jet p_T in the central region of the detector. The figures are taken from Ref. [16].

As the PF jets are formed with a set of particles depending on the different detector responses which are not linear. Further, there are dead materials in the detector system. Hence it is not straightforward to translate the measured energy of the PF jet to the true particle or parton energy. Consequently the jets are required to be calibrated via multi-step procedure which is described below.

6.3.1 Jet energy calibration

The jet energy corrections (JEC) are used to calibrate the energy of the jet and correct the value as much as possible to match the corresponding particle level jets. The jet constituents which are clustered from the PF objects are corrected at the very first step by subtracting the momenta of the tracks of the charged hadrons clustered inside the jets originating from the pileup vertices, the corrected jets are now called charged hadron subtracted (CHS) PF jets and this step is limited only upto $|\eta| < 2.5$ due to the tracker acceptance.

Even after the CHS correction a significant contribution from the pileup remains in the jets. The

first step of it is called L1 Pileup correction to remove the energy coming from the pileup particles clustered inside the jet. The pileup offset correction is determined from the simulation of a QCD dijet event sample with and without the pileup contribution, parameterized as a function of jet p_T and η of the jet, the energy density (ρ) of the event and jet area. The residual difference between the data and the MC is corrected using a random cone method applied in Zero-Bias data, which is collected by the CMS experiment without using any external trigger. In the random cone method, many jets are reconstructed by clustering particles in randomly placed cone, the average energy of these jets are mostly due to the detector response and the pileup effects. An uncertainty of 5% on the jet response is applied on data-simulation comparison in the random cone method.

The second step is called L2L3 MC-truth corrections, which are obtained from the simulation as a function of the p_T and η of the jet, by comparing the average energy of the reconstructed jet with a geometrically matched generator level jet energy. The correction factors are then applied on the jets both in the simulation and the data. After application of the simulation based L2L3 correction factors, the residual differences between the data and the simulation is accounted for in terms of jet energy scale (JES). JES is determined in two steps. If the jets are in the central region of $|\eta| < 1.3$, the JES is determined from the photon+jet, Z ($\to$ e$^+$e$^-$/$\mu^+$$\mu^-$)+jet events where p_T of the photons and the Z boson can be well measured. The JES at the forward regions are obtained from a QCD dijet event where two jets are expected to be balanced by each other and the leading jet is required to be at the central region. The JES is estimated for the forward jet with respect to the well callibrated central jet. The residual discrepancy of JES between data and simulation is appeared as data-to-simulation correction factor.

6.3.2 Jet energy resolution

The jet energy resolution (JER) is derived using the principle of p_T balance in dijet, photon+jet and Z +jet events, where the jet energy response has been derived with respect to a reference object i.e central jet, photon and Z boson respectively. For each jet, the response in a given p_T and η range is modelled with a Double-sided Crystal Ball function (DSCB); the width of the DSCB gives the measurement of JER. The JER in data is found to be worse than in simulation. To match the JER in simulation with data, a correction factor is added to the resolution. The data-simulation correction factor is derived for a jet in different η regions and shown in Fig. 6.6.

Figure 6.6: JER scale factor as a function of absolute value of pseudorapidity ($|\eta|$) derived from QCD dijet event topology. Ref. [17].

6.3.3 Heavy flavor jet identification

A jet containing at least one bottom or charm hadron decay products is called heavy flavor jet. Identification and classification of these heavy flavor jets from the light-flavour or gluon initiated jets (u d s g) is a very crucial task in the CMS experiment in the context of mulitple interesting physics programs, eg. for top quark related studies and for the Higgs boson candidate decaying to a pair of b-quarks (H $\rightarrow$ b$\bar{\text{b}}$). The two major physics analyses described in this thesis heavily utilized this b-jet identification scheme to construct the H $\rightarrow$ b$\bar{\text{b}}$ candidate. We refere to Sec. 7.6 and 9.5 for the details.

A bottom and charm hadron has a larger lifetime, on average about of 1.5 ps (0.45 to 1 ps). Due to the larger life time they can traverse a distance of about few mm to one cm after originating from the primary vertex before decaying at the secondary vertex (SV). So, the tracks clustered inside the jets have non-zero positive value of impact parameters (IP), the distance from the primary vertex to the point of closest approach. The mass of the b and c quarks being higher, the corresponding hadrons have larger fragmentation compared to the light quarks. Thus the decay products have a harder p_{T} spectrum compared to the udsg-jets. For almost 20% (10%) cases heavy flavor jet contains soft leptons (muons or electrons) arising the leptonic decay chain of the bottom or charm hadrons. CMS has exploited this information described above in a machine learning algorithm method to discriminate the true b and c jets against udsg jets using a simulated t$\bar{\text{t}}$ event sample in a large range of p_{T} spectrum. As the training mostly relies on the track multiplicity and other information related to the tracks, so the training is limited

Figure 6.7:   The b-jet efficiency versus light-flavored (c) jet mis-tag rate is shown in blue & red solid (dashed) lines for DEEPJET and DEEPCSV b-tagging algorithms for the jet with $p_T >$ 30 GeV in the $t\bar{t}$ MC simulated process. The figure is adapted from Ref. [18].

only for the jets upto $|\eta| < 2.5$, the acceptance provided by the tracker coverage.

The DEEPJET [18] is one of the recent versions of the b-jet classifier, which uses a large number (approximately 650) of jet variables as input through a deep neural network algorithm. The inputs used here, contain the information of all the PF objects clustered inside the jet and they can be classified into four different types (a) The global variables related to the jet level information: jet 4-momentum, number of constituent particles, primary and secondary vertices, (b) Charge particle information: charged track p_T, η, 2D and 3D impact parameters etc., (c) neutral hadron information, and (d) the variables related to the secondary vertex: p_T and mass of SV, tracks associated with it etc. It is a multiclass discriminator which can differentiate b, c, uds and gluon jets by utilizing the same set of input variables. The b-jets are trained in 3 different sub-classes (a) bb: containing two bottom hadrons, (b) b: containing one bottom hadron and decaying hadronically and (c) b_{lept}: bottom hadron decaying leptonically. The performance of the DEEPJET algorithm has been shown in Fig. 6.7, along with the performance of the previously used DNN based classifier DEEPCSV [93]. From the figure it is evident that for a particular mis-tag rate the efficiency of b-jets are highly improved using DEEPJET algorithm.

The calibration of b-tagging is performed by the classifier response in simulated events and data in the dedicated b, c and light-flavor jets controlled regions. The difference between the data and the simulation has been applied to the latter as b-tagging scale factors (SF) [94] and the uncertainties of the determination of the SF have also been propagated.

Figure 6.8: The distribution of $p_T^{\text{gen}}/p_T^{\text{reco}}$, which is used as the target variable of the b-jet energy regression. The figure is adapted from Ref. [19].

6.3.4 b-jet energy regression

In case of b-jet, there is a large probability of a bottom hadron decaying to a lepton and neutrino. Since the neutrinos escape any detection without depositing any energy, the effect leads to the underestimation of the reconstructed b-jet energy. Also for the b-quarks, due to relatively larger mass and higher soft radiation, the spread of the jets in $\eta - \phi$ plane is wide, and it leads to the leakage of energy at the jet clustering. A multivariate regression algorithm [19] based on a deep feed-forward neural network is exploited using jet composition and shape information, to correct the jet energy and resolution.

A $t\bar{t}$ simulated process has been used here, where it is easy to find a large number of genuine b-jets with a wide range of p_T. The target variable (y) in this regression is the ratio of the generator level jet p_T clustered from all the generator level particles, including neutrino, to the geomatrically matched reconstructed jet p_T ($y = p_T^{\text{gen}}/p_T^{\text{reco}}$). The shape of the target variable, y is shown in Fig. 6.8, the expected value is around 1. Longer tail to the higher values comes from the underestimation of energy of the reconstructed jets.

Some of the important feature variables used in the training are: p_T, η, mass and transverse mass of the jet, soft lepton information from the semileptonic decays of the b-hadron, secondary vertex information, energy fraction carried out by the different type of PF objects in different

annular region in $\eta - \phi$ plane, energy sharing of the jet constituents inside the jets etc. In output the regression gives a correction factor and a resolution estimator for all the jets, which corrects the jet 4-momentum. The improvement of the application of the b-jet energy regression can be utilized in the di-jet invariant mass of a resonance particle decaying to a pair of b-jets, which not only helps to shift the peak of the invariant mass distribution ($m_{b\bar{b}}$) to the true value of the resonant particle decaying, but also the resolution of the $m_{b\bar{b}}$ spectrum. This b-jet energy regression has been exploited in the two analyses of this thesis work to deal with a H $\rightarrow$ b$\bar{\text{b}}$ candidate as described in Sec. 7.6 and 9.6.

6.3.5   Quark-gluon likelihood (QGL)

To discriminate a jet originating from a quark and from a gluon is a very important task at the LHC, which plays a crucial role in many physics analysis in full-hadronic final state. In this thesis the two physics analysis, exploits the VBF topology where the two forward-backward jets are coming from quarks, and possible source of background is QCD processes, where the gluon initiated jets are predominant. CMS has performed a likelihood based discriminator to distinguish between the quark and gluon jets [20], trained on a simulated QCD multijet process and validate in data. It utilizes three physical observables with the most discriminating powers, mentioned below.

- Due to the large color factor, gluons jets has larger particle multiplicity

- The jet momentum fragmentation function (p_T^D). For the gluon jets, p_T^D is expected to be small due to the uniform sharing of the transverse momentum of the jet among its constituents.

$$p_T^D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}} \tag{6.3}$$

  where $p_{T,i}$ is the transverse momenta of the i-th jet constituent.

- A quark initiated jet is much more collimated than the gluon jets. This is measured from the minor axis (σ_2) of the elliptical shape of the jet in (η, ϕ) space, after weighting each constituent by its transverse momentum. It is found that the negative logarithm of σ_2 ($-\ln\sigma_2$) is small for a gluon jet compared to a quark jet.

The separate trainings have been performed for the jets in three different pseudorapidity regions. The distributions of the variables are shown in Fig. 6.9 in the region of $0 < |\eta| < 1.3$,

where the most distinguished features are observed due to the tracker availability and low pileup contamination. The performance of the QGL has been validated in data in two dedicated control regions (i) Z+jets (quark jet enriched), and (ii) QCD dijet (gluon enriched). Figure. 6.10 shows the QGL variable distribution in data and MC simulation, the quark and gluon jets perform separately. The QGL score has been utilized in the two analysis of this thesis mostly to distinguish VBF-tagged jets from the QCD induced multijet backgrounds, described in Secs. 7.8 and 9.7.1.

Figure 6.9: Normalized distributions of the three training variables for quark (in blue) and gluon (in red) initiated jets with $80 < p_\mathrm{T} < 100\,\mathrm{GeV}$ and inside $0 < |\eta| < 1.3$. Ref. [20].

6.4   Missing transverse energy

The particles which do not interact directly with the detector material, eg. neutrinos, escape the detector without depositing any energy on it, this may lead imbalance in momentum between the initial and the final state. The vector momentum imbalance in the transverse plane of the beam direction is known as the missing transverse energy or momentum ($p_\mathrm{T}^\mathrm{miss}$). The PF raw $p_\mathrm{T}^\mathrm{miss}$ is defined by the vector sum of the transverse momenta of all the particles reconstructed through PF algorithm.

$$\vec{p}_\mathrm{T,PF}^{\,\mathrm{miss}}(\mathrm{raw}) = - \sum_{i=1}^{N_\mathrm{particles}} \vec{p}_{\mathrm{T},i}. \tag{6.4}$$

Figure 6.10: Data and MC comparison of QGL discriminator score in Z+jets and QCD dijet events for the jets at the central region of the detector ($0 < |\eta| < 2.0$) with with $80 < p_\mathrm{T} < 100\,\mathrm{GeV}$. The composition of quarks and gluon jets are shown in blue and red solid area respectively, while data is shown by black marker. Ref. [20].

We have already discussed about the jet energy correction, which corrects the energy of the jets clustered from the particles reconstructed by the PF algorithm. So it is important to propagate this Jet energy correction (JEC) at the $p_\mathrm{T}^\mathrm{miss}$ reconstruction. The JEC included missing transverse momentum is defined as,

$$\vec{p}_\mathrm{T,PF}^\mathrm{miss} = - \sum_{i=1}^{N_\mathrm{particles}} \vec{p}_{\mathrm{T},i} - \sum_{j=1}^{N_\mathrm{PF\,jets}} (\vec{p}_{\mathrm{T},j}^\mathrm{corr} - \vec{p}_{\mathrm{T},j}), \tag{6.5}$$

which includes a term that replaces the raw momentum $\vec{p}_{\mathrm{T},j}$ of each PF jet with $\vec{p}_{\mathrm{T},j} > 10\,\mathrm{GeV}$ by its corrected value $\vec{p}_{\mathrm{T},j}^\mathrm{corr}$. This is called the Type-I correction applied on the PF $p_\mathrm{T}^\mathrm{miss}$. In this thesis very detailed study about the $p_\mathrm{T}^\mathrm{miss}$ response and resolution has been described in Sec. 10.5, which shows the importance of Type-I correction.

6.5   Pileup per particle identification

From the above discussion it can be realized that there are several techniques employed at various stages of the physics object reconstructions to deal with the pileup effect. Pileup Per Particle Identification or PUPPI [21], is another approach which tries to remove the pileup effect from the event itself, rather than just correcting the jets like CHS mechanism mentioned above. It is a method to re-scale each particle's four momenta based on their nature of origin, whether

they are coming from the hard scatter or from the pileup effects. Ideally, particles coming from the pileup should get a weight factor close to zero and one for the particles coming from the hard scatter. This leads to the correction factor. Similar to the case of particle flow algorithm, the p_T^{miss} quantity can be formed from the PUPPI particles as well. Figure. 6.11 shows the resolution of x-component of the p_T^{miss} constructed from different algorithms and it is found that the resolution of PUPPI p_T^{miss} is better compared to that from PF method. For the works described in this thesis, we have used PUPPI p_T^{miss} along with the PF p_T^{miss} definitions to estimate the missing transverse momentum in the event containing a photon and jets (γ + jets).

Figure 6.11: The resolution of x-component of p_T^{miss} in a simulated Z +jets events with average pileup 80, constructed from different p_T^{miss} reconstrcution algorithms PF (blue solid), PF with CHS correction (red dashed line) and PUPPI (in magenta). Ref. [21].

Chapter 7

Non-resonant Higgs pair production and subsequent decay to a pair of photons and b-quarks

7.1 Introduction

The major contribution of this thesis work describes a search for non-resonant production of Higgs bosons pairs through the vector boson fusion mode (VBFHH) decaying to $b\bar{b} + \gamma\gamma$ final state using a data sample of $137\,\text{fb}^{-1}$ collected by the CMS experiment from 2016 to 2018 (Run 2). The $b\bar{b}\gamma\gamma$ final state has a combined branching ratio of $2.63 \pm 0.06 \times 10^{-3}$ [26] for $M_\text{H} = 125\,\text{GeV}$, which can be derived from the individual branching ratio of the $\text{H} \to b\bar{b}$ and $\text{H} \to \gamma\gamma$ decay, $2 \times \mathcal{B}r(\text{H} \to b\bar{b}) \times \mathcal{B}r(\text{H} \to \gamma\gamma) = 2 \times 0.58 \times 0.00223$. This channel is one of the most sensitive to HH production because of the large SM branching fraction of $\text{H} \to b\bar{b}$, the good mass resolution of the $\text{H} \to \gamma\gamma$ candidate, and relatively low background rates. Finally the results are interpreted by combining with the parallel analysis utilizing the gluon-gluon fusion (ggHH) production mode, which is not a part of this thesis. An orthogonality in the event selection criteria for the two processes has been maintained properly between the VBFHH and ggHH analysis and described in the following sections whenever needed.

The details of VBFHH process has already been discussed in Sec. 3.5. The key coupling parameters which can be accessed from the VBFHH process are C_V, C_{2V} and λ as evident from the tree level diagrams presented in Fig. 3.12. Most generally, the actual values of these

couplings chosen by Nature and measured from data can be different from the prediction of SM. Hence experimental determination of these couplings is extremely important for Higgs characterization and their precision measurements would provide a possible insight into beyond SM physics. Various types of Higgs coupling are measured in terms of the coupling modifiers with respect to SM in κ-framework described in sec. 3.3.2, the relevant coupling modifiers are (a) κ_V for C_V, where V= W or Z, (b) κ_λ for λ and (c) κ_{2V} for C_{2V}. It is to be noted that κ_V is already very well constrained from single Higgs production and decay to H $\rightarrow$ WW or H $\rightarrow$ ZZ measurements at the LHC. The HH production is of prime importance for establishing Higgs trilinear self coupling and the Higgs pair coupling with a pair of vector bosons, κ_{2V}. Though ggHH is the main mode for constraining κ_λ, VBFHH provides uniquely, the access of C_{2V} due to the diagram on the Fig. 3.12 (right). As evident from the diagram on the left in Fig. 3.12 , VBFHH uniquely provides direct access of the κ_{2V} coupling modifier corresponds to the C_{2V} coupling. The main goal of the VBFHH analysis is to establish the presence of κ_{2V} mediated process and consequently set a stringent limit on it as a probe to *new physics* beyond SM. Presence of any anomalous coupling due to only *new physics* is likely to modify (mostly enhance) the production rate. The VBF process has very special event topology where the two jets are moving back to back almost along the beam line and the two H candidates are at the central part of the detector produced from the fusion of the two vector bosons radiated out from the two incoming partons. A cartoon of VBFHH event topology is shown in Fig. 7.1. It must be noted that the final state being accessed though inclusive HH + jj. Hence the ggHH process at N^{2}LO is a very important, almost irreducible background for VBFHH analysis. We shall discuss this later in detail.

Figure 7.1: A cartoon of the VBFHH type event topology where the two Higgs boson are at the central part of the detector and decay to a pair of photons and a pair of b-jets and two forward-backward jets (VBF-jets) are moving opposite direction along the beam direction.

The VBFHH cross section and the kinematics of the final state particles depend on the Higgs coupling modifiers present in the process κ_V, κ_{2V} and κ_λ. This dependency can be interpreted by six terms coming from the three LO Feynman diagrams and their interferences. Therefore, a VBFHH sample corresponding to any point in the (κ_V, κ_{2V} and κ_λ) parameter space can be obtained from the linear combination of any six MC samples with different values of κ_V, κ_{2V} and κ_λ. The mechanism of the production of VBFHH sample with arbitrary (κ_V, κ_{2V} and κ_λ) parameters is explained in the next section.

7.2   Signal sample production and morphing mechanism

Monte Carlo (MC) events for VBFHH signal are generated at the leading order (LO) accuracy using MADGRAPH5_AMC@NLO [95] matrix element generator, utilizing effective field theory calculations considering the values of κ_V, κ_{2V} and κ_λ as free parameters. The cross section for VBFHH can be described as an analytic expression from the three LO Feynman diagrams in terms of the coupling parameters as presented in Eq. 7.1, and further expansion of it gives rise the interference terms between the each pair of diagrams.

$$
\begin{aligned}
\sigma(\kappa_V, \kappa_{2V}, \kappa_\lambda) &= |\kappa_V \kappa_\lambda A + \kappa_V^2 B + \kappa_{2V} C|^2 \\
&= \kappa_V^2 \kappa_\lambda^2 A^2 + \kappa_V^4 B^2 + \kappa_{2V}^2 C^2 + \kappa_V^3 \kappa_\lambda (AB^* + A^*B) + \kappa_V^2 \kappa_{2V}(BC^* + B^*C) \\
&\quad + \kappa_V \kappa_\lambda \kappa_{2V}(CA^* + C^*A) \\
&= \kappa_V^2 \kappa_\lambda^2 a + \kappa_V^4 b + \kappa_{2V}^2 c + \kappa_V^3 \kappa_\lambda i_{ab} + \kappa_V^2 \kappa_{2V} i_{bc} + \kappa_V \kappa_\lambda \kappa_{2V} i_{ca}
\end{aligned}
\tag{7.1}
$$

Total six terms appear in the above equation, where a, b and c denote the individual square terms of the 3 diagrams, while i_{ab}, i_{bc} and i_{ca} are the interference terms among them. Equation 7.1 can also be expressed in a matrix form as,

$$
\sigma(\kappa_V, \kappa_{2V}, \kappa_\lambda) = \begin{pmatrix} \kappa_V^2 \kappa_\lambda^2 & \kappa_V^4 & \kappa_{2V}^2 & \kappa_V^3 \kappa_\lambda & \kappa_V \kappa_\lambda \kappa_{2V} & \kappa_V^2 \kappa_{2V} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ i_{ab} \\ i_{bc} \\ i_{ca} \end{pmatrix}
\tag{7.2}
$$

Hence, to evaluate the cross section and extract the event kinematics for any arbitrary set of $(\kappa_V', \kappa_{2V}', \kappa_\lambda')$, six samples with the known parameter values and the corresponding cross sections

are needed. By knowing the cross section of these six samples from MADGRAPH5_AMC@NLO and putting them as a input of the Eq. 7.2 can be expressed as:

$$
\begin{pmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_1 \\ \sigma_2 \\ \sigma_1 \\ \sigma_2 \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} \end{pmatrix} \cdot \begin{pmatrix} \kappa_V^2 \kappa_\lambda^2 \\ \kappa_V^4 \\ \kappa_{2V} \\ \kappa_V^3 \kappa_\lambda \\ \kappa_V \kappa_\lambda \kappa_{2V} \\ \kappa_V^2 \kappa_{2V} \end{pmatrix} \tag{7.3}
$$

In matrix notation:

$$
\sigma(\kappa_V, \kappa_{2V}, \kappa_\lambda) = C \, \mathbf{v}(\kappa_V, \kappa_{2V}, \kappa_\lambda) \tag{7.4}
$$

hence,

$$
\mathbf{v}(\kappa_V, \kappa_{2V}, \kappa_\lambda) = C^{-1}(\kappa_V, \kappa_{2V}, \kappa_\lambda) \, \sigma(\kappa_V, \kappa_{2V}, \kappa_\lambda) \tag{7.5}
$$

Hence we can write the cross section of the arbitrary set of $(\kappa_V', \kappa_{2V}', \kappa_\lambda')$,

$$
\sigma(\kappa_V', \kappa_{2V}', \kappa_\lambda') = C(\kappa_V', \kappa_{2V}', \kappa_\lambda') C^{-1}(\kappa_V, \kappa_{2V}, \kappa_\lambda) \, \sigma(\kappa_V, \kappa_{2V}, \kappa_\lambda) \tag{7.6}
$$

Ultimately, Eq. 7.1 can be written as:

$$
\sigma(\kappa_V', \kappa_{2V}', \kappa_\lambda') = \sum_{i=1}^{6} f_i(\kappa_V, \kappa_{2V}, \kappa_\lambda, \kappa_V', \kappa_{2V}', \kappa_\lambda') \, \sigma_i(\kappa_V, \kappa_{2V}, \kappa_\lambda) \tag{7.7}
$$

Here $f_i(\kappa_V, \kappa_{2V}, \kappa_\lambda, \kappa_V', \kappa_{2V}', \kappa_\lambda')$ are the set of functions of $\kappa_V, \kappa_{2V}, \kappa_\lambda, \kappa_V', \kappa_{2V}'$ and κ_λ' corresponding to the six known samples given as input in Eq. 7.3 and they are orthogonal to each other. The explicit form of the functions f_i will depend on the input values of $\kappa_V, \kappa_{2V}, \kappa_\lambda$ and σ_i. Finally Eq. 7.7 is used to determine the cross section for any set of parameters.

7.2.1   Validation of the VBFHH sampling mechanism

Seven samples (s_i) have been generated from MADGRAPH5_AMC@NLO with a cross section (σ_i) corresponds to each one. Six of them are used here in VBFHH sample parametrization mechanism and one for validation. The VBFHH produced samples with the values of the $\kappa_V, \kappa_{2V}, \kappa_\lambda$ parameters along with their LO cross section are presented in Table 7.1. It is to be noted that the cross section for $\kappa_{2V} = 0$ is almost 20 times larger than the SM value. Since we want to test the

Table 7.1: Samples from Madgraph generator with specified values of the coupling modifiers and the corresponding cross sections.

No.	$\kappa_V, \kappa_{2V}, \kappa_\lambda$	σ [fb]
1	1,1,1	1.67
2	1,2,1	13.74
3	1,1,2	1.38
4	1,1,0	4.45
5	0.5,1,1	10.46
6	1.5,1,1	63.80
7	1,0,1	26.17

hypothesis of $\kappa_{2V} = 0$, we have designed the detailed analysis using this sample as explained later. As expected, some of the kinematic distributions for $\kappa_{2V} = 0$ sample are different than those from SM sample. This aspect has been used to optimize the analysis strategy.

It has been checked and as expected from the definitions that the invariant masses of the decay products of the individual H candidates, $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ are not distorted at the time of reconstruction of VBFHH samples with arbitrary values of coupling parameters. This ensures the stability of the morphing mechanism to reconstruct VBFHH sample with any set $\kappa_V, \kappa_{2V}, \kappa_\lambda$. Figure 7.2 displays the invariant mass distributions for VBFHH signals having different κ_{2V} (left) and κ_λ(right). The variation of the VBFHH cross section in 2-dimensional plane of κ_{2V}-κ_λ and κ_{2V}-κ_V are presented in Fig. 7.3.

7.3   Collision dataset and simulated samples

The analyzed data volume corresponds to a total integrated luminosity of 137 fb^{-1} collected in the LHC Run 2. Events are first selected with the first tier of hardware based trigger system, Level-1 (L1), requiring presence of two e/γ objects in the event with transverse energy thresholds of 23(25) & 10(14) GeV for the leading and subleading objects in 2016 (2017 & 2018). Due to the unavailability of tracker information at the L1 trigger system both electron and photon are treated as the same object by relying only the amount of energy deposition in electromagnetic calorimeter.

The background processes relevant for this analysis are of two types: (a) single Higgs produc-

Figure 7.2: Upper row: $m_{\gamma\gamma}$ (left) and $m_{b\bar{b}}$ (right) distributions for variations in κ_{2V}, keeping κ_{λ} and κ_V as according to SM. Lower row: $m_{\gamma\gamma}$(left) and $m_{b\bar{b}}$(right) distributions with the variations of κ_{λ}, keeping κ_{2V} and κ_V as according to SM. All are area normalized distributions.

tion and decaying to a pair of photons, leading to a resonant peak of H $\rightarrow \gamma\gamma$ candidate in $m_{\gamma\gamma}$ spectrum; so these processes are called resonant background, (b) non-resonant backgrounds which have a continuum $m_{\gamma\gamma}$ distribution. The main non-resonant background contribution comes from the QCD induced processes with well-isolated photons coming from the hard scatter (prompt photons) or multijet processes where jets are misidentified as photons (fake photons). Although the probability for a jet to be faked as a photon is very low, but due to the large production cross section of QCD processes, the contamination of this type of background is not negligible. These non-resonant backgrounds are classified according to the number of prompt and fake photons in the selected diphoton candidates.

Figure 7.3: The cross section of VBFHH process for the 2-dimensional variation in κ_{2V}-κ_λ (left) and κ_{2V}-κ_V (right) plane. The colors shades represent the cross sections in fb.

The prompt-prompt background ($\gamma\gamma$ + jets) events are simulated with the SHERPA v.2.2.1 [96] at LO; it includes the born processes with additional jets at LO accuracy. To increase the statistics of events, additional samples forcing the additional jets to originate from b-quarks ($\gamma\gamma$ + b-jets) have been generated. The prompt-fake (γ + jets) and the fake-fake contributions are simulated with PYTHIA 8.212 [97], with a filter to get a neutral hadrons(π^0, η) with large probability of jet faking as photon.

Although the VBFHH event topology differ from ggHH at LO, due to presence of two additional light quark jets, but at N^2LO due to the additional gluon radiations ggHH can mimic the VBFHH signals and almost 30% of the ggHH events are migrated to VBFHH selection. The ggHH signal samples are simulated at NLO [98–102] including the full top quark mass dependence [103] using POWHEG v-2.0.

For resonant background, we consider all possible Higgs production mechanisms: gluon-gluon fusion (ggH), associated production with a vector boson (VH), vector-boson fusion (VBFH), associated production with top quarks ($t\bar{t}$H) and associate production of top and a light quark (tHq), which are produced by POWHEG v2.0.

All simulated event at hard scattering level, the samples are interfaced with PYTHIA 8.212 for parton showering and fragmentation with the standard p_T-ordered parton shower (PS) scheme. The underlying event is modeled with PYTHIA 8.212, tuned according to data [104,105]; CUETP8M1 for 2016 and the CP5 tune for 2017 and 2018. PDFs are taken from the NNPDF3.0 [106] NLO (2016) or NNPDF3.1 [107] NNLO (2017 and 2018) set for all simulated samples except for the signal simulated at LO, for which the PDF4LHC15_NLO_MC set at NLO [106, 108–111]

is used. The response of the CMS detector is modeled using the GEANT4 [112] package. The simulated events include additional p-p interactions within the same or nearby bunch crossings (pileup), as observed in the data.

7.4 Analysis workflow

The analysis will follow the flowchart shown in Fig. 7.4. The detailed description of each step has been described in the following sections.

Figure 7.4: A cartoon of analysis flowchart describing the each step.

7.5 High level trigger (HLT)

7.5.1 HLT requirements

The events which are passed the double e/γ L1 trigger discussed above, are needed to pass the second tier of selection implemented with software based High Level Trigger (HLT) paths. The leading photon is required to have $E_T > 30$ GeV and for subleading photon $E_T > 18$ (22) GeV for 2016 (2017 and 2018). To control the rate of the events passing HLT path and to increase the purity of the photons another set of additional requirements have been imposed on the photon candidates isolation and identification, which are tabulated in Tab. 7.2. The requirements on the ECAL barrel (EB) and ECAL endcap (EE) photons are slightly different due to the different noise level in the EB and EE crystals.

Table 7.2: The quality requirements applied on the photon candidates at the HLT level.

	H/E	$\sigma_{i\eta i\eta}$	R_9	Iso_{ph} (lead)	Iso_{ch}
EB	< 0.12	–	> 0.85	–	–
EB	< 0.12	< 0.015	< 0.85	< (6.0 + 0.012 E_T)	< (6.0 + 0.002 E_T)
EE	< 0.1	–	> 0.9	–	–
EE	< 0.1	< 0.035	< 0.9	< (6.0 + 0.012 E_T)	< (6.0 + 0.002 E_T)
Other trigger requirements					
Lead photon $E_T > 30$ GeV	Sublead photon $E_T > 18$ GeV in 2016, Sublead photon $E_T > 22$ GeV in 2017-18			$m_{\gamma\gamma} > 90$ GeV	

The definitions of the variables used in the trigger algorithms are given below (detailed definitions are in Ref. [8, 14]):

1. Hadronic over electromagnetic energy ratio (H/E): The H/E ratio is defined as the energy deposition (H) on the HCAL tower within a cone of $\Delta R = 0.15$ around the super cluster of the photon or electron candidate at the ECAL (E). For good quality of photons the value of H/E is expected to be small.

2. $\sigma_{i\eta i\eta}$: The second moment of the log-weighted distribution of crystal energies in η, calculated in 5×5 matrix around the most energetic crystal inside a super cluster. The

definition of $\sigma_{i\eta i\eta}$ is given by:

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5\times5} w_i(\eta_i - \bar{\eta}_{5\times5})^2}{\sum_i^{5\times5} w_i}} \tag{7.8}$$

Where η_i is the position of the i-th crystal, and $\bar{\eta}_{5\times5}$ is the mean position of the 5×5 cluster. The weight factor $w_i = \max(0, 4.7 + \ln(E_i/E_{5\times5}))$, is nonzero when $\ln(E_i/E_{5\times5}) > -4.7$, which corresponds to the $E_i > 0.9\%$ of the total energy of 5×5 cluster. This ensures that only the crystals above a noise thresholds are included in this variables. The distributions of $\sigma_{i\eta i\eta}$ of a prompt photon or electrons are narrow compared to the fake photons coming from jets.

3. R_9: The sum of the energy deposition of the 3×3 crystals centered on the most energetic seed crystal in the supercluster divided by the total energy deposition on that supercluster. The shape of R_9 distribution for the unconverted photons has high value close to unity.

4. Iso_{ph} and Iso_{ch}: These two are the isolation variables obtained by summing the p_T of the electromagnetic candidates and charged hadrons around a cone of radius 0.3 of a photon. The larger energy of the incoming photon has the larger energy deposition around its direction, for this reason the isolation variables often used as a function of E_T and called relative isolation.

The HLT requirements on the photon's shower shape and isolation variables, described above, have been applied to pick a good quality identified and isolated photon from the background sources. For a prompt photon, the genuine source of backgrounds are jets fragmenting into π^0 or η, which subsequently decays to two closely spaced non-isolated photons. The distributions of these shower shape and isolation variables are shown in Figs. 7.5 (for EB) and 7.6 (for EE), where a reconstructed photon matched with a generator level hard scatter photon within a cone of 0.13 is referred to as prompt photon while other unmatched photons are mostly coming from the jets faked as photons in a simulated γ + jets sample.

7.5.2   Trigger performance

The efficiency for an event with 2 good photons to pass the trigger criteria has been derived only from the data, and the same efficiency factor has been used for MC sample as trigger scale

Figure 7.5: The distributions of the shower shape variables like H/E, $\sigma_{i\eta i\eta}$ (upper row) and two isolation variables (lower row) for a prompt photon (in red line) and a fake photon (in blue area) at the ECAL barrel.

factor (SF)[1], since for the MC, the HLT conditions are not well modeled.

To get rid of any bias in the trigger study the efficiency measurements are carried out by tag-and-probe technique on an event targeting $Z \rightarrow e^+e^-$ events, employing the same selections on the shower shape and the calorimetric isolation criteria mentioned above. The efficiency for the required diphoton trigger is computed separately for the leading and the subleading photons. The final efficiency is the product of these two which are directly applied to the simulation at the time of event selection. We define

$$\text{Efficiency} = \frac{\text{Number of probes matched with HLT requirements}}{\text{Total number of probes satisfying kinematic conditions}} \qquad (7.9)$$

In the tag-and-probe technique general idea is to impose very stringent or tight requirements

---

[1]this same scale factor commonly used for all analysis having a SM H of $M_H \sim 125$ GeV decays to a pair of photons

Figure 7.6: The distributions of the shower shape variables like H/E, $\sigma_{i\eta i\eta}$ (upper row) and two isolation variables (lower row) for a prompt photon (in red line) and a fake photon (in blue area) at the ECAL endcap.

on the the 'tag' object (in this case electron) and comparatively looser selections on the 'probe' object. The tagged electron is required to have $p_\text{T} > 40$ GeV, within $|\eta| < 2.1$. For the leading object of the diphoton trigger, probe electron must have $|\eta| < 2.5$, $p_\text{T} > 30$ GeV. The collection of probes passing this selection becomes the denominator of the leading photon efficiency, while the numerator is the the number of probes satisfying the HLT conditions mentioned in Tab. 7.2, the same prescription has also been followed for the subleading object of the HLT trigger just by changing the $p_\text{T} > 18$ (22) GeV for 2016 (2017 & 2018).

The electrons have different shower shapes in ECAL with respect to the photons, due to their interactions with the material upstream of the calorimeter, and, coming from a Z decay (spin 1 boson), they have also a different η distribution with respect to photons coming from a Higgs decay (spin 0 boson), the entries to the efficiencies are weighted in R_9 and η from the respective simulated samples (H $\rightarrow \gamma\gamma$ at M_H 125 GeV and Z $\rightarrow$ e$^+$e$^-$ processes), to match the H $\rightarrow \gamma\gamma$

distributions. The re-weighting factors in (R_9, η) plane for the year of 2018 has been shown in Fig. 7.7.

Figure 7.7: Re-weighting factors in (R_9, η) for $Z \rightarrow e^+e^-$ selected events with respect to $H \rightarrow \gamma\gamma$ events.

The HLT efficiencies have been derived for both leading and subleading photons with respect to different p_T, R_9 regions separately in two η regions ECAL barrel (EB) and ECAL endcap (EE) directly from data. The efficiencies for 2018 leading and subleading photons are shown in Fig. 7.8.

Figure 7.8: Diphoton trigger efficiency measured on 2018 data for $Z \rightarrow ee$ events using tag-and-probe method. Efficiency with respect to offline probe E_T, shown in the categories defined according to probe R_9 and $|\eta|$.

7.5.3   Determination of uncertainties in trigger efficiency

In data, the selection of tag-and-probe method accumulates a certain amount of background processes along with a $Z \rightarrow e^+e^-$ data. In each bin (p_T, R_9, η) the contributions of $Z \rightarrow e^+e^-$ has been determined by fitting the $m_{e^+e^-}$ distribution by a signal shape which is a function of Breit-Wigner [113] convoluted with a crystal ball (CB) function [114] along with a background PDF. Figure. 7.9 shows the fit of the signal and the background in one bin where "passing probe" refers to them who pass the HLT requirements, and "failing probe" refers to the probes who could not. Two different choice of background shape functions yields a difference of $Z \rightarrow e^+e^-$ event estimation. This difference is propagated as a systematic unceratinties along with the statistical uncertainties of $Z \rightarrow e^+e^-$ events in each bins. The average systematic uncertainties for the trigger efficiency measurements found to be 1-2%.

Figure 7.9: The fit of the m_{ee} distributions of the electrons selected by tag-and-probe technique for the both two cases where probes pass/fail the HLT selection. The black points are the selected data, red solid line is the signal shape, while blue solid line describes the background contributions.

7.6 Offline event selection

7.6.1 Higgs candidates selection

In order to select the events at least two well isolated photon candidates are required to recon-struct the H decaying to a pair of photons (H → γγ). To reduce any kind of possible distortion on the $m_{\gamma\gamma}$ distribution, no direct p_{T} thresholds are applied for the photon candidates. Two well isolated photon candidates inside $|\eta| < 2.5$ has been selected along with some additional requirements:

- $100 < m_{\gamma\gamma} < 180$ GeV.

- $p_{\mathrm{T}}^{\gamma 1}/m_{\gamma\gamma}$ ($p_{\mathrm{T}}^{\gamma 2}/m_{\gamma\gamma}$) > 1/3(1/4)

Where, $p_{\mathrm{T}}^{\gamma 1}$ and $p_{\mathrm{T}}^{\gamma 2}$ are the transverse momentum of the two selected photons for H → γγ candidate. The transition region between the EB and EE ($1.446 < |\eta| < 1.57$) has been excluded from the photon selection due to poor reconstruction efficiency. If there is more than one pair of photons satisfying the above selection criteria, then the pair with maximum diphoton transverse momentum ($p_{\mathrm{T}}^{\gamma\gamma}$) is selected as H → γγ candidate.

The H candidate which is decaying to a pair of b-quarks is selected from a collection of jets with $p_{\mathrm{T}} > 25$ GeV, $|\eta| < 2.4$ (2.5) inside the tracker acceptance region during 2016 (2017 & 2018) data taking. To remove the overlap between the photons and jets an angular seperation is required $\Delta R_{j,\gamma} > 0.4$, where $\Delta R_{j,\gamma} = \sqrt{(\eta_j - \eta_\gamma)^2 + (\phi_j - \phi_\gamma)^2}$. In addition jets originating from the hadronization of the b-quarks are distinguished from the c or light quark jets based on a Deep Neural Network (DNN) algorithm, called DEEPJET utilizing the PF level information of all constituents clustered inside the jet along with the information of the secondary vertex. The two jets with maximum sum of DEEPJET b-tag scores are selected as H → b$\bar{\text{b}}$ candidate with an invariant mass range of $70 < m_{b\bar{b}} < 190$ GeV. A two step b-jet energy regression has been applied. In the first step a DNN based b-jet energy regression is applied to correct energy and resolution of the each individual jets, which is described in Sec. 6.3.4, which also improves the $m_{b\bar{b}}$ spectrum as well. Additionally, an regression technique has been developed and applied specifically for the b$\bar{\text{b}}$γγ final states to further improve the $m_{b\bar{b}}$ resolution to have an overall improvement by about 20%. Further the $m_{b\bar{b}}$ peak position is shifted by 5.5 GeV (5%) closer to the expected Higgs boson mass.

A summary of the preselection criteria for jets and photons can be found in Tab. 7.3.

Table 7.3: Summary of the baseline selection criteria.

Photons		b-jets	
Variable	Selection	Variable	Selection
$p_\mathrm{T}^{\gamma 1} [GeV]$	$> m_{\gamma\gamma}/3$	p_T [GeV]	> 25
$p_\mathrm{T}^{\gamma 2} [GeV]$	$> m_{\gamma\gamma}/4$	$\Delta R_{\gamma j}$	> 0.4
$|\eta|$	< 2.5	$|\eta|$	< 2.4
$m_{\gamma\gamma}$ [GeV]	$[100, 180]$	$m_{b\bar{b}}$ [GeV]	$[70, 190]$
		DEEPJET score	> 0

7.6.2 VBF jet selection

Keeping the topology of the VBF type of events in mind, two additional jets originating from the outgoing partons scattered away are selected. The conditions imposed on the light-quark anti-k_T jets are $p_\mathrm{T}^{VBF} > 40$ and 30 GeV for the leading and subleading ones and $|\eta| < 4.7$. As these VBF-tagged jets are mostly populated in the forward-backward regions of the detector and without any supplementary tracker information, there is a large probability of pileup jets (jets clustered from the pileup particles) to be faked as VBF jets. A dedicated pileup jet identification (PUID) criterion has been applied on per-jet basis to discriminate the jets coming from pileup and the jets from the hard-scatter processes. The PUID is a Boosted Decision Tree (BDT) based identification algorithm trained on true pilup and hard-scatter jets based on the jet constituents multiplicity, shape distribution and charged particle track information (only upto $|\eta| < 2.5$).

7.6.3 Event yields

After the HLT and offline selection the event yields coming from the signal and from different contributing backgrond process estimated from MC simulations are presented in Tab. 7.4 along with the collision data obtained from each individual year of data taking periods. As expected at this stage and evident from Tab. 7.4, the selected events are completely dominated by the non-resonant background. The $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distribution in full Run2 data and the resonant single H background processes after the VBF like preselection is presented in Fig. 7.10 left and right respectively, where the VBFHH signal is magnified by a factor of 10^4 for the purpose of

Table 7.4: Expected event yields of signal, different background processes estimated from simulations and data events in each of the Run 2 data taking year.

Process	2016 (35.9 fb^{-1})	2017 (41.5 fb^{-1})	2018 (59.7 fb^{-1})
VBFHH	0.02	0.03	0.04
ggHH	0.28	0.32	0.54
ggH	21.62	23.14	30.79
VBFH	1.97	2.48	3.53
VH	4.63	5.85	7.9
t$\bar{\text{t}}$H	15.53	17.93	26.24
tHq	1.11	1.54	2.23
γ+jets	1278.33	1357.35	2626.13
$\gamma\gamma$ + jets	28154.18	36855.65	53239.13
Data	29725	38213	55936

display. This again underlines the huge contribution of non-resonant background after simple kinematic selection.

Figure 7.10: The distributions of $m_{\gamma\gamma}$ (left) and $m_{b\bar{b}}$ (right) in full Run2 data superposed with expected contribution from MC samples of resonant single Higgs processes after preselection. For ease of visual inspection, SM VBFHH signal has been scaled up by a factor of 10^4. The data is mostly dominated by non-resonant $\gamma\gamma$+jets and γ+jets backgrounds.

7.6.4   Kinematic variation of VBFHH due to coupling parameters

Just like the cross section, the kinematics of the VBFHH process also depends on the coupling parameters. The reconstructed object from the decay products of the H, have different kinematic properties and efficiency factors, as a result, the four body invariant mass of the four H decay objects (two photons and two b-jets), $m_{\gamma\gamma bb}$ is the best variable to probe the difference of kinematics between the different VBFHH samples. Thus $m_{\gamma\gamma bb}$ is the invariant mass of the two H candidate. Higgs pair system which is particularly sensitive to different values of the couplings. Due to the reconstruction issue of the different objects the $m_{\gamma\gamma bb}$ can be distorted from its true value. A new variable, $\widetilde{M}_X$ is defined from the reconstructed $m_{\gamma\gamma bb}$ distribution by subtracting each reconstructed čandidates mass, $m_{\gamma\gamma}$ & $m_{b\bar{b}}$ and adding the true H mass (125 GeV) [115]. The definition of $\widetilde{M}_X$ is:

$$\widetilde{M}_X = m_{\gamma\gamma bb} - (m_{b\bar{b}} - M_H) - (m_{\gamma\gamma} - M_H), \tag{7.10}$$

In Fig. 7.11, the distribution of $\widetilde{M}_X$ variable is shown for different values of κ_{2V} appearing in the VBFHH process. The signals with $\kappa_{2V} = 0$ and $\kappa_{2V} = 2$ have a much harder spectrum than the SM VBF HH signal which also ensure, that the acceptance for the VBFHH processes with these anomalous values of coupling parameters has larger acceptance than the SM one.

Figure 7.11: The shape normalized distributions of $\widetilde{M}_X$ variable of VBFHH processes due to different values of coupling parameter κ_{2V}. The SM ($\kappa_{2V} = 1$) production is shown by red solid line, while $\kappa_{2V} = 0$ and $\kappa_{2V} = 2$ are described in orange and green dashed lines respectively.

7.7   ttH background rejection

Events produced from the single Higgs production processes constitute a significant component of the background around $M_\text{H} = 125\,\text{GeV}$. This motivates the development of a dedicated discriminator to reject these background events. Among the various single-H processes $t\bar{t}H$, with $H \rightarrow \gamma\gamma$ is the most dominant background in the most pure signal region, as the two genuine b-jets, two photons and additional jets can come in the final state of $t\bar{t}H$ process, which makes it easy to mimic the signal topology. To discriminate it from HH processes a dedicated DNN training, refereed to as ttHKiller, has been performed.

The basic feature variables for the ttHKiller training has been classified into class (a) Low level: The four momentum of all the PF reconstructed objects (b) High level kinematic variables: describing the kinematics of the event topology.

Variables to reject events with a leptonic-decay of the W boson

- Number of jets with $p_\text{T} > 15$ GeV and $|\eta| < 2.5$

- Scalar sum S_T of the transverse momentum if all jets mentioned above.

$$S_T = \sum_j |\vec{p}_T^{\,j}| \tag{7.11}$$

- In the leptonic decay of the W boson produced in the decay of top quark for $t\bar{t}H$ process, a neutrino is always produced along with a lepton, which generates a substantial amount of p_T^miss.

- $\Delta\phi(b, p_\text{T}^\text{miss})$: The azimuthal angle separations between the p_T^miss and the two b-tagged jets

- 4-momentum of the electron and muon: Leptons are expected in the semileptonic decay of the W boson for $t\bar{t}H$ process.

- The 4-momenta of the selected di-b jets and the diphoton pair candidates.

Variables to reject events with a hadronic-decay of the W boson

- A quantity χ_t^2, based on the kinematic properties of selected b-jets, which is defined as:

$$\chi_t^2 = \left(\frac{M_W - m_{jj}}{0.1 \times M_W}\right)^2 + \left(\frac{m_t - m_{bjj}}{0.1 \times m_t}\right)^2 \tag{7.12}$$

Where M_W and m_t are the true masses of the W and t. The χ_t^2 is expected to be zero for the hadronic decay of W from the t-quark in $t\bar{t}H$ process. Here "j" refers to as the

additional jet other than the selected b-jets, so this variable is only possible to construct for the case where atleast two additional jets are available.

The distribution of the kinematic variables used in the ttHKiller training are shown in Fig. 7.12, 7.13 and 7.14. The transverse momenta of the leading and subleading photons (b-jets) are denoted as: $p_T^{\gamma 1}$, $p_T^{\gamma 2}$ (p_T^{b1}, p_T^{b2}) while the transverse momentum of diphoton (di-b-jets) candidate is $p_T^{\gamma\gamma}$ (p_T^{bb}).

Figure 7.12: The distributions of variables used in the ttHKiller , where the VBFHH, ggHH & ttH̄ → γγ processes are shown in red solid line, blue dashed line & red green filled respectively.

Figure 7.13: The distributions of variables used in the ttHKiller , where the VBFHH, ggHH & t̄tH processes are shown in red solid line, blue dashed line & red green filled respectively.

Figure 7.14: The distributions of variables used in the ttHKiller , where the VBFHH, ggHH & t̄tH processes are shown in red solid line, blue dashed line & red green filled respectively.

7.7.1 Performance

The distribution of the ttHKiller output score is shown in Fig. 7.15, which displays the clear discrimination of the $t\bar{t}$H background against both the HH processes.

Figure 7.15: The ttHKiller output score distribution of the VBFHH, ggHH & $t\bar{t}$H processes are shown in red solid line, blue dashed line & red green filled respectively.

The threshold value of the ttHKiller score is found to be at 0.26, which is optimized by keeping the background rejection at 85% corresponding to the VBFHH signal selection efficiency at 90%. For this the fractional rejection of the other backgrounds is presented in Tab. 7.5, which indicates that the use of ttHKiller is not an optimum option for the reduction of the other background contributions.

Table 7.5: The rejection fraction of the event yields (in %) for the different contributing background processes after application of ttHKiller threshold value.

ggH	VBFH	VH	$t\bar{t}$H	tHq	γ + jets	$\gamma\gamma$ + jets
16	10	35	83	38	16	30

7.8   Non-resonant background rejection

From the Fig. 7.10, it is evident that the contribution of the most dominating $\gamma\gamma$ + jets/γ + jets background is overwhelmed on the SM VBFHH process by a factor of almost 10^6 in the very narrow region of $m_{\gamma\gamma}$ of our interact i.e [115-135 GeV]. On the other hand, a substantial amount of ggHH events migrated to the VBFHH phase space, which is treated as a background for the κ_{2V} coupling determination. A multiclass XGBoost training is performed separately for 2 regions: for $\widetilde{M}_X > 500$ GeV and $\widetilde{M}_X < 500$ GeV. For each case, a mixture of Monte Carlo samples for VBFHH SM and $\kappa_{2V} = 0$ has been used as signal. Both the samples have been normalized to total weights for giving same importance in training. There are 2 background classes: (i) ggHH production and (ii) the non-resonant processes: exclusive $\gamma\gamma$ + jets and inclusive γ + jets. Separate trainings have been performed for three data taking years of LHC to take into account variations in detector configuration, eg., in tracker, as well as variations in machine environment, eg., pileup.

The total set of input variables can be categorized into two groups (a) kinematic, angular and identification variables related to the photons and b-jets and (b) kinematic variables for the VBF-tagged jets describing the typical VBF topology. The second set of variables has an important impact to distinguish between the VBFHH and ggHH processes. The detailed description of the variables are listed below:

Photons and b-jets related variables

- $p_T^{\gamma 1}/m_{\gamma\gamma}$ & $p_T^{\gamma 2}/m_{\gamma\gamma}$: The transverse momentum of the photons divided by the invariant mass of the H → $\gamma\gamma$ candidate.

- $p_T^{b1}/m_{b\bar{b}}$ & $p_T^{b2}/m_{b\bar{b}}$: The transverse momentum of the two b-jets scaled by the invariant mass of the H → $b\bar{b}$ candidate.

- $p_T^{\gamma\gamma}/m_{\gamma\gamma bb}$ & $p_T^{bb}/m_{\gamma\gamma bb}$: The transverse momenta of the diphoton and di b-jet candidate scaled with the H H invariant mass.

- $|\cos(\theta_{CS}^*)|$: θ_{CS}^* is defined as the Colin-Sopper angle between the Higgs boson decaying to diphoton pair (H → $\gamma\gamma$) with the average beam direction in the HH centre-of-mass system.

- $|\cos(\theta_{\gamma\gamma}|$ and $|\cos(\theta_{bb})|$: These are the angles between one of the H decay products chosen arbitrarily (either photons or b-jets) with respect to the direction of the H boson candi-

date.

- Miminum $\Delta R_{b\gamma}$: minimum angular distance between the selected photons and b-jets.

- Another $\Delta R_{b\gamma}$: The angular distance between the other photon and b-jet different from miminum $\Delta R_{b\gamma}$.

- Photon MVA ID: The identification scores for the leading and subleading photons. The several shower shape and isolation variables are used in identification training to discriminate prompt photons from the fake photons.

- DEEPJET score: The b-tagging score of the leading and subleading b-tagged jets.

- Resolution of the b-jets: jet energy resolution for leading and sub-leading jets ($\sigma^b_{p_T}/p^b_T$) are obtained from output of the b-jet energy regression.

- Photon resolution variables: the energy resolution normalized to the energy, for the leading and subleading photon (σ_E/E), and the diphoton mass resolutions for the two selected photons ($\sigma_{m_{\gamma\gamma}}/m_{\gamma\gamma}$)

VBF-tagged jet related variables

- Leading VBF-jet and subleading VBF-jet transverse momentum over VBF-jet invariant mass, $p^{VBF}_{T1,T2}/m^{VBF}_{jj}$.

- Pseudo-rapidity of leading and subleading VBF-jet, $\eta^{VBF}_{1,2}$.

- Product and difference of η^{VBF} between two VBF-jets: $\eta^{VBF}_1 \times \eta^{VBF}_2$ and $\Delta\eta = \eta^{VBF}_1 - \eta^{VBF}_2$.

- Quark-gluon likelihood (QGL) score of two VBF-jets. The details of the QGL is discussed in Sec. 6.3.5.

- Minimum angular distance between one VBF-jet and one photon, $\Delta R_{\gamma j}$.

- Minimum angular distance between one VBF-jet and one b-jet, ΔR_{bj}

- Invariant mass of two VBF-jets (m^{VBF}_{jj}).

- Centrality variables for di-photon and di-bjet systems, representing the relative positions of the two Higgs candidates with respect to the VBF jets, $C_{\gamma\gamma}$ and C_{bb} respectively. The

centrality variable is defined as,

$$C_{xx} = \exp.\left[-\frac{4}{(\eta_1^{VBF} - \eta_2^{VBF})^2}\left(\eta_{xx} - \frac{\eta_1^{VBF} + \eta_2^{VBF}}{2}\right)^2\right]$$

where x is γ or b, and η_1 and η_2 are the pseudo-rapidities of leading and subleading VBF-jets.

The distribution of variables mentioned above are shown in Fig. 7.16, 7.17, 7.18, 7.19 and 7.20. The variables used in the MVA have very minimal correlation with the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distribution, as those two are the final search variables for this analysis.
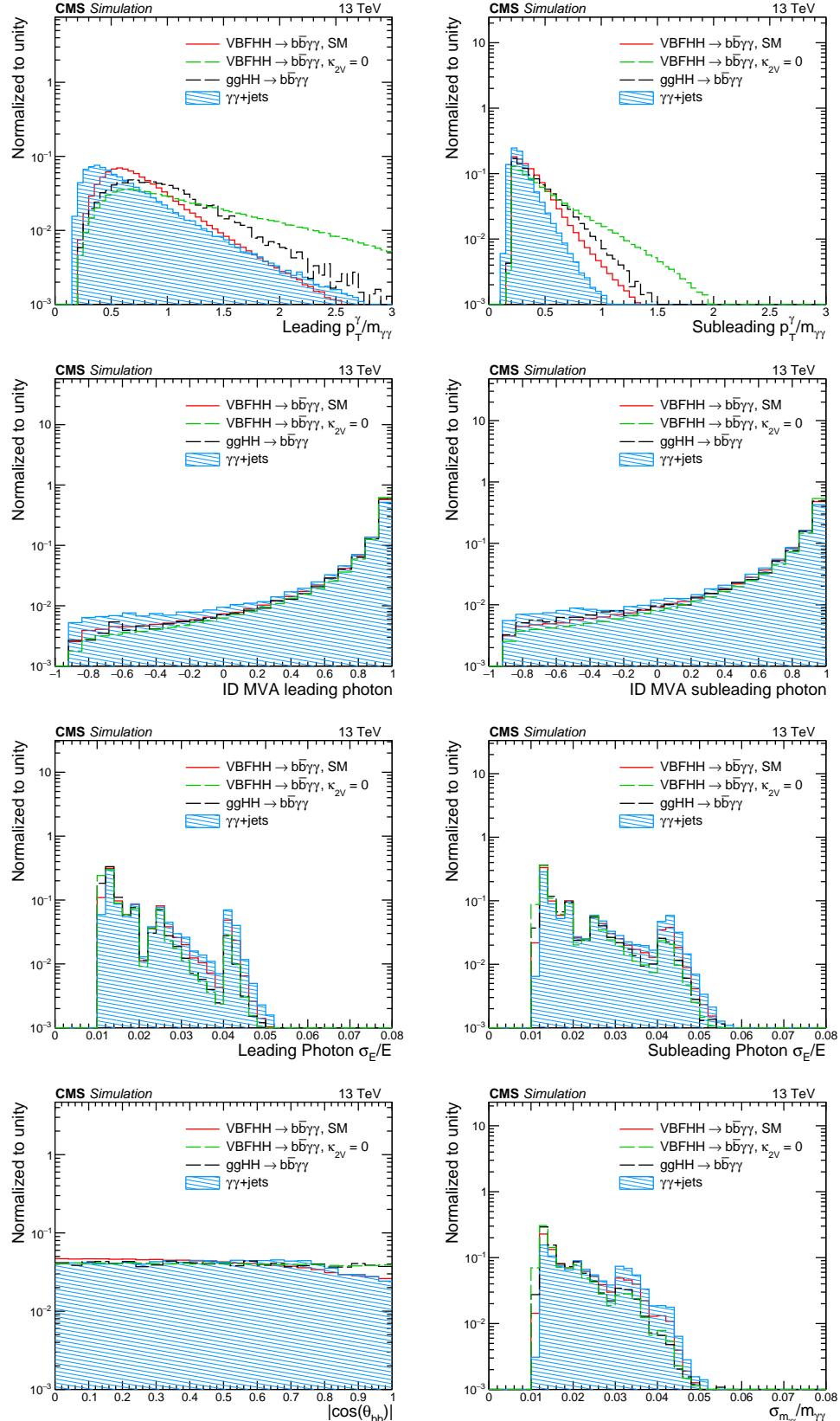
Figure 7.16: The variables related to the photons used in the MVA training. VBFHH SM , VBFHH $\kappa_{2V} = 0$, ggHH SM processes and $\gamma\gamma$ + jets background are shown in red solid line, green dashed line, black dashed line and blue filled area respectively.
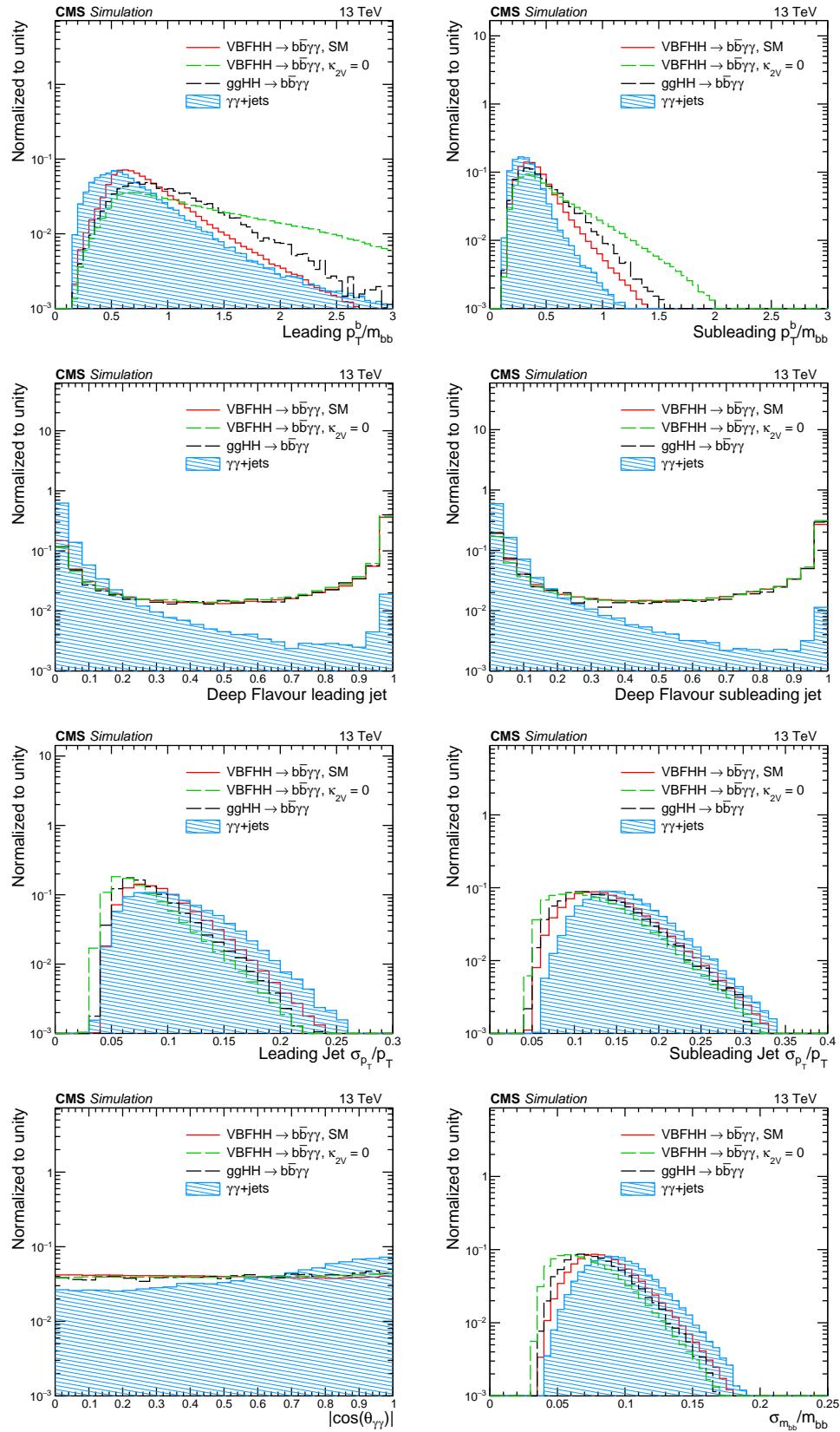
Figure 7.17: The variables related to the b-jets used in the MVA training. The description of the plots are the same as for Fig. 7.16.
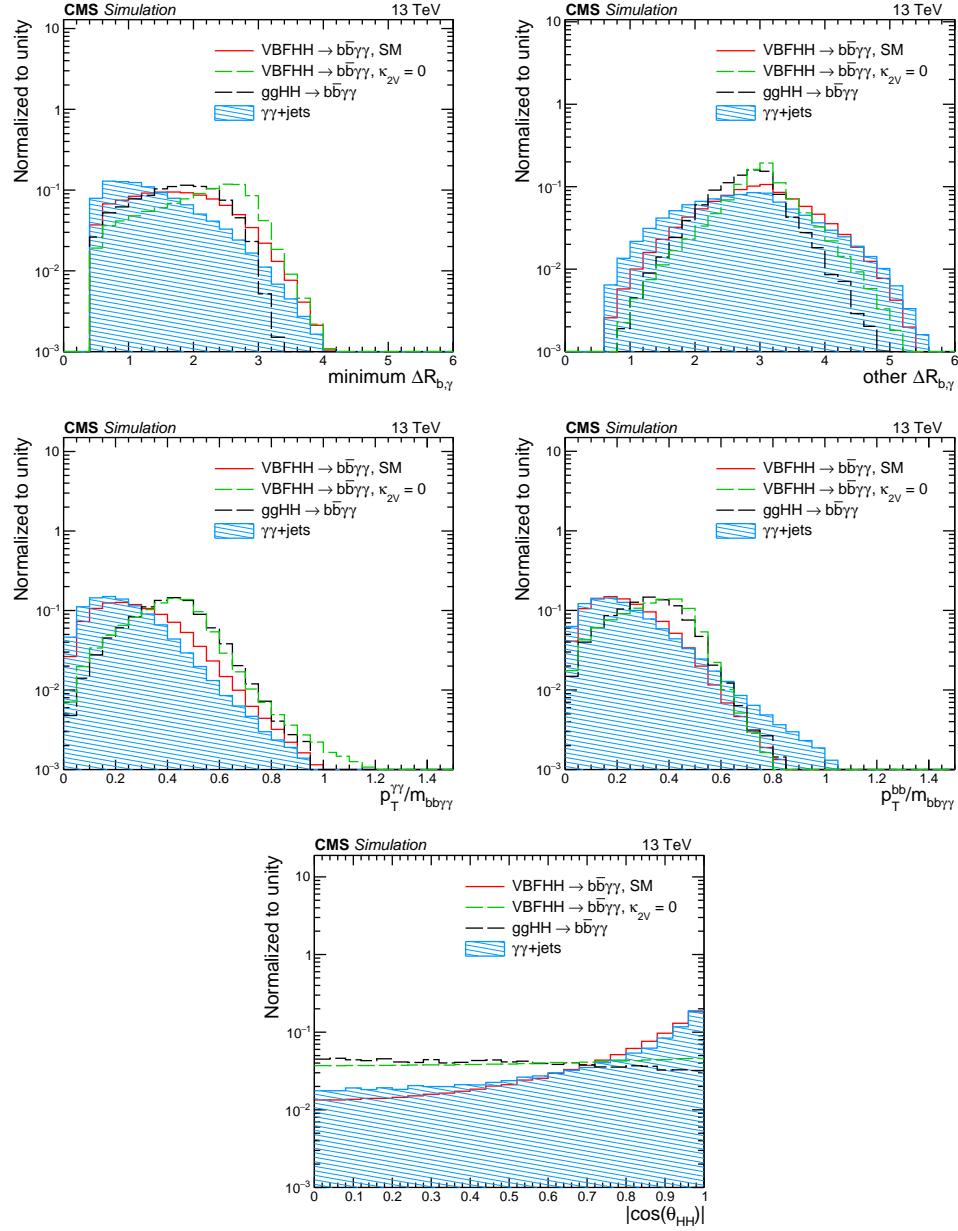
Figure 7.18: The variables related to the reconstructed H candidate. The description of the plots are the same as for Fig. 7.16.
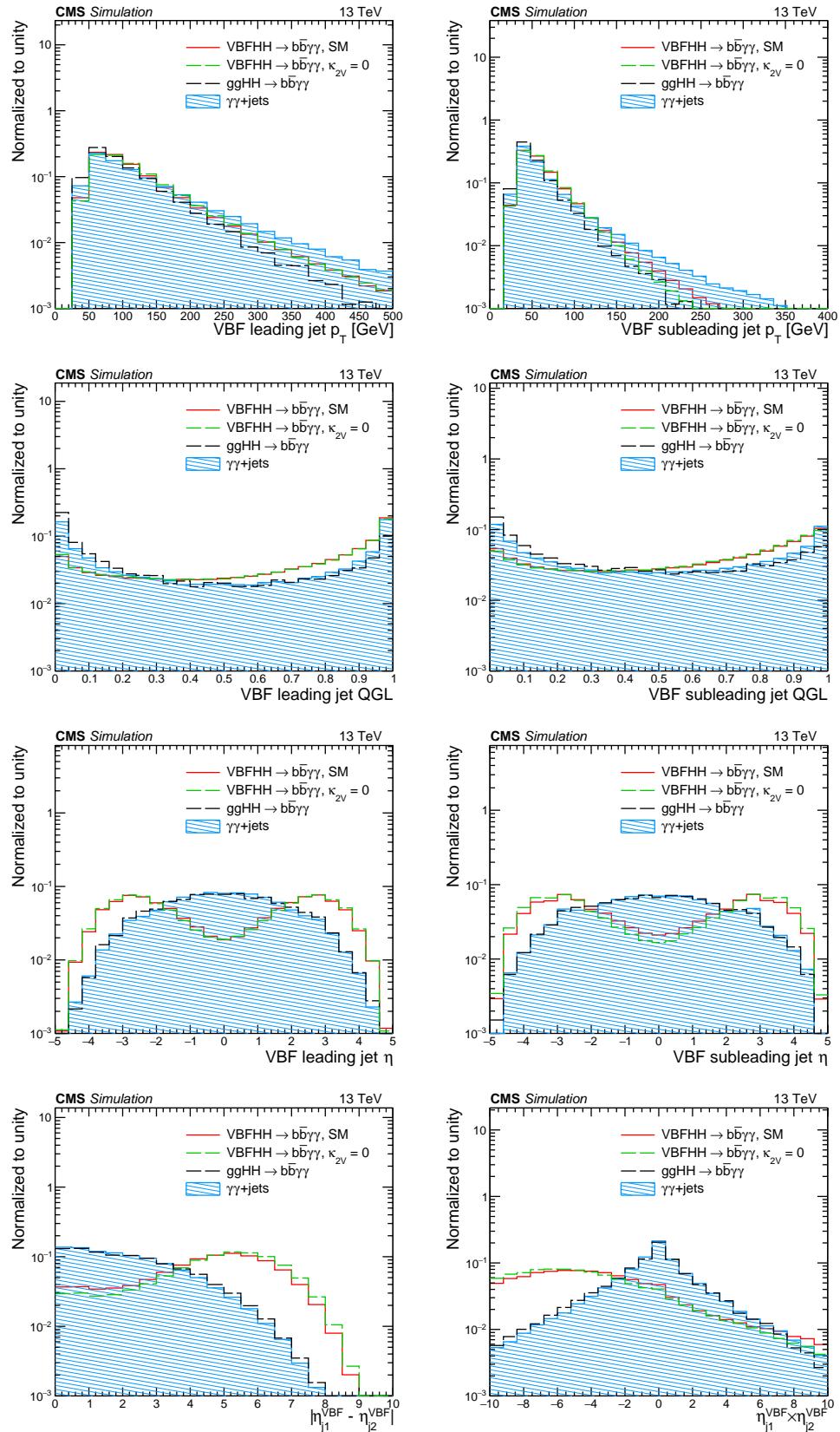
Figure 7.19: VBF-tagged jets related variables. The description of the plots are the same as for Fig. 7.16.
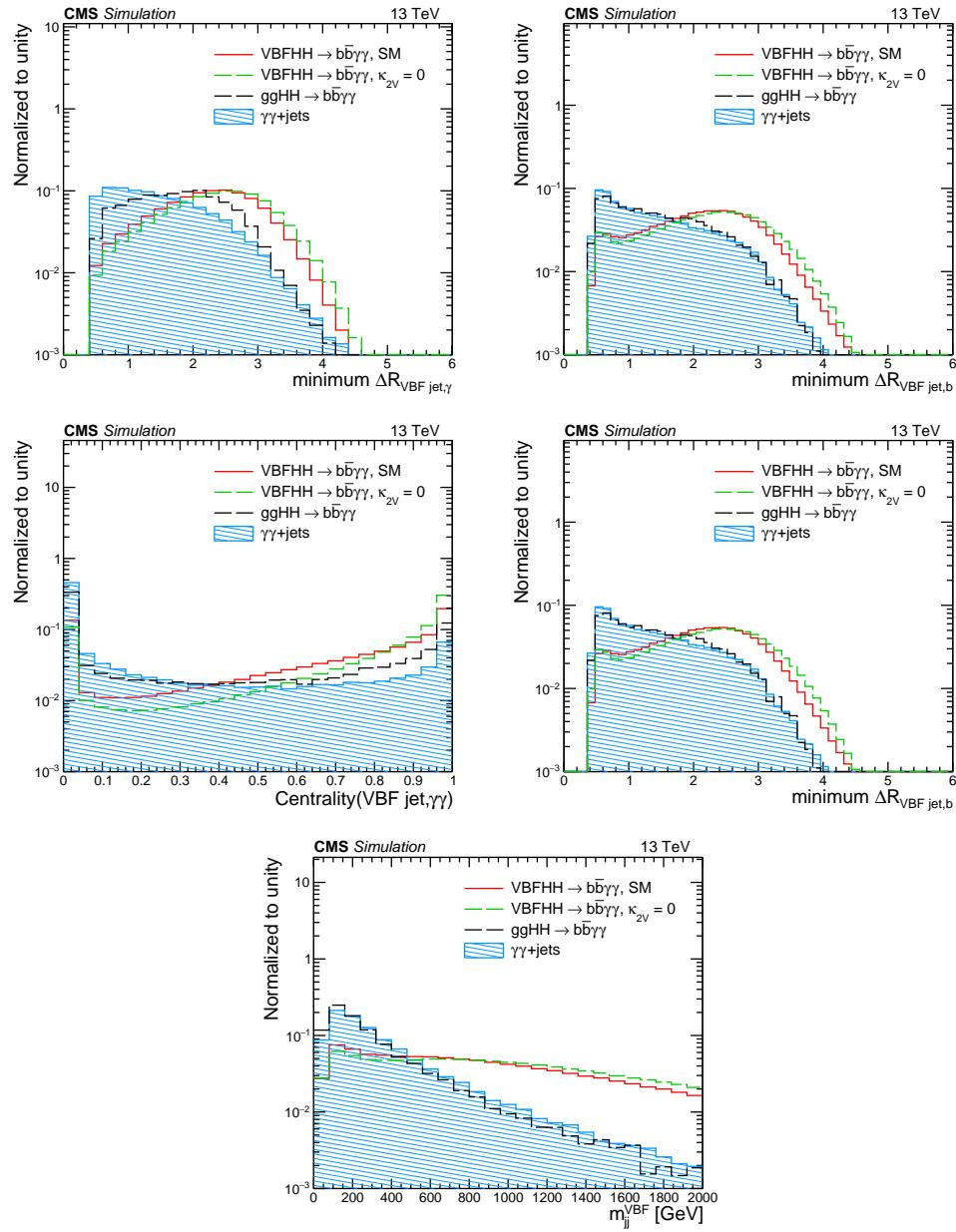
Figure 7.20: VBF-tagged jets related variables. The description of the plots are the same as for Fig. 7.16.

7.8.1 NLO re-weighting on the ggHH sample

As mentioned already, ggHH process is one of the most important backgrounds of the VBFHH analysis. However, for a given year, the available statistics of the MC sample of ggHH with accuracy of the next-to-leading-order (NLO) is quite low; hence it can not be be used for BDT training for VBFHH classifier. Hence the SM the leading-order (LO) sample of ggHH with enough statistics is re-weighted to produce an *improved* LO samples for the kinematic variables displaying major differences between LO and NLO processes. Two variables among the full set of inputs, the transverse momentum of the diHiggs system (p_T^{HH}) and the minimum angular distance between one photon and b-jets ($\Delta R_{b,\gamma}^{min}$), are found to have the maximum disagreement between the LO and NLO ggHH sample, so the LO ggHH sample is re-weighted in p_T^{HH}-$\Delta R_{b,\gamma}^{min}$ plane to the NLO process. Figure 7.21 presents the distributions of these two variables from the samples produced at LO and NLO. Figure 7.22 displays a number of distributions after the re-weighting to underline the fact that other variables used in the training are not affected in this re-weighting process.
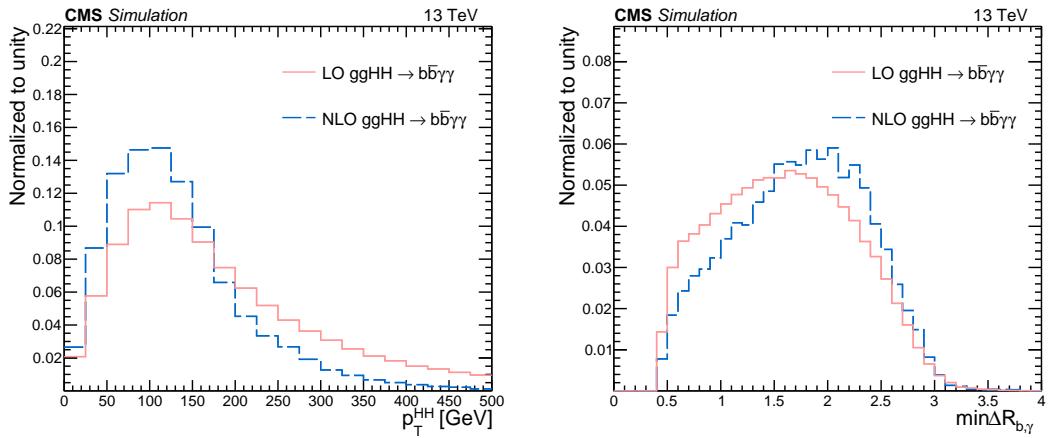
Figure 7.21: Normalized distributions for HH transverse momentum (left) and minimum angular distance between one photon and b-jet (right) in ggHH process before re-weighting; LO (black) and NLO (blue).

7.8.2 Validation of Training

Event samples are split into two for the training and the testing purposes. Upper panel of Fig. 7.23 and 7.24 present the validation of the training where the distributions from the test and the training samples for the signal and the backgrounds are superimposed on the same
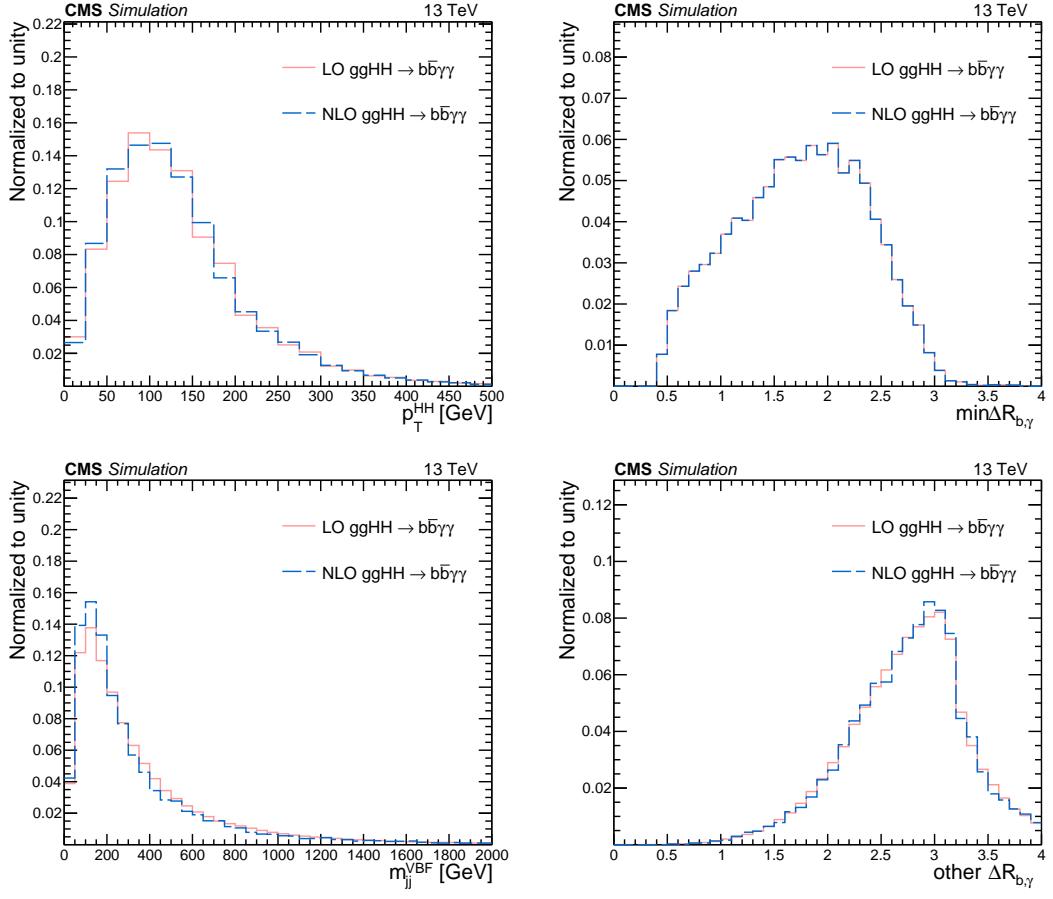
Figure 7.22: Normalized distributions of variables after re-weighting of LO ggHH (black) process with respect to NLO (blue).

canvas for $\widetilde{M}_X < 500$ GeV and $\widetilde{M}_X > 500$ GeV for the separation of VBFHH signal from non-resonant background and ggHH process respectively. Further in these plots the distributions for the three years have been merged according to the luminosity after checking that the individual distributions do not bear any specific difference. As expected, the VBFHH signal is very well-separated from the non-resonant process of $\gamma\gamma$ + jets, while the ggHH is less separated from the VBFHH signal due to the same behavior related to the H $\rightarrow \gamma\gamma$ and H $\rightarrow$ b$\overline{\text{b}}$ kinematics. The ROC curve corresponding to the non-resonant and the ggHH backgrounds for $\widetilde{M}_X < 500$ GeV and $\widetilde{M}_X > 500$ GeV are being presented. No over-training has been found in the model.

7.8.3 Transformation of MVA score

The VBFHH categories are optimized to maximize the expected sensitivity of the analysis. From Fig. 7.23 it is visible that the signal events are mostly populated at the very high value (close to 1) of the MVA scores, that may produce huge uncertainties at the time of optimization
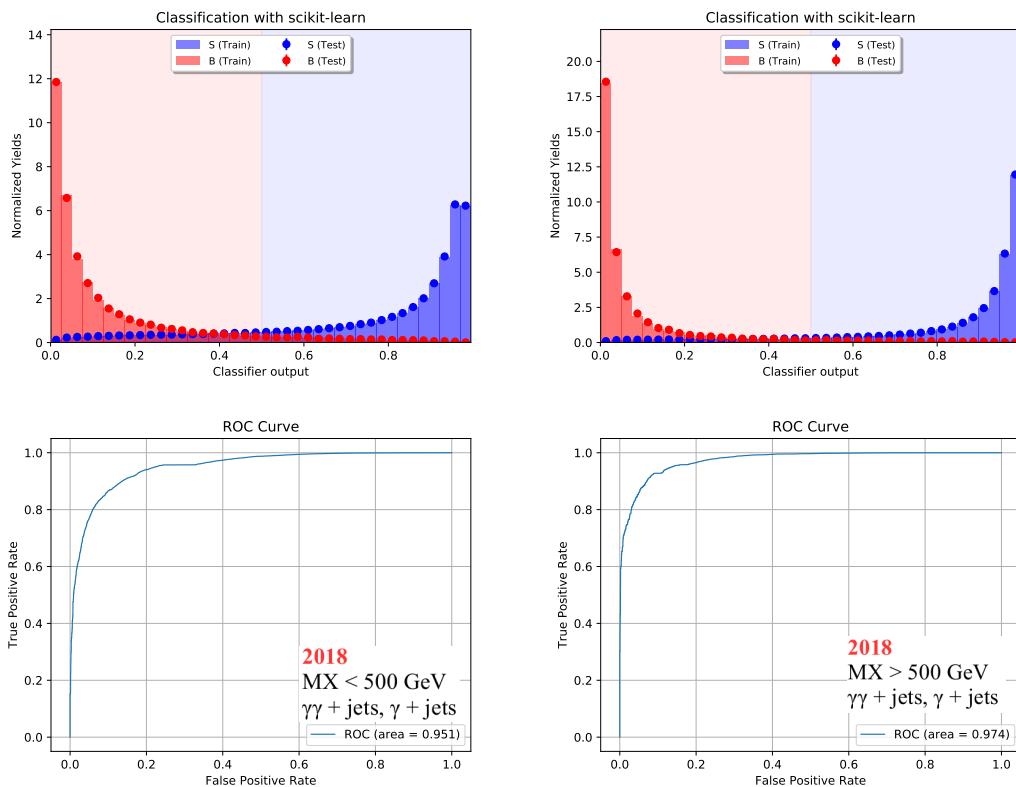
Figure 7.23: Upper row: MVA distributions of training and testing sampling of VBFHH signal (blue) and non-resonant continuum background (red) events in the two region: $\widetilde{M}_{\mathrm{X}} < 500$ GeV (left) and $\widetilde{M}_{\mathrm{X}} > 500$ GeV (right). Lower row: The ROC distributions of the training performance in the two $\widetilde{M}_{\mathrm{X}}$ regions.
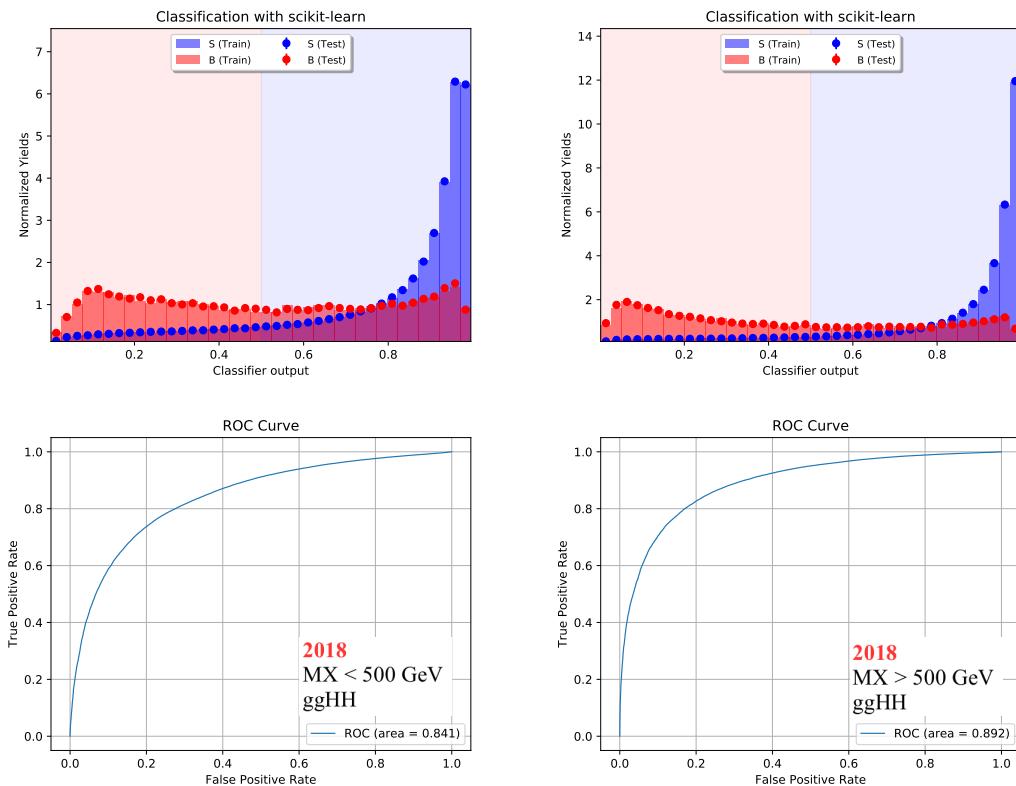
Figure 7.24: Upper row: MVA distributions of training and testing sampling of the VBFHH signal (blue) and ggHH (red) events in the two region: $\widetilde{M}_X < 500$ GeV (left) and $\widetilde{M}_X > 500$ GeV (right). Lower row: same as described in Fig. 7.23.

of the MVA threshold value. For this reason, the signal distribution is flattened by a cumulative transformation to optimization procedure. The same functional form of this transformation is applied for all the backgrounds and the data. The transformed distributions for the backgrounds are presented in Fig. 7.25 respectively by merging three years according to the integrated luminosity of each data taking year.

The distributions of the VBFHH MVA score in the region $\widetilde{M}_X > 500$ GeV and $\widetilde{M}_X < 500$ GeV are shown in Fig. 7.25.
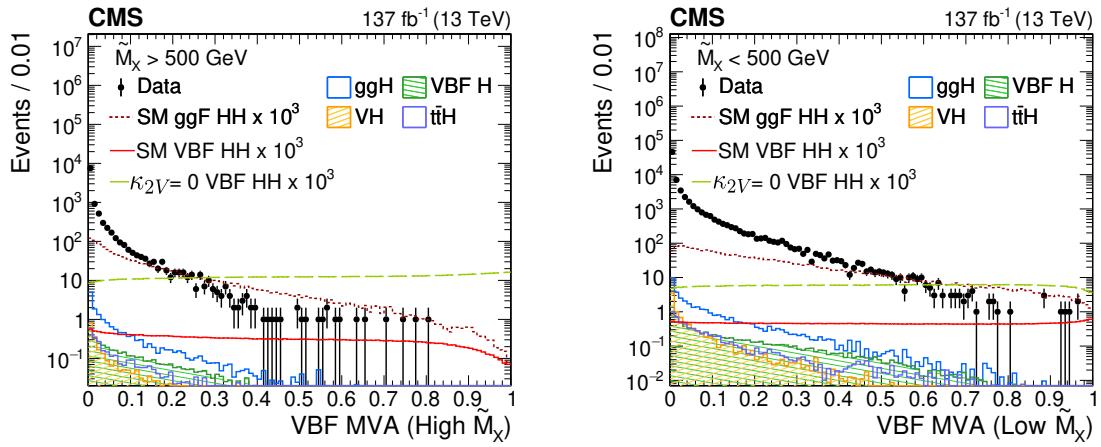
Figure 7.25: The distribution of the two MVA outputs is shown in data and simulated events in VBF regions for $\widetilde{M}_X > 500$ GeV (left) and $\widetilde{M}_X < 500$ GeV (right). Data, dominated by the $\gamma\gamma$ + jets and γ + jets backgrounds, are compared to the VBFHH signal samples with SM couplings and $\kappa_{2V} = 0$, SM ggHH and single H samples ($t\bar{t}H$, ggH, VBFH, VH). The error bars on the data points indicate statistical uncertainties.

7.8.4 Optimization of threshold for VBFHH categorization

Two threshold values for the transformed MVA are required to make the categorization for the VBFHH signal corresponding to 2 regions: $\widetilde{M}_X < 500$ GeV and $\widetilde{M}_X > 500$ GeV. These values are optimized by maximizing the expected signal significance, defined as $\frac{S}{\sqrt{B}}$, where S and B are the signal (combined contribution of VBFHH SM and $\kappa_{2V}=0$ signals) and non-resonant background of $\gamma\gamma$ + jets/γ + jets processes respectively, in the narrow di-photon invariant mass region of $115 < m_{\gamma\gamma} < 135$ GeV, while keeping a minimum of 8 MC background events in the side-band region of $m_{\gamma\gamma}$ defined as [100-115 GeV] and [135-180 GeV]. This optimization is done by merging MC samples for three years according to the respective integrated luminosity.

The optimized values of MVA score is found at 0.86 for $\widetilde{M}_X < 500$ GeV and 0.52 for $\widetilde{M}_X > 500$ GeV, which are shown in Fig 7.26. The score for ggHH process has been magnified by a factor of 100 for the purpose of display only.
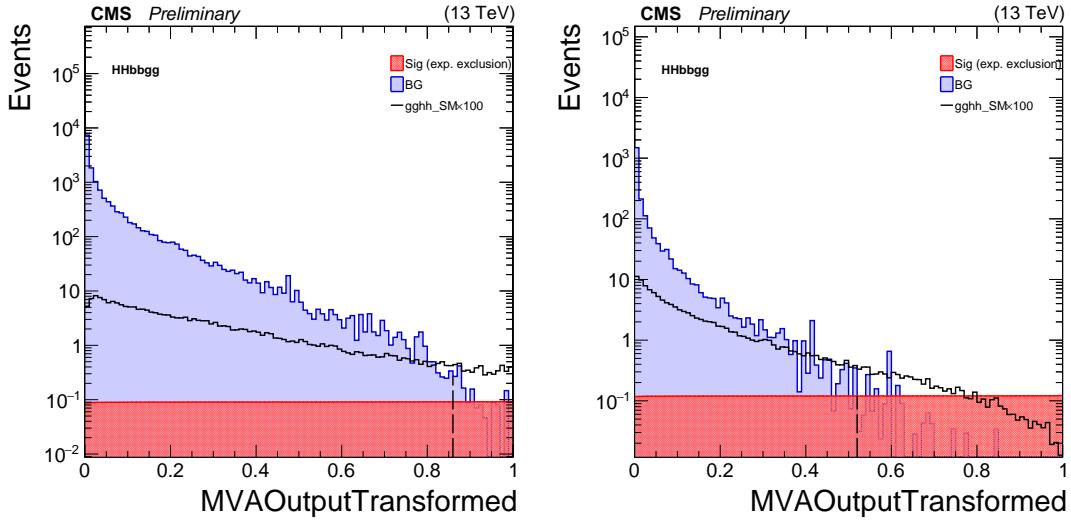
Figure 7.26: Optimization of threshold value for transformed MVA scores for selection of VBFHH events for $\widetilde{M}_X < 500$ GeV and $\widetilde{M}_X > 500$ GeV. The score for ggHH process has been magnified by a factor of 100.

The definition of the two VBFHH categories are given below, along with a cartoon of the categorization map in $\widetilde{M}_X$-MVA plane, in Fig. 7.27 (left).

- VBFHH CAT-0 : $\widetilde{M}_X > 500$ GeV and VBFHH MVA > 0.52

- VBFHH CAT-1 : $\widetilde{M}_X < 500$ GeV and VBFHH MVA > 0.84

The events which failed to pass the VBFHH category threshold are treated further in ggHH analysis, which is out of the scope of these thesis. For ggHH analysis a total of 12 (3 MVA $\times$ 4 $\widetilde{M}_X$) categories have been made, which is in Fig. 7.27 (right). Thus a total 14 analysis categories (2 from VBFHH & 12 from ggHH) are used simultaneously to extract results.

Table 7.6 presents the event yield for SM VBFHH signal, SM ggHH and other resonant single Higgs background processes along with the data observed at the side band region of the $m_{b\bar{b}}$ spectrum, which is dominated by the $\gamma\gamma$ + jets and γ + jets processes. The non-resonant backgrounds are underestimated in MC and hence its contribution is determined from data directly as explained later. The collective yields in the signal side band in data, for 3 years, corresponding to the integrated luminosity of 137 fb^{-1}, are also mentioned in the last column of Tab. 7.6
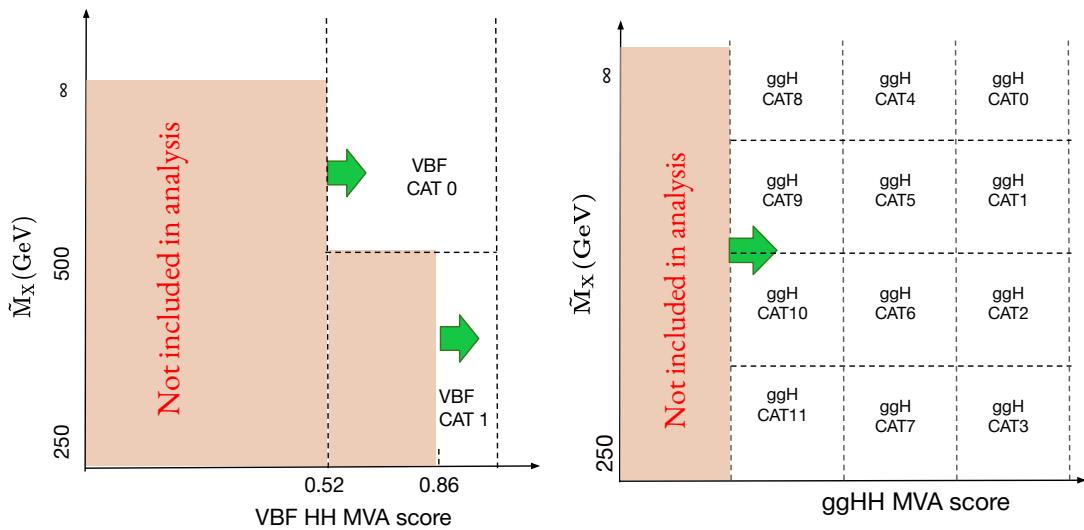
Figure 7.27: Cartoons of the categorization scheme for the VBFHH (left) and ggHH (right) analysis.

after blinding the region of $115 < m_{\gamma\gamma} < 135$ GeV. It is also checked that the migration of ggHH events to the VBFHH dedicated categories is less than 1%.

Table 7.6: Expected event yields for SM signal and backgrounds other than non-resonant ones and actual number of events in complete Run2 data excluding the blinded signal region of $115 < m_{\gamma\gamma} < 135$ GeV.

Category	VBFHH SM	ggHH SM	ggH	VBFH	VH	ttH	data in side band
VBFHH CAT-0	0.0116	0.0145	0.2433	0.1182	0.0025	0.1226	6
VBFHH CAT-1	0.0068	0.0052	0.0145	0.03	0.0004	0.0222	5

7.9 Statistical analysis

7.9.1 Signal modeling

For each of 14 categories (2 VBFHH + 12 ggHH) the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distributions are fit led to extract the signal shape parameters separately for the VBFHH and ggHH processes from simulation. The spectrum of $m_{\gamma\gamma}$ in the region $115 < m_{\gamma\gamma} < 135$ GeV is fitted with multi-Gaussian. The number of Gaussian functions needed to fit in each signal sample and in each category are determined by Fisher test (F-Test) method. The $m_{b\bar{b}}$ spectra is fitted with a double-sided crystal

ball (DSCB) function. Due to the poor resolution of the $m_{b\bar{b}}$ spectrum,, the fit is performed in the entire $m_{b\bar{b}}$ mass window ($70 < m_{b\bar{b}} < 190\,\text{GeV}$). The considered region is larger in this case than for $m_{\gamma\gamma}$ to account for the worse b-jet energy resolution. Figure. 7.28 shows the parametric fit of the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distribution of the VBFHH signal in the two VBFHH categories.

For the HH signal the total probability density function (PDF) is product of the individual
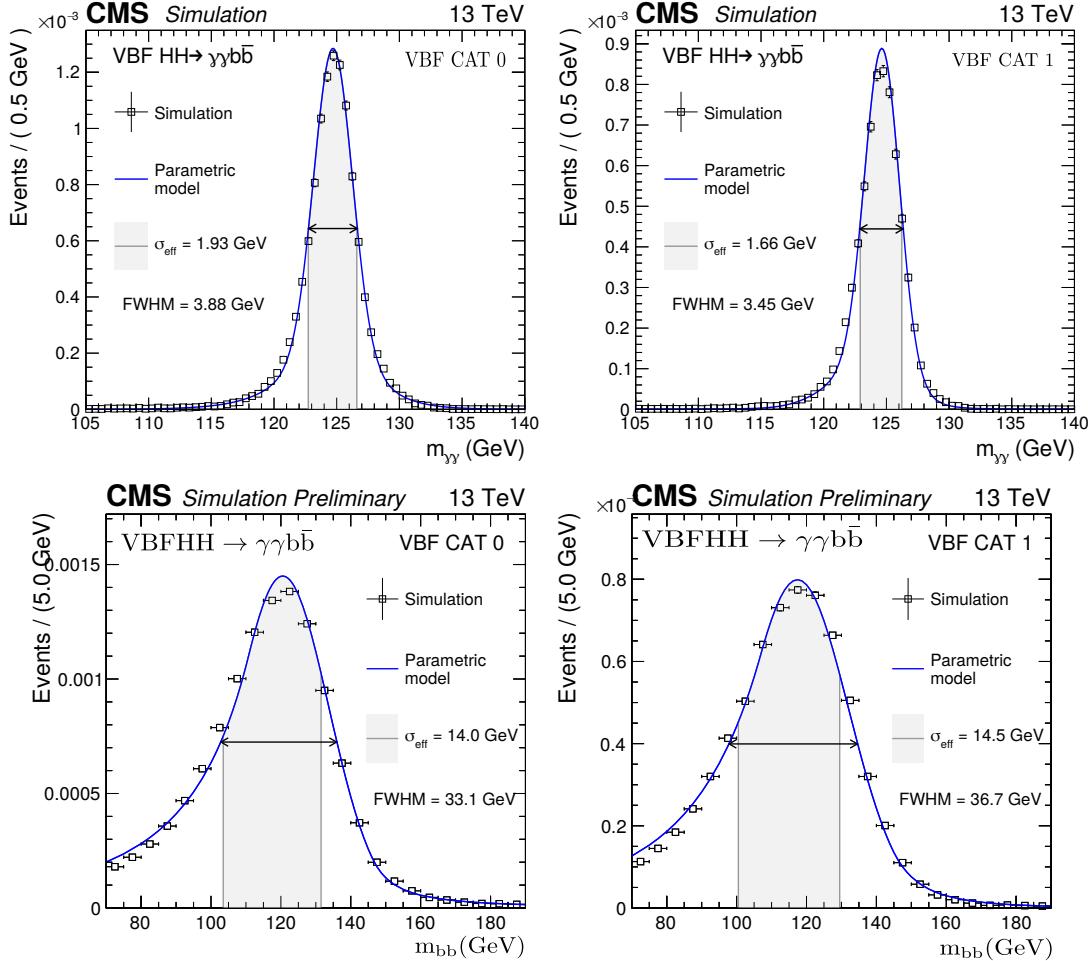
Figure 7.28: Parametrized signal shape for $m_{\gamma\gamma}$ (upper row) and $m_{b\bar{b}}$ (bottom row) of the VBFHH signal in the VBF CAT0 and VBF CAT1.

fitting functions of the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ spectrum, assuming the two distribution $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ have very minimal correlation as the H $\rightarrow \gamma\gamma$ and H $\rightarrow b\bar{b}$ are independent of each other. Figure. 7.29 shows the correlation between $m_{\gamma\gamma}$-$m_{b\bar{b}}$ distributions for VBFHH and ggHH process, for the both cases the correlation is found to be less than 0.1%.
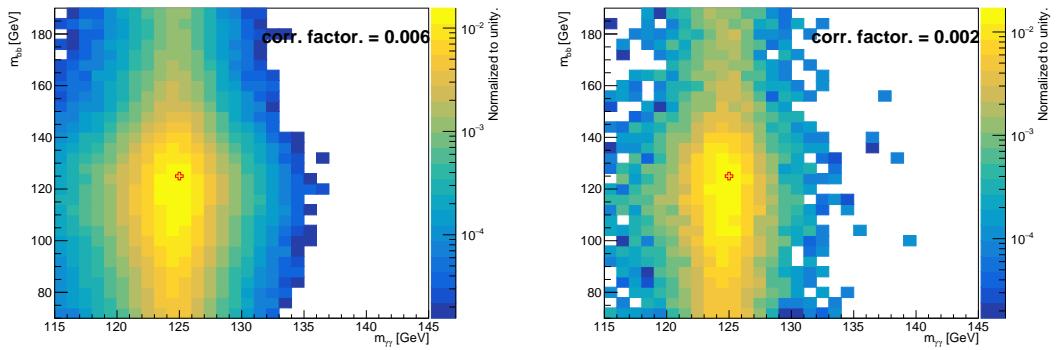
Figure 7.29: 2D scatter plot of VBFHH SM (left) and ggHH (right) in $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distribution. The correlation factor has been printed on the figures.

7.9.2 Resonant single-H background modeling

The contributions of the single-H background has been estimated from simulation by fitting $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distribution. In the SM single-H backgrounds, with H → γγ, the $m_{\gamma\gamma}$ spectrum is similar to that obtained from the HH processes. Hence to fit the $m_{\gamma\gamma}$ distribution, from resonant background of single production the same approach has been taken like the signal modeling described above in Sec. 7.9.1.

However the shape of the $m_{b\bar{b}}$ distributions of different single-H processes are different and modeled with separate functional forms. The $m_{b\bar{b}}$ distributions for ggH and VBFH processes are modeled using a Bernstein polynomial, for t$\bar{t}$H using a Gaussian function and for VH process $m_{b\bar{b}}$ is fitted by a Crystal Ball(CB) function. Interestingly, the core of the CB function is able to fit the hadroinc decay of the vector boson, specially Z → b$\bar{b}$, near 90 GeV. Figure. 7.30 shows the fit of the $m_{b\bar{b}}$ distribution for the different single-H processes for Run 2 data corresponding to the integrated luminosity of 137 fb^{-1}.

7.9.3 Non-resonant continuum background distribution

The non-resonant background contribution is modeled directly from data by fitting the sideband of the continuum distribution of $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ in each analysis categories separately. Three different families of function: Bernstein, Power law and exponential have been used as trails, to fit the distributions, finally the optimized choice of the polynomial function and it's order has been determined by the F-Test method, which is described below. In each analysis category the bias for the choice of the polynomial has been checked by fitting the background with another
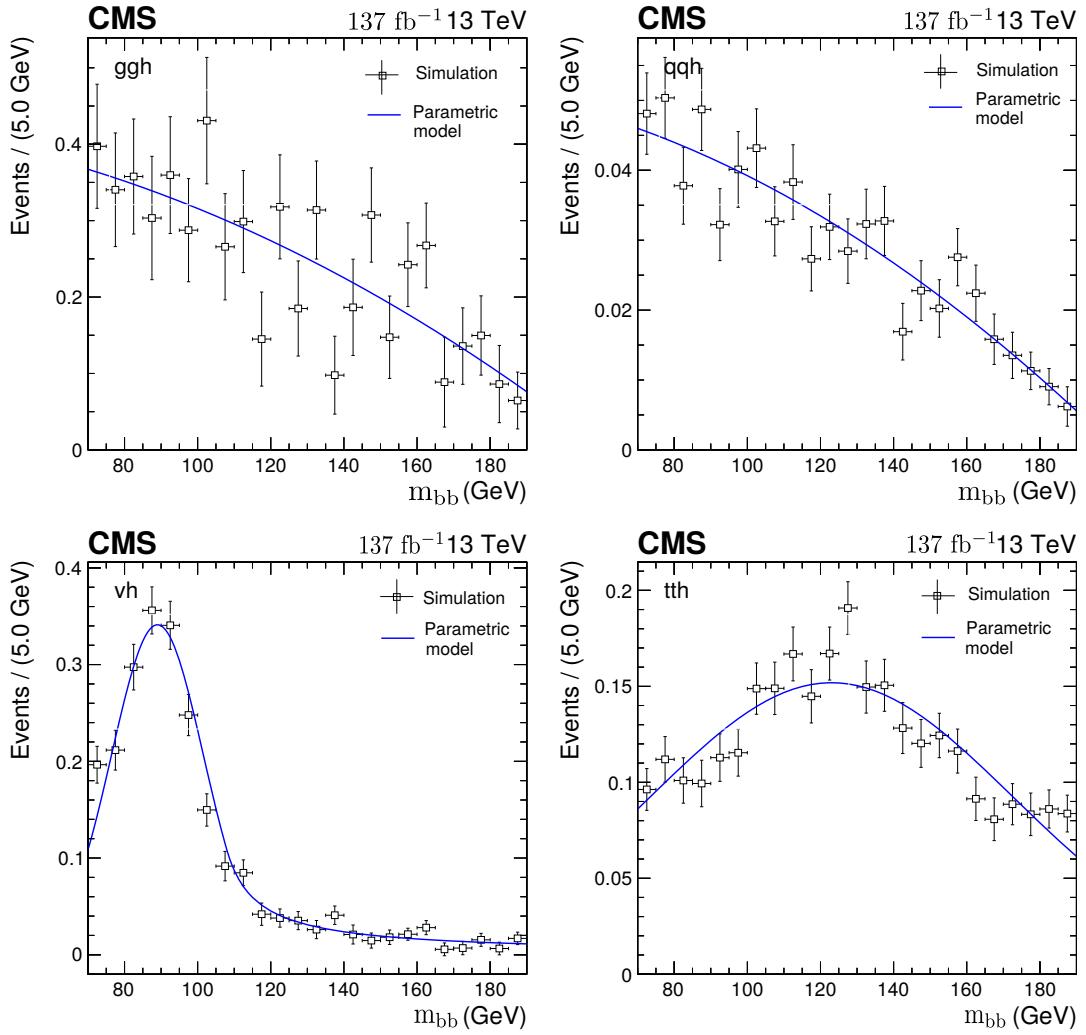
Figure 7.30: The parametric fit of the $m_{b\bar{b}}$ distributions for the single-H processes: ggH (top left), VBFH (top right), VH (bottom left) & $t\bar{t}$H (bottom right).

type of polynomial function, and the bias is found to be negligible.

In the F-Test mechanism, χ^2 value has been derived for each type of function after fitting the data sidebands. The number of degrees of freedom (n.d.f) is estimtaed from the number of bins of the fitted spectrum and parameters of the fitting function. The probability of that χ^2 value corresponding to that number of degrees of freedom $P(\chi^2,n.d.f)$ is calculated for each function. Finally the function to be used for the fit is chosen by maximizing the probability score $P(\chi^2,n.d.f)$.

Figure 7.31 displays the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ fits for the two VBFHH categories, where data are blinded in the region of $115 < m_{\gamma\gamma} < 135$ GeV and $100 < m_{b\bar{b}} < 140$ GeV respectively. The polynomial and its order corresponding to the best fit function of the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distributions in the two

VBFHH categories has been mentioned in Tab. 7.7.
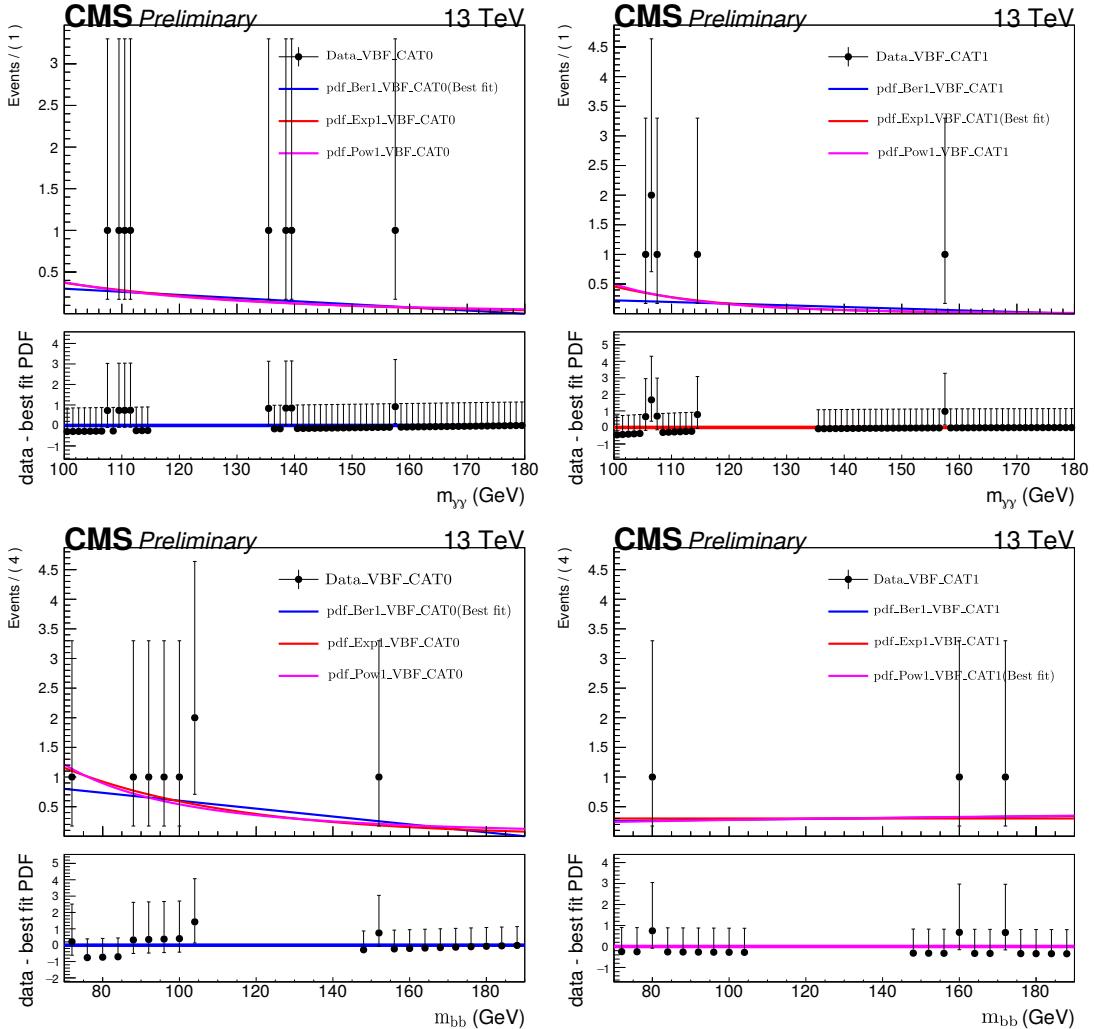
Figure 7.31: $m_{\gamma\gamma}$ (upper row) and $m_{b\bar{b}}$ (lower lower) fits for background modeling in 2 VBFHH categories each $\widetilde{M}_X < 500$ GeV and $\widetilde{M}_X > 500$ GeV. Data are blinded in the region of $115 < m_{b\bar{b}} < 135$ GeV and $100 < m_{b\bar{b}} < 140$ GeV respectively.

Table 7.7: The polynomial function and its order, used to fit the side band of the $m_{\gamma\gamma}$ and the $m_{b\bar{b}}$ distributions of continuum non-resonant background in the two VBFHH categories.

Category	$m_{\gamma\gamma}$	$m_{b\bar{b}}$
VBFHH CAT-0	Bernstein ord. 1	Bernstein ord. 1
VBFHH CAT-1	Exponential ord. 1	Power law ord. 1

7.10   Systematic uncertainties

The systematic uncertainties arise from both theoretical and experimental sources, which are applied only on the signal and resonant single-H background processes as the contribution for these background processes have been estimated from the MC simulation. The uncertainties due to the fitting of the continuum backgrounds in data-driven way have been propagated in the analysis using discrete profile method [116] technique.

The impact of the systematic uncertainties affect the results in two ways (i) normalization uncertainty: changes only the total event yields as a multiplicative factors, and (ii) shape uncertainty: it has effects on the kinematics which leads to the migration of events from one category to the other.

Theoretical uncertainties All the theoretical uncertainties are applied as normalization uncertainty. The major sources can be classified as follows,

- *Cross section uncertainties*: The major sources of the theoretical uncertainties on the calculation of cross sections for the HH and single H processes due to the QCD scale variation, the variation of PDF set and α_S. The cross sections and associate uncertainties of these processes are tabulated in Tab. 7.8.

Table 7.8: Summary of the HH and H production cross sections along with the associated uncertainties (in %) due the QCD scale and PDF+α_S at $\sqrt{s}$=13 TeV.

Process	Prod. mode	Cross section
HH	ggHH	$31.05^{+6\%}_{-23\%} \pm 3.0\%$ (fb)
	VBFHH	$1.726^{+0.03\%}_{-0.04\%} \pm 2.1\%$ (fb)
Single H	ggH	$48.58^{+4.6\%}_{-6.7\%} \pm 3.2\%$ (pb)
	VBFH	$3.782^{+0.4\%}_{-0.3\%} \pm 2.1\%$ (pb)
	VH	$2.256^{+3.8\%}_{-3.1\%} \pm 1.6\%$ (pb)
	$t\bar{t}$H	$0.507^{+5.8\%}_{-9.2\%} \pm 3.6\%$ (pb)

- *Br(HH $\rightarrow$ bb$\gamma\gamma$) uncertainties* : Uncertainties on the branching ratio calculation for the H $\rightarrow \gamma\gamma$ and H $\rightarrow$ b$\bar{\text{b}}$ decay modes are shown in Tab. 7.9.

- *Parton shower uncertainty in VBFHH event simulation* : The systematic uncertainties due to the choice of the PYTHIA parton shower scheme for the VBFHH process has been

Table 7.9: The branching ratio of the H → γγ and H → b$\overline{\text{b}}$ decay modes with the associated uncertainties for M_H = 125 GeV.

Branching mode	$\mathcal{B}r$ and uncertainty
H → γγ	$0.00227^{+1.73\%}_{-1.72\%}$
H → b$\overline{\text{b}}$	$0.58^{+0.78\%}_{-0.80\%}$

deployed by propagating the relative difference of the event yields corresponds to the standard p_T ordered and dipole recoil parton shower [117] scheme in each category.

Experimental uncertainties All the experimental uncertainties appear as shape uncertainty and they are mostly uncorrelated across the years. The major sources of the experimental uncertainties are listed below.

- *Photon preselection*: All diphoton candidates are required to satisfy the various offline preselection criteria designed more stringently than the HLT requirements described in Tab. 7.2. The preselection efficiency has been measured both in data and MC using $Z \to \text{e}^+\text{e}^-$ events. The systematic uncertainty for the preselection efficiency is less than 1%.

- *Photon identification*: The uncertainties come from the application of the scale factors for photon identification BDT score to cover the discrepancy between data and simulation from $Z \to \text{e}^+\text{e}^-$ events and also from the $Z \to \mu^+\mu^- \gamma$ events. The amount on unceratianty is found to be 3% of the per photon MVA ID scale factor.

- *Photon energy scale and resolution*: A regression has been applied to correct the photon energy scale and resolution, and it is found that the performance of the regression is different in data and MC. So a additional 1 to 2% of uncertainties applied on the regression correction factor to cover the data-simulation discrepancy.

- *Trigger scale factor*: The triggere SF uncertainty is about 1-2% applied on each photons as a function of p_T in each η, R_9 region which is described in Sec. 7.5

- *Jet energy scale and resolution*: The jet energy scale of each jet is a correction factor determined from the p_T balance of the jets with Z boson, decaying to $\ell\ell$, γ+ jets and with respect to a di-jet system. The uncertainties due to the jet energy scale and resolution are found to be 5-10% relative to their p_T and propagate to the analysis by varying jet energy

within their uncertainties.

- *b tagging*: The uncertainties in b-tagging efficiency are evaluated with the control samples of the semileptonic $t\bar{t}$ decays, Z+jets events, and the inclusive QCD multijet events. The uncertainties associated with the selection of the working point of the DEEPJET score for the tagger vary between 4 and 8% depending on jet flavor, p_T, and η. The DEEPJET discriminants of the two b-tagged jets are also used as an input to the BDT classifiers. An assessment of the control samples of QCD multijet and leptonic $t\bar{t}$ events revealed only small differences in the DEEPJET discriminant distribution between data and simulated events above the threshold defining the working point. These differences are found to have subdominant effects on the response of BDT classifiers, leading to migration of events between categories at a percent level. .

- *Integrated luminosity*: The uncertainty in the luminosity measurement has both correlated and uncorrelated components, which vary across different data-taking periods. The year-wise uncertainties are determined by the luminosity monitors of the CMS detector [118–120]. For the data taking periods of 2016 and 2018, the uncorrelated uncertainties are 2.3 and 2.5% respectively. The correlated part of the uncertainty is 1.8%.

- *Pileup modeling*: The number of primary interactions per bunch crossing varies with the instantaneous luminosity during data taking operation. In order to match the distribution of pileup in data with that in the simulated samples, weights are applied that are determined by studying minimum bias data sets. A normalization uncertainty is derived by altering the pileup weights obtained by changing the minimum bias cross section by ±4.6% of its nominal value 69.7 mb.

- *Pileup jet identification*: The pileup jet id is applied for the jets produced in the hard scatter to discriminate them from the pileup jets. The corresponding uncertainty depends on the p_T and η of the jets to cover the difference of the pileup id score between data and MC.

The accuracy of this analysis is statistically limited. The observed best fit value of the signal strength of the inclusive HH production ($\hat{\mu}_{HH}$) and the impacts of the first thirty major systematic uncertainties are shown in Fig. 7.32. The normalized pull of a particular systematic uncertainty is defined as the difference of the of the final ($\hat{\theta}$) and initial(θ_0) values of the

uncertainty parameter, normalized with the error ($\Delta\theta$) of that:

$$\text{pull of NP} = \frac{\hat{\theta} - \theta_0}{\Delta\theta} \tag{7.13}$$

The ranking of the systematic uncertainties are based on their impact on $\hat{\mu}_{HH}$.
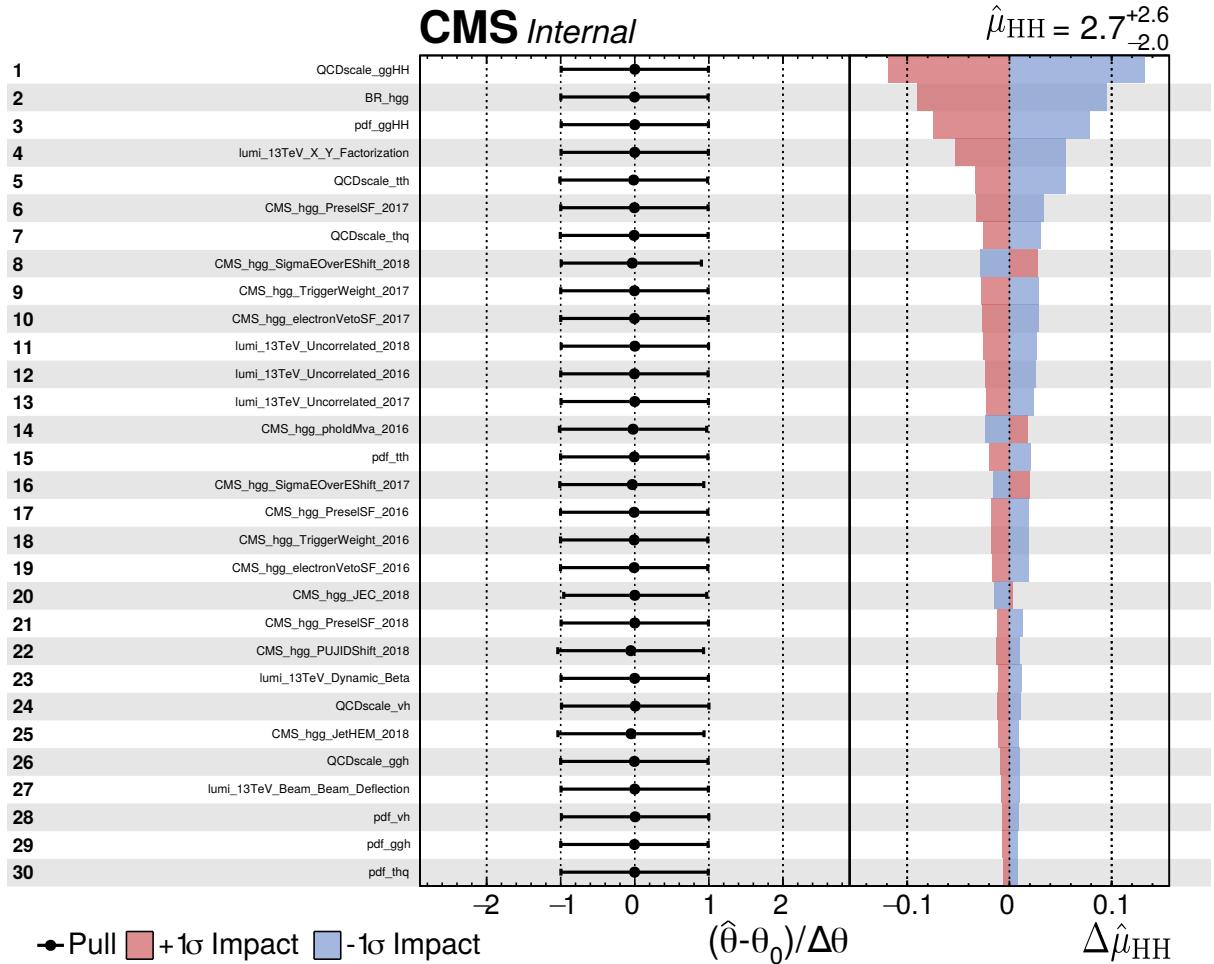
Figure 7.32: Ranking of the first thirty major sources of systematic uncertainties included in the fit according to their impact on the measured signal strength $\hat{\mu}_{HH}$.

7.11 Results

A 2-dimensional unbinned maximum likelihood fit to the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distributions is performed simultaneously in the 14 HH (12 from ggHH and 2 from VBFHH) categories to extract the HH signal strength. The data and the signal-plus-background fits for the two VBFHH categories are shown in Fig. 7.33.
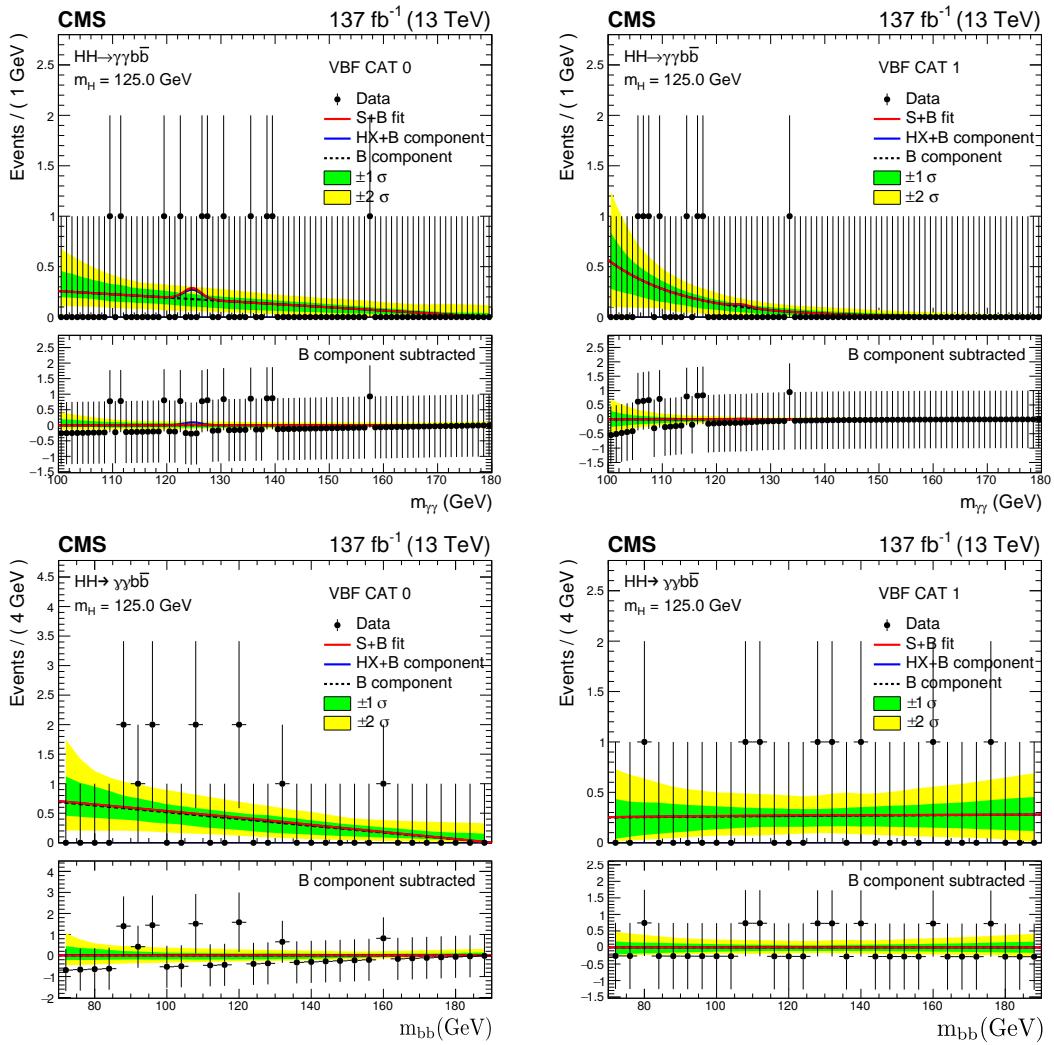
Figure 7.33: Invariant mass distributions $m_{\gamma\gamma}$ (upper row) and $m_{b\bar{b}}$ (bottom row) in VBF CAT-0 (left) and VBF CAT-1 (right) for the selected events in data (black points). The solid red line shows the sum of the fitted signal and single-H and continuum background (HH+H+B), the solid blue line shows the background component from the single Higgs boson and the non-resonant processes (H+B), and the dashed black line shows the non-resonant background only component (B). The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel in each plot shows the residual signal yield after the background subtraction.

The weighted distribution of events while merging all categories according to the factor $S/(S + B)$ (where S is the number of the signal events and B is the number of background events derived from the signal-plus-background fit), is shown in Fig. 7.34.
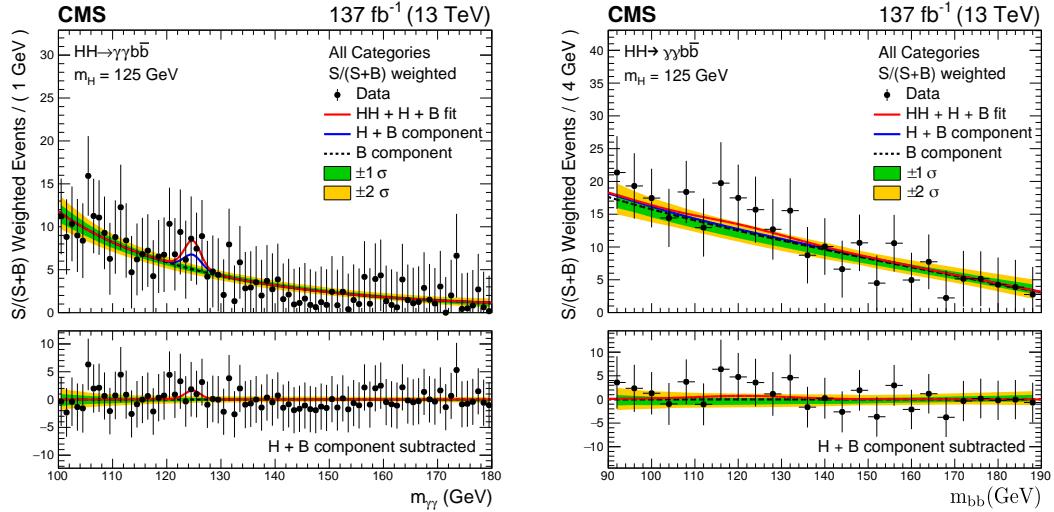
Figure 7.34: Invariant mass distributions $m_{\gamma\gamma}$ (left) and $m_{b\bar{b}}$ (right) for the selected events in data (black points) weighted by S/(S+B), where S (B) is the number of signal (background) events extracted from the signal-plus-background fit. The description of the plot is same as Fig. 7.33.

No deviation from the background only hypothesis is observed in the analysis. An upper limit at 95% confidence level has been derived on the signal strength of the HH production due to the ggHH and VBFHH processes (μ_{HH}) and only VBFHH (μ_{VBFHH}) process.

The observed (expected) 95% CL upper limit for the inclusive HH production corresponds to 7.7 (5.2) times the SM prediction while the corresponding quantity for only VBFHH production is 225 (208) times the SM prediction.

The 95% CL upper limit on the inclusive HH signal (VBFHH) cross section times branching fraction is presented as a function of κ_λ (κ_{2V}) in Fig. 7.35. The theoretical prediction of the cross sections are superimposed on the plots, which provide the observed (expected) allowed ranges for the coupling modifiers,

$$-3.5 < \kappa_\lambda < 8.5 \ (-2.5 < \kappa_\lambda < 8.2)$$

$$-1.3 < \kappa_{2V} < 3.5 \ (-0.9 < \kappa_{2V} < 3.1)$$

.

Although 14 HH categories have been used simultaneously to set the limit of κ_λ and κ_{2V} both, for the κ_λ scan the main sensitivity is driven by the 12 ggHH categories where as the sensitivity for κ_{2V} is due to the 2 VBFHH categories, where the ggHH is treated as a background. It can be seen in Fig. 7.35(right) that this analysis is more sensitive to anomalous values of κ_{2V} than to the region around the SM prediction. This is related to the fact that, for anomalous values of κ_{2V}, the total cross section is enhanced and the $\widetilde{M}_X$ spectrum is harder. The shape of the observed and expected curves in these two figures are exactly due to the efficiency times acceptance ($\epsilon \times A$) of the signal samples at the each of the individual points of κ_λ or κ_{2V}. The $\epsilon \times A$ of the ggHH signal is maximum at $\kappa_\lambda \sim 2$, which corresponds to the lower upper limit, as evident in Fig. 7.35 (left). Similarly, the $\epsilon \times A$ of the VBFHH signal is minimum near the value of $\kappa_{2V} \sim 1$, this causes the higher upper limit at the value of $\kappa_{2V} \sim 1$, as shown in Fig. 7.35 (right). For the higher values of the coupling modifiers both κ_λ and κ_{2V}, the $\epsilon \times A$ of the signal samples are almost saturated, which is understood from the flat curves of the upper limits at the either side of the SM value of the parameters.
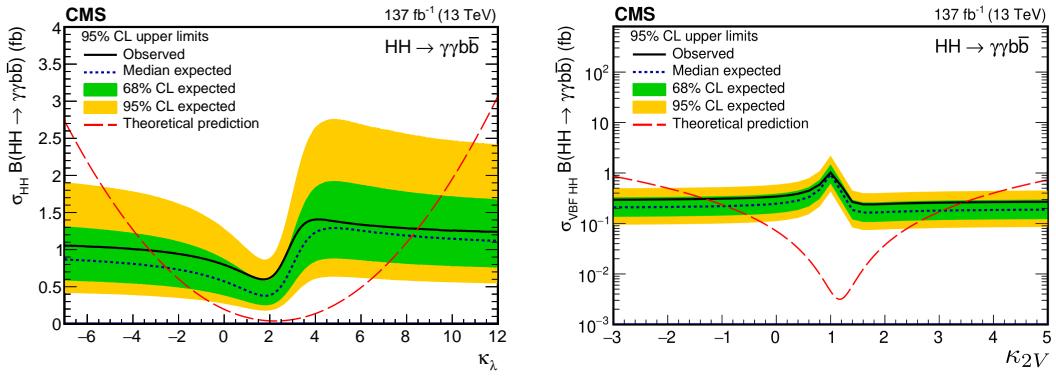
Figure 7.35: Expected and observed 95% CL upper limits on the product of the inclusive HH (left) and VBF HH (right) production cross sections multiplied with $\mathcal{B}(HH \rightarrow b\overline{b}\gamma\gamma)$ obtained for different values of κ_λ (keeping $\kappa_{2V} = 1$) and κ_{2V} (keeping $\kappa_\lambda = 1$). The green and yellow bands represent, respectively, the one and two standard deviations bands across the expected limit. The long-dashed red line shows the theoretical prediction.

In the context of the VBFHH and ggHH analysis, the sensitivity for the κ_{2V} and κ_λ parameter is shown as 2-dimensional negative log likelihood scan shown in Fig. 7.36.
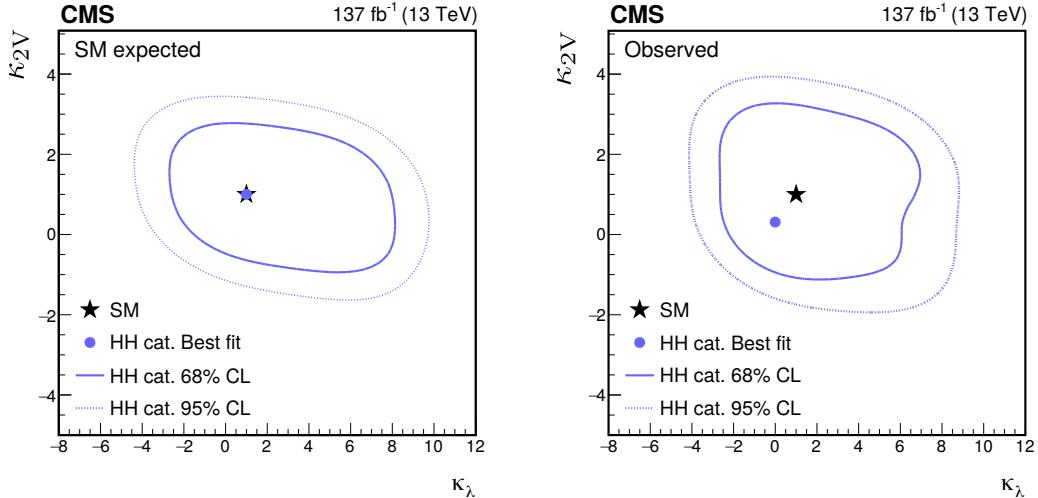
Figure 7.36: Expected (left) and observed (right) 2D negative log-likelihood scan for κ_λ-κ_{2V}. The blue solid and dashed lines show the 68% and 95% CL contour repectively whereas the black star and blue filled circle indicate the SM prediction and the best fit value.

7.12   Summary

A search for the non-resonant production of a pair of Higgs bosons in the decay channel $b\bar{b}\gamma\gamma$ using full Run 2 data with an integrated luminosity of 137 fb^{-1} is performed. The VBFHH production is included for the first time in CMS Run 2 HH physics analysis program. The observed (expected) upper limit at 95% CL on inclusive HH production is 7.7 (5.2) times the SM prediction, while for the VBFHH production is 225 (208). This work is publsihed in Ref. [25].

Chapter 8

Projection of non-resonant pair production of Higgs boson and decays to bb$\gamma\gamma$ final state at the HL-LHC

8.1   Introduction

Since in the SM, the production rate of Higgs boson pair (HH) is extremely low, more than one thousand times weaker than the rate of the single H production, it is anticipated that during the LHC operations in the near future, corresponding to an integrated luminosity ($\mathcal{L}$) of few hundred fb^{-1}, the experiments will be able to establish only the evidence of HH production. However, it is the high luminosity operation of the LHC (HL-LHC) which will pave the way for the observation of the HH production. This motivates the present study to evaluate the potential of the CMS experiment for the SM HH signal at the HL-LHC condition, in the context of the Snowmass Workshop [121].

The analysis strategy is guided by Run 2 study [25] as well as the work reported in [122] in the context of the HL-LHC. The present study corresponds to an updated detector geometry and reconstruction algorithm. Further, in addition to the dominant gluon-gluon fusion (ggHH) process, this study includes the vector boson fusion (VBFHH) production mode of HH production as well; in the earlier study [122] the VBF process was not considered. The study is performed for the Phase-2 upgraded CMS detector emulated with DELPHES [123] fast and parametric detector simulation package tuned for $\sqrt{s} = 14$ TeV; the average number of minimum

bias interactions per bunch crossing (pileup) is considered to be 200. A detailed overview of the CMS detector upgrade program is available in Ref. [124–131], while the expected performance of the reconstruction algorithms and pile-up mitigation with the CMS detector is summarized in Ref. [132].

8.2 Analysis flowchart

Since this study includes both ggHH and VBFHH production mode, from the steps of the event selection to the categorization of events the analysis is designed to proceed in parallel for ggHH and VBFHH stream by maintaining orthogonality
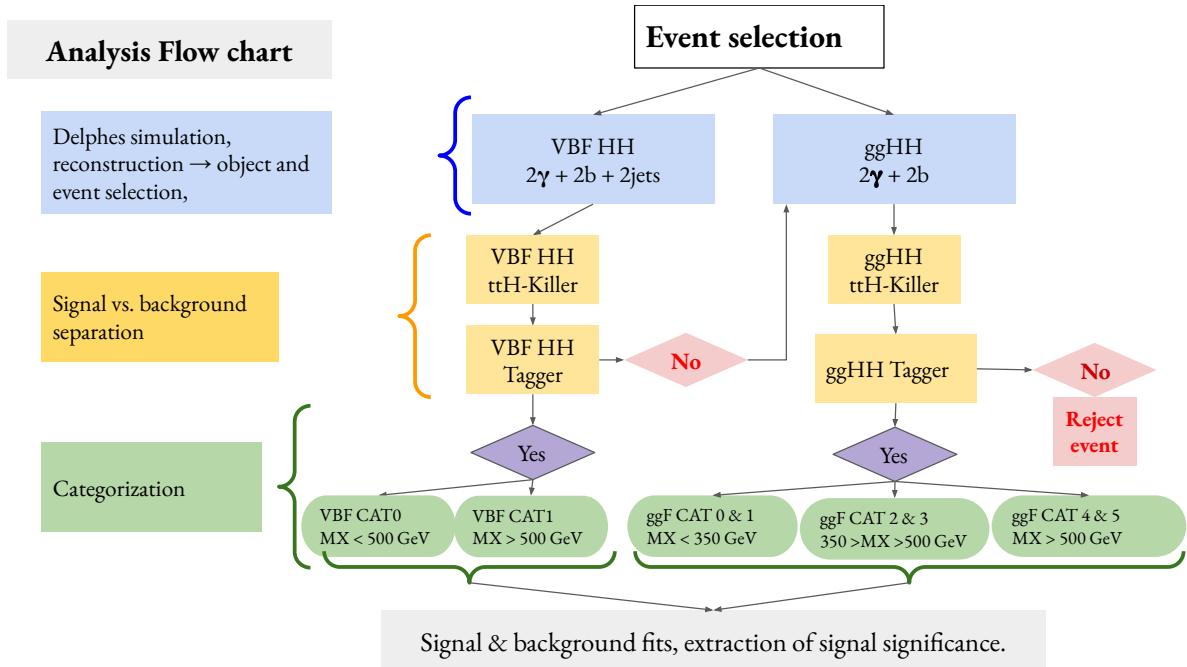
Figure 8.1: Analysis flowchart of the HH → b$\bar{\text{b}}$γγ projection study at the HL-LHC scenario.

8.3 Simulated samples and event reconstruction

The MC simulation of the signals and backgrounds have been done using the state-of-the-art event generators. The production of signals, non-resonant backgrounds like γγ+jets and γ+jets and single-H processes follow the same procedure as described in Sec. 7.3.

Since this is only a projection study, there is no real collision data; hence it is important to estimate the possible contribution of backgrounds as precise by as possible by only relying on

simulation. So the sub-dominant continuum background contributions are also included here, which is not important in Run 2 analysis due to their small contributions. from $\gamma\gamma$ + jets/γ + jets, as the continuum background contribution is estimated from the side-band of real data in Run 2 analysis. Such processes are the inclusive single top (t + X) or top pair ($t\bar{t}$ + X) productions in association with one or two photons where the b-jets originating from the decay of the top quarks can potentially mimic the final state. Hence it is important to take into account the proper background processes corresponding to this class of backgrounds. Accordingly, the processes $t\bar{t}$ + $\gamma\gamma$, $t\bar{t}$ + γ+ jets, tγ+ jets, as well as inclusive $t\bar{t}$ + X are produced by the MG5_aMC @NLO version-2.7 while the decay of top quarks and further hadronisation are implemented through PYTHIA 8.212.

All the simulated samples are processed with DELPHES [133] fast and parameterized simulation of the simplified Phase-2 upgraded CMS detector with an average number of pileup collisions of 200 superposed on the hard scatter event. Subsequent analysis is carried out with DELPHES reconstructed events which use particle flow algorithm. Pile-up subtraction is implemented at the reconstructed particle level via "pileup per particle identification", PUPPI [21] algorithm, that was specifically tuned to reduce the dependence of the missing transverse energy variable on the number of pileup interactions in a given event. The jets are reconstructed from PUPPI particles using the anti-k_T jet clustering algorithm [15] with a distance parameter of 0.4. Additional momentum smearing and re-scaling has been applied to the reconstructed jets to match the observed momentum resolution using full simulation. The b-tagging method uses a deep neural network (DNN) based secondary vertex algorithm, DEEPJET [18]; the tagging efficiency and fake-rates are parameterised based on the full simulation and using the parton flavor information from Monte Carlo events.

For the reconstructed photons the momentum resolution and scale is derived from the full simulation as a function of p_T and $|\eta|$. A relative isolation variable is defined by summing the p_T of all the PUPPI particles within a $\eta - \phi$ cone of radius 0.3 scaled by the p_T of the photon. An additional identification efficiency factor, dependent on p_T and $|\eta|$ is folded with the isolation efficiency to match that derived from the full simulation.

8.4   Event selection

The selected events are required to have a pair of reconstructed photons and a pair of b-jets, satisfying a set of criteria similar to the analysis reported in [25, 122] which is also described in Sec. 7.6. The analysis relies on a di-photon trigger, whose efficiency in this analysis is assumed to be 100% for the given offline selection requirements. The reconstructed photons are required to pass stringent isolation and lenient identification criteria. The leading and sub-leading photons should have transverse momenta p_T above 30 and 20 GeV respectively, and pseudo-rapidity $|\eta| < 2.5$. If at least one of the photons is in the transition gap between the barrel and endcap portions of the ECAL ($1.442 < |\eta| < 1.556$), the event is rejected. The diphoton invariant mass ($m_{\gamma\gamma}$) is required to be in the range of [100,180] GeV while the leading (subleading) photons should satisfy $p_\mathrm{T}/m_{\gamma\gamma} > 0.33$ (0.25).

With the presently available detector description and the object reconstruction algorithm in DELPHES, the extension of the pseudorapidity range for photons up to $|\eta| \sim 3$ instead of 2.5 results in about 2% increase in the signal acceptance and about 16% increase in the non-resonant background yields; hence, in this study, the photon selection is restricted to $|\eta| < 2.5$. Subsequent implementation of the actual capability of the CMS Phase-2 detector, in the DELPHES software, is expected to result in an improvement of the signal-to-background ratio in this region.

The events should have at least two jets and the jets are required to have $p_\mathrm{T} > 30$ GeV and $|\eta| < 2.5$, as well as satisfy the tight jet identification criteria. Each jet is checked for its b-tag identification score. No additional b-jet energy calibration has been applied in this study. The pair with the highest sum of the b-tag scores is selected. Finally, the invariant mass of the selected b-jet pair ($m_{b\bar{b}}$) is required to be within the range [70,190] GeV.

After the photons and b-jets are selected, an additional requirement is imposed on the angular separation $\Delta R > 0.4$ between any the two selected photons and b-jets. This removes the spatial overlap of the reconstructed objects.

To select the events corresponding to HH production via VBF, additional requirements are imposed. The VBFHH process is characterized by the presence of two additional energetic jets at the high $|\eta|$ region, corresponding to the two quarks from each of the colliding partons scattered away after the collision. These jets are required to be separated from the selected

photons and b-jets ($\Delta R_{j\gamma} > 0.4$ and $\Delta R_{jb} > 0.4$), to lie inside $|\eta| < 5$ and to have transverse momenta greater than 40 (30) GeV for leading (subleading) one. The jet pair with the highest invariant mass (m_{jj}) is selected as the VBF-tagged jets. We call this VBFHH selection where the event is required to have a pair of photon, a pair of b-jets and a pair of additional jets satisfying VBF-like criteria described above. The event without the VBF jets goes into the ggHH-selected sample. Both classes of events undergo further selections as discussed below.

The distributions of $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ from the signal and the background processes for the ggHH and VBFHH selected events are presented in Figs. 8.2 and 8.3 respectively. The contribution of VBFHH signal in the ggHH-selected events is negligible. However ggHH signal events may mimic the topology of the VBFHH process due to QCD higher order processes as well as final state radiations. Hence the VBFHH-selected sample is likely to have considerable amount of contamination from the ggHH signal process. This is tackled by a dedicated method, described later. Due to the excellent photon energy resolution, the peaking structure at the Higgs mass is well-pronounced in case of the $m_{\gamma\gamma}$ distributions. From both figures, it is evident that the signal rate is overwhelmed by the background. The salient features of the subsequent analysis are described below. The evnt yeilds contributing in this analysis both in ggHH and VBFHH
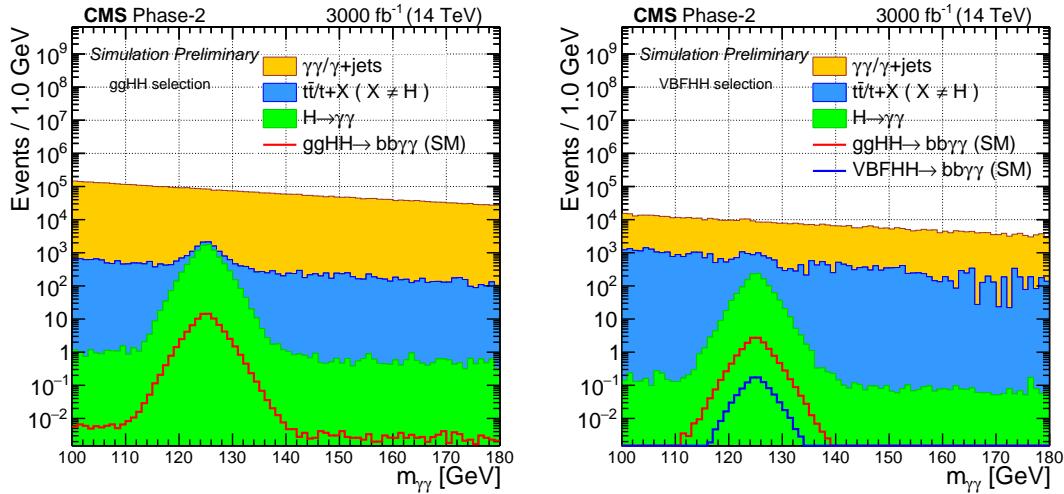
Figure 8.2: Distributions of the $m_{\gamma\gamma}$ variable for the ggHH- (left) and VBFHH- (right) selected events corresponding to $\mathcal{L}$ = 3000 fb^{-1}; ggHH signal (in red), VBFHH signal (in blue) are overlaid on different background processes (filled stacks in different colours).

phase spaces are presented inTab. 8.1 Figure 8.4 shows the $\widetilde{M}_{\mathrm{X}}$ distribution for the signal and the background events in the ggHH and VBFHH event samples; it is obvious that the shapes of the spectra are different. Therefore, it is expected that the relative signal significance against
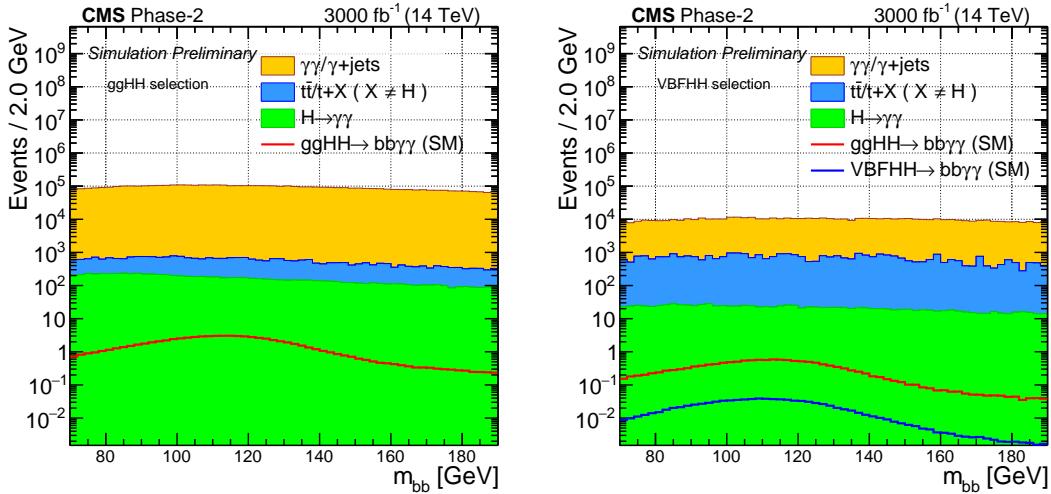
Figure 8.3: Distributions of the $m_{b\bar{b}}$ variable for the ggHH- (left) and VBFHH- (right) selected events corresponding to $\mathcal{L} = 3000$ fb^{-1}; ggHH signal (in red), VBFHH signal (in blue) are overlaid on different background processes (filled stacks in different colours).

Table 8.1: Event yields from different processes contributing in the ggHH & VBFHH event selection for an integrated luminosity of 3000 fb^{-1}.

Phase space	$\gamma\gamma/\gamma$+jets	$t\bar{t}$ +X (not H)	single-H	VBFHH	ggHH
ggHH	4476008.85	185179.89	7396.09	-	61.12
VBFHH	550853.14	39059.95	1252.03	1.49	15.53

the collective background will be different for different regions of $\widetilde{M}_X$. This feature is utilized at a later stage of the analysis.
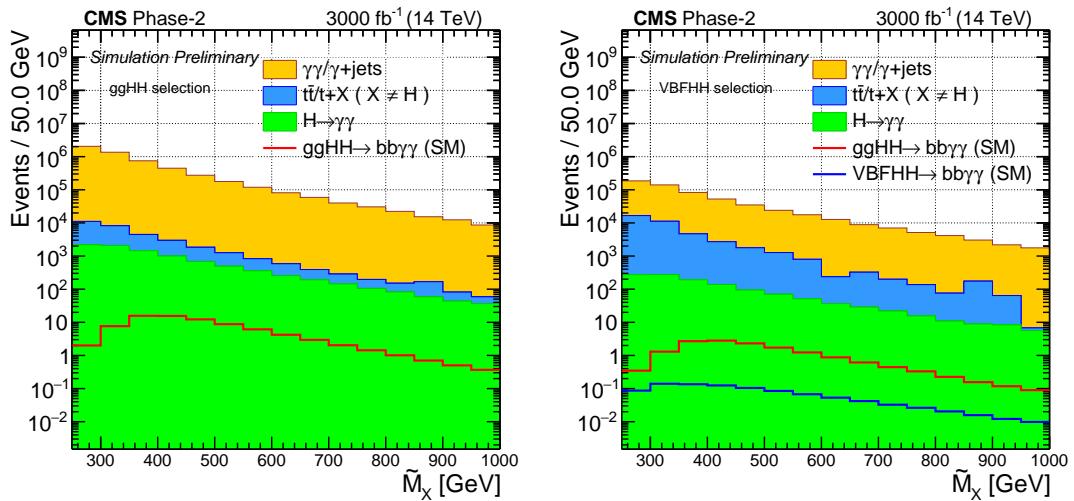
Figure 8.4: Distributions of the $\widetilde{M}_X$ variable for the ggHH- (left) and VBFHH- (right) selected events corresponding to $\mathcal{L} = 3000$ fb^{-1}; the ggHH (in red), VBFHH (in blue) signal components are overlaid on different background processes (filled stacks in different colours).

8.4.1   t̄tH background rejection

The ggHH and VBFHH signals are much smaller than the background rates. Hence an additional background rejection method was employed by using TMVA [134] based Gradient Boosted Decision Tree (BDTG) discriminant. Among all the single Higgs boson production processes(H+X), where H decays into γγ, the most dominant process in the signal region is t̄tH, since it replicates the same final state where two b-jets are likely to originate from the decays of the top quarks. Hence two separate multivariate discriminants are developed against the t̄tH background process: ggHH-*ttHKiller* and VBFHH-*ttHKiller* for the ggHH and VBFHH selected events respectively. The high impact feature variables used for the training include the transverse momenta of the selected b-jets, the multiplicities of the b-jets and the isolated leptons, the angular separation between the missing transverse momentum and the two selected b-jets etc., the detailed list is given in Sec. 7.7. The *ttHKiller* scores for the ggHH- and VBFHH-selected events are presented in Fig. 8.5. Events with high scores are selected for further analysis to reduce the t̄tH background process. The threshold values are -0.2 and 0.05 which correspond to 90% efficiencies for ggHH and VBFHH signal selection while the t̄tH background rejection efficiencies are about 85% and 90% for the two cases respectively. It is to be noted that the *ttHKiller* rejects a significant amount of the other t̄t + X events as well and the rejection factors is given in Tab. 8.2.

Table 8.2: Rejection (in %) of top induced backgrounds using ttHKiller.

Process	Rejection (in %)
t̄tH	85-90
t̄t γγ	82
t̄t γ	75
t̄t +X (not H)	79
tγ+jets	40

8.4.2   Non-resonant background rejection

The amount of the nonresonant background component, due to γγ + jets and γ + jets processes, is much higher than the signal processes by several orders of magnitude. Hence another set discriminant is used to distinguish ggHH and VBFHH signals from the non-resonant back-
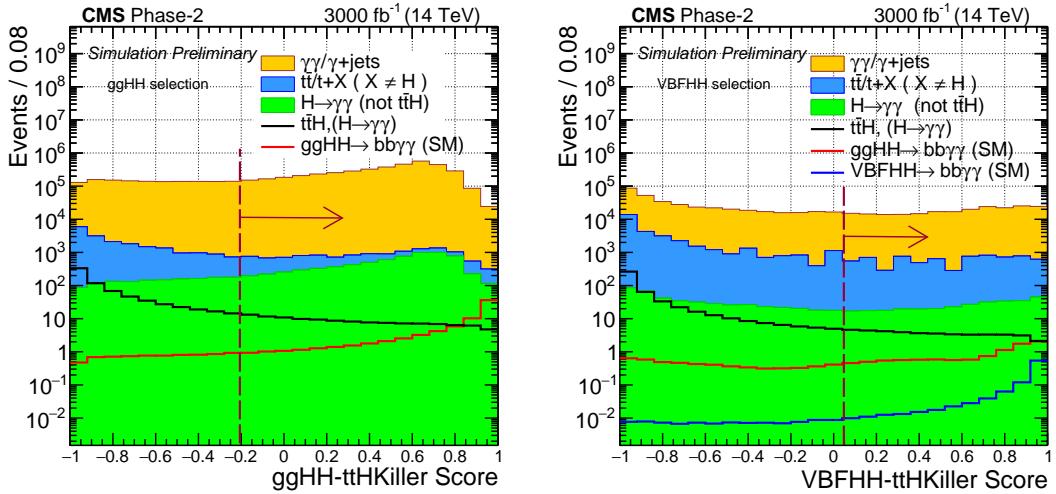
Figure 8.5: Distribution of the ggHH-*ttHKiller* and the VBFHH-*ttHKiller* discriminant scores for the ggHH- (left) and VBFHH- (right) selected events corresponding to $\mathcal{L} = 3000$ fb^{-1}; ggHH signal (in red), VBFHH signal (in blue) are overlaid on different background processes (filled stacks in different colours).

grounds. The ggHH selected event sample essentially has ggHH signal, albeit dominated by the non-resonant background. The contamination due to VBFHH signal is negligible in this case. Hence a multivariate analysis neural network, ggHH-*Tagger*, is trained with ggHH signal against the nonresonant background. However, it is to be noted that the VBFHH selected sample contains both the ggHH and VBFHH type of signal events though the total signal component is tiny compared to the background events. Hence, a multiclass deep neural network (DNN) based discriminant, VBFHH-*Tagger*, is trained to separate the VBFHH signal simultaneously from the nonresonant background as well as the ggHH signal. The distinctive topology of the VBF selection criteria helps to well-separate VBFHH from ggHH and increases the purity of the VBFHH signal (migration of ggHH signal events is at a level lower than 1%). The discriminating variables are used in both ggHH-*Tagger* and VBFHH-*Tagger* are same as those used in Run 2 analysis described in Sec. 7.9.3.

In the final stage of the analysis the distributions of $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ are used to establish the presence of inclusive HH signal; so these variables have not been used in the training to get rid of any mass sculpting. Figure. 8.6 shows the normalized distribution of the $m_{\gamma\gamma}$ variable for different values of ggHH-*Tagger* (left) and VBFHH-*Tagger* (right) scores; no sculpting of mass has been found due to the discriminator score. Thus, the discriminants mentioned

above are not correlated with the variables $m_{\gamma\gamma}$, $m_{b\bar{b}}$ as well as $\widetilde{M}_X$. Figure 8.7 shows the score distributions for ggHH-*Tagger* (left) and VBFHH-*Tagger* (right) for the ggHH- and the VBFHH-selected event samples respectively.
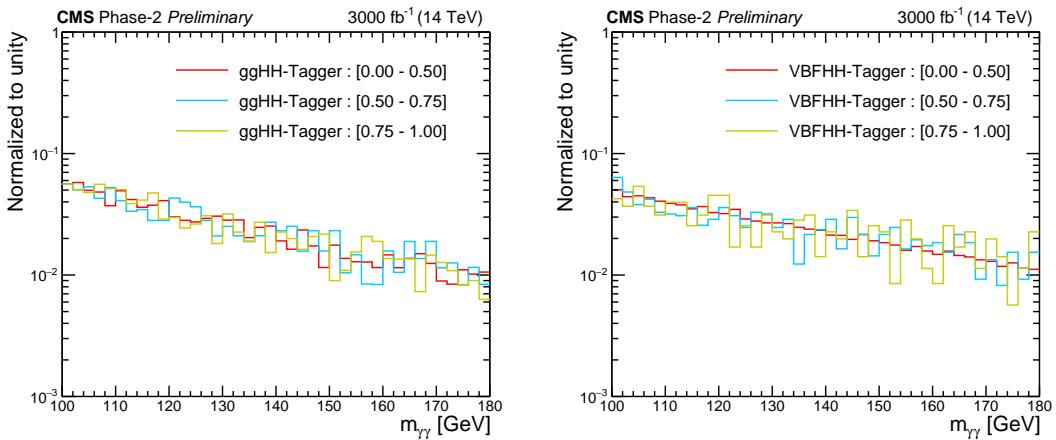
Figure 8.6: Distributions of the $m_{\gamma\gamma}$ spectrum for the γγ + jets background process for different values of the ggHH-*Tagger* and the VBFHH-*Tagger* discriminant scores.

8.4.3   Category optimization

In order to maximize the sensitivity of the analysis, the event collection is split into different categories according to the discriminant scores and the $\widetilde{M}_X$ value, as summarized in Table 8.3. The ggHH-selected events are classified into a total of six categories based on 3 ranges of $\widetilde{M}_X$: [250, 350] GeV, [350, 500] GeV and greater than 500 GeV; each of these are further subdivided into medium and high signal purity categories, depending on the ggHH-*Tagger* score. The VBFHH-selected events are classified into two high purity VBFHH categories (high VBFHH *Tagger* score): $\widetilde{M}_X < 500$ GeV and $\widetilde{M}_X > 500$ GeV. The category boundaries for both ggHH-*Tagger* and VBFHH-*Tagger* scores are optimized by maximizing the total expected significance $S/\sqrt{B}$ in each $\widetilde{M}_X$ region, where S is the number of signal and B is the number of non-resonant background events. The total significance is calculated by summing in quadrature the expected signal significance over all categories for individual $\widetilde{M}_X$ regions. It is evident that out of the total 8 categories shown in Table 8.3, 3 ggHH (ggHH CAT1, ggHH CAT3 and ggHH CAT5) and 2 VBFHH categories correspond to the higher S/B values due to the tighter thresholds of the MVA score.
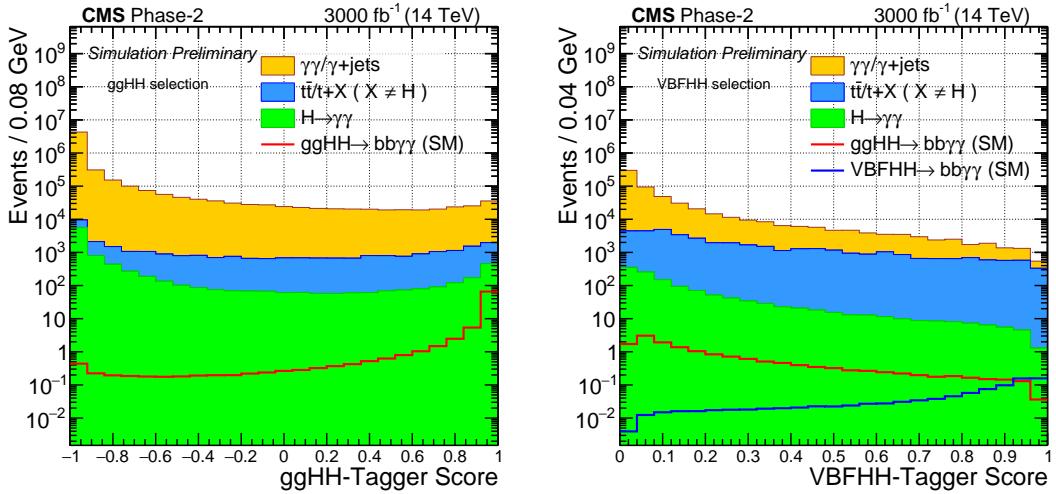
Figure 8.7: Distributions of the ggHH-*Tagger* and the VBFHH-*Tagger* discriminant scores for the ggHH- (left) and VBFHH- (right) selected events corresponding to $\mathcal{L} = 3000$ fb^{-1}; the ggHH (in red), VBFHH (in blue) signal components are overlaid on different background processes (filled stacks in different colours).

8.5 Systematic uncertainties

Two classes of systematic uncertainties are included in this study: experimental and theoretical. The experimental uncertainties considered here are assumed to be same as in Run 2 described in Ref. [122]. The reference values are quoted in Table 8.4 and Table 8.5. The theoretical uncertainties are applied for the ggHH and the VBFHH signal and all the single Higgs boson processes; the nominal values of the theoretical uncertainties correspond to the latest recommendation at $\sqrt{s} = 14$ TeV, described in the CERN Yellow Report IV [26].

8.6 Results

The extraction of the signal significance is performed via a simultaneous fit of the $m_{\gamma\gamma}$ and $m_{b\bar{b}}$ distributions. The correlation between the $m_{\gamma\gamma}$ and the $m_{b\bar{b}}$ variables for both ggHH and VBFHH signals has been checked to be negligible. For each category, the shape of the two mass distributions, $m_{\gamma\gamma}$ and $m_{b\bar{b}}$, are determined individually. The shapes of the $m_{\gamma\gamma}$ due to the HH signals (ggHH and VBFHH) and the single Higgs boson background are modeled by a Crystal Ball (CB) function in the range [115,135] GeV. However, in order to properly describe the lower tail of the $m_{b\bar{b}}$, a single CB function is not sufficient; hence a more complex

Table 8.3: Definitions of the analysis categories.

Category	$\widetilde{M}_X$ range	HH-*Tagger* range
VBFHH CAT0	$\widetilde{M}_X < 500$ GeV	VBFHH-*Tagger* > 0.95
VBFHH CAT1	$\widetilde{M}_X > 500$ GeV	VBFHH-*Tagger* > 0.98
ggHH CAT0	$\widetilde{M}_X < 350$ GeV	$0.85 <$ ggHH-*Tagger* < 0.95
ggHH CAT1		$0.95 <$ ggHH-*Tagger* < 1.00
ggHH CAT2	$350 < \widetilde{M}_X < 500$ GeV	$0.89 <$ ggHH-*Tagger* < 0.95
ggHH CAT3		$0.95 <$ ggHH-*Tagger* < 1.00
ggHH CAT4	$\widetilde{M}_X > 500$ GeV	$0.89 <$ ggHH-*Tagger* < 0.95
ggHH CAT5		$0.95 <$ ggHH-*Tagger* < 1.00

Table 8.4: Sources of experimental uncertainties and impacts on event yields (in %) estimated for the Run 2 [25] studies and used in this analysis.

Systematic unc.	Impacts on yields (%)
$m_{\gamma\gamma}$ resolution	1.0
$m_{b\bar{b}}$ resolution	5.5
Diphoton trigger efficiency	0.2
Photon energy scale	2.0
Jet energy scale	1.0
b-tag efficiency	1.0
Photon Id	0.5 / photon
$m_{\gamma\gamma}$ scale	0.3
$m_{b\bar{b}}$ scale	1.2
Luminosity	2.5

model, consisting of the sum of a CB and a Gaussian function, has been employed. The results of the fits for the ggHH signal in one of the analysis categories (high $\widetilde{M}_X$ and medium purity) are shown in Fig. 8.8. The shapes of the continuum backgrounds in the $m_{\gamma\gamma}$ and the $m_{b\bar{b}}$ distributions are modeled with separate exponential functions. The uncertainty associated with this modeling is taken into account in extracting the result for signal significance. This analysis is based only on simulation, the (pseudo-)data correspond to the total contribution from all the processes according to their cross section values, normalized to $\mathcal{L} = 3000$ fb^{-1}.

Table 8.5: Sources of theoretical uncertainties on the production cross section of HH signals and single-H backgrounds (in %) estimated for $\sqrt{s}=14\,$TeV and used in this analysis according to CERN Yellow Report IV [26].

Process	QCD scale unc.	PDF+α_s unc.
ggHH	2.1/-4.9	±3.0
VBFHH	0.3/-0.4	±2.1
ggH	4.6/-6.7	±3.2
VBF	0.5/-0.3	±2.1
VH	0.4/-0.7	±1.8
ttH	6.0/-9.2	±3.5
tHQ	6.4/-14.7	±3.6
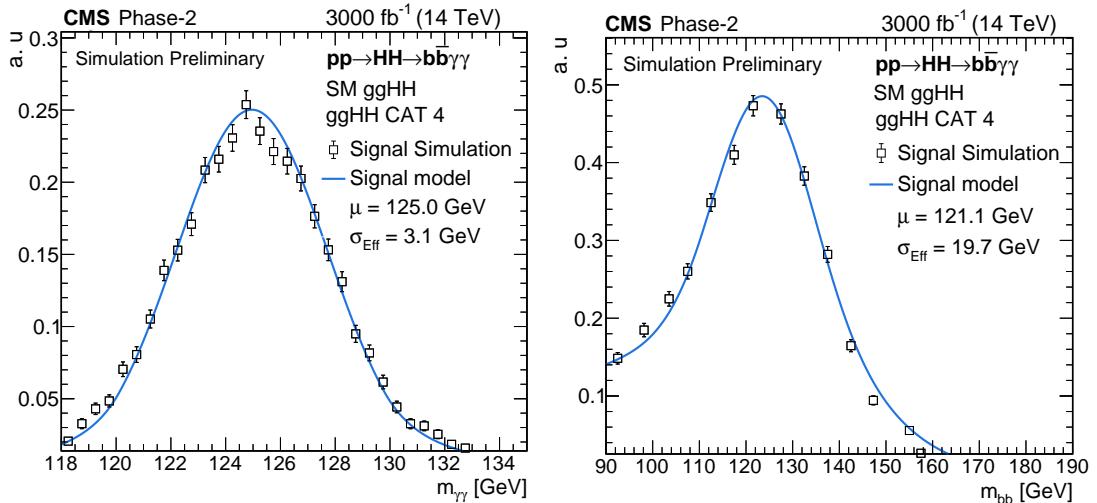
Figure 8.8: Expected distributions of the $m_{\gamma\gamma}$ (left) and the $m_{b\bar{b}}$ (right) for the ggHH signal sample in ggHH CAT4 (high $\widetilde{M}_{\mathrm{X}}$ and medium purity) at HL-LHC for $\mathcal{L} = 3000\,\mathrm{fb}^{-1}$.

Unfortunately, the available Monte Carlo simulated samples are statistically limited. This leads to considerable fluctuations in the $m_{\gamma\gamma}$ and the $m_{b\bar{b}}$ spectra. To circumvent this problem, first the probability distribution functions (PDF) are extracted by fitting the continuum distributions. The pseudo-data points are subsequently generated using toys from those PDFs for each of the analysis categories. A combined fit of the total background (single Higgs boson + continuum) is performed, as shown in Fig. 8.9. Taking into account all the eight categories simultaneously, the presence of inclusive HH signal (ggHH + VBFHH) in the standard model hypothesis is established in pseudo-data. The extracted significance for the inclusive HH signal is 2.16 stan-
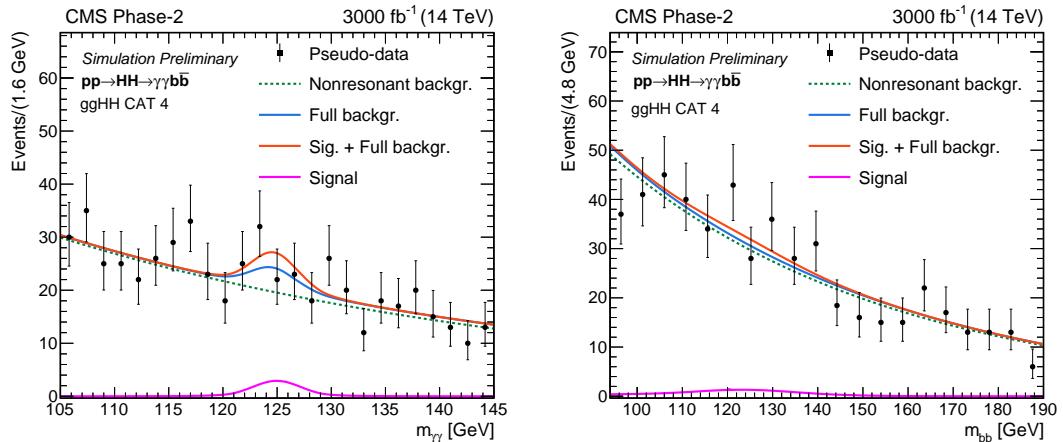
Figure 8.9: Distributions of the $m_{\gamma\gamma}$ (left) and the $m_{b\bar{b}}$ (right) for the selected pseudo-data events (black points) corresponding to $\mathcal{L} = 3000 \ \mathrm{fb}^{-1}$ shown for ggHH CAT4 (high $\widetilde{M}_X$ and medium purity) along with the expectations as estimated from the simulation. The curves correspond to continuum background only (green dashed), total background (continuum + single Higgs boson) (solid blue), and the signal + background (solid red). The signal contribution is shown in solid magenta line at the bottom of the plot.

dard deviations (σ) including systematic uncertainties. This can be compared to 1.83σ reported in the previous projection [122]. The 17% improvement in the signal significance arises from the improved strategy of tackling various types of backgrounds and modification of the selection criteria for the jets (updated p_T threshold), combined with dedicated analysis categories sensitive to VBFHH signal.

8.6.1   A rough estimation from the Run-2 result

In the Run-2 analysis we have estimated the upper limit (UL) on the inclusive HH cross section times branching fraction, and it is found to be 5.2 times the SM prediction at 95% CL level. Roughly this UL can be interpreted in terms of signal significance of about 1.96 standard deviations over the background only hypothesis.

Now if we assume that at the HL-LHC the overall efficiency, acceptance and the pileup conditions are also the same as in Run2, except the increased value of the integrated luminosity (projected luminosity at the HL LHC $\sim$ 3000 fb-1), then from a naive scaling, the expected

signal significance of the HH signal at the HL-LHC can be estimated.

$$
\begin{aligned}
Z_{Run2} &= \frac{S_{\text{Run2}}}{\sqrt{B_{\text{Run2}}}} = \frac{\sigma_{\text{SM}}.\text{UL}_{\text{Run2}}.L_{\text{Run2}}.\epsilon.A}{\sqrt{B_{\text{Run2}}}} \\[2mm]
Z_{HLLHC} &= \frac{S_{\text{HLLHC}}}{\sqrt{B_{\text{HLLHC}}}} = \frac{\sigma_{\text{SM}}.L_{\text{HLLHC}}.\epsilon.A}{\sqrt{B_{\text{HLLHC}}}} \\[2mm]
Z_{\text{HLLHC}} &= Z_{\text{Run2}} \frac{\sqrt{B_{\text{Run2}}}}{\sqrt{B_{\text{HLLHC}}}} \times \frac{L_{\text{HLLHC}}}{\text{UL}_{\text{Run2}}.L_{\text{Run2}}} \\[2mm]
Z_{\text{HLLHC}} &= Z_{\text{Run2}} \times \frac{1}{\text{UL}_{\text{Run2}}} \sqrt{\frac{L_{\text{HLLHC}}}{L_{\text{Run2}}}} \\[2mm]
Z_{\text{HLLHC}} &= 1.96 \times \frac{4.8}{5.2} \sim 1.8
\end{aligned}
\tag{8.1}
$$

Where, Z_{Run2} is the derived HH signal significance at Run2 corresponding a cross section 5.1 times the SM prediction and Z_{HLLHC} is the projected significance at the HL-LHC. $S_{\text{Run2(HLLHC)}}$ and $B_{Run2(HLLHC)}$ are the expected number of signal and background events at the Run2 (HL-LHC).The derived signal significance is well matched to our reported results in the projection study.

8.7 The sensitivity of the analysis in different pileup density condition

The instantaneous luminosity is directly coupled with the average pileup in an event. After the upgrade HL-LHC machine is expected to have 140 to 200 pileup events per bunch crossing. These values correspond to instantaneous luminosity of approximately 5 - 7.5 $\times 10^{34} \text{cm}^{-2}\text{s}^{-1}$. Due to increase of pileup conditions it deteriorates the reconstruction and identification efficiency of the objects like photon and b-jets, relevant for this analysis. This has direct effect on the sensitivity of the analysis both in terms of the signal selection efficiency and in terms of enhancement of the background contamination. The sensitivity of the HH $\rightarrow$b$\bar{\text{b}}\gamma\gamma$ analysis has been estimated with respect to the pileup event density. The pileup event density is defined as the number of pileup events per bunch crossing length along z-direction, average bunch crossing length is 20 mm. The signal significance of HH $\rightarrow$b$\bar{\text{b}}\gamma\gamma$ process has been estimated for different pileup density scenario and is given in Tab. 8.6. The result of this study is reported in Ref. [135].

Table 8.6: Signal significance of HH $\rightarrow b\bar{b}\gamma\gamma$ in standard deviation for different pileup density scenario.

PU density (events/mm)	Run 2 : 0.33	1.0-1.5	1.5-2.0	2.0-3.0	3.0-4.0
Significance (in σ)	–	2.18	2.16	2.12	2.06

8.8 Summary

The expected signal sensitivity for the standard model production of non-resonant Higgs boson pair in the inclusive $b\bar{b}\gamma\gamma$ final state is determined to be 2.16 standard deviation. The study has been carried out using DELPHES fast simulation of the Phase-2 upgraded CMS detector and corresponding to an average pileup scenario of 200 and integrated luminosity of 3000 fb^{-1} of b-b collisions at $\sqrt{s}$ = 14 TeV. The study has also been extended for the different pileup density conditions and it has been found that the analysis sensitivity reduces with higher pileup density. Anticipating better analysis techniques and reduced systematic uncertainties, it is expected that the combination of various final states will pave the way for the CMS and ATLAS experiment towards a 5σ discovery for the Higgs boson pair production at the HL-LHC.

Chapter 9

Measurement of the Higgs boson production via vector boson fusion and decaying to bottom quarks

9.1 Introduction

The SM Higgs boson with mass of 125 GeV decays most frequently into a bottom quark-antiquark ($b\bar{b}$) pair with a branching fraction of about 58% [136, 137]. However, it is challenging to explore this decay mode experimentally. Further, in the dominant gluon-gluon fusion (ggH) production mode the H $\rightarrow$ b$\bar{b}$ signal is overwhelmed by background consisting of b$\bar{b}$ pairs produced through the strong interaction, referred to as quantum chromodynamics (QCD) induced multijet events. At the LHC, a moderate sensitivity to the H $\rightarrow$ b$\bar{b}$ decay in the ggH process can be achieved by exploiting boosted production of the Higgs boson [138]. The most promising production mechanism to study the H $\rightarrow$ b$\bar{b}$ decay is the Higgs boson production in association with a leptonically decaying Z or W boson (VH). Although VH production cross section is much less than an order of magnitude than that of ggH, leptonic decays of W and Z boson provide a handle to reduce backgrounds also larger trigger efficiency, thereby making the H $\rightarrow$ b$\bar{b}$ decay accessible for detection. The VH production mode has contributed with the largest sensitivity to the observation of the H $\rightarrow$ b$\bar{b}$ decay by the CMS and ATLAS Collaborations using collision data at $\sqrt{s} = 13$ TeV. The ATLAS Collaboration has measured the H $\rightarrow$ b$\bar{b}$ signal yield relative to the SM prediction to be $\mu_{\mathrm{Hb\bar{b}}}^{\mathrm{incl.}} = 1.02 \pm 0.12\,\mathrm{(stat)} \pm 0.14\,\mathrm{(syst)}$,

corresponding to a significance of 6.7 standard deviations (σ) [139]. The measurement by the CMS collaboration is $\mu_{\mathrm{Hb\bar{b}}}^{\mathrm{incl.}} = 1.04 \pm 0.14\,(\text{stat}) \pm 0.14\,(\text{syst})$ and corresponds to a significance of $5.6\,\sigma$ [47]. This analysis considers an alternative process vector boson fusion (VBF) for the detection of the $H \rightarrow b\bar{b}$ signal. The VBF production of the Higgs boson (VBFH) has the second-largest cross section at the LHC and attracts particular attention because it involves large momentum transfer and provides a sensitive probe of momentum-dependent anomalous coupling [140]. The $H \rightarrow b\bar{b}$ decay from the VBFH production, $qqH \rightarrow qqb\bar{b}$, gives rise to a four-jet final state as depicted in Fig. 9.1. Two of the jets, from the $H \rightarrow b\bar{b}$ decay, typically lie in the central region of the detector. The other two jets, from the light quarks, are produced mainly in the forward and backward directions relative to the beam line and, consequently, have a large rapidity seperation between them as well as the high dijet invariant mass. We refer to the latter as VBF jets. As the reaction proceeds via the exchange of colorless particles (W and Z bosons), the color connection between outgoing light quarks is suppressed, leading to a relatively small amount of hadronic activity in the rapidity interval between the VBF jets and b-tagged jets originating from the Higgs boson decay. These distinct features allow for the suppression of the large QCD induced multijet background and identification of the signal process.
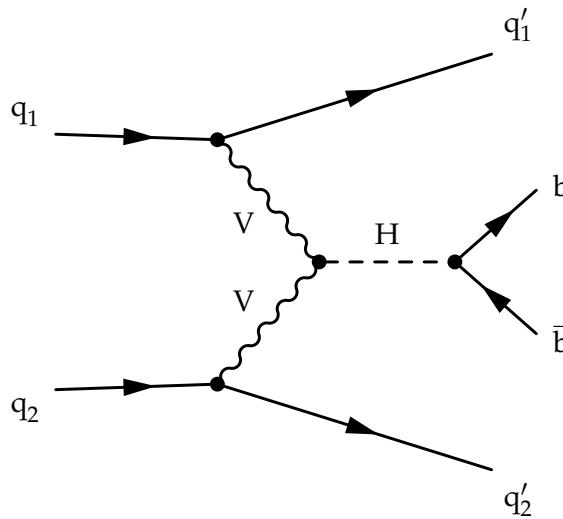
Figure 9.1: Representative Feynman diagram of the leading order VBF production of a Higgs boson followed by its decay to a pair of b quarks.

The previous measurement of the $qqH \rightarrow qqb\bar{b}$ process by the CMS Collaboration is based on a data set of proton-proton (pp) collisions at $\sqrt{s} = 8\,\text{TeV}$ and corresponding to an integrated luminosity of about $20\,\text{fb}^{-1}$ [141]. The signal strength was observed to be $\mu_{\mathrm{Hb\bar{b}}}^{\mathrm{qqH}} = 2.8_{-1.4}^{+1.6}$ with

a significance of $2.2\,\sigma$, while the expected significance was $0.8\,\sigma$. The ATLAS Collaboration recently reported a measurement of the $\mathrm{qqH} \to \mathrm{qqb\overline{b}}$ process using $127\,\mathrm{fb}^{-1}$ of data collected in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ [142] and the signal strength was measured to be $\mu^{\mathrm{qqH}}_{\mathrm{Hb\overline{b}}} = 0.95^{+0.38}_{-0.36}$, corresponding to an observed (expected) significance of $2.6\,(2.8)\,\sigma$. In this work, we present an updated CMS measurement of the $\mathrm{qqH} \to \mathrm{qqb\overline{b}}$ process using data collected at $\sqrt{s} = 13\,\mathrm{TeV}$.

9.1.1 Kinematic properties of the signal process

As visible from the Feynman diagram of the VBF process in Fig. 9.1, a valence quark of each one of the colliding protons radiates a W or Z boson that subsequently interact (via their longitudinal components) with each other to produce H. So, for $\mathrm{H} \to \mathrm{b\overline{b}}$ decay channel in the final state consists of two light quarks (q_1, q_2) and two b-quarks (b_1, b_2). In this way the two valence quarks are typically scattered away at a reasonably small angle with respect to the beam line. The resulting jets can be mostly detected within the detector acceptance. The prominent signature of the VBF production is therefore characterized by the two energetic light quark jets, roughly in forward and backward directions, separate moderately in pseudorapidity (η) space. The topology of the two VBF jets are utilized in many way in this analysis. The two b jets originating from the bottom quark antiquark pair are mostly in the central region of the detector.

Some of the kinematic properties of the Higgs boson of the signal qqH sample at the generator level are shown in Fig. 9.2 comparing with the ggH production. The transverse momentum and pseudorapidity of the light-quarks and the b-quarks form the H decay are presented in Fig. 9.3, as already mentioned that the light-quarks have only moderate transverse momentum leading to the interesting shape in the η distributions. The b quarks are produced centrally with reasonably high transverse momentum.

The kinematics of the VBF jets are shown in Fig. 9.4 at the parton level. As the two forward-backward quarks are scattered almost back to back, they have high longitudinal momentum. This leads to their invariant mass being high as well as large pseudorapidity gap between them. These distinct features are utilized in our analysis at the trigger level for on-line selection and also at the off-line level with the reconstructed jets.
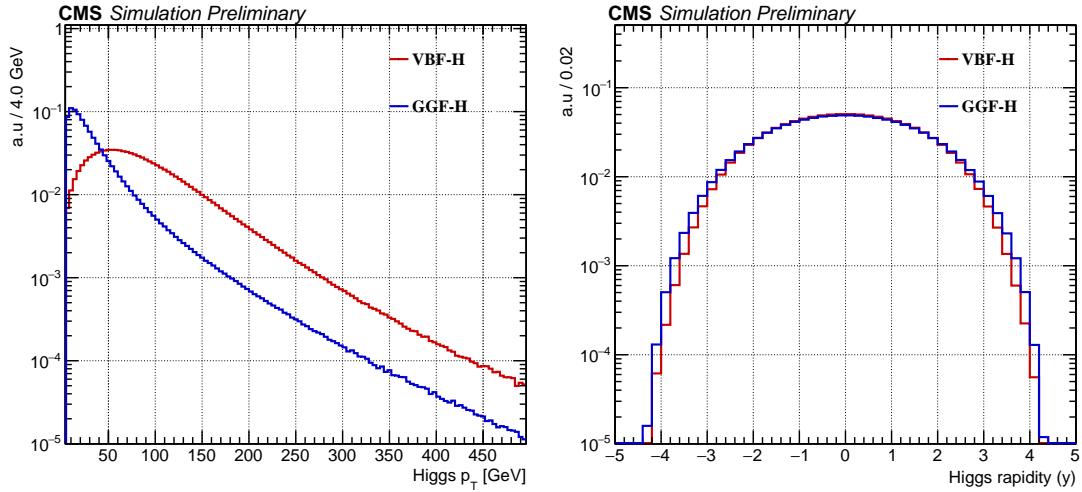
Figure 9.2: Transverse momentum p_T (left) and rapidity (right) distributions of the Higgs boson at the partonic state for VBF signal sample (in blue) and ggH process (in Red).
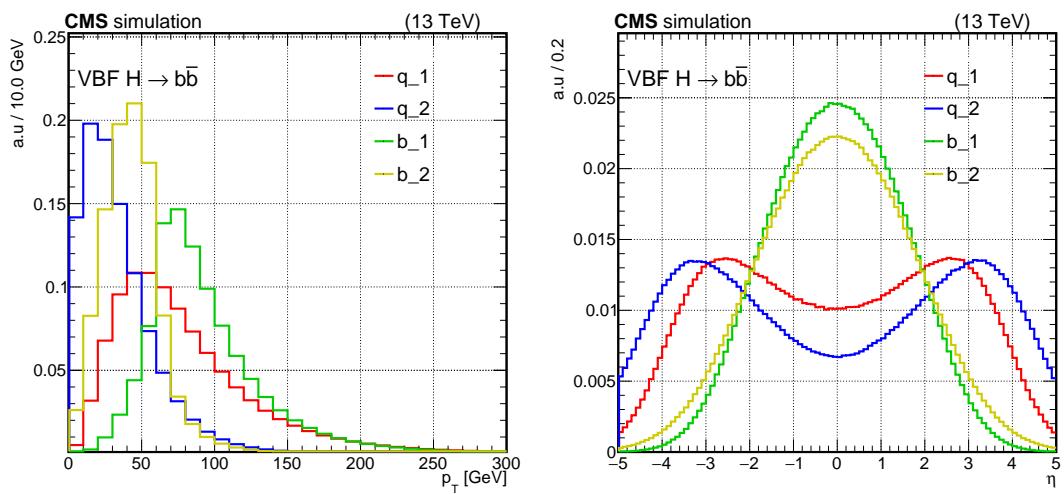
Figure 9.3: Transverse momentum p_T (left) and pseudorapidity (right) distributions of the four final state quarks of the VBF signal sample without any selection cut; here the leading light quark (in red), sub-leading light quark (in blue), leading b-quark (in green) and sub-leading b-quark (in yellow) are denoted as q_1, q_2, b_1 and b_2 respectively.
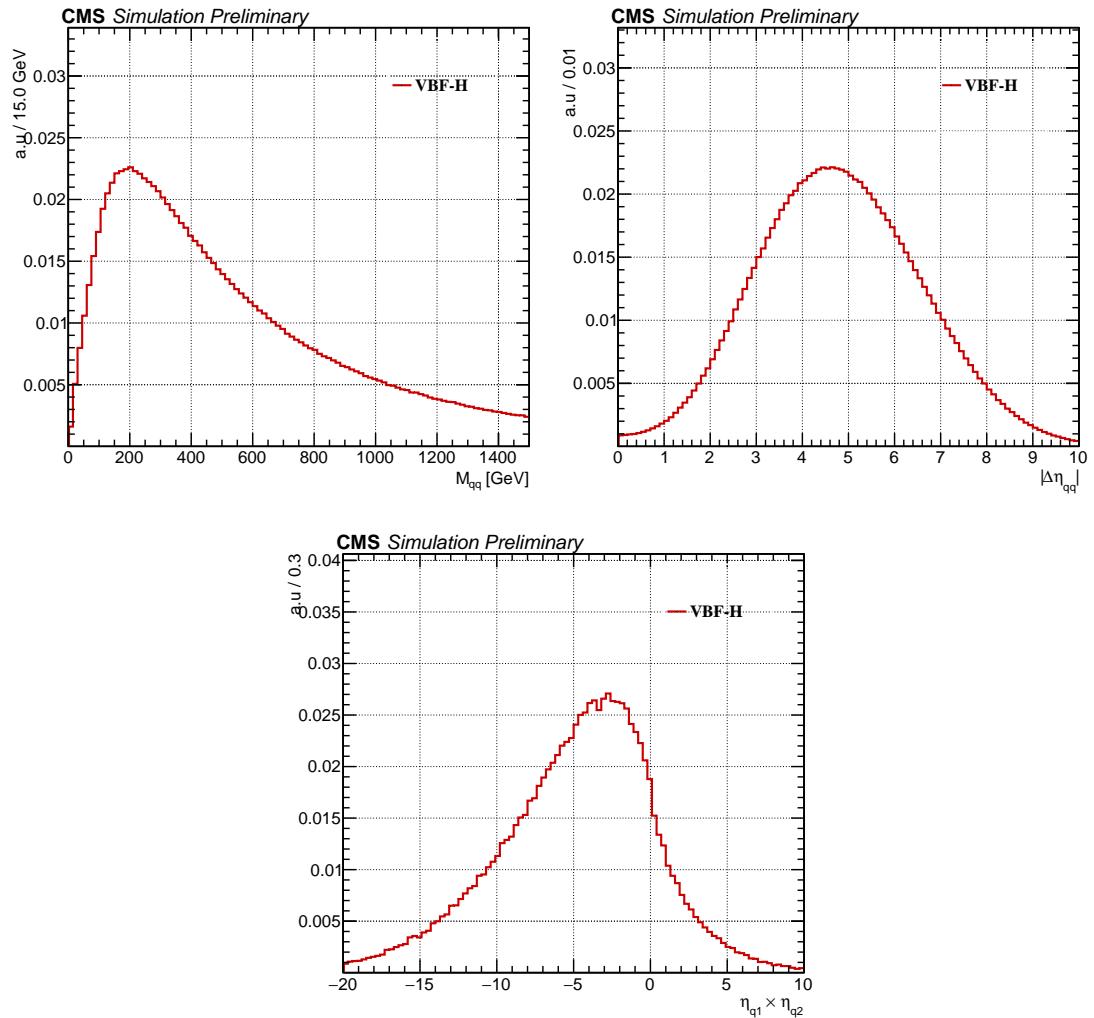
Figure 9.4: The invariant mass (M_{qq}), absolute value of the difference of pseudorapidity ($|\Delta\eta_{qq}|$) and product of pseudorapidity ($\eta_{q1} \times \eta_{q2}$) of the two forward-backward quarks.

Study with $\Delta\phi_{b\bar{b}}^-$

The VBFH events are produced with a topology of H in the central region and along the two forward-backward quarks. The azimuthal angle separations ($\Delta\phi$) between the three objects H, q1 and q2 have been shown in Fig. 9.5, 9.6 to visualize the angular correlation among them at the hard-scatter level in the transverse plane perpendicular to the beam direction.

Figure 9.5 shows the 2-dimensional distribution of $\Delta\phi$ between the H with the scattered quarks
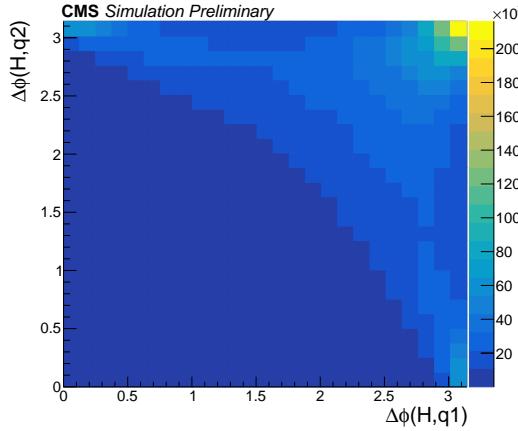
Figure 9.5: 2-dimensional distribution of $\Delta\phi_{H,q1}$-$\Delta\phi_{H,q2}$, H candidate is found to be opposite to the both VBF quarks in transverse plane for the maximum cases.
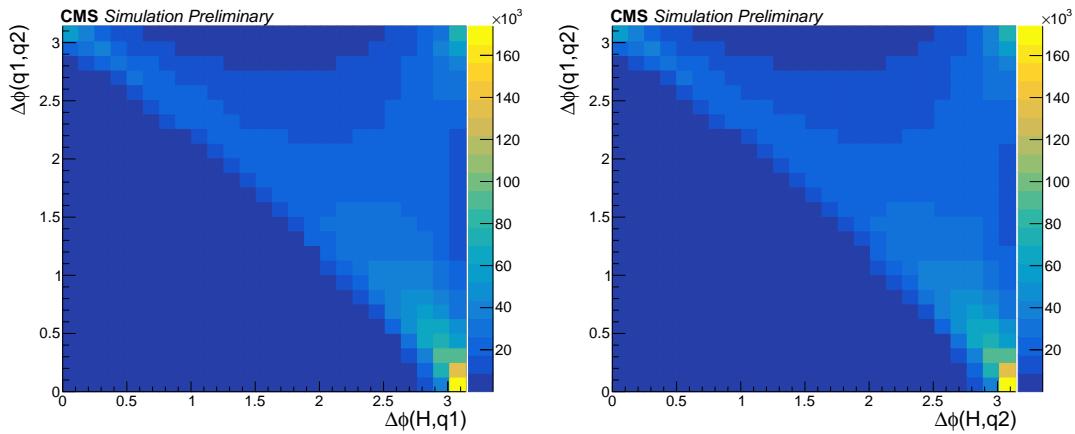
Figure 9.6: 2-dimensional distribution of $\Delta\phi_{H,q1/q2}$ and $\Delta\phi_{q1,q2}$.

($\Delta\phi_{H,q1}$,$\Delta\phi_{H,q2}$). It is observed that for most of the events, H is laying almost opposite to both the two quarks ($\Delta\phi_{H,q} \sim \pi$). Hence these two VBF quarks are close to each other in the transverse plane corresponding to small $\Delta\phi$ separation. This is shown in Figs. 9.6, where the 2-dimensional distributions of the $\Delta\phi$ between the H and any of the VBF quarks is presented

with the $\Delta\phi$ between the two VBF quarks ($\Delta\phi_{q1,q2}$). The event topology can now be understood by a cartoon in Fig. 9.7, where the H is balanced by the two VBF quarks.
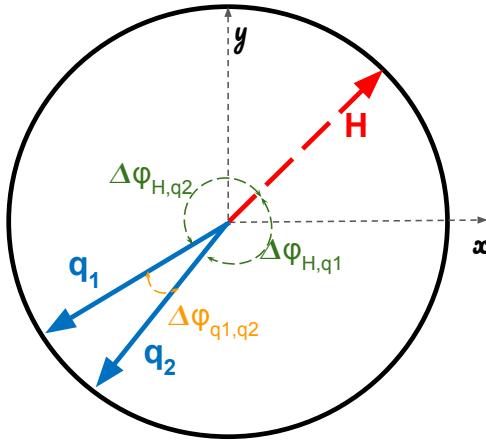
Figure 9.7: A cartoon of VBF Higgs production process in the transverse plane, where the Higgs boson is balanced by the two VBF quarks.

Since the VBF quarks have small angular separation in transverse plane, the resultant transverse momentum of these quarks (p_{T}^{qq}) is almost equal to the scalar p_{T} sum of the individual VBF quarks, as shown in Fig. 9.8 (left). Figure. 9.8(right) shows that the transverse momenta of the H is balanced by the p_{T}^{qq}, as described above.
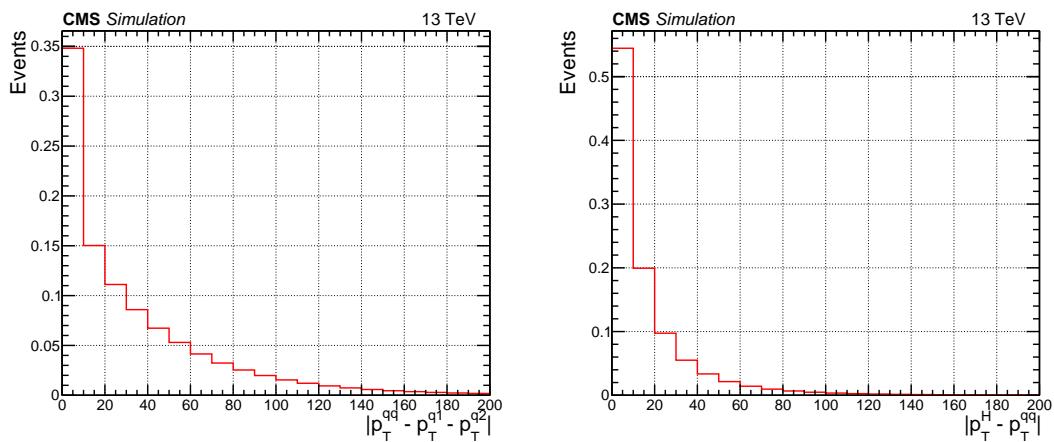
Figure 9.8: The quantities $|p_{\mathrm{T}}^{qq} - p_{\mathrm{T}}^{q1} - p_{\mathrm{T}}^{q2}|$ (left) and $|p_{\mathrm{T}}^{H} - p_{\mathrm{T}}^{qq}|$ (right) at the hard-scatter level

In general, the transverse momentum of the H produced through the VBF production mode is relatively small, hence the angular separation in the azimuthal (ϕ) plane, of the two b-quarks decaying from the H is large. Figure. 9.9 (left) shows the unit normalized distribution of the

Higgs boson transverse momentum (p_T^H) at the hard-scatter level, while Fig. 9.9 (right) presents the $\Delta\phi_{b\bar{b}}$. As expected, the quantity $\Delta\phi_{b\bar{b}}$ is decreasing with the increment of p_T^H, and shown in a 2-dimensional representation in Fig. 9.10.
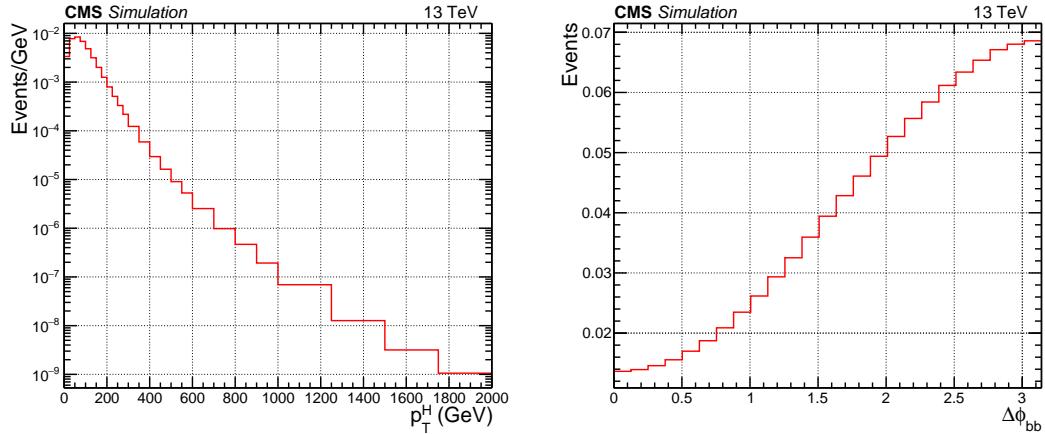
Figure 9.9: Left: Normalized distribution of the Higgs boson transverse momentum produced through the VBF mode at the hard-scatter level. Right: distribution of the separation in azimuthal angle of the two b-quarks from the H decay.
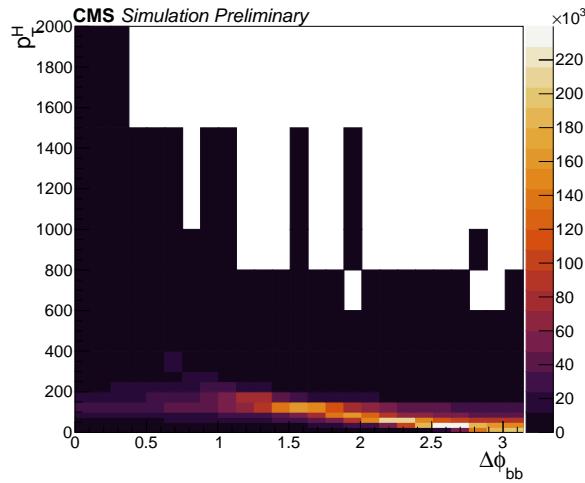
Figure 9.10: 2-dimensional distribution of $\Delta\phi_{b\bar{b}}$ with the Higgs boson transverse momentum (p_T^H).

Due to the large production rate of QCD-multijet processes at the LHC, the trigger system of the CMS experiment imposed high p_T threshold requirements on the jets both at L1 and HLT levels to control the event acceptance rate. On average, the threshold values of the quad-jet triggers are about 100, 80, 70 and 30 GeV during the two years 2016 and 2018. The actual trigger conditions used in the analysis has been demonstrated in Sec. 9.4. It has been found

that the p_T threshold requirements on the reconstructed jets has a direct impact on p_T^H. From the Fig. 9.11 (left), it is also evident that the applied trigger p_T thresholds allow to reconstruct a slightly boosted Higgs boson, and the separation in the azimuthal angle of its decay products (b-quarks) is small, as shown in Fig. 9.11 (right).
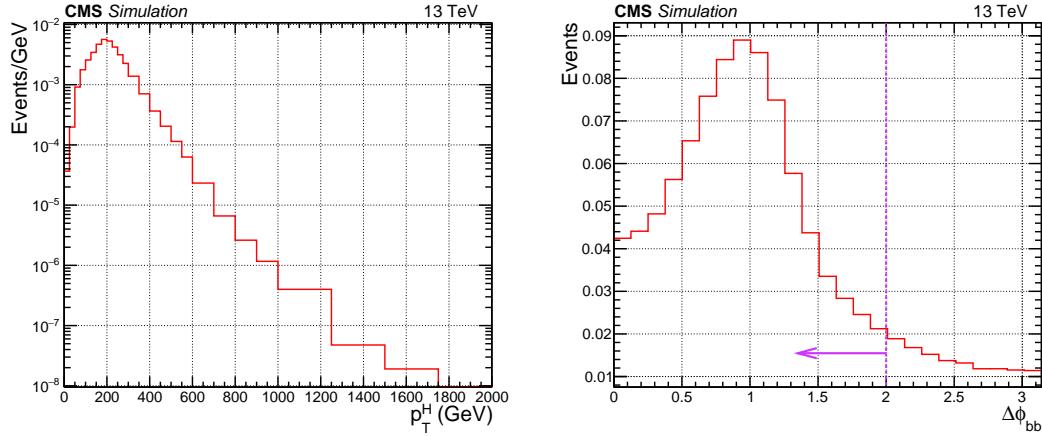
Figure 9.11: Left: Normalized distribution of the Higgs boson transverse momentum produced through the VBF mode at the hard-scatter level and the separation in azimuthal angle of the two b-quarks from the H decay after the quad-jet p_T threshold requirements.

This inspires to impose additional requirements of $\Delta\phi_{b\bar{b}}$ less than around 2.0 at the HLT level to reduce the event acceptance rate for offline analysis.

9.2 Analysis flowchart

The flowchart of the qqH → qqb$\overline{\text{b}}$ analysis using Run 2 data, describing the different steps are shown in Fig. 9.12.
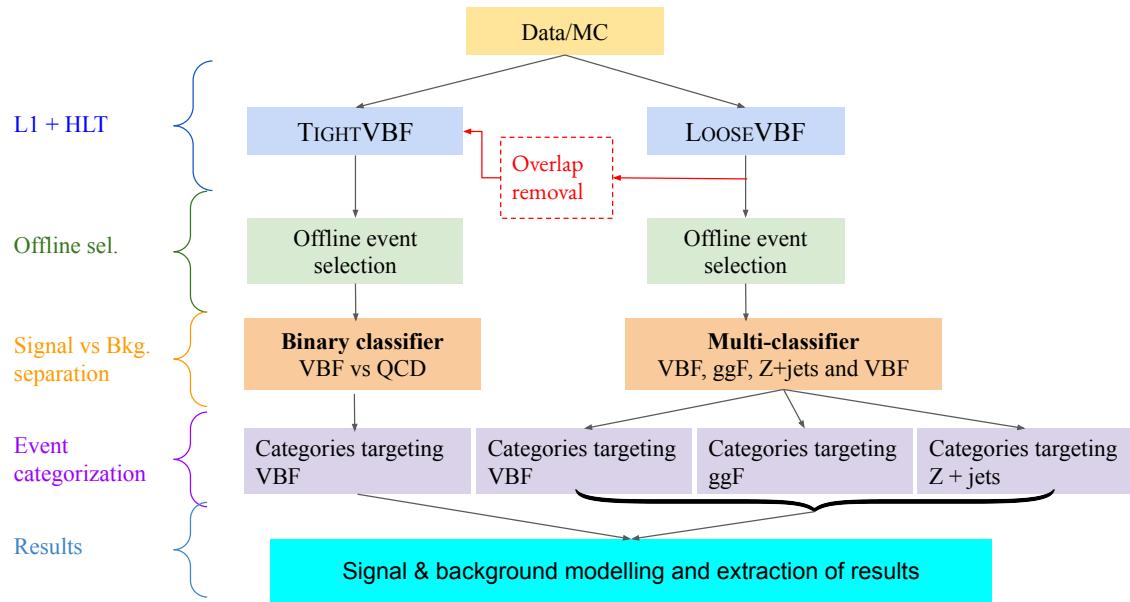
Figure 9.12: Analysis flowchart of the qqH → qqb$\overline{\text{b}}$ analysis with Run 2 data.

9.3   Dataset and simulated samples

This analysis uses data collected by the CMS experiment in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ and corresponding to an integrated luminosity of $90.8\,\text{fb}^{-1}$. The total analyzed data volume comprises two sets recorded under different experimental conditions. The first set was collected in 2016 and corresponds to $36.3\,\text{fb}^{-1}$. The second set was collected in 2018, when the LHC delivered pp collisions with higher instantaneous luminosity compared to 2016, and corresponds to $54.5\,\text{fb}^{-1}$. No trigger path suitable for this analysis was available during the data taking period in 2017.

The analysis relies on the simulated Monte Carlo (MC) samples for estimation of the signal acceptance and efficiency. Additionally, the contributions from various subdominant background processes are determined from simulation.

The $\text{qqH} \rightarrow \text{qqb}\bar{\text{b}}$ signal is generated using POWHEG v-2.0 at next-to-leading order (NLO) accuracy in the QCD coupling constant α_S [143]. A dipole parton shower model [144] is used for modeling of the initial-final state color flow that takes into account the color connection between the incoming and the outgoing partons. An alternative $\text{qqH} \rightarrow \text{qqb}\bar{\text{b}}$ sample is also prepared using the POWHEG matrix element generator interfaced with HERWIG 7 [145] for fragmentation and hadronization. This sample is used only to assess the systematic uncertainty related to the choice of the showering and hadronization model. The ggH production of the Higgs boson with at least two accompanying jets has a non-negligible contribution to the kinematic phase space considered in this analysis. The ggH process is generated using the MiNLO event generator at next-to-NLO (NNLO) [146, 147] precision in α_S [98] including finite top quark mass effects. Contributions from the weak gauge boson (V = W/Z) associated (VH) and top-quark-pair associated ($\text{t}\bar{\text{t}}\text{H}$) production of the Higgs boson are found to be negligible.

The dominant continuum background for this analysis is QCD induced multijet production. To assess properties of these events and validate the analysis strategy, QCD multijet events are generated with MADGRAPH5_AMC@NLO [95] at leading order (LO) precision in α_S. The matrix element is matched to the hadron showers generated by PYTHIA using the MLM prescription [148].

The dominant resonant background is the inclusive production of Z bosons (Z+jets), where about 70% of the time Z decays via a quark-antiquark pair of the same flavor, including bottom.

Hence, the main component of the resonant background in the event sample corresponds to the $Z \to b\bar{b}$ decay mode, although there are contributions from charm quark, light quark, and gluon jets, which can be being misidentified as b jets. There are two different mechanisms of inclusive Z boson production: via QCD induced or pure electroweak (EWK) processes, as depicted in the representative Feynman diagrams in Fig. 9.13.
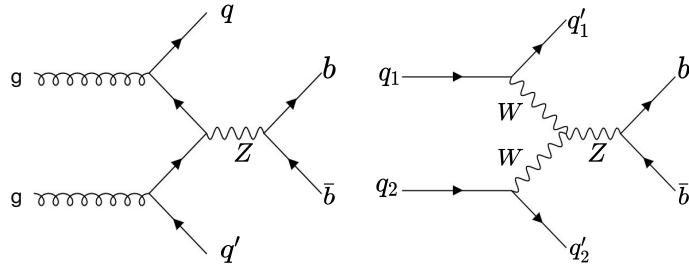
Figure 9.13: Representative Feynman diagrams of the Z + 2jets process produced from QCD induced (left) and pure electroweak (right) modes and subsequent decay of Z boson to a pair of bottom quarks .

Though the latter has a much lower rate, the event topology is the same as the signal. Similarly, W bosons can be produced in association with jets through a fusion of quark-antiquark pairs or vector bosons. The inclusive W and Z samples are generated using MADGRAPH5_AMC@NLO at LO precision in α_S. For the QCD induced Z+jets sample, correction factors have been applied to match the generator-level p_T distributions from the analytical predictions available with the highest order accuracy in the perturbative expansion [149]. In fact two individual correction factors are used for each p_T bin; first to emulate the spectrum predicted by QCD NNLO effect and then further re-weighting was done to incorporate the higher-order electroweak effects.

Other important background contributions in the signal region arise from inclusive single top quark ($t/\bar{t}+X$) and top quark-antiquark pair ($t\bar{t}+X$) productions. These are modeled by POWHEG with the NLO QCD accuracy. For the $t\bar{t}$ process all possible, combinations of the decay modes of the two W bosons from the top quark and antiquark decays are considered leading to hadronic, leptonic, and semileptonic final states.

All the simulated samples are interfaced with PYTHIA 8.212 [97] for parton showering and fragmentation with the standard p_T-ordered parton shower scheme. The underlying event is modeled with PYTHIA, using the CP5 tune [105] for both of the years. The parton density functions (PDFs) are taken from the sets of NNPDF3.0 [106] for 2016 samples and NNPDF3.1 [107] for

2018 samples. The response of the CMS detector is modeled using the GEANT4 [112] package. The event reconstruction is performed with the same algorithms as are used for data. Additional pileup interactions in each bunch crossing are generated with PYTHIA and added to the simulated samples following a Poisson distribution with the mean value determined in data. The simulated events are weighted such that the pileup distribution in the simulation matches the one observed in data.

9.4   Triggers

9.4.1   L1 and HLT trigger

Events are selected with dedicated L1 and HLT selections optimized separately for 2016 and 2018. At the L1 stage, events are required to have at least three jets with p_T above certain thresholds that were optimized according to the instantaneous luminosity. The p_T thresholds of 90, 76, and 64 GeV were imposed for 2016 and 100, 80, and 70 GeV for 2018. The presence of a fourth jet is not required at the L1 stage.

An event is accepted by the HLT if it contains at least four jets reconstructed online with the PF algorithm. Jets are required to have p_T greater than 92, 76, 64, and 16 (105, 88, 76, and 15) GeV in 2016 (2018). Two complementary online requirements (HLT paths), as explained below, are also implemented to select events in each of the two sets. In the following we refer to these HLT paths as HLT_TIGHT and HLT_LOOSE.

The HLT_TIGHT path selects events with at least one b-tagged jet among four leading jets. The working point chosen for the online b tagging in the HLT_TIGHT path corresponds to b jet efficiency of roughly 55 (60)% and a misidentification rate for light quark and gluon jets of 2 (1)% in 2016 (2018). If more than one jet is tagged as a b jet the one with the highest b-tag discriminant is selected and out of the three remaining jets, a jet pair with the largest $\Delta\eta_{jj}$ is chosen as the pair of VBF-tagged jets. The HLT_TIGHT path imposes very stringent conditions on the VBF-tagged jets: they must be separated in pseudorapidity by $|\Delta\eta_{jj}| > 4.1$ (3.5) and have invariant mass $m_{jj} > 500$ (460) GeV in 2016 (2018). For the other two jets, which are assigned to the H $\rightarrow$ b$\bar{\text{b}}$ decay, the separation in azimuthal angle is required to be $\Delta\phi_{b\bar{b}} < 1.6$ (1.9) radian for 2016 (2018).

In contrast to the HLT_TIGHT path, the HLT_LOOSE path selects events with at least two b-

tagged jets but imposes comparatively lenient requirements on the VBF-tagged jets: $m_{jj} > 240$ (200) GeV and $|\Delta\eta_{jj}| > 2.3$ (1.5) for 2016 (2018). The azimuthal separation of b-tagged jets is required to be $\Delta\phi_{b\bar{b}} < 2.1$ (2.8) rad in 2016 (2018). The efficiencies to select the signal with the HLT_Tight and HLT_Loose paths are 3.1 (2.3) and 3.5 (2.5)% in 2016 (2018), respectively. A summary of the trigger conditions used in the analysis is mentioned in Tab. 9.1.

Table 9.1: The HLT selection requirements in the four analyzed samples.

Requirements	2016 (36.3 fb^{-1})		2018 (54.5 fb^{-1})	
	Tight	Loose	Tight	Loose
p_T thresholds	92, 76, 64, 16 GeV		105, 88, 76, 15 GeV	
Number of b tags	≥ 1	≥ 2	≥ 1	≥ 2
$\Delta\phi_{b\bar{b}}$ (rad)	≤ 1.6	≤ 2.1	≤ 1.9	≤ 2.8
$|\Delta\eta_{jj}|$	≥ 4.1	≥ 2.3	≥ 3.5	≥ 1.5
m_{jj}	≥ 500 GeV	≥ 240 GeV	≥ 460 GeV	≥ 200 GeV

The overlap removal is applied at the HLT level, if an event passes both HLT_Tight and HLT_Loose HLT requirements, it will be assigned only to the HLT_Tight. A cartton of such HLT overlap removal step is shown in Fig. 9.14
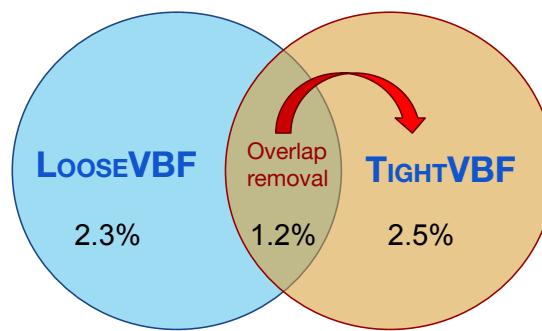
Figure 9.14: A cartoon of overlap removal of events at the HLT step, the numbers mentioned in the plots are the corresponding efficiencies for the 2018 HLT paths.

9.4.2 Trigger p_T scale factors

Because of mis-modeling of different effects (detector subsystems response, reconstruction algorithm performance, etc.) the actual efficiency of particular HLT path is different for MC-simulation and experimental data. To provide correct responses in Data and MC, these differences should be estimated and corrected by introducing HLT scale factors (SF).

There are four jet p_T-threshold requirements at HLT level for QuadJet path mentioned above. While for the first three thresholds there are corresponding requirements at L1 level trigger, 4th jet p_T-threshold requirement is applied only at the HLT. The trigger p_T scale factors are calculated with strict identification and matching of the HLT trigger objects to their corresponding trigger objects at L1 (called L1 seed) stage. The scale factors are calculated for each individual jet employed in the HLT path, using tag-and-probe method.

Events are selected with single jet HLT trigger p_T threshold of 60 (80) GeV with an additional requirement on leading p_T offline jet to have $p_T > 110$ (130) GeV and $|\eta| < 2.2$; if this jet matches to the trigger object used at the L1, it is refereed to as the tagged jet. Now in the same event a subleading p_T jet is considered as the probe jet. To ensure that the event has di-jet topology, the following additional conditions are imposed: the difference of the azimuthal angle between the tag and probe jets should have $|\Delta\phi| > 2.5$ and the p_T of the third jet should be smaller than 30% of the mean p_T of the first two jet, $p_T^{j3} < 0.3 \left(\frac{p_T^{j1} + p_T^{j2}}{2} \right)$. If a probe jet is able to pass the HLT p_T threshold and have a macthed L1 , Calo and HLT jets with in $\Delta R < 0.4$, it is called matched-probe. The efficiency is now defined as:

$$\text{Efficiency} = \frac{\text{no of events with tag jet and a matched-probe jet}}{\text{no of events with tag jet and a probe jet}} \tag{9.1}$$

This procedure repeated by changing the p_T thresholds according to the HLT requirements for the first three jets. For the fourth jet it is assumed that the efficiency is 100% as there is no restrictions at the L1 level. Efficiency has been measured in data and in QCD multi-jet MC simulated event sample following the same strategy mentioned above. The difference of efficiency between the data and MC for each particular p_T threshold has been treated as the Trigger SF, Finally the total SF is:

$$\text{Trigger SF} = \text{SF}_1 \times \text{SF}_2 \times \text{SF}_3 \tag{9.2}$$

The trigger efficiency from the data and the QCD events and the corresponding SF for the leading HLT jet in four different η regions for 2018 are shown in Fig. 9.15.

9.4.3   Validation of trigger SF

Validation of the derived trigger scale factors (p_T dependent) are checked with the use of a control `QuadPFJet` HLT triggers with the same p_T thresholds but without any b-tagging and VBF requirements. Recorded integrated luminosity corresponding to these paths at 2016 and 2018 data taking are about 31.8 pb^{-1} and 108.4 pb^{-1} respectively.

The data and QCD MC events (contribution of other processes is negligible) are first selected using the control HLT trigger and subsequently offline jet selection conditions are applied. To check the impact of the trigger SF, MC events are weighted as per event basis with the SF derived by Eq. 9.2 and a comparison has been made before and after the application of the SF which is shown in Fig. 9.16. From the data MC comparison before and after the application of trigger SF, it is evident that the introduction of HLT trigger p_T scale factors can fix the existing disagreement of kinematic variables distributions between the data and MC, more particularly in the low-p_T regions.
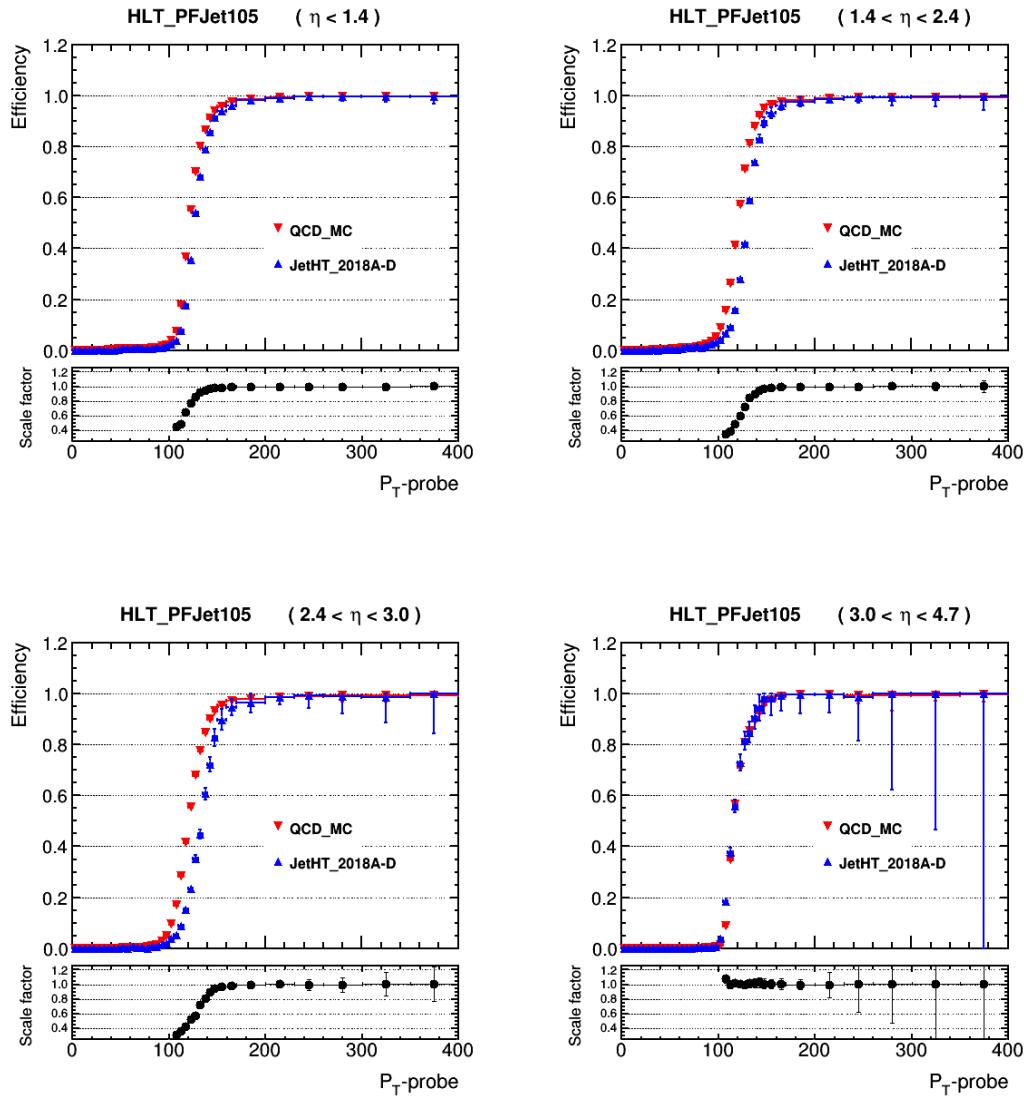
Figure 9.15: Upper panel: `HLT_Jet105` trigger efficiency as a function of "Probe"-jet p_T for MC QCD and 2018 collision dataset in various η-regions. Lower panel shows the ratio of these efficiencies (data to MC) as HLT scale factors.
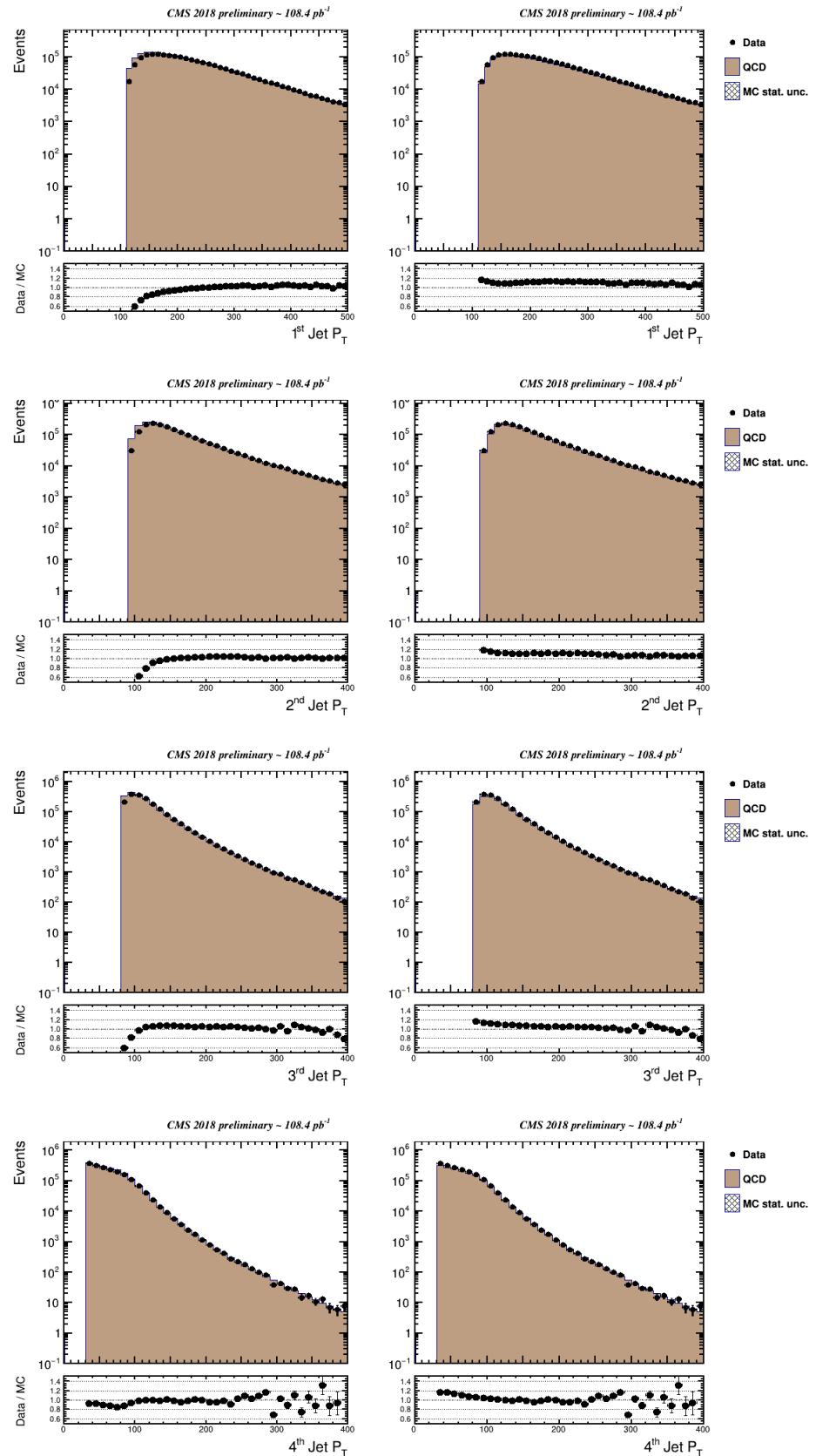
Figure 9.16: The p_T distribution of the four p_T leading jets in Data (in black) & QCD MC events (gray filled area) before (left) and after (right) applying trigger p_T scale factors. The lower panel of each plot shows the agreement between data & MC.

9.5   Offline event selection

An event is discarded if it contains any isolated muon or electron identified using a loose working point, corresponding to a selection efficiency of 95% and a misidentification rate of 1–2%.This requirement suppresses contributions from inclusive top pair and single top production processes events with leptonic decays of the W bosons from the cascade decay of the top quarks. The event is selected if it contains four jets with p_T greater than 95, 80, 65 and 30 GeV in 2016 and 110, 90, 80, 30 GeV in 2018 following the HLT p_T thresholds described in Sec. 9.4.2. At least two jets are required to be b-tagged with the DEEPJET algorithm. Out of the four p_T-leading jets, the two jets with the highest DEEPJET b tag scores are used to reconstruct the H $\rightarrow$ b$\bar{\text{b}}$ decay candidate. After the Higgs boson candidate is selected, the two remaining jets are considered as VBF-tagged jet candidates. The jets with $p_T < 50$ GeV must also pass an identification criterion designed to reduce the number of selected jets originating from pileup interactions [150].

Depending on the data-taking period and the fired trigger, events are split into four non-overlapping samples that are analyzed independently. Events from 2016 (2018) that pass HLT_TIGHT path are assigned to the TIGHT 2016 (TIGHT 2018) sample. Events from 2016 (2018) that fail HLT_TIGHT path but pass HLT_LOOSE path constitute the LOOSE 2016 (LOOSE 2018) sample. The offline selection criteria applied to the VBF- and b-tagged jets in each sample are:

- TIGHT 2016: $\Delta\phi_{\text{b}\bar{\text{b}}} < 1.6$ rad, $|\Delta\eta_{\text{jj}}| > 4.2$, $m_{\text{jj}} > 500$ GeV;

- LOOSE 2016: $\Delta\phi_{\text{b}\bar{\text{b}}} < 2.1$ rad, $|\Delta\eta_{\text{jj}}| > 2.5$, $m_{\text{jj}} > 250$ GeV;

- TIGHT 2018: $\Delta\phi_{\text{b}\bar{\text{b}}} < 1.6$ rad, $|\Delta\eta_{\text{jj}}| > 3.8$, $m_{\text{jj}} > 500$ GeV; and

- LOOSE 2018: $\Delta\phi_{\text{b}\bar{\text{b}}} < 2.1$ rad, $|\Delta\eta_{\text{jj}}| > 2.5$, $m_{\text{jj}} > 250$ GeV.

Details of the selection requirements imposed by the HLT paths and in the offline analysis are summarized in Table 9.2. The data-MC distribution of the LOOSE 2018 selected events have been shown in Fig. 9.17 for the p_T, η and DEEPJET scores of the two selected b-jets. Moderate agreement between data and MC has been found.

Table 9.3 presents the event yields for the qqH $\rightarrow$ qqb$\bar{\text{b}}$ signal after the HLT and offline selections, the contributing background processes for each year and each analysis classes estimated from the different MC simulated samples and in the Data for the fitting mass region of

Table 9.2: Offline selection requirements in the four analyzed samples.

Requirements	2016 (36.3 fb^{-1})		2018 (54.5 fb^{-1})	
	TIGHT	LOOSE	TIGHT	LOOSE
p_{T} thresholds	95, 80, 65, 30 GeV		110, 90, 80, 30 GeV	
Jet $|\eta| < 4.7$	✓	✓	✓	✓
Lepton veto	✓	✓	✓	✓
Number of b tags ≥ 2	✓	✓	✓	✓
b jet $|\eta| < 2.4$	✓	✓	✓	✓
$\Delta\phi_{b\bar{b}}$	≤ 1.6	≥ 2.1	≥ 1.6	≥ 2.1
$\Delta\eta_{jj}$	≥ 4.2	≥ 2.5	≥ 3.8	≥ 2.5
m_{jj}	≥ 500 GeV	≥ 250 GeV	≥ 500 GeV	≥ 250 GeV

$80 < m_{b\bar{b}} < 200$ GeV. The leading order (LO) estimation of the QCD multijet process in this particular topology is found to be poor to describe the data. Further the MC sample is statistically limited. Hence the actual contribution of this process, along with the other continuum backgrounds like $t\bar{t}$, single-top and W+jets processes are estimated from data in an unbiased way which is described later. Essentially the side band of the Higgs mass region of the invariant mass of the b-jet pair ($m_{b\bar{b}}$) has been used. The signal and the peaking backgrounds of Z+jets are estimated from the MC simulation.

Table 9.3: Event yields of different contributing processes estimated from MC simulation and DATA events in the $m_{b\bar{b}}$ fitted mass range

Process	2016 (36.3 fb^{-1})		2018 (54.5 fb^{-1})	
	TIGHT	LOOSE	TIGHT	LOOSE
VBF H	525 ± 2	352 ± 2	531 ± 2	184 ± 2
ggF H	139 ± 10	444 ± 5	205 ± 10	335 ± 6
QCD	353599 ± 6222	1344290 ± 16872	367552 ± 7238	743235 ± 28808
t$\bar{t}$	1781 ± 10	12905 ± 27	2400 ± 10	8467 ± 20
singe-t	417 ± 5	2396 ± 13	622 ± 6	1267 ± 10
W+jets	116 ± 10	767 ± 33	167 ± 16	528 ± 21
DY Z+jets	1167 ± 29	8604 ± 73	1546 ± 26	6338 ± 51
EWK Z+jets	284 ± 3	499 ± 4	373 ± 4	382 ± 4
Total MC	358028±13201	1370257±17297	373396±7982	760736±29025
DATA	376120 ± 613	1377160 ± 1170	389658 ± 624	750594 ± 866
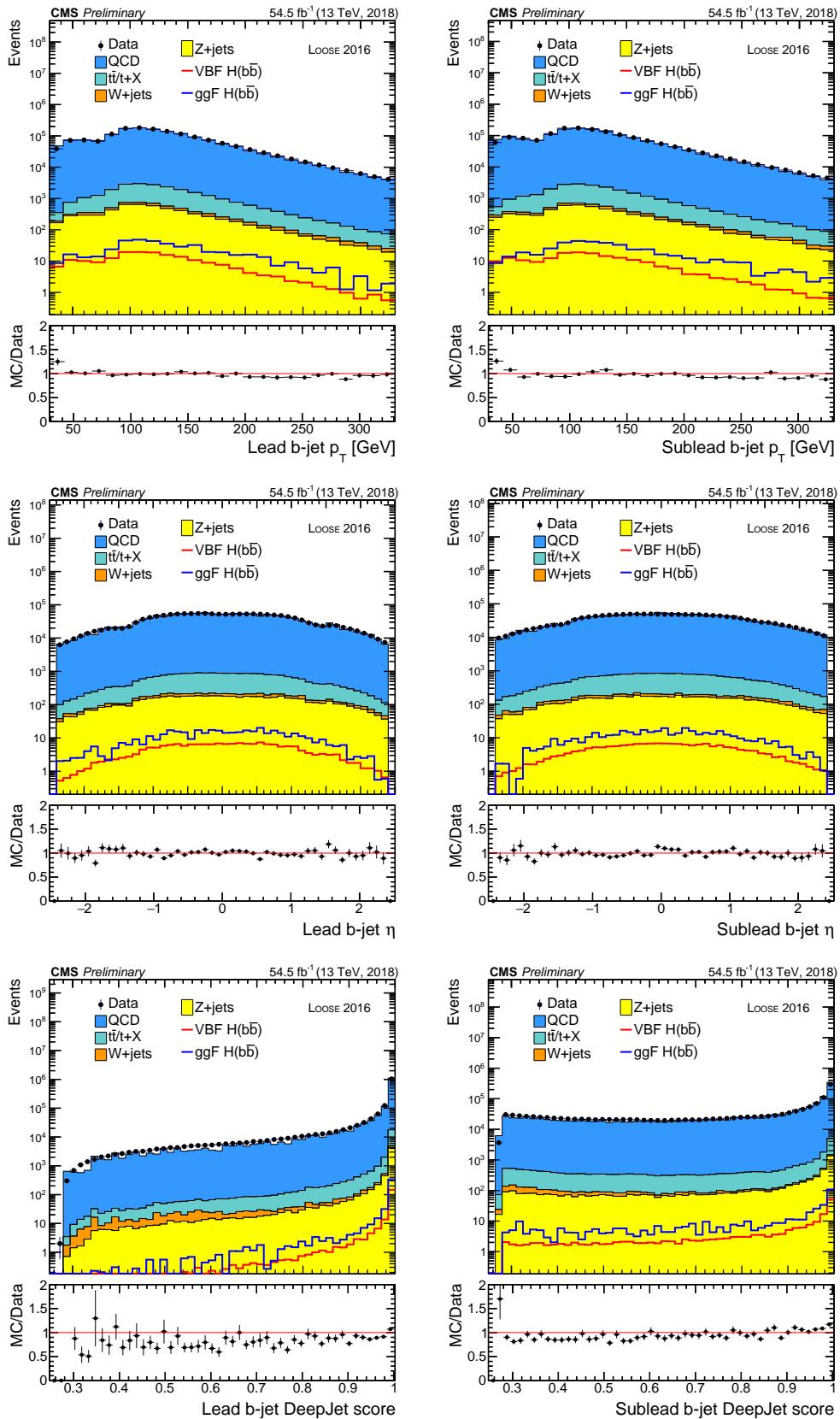
Figure 9.17: Data-MC distribution of the p_T, η, and DEEPJET distribution of the two selected b-tagged jets selected with HLT_LOOSE in 2018.

9.6 b-jet energy regression

The same b-jet energy regression applied on the selected b-jets is the same as described in Sec. 6.3.4. This regression not only improves the H $\rightarrow$ b$\bar{\text{b}}$, but also Z $\rightarrow$ b$\bar{\text{b}}$ invariant mass spectra as well. The improvement is measured in terms of the mass resolution as the quantity σ/μ of the $m_{\text{b}\bar{\text{b}}}$ distribution, where σ is the dispersion in the measurement and μ is the nominal value of $m_{\text{b}\bar{\text{b}}}$, by fitting the $m_{\text{b}\bar{\text{b}}}$ spectrum with a Crystal Ball function. The amount of the improvement is specific to the analysis phase space and for this particular analysis it is given in Table. 9.4. Figure 9.18 represents the effect of the b-jet energy regression on the $m_{\text{b}\bar{\text{b}}}$ distribution for qqH $\rightarrow$ qqb$\bar{\text{b}}$ signal for four analyzed samples.
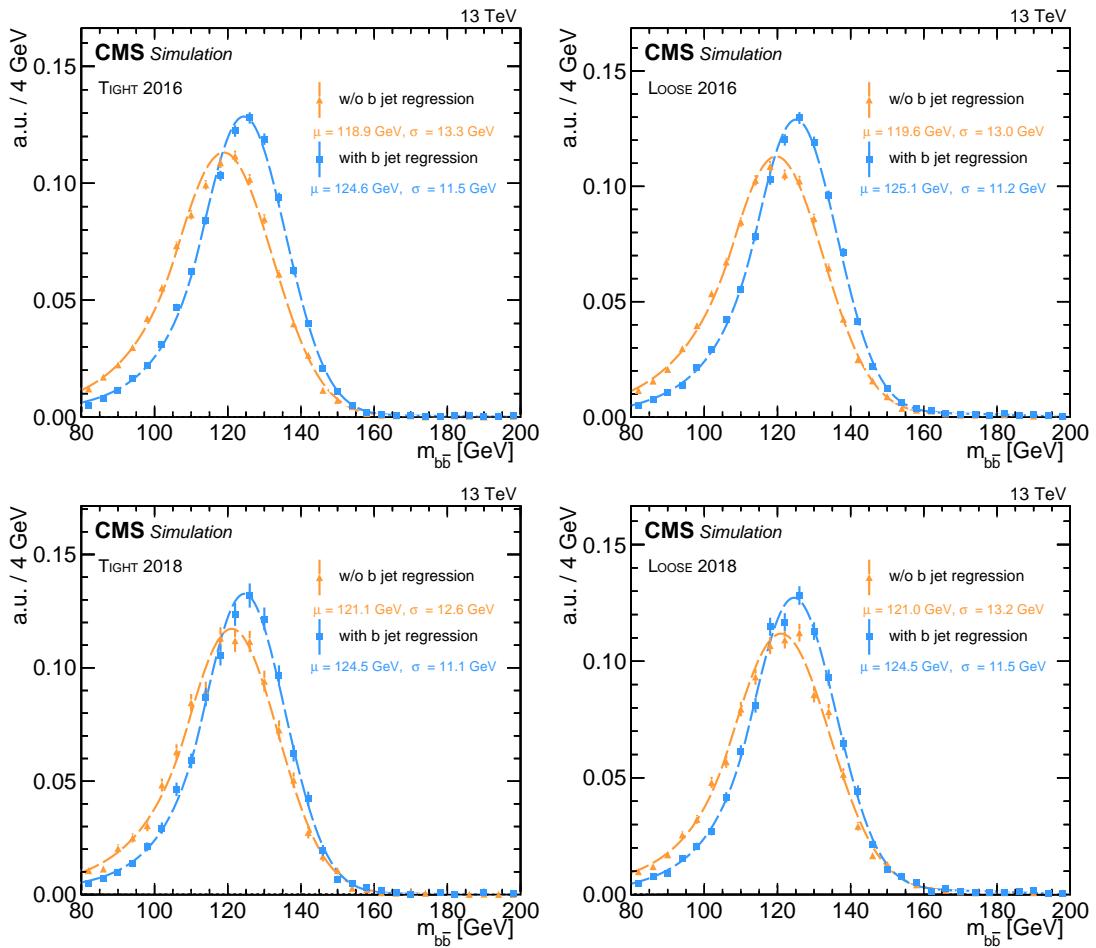
Figure 9.18: The invariant mass $m_{\text{b}\bar{\text{b}}}$ of the b-jet pair in simulated qqH $\rightarrow$ qqb$\bar{\text{b}}$ signal events before (orange) and after (blue) the application of the b jet energy regression in the TIGHT (left) and LOOSE (right) samples for 2016 (top row) and 2018 (bottom row). A one-sided Crystal Ball function has been used to fit the distributions.

Table 9.4: Impact of b-jet energy regression for $qqH \rightarrow qq b\bar{b}$ signal on $m_{b\bar{b}}$ peak position and resolution. Results presented for TIGHT and LOOSE classes for each year.

Year	mean (μ) [GeV]		dispersion (σ [GeV])		Improvement on
	w/o reg.	with reg.	w/o reg.	with reg.	σ/μ %
TIGHT 2016	117.9	123.8	14.2	12.3	17.5
TIGHT 2018	120.1	124.4	14.8	12.6	17.8
LOOSE 2016	118.8	124.4	13.2	11.9	14.0
LOOSE 2018	120.5	124.5	13.6	11.8	16.0

9.6.1   Effect of scale and smearing applied after after b-jet energy regression

The correlations among the input variables for the b-jet energy regression training are different for MC and data, they are propagated differently to the evaluation of energy of the b-jets using the regression technique. In order to correct the reconstructed b-jet energy in MC with respect to the data a scale as well as a smearing factor are applied on each selected b-jet. The detailed description of the determination of the scale and smearing terms is presented in CMS Run 2 VH (b$\bar{b}$) analysis described in Ref. [27]. After the application of the b-jet energy regression, a dedicated smearing is applied to the invariant mass distribution of Higgs candidate (H $\rightarrow$ b$\bar{b}$) to equalize the performance in data and in simulation. The smearing parameters are extracted in events where a jet recoils against a Z boson that decays into leptons as shown via a representative Feynman diagram in Fig. 9.19.

Because the p_T of the Z boson (p_T^Z) is balanced with the jet p_T (p_T^j), and given that the lepton momentum measurement is extremely precise, the ratio of the reconstructed p_T^j to p_T^Z makes the measurement of the jet energy resolution possible. The selected events are divided into four regions of $\alpha = p_T^j/p_T^Z$. For each range of α, ratio of the jet to the Z boson p_T, the mean and the standard deviation in data and simulation are extracted and used as scale and smearing corrections to the simulated regressed jets.

Figure. 9.20 represents the transverse momenta distributions for the leading and the subleading b-quark from the decay of the Higgs boson through VBF and VH production mode at the partonic level. It is evident that there is no significant difference between the two processes. This inspires that, the scale and smearing terms derived for the VH analysis can be used safely
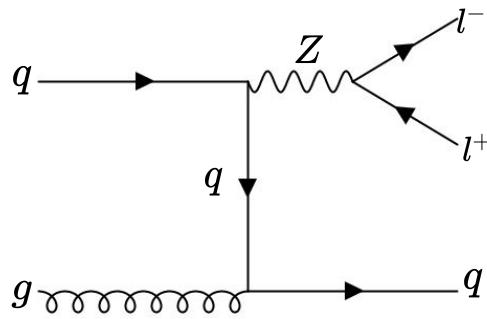
Figure 9.19: A representative feynman diagram of Z + 1 jet process, subsequently Z boson is decaying to a pair of leptons.

in the VBF analysis for the selected b-jets.

The scaling and smearing terms (in % of b-jet p_T) and the corresponding uncertainties are presented in Tab. 9.5, where the scale and smearing terms behave oppositely for 2016 and 2018. Scale is applied directly on the 4-momenta of the selected b-jets which, in turn, affects the position of the peak of the $m_{b\bar{b}}$ distribution, while smearing mainly modifies the resolution of the peak structure in the $m_{b\bar{b}}$ spectrum. A Crystal Ball function is used to fit the distribution to obtain the mean and the dispersion. After the application of the scaling and the smearing correction the $m_{b\bar{b}}$ peak has shifted towards the higher value with better resolution for 2016, while for 2018 the peak has shifted to the lower value with a worse resolution. This tendency can be expected from the derived values of the scale and the smearing factors as presented in Tab. 9.5.

Table 9.5: The extracted scale and smearing terms for modification of p_T of jet in H $\rightarrow$ b$\bar{b}$ decays, for the year 2016 and 2018, taken from CMS VH(bb) analysis. Ref. [27].

year	Scale	Smearing
2016	+0.4±1.8%	-4.4±6.1%
2018	-1.8±1.9%	+5.0±7.9%
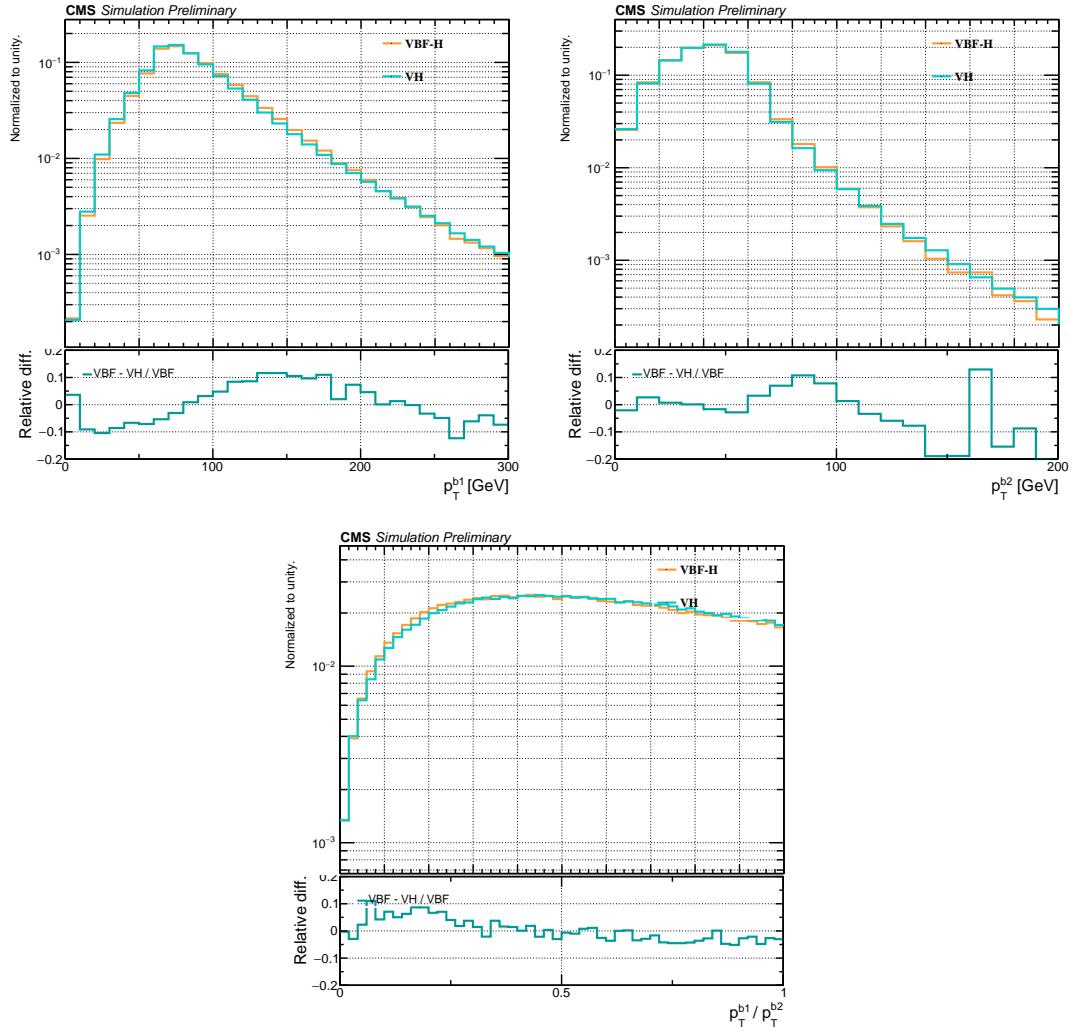
Figure 9.20: The p_T distributions for the leading and the subleading b-quarks from the decay of the Higgs boson, and the ratio of their p_T in VBF (in orange line) and VH (in cyan line) production modes. The lower panel shows the relative difference between the two processes for the corresponding variables.

9.7 Background Rejection

As the HLT and the offline selections are imposed with different conditions between the TIGHT and LOOSE events for the two years separately, the background reduction through MVA method is deployed in the analysis separately.

9.7.1 MVA discriminator in TightVBF event selection

After the selection of events it is found that the contribution of signal is submerged in the overwhelming backgrounds which is dominated by the QCD induced multijet production. To increase the sensitivity of our analysis it is important to separate the signal from the backgrounds based on the discriminating variables. Hence a multivariate analysis (MVA) technique is used where the discriminator is trained with Gradient Boosting algorithm. The MVA training has been performed separately for the years 2016 and 2018. The choice of the feature variables are based not only on their high discriminating power, but also on their weak dependence on the invariant mass of the two b-jets ($m_{b\bar{b}}$), as this is the final search variable. This requirement is utilized later in the analysis where the events are categorized based on the score of the discriminator (purity of the sample of events with discriminator value above a threshold) and the $m_{b\bar{b}}$ spectrum in each analysis category is fitted separately. For this reason $m_{b\bar{b}}$ should not be correlated with the discriminator. Further, possible mass sculpting due to the MVA score is also avoided.

The optimum set of input variables are determined by pruning interactively their effectiveness with N-1 tests and estimating the correlation with the $m_{b\bar{b}}$ variable. The 4-momenta of the two b-jets are not used in the discriminator due to the high correlation with the $m_{b\bar{b}}$. The final variables are listed below and the normalized distributions of qqH $\rightarrow$ qqb$\bar{b}$ signal and the background in the TIGHT 2018 analysis event samples are presented in Figs. 9.21 and 9.22.

- Properties of the VBF-tagged jets: invariant mass (m_{jj}), absolute difference of the pseudorapidity ($|\Delta\eta_{jj}|$) and the azimuthal angle ($|\Delta\phi_{jj}|$) of the two VBF jets;

- The minimal opening angle between the momentum vector of any of the two VBF-tagged jets and the momentum vector of the dijet system composed by the VBF-tagged jet pair (α_{jj});

- The quark-gluon discriminator [151, 152] score of the two VBF-tagged jets. This is a

likelihood discriminator constructed using the information on the charged and the neutral constituents of a given jet and the jet shape variables to distinguish between the jets originating from quarks and from gluons;

- The DEEPJET b-tagging scores of the two selected b-tagged jets assigned to the H → b$\overline{\text{b}}$ decay;

- The ratio of the magnitude of the transverse momentum vector of the selected four-jet system to the scalar p_T-sum of the four selected jets $(|\sum_{i=1}^{4} \vec{p}_\text{T}|/ \sum_{i=1}^{4} p_\text{T})$;

- The vector sum of the longitudinal component of the momentum vector of the selected four-jet system $(\sum_{i=1}^{4} p_z)$;

- The absolute value of the difference in the azimuthal angle between the Higgs boson candidate and the dijet system composed of the VBF-tagged jets $(|\phi_{\text{b}\overline{\text{b}}} - \phi_{\text{jj}}|)$;

- The multiplicity and p_T-sum of extra jets in the event with $p_\text{T} > 30\,\text{GeV}$ and $|\eta| < 2.4$, excluding the selected b-tagged and VBF-tagged jets.

9.7.2   Training architecture

The training of the multivariate discriminant is performed separately for 2016 and 2018 using TMVA [134] based Gradient Boosted Decision Tree (BDTG) corresponding to the TIGHT class. The qqH → qqb$\overline{\text{b}}$ signal sample is used for the training. Due to the insufficient statistics and poor modelling of the QCD MC sample, a small subset of the selected collision data (about 5%) has been used as a proxy of the QCD background. It is found that in the selected data, almost 98-99% contamination coming from the QCD multijet process; so the signal contamination in the selected 5% of data used for the training is negligible.

The parameters of the training are optimized based on the following consideration (i) max-imisng the ROC-integral value, (ii) reasonable Kolmogrov-Smironov test score for both signal and background (KS test score > 0.1 at least) and, finally, (iii) the training-testing behavior. The ROC curve, shown in Fig. 9.23 indicates that the signal and background separation achieved.

Figure 9.24 shows the discriminator score (D) distributions for the VBFH, ggH, and Z → b$\overline{\text{b}}$ processes and the data in the TIGHT 2016 and TIGHT 2018 samples. Even in the high purity region with high MVA values, the QCD background is still much larger than the signal by at least two orders of magnitude.
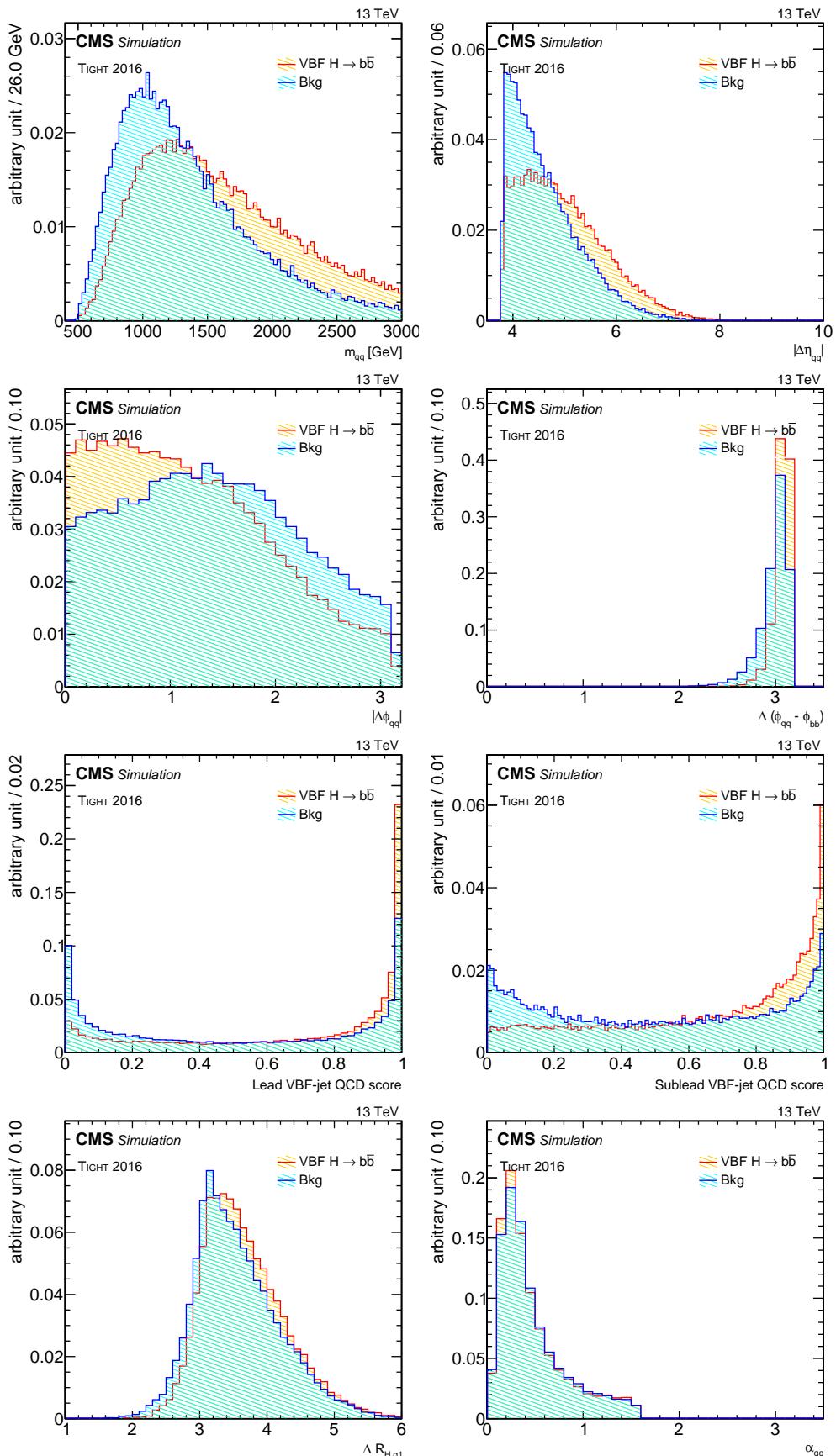
Figure 9.21: Distributions of the input variables for qqH → qqb$\bar{\text{b}}$ signal and background for TIGHT 2018 sample. All distributions are normalized to unity.
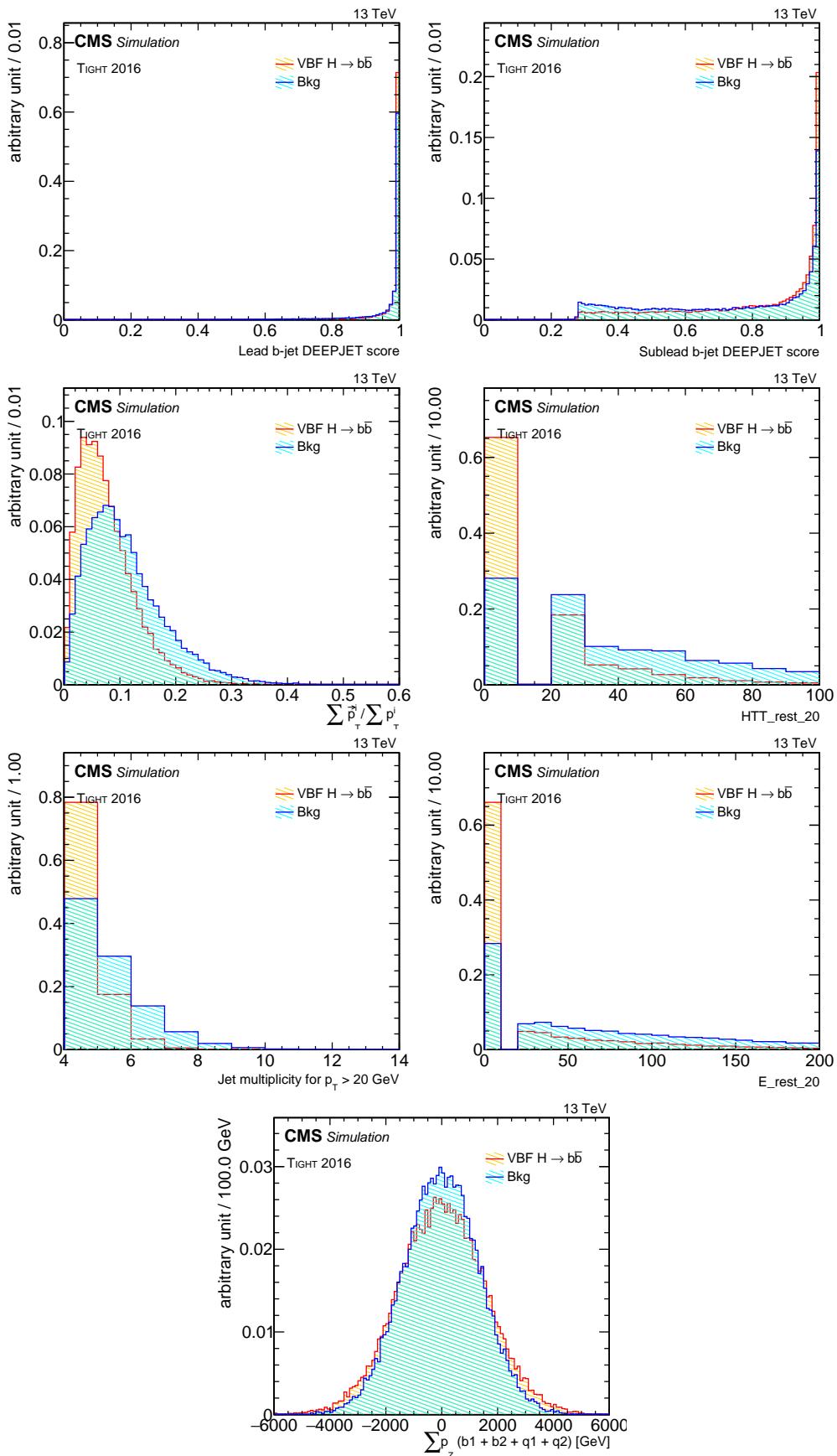
Figure 9.22: Distributions of the input variables for qqH → qqb$\bar{\text{b}}$ signal and background for TIGHT 2018 sample. All distributions are normalized to unity.
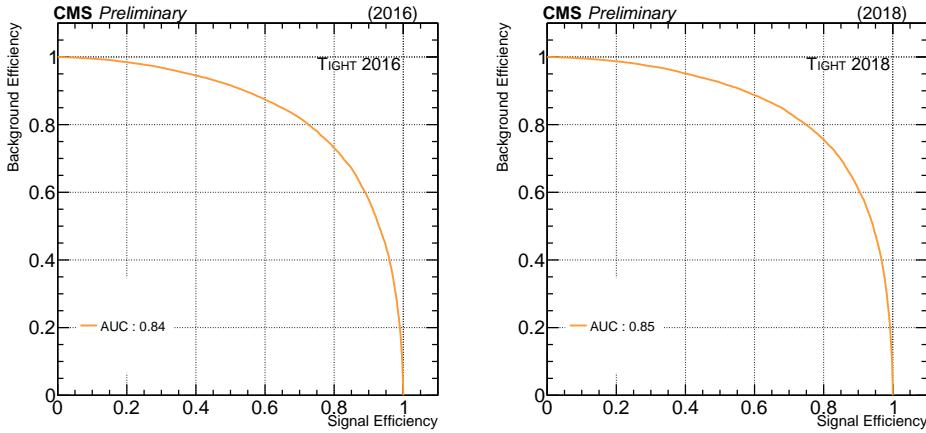
Figure 9.23: ROC curves obtained after training for TIGHT 2016 (left) & 2018 (right).
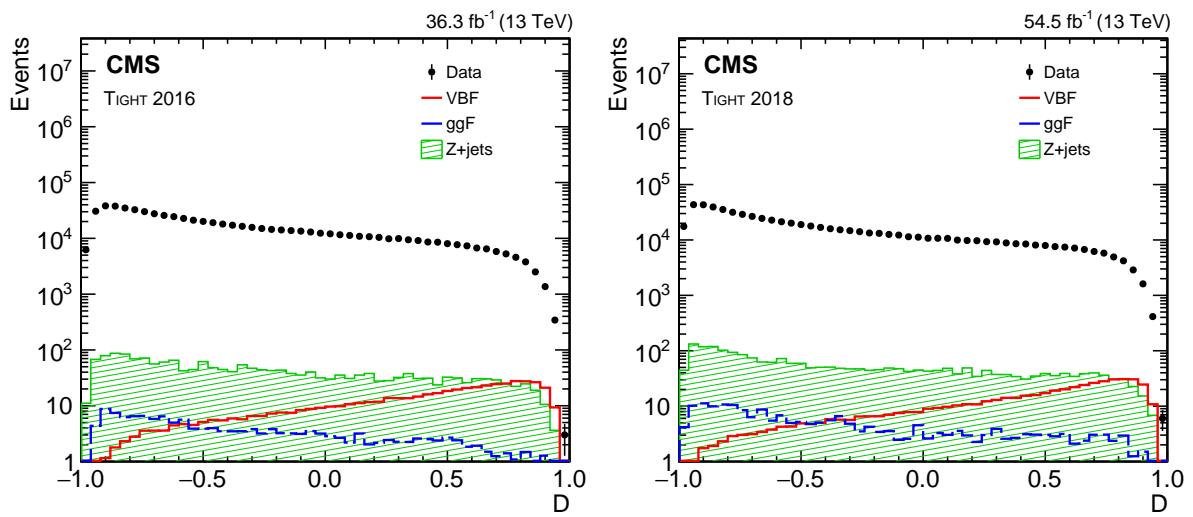
Figure 9.24: The distributions of the VBF BDT outputs in data and simulated samples in the TIGHT 2016 (left) and TIGHT 2018 (right) analysis samples. Data events (points), dominated by the QCD multijet background, are compared to the VBF (red solid line), ggH (blue dashed line) and Z+jets (green hatched area) processes.

9.7.3   Checks with MVA training

Some additional checks have been performed to test the stability of the MVA training.

1.   Training with a different set of 5% of Data

To check the stability and consistency of the BDT training, an alternative training has been performed using a second set of 5% of the selected data, different from the 5% of data already used in the nominal training which is used in the analysis. Figure. 9.25 represents the MVA score for the two independent sets (each 5% of data) and establish the consistency between them.
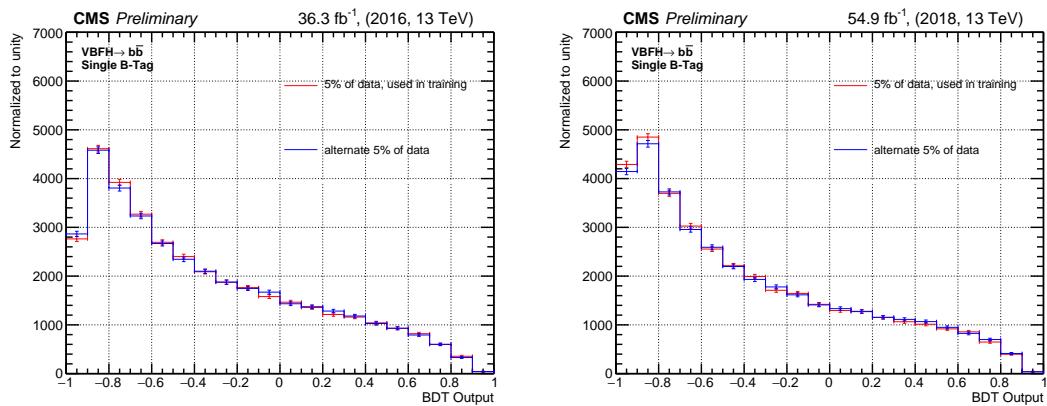
Figure 9.25: MVA score distribution trained with the two independent sets (of 5%) data used in the training for TIGHT 2016 & TIGHT 2018 event selection.

2.   Check performed for possible mass sculpting

The main contributing background in this analysis is the QCD-induced multijet production leading to a continuum of falling distribution for the $m_{b\bar{b}}$ variable. The mean of the shape of the $m_{b\bar{b}}$ spectra should be independent from the BDT output score and this is shown in Fig. 9.26, where the mass sculpting has been checked both in the data and in the QCD MC samples.

9.7.4   MVA discrimination in LooseVBF event selection

In the LOOSE event selection, the contributions of the Z+jets and ggH processes are comparable with the qqH $\rightarrow$ qqb$\bar{b}$ signal due to the looser VBF topological cuts at the HLT, and subsequently, also at the offline level. So a multi-class MVA has been trained to distinguish the qqH $\rightarrow$ qqb$\bar{b}$ signal from the QCD multijet background as well as the Z+jets and ggH
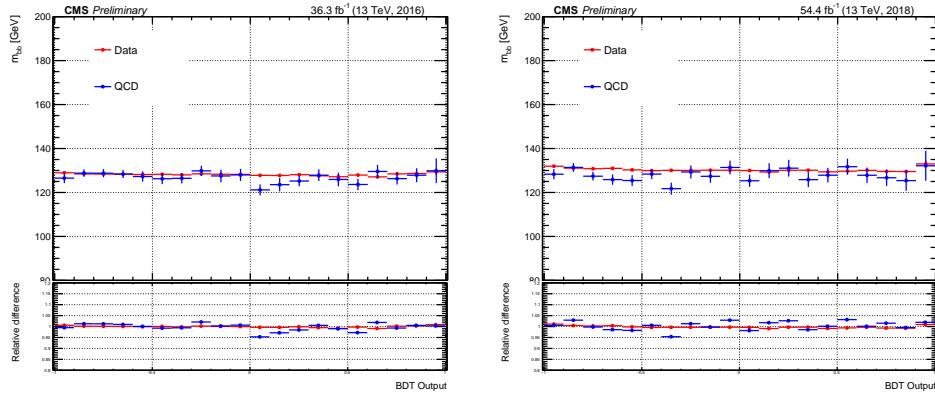
Figure 9.26: The mean of the $m_{b\bar{b}}$ mass spectrum for the aata (in red) and QCD MC (in blue) events corresponding to the different BDT output score ranges for 2016 (left) and 2018 (right). The Data and QCD MC samples show similar tendency with flat distribution over the full range of BDT score. The lower panel shows the relative difference with of the individual BDT range with respect to the full BDT range.

processes simultaneously. Due to the poor MC modelling of the QCD process, 5% of the selected collision data is used as QCD background profile in the training, while for Z+jets and ggH processes respective MC simulated samples are used. The training strategy is the same as opted for the TIGHT analysis, with the same set of feature variables that have been used as input for the training and listed in Sec. 9.7.1.

The distributions of the feature variables used in the multi-class training for LOOSE event selection are presented in Figs. 9.27 and 9.28 corresponding to the respectively.

The multiclass BDT has four outputs, D_{VBF}, D_{ggH}, D_{Z}, and D_{QCD}, quantifying the compatibility of each event with the VBFH, ggH, Z+jets, and QCD multijet production hypotheses, respectively. The corresponding distributions are shown in Fig. 9.29 for two different years.
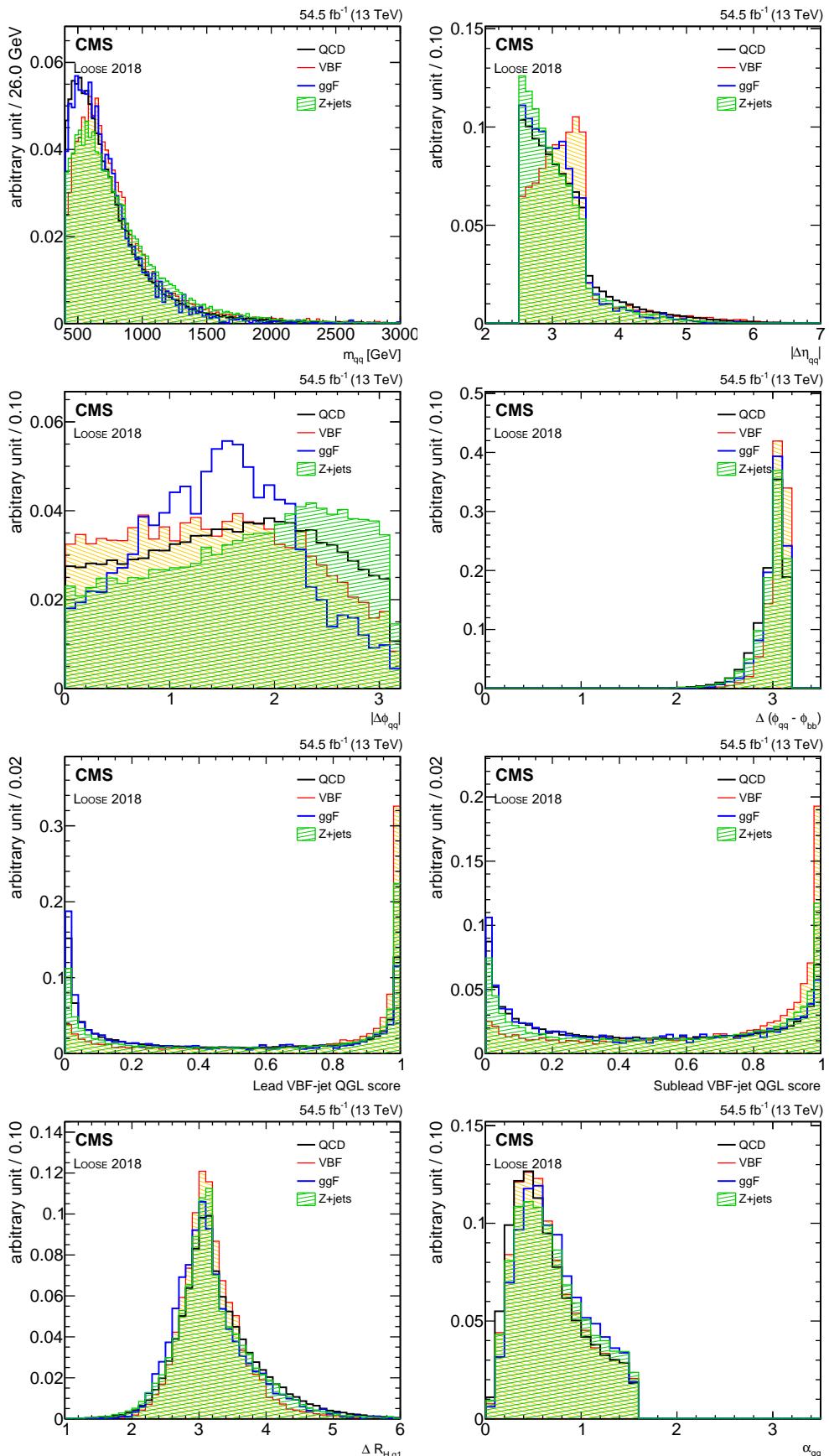
Figure 9.27: Distributions of the input variables for qqH → qqb$\bar{\text{b}}$ signal (in green), ggH (in red), Z+jets (in blue) and QCD (in black) for LOOSE 2018 sample. All distributions are normalized to unity.
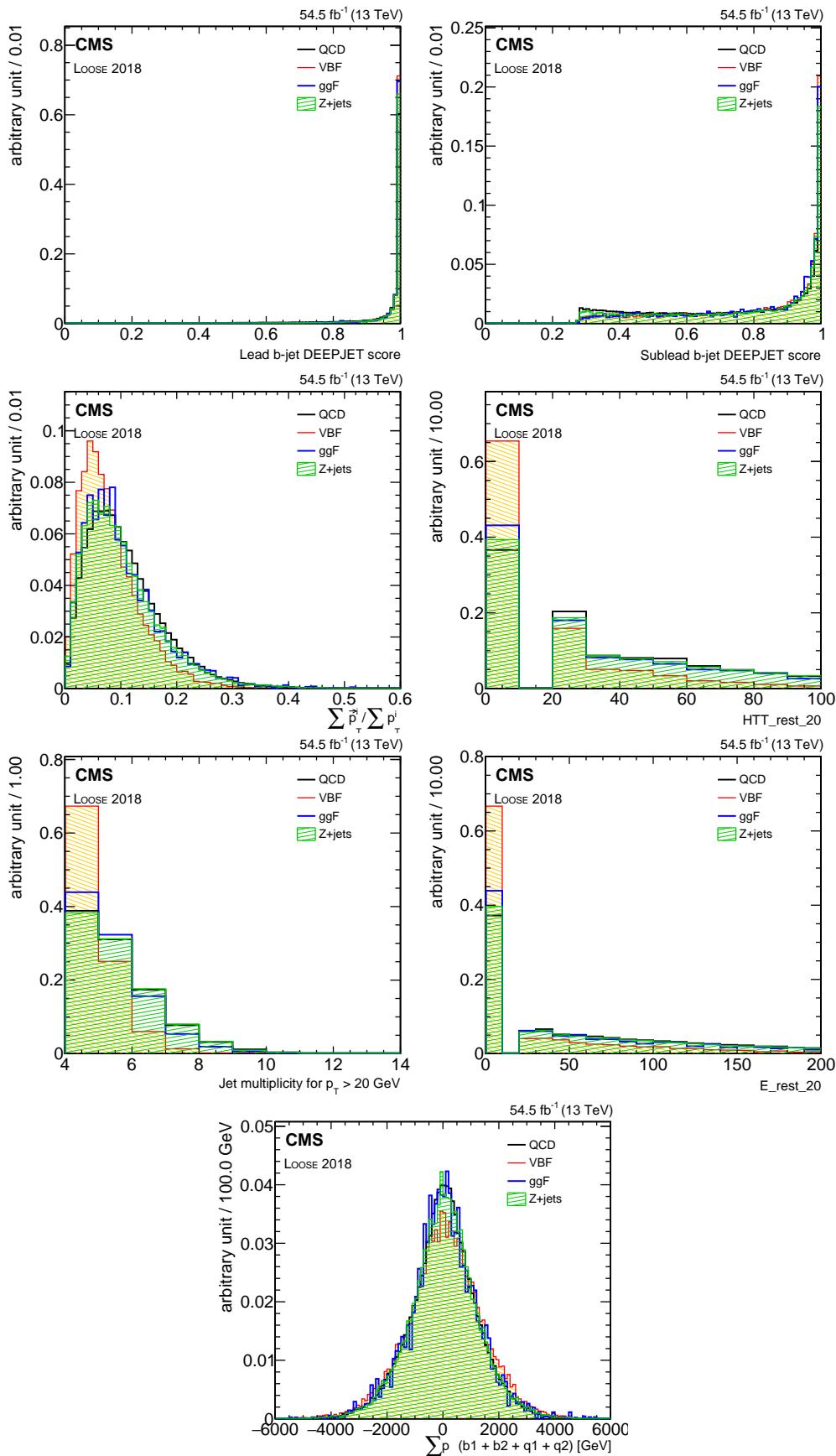
Figure 9.28: Distributions of the input variables for qqH → qqb$\bar{\text{b}}$ signal (in green), ggH (in red), Z+jets (in blue) and QCD (in black) for LOOSE 2018 sample. All distributions are normalized to unity.
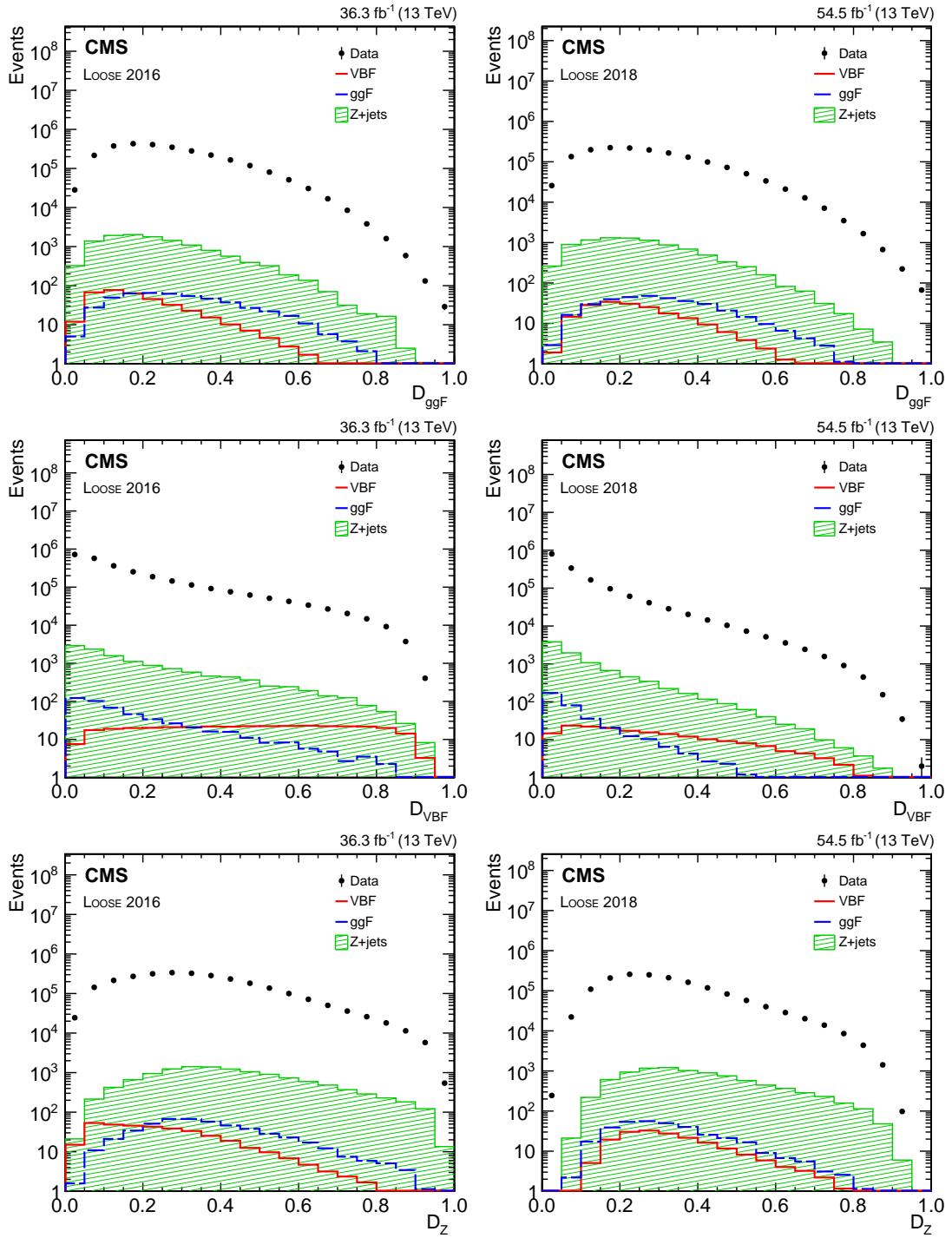
Figure 9.29: The distributions of the BDT outputs: D_{ggH} (upper), D_{VBF} (middle) and D_Z (lower) in data and simulated samples in the LOOSE 2016 (left) and LOOSE 2018 (right) analysis samples. The description of the figures are same as in Fig. 9.24.

9.8   Categorization of events

Based on their BDT scores, events are classified into multiple exclusive categories, targeting the VBFH, ggH, and Z+jets processes. Details of the event categorization and the naming convention of the event categories are given in Table 9.6. The BDT thresholds defining each category, are optimized by maximizing the quantity $S/\sqrt{B}$, where S is the number of expected events of the targeted process and B is the number of QCD multijet background events approximated as the observed data events in the $m_{b\bar{b}}$ interval populated by the targeted process. As evident S is significantly smaller than the value of B. The $m_{b\bar{b}}$ interval of 80–100 GeV is used for the Z+jets process and 104–146 GeV for the VBF and ggH processes. In total 18 categories are introduced, three in each of the two TIGHT analysis samples, and six in each of the two LOOSE analysis samples.

The aim of introducing distinct categories sensitive to the production of the Z boson is two-fold. First, these categories are intended to establish the signal from the $Z \rightarrow b\bar{b}$ standard candle, thereby validating the analysis techniques employed in this study. Second, the tail of each $m_{b\bar{b}}$ distribution in the Z+jets sample extends to the region partially populated by the signal events, thus affecting the precision of the measurement. It is therefore important to constrain the background from the Z+jets process with dedicated categories, thus improving the sensitivity to the signal. Further, introduction of separate categories targeting the ggH process improves the sensitivity of the analysis to inclusive Higgs boson production.

The number of selected events in data along with the expected background and signal yields in each analysis category are detailed in Tables 9.7 and 9.8.

Table 9.6: Event categorization used in the analysis for a total of 18 categories accounting for 2 years 2016 and 2018. The name of the categories are given in the first column. The BDT score boundaries defining each category are given in the second column and the targeted process is indicated in the third column.

Category	BDT score boundaries	Targeted process
year : 2016		
TIGHT 1	$0.25 \leq D < 0.50$	VBFH
TIGHT 2	$0.50 \leq D < 0.75$	VBFH
TIGHT 3	$0.75 \leq D$	VBFH
LOOSE G1	$0.50 \leq D_{ggH} < 0.55$	ggH
LOOSE G2	$0.55 \leq D_{ggH}$	ggH
LOOSE V1	$D_{ggH} < 0.50, 0.80 \leq D_{VBF} < 0.85$	VBFH
LOOSE V2	$D_{ggH} < 0.50, 0.85 \leq D_{VBF}$	VBFH
LOOSE Z1	$D_{ggH} < 0.50, D_{VBF} < 0.80, 0.60 \leq D_Z < 0.75$	Z+jets
LOOSE Z2	$D_{ggH} < 0.50, D_{VBF} < 0.80, 0.75 \leq D_Z$	Z+jets
year : 2018		
TIGHT 1	$0.25 \leq D < 0.50$	VBFH
TIGHT 2	$0.50 \leq D < 0.75$	VBFH
TIGHT 3	$0.75 \leq D$	VBFH
LOOSE G1	$0.55 \leq D_{ggH} < 0.60$	ggH
LOOSE G2	$0.60 \leq D_{ggH}$	ggH
LOOSE V1	$D_{ggH} < 0.55, 0.50 \leq D_{VBF} < 0.55$	VBFH
LOOSE V2	$D_{ggH} < 0.55, 0.55 \leq D_{VBF}$	VBFH
LOOSE Z1	$D_{ggH} < 0.55, D_{VBF} < 0.50, 0.60 \leq D_Z < 0.70$	Z+jets
LOOSE Z2	$D_{ggH} < 0.55, D_{VBF} < 0.50, 0.70 \leq D_Z$	Z+jets

Table 9.7:   Event yields for various categories of the analyzed 2016 data corresponding to 36.3 fb^{-1}, compared to the expected number of events from the simulated samples of signal and background other than the QCD multijet process. The quoted uncertainties are statistical only.

Category	VBFH	ggH	Z+jets	t$\bar{\text{t}}$ & t/$\bar{\text{t}}$	W+jets	Data
LOOSE G1	4.5 ± 0.2	27.5 ± 1.2	275 ± 16	116.7 ± 2.6	10.6 ± 3.0	41430 ± 200
LOOSE G2	6.1 ± 0.3	51.6 ± 1.8	407 ± 19	127.3 ± 2.7	8.2 ± 2.4	58890 ± 240
LOOSE V1	19.9 ± 0.4	2.7 ± 0.4	45 ± 2	10.6 ± 0.8	1.9 ± 1.3	4330 ± 70
LOOSE V2	17.4 ± 0.4	1.7 ± 0.3	31 ± 5	4.7 ± 0.4	0.5 ± 0.3	1900 ± 40
LOOSE Z1	9.6 ± 0.3	34.8 ± 1.4	1150 ± 20	226.3 ± 3.3	40.8 ± 6.1	78850 ± 280
LOOSE Z2	3.1 ± 0.2	14.1 ± 0.7	650 ± 10	199.0 ± 3.0	35.0 ± 5.9	29990 ± 170
TIGHT 1	92.7 ± 1.0	15.6 ± 1.0	161 ± 8	37.6 ± 1.4	7.7 ± 2.5	29860 ± 170
TIGHT 2	136.2 ± 1.5	13.9 ± 0.9	151 ± 6	22.5 ± 1.2	4.2 ± 1.4	21830 ± 150
TIGHT 3	117.3 ± 1.1	6.2 ± 0.7	75 ± 3	5.5 ± 0.5	3.1 ± 1.1	7230 ± 90

Table 9.8:   Event yields for various categories of the analyzed 2018 data corresponding to 54.5 fb^{-1}, compared to the expected number of events from the simulated samples of signal and background other than the QCD multijet process. The quoted uncertainties are statistical only.

Category	VBFH	ggH	Z+jets	t$\bar{\text{t}}$ & t/$\bar{\text{t}}$	W+jets	Data
LOOSE G1	2.4 ± 0.2	13.5 ± 2.1	137 ± 9	32.4 ± 1.4	2.0 ± 0.5	17290 ± 130
LOOSE G2	2.9 ± 0.2	24.3 ± 1.5	180 ± 10	33.1 ± 1.4	2.1 ± 0.5	24880 ± 160
LOOSE V1	6.4 ± 0.3	1.0 ± 0.3	22 ± 3	4.4 ± 0.5	0.5 ± 0.3	1910 ± 40
LOOSE V2	11.0 ± 0.4	1.7 ± 0.4	25 ± 3	4.7 ± 0.5	0.3 ± 0.2	2450 ± 50
LOOSE Z1	7.0 ± 0.3	10.8 ± 0.9	506 ± 10	59.0 ± 1.7	19.5 ± 2.7	24550 ± 160
LOOSE Z2	4.0 ± 0.2	7.1 ± 0.7	445 ± 7	99.0 ± 2.2	24.0 ± 2.8	14530 ± 120
TIGHT 1	89.5 ± 1.1	18.0 ± 3.2	190 ± 8	49.3 ± 1.5	8.0 ± 2.5	29260 ± 170
TIGHT 2	139.8 ± 1.4	17.9 ± 3.3	202 ± 7	31.6 ± 1.3	7.2 ± 2.4	23390 ± 150
TIGHT 3	134.5 ± 1.4	8.7 ± 2.6	104 ± 5	7.6 ± 0.6	3.4 ± 1.6	8200 ± 90

9.9 Statistical analysis

The test statistic chosen to determine the signal yield is based on the profile likelihood ratio [153, 154]. A detailed description of the test statistic is given in Sec. B. The signal strength is extracted from the simultaneous binned maximum likelihood fit of the $m_{b\bar{b}}$ distribution in all categories obtained from data. In each category, the $m_{b\bar{b}}$ distribution is fitted with a superposition of three parametric analytical functions accounting for: (i) signal (ii) resonant Z+jets background and, (iii) the continuum background, dominated by QCD multijet events.

9.9.1 Signal modelling

The combined contribution qqH and ggH signals is estimated from the MC simulations, modeled by a Cryastal Ball (CB) [114] function at the peak and a Bernstein polynomial of 2nd order for the tail at the higer value of $m_{b\bar{b}}$ distributions which may come from the wrong choice of the jets at the Higgs candidate reconstruction. Figure. 9.30 shows the signal modeling in the TIGHT and LOOSE analysis class for both the two years by merging all categories.

9.9.2 Z+jets background modelling

The peak of the $m_{b\bar{b}}$ distribution of the Z $\rightarrow$ b$\bar{b}$ process is very close to the H $\rightarrow$ b$\bar{b}$ decay peak; so a proper modeling of this process is very important. Hence the contribution of the Z+jets process due to the QCD and pure electroweak production modes are considered from the dedicated MC simulations and the parametric modeling of the shape of the $m_{b\bar{b}}$ distribution has been derived using the same approach used for the signal modeling. The Crystal Ball function used around the Z boson mass at 91 GeV and 2nd order Bernstein polynomial for the tail of the high $m_{b\bar{b}}$ values. Figure. 9.31 shows the signal modeling in the TIGHT and LOOSE analysis class for both the two years merging all categories.
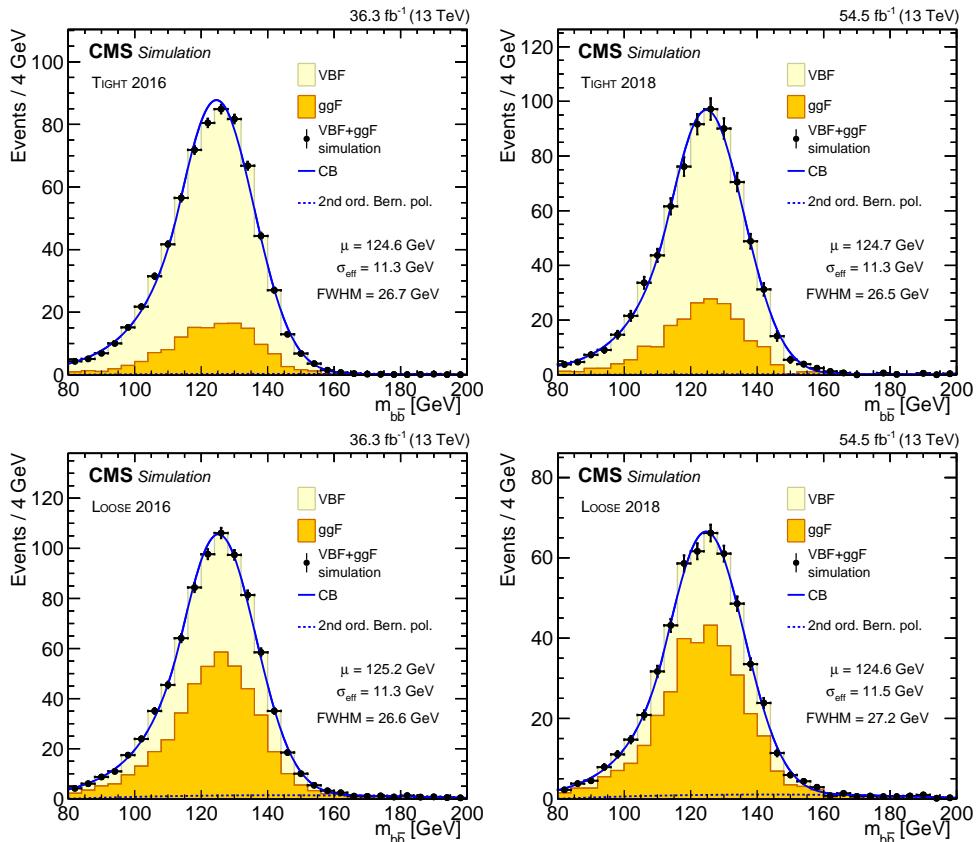
Figure 9.30: The $m_{b\bar{b}}$ modeling of the signal: VBF contribution (in yellow), ggH contribution (in orange) and total Higgs (VBF + ggH) contribution (in balk points) with statistical uncertainties and the fitted model (with blue solid lines) in TIGHT (upper row) and LOOSE (bottom row) event selection for 2016 (left) and 2018 (right). A parametrization by 2nd order Bernstein polynomial has been used to account for the continuum background as discussed in Sec. 9.9.3
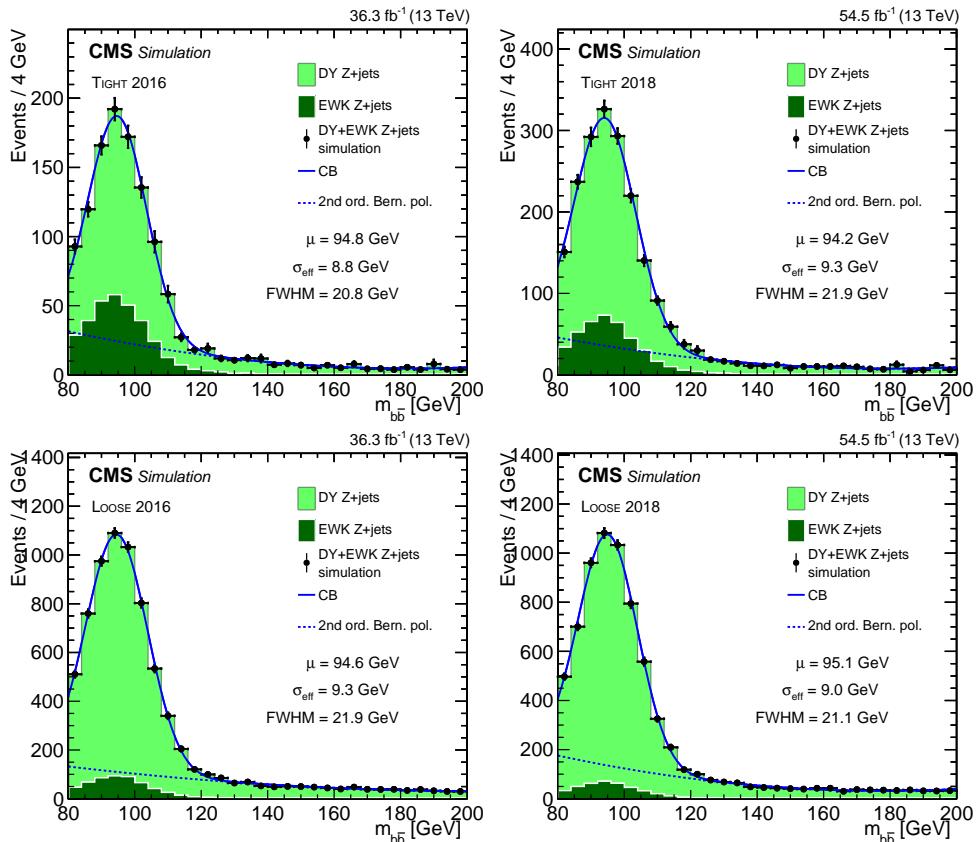
Figure 9.31: The $m_{b\bar{b}}$ modeling of the signal: DY Z+jets contribution (in light green), EWK Z+jets contribution (in dark green) and total Z+jets contribution (in balk points) with statistical uncertainties and the fitted model (with blue solid lines) in TIGHT (upper row) and LOOSE (bottom row) for 2016 (left) and 2018 (right). A parametrization by 2nd order Bernstein polynomial has been used to account for the continuum background as discussed in Sec. 9.9.3

9.9.3 Continuum background modeling

The shape and the normalization of the continuum background is estimated directly from the data by fitting side-bands of the resonance region in the $m_{b\bar{b}}$ distribution. The shape of continuum background is modeled individually for each event category "i" by a convolution of an exponential and a polynomial functions:

$$F_i^{QCD} = \exp\left(-b_i \cdot m_{bb}\right) \cdot \left(1 + \sum_{i=1}^{n} a_i \cdot m_{bb}^i\right) \tag{9.3}$$

In each category the choice of the polynomial function is guided by the combined fit of two sideband regions, $80 < m_{b\bar{b}} < 104$ GeV and $146 < m_{b\bar{b}} < 200$ GeV. Sequential fits with increasing order of polynomial are performed followed by the Fisher-Test to select the optimum choice of the polynomial order with the least number of parameters which ensures the fit to be of good quality. The fit function of the continuum $m_{b\bar{b}}$ spetctra in each analysis categories are given in Tab. 9.9.

Table 9.9:   The functional forms used to fit the continuum component of the background in various analysis categories. The notation "exp" stands for the exponential function, "exp⊗pol1 (pol2)" denotes the product of an exponential function and a first-order (second-order) polynomial as given in eq. 9.3.

Category	Analysis class	
	TIGHT 2016	TIGHT 2018
1	exp	exp
2	exp	exp
3	exp	exp
	LOOSE 2016	LOOSE 2018
G1	exp⊗pol2	exp⊗pol1
G2	exp⊗pol2	exp⊗pol1
V1	exp	exp
V2	exp	exp
Z1	exp⊗pol1	exp⊗pol1
Z2	exp⊗pol1	exp⊗pol1

The fits of the $m_{b\bar{b}}$ distribution in the data side-band region for each of the analysis categories are shown in Figs. 9.32 (TIGHT 2016), 9.33 (TIGHT 2018), 9.34 (LOOSE 2016) and 9.35 (LOOSE 2018). The values of χ^2 for per degrees of freedom (χ^2/n.d.f) and the probability value corresponding to the χ^2 ($P(\chi^2)$) are mentioned in each plot to quantify the goodness of the fit.
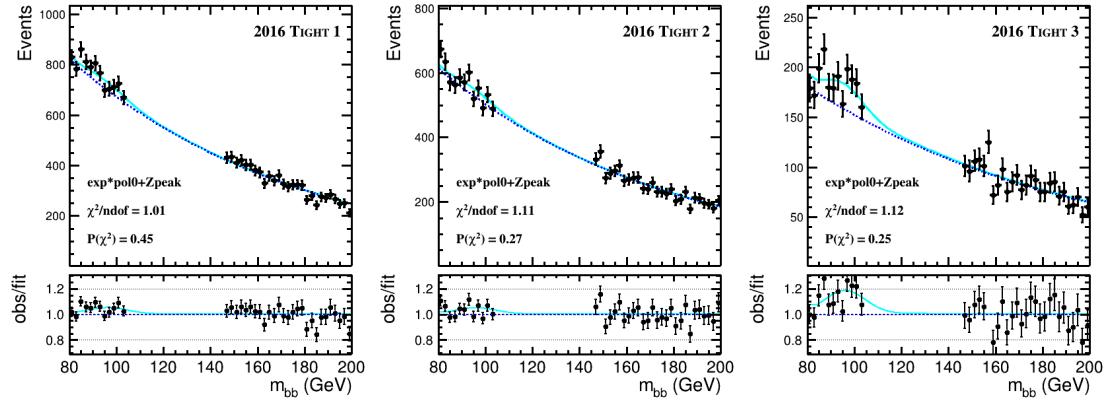
Figure 9.32: The blinded $m_{b\bar{b}}$ modelling from sideband data: black dots represents the data selected in TIGHT 2016 with statistical uncertainties, blue dashed curves represents the exponential fit for the continuum and cyan color follows the Z-peak + continuum. Lower panel presents the data over parametric fit.
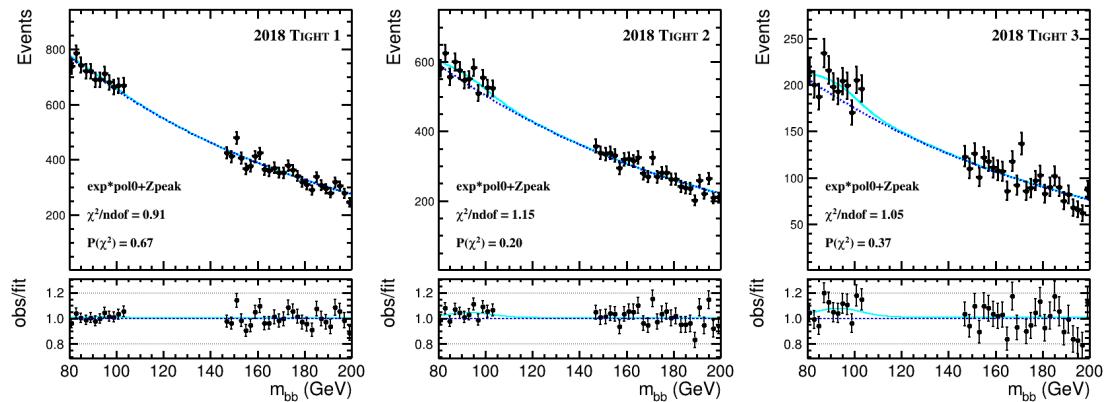
Figure 9.33: The blinded $m_{b\bar{b}}$ modelling from sideband data: black dots represents the data selected in TIGHT 2018 with statistical uncertainties, blue dashed curves represents the exponential fit for the continuum and cyan color follows the Z-peak + continuum. Lower panel presents the data over parametric fit.
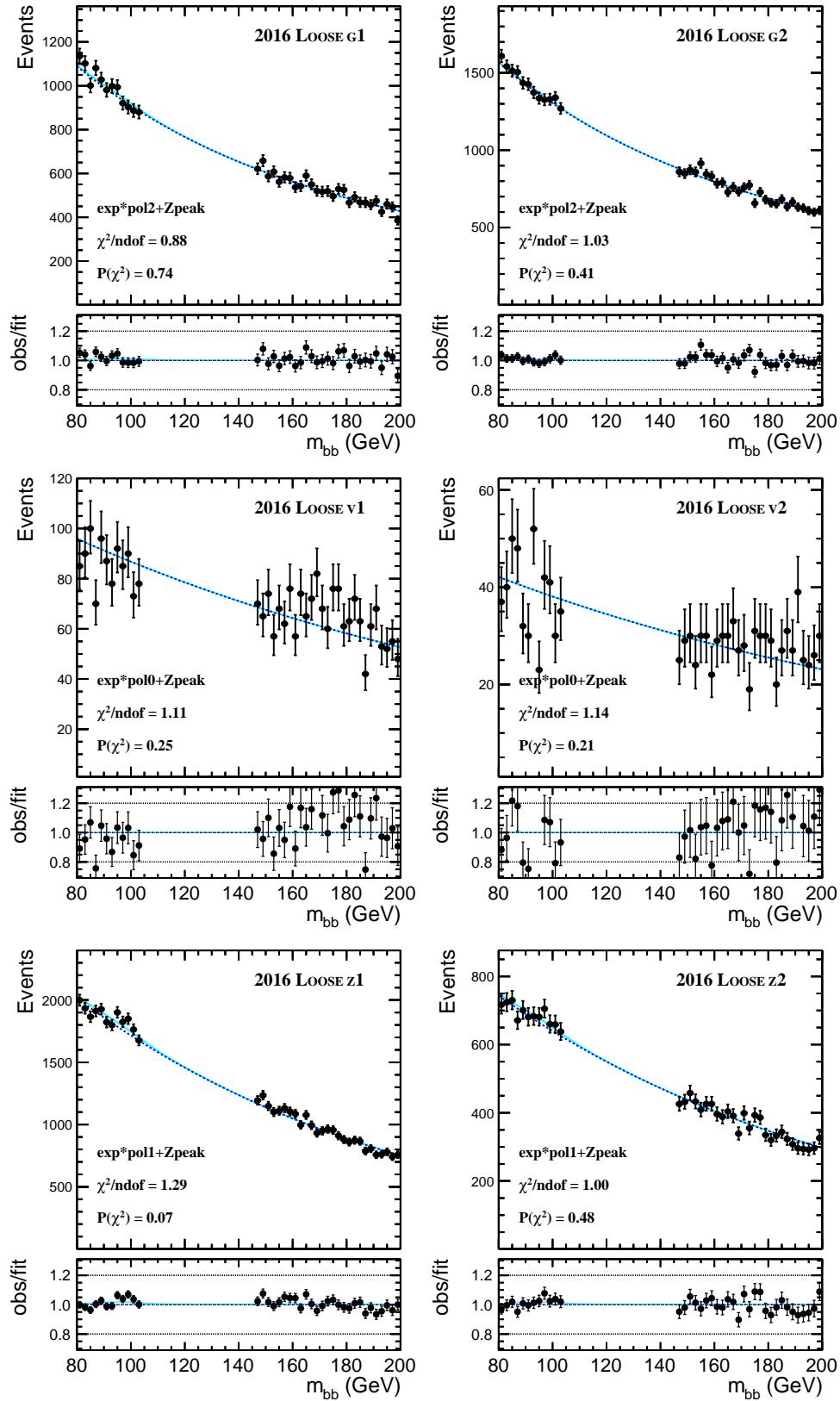
Figure 9.34: The blinded $m_{b\bar{b}}$ modelling from sideband data: black dots represents the data selected in LOOSE 2016 with statistical uncertainties, blue dashed curves represents the exponential fit for the continuum and cyan color follows the Z-peak + continuum. Lower panel presents the data over parametric fit.
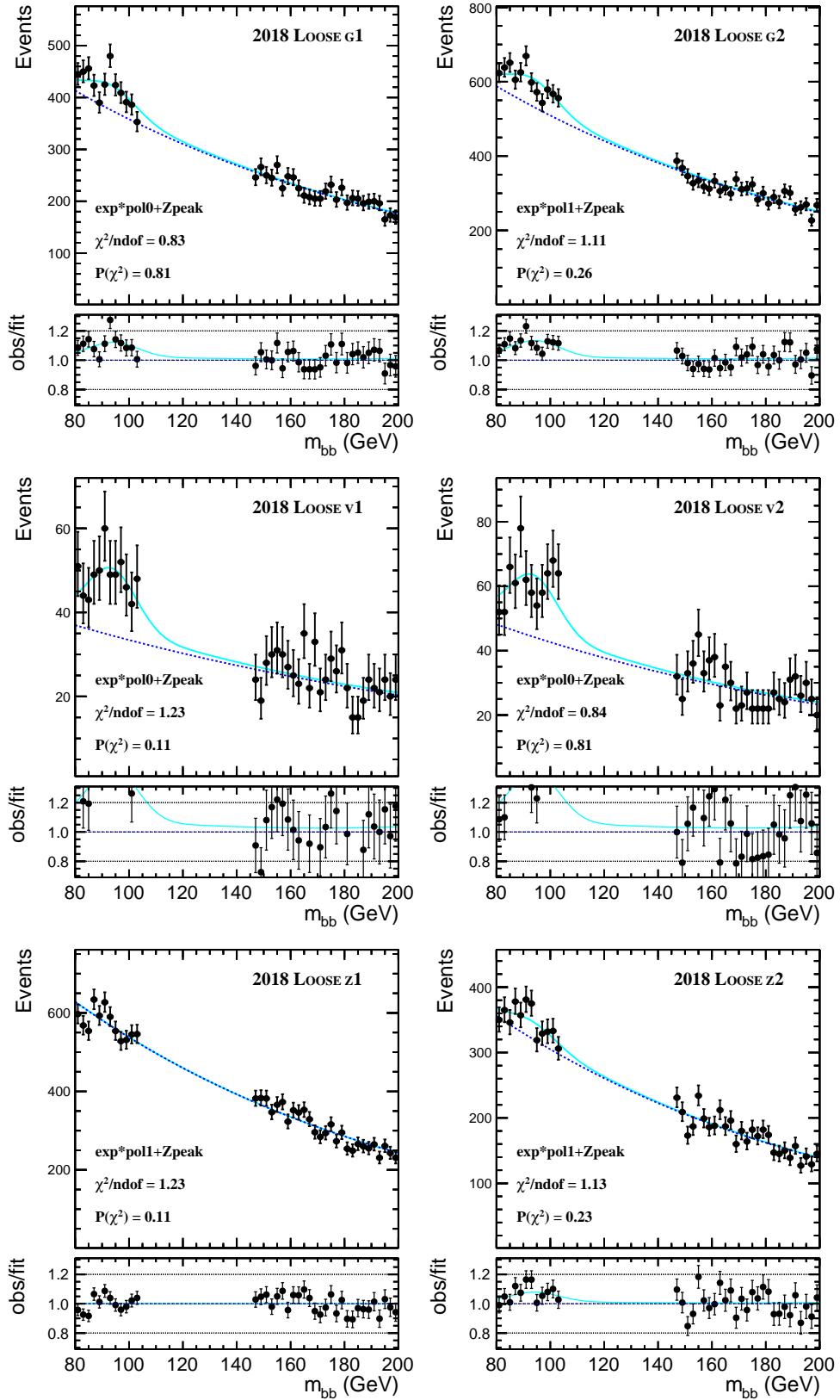
Figure 9.35: The blinded $m_{b\bar{b}}$ modelling from sideband data: black dots represents the data selected in LOOSE 2016 with statistical uncertainties, blue dashed curves represents the exponential fit for the continuum and cyan color follows the Z-peak + continuum. Lower panel presents the data over parametric fit.

9.9.4 Bias study

Our continuum background is estimated directly by fitting data using some parametric function of $m_{b\bar{b}}$. The final result is expected to be independent on the choice of the parametric functional form. So a bias study has been performed for each individual categories, where an alternative model is used to generate toys for an injected signal strength and in the next step generated toys are fitted with the nominal model. For each category 1000 toys have been generated. The alternative and nominal models, which are used for the bias study in the individual categories are listed in Tab. 9.10. The bias has been measured from the quantity $(\mu - \mu_{\text{inj}})/\Delta\mu$, where μ_{inj} is the injected signal strength with the alternative model, μ is the measured signal strength using the nominal model and $\Delta\mu$ is the associated errors of the measurements. The amount of bias in each category has been found to be less than 20% which is within the acceptable range. The measurements of the bias in TIGHT 2018 analysis categories are presented in Fig. 9.36 as example.

Table 9.10: The parametric function for the alternative and the nominal form of continuum background $m_{b\bar{b}}$ modeling in the TIGHT 2018 categories.

	TIGHT 2016		TIGHT 2018	
Categories	Alt.	Nom.	Alt.	Nom.
1	exp.pol1	exp.	exp.pol1	exp.
2	exp.pol1	exp.	exp.pol1	exp.
3	exp.pol1	exp.	exp.pol1	exp.
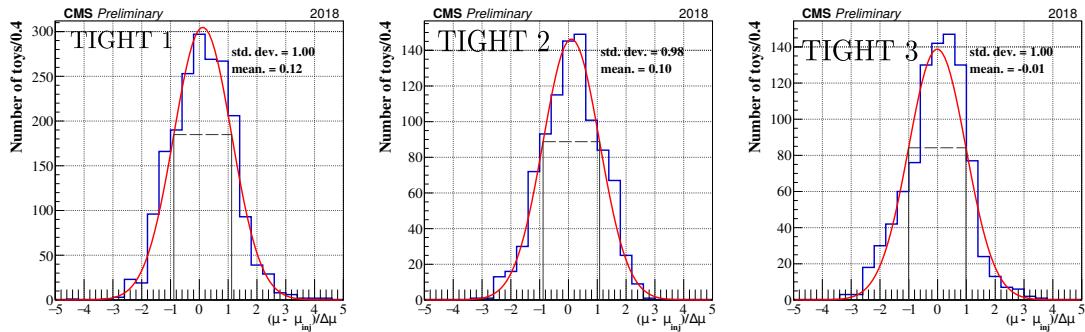
Figure 9.36: Distribution of pseudo-data fit results for TIGHT 2018 analysis categories, by using an alternative model to generate toys and fit the toys by the nominal model, while the injected signal strength is assumed unity.

9.9.4.1 Bias test with a different family of function

For the completeness of the bias study, another family of function, inverse polynomial has been used as the alternative model to fit the continuum. The order of the inverse polynomial is derived by the F-Test method in the each individual category. The measurements of the bias in all TIGHT 2018 analysis categories are presented in Fig. 9.37.
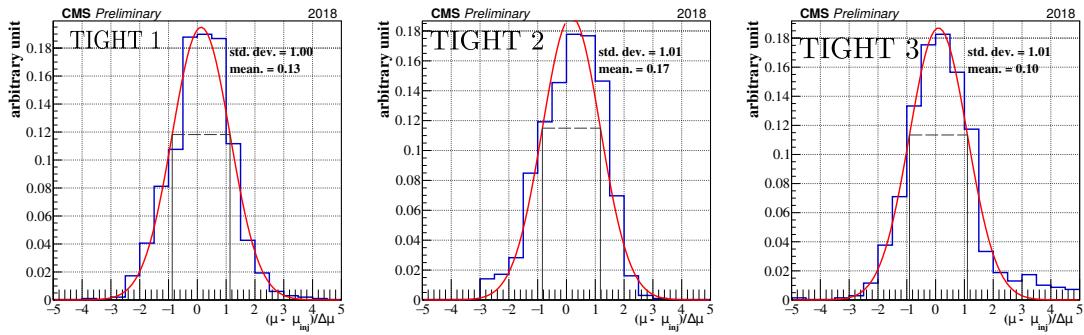
Figure 9.37: Distribution of pseudo-data fit results for 2018 using an alternative model to generate and fit by the nominal model, while the injected signal strength is assumed unity.

9.10   Systematic uncertainties

Several systematic uncertainties affect the derived results in two different ways, by impacting on the overall normalization of the different processes in each categories (normalization uncertainties) and by changing the shapes of the $m_{b\bar{b}}$ distributions (shape uncertainties) for the signal and the backgrounds.

Systematic uncertainties are incorporated in the signal extraction procedure via nuisance parameters (NP) with Gaussian or log-normal probability density functions (pdfs), which are treated according to the frequentest paradigm. The most dominant systematic uncertainties affecting this measurement are same as the HH analysis and described in details Sec. 7.10. Theoretical uncertainties arise from the limited precision in the computation of the inclusive and differential cross sections of the modeled processes and are fully correlated between event categories and the data taking periods. Experimental uncertainties are related to imperfect simulation of the detector response and the consequent inaccurate modeling of the reconstruction of physics objects and observable in the simulated samples.

The impact of the first 30 uncertainties on the signal strength of the inclusive measurement of

$H \rightarrow b\bar{b}$ process, $\mu_{Hb\bar{b}}^{incl.}$, is shown in Fig. 9.38. The most dominating systematic uncertainties for this analysis is due to the choice of the parton showering and the hadronization model for the VBF production. To evaluate this uncertainty arising from the choice of the showering and hadroinzation model we compared results obtained with two types of simulation: PYTHIA (nominal model) and HERWIG (alternative). The difference in the signal acceptance in each category, as predicted by the nominal and the alternative model, is propagated as the systematic uncertainty.
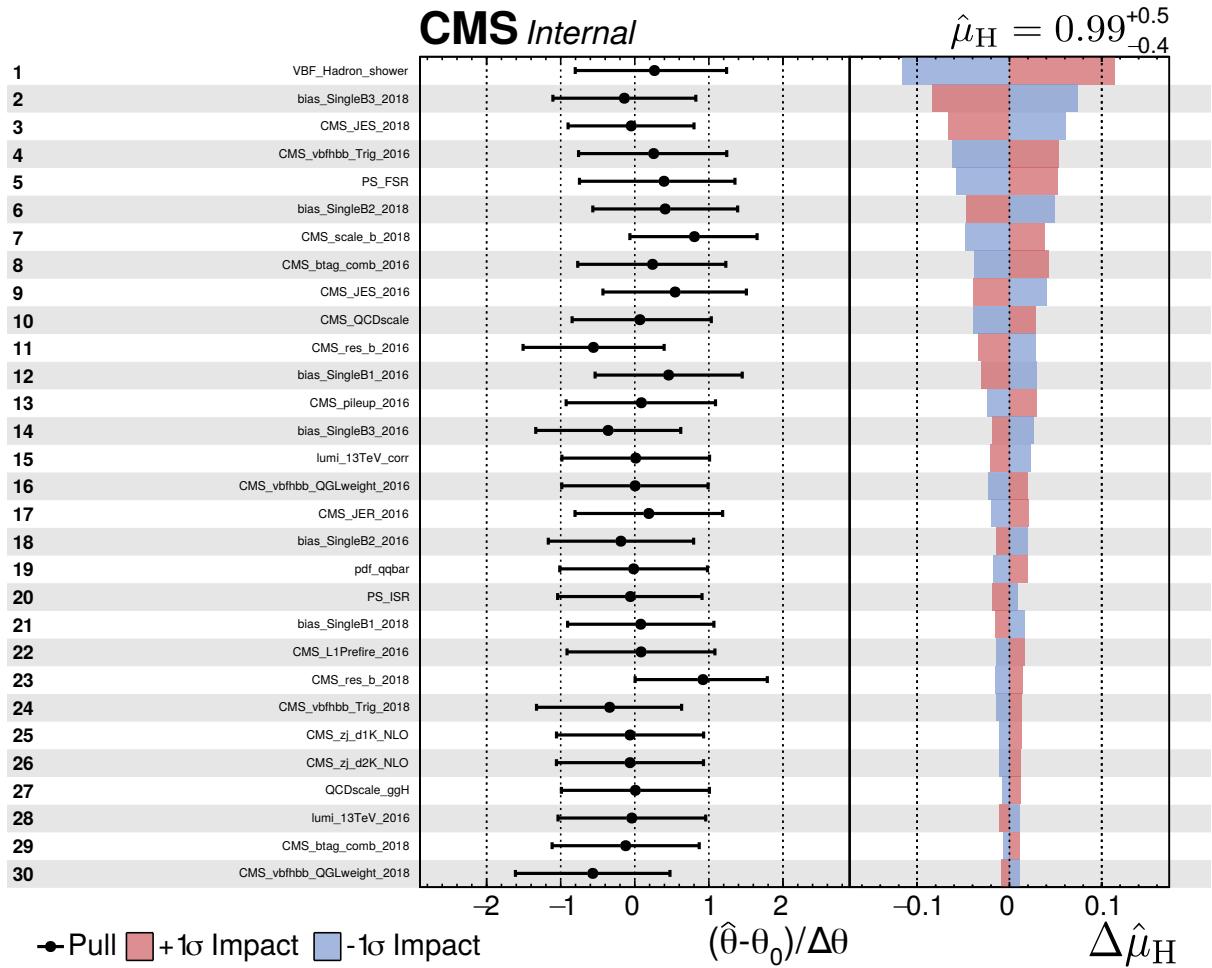
Figure 9.38: Ranking of the major systematic uncertainties appearing in this analysis based on their impact on the signal strength of the inclusive Higgs boson production ($\mu_{Hb\bar{b}}^{incl.} = \hat{\mu}_H$).

9.11   Results

This analysis is primarily sensitive to the VBF Higgs boson production followed by H $\rightarrow$ b$\overline{\text{b}}$ decay. The outcome of the measurement depends on the way the contribution from the ggH process is accounted for, and three scenarios are presented differing in the way the ggH process is treated in the signal extraction procedure.

9.11.1   Measurement of inclusive Higgs boson production

In the measurement of the inclusive Higgs boson production rate, the ggH process is considered as part of the signal. The fit is performed with an unconstrained signal strength modifier $\mu_{\text{Hb}\overline{\text{b}}}^{\text{incl.}}$ that simultaneously scales the yields of the VBF and the ggH events in all event categories. The parameter $\mu_{\text{Hb}\overline{\text{b}}}^{\text{incl.}}$ is the product of the inclusive production cross section of the Higgs boson and the H $\rightarrow$ b$\overline{\text{b}}$ branching fraction relative to the SM expectation. In the measurement, we have allowed the overall normalization of the Z+jets process $\mu_{\text{Zb}\overline{\text{b}}}$ to vary unconstrained. The two signal strengths $\mu_{\text{Hb}\overline{\text{b}}}^{\text{incl.}}$ and $\mu_{\text{Zb}\overline{\text{b}}}$ can be defined as,

- $\mu_{\text{Hb}\overline{\text{b}}}^{\text{incl.}} = \dfrac{\left[(\sigma_{qqH}+\sigma_{ggH}).\mathcal{B}r(\text{H}\rightarrow\text{b}\overline{\text{b}})\right]^{obs}}{\left[(\sigma_{qqH}+\sigma_{ggH}).\mathcal{B}r(\text{H}\rightarrow\text{b}\overline{\text{b}})\right]^{SM}}$

  The SM predicted cross sections of the VBFH (σ_{qqH}) and ggH (σ_{ggH}) processes and the value of the $\mathcal{B}r(\text{H}\rightarrow\text{b}\overline{\text{b}})$ are mentioned in Tab. 7.8 and 7.9 respectively along with the uncertainties.

- $\mu_{\text{Zb}\overline{\text{b}}} = \dfrac{\left[\sigma_{Z+jets}.\mathcal{B}r(\text{Z}\rightarrow\text{b}\overline{\text{b}})\right]^{obs}}{\left[\sigma_{Z+jets}.\mathcal{B}r(\text{Z}\rightarrow\text{b}\overline{\text{b}})\right]^{SM}}$

The analytic function used to fit the $m_{\text{b}\overline{\text{b}}}$ spectrum in the ith category is given by

$$F_i(m_{\text{b}\overline{\text{b}}}|\vec{\theta}) = N_i^{\text{QCD}}F_i^{\text{QCD}}(m_{\text{b}\overline{\text{b}}}|\vec{\alpha}_i) + \mu_{\text{Zb}\overline{\text{b}}}N_i^{\text{Zb}\overline{\text{b}}}(\vec{\theta})F_i^{\text{Zb}\overline{\text{b}}}(m_{\text{b}\overline{\text{b}}}|\vec{\theta}_S)$$
$$+ \mu_{\text{Hb}\overline{\text{b}}}^{\text{incl.}}(N_i^{\text{qqH}}(\vec{\theta}) + N_i^{\text{ggH}}(\vec{\theta}))F_i^{\text{Hb}\overline{\text{b}}}(m_{\text{b}\overline{\text{b}}}|\vec{\theta}_S). \tag{9.4}$$

The function includes the following category-dependent components:

- N_i^{QCD}: the normalization of the QCD multijet background extracted from the fit.

- $F_i^{\text{QCD}}(m_{\text{b}\overline{\text{b}}}|\vec{\alpha}_i)$: analytical parametric function modeling the QCD multijet background. The parameters $\vec{\alpha}_i$ of the function are obtained from the fit. The types of functions used to model the QCD multijet background are discussed in Sec. 9.9.3.

- $N_i^{\mathrm{qqH,ggH,Zb\bar{b}}}(\vec{\theta})$: predicted yields of the VBFH, ggH and Z+jets processes respectively. These yields depend on the nuisance parameters $\vec{\theta}$ that incorporate systematic uncertainties in the fit.

- $F_i^{\mathrm{Hb\bar{b},Zb\bar{b}}}(m_{\mathrm{b\bar{b}}}|\vec{\theta}_S)$: analytic functions modeling the $m_{\mathrm{b\bar{b}}}$ distribution in the samples of H $\rightarrow$ b$\bar{\mathrm{b}}$ and Z $\rightarrow$ b$\bar{\mathrm{b}}$ decays. The parameters of the analytic function are influenced by the nuisance parameters $\vec{\theta}_S$ associated with the uncertainties in the scale and resolution of the b jet energy regression.

The measurement of the inclusive Higgs boson production yields the best fit values

$$\mu_{\mathrm{Hb\bar{b}}}^{\mathrm{incl.}} = 0.99_{-0.24}^{+0.33}\,(\text{syst}) \pm 0.33\,(\text{stat}),$$

$$\mu_{\mathrm{Zb\bar{b}}} = 0.96 \pm 0.22\,(\text{syst}) \pm 0.22\,(\text{stat}).$$

The signal strengths, thus match with SM expectations very well within the uncertainties. The H $\rightarrow$ b$\bar{\mathrm{b}}$ signal, including contributions from the VBFH and ggH processes, is observed with a statistical significance of $2.6\,\sigma$, compared to the expected significance of $2.9\,\sigma$. The sensitivity of the measurement is driven by the TIGHT categories, while the LOOSE categories constrain the Z+jets background. Figures 9.39 and 9.40 show the results of the fit in the TIGHT 2016 and TIGHT 2018 categories. The results of the fit in the LOOSE 2016 Z2 and LOOSE 2018 Z2 categories are shown in Fig. 9.41.

Figure 9.42 shows the $m_{\mathrm{b\bar{b}}}$ spectrum combining all 18 analysis categories. Each category enters the combination with a weight $S/(S + B)$, where S and B are total signal and background yields obtained in a given category from the combined fit, respectively. The yields are computed by integrating the $m_{\mathrm{b\bar{b}}}$ distribution over the entire fitted range from 80 to 200 GeV. The distribution in data is compared with the fitted background-only and signal-plus-background models.

9.11.2 Measurement of VBF production when ggH production is constrained to SM expectations

The measurement of the exclusive VBFH production rate has been performed with the contribution from the ggH process constrained within theoretical and experimental uncertainties to the SM expectation. In this case, the analytic function employed to fit the $m_{\mathrm{b\bar{b}}}$ spectrum in
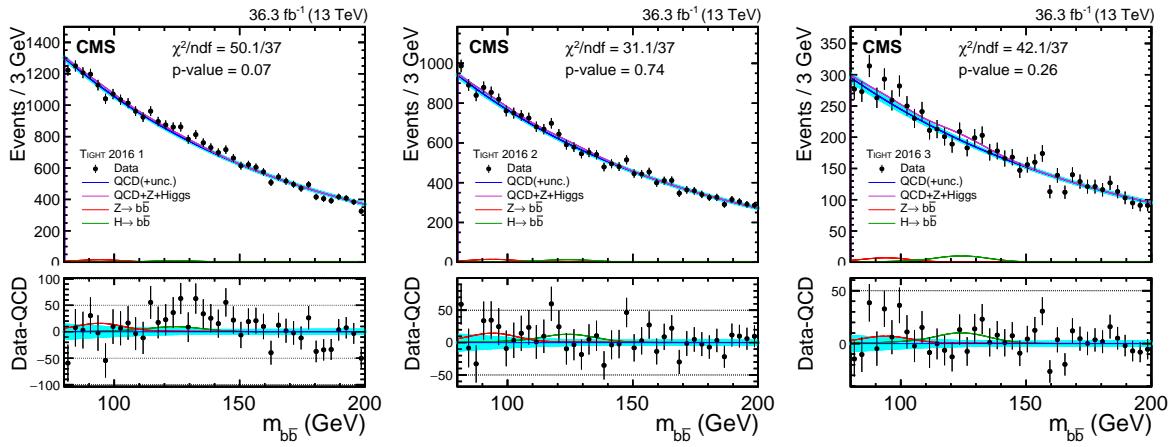
Figure 9.39: The $m_{b\bar{b}}$ distributions in three event categories: TIGHT 2016 1 (left), TIGHT 2016 2 (center), and TIGHT 2016 3 (right). The circles with error bars indicate data, the blue solid curve corresponds to the fitted nonresonant component of the background, dominated by the QCD multijet events; the shaded (cyan) band represents 1σ uncertainty band. The total signal-plus-background model includes contributions from $Z \rightarrow b\bar{b}$, $H \rightarrow b\bar{b}$, and the nonresonant component; it is represented by the magenta curve. The lower panel compares the distribution of the data after subtracting the nonresonant component with the resonant contributions of the $Z \rightarrow b\bar{b}$ background (red curve) and $H \rightarrow b\bar{b}$ signal (green curve).
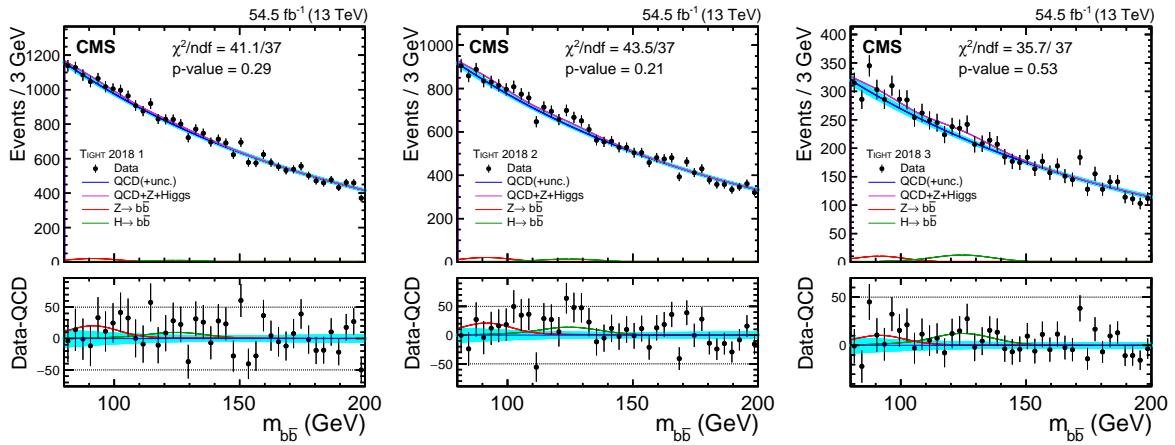
Figure 9.40: The $m_{b\bar{b}}$ distributions in three event categories: TIGHT 2018 1 (left), TIGHT 2018 2 (center), and TIGHT 2018 3 (right). A complete description is given in Fig. 9.39.
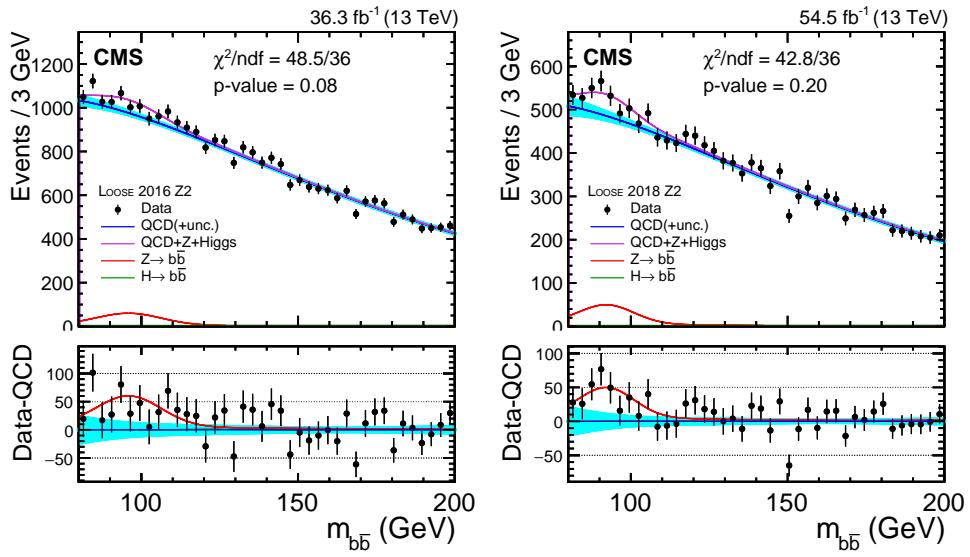
Figure 9.41: The $m_{b\bar{b}}$ distributions in two event categories: LOOSE 2016 Z2 (left) and LOOSE 2018 Z2 (right). A complete description is given in Fig. 9.39.
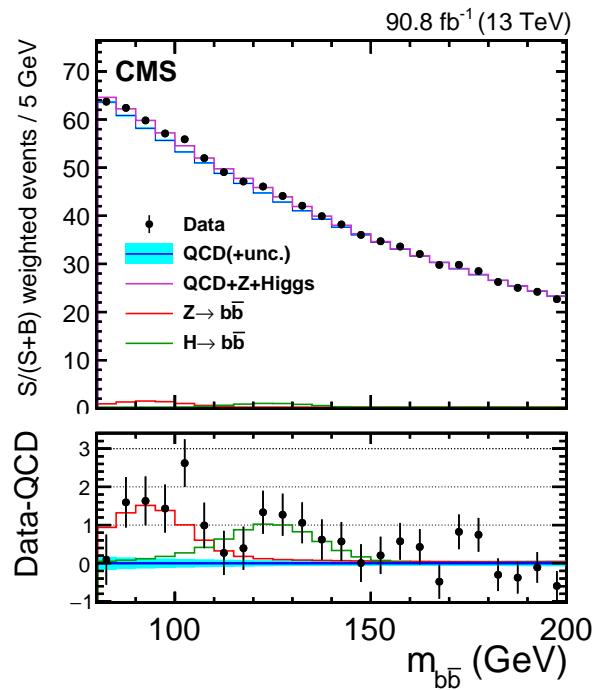
Figure 9.42: The $m_{b\bar{b}}$ distribution derived after weighted combination of all categories in the analysis. A complete description is given in Fig. 9.39.

category i is modified to be

$$F_i(m_{b\bar{b}}|\vec{\theta}) = N_i^{QCD}F_i^{QCD}(m_{b\bar{b}}|\vec{\alpha}_i) + \mu_{Zb\bar{b}}N_i^{Zb\bar{b}}(\vec{\theta})F_i^{Zb\bar{b}}(m_{b\bar{b}}|\vec{\theta}_S)$$

$$+ (\mu_{Hb\bar{b}}^{qqH}N_i^{qqH}(\vec{\theta}) + N_i^{ggH}(\vec{\theta}))F_i^{Hb\bar{b}}(m_{b\bar{b}}|\vec{\theta}_S). \quad (9.5)$$

The fit is performed with two unconstrained parameters: the signal strength modifier for the VBFH process ($\mu_{Hb\bar{b}}^{qqH}$) and $\mu_{Zb\bar{b}}$. The measurement yields

$$\mu_{Hb\bar{b}}^{qqH} = 1.01_{-0.28}^{+0.39}\,(\text{syst}) \pm 0.36\,(\text{stat}),$$

$$\mu_{Zb\bar{b}} = 0.96 \pm 0.22\,(\text{syst}) \pm 0.22\,(\text{stat}).$$

Again no deviations from the SM expectations has been found in the measurement. The VBFH signal is observed with a significance of $2.4\,\sigma$, while the expected significance is $2.7\,\sigma$. The best fit values of the signal strength modifiers for the different processes are shown in Fig. 9.43 (left). Figure 9.43 (right) presents the comparison of this result with that of CMS Run-1 and ATLAS Run-2.
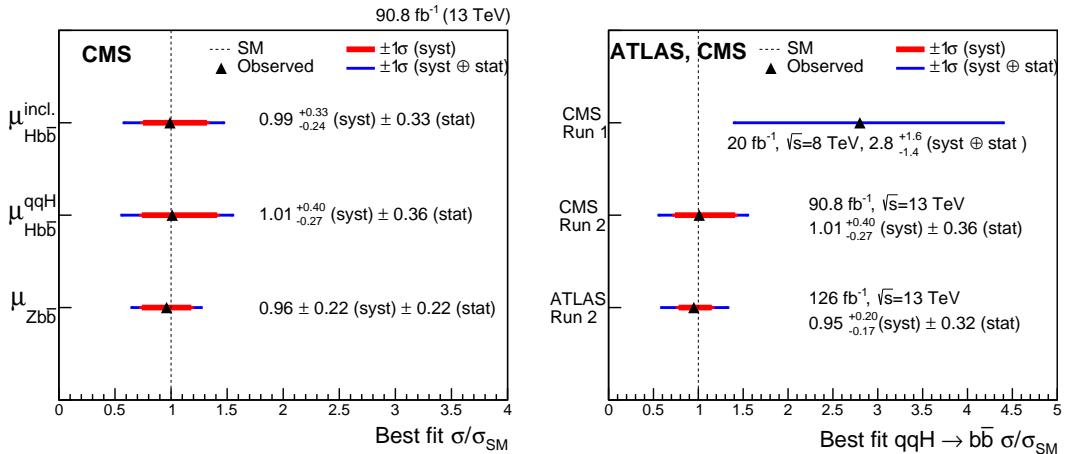
Figure 9.43: Left: The best fit values of the signal strength modifier for the different processes. The horizontal bars in blue and red colors represent the $\pm 1\,\sigma$ total uncertainty and its systematic component respectively. The vertical dashed line corresponds to the standard model prediction, ie $\mu = 1$. Right: The best fit values of the qqH signal strength modifiers for the CMS Run-1, ATLAS Run-2 and the new CMS Run-2 analyses.

9.11.3   Independent measurement of VBFH and ggH production

Additionally independent measurement of the VBFH and ggH production rates have been also estimated by fitting the $m_{b\bar{b}}$ spectra with three unconstrained parameters, $\mu_{Zb\bar{b}}$, $\mu_{Hb\bar{b}}^{qqH}$, and $\mu_{Hb\bar{b}}^{ggH}$,

with the latter being the signal strength for the ggH process. In this measurement the $m_{b\bar{b}}$ spectrum in the ith category is fitted with the function

$$
\begin{aligned}
F_i(m_{b\bar{b}}|\vec{\theta}) = {} & N_i^{\text{QCD}} F_i^{\text{QCD}}(m_{b\bar{b}}|\vec{\alpha}_i) + \mu_{Zb\bar{b}} N_i^{Zb\bar{b}}(\vec{\theta}) F_i^{Zb\bar{b}}(m_{b\bar{b}}|\vec{\theta}_S) \\
& + (\mu_{Hb\bar{b}}^{qqH} N_i^{qqH}(\vec{\theta}) + \mu_{Hb\bar{b}}^{ggH} N_i^{ggH}(\vec{\theta})) F_i^{Hb\bar{b}}(m_{b\bar{b}}|\vec{\theta}_S).
\end{aligned}
\tag{9.6}
$$

The fit yields

$$
\mu_{Hb\bar{b}}^{qqH} = 1.53^{+0.62}_{-0.70}\,(\text{syst}) \pm 0.54\,(\text{stat}),
$$

$$
\mu_{Hb\bar{b}}^{ggH} = -2.7^{+5.0}_{-2.6}\,(\text{syst}) \pm 3.5\,(\text{stat}),
$$

$$
\mu_{Zb\bar{b}} = 0.94 \pm 0.29\,(\text{syst}) \pm 0.25\,(\text{stat}).
$$

The best fit values of the signal strength modifiers for the different processes are shown in Fig. 9.44 (left). The negative best fit value for the signal strength modifier of the ggH process, i.e, $\mu_{Hb\bar{b}}^{ggH}$, is caused by under-fluctuations of the data for $m_{b\bar{b}}$ region between 100 and 150 GeV observed in some of the event categories targeting the ggH or Z+jets processes. A two-dimensional likelihood scan of $\mu_{Hb\bar{b}}^{qqH}$ and $\mu_{Hb\bar{b}}^{ggH}$ presented in Fig. 9.44 (right), shows that the measured value of $\mu_{Hb\bar{b}}^{ggH}$ is consistent with the SM prediction within $2\,\sigma$. Because of the degeneracy of the $m_{b\bar{b}}$ shape due to the VBFH and ggH processes and the non-negligible contributions from the ggH process in the categories targeting VBFH production, $\mu_{Hb\bar{b}}^{qqH}$ and $\mu_{Hb\bar{b}}^{ggH}$ exhibit strong anti-correlation. As a consequence, a downward shift in the measured value of $\mu_{Hb\bar{b}}^{ggH}$ causes an upward shift in the measured value of $\mu_{Hb\bar{b}}^{qqH}$.

9.12   Summary

A measurement of the Higgs boson (H) produced in the vector boson fusion (VBF) process and decaying to bottom quark-antiquark pair ($b\bar{b}$) was performed on proton-proton collision data sets collected by the CMS experiment at $\sqrt{s} = 13$ TeV and corresponding to a total integrated luminosity of 90.8 fb^{-1}. The analysis employs boosted decision trees (BDTs) to discriminate the signal against major background processes QCD multijet production and Z+jets events. The BDTs exploit the kinematic properties of the VBF jets, information of the b-tagged jets assigned to the H $\rightarrow$ b$\bar{b}$ decay, and the global event shape variables. Based on the BDT response, multiple event categories are introduced, targeting the VBF, gluon-gluon fusion (ggH), and Z+jets processes to achieve maximum sensitivity for the signal. While the VBF categories
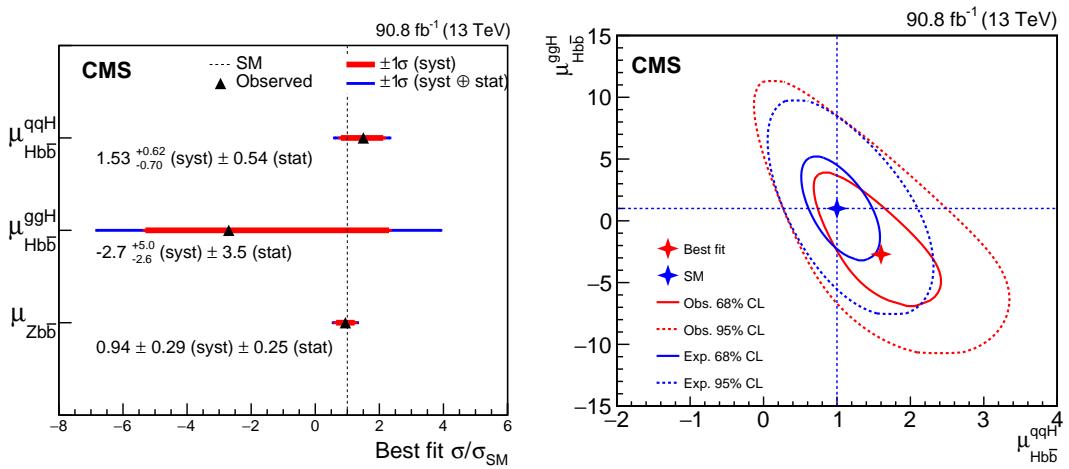
Figure 9.44: The best fit values of the signal strength modifier for the different processes, the horizontal bars in blue and red colors represent the $\pm 1\,\sigma$ total uncertainty and its systematic component and the vertical dashed line shows the SM prediction (left). The two-dimensional likelihood scan of $\mu_{\text{Hb}\bar{\text{b}}}^{\text{qqH}}$ and $\mu_{\text{Hb}\bar{\text{b}}}^{\text{ggH}}$, the red (blue) solid and dashed lines correspond to the observed (expected) 68% and 95% CL contours in the ($\mu_{\text{Hb}\bar{\text{b}}}^{\text{qqH}}, \mu_{\text{Hb}\bar{\text{b}}}^{\text{ggH}}$) plane (right). The SM predicted and observed best fit values are indicated by the blue and red crosses.

have the highest signal-to-background ratio, the Z+jets categories constrain the largest resonant background. The ggH categories enhance the sensitivity to the inclusive production of the Higgs boson in association with two jets.

The rate of VBFH production followed by $H \to b\bar{b}$ decay was measured with the ggH contribution constrained within the theoretical and the experimental uncertainties to the SM prediction. The qqH $\to$ qqb$\bar{\text{b}}$ signal strength, defined as the signal rate of the qqH $\to$ qqb$\bar{\text{b}}$ process relative to the prediction in the SM, is measured to be $\mu_{\text{Hb}\bar{\text{b}}}^{\text{qqH}} = 1.01_{-0.46}^{+0.55}$. The signal was observed with a significance of $2.4\,\sigma$, compared to the expected significance of $2.7\,\sigma$. In addition, inclusive Higgs boson production in association with two jets, followed by $H \to b\bar{b}$ decay, was measured by treating the ggH contribution as part of the signal. The inclusive signal strength was measured to be $\mu_{\text{Hb}\bar{\text{b}}}^{\text{incl.}} = 0.99_{-0.41}^{+0.48}$, corresponding to an observed (expected) significance of $2.6\,(2.9)\,\sigma$.

Chapter 10

Performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ in Run 2 data

10.1 Introduction

The missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) is the magnitude of negative vector sum of the transverse momentum of all the measured particles in the detector, which can be inferred from the momentum imbalance in the plane transverse to the beam direction. This undetectable energy may be caused either due to the neutrinos or by possibly undiscovered particles predicted by any new physics model such as a dark matter candidate. Thus a precise measurements of $p_{\mathrm{T}}^{\mathrm{miss}}$ is not only crucial for precision studies of SM physics (like, measurement of mass of W boson), but it is also essential for a large variety of searches targeting BSM physics (eg, production of dark matter particle or supersymmetric particles, like neutralino). The reconstruction of $p_{\mathrm{T}}^{\mathrm{miss}}$ object by the Particle Flow (PF) and also by the PUPPI algorithm has been described in Sec. 6.4.

10.2 Study of the $p_{\mathrm{T}}^{\mathrm{miss}}$ filters

Anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ events can also arise due to the detector noises from the detector subsystems, misalignment of the detector components and dead-cells which leads the anomalous hits or energy deposits in the detectors. These events can have large fake $p_{\mathrm{T}}^{\mathrm{miss}}$ and it can be a serious issue for searches and measurements of any physics phenomena involving $p_{\mathrm{T}}^{\mathrm{miss}}$ in the final state. To reject these noisy events with spurious signals and machine induced backgrounds, a set of dedicated algorithms have been employed using information from various subdetectors. These algorithms are called $p_{\mathrm{T}}^{\mathrm{miss}}$ filters which are recommended to be used in each physics analysis. The performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ filters in Run 2 is discussed in this section. The main purpose of $p_{\mathrm{T}}^{\mathrm{miss}}$

filters is to identify events with anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ in the pp collision data and to reject them. The list of $p_{\mathrm{T}}^{\mathrm{miss}}$ filters, recommended in Run 2 data analysis are listed below.

- **Primary vertex filter**: Events are slected with at least one well reconstructed primary vertex, based on quality criteria such as: minimum numbers of degrees of freedom > 4, maximum distance along z-direction (dz) from the interaction point < 24 cm and in the transverse direction (dxy) < 2 cm.

- **Beam halo filter**: Whenever two protons are colliding with each other, they are always surrounded by a cloud of other particles, mostly coming due to the collision of protons with the residual gas inside the beam pipe or the interaction with the beam pipe itself. To reduce this contamination the LHC machine working group is taking several measures in terms of radiation shielding. However the Halo muons are still able to reach the detector with sufficient energy (upto several TeV) and can deposit a substantial amount of energy (upto several hundred GeV) in ECAL, HCAL as well as produce hits in the muon chamber and cause momentum imbalance in the measurement. This leads to the production of large $p_{\mathrm{T}}^{\mathrm{miss}}$. The beam halo muon moves almost parallel to beam axis, as shown in Fig. 10.1. The beam halo filter searches for halo-like tracks by requiring a muon track approximately parallel to the beam that traverses both muon endcaps.
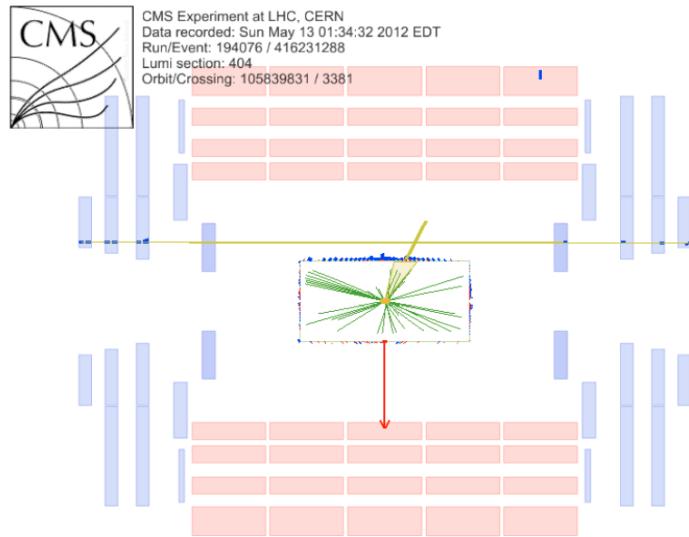
Figure 10.1: A display of an event in r-z view, having an halo-muon traversing parallel to the beam direction. Ref. [22].

- **HCAL noise filter**: Rejects sporadic noise from HCAL energy deposits, originating from

hybrid photodiode (HPD) pixel sensors, readout box (RBX) noise pulses etc. This kind of filter can be applied on the event in two ways either by cleaning the noise from the event or by rejecting the full event from the dataset having the noise.

- **ECAL trigger-primitive filter**: In some ECAL towers where precision readout is unavailable the trigger primitive (TP) information is used to determine the energy loss. If the TP value is saturated it indicates that the lost energy is significantly high enough to contribute to the higher end of $p_\mathrm{T}^\mathrm{miss}$ spectrum, such events are rejected by ECAL trigger-primitive filter.

- **ECAL endcap noise filter**: ECAL endcap super crystals giving anomalous pulses (mainly due to the issue from the applied High Voltage) to the vacuum phototriodes (VPTs) are identified and events where this occurs are rejected.

- **ECAL bad calibration filter**: Some seed crystals of ECAL clusters are known to receive large laser correction (mainly in EE region). These crystals are identified using a minimum energy threshold and subsequently events, where they appear, are rejected.

- **Bad PF muon filter**: Events with low-quality muon reconstruction that contribute to the $p_\mathrm{T}^\mathrm{miss}$ calculation as mis-reconstructed PF muon or PF charged hadron, are identified and rejected from the dataset.

10.3 Behavior of the $p_\mathrm{T}^\mathrm{miss}$ filters in data

The effect of $p_\mathrm{T}^\mathrm{miss}$ filters for the events with anomalous $p_\mathrm{T}^\mathrm{miss}$ are studied separately for the three years of data-taking during Run-2: 2016, 2017 and 2018, and reported in Ref. [155]. The year-wise distributions of $p_\mathrm{T}^\mathrm{miss}$ reconstructed by two different algorithms, PF and PUPPI before and after application of $p_\mathrm{T}^\mathrm{miss}$ filters are presented in Fig.s 10.2, 10.3 and 10.4 for 2016, 2017 and 2018 data respectively. The events are selected in the noise enriched region where an event should have at least one jet of $p_\mathrm{T} > 200$ GeV. From the above figures, it is evident that the fraction of events passing the filters is close to unity at the low $p_\mathrm{T}^\mathrm{miss}$ region while the application of filters have made the distributions falling smoothly and faster in the high $p_\mathrm{T}^\mathrm{miss}$ regime. Hence a substantial amount of events with high $p_\mathrm{T}^\mathrm{miss}$ value are rejected, where the effects of the detector or reconstruction issues are expected to be more pronounced. Subsequently, jet identification criteria are also applied on the leading jet of the filtered events
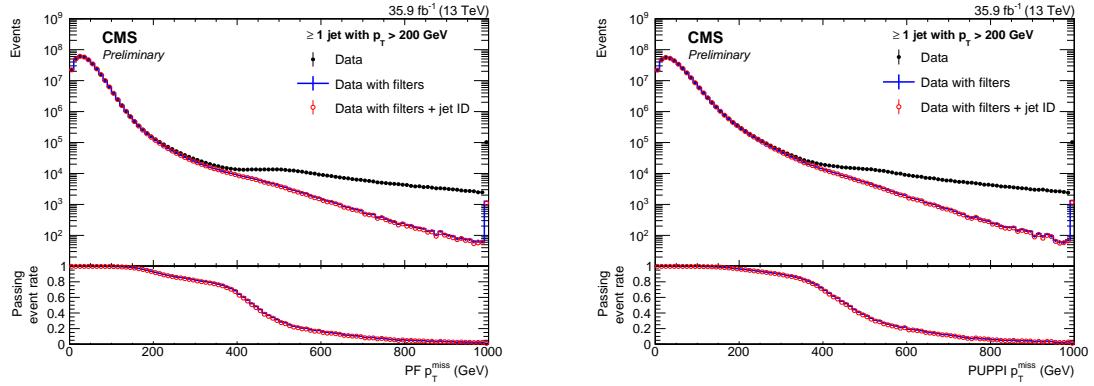
Figure 10.2: Distribution of p_T^{miss} in 2016 collision data, reconstructed with 2 algorithms: PF and PUPPI presented in left and right plots, in events with at least one jet with $p_T > 200$ GeV. The black dots and blue line correspond to values before and after the application of p_T^{miss} filters. The red markers correspond to the events where the highest p_T jet additionally satisfies the jet identification criteria. Last bin indicates overflow.
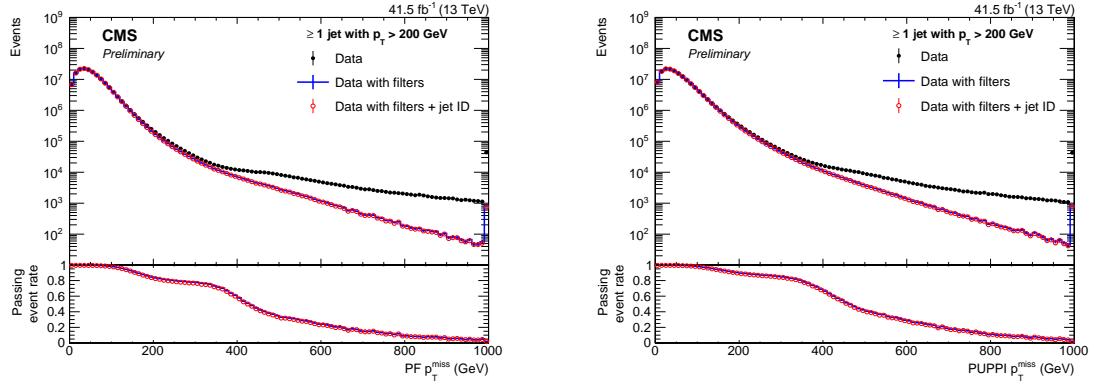
Figure 10.3: Distribution of p_T^{miss} in 2017 collision data, reconstructed with 2 algorithms: PF and PUPPI presented in left and right plots, in events with at least one jet with $p_T > 200$ GeV. The descriptions of the plots is same as in Fig. 10.2.
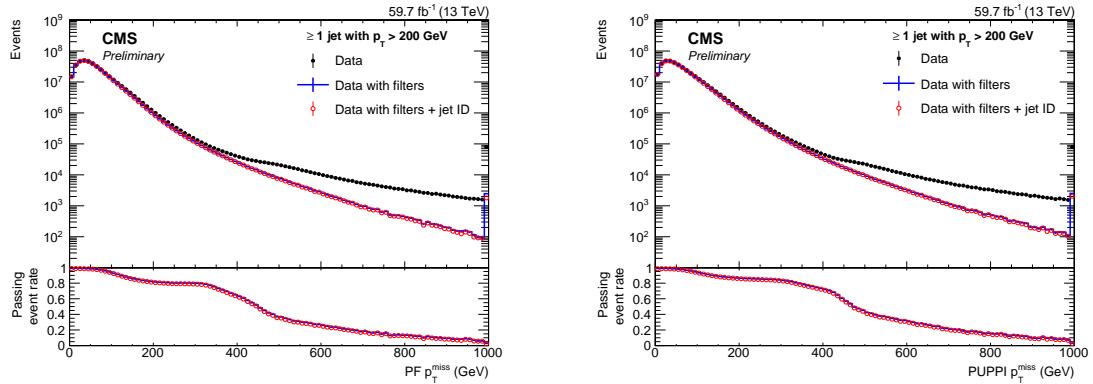
Figure 10.4: Distribution of p_T^{miss} in 2018 collision data, reconstructed with 2 algorithms: PF and PUPPI presented in left and right plots, in events with at least one jet with $p_T > 200$ GeV. The descriptions of the plots is same as in Fig. 10.2.

to clean up the events more; however no further difference is observed with respect to the situation with the application of only p_T^{miss} filters. From this study it is concluded that the p_T^{miss} filters are capable enough to filter out noisy events.

10.3.1 Some event displays of events rejected by a particular p_T^{miss} filter

In the following some of the event displays have been shown, which are rejected by different type of p_T^{miss} filters having large p_T^{miss}, caused by different noises from detector subsystems.
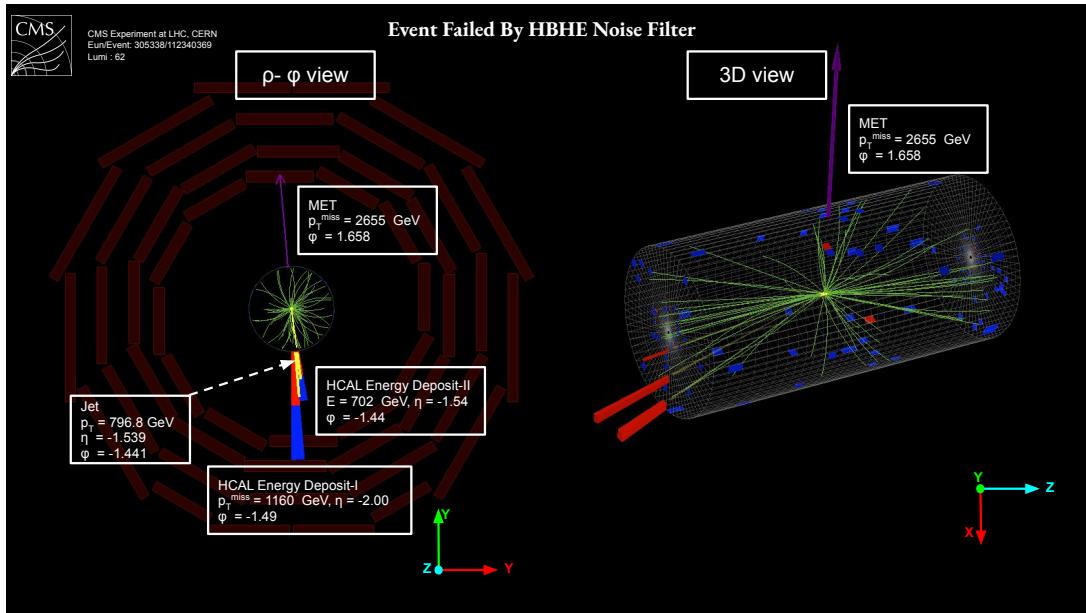
Figure 10.5: The event rejected by the *HBHE_Noise rejection* filter, where the large p_T^{miss} is originated due to the energy deposition of a jet in the HCAL, at the opposite hemispheres of the detector.
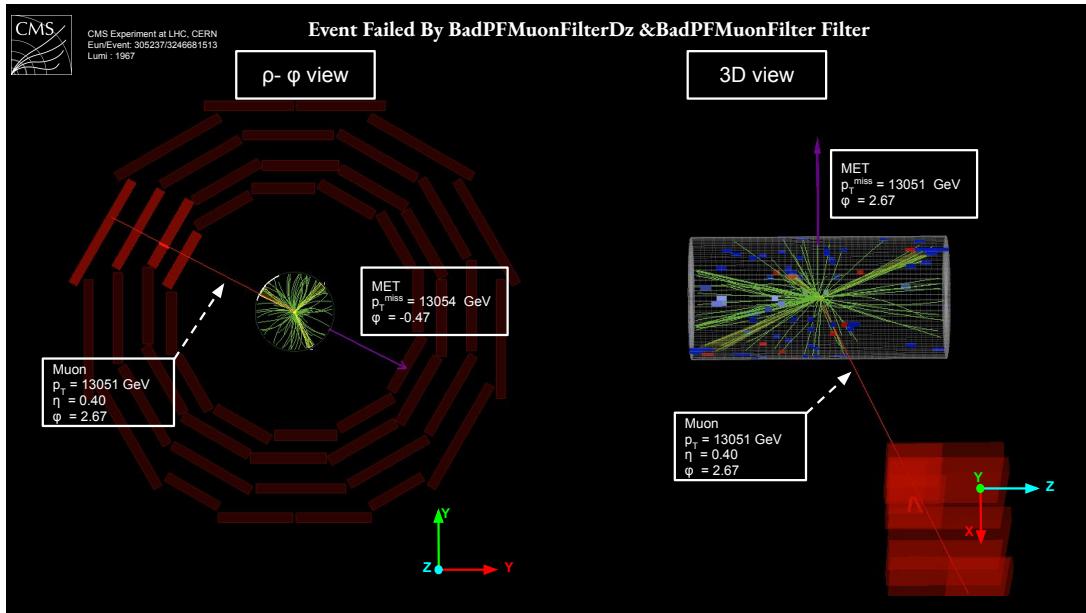
Figure 10.6: The event rejected by the *BAD PF-Muon candidate* filter identifying muons mistakenly reconstructed with very high p_T^{miss}. The filter decision is based on the presence of a high p_T muon with a poor quality track and a large momentum measurement, which produces a large p_T^{miss}.
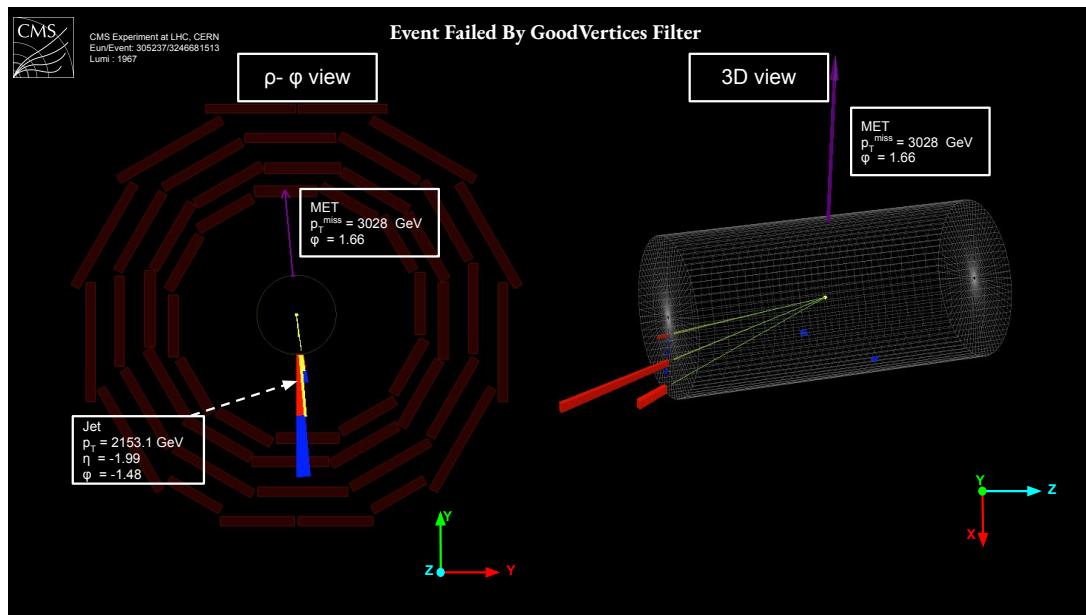
Figure 10.7: An event with high p_T^{miss}, rejected by a *Good Vertices* filter. Here no track is reconstructed which can be associated with the primary vertex.
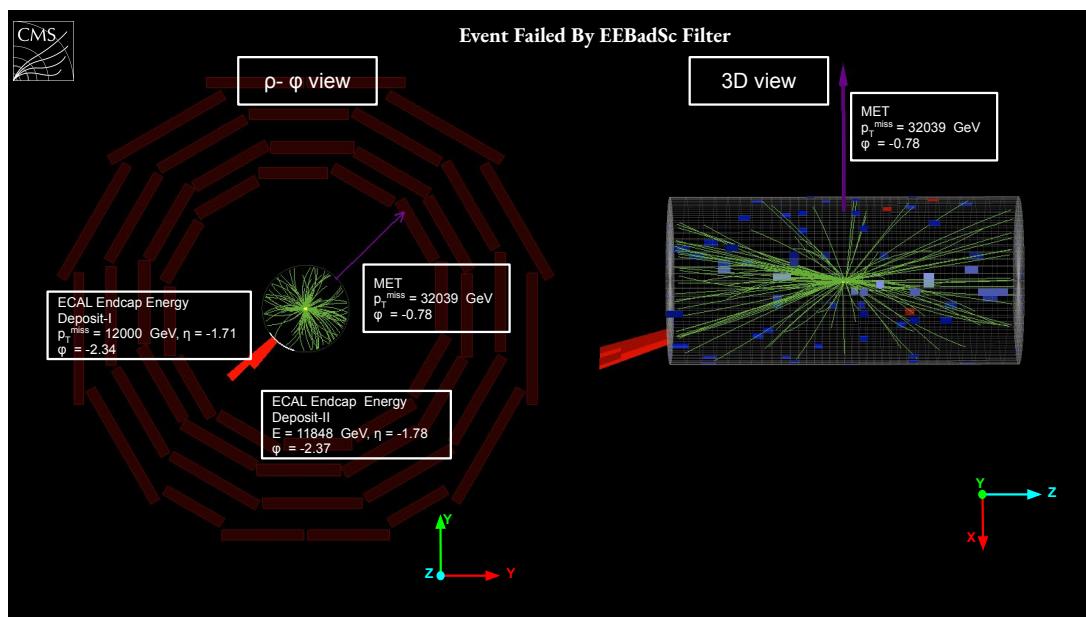
Figure 10.8: An event with high p_T^{miss}, rejected by *EE Bad SuperCluster* filter. The high p_T^{miss} in the event is produced due to the high energy deposit in the ECAL endcap (EE), which is detected and removed by this filter.

10.3.2 Good Event

A good mono-jet event which is passed by all recommended filters is shown in Fig. 10.9. Here the event consists only one high p_T (> 1000 GeV) jet and no other jets with $p_\mathrm{T} > 50$ GeV. The reason of $p_\mathrm{T}^\mathrm{miss}$ in this event is only due to the leading jet and it is back-to-back to it.
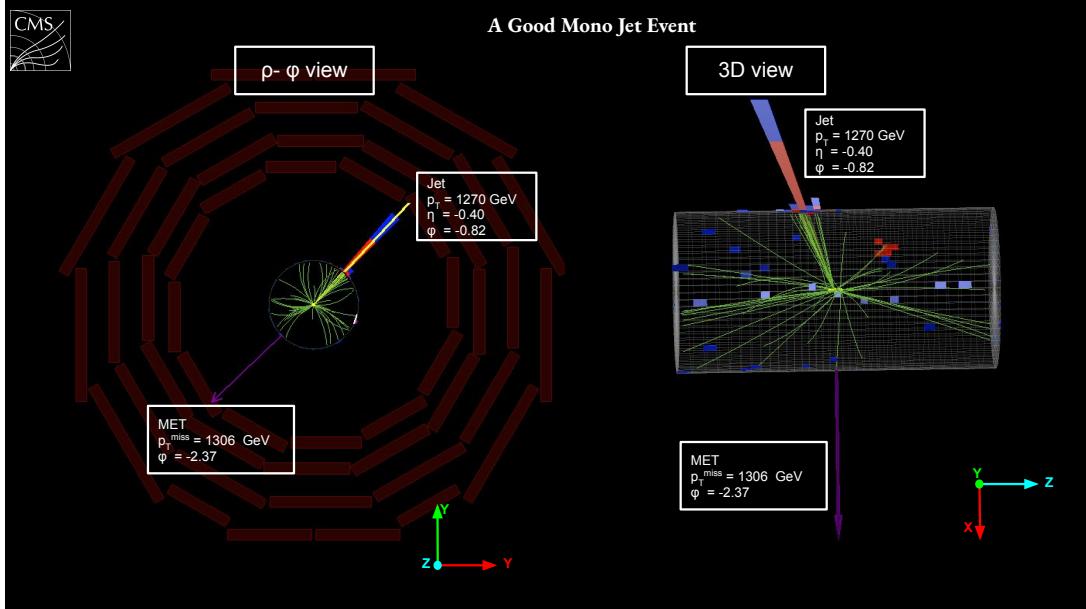
Figure 10.9: A good mono-jet event with high p_T jet and high $p_\mathrm{T}^\mathrm{miss}$ and passed by all recommended filters.

10.4 Performance of $p_\mathrm{T}^\mathrm{miss}$ filter

The performance of these filters is studied in Run 2 data. These filters reject many of the noisy and anomalous $p_\mathrm{T}^\mathrm{miss}$ events. Many of these effects have been simulated in the MC samples as well, so the $p_\mathrm{T}^\mathrm{miss}$ filters have been applied both in data and MC. Figure 10.10 shows PF $p_\mathrm{T}^\mathrm{miss}$ and PUPPI $p_\mathrm{T}^\mathrm{miss}$ distributions in data and MC before and after application of $p_\mathrm{T}^\mathrm{miss}$ filters. Without the use of these filters, the $p_\mathrm{T}^\mathrm{miss}$ distribution has a long tail both in data and MC, although the effect is more visible in data compared to MC. After the recommended filters have been applied, there is better agreement between data and simulated samples. The remaining discrepancies between data and MC can be attributed to use of the effect pf the LO samples for some of the MC processes. It can also be seen that the uncleaned samples have small peaks in the leading jet ϕ distribution near 0 and π, which is clearly a sign of anomalous events. Many of these events are rejected by beam halo filter.
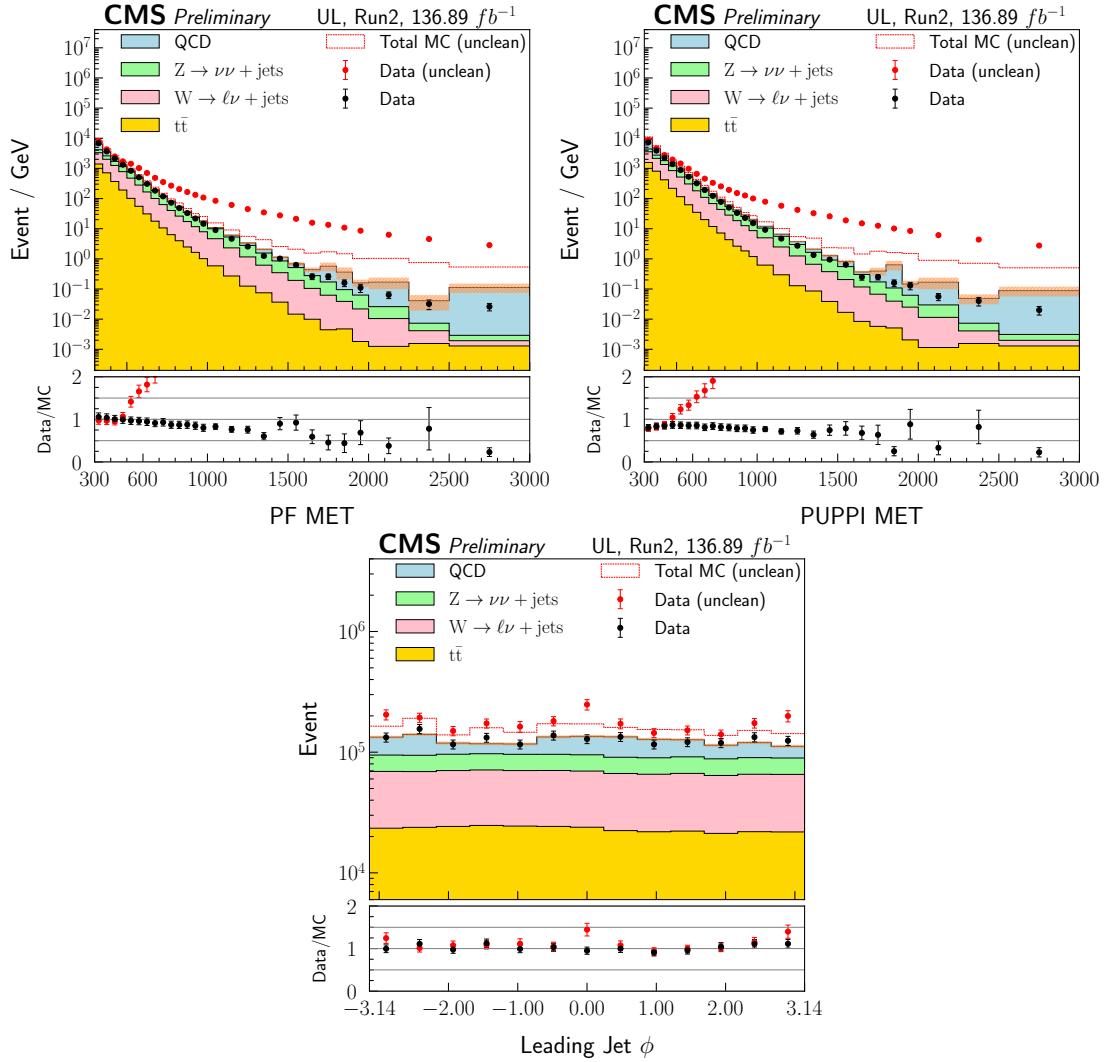
Figure 10.10: Dirstribution of $p_\mathrm{T}^\mathrm{miss}$ in data and MC before and after cleaning. The top, middle and bottom rows correspond to PF $p_\mathrm{T}^\mathrm{miss}$, PUPPI $p_\mathrm{T}^\mathrm{miss}$ and leading jet ϕ distributions, respectively.

10.5   Study of the $p_{\mathrm{T}}^{\mathrm{miss}}$ in γ+jets events

10.5.1   Introduction

The performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ can be studied in events where there is no genuine source of missing momentum in the event and the final state particles can be precisely calibrated. Hence, excellent measurement capability of photon energy in CMS detector can be utilized to measure the $p_{\mathrm{T}}^{\mathrm{miss}}$ in the inclusive single photon and jets final state, with restricted kinematics.

If the photon transverse momentum is $\vec{q}_T$, and transverse momentum of the hadronic recoil, defined as the vectorial sum of the transverse momenta of all particles except the photon, is $\vec{u}_T$, then from the momentum conservation in the transverse plane, $p_{\mathrm{T}}^{\mathrm{miss}}$ can be expressed using the relation,

$$\vec{q}_T + \vec{u}_T + \vec{p}_{\mathrm{T}}^{\mathrm{miss}} = 0 \tag{10.1}$$

The event topology of a γ +jets event is illustrated as a cartoon in Fig. 10.11.
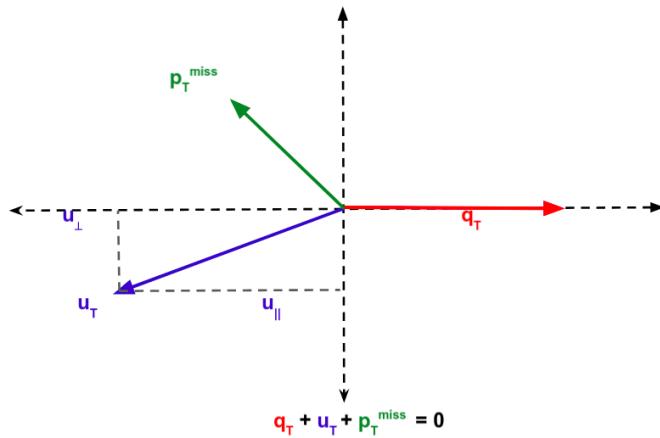
Figure 10.11: Illustration of γ+jets event kinematics.

The photon provides a very precise momentum scale of the event and an unique event axis along the direction of itself ($\vec{q}_T$). The projections of the hadronic jet activities parallel ($u_{\parallel}$) or perpendicular ($u_{\perp}$) to this axis are measures of the momentum balance of the event. The variables $u_{\parallel}$ and $u_{\perp}$ are expressed in terms of the x and y components of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and $\vec{q}_T$:

$$u_{\parallel} = (u_x q_x + u_y q_y)/q_T$$
$$u_{\perp} = (u_x q_y - u_y q_x)/q_T \tag{10.2}$$

where $u_x = -\left(p_x^{miss} + q_x\right)$ and $u_y = -\left(p_y^{miss} + q_y\right)$ are the components of the hadronic recoil in the transverse plane.

In the ideal condition, where there is no $\vec{p}_T^{\text{miss}}$ in the final state, from Eq. 10.2 it is found that $|u_\parallel| = |q_T|$ and $u_\perp = 0$.

10.5.2 Corrections and uncertainties

In Sec. 6.4 it has been explained that negative of the vector sum of all the reconstructed PF objects in an event is called "Raw PF $\vec{p}_T^{\text{miss}}$ ". The propagation of Jet Energy Scale (JES) correction into the $\vec{p}_T^{\text{miss}}$ reconstruction leads to "Type-1 corrected PF $\vec{p}_T^{\text{miss}}$ ".

XY-shift correction

An asymmetry in the ϕ distributions of $\vec{p}_T^{\text{miss}}$ is observed due to shifts of its x (p_X^{miss}) and y (p_Y^{miss}) components in the transverse plane. This effect is mainly due to anisotropic detector responses, inactive calorimeter cells or tracking regions, the detector misalignment, the displacement of the beam spot. Ideally $\vec{p}_T^{\text{miss}}$ should be isotropically distributed in azimuthal distribution in ϕ around the beam axis. But we observe the reconstructed $\vec{p}_T^{\text{miss}}$ does depend on the ϕ and the shape of this ϕ modulation is different in each year and even in different eras of same year within the data-taking.

The effect often increases with pile up, since more particles are counted in the $\vec{p}_T^{\text{miss}}$ reconstruction. It is particularly relevant for algorithms that do not discard particles associated with the pile up vertices such as the PF p_T^{miss}. A simple mitigation consists of evaluating these average X and Y shifts as a function of the number of the reconstructed number of primary vertices.

Fig. 10.12 shows the x and y components of $\vec{p}_T^{\text{miss}}$ as a function of the number of primary vertices, and a linear trend is observed. A fit to a first order polynomial ($a \times N_{PV} + b$) is performed. The x or y of a $\vec{p}_T^{\text{miss}}$ corrections are then simply defined as follows.

$$p_X^{miss}(corr) = p_X^{miss} - (a_1 * N_{PV} + b_1) \tag{10.3}$$

$$p_Y^{miss}(corr) = p_Y^{miss} - (a_2 * N_{PV} + b_2) \tag{10.4}$$

where N_{PV} is the number of reconstructed vertices in the event and the constants a_1, b_1, a_2 and b_2 depend on the run period.
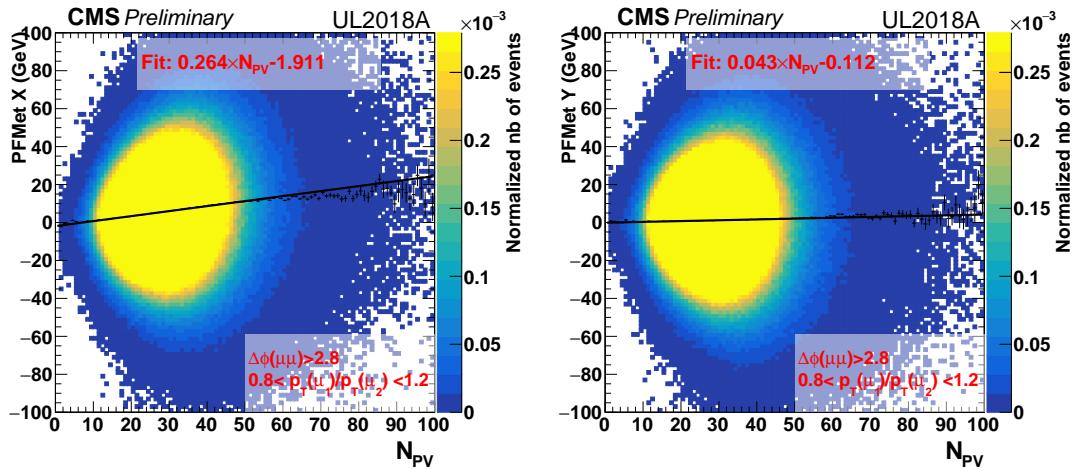
Figure 10.12: Distribution of the x and y component of the Type-1 PF p_T^{miss}, as a function of the number of reconstructed vertices for the various 2018 data taking period of era-A. The figures are taken from CMS_AN_ 2021/013 by Laurent Thomas.

10.5.3 Data and simulated samples

The full dataset collected by CMS experiment at the LHC in p p collision at the centre of mass energy of 13 TeV during the whole Run-2 corresponds to the total integrated luminosity ($\mathcal{L}$) of 137 fb^{-1} has been used.

The γ + jets signal sample, where one prompt photon produced in association with jets, has been generated using MADGRAPH5_AMC@NLO v-2.7 at the leading order (LO) accuracy, k-factor depending on photon p_T and η is applied to scale the cross section to the next-to-leading order (NLO) accuracy. To take the fake contribution of photons coming from jets in QCD induced multijet production process, a simulated sample has also been generated at LO using MAD-GRAPH5_AMC@NLO. Both the γ + jets and the QCD mulitjet simulations have been performed in the different bin of H_T (scalar sum of the p_T of all final state objects) to increase the event statistics in the samples.

The events samples corresponding to the single prompt photon production in association with a pair of top quarks ($t\bar{t}$ +γ), with single top quark (t +γ) and with a vector boson (V+γ, V = W/Z), are produced at NLO accuracy using MADGRAPH5_AMC@NLO. The possibility of a lepton faking as a photon has been taken into account by V+jets production at NLO, where V = W/Z decays to lepton.

Subsequently that all the simulated processes are hadronized by PYTHIA v-8.212 and are passed through the full detector simulation using by GEANT4 package, within the framework of standard CMS software.

10.5.4 Event selection

Both the data and MC events are selected with a set of HLT triggers requiring at least one isolated photon having a transverse momentum, denoted as q_T, greater than 50/75/90/120/165 GeV. As the production cross section of photons at the LHC is very high at the relevant p_T range, to control the rates of the HLT all of them are pre-scaled. A pre-scale factors are applied an event-by-event basis depending on the HLT, which is fired to select the event.

After the HLT requirements events are selected with a set of offline requirements on photons and jets. The selection conditions are mentioned below.

Selection of photon

Events are first selected with exactly one tightly identified and isolated photon. Photons with transverse momentum $q_T > 50$ GeV and $|\eta| < 1.44$ are selected to pick events with highly energetic photons only in the barrel region; endcap photons are not considered in order to suppress the noise.

The photon is required to pass an MVA-based photon identification score corresponding to the 70.24% efficiency of selecting prompt photons with 90% background rejection. Here backgrounds are generally the jets faking as photons.

In addition with the photon identification and isolation, the pixel track veto is applied to ensure that it is not an electron. This requirement rejects any photon having at least two hits found at the inner layers of the tracker (pixel detectors).

Selection of jets and p_T^{miss}

The PF jets reconstructed by anti-k_T jet clustering algorithm with the distance parameter D= 0.4. Several quality criteria, called PF Jet identification and mentioned in Tab. 10.1, is applied on the jets to reject fake, badly reconstructed and noise jets. This retains 98 - 99% of real jets. Jets are required to have $p_T > 40$ GeV and inside the tracker acceptance region of $|\eta| < 2.4$. The jets are corrected by the recommended Jet Energy Corrections (JEC) for each year separately.

No requirements on the p_T^{miss} is applied for this study.

Table 10.1: PF jet identification requirements

Variables	Values
Neutral Hadron fraction	< 0.90
Neutral EM fraction	< 0.90
Number of constituents	> 1
Muon fraction	< 0.80
Charged Hadron fraction	> 0
Charged Multiplicity	> 0
Charged EM fraction	> 0.80

Electrons and Muons

To reduce the contamination from top and electroweak (EW) backgrounds, where a lepton can fake as photons, an event is rejected if it contains an identified electrons or muons with $p_T > 10$ GeV. The basic requirements for the identification of electrons and muons are listed in Tab. 10.2 and Tab. 10.3 respectively.

Table 10.2: Identification requirements for electrons

Variables	$|\eta| < 1.479$	$1.479 < |\eta| < 2.5$
full 5×5 $\sigma_{i\eta i\eta}$	<0.0126	<0.0457
H/E	<0.05 + 1.16/E	<0.05+2.54/E
Relative isolation	<0.198+0.506/p_T	<0.203 + 0.963/p_T
Number of missing hits	≤ 2	≤ 3
Pass convertion veto	yes	yes
d_{xy}(cm)	< 0.05	< 0.1
d_z (cm)	< 0.1	< 0.2

The definitions of the variables mentioned in Tab. 10.2 and Tab. 10.3, are explained in Sec. 7.5.

10.5.5 Systematic uncertainties

p_T^{miss} is a composite object which can not be measured directly, while it is reconstructed from all of the observed particles in an event. So the systematic uncertainties are highly dependent on the final state objects and topology of the events. In this case two most dominant systematic

Table 10.3: Identification requirements for muons

Variable	$|\eta| < 2.5$
d_{xy} (cm)	< 0.05
d_z (cm)	< 0.1
Relative isolation ($\Delta R = 0.3$)	< 0.4
tracker/ global muon	yes

uncertainties, taken into account, are:

- Jet energy scale (JES) uncertainties related to the selected jet in the event.

- Unclustered energy uncertainties due to the energy scale of low-energy particles.

10.5.6 Data-MC distribution

In this section the data and MC are compared while reconstructing p_T^{miss} with different algorithms. Figures .10.13, 10.14, 10.15 and 10.16 show the "Raw PF", "Type-1 corrected PF", "XY-shift corrected PF" and "Puppi" p_T^{miss} distributions (left) and their ϕ distributions for all three years.

Figure. 10.15 (left) shows the reduction in the ϕ modulation of the p_T^{miss} after the XY-shift correction while it has very small impact on the p_T^{miss} distribution itself. From the Fig. 10.16 it is evident that Puppi algorithm already takes into account the linear dependnce of p_T^{miss} x and y components with the pile up vertices, thus ensuring that no additional XY-shift correction is needed to correct the ϕ distribution.
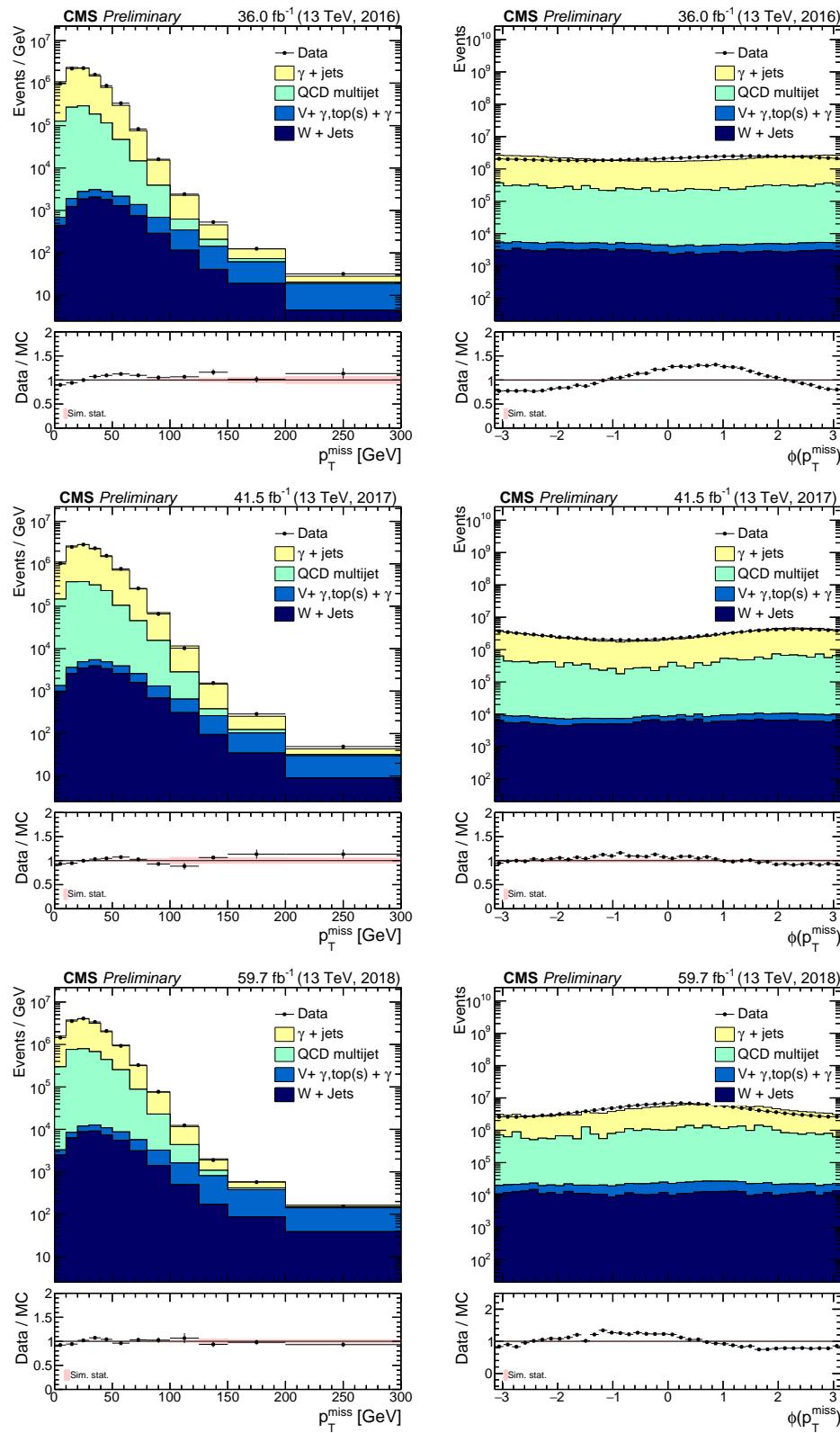
Figure 10.13: Raw PF p_T^{miss} distributions for 2016 (top), 2017 (middle) and 2018 (bottom), different colours represents different processes. Lower panels show the uncertainties.
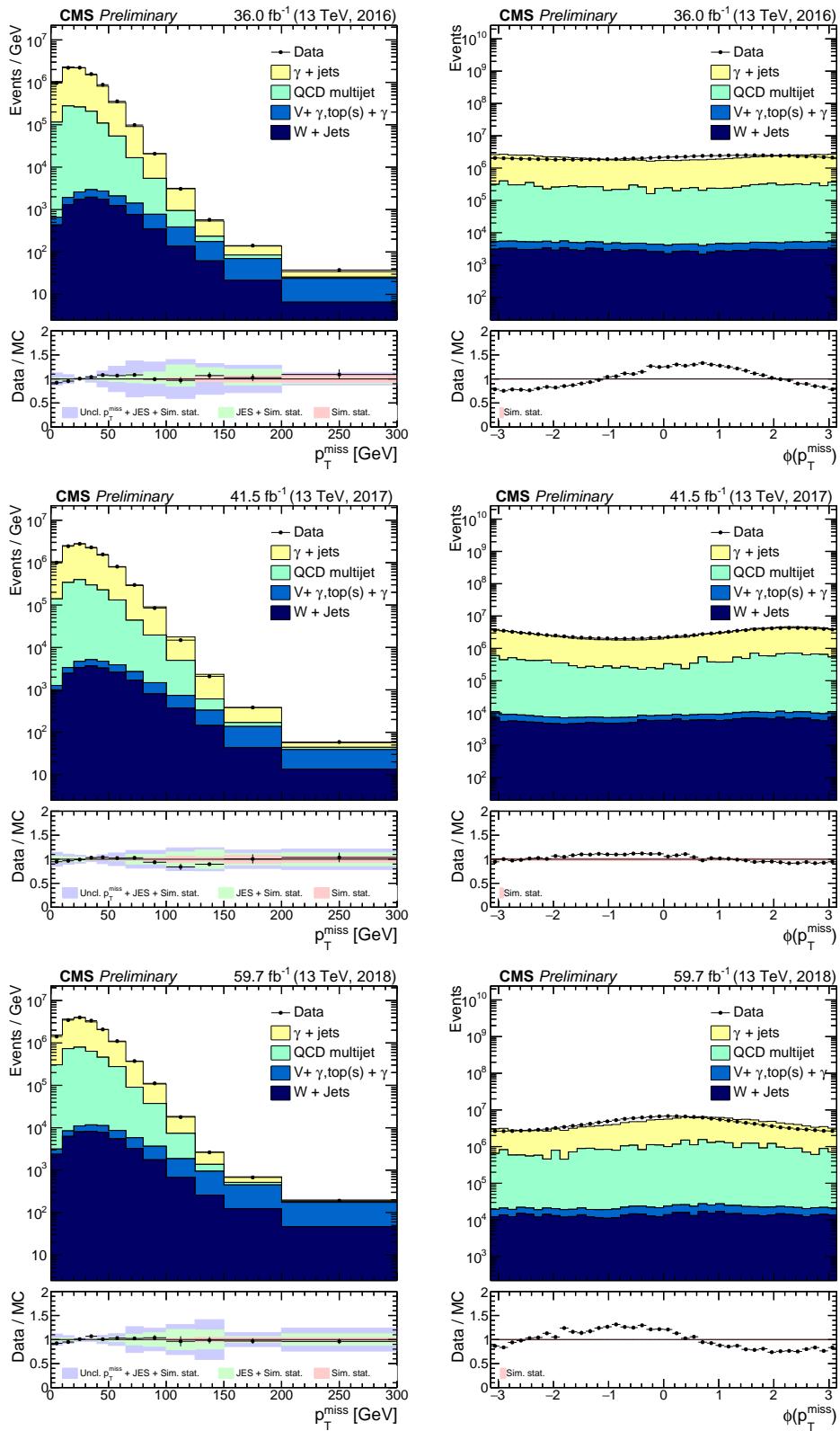
Figure 10.14: Type-1 corrected PF p_T^{miss} distributions for 2016 (top) , 2017 (middle) and 2018 (bottom), different colour represents different processes. Lower panels show the uncertainties.
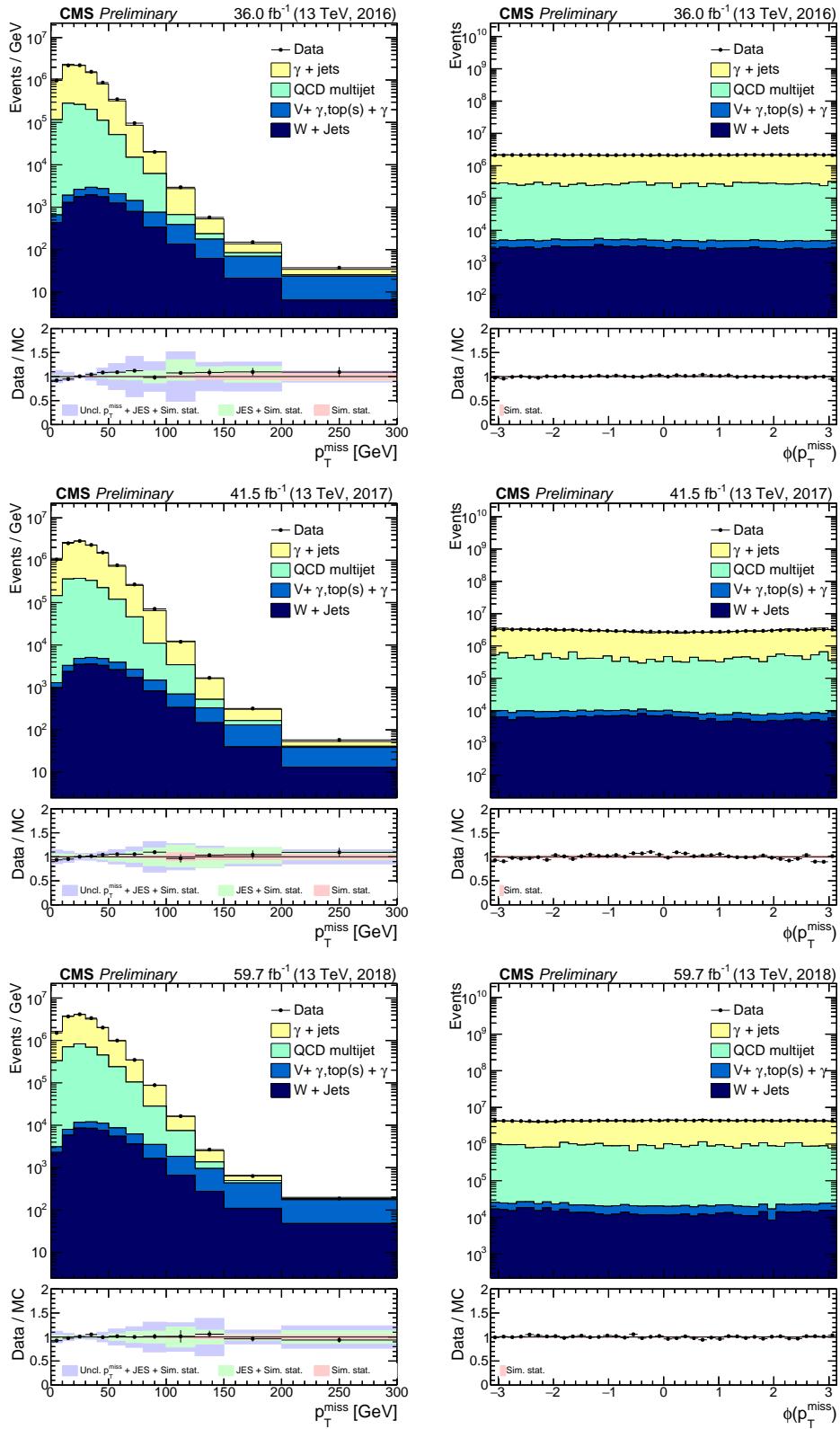
Figure 10.15: PF p_T^{miss} distributions after XY-shift correction for 2016 (top) , 2017 (middle) and 2018 (bottom), different colour represents different processes. Lower panels show the uncertainties.
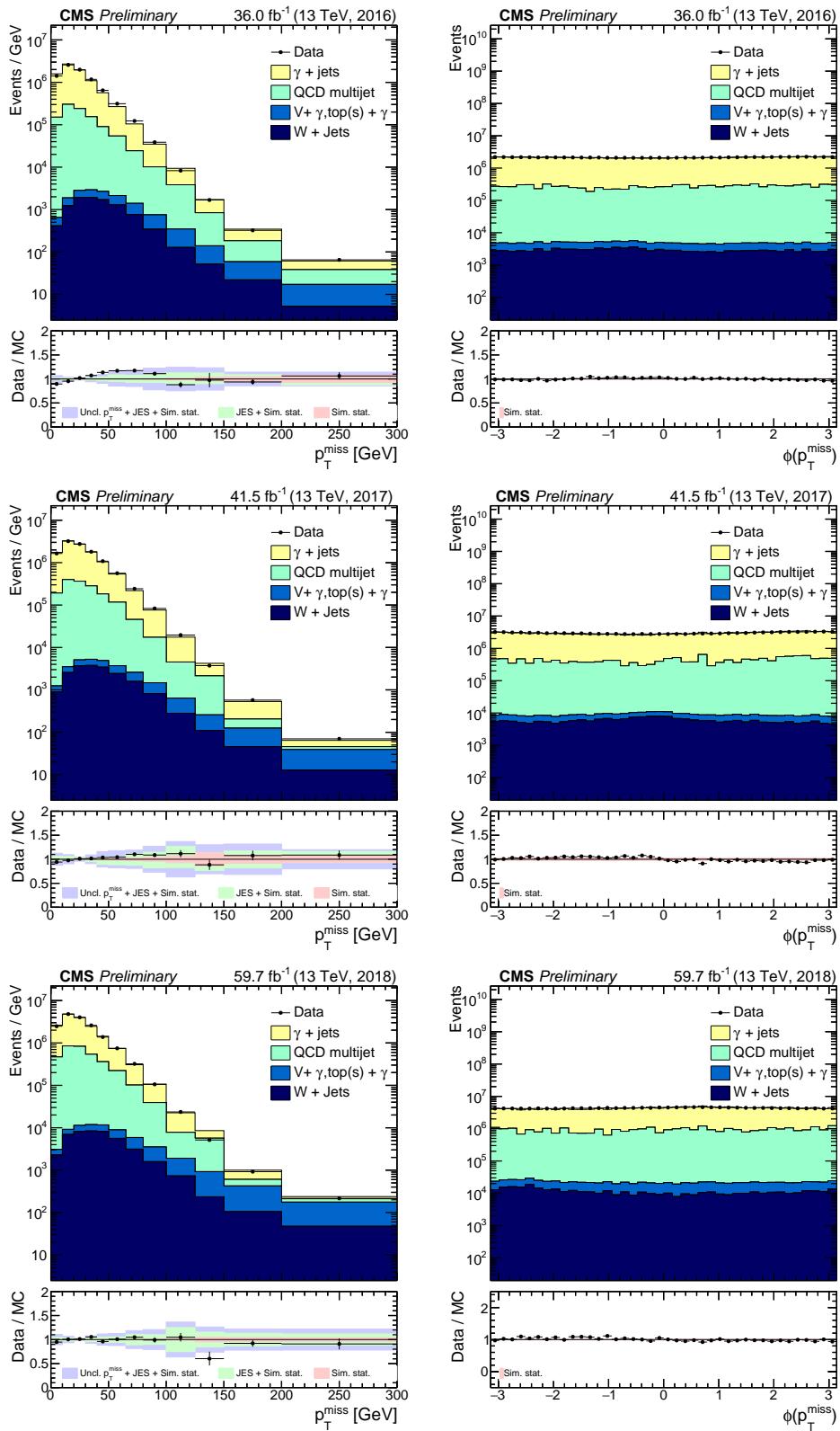
Figure 10.16: PUPPI p_T^{miss} distributions for 2016 (top) , 2017 (middle) and 2018 (bottom), different colour represents different processes. Lower panels show the uncertainties.

10.5.7 Performance of the p_T^{miss} reconstruction

p_T^{miss} response

The quantity $-\langle u_\parallel \rangle / q_T$ is denoted as the scale or response of the p_T^{miss} reconstruction algorithm. The distribution of this variable against photon transverse momentum q_T is referred to as the response curve, ideally which should be a straight line with zero slope at $-\langle u_\parallel \rangle / q_T = 1$. Figure. 10.17 represents the distribution of the response as function of q_T for three different data-taking years. It has been found that the response of "Type-1 corrected PF" p_T^{miss} is significantly better compared to the "Raw PF" p_T^{miss}, which ensures the importance of the propagation of Jet energy corrections. The response of "Puppi" p_T^{miss} is slightly better compared to the "Type-1 corrected PF" p_T^{miss} at high values of q_T, above 150 GeV.
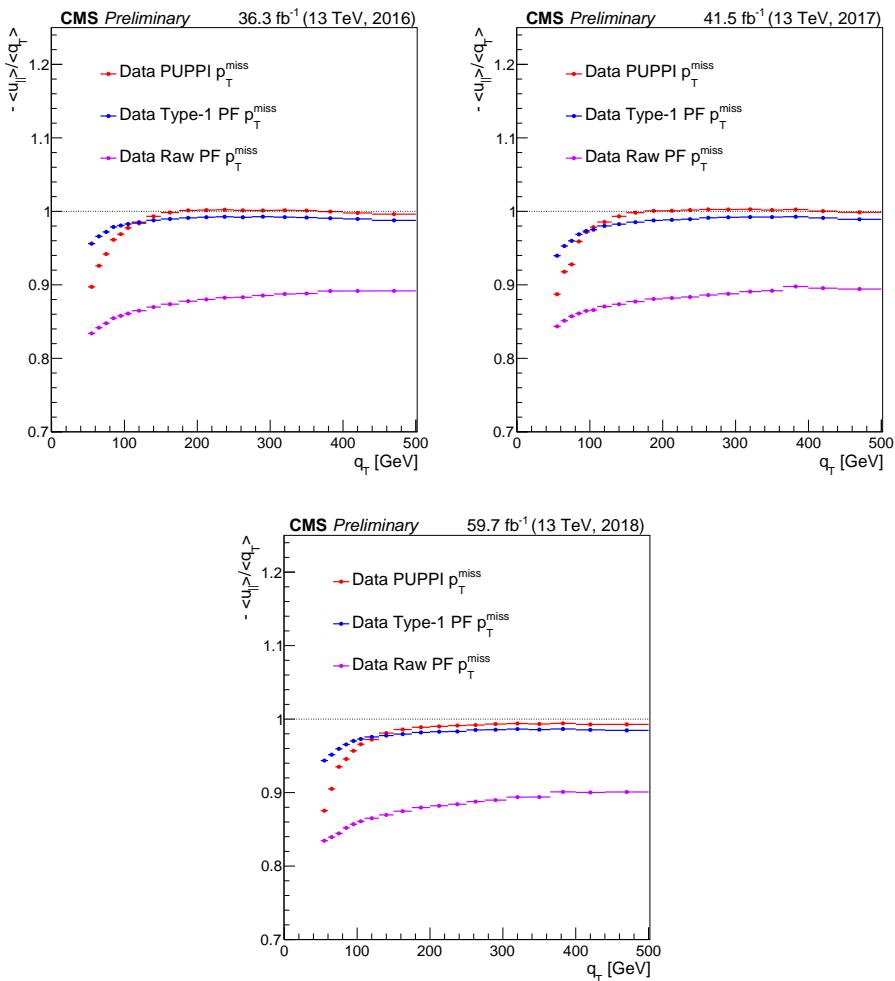
Figure 10.17: The p_T^{miss} responses against photon transverse momentum (q_T) for "Raw PF" (with magenta line), "Type-1 corrected PF" (with blue line) and "Puppi" (with red line) p_T^{miss} in three data-taking years.

p_T^{miss} resolution

The two quantities, $u_\parallel + q_T$ and $u_\perp$ are each ideally equal to zero, but due to the momentum imbalance, these distributions have a finite width across zero. These widths, $\sigma(u_\parallel)$ and $\sigma(u_\perp)$ represent the resolution of the p_T^{miss} both in parallel and perpendicular direction to the photon transverse momentum. These can be parametrized as a function of either q_T or the number of reconstructed vertices (N_{vtx}) in the selected events.

Figures 10.18 and 10.19 show the parallel and perpendicular components of p_T^{miss} with respect to the q_T as a function of number of reconstructed primary vertices, N_{vtx}.
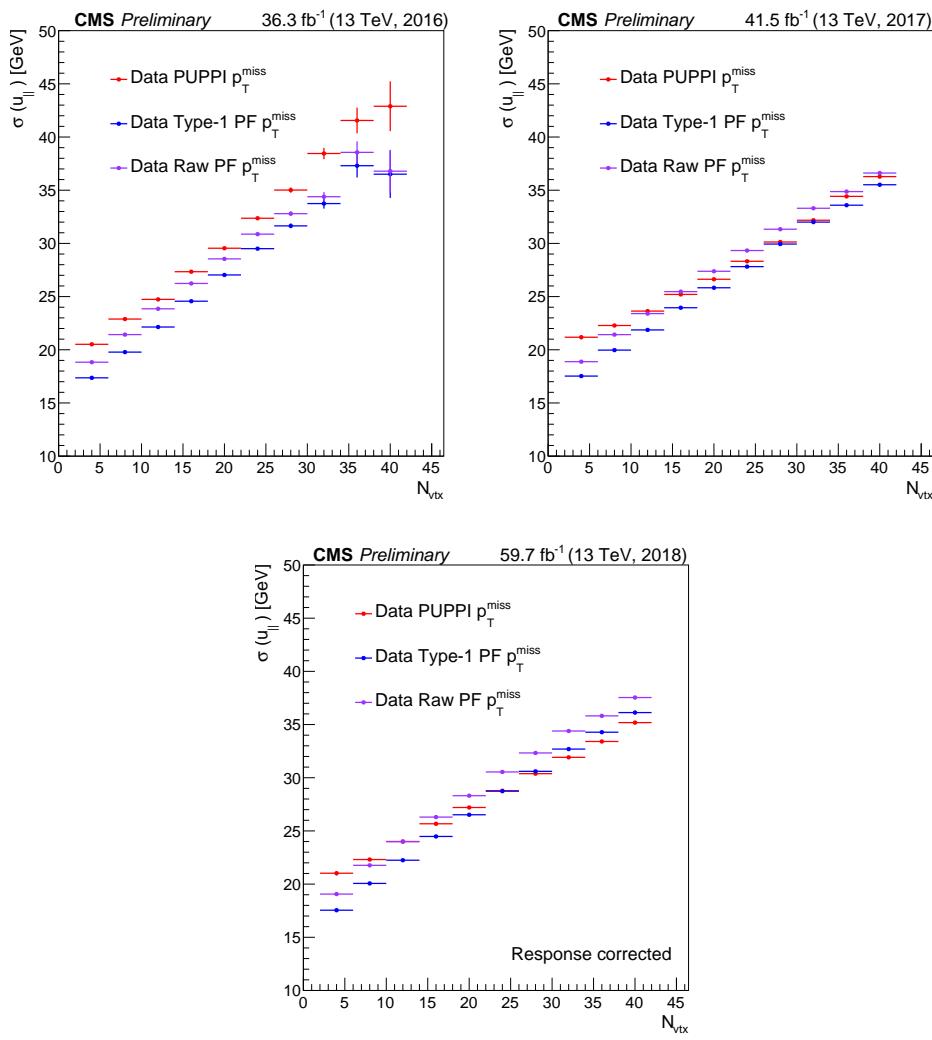
Figure 10.18: Parallel component of p_T^{miss} resolution $\sigma(u_\parallel)$, as a function of the number of reconstructed number of primary vertices (N_{vtx}) for "Raw PF" (with magenta line), "Type-1 corrected PF" (with blue line) and "PUPPI" (with red line) p_T^{miss} in three data-taking years.
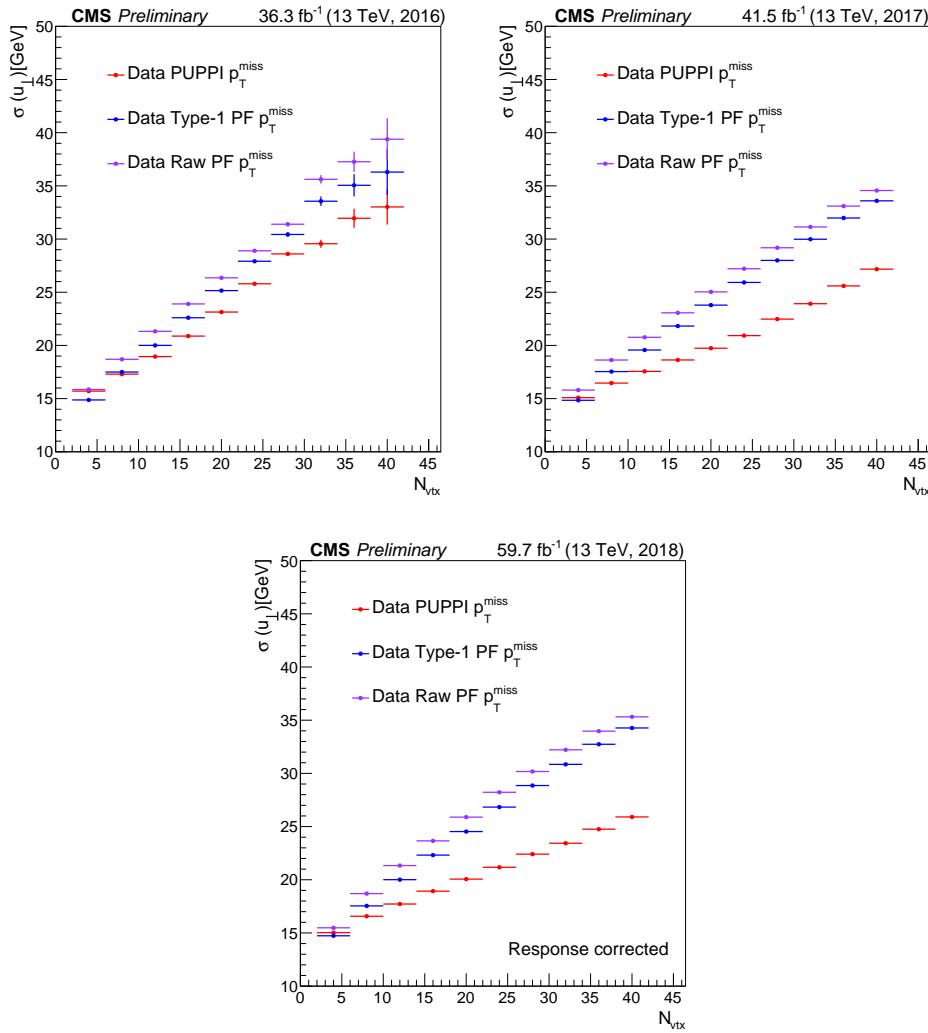
Figure 10.19: Perpendicular component of p_T^{miss} resolution $\sigma(u_\perp)$, as a function of number of reconstructed number of primary vertices for "Raw PF" (with magenta line), "Type-1 corrected PF" (with blue line) and "PUPPI" (with red line) p_T^{miss} in three data-taking years.

10.6   Summary

In this study, the efficiency of the p_T^{miss} filters has been derived. It is found that they reject the events with anomalous high p_T^{miss} due to the noisy detector responses and the mis-reconstructions of different objects. After the application of p_T^{miss} filters good data-mc agreement has been achieved. These filters are recommended to be used for all physics analyses.

In addition the several estimation of p_T^{miss} are also studied in the events containing a photon and jets (γ + jets). It has been found that "Type-1 corrected PF" p_T^{miss} has the response close to unity, which ensures the importance of jet energy scale (JES) correction to the PF p_T^{miss} reconstruction. It has also been found that the better pileup mitigation of the "PUPPI" p_T^{miss} has better resolution than the PF reconstruction.

Chapter 11

Conclusion

This thesis covers an extensive study of the measurement of two important properties of the Higgs boson using the p-p collision data at $\sqrt{s}$=13 TeV collected by the CMS experiment at the CERN LHC during Run 2. They are both timely and supplement the wisdom being gained about the Higgs physics by the LHC community.

For the first project of this thesis, a search for the non resonant Higgs pair production via vector boson fusion mode (VBFHH) has been studied through the final state of a pair of bottom quark-antiquark pair and a pair of photons. The inclusion of VBFHH production along with the gluon gluon fusion Higgs pair production (ggHH) has been performed for the first time in the CMS. The analysis technique is optimized to be sensitive of beyond standard model physics also. Dedicated multivariate analysis (MVA) methods have been used to make the study produce the best result of that time. The 95% CL upper limits on the production cross section and branching ratio of inclusive ggHH and VBFHH processes have been measured along with the current allowed range of the coupling modifiers of trilinear Higgs coupling and a pair of Higgs coupling with a pair of vector bosons.

Due to the very small cross section of the non-resonant Higgs boson pair production, current volume of the LHC data is not sensitive enough to establish the evidence of the HH production. Hence a dedicated projection study of non-resonant HH search is performed in the same final state at the high luminosity LHC (HL-LHC) condition corresponding to $\sqrt{s}$ =14 TeV and an accumulated integrated luminosity of 3000 fb^{-1}. The Phase-2 upgraded CMS detector conditions are utilized here for the physics object reconstruction at the high pileup scenario (PU ~

200). The signal significance of the inclusive HH process is found to be 2.16 standard deviation which ensures that the inclusion of other HH decay channels and the combination of the data from the CMS and the ATLAS experiments will surely bring the 5 standard deviation discovery of the HH process at the HL-LHC. However, given the current pace of improvement in physics efficiency, diHiggs production process may be established earlier.

The second part of the thesis highlights the measurement of the Higgs boson production through vector boson fusion process and decaying to a bottom quark-antiquark pair. Dedicated high level trigger (HLT) path using the VBF event topology and b-tagging requirements along with the rigorous use of machine learning techniques help to reduce the overwhelming fully hadronic final state backgrounds dominated by QCD multijet process. The inclusive production of the Higgs boson, along with two extra jets, has been observed with 2.6 standard deviations, while the signal significance of only VBF Higgs production is 2.4 standard deviation. This study complements the exploration of $H \rightarrow b\bar{b}$ decay mode via other production processes.

Final part of this thesis is for the study of the missing transverse momentum p_T^{miss}, a very important physics object required in the measurement for the standard model physics containing neutrinos as well as to search for the new physics models. The response and the resolution of p_T^{miss}, reconstructed from particle flow and PUPPI algorithms are estimated in the events containing a photon and jets, where the activity of the jets is recoiled against precisely measured photons.

Bibliography

[1] **LEP Working Group for Higgs boson searches, ALEPH, DELPHI, L3, OPAL** Collaboration, R. Barate *et al.*, "Search for the standard model Higgs boson at LEP," *Phys. Lett. B* **565** (2003) 61–75, `arXiv:hep-ex/0306033`.

[2] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B* **716** (2012) 30–61, `arXiv:1207.7235 [hep-ex]`.

[3] **CMS** Collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery," *Nature* **607** no. 7917, (2022) 60–68, `arXiv:2207.00043 [hep-ex]`.

[4] J. Alison *et al.*, "Higgs boson potential at colliders: Status and perspectives," *Rev. Phys.* **5** (2020) 100045, `arXiv:1910.00012 [hep-ph]`.

[5] M. Gouzevitch and A. Carvalho, "A review of higgs boson pair production," *Reviews in Physics* **5** (2020) 100039. `https://www.sciencedirect.com/science/article/pii/S2405428320300022`.

[6] **CMS** Collaboration, L. Viliani, "CMS tracker performance and readiness for LHC Run II," *Nucl. Instrum. Meth. A* **824** (2016) 67–69.

[7] C. Collaboration, C. Amsler, V. Chiochia, S. Visscher, and R. Romaniuk, "Performance and operation of the cms electromagnetic calorimeter," *Journal of Instrumentation* **5** (03, 2010) T03010.

[8] **CMS** Collaboration, V. Khachatryan *et al.*, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV," *JINST* **10** no. 06, (2015) P06005, `arXiv:1502.02701 [physics.ins-det]`.

[9] **CMS** Collaboration, V. Khachatryan *et al.*, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV," *JINST* **10** no. 08, (2015) P08010, `arXiv:1502.02702 [physics.ins-det]`.

[10] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s}$ = 13 TeV," *JINST* **15** no. 05, (2020) P05002, `arXiv:1910.00079 [physics.ins-det]`.

[11] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV," *JINST* **13** no. 06, (2018) P06015, `arXiv:1804.04528 [physics.ins-det]`.

[12] C. Collaboration, "Performance of the CMS drift tube chambers with cosmic rays," *Journal of Instrumentation* **5** no. 03, (Mar, 2010) T03015–T03015. `https://doi.org/10.1088%2F1748-0221%2F5%2F03%2Ft03015`.

[13] **CMScollaboration** Collaboration, A. Tapper and D. Acosta, "CMS Technical Design Report for the Level-1 Trigger Upgrade," tech. rep., 2013. `https://cds.cern.ch/record/1556311`. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

[14] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC," *JINST* **16** no. 05, (2021) P05014, `arXiv:2012.06888 [hep-ex]`.

[15] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP* **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[16] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Particle-flow reconstruction and global event description with the CMS detector," *JINST* **12** no. 10, (2017) P10003, `arXiv:1706.04965 [physics.ins-det]`.

[17] **CMS** Collaboration, "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018,". `https://cds.cern.ch/record/2715872`.

[18] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, "Jet Flavour Classification Using DeepJet," *JINST* **15** no. 12, (2020) P12012, `arXiv:2008.10519 [hep-ex]`.

[19] **CMS** Collaboration, A. M. Sirunyan *et al.*, "A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution," *Comput. Softw. Big Sci.* **4** no. 1, (2020) 10, `arXiv:1912.06046 [physics.data-an]`.

[20] **CMS** Collaboration, "Performance of quark/gluon discrimination in 13 TeV data,". `https://cds.cern.ch/record/2234117`.

[21] **CMS** Collaboration, "Pileup mitigation at CMS in 13 TeV data," *JINST* **15** no. 09, (2020) P09018, `arXiv:2003.00503 [hep-ex]`.

[22] **CMS** Collaboration, "MET performance in 8 TeV data," tech. rep., CERN, Geneva, 2013. `https://cds.cern.ch/record/1543527`.

[23] **Particle Data Group** Collaboration, "Review of Particle Physics," *Progress of Theoretical and Experimental Physics* **2022** no. 8, (2022) .

[24] **LHC Higgs Cross Section Working Group** Collaboration, J. R. Andersen *et al.*, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties," `arXiv:1307.1347 [hep-ph]`.

[25] **CMS** Collaboration, "Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Journal of High Energy Physics* **2021** no. 3, (Mar, 2021) . `http://dx.doi.org/10.1007/JHEP03(2021)257`.

[26] D. de Florian *et al.*, "Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector," CERN Report CERN-2017-002-M, 2016. `arXiv:1610.07922 [hep-ph]`.

[27] **CMS** Collaboration, "Simplified template cross section measurements of Higgs boson produced in association with vector bosons in the H→b$\bar{\text{b}}$ decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV," tech. rep., CERN, Geneva, 2022. `http://cds.cern.ch/record/2827421`.

[28] F. Englert and R. Brout, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.* **13** (1964) 321.

[29] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.* **13** (1964) 508.

[30] T. Nakano and K. Nishijima, "Charge Independence for V-particles*," *Progress of Theoretical Physics* **10** no. 5, (11, 1953) 581–582, `https://academic.oup.com/ptp/article-pdf/10/5/581/5364926/10-5-581.pdf`. `https://doi.org/10.1143/PTP.10.581`.

[31] K. Nishijima, "Charge Independence Theory of V Particles*," *Progress of Theoretical Physics* **13** no. 3, (03, 1955) 285–304, `https://academic.oup.com/ptp/article-pdf/13/3/285/5425869/13-3-285.pdf`. `https://doi.org/10.1143/PTP.13.285`.

[32] M. Gell-Mann, "The interpretation of the new particles as displaced charge multiplets," *Nuovo Cim.* **4** no. S2, (1956) 848–866.

[33] S. Forte, L. Garrido, J. I. Latorre, and A. Piccione, "Neural network parametrization of deep inelastic structure functions," *JHEP* **05** (2002) 062, `arXiv:hep-ph/0204232`.

[34] G. Ridolfi, "Search for the Higgs boson: theoretical perspectives," *Frascati Phys. Ser.* **22** (2001) 291–304. `https://cds.cern.ch/record/506136`.

[35] **CDF, D0** Collaboration, "Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with up to 2.4 fb^{-1} of data," `arXiv:0804.3423 [hep-ex]`.

[36] **CDF, D0** Collaboration, T. T. E. V. N. P. H. W. Group, "Combined CDF and D0 Upper Limits on Standard Model Higgs-Boson Production with 2.1 - 5.4 fb**-1 of Data," in *20th Hadron Collider Physics Symposium 2009 (HCP 2009)*. 11, 2009. `arXiv:0911.3930 [hep-ex]`.

[37] **D0 Collaboration** Collaboration, V. M. Abazov and et al., "Search for the standard model higgs boson in $\ell\nu$+jets final states in 9.7 fb^{-1} of $p\bar{p}$ collisions with the d0 detector," *Phys. Rev. D* **88** (Sep, 2013) 052008. `https://link.aps.org/doi/10.1103/PhysRevD.88.052008`.

[38] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV," *Eur. Phys. J. C* **81** no. 4, (2021) 378, `arXiv:2011.03652 [hep-ex]`.

[39] **CMS** Collaboration, A. Tumasyan *et al.*, "Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production," *Nature Phys.* **18** no. 11, (2022) 1329–1334, arXiv:2202.06923 [hep-ex].

[40] **ATLAS** Collaboration, G. 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** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[41] **CMS** Collaboration, V. Khachatryan *et al.*, "Observation of the Diphoton Decay of the Higgs Boson and Measurement of Its Properties," *Eur. Phys. J. C* **74** no. 10, (2014) 3076, arXiv:1407.0558 [hep-ex].

[42] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of the Properties of a Higgs Boson in the Four-Lepton Final State," *Phys. Rev. D* **89** no. 9, (2014) 092007, arXiv:1312.5353 [hep-ex].

[43] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of Higgs Boson Production and Properties in the WW Decay Channel with Leptonic Final States," *JHEP* **01** (2014) 096, arXiv:1312.1129 [hep-ex].

[44] **CMS** Collaboration, V. Khachatryan *et al.*, "Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs," *Phys. Lett. B* **736** (2014) 64–85, arXiv:1405.3455 [hep-ex].

[45] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector," *Phys. Lett. B* **779** (2018) 283–316, arXiv:1708.00373 [hep-ex].

[46] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Observation of $t\bar{t}H$ production," *Phys. Rev. Lett.* **120** no. 23, (2018) 231801, arXiv:1804.02610 [hep-ex].

[47] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Observation of Higgs boson decay to bottom quarks," *Phys. Rev. Lett.* **121** no. 12, (2018) 121801, arXiv:1808.08242 [hep-ex].

[48] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Evidence for Higgs boson decay to a pair of muons," *JHEP* **01** (2021) 148, arXiv:2009.04363 [hep-ex].

[49] M. Grazzini, G. Heinrich, S. Jones, S. Kallweit, M. Kerner, J. M. Lindert, and J. Mazzitelli, "Higgs boson pair production at NNLO with top quark mass effects," *JHEP* **05** (2018) 059, `arXiv:1803.02463` `[hep-ph]`.

[50] S. Amoroso *et al.*, "Les Houches 2019: Physics at TeV Colliders: Standard Model Working Group Report," in *11th Les Houches Workshop on Physics at TeV Colliders: PhysTeV Les Houches*. 3, 2020. `arXiv:2003.01700` `[hep-ph]`.

[51] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca, and M. Spira, "$gg \rightarrow HH$: Combined uncertainties," *Phys. Rev. D* **103** no. 5, (2021) 056002, `arXiv:2008.11626` `[hep-ph]`.

[52] L.-S. Ling, R.-Y. Zhang, W.-G. Ma, L. Guo, W.-H. Li, and X.-Z. Li, "NNLO QCD corrections to Higgs pair production via vector boson fusion at hadron colliders," *Phys. Rev. D* **89** no. 7, (2014) 073001, `arXiv:1401.7754` `[hep-ph]`.

[53] F. A. Dreyer and A. Karlberg, "Fully differential Vector-Boson Fusion Higgs Pair Production at Next-to-Next-to-Leading Order," *Phys. Rev. D* **99** no. 7, (2019) 074028, `arXiv:1811.07918` `[hep-ph]`.

[54] J. Baglio, A. Djouadi, R. Gröber, M. M. Mühlleitner, J. Quevillon, and M. Spira, "The measurement of the Higgs self-coupling at the LHC: theoretical status," *JHEP* **04** (2013) 151, `arXiv:1212.5581` `[hep-ph]`.

[55] **ATLAS** Collaboration, "Search for Higgs boson pair production in association with a vector boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," `arXiv:2210.05415` `[hep-ex]`.

[56] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli, E. Vryonidou, and M. Zaro, "Higgs pair production at the LHC with NLO and parton-shower effects," *Phys. Lett. B* **732** (2014) 142–149, `arXiv:1401.7340` `[hep-ph]`.

[57] L. Li, Y.-Y. Li, and T. Liu, "Anatomy of *tthh* physics at the HL-LHC," *Phys. Rev. D* **101** no. 5, (2020) 055043, `arXiv:1905.03772` `[hep-ph]`.

[58] ATLAS Collaboration, "Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state using pp collision data at $\sqrt{s} = 8$ TeV from the ATLAS detector," *Phys. Rev. Lett.* **114** (2015) 081802, `arXiv:1406.5053` `[hep-ex]`.

[59] ATLAS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector," *Eur. Phys. J. C* **75** (2015) 412, arXiv:1506.00285 [hep-ex].

[60] ATLAS Collaboration, "Searches for Higgs boson pair production in the $HH \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector," *Phys. Rev. D* **92** (2015) 092004, arXiv:1509.04670 [hep-ex].

[61] ATLAS Collaboration, "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector," *Phys. Rev. D* **94** (2016) 052002, arXiv:1606.04782 [hep-ex].

[62] ATLAS Collaboration, "Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector," *Phys. Lett. B* **800** (2020) 135103, arXiv:1906.02025 [hep-ex].

[63] CMS Collaboration, "Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV," *Phys. Rev. D* **94** (2016) 052012, arXiv:1603.06896 [hep-ex].

[64] ATLAS Collaboration, "Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment," *JHEP* **11** (2018) 040, arXiv:1807.04873 [hep-ex].

[65] CMS Collaboration, "Search for Higgs boson pair production in the $b\bar{b}\tau\tau$ final state in proton-proton collisions at $\sqrt{s}$ = 8 TeV," *Phys. Rev. D* **96** (2017) 072004, arXiv:1707.00350 [hep-ex].

[66] CMS Collaboration, "Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s}$ = 13 TeV," *Phys. Lett. B* **778** (2018) 101, arXiv:1707.02909 [hep-ex].

[67] CMS Collaboration, "Search for resonant and nonresonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV," *JHEP* **01** (2018) 054, arXiv:1708.04188 [hep-ex].

[68] CMS Collaboration, "Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state in pp collisions at $\sqrt{s}$ = 13 TeV," *Phys. Lett. B* **788** (2019) 7, arXiv:1806.00408

[hep-ex].

[69] **CMS** Collaboration, "Search for Higgs boson pair production via vector boson fusion with highly Lorentz-boosted Higgs bosons in the four b quark final state at $\sqrt{s} = 13$ TeV," tech. rep., CERN, Geneva, 2021. `http://cds.cern.ch/record/2776802`.

[70] **CMS** Collaboration, "Search for Higgs boson pairs decaying to WWWW, WW$\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV," `arXiv:2206.10268 [hep-ex]`.

[71] **CMS** Collaboration, "Search for nonresonant Higgs boson pair production in the WW$\gamma\gamma$ channel in pp collisions at $\sqrt{s} = 13$ TeV," tech. rep., CERN, Geneva, 2022. `http://cds.cern.ch/record/2840773`.

[72] **ATLAS** Collaboration, "Search for nonresonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," `arXiv:2301.03212 [hep-ex]`.

[73] **ATLAS** Collaboration, "Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector," `arXiv:2209.10910 [hep-ex]`.

[74] **ATLAS** Collaboration, G. Aad *et al.*, "Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV," *Phys. Lett. B* **801** (2020) 135145, `arXiv:1908.06765 [hep-ex]`.

[75] **ATLAS** Collaboration, G. Aad *et al.*, "Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Rev. D* **106** no. 5, (2022) 052001, `arXiv:2112.11876 [hep-ex]`.

[76] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP* **03** (2021) 257, `arXiv:2011.12373 [hep-ex]`.

[77] **CMS** Collaboration, "Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV," `arXiv:2206.09401 [hep-ex]`.

[78] **CMS** Collaboration, "Search for nonresonant Higgs boson pair production in the four leptons plus two b jets final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV," `arXiv:2206.10657 [hep-ex]`.

[79] **CMS** Collaboration, A. Tumasyan *et al.*, "Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at s=13 TeV," *Phys. Rev. Lett.* **129** no. 8, (2022) 081802, `arXiv:2202.09617 [hep-ex]`.

[80] **ATLAS** Collaboration, "Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using *pp* collisions at $\sqrt{s}$ = 13 TeV," `arXiv:2211.01216 [hep-ex]`.

[81] "LHC Machine," *JINST* **3** (2008) S08001.

[82] **ATLAS** Collaboration, G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST* **3** (2008) S08003.

[83] **ALICE** Collaboration, K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC," *JINST* **3** (2008) S08002.

[84] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS Experiment at the CERN LHC," *JINST* **3** (2008) S08004.

[85] **LHCb** Collaboration, A. A. Alves, Jr. *et al.*, "The LHCb Detector at the LHC," *JINST* **3** (2008) S08005.

[86] K. Hubner, "Design and construction of the ISR," in *40th Anniversary of the First Proton-Proton Collisions in the CERN Intersecting Storage Rings (ISR)*. 6, 2012. `arXiv:1206.3948 [physics.acc-ph]`.

[87] **CMS** Collaboration, V. Karimki, M. Mannelli, P. Siegrist, H. Breuker, A. Caner, R. Castaldi, K. Freudenreich, G. Hall, R. Horisberger, M. Huhtinen, and A. Cattai, *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. `https://cds.cern.ch/record/368412`.

[88] **CMS Tracker Group** Collaboration, W. Adam *et al.*, "The CMS Phase-1 Pixel Detector Upgrade," *JINST* **16** no. 02, (2021) P02027, `arXiv:2012.14304 [physics.ins-det]`.

[89] Dominguez *et al.*, "CMS Technical Design Report for the Pixel Detector Upgrade," tech. rep., 2012. `https://cds.cern.ch/record/1481838`. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

[90] **CMS** Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report.* Technical design report. CMS. CERN, Geneva, 1997. `https://cds.cern.ch/record/349375`.

[91] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev. D* **48** (1993) 3160–3166, `arXiv:hep-ph/9305266`.

[92] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms," *JHEP* **08** (1997) 001, `arXiv:hep-ph/9707323`.

[93] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *JINST* **13** no. 05, (2018) P05011, `arXiv:1712.07158 [physics.ins-det]`.

[94] **CMS** Collaboration, "B-tagging performance of the CMS Legacy dataset 2018.,". `https://cds.cern.ch/record/2759970`.

[95] 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** (2014) 079, `arXiv:1405.0301 [hep-ph]`.

[96] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP* **02** (2009) 007, `arXiv:0811.4622 [hep-ph]`.

[97] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.* **191** (2015) 159, `arXiv:1410.3012 [hep-ph]`.

[98] 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** (2012) 088, `arXiv:1111.2854 [hep-ph]`.

[99] G. Heinrich, S. P. Jones, M. Kerner, G. Luisoni, and E. Vryonidou, "NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers," *JHEP* **08** (2017) 088, `arXiv:1703.09252` `[hep-ph]`.

[100] G. Heinrich, S. P. Jones, M. Kerner, G. Luisoni, and L. Scyboz, "Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects," *JHEP* **06** (2019) 066, `arXiv:1903.08137` `[hep-ph]`.

[101] S. P. Jones and S. Kuttimalai, "Parton shower and NLO-matching uncertainties in Higgs boson pair production," *JHEP* **02** (2018) 176, `arXiv:1711.03319` `[hep-ph]`.

[102] G. Heinrich, S. P. Jones, M. Kerner, and L. Scyboz, "A non-linear EFT description of $gg \to HH$ at NLO interfaced to POWHEG," *JHEP* **10** (2020) 021, `arXiv:2006.16877` `[hep-ph]`.

[103] G. Buchalla, M. Capozi, A. Celis, G. Heinrich, and L. Scyboz, "Higgs boson pair production in non-linear Effective Field Theory with full m_t-dependence at NLO QCD," *JHEP* **09** (2018) 057, `arXiv:1806.05162` `[hep-ph]`.

[104] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements," *Eur. Phys. J. C* **76** (2016) 155, `arXiv:1512.00815` `[hep-ex]`.

[105] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements," *Eur. Phys. J. C* **80** (2020) 4, `arXiv:1903.12179` `[hep-ex]`.

[106] **NNPDF** Collaboration, R. D. Ball *et al.*, "Parton distributions for the LHC Run II," *JHEP* **04** (2015) 040, `arXiv:1410.8849` `[hep-ph]`.

[107] **NNPDF** Collaboration, R. D. Ball *et al.*, "Parton distributions from high-precision collider data," *Eur. Phys. J. C* **77** (2017) 663, `arXiv:1706.00428` `[hep-ph]`.

[108] S. Carrazza, J. I. Latorre, J. Rojo, and G. Watt, "A compression algorithm for the combination of PDF sets," *Eur. Phys. J. C* **75** (2015) 474, `arXiv:1504.06469` `[hep-ph]`.

[109] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II," *J. Phys. G* **43** (2016) 023001, `arXiv:1510.03865` `[hep-ph]`.

[110] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, "New parton distribution functions from a global analysis of quantum chromodynamics," *Phys. Rev. D* **93** (2016) 033006, `arXiv:1506.07443 [hep-ph]`.

[111] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, "Parton distributions in the LHC era: MMHT 2014 PDFs," *Eur. Phys. J. C* **75** (2015) 204, `arXiv:1412.3989 [hep-ph]`.

[112] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4—a simulation toolkit," *Nucl. Instrum. Meth. A* **506** (2003) 250.

[113] G. Breit and E. Wigner, "Capture of slow neutrons," *Phys. Rev.* **49** (Apr, 1936) 519–531. `https://link.aps.org/doi/10.1103/PhysRev.49.519`.

[114] M. J. Oreglia, *A study of the reactions $\psi' \to \gamma\gamma\psi$*. PhD thesis, Stanford University, 1980. `http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-236.pdf`. SLAC Report SLAC-R-236.

[115] N. Kumar and S. P. Martin, "LHC search for di-Higgs decays of stoponium and other scalars in events with two photons and two bottom jets," *Phys. Rev. D* **90** (2014) 055007, `arXiv:1404.0996 [hep-ph]`.

[116] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, "Handling uncertainties in background shapes: the discrete profiling method," *JINST* **10** no. 04, (2015) P04015, `arXiv:1408.6865 [physics.data-an]`.

[117] B. Cabouat and T. Sjöstrand, "Some Dipole Shower Studies," *Eur. Phys. J. C* **78** no. 3, (2018) 226, `arXiv:1710.00391 [hep-ph]`.

[118] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS," *Eur. Phys. J. C* **81** (2021) 800, `arXiv:2104.01927 [hep-ex]`.

[119] **CMS** Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV," tech. rep., CERN, Geneva, 2018. `https://cds.cern.ch/record/2621960`.

[120] **CMS** Collaboration, CMS Collaboration, "CMS luminosity measurement for the 2018
      data-taking period at $\sqrt{s} = 13\,\text{TeV}$," CMS Physics Analysis Summary
      CMS-PAS-LUM-18-002, 2019. `https://cds.cern.ch/record/2676164/`.

[121] "Snowmass 2021." Https://snowmass21.org/.

[122] **CMS** Collaboration, "Prospects for HH measurements at the HL-LHC," Tech. Rep.
      CMS-PAS-FTR-18-019, CERN, Geneva, 2018.
      `http://cds.cern.ch/record/2652549`.

[123] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco,
      V. Lemaître, A. Mertens, and M. Selvaggi, "DELPHES 3, A modular framework for
      fast simulation of a generic collider experiment," *JHEP* **02** (2014) 057,
      `arXiv:1307.6346 [hep-ex]`.

[124] **CMS** Collaboration, "Technical Proposal for the Phase-II Upgrade of the CMS
      Detector," CMS Technical Proposal CERN-LHCC-2015-010. LHCC-P-008.
      CMS-TDR-15-02, 2015. `https://cds.cern.ch/record/2020886`.

[125] **CMS** Collaboration, C. Collaboration, "The Phase-2 Upgrade of the CMS Tracker,"
      CMS Technical Design Report CERN-LHCC-2017-009. CMS-TDR-014, 2017.
      `https://cds.cern.ch/record/2272264`.

[126] **CMS** Collaboration, C. Collaboration, "The Phase-2 Upgrade of the CMS Barrel
      Calorimeters Technical Design Report," CMS Technical Design Report
      CERN-LHCC-2017-011. CMS-TDR-015, 2017.
      `https://cds.cern.ch/record/2283187`.

[127] **CMS** Collaboration, C. Collaboration, "The Phase-2 Upgrade of the CMS Endcap
      Calorimeter," CMS Technical Design Report CERN-LHCC-2017-023. CMS-TDR-019,
      2017. `https://cds.cern.ch/record/2293646`.

[128] **CMS** Collaboration, C. Collaboration, "The Phase-2 Upgrade of the CMS Muon
      Detectors," CMS Technical Design Report CERN-LHCC-2017-012. CMS-TDR-016,
      2017. `https://cds.cern.ch/record/2283189`.

[129] **CMS** Collaboration, "A MIP Timing Detector for the CMS Phase-2 Upgrade," Tech.
      Rep. CERN-LHCC-2019-003, CMS-TDR-020, CERN, Geneva, Mar, 2019.

`https://cds.cern.ch/record/2667167`.

[130] **CMS** Collaboration, "The Phase-2 Upgrade of the CMS Level-1 Trigger," Tech. Rep. CERN-LHCC-2020-004, CMS-TDR-021, CERN, Geneva, Apr, 2020. `https://cds.cern.ch/record/2714892`. Final version.

[131] **CMS** Collaboration, "The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger," Tech. Rep. CERN-LHCC-2021-007, CMS-TDR-022, CERN, Geneva, Mar, 2021. `https://cds.cern.ch/record/2759072`.

[132] **CMS** Collaboration, CMS Collaboration, "Expected performance of the physics objects with the upgraded CMS detector at the HL-LHC," Tech. Rep. CMS-NOTE-2018-006. CERN-CMS-NOTE-2018-006, 2018. `https://cds.cern.ch/record/2650976`.

[133] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi, "Delphes 3: a modular framework for fast simulation of a generic collider experiment," *Journal of High Energy Physics* **2014** no. 2, (Feb, 2014) . `http://dx.doi.org/10.1007/JHEP02(2014)057`.

[134] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. K. J. au2, M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla, "Tmva - toolkit for multivariate data analysis," 2009.

[135] X. Buffat, I. Efthymiopoulos, L. D. Medina Miranda, R. Tomas Garcia, J. Wenninger, S. M. Demers Konezny, K. F. Einsweiler, R. Hawkings, B. Giacobbe, C. Bohm, B. Petersen, A. Polini, A. Sfyrla, S. Pagan Griso, T. Strebler, D. Bloch, D. N. Taylor, A. Dabrowski, F. Hartmann, A. Mohammadi, S. Mukherjee, M. Narain, G. Ortona, G. Pasztor, A. Savin, S. Sekmen, D. Stickland, L. Dufour, and M. R. J. Williams, "HL-LHC Experiment Data Quality Working Group Summary Report," tech. rep., CERN, Geneva, 2022. `https://cds.cern.ch/record/2802720`.

[136] 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** (1998) 56, `arXiv:hep-ph/9704448`.

[137] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the nature of the Higgs sector," *CERN Yellow Rep. Monogr.* **2** (2017) , `arXiv:1610.07922 [hep-ph]`.

[138] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$," *JHEP* **12** (2020) 085, `arXiv:2006.13251 [hep-ex]`.

[139] **ATLAS** Collaboration, G. Aad *et al.*, "Measurements of WH and ZH production in the H $\rightarrow$ b$\bar{\text{b}}$ decay channel in pp collisions at 13 TeV with the ATLAS detector," *Eur. Phys. J. C* **81** (2021) 178, `arXiv:2007.02873 [hep-ex]`.

[140] F. Maltoni, K. Mawatari, and M. Zaro, "Higgs characterisation via vector-boson fusion and associated production: NLO and parton-shower effects," *Eur. Phys. J. C* **74** (2014) 2710, `arXiv:1311.1829 [hep-ph]`.

[141] **CMS** Collaboration, V. Khachatryan *et al.*, "Search for the standard model Higgs boson produced through vector boson fusion and decaying to b$\bar{\text{b}}$," *Phys. Rev. D* **92** (2015) 032008, `arXiv:1506.01010 [hep-ex]`.

[142] **ATLAS** Collaboration, G. Aad *et al.*, "Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13\,\text{TeV}$," *Eur. Phys. J. C* **81** (2021) 537, `arXiv:2011.08280 [hep-ex]`.

[143] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG," *JHEP* **02** (2010) 037, `arXiv:0911.5299 [hep-ph]`.

[144] B. Cabouat and T. Sjöstrand, "Some dipole shower studies," *Eur. Phys. J. C* **78** (2018) 226, `arXiv:1710.00391 [hep-ph]`.

[145] J. Bellm *et al.*, "HERWIG 7.0/HERWIG++3.0 release note," *Eur. Phys. J. C* **76** (2016) 196, `arXiv:1512.01178 [hep-ph]`.

[146] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, "HW$^{\pm}$/HZ+0 and 1 jet at NLO with the POWHEG box interfaced to GoSam and their merging within MiNLO," *JHEP* **10** (2013) 083, `arXiv:1306.2542 [hep-ph]`.

[147] K. Hamilton, P. Nason, C. Oleari, and G. Zanderighi, "Merging H/W/Z+0 and 1 jet at NLO with no merging scale: A path to parton shower + NNLO matching," *JHEP* **05** (2013) 082, arXiv:1212.4504 [hep-ph].

[148] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions," *JHEP* **01** (2007) 013, arXiv:hep-ph/0611129.

[149] J. M. Lindert *et al.*, "Precise predictions for V+jets dark matter backgrounds," *Eur. Phys. J. C* **77** (2017) 829, arXiv:1705.04664 [hep-ph].

[150] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Pileup mitigation at CMS in 13 TeV data," *JINST* **15** (2020) P09018, arXiv:2003.00503 [hep-ex].

[151] **CMS** Collaboration, "Performance of quark/gluon discrimination in 8 TeV pp data," CMS Physics Analysis Summary CMS-PAS-JME-13-002, 2013. https://cds.cern.ch/record/1599732.

[152] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Jet algorithms performance in 13 TeV data," CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, 2017. https://cds.cern.ch/record/2256875.

[153] ATLAS and CMS Collaborations, and LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011," Tech. Rep. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, 2011. https://cds.cern.ch/record/1379837.

[154] G. Cowan, *Statistical data analysis*. Oxford University Press, USA, 1998.

[155] **CMS** Collaboration, "Mitigation of anomalous missing transverse momentum measurements in data collected by CMS at sqrt(s) = 13 TeV during the LHC Run 2,". https://cds.cern.ch/record/2714938.

[156] **TheATLAS, TheCMS, TheLHCHiggsCombinationGroup** Collaboration, "Procedure for the LHC Higgs boson search combination in Summer 2011," tech. rep., CERN, Geneva, 2011. http://cds.cern.ch/record/1379837.

[157] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J. C* **71** (2011) 1554, `arXiv:1007.1727 [physics.data-an]`. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[158] E. Gross, "Practical Statistics for High Energy Physics," *CERN Yellow Rep. School Proc.* **3** (2018) 199–221.

[159] A. L. Read, "Presentation of search results: the cls technique," *Journal of Physics G: Nuclear and Particle Physics* **28** no. 10, (Sep, 2002) 2693. `https://dx.doi.org/10.1088/0954-3899/28/10/313`.

Appendices

Appendix A

Event displays

VBF HH → b$\bar{\text{b}}$γγ
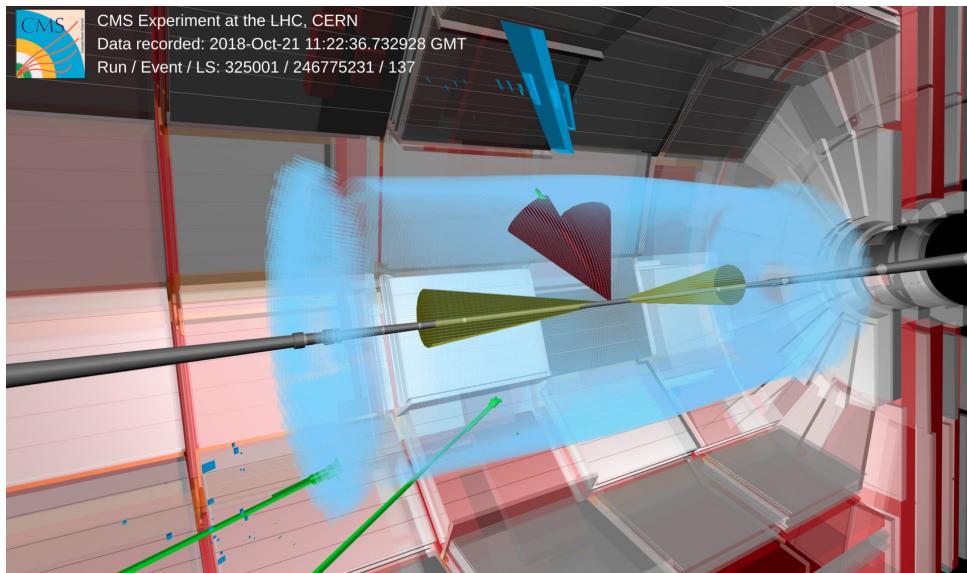
Figure A.1: An event display of a real collision data with a potential VBF Higgs pair production event and subsequent decay to a pair of bottom quarks and a pair of photons, where two forward jets (in deep green colored cone) are back to back along the beam direction and two b-tagged jets (in violate colored cone) and two photons (in green lines) from the decay of the Higgs bosons are in the central part of the detector. The event is recorded in 2018 Run 2 data taking period by rhe CMS detector.

VBF H → b$\bar{\text{b}}$
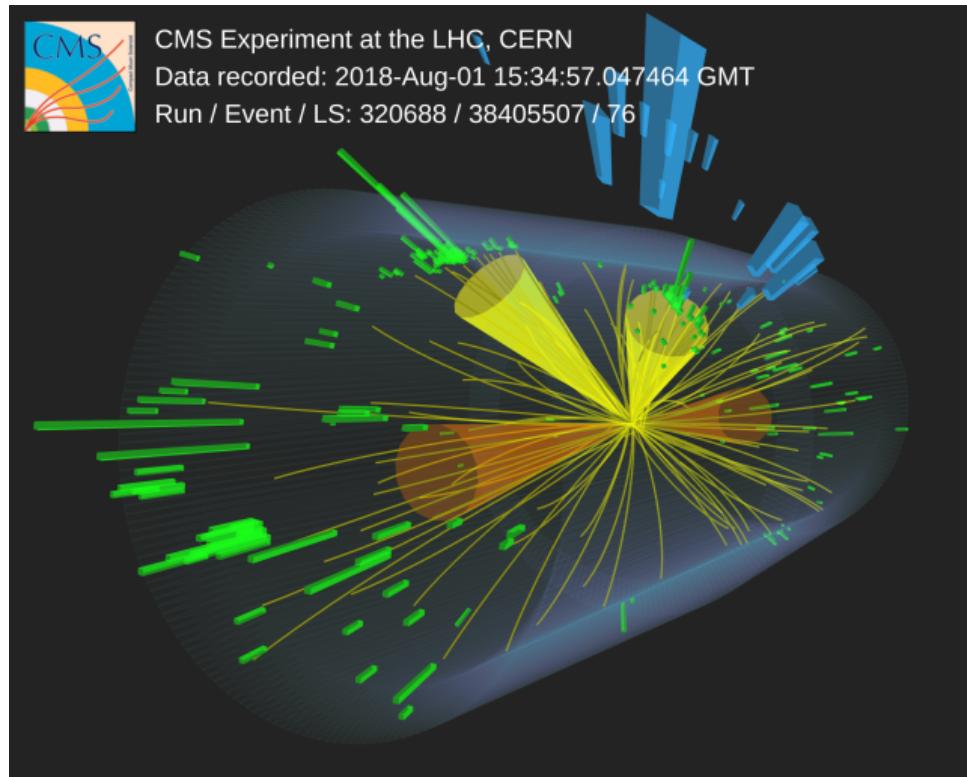

Figure A.2: An event display of a real collision data with VBF Higgs production and subsequent decay to a pair of bottom quarks, where two forward-backward jets (in orange colored cone) are back to back along the beam direction and two b-tagged jets (in yellow) from the Higgs decay are at the central part of the detector. The event is collect in 2018 Run 2 data taking period.

Appendix B

Statistical interpretation

B.1   Introduction

The main motivation of the statistical interpretation of Higgs analysis is to establish the presence of a signal or to exclude the assumption of that signal in the observed data. For this purpose, one generally follows the "frequentist" approach as a common statistical framework, which has been widely used in the ATLAS and CMS experiments for the Higgs boson discovery [156] at the LHC.

Let us consider a process, like, Higgs boson production of a particular mass M_{H} with the standard model predicted cross section of $\sigma_{SM}(M_{\mathrm{H}})$. The expected signal event yield ($s(M_{\mathrm{H}})$) can be estimated according to the SM, using the following formula corresponding to an integrated luminosity of $\mathcal{L}$,

$$s(M_{\mathrm{H}}) = \mathcal{L}.\sigma_{\mathrm{SM}}(M_{\mathrm{H}}).\epsilon.A \tag{B.1}$$

where, ϵ and A are the efficiency and the acceptance of the detector. If b is the expected background predicted from the SM then the total number of observed

events in data can be interpreted as:

$$n = \mu s(M_{\mathrm{H}}) + b \qquad\qquad\qquad (B.2)$$

where μ is the signal strength, defined as : $\mu = \sigma^{\mathrm{obs}}/\sigma_{\mathrm{SM}}$

Now we can assume two different hypothesis : (i) Null hypothesis (H_0) : background only hypothesis where there is no presence of signal, and (ii) alternative hypothesis (H_μ), signal-plus-background hypothesis, i.e for the procedure mentioned above a Higgs boson of mass M_{H} with a signal strength μ is present in data on top the background. To test the hypothesis we have to compute an function called "test statistic" (q_{obs}) [157,158] in the observed data. Finally we need to judge, based on the q_{obs}, whether we are able to reject the null hypothesis in favor of the alternative hypothesis or fail to reject the null hypothesis. The "test statistic" can be defined in many ways, one of the common approach is in terms of Likelihood ratios.

In an observation of n data events, the likelihood function, $L(n|\mu, \theta)$ is defined as,

$$L(n|\mu, \theta) = P\left(n|\mu.s(\theta) + b(\theta)\right).p\left(\theta|\bar{\theta}\right) \qquad\qquad (B.3)$$

Here P is the probability density function; for the counting experiment it is a Poissonian probability function. The systematic uncertainties are included in the Likelihood function as nuisance parameters (NP) accounting for their effects in both signal ($s(\theta)$) and background $b(\theta)$. The quantity $p\left(\theta|\bar{\theta}\right)$ is the prior knowledge of the nuisance parameters θ, given there initial estimation $\bar{\theta}$. Typically, a binned likelihood fit is performed including all analysis region simultaneously to determine the parameter of interest. For each, analysis bin a Poissonian distribution with corresponding signal and background events are considered for a global value of the signal strength (μ). The likelihood in ag-

gregate is the product of individual bin likelihood and is given by,

$$L(n|\mu, \theta) = \prod_{i \epsilon \text{bins}} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} \exp\left[-\left(\mu s_i + b_i\right)\right].p\left(\theta|\bar{\theta}\right) \tag{B.4}$$

Now the LHC profile likelihood ratio as test statistics is defined as:

$$\hat{q}_\mu = -2\ln\frac{L\left(n, \bar{\theta}|\mu, \hat{\theta}_\mu\right)}{L\left(n, \bar{\theta}|\hat{\mu}, \hat{\theta}\right)} = -2\ln\lambda(\mu), \ 0 \le \hat{\mu} \le \mu \tag{B.5}$$

where $\hat{\mu}$ and $\hat{\theta}$ are the best fit value of μ and θ respectively, derived from the data, and $\hat{\theta}_\mu$ is the best fit value of θ for a particular value of μ.

B.2   Upper limit

For each value of μ, that we need to test, the test statistic $\tilde{q}_\mu^{obs}$ can be evaluated from the observed data. The p-values, p_b and p_μ, can be constructed from the test statistic associated with both the null and the alternative hypothesis, H_0 and H_μ.

$$1 - p_b = P\left(\tilde{q}_\mu \ge \tilde{q}_\mu^{obs}|H_0\right) = \int_{\tilde{q}_\mu^{obs}}^\infty f\left(\tilde{q}_\mu|0, \tilde{\theta}_0^{obs}\right) d\tilde{q}_\mu$$

$$p_\mu = P\left(\tilde{q}_\mu \ge \tilde{q}_\mu^{obs}|H_\mu\right) = \int_{\tilde{q}_\mu^{obs}}^\infty f\left(\tilde{q}_\mu|\mu, \tilde{\theta}_\mu\right) d\tilde{q}_\mu^{obs} \tag{B.6}$$

In the modified frequentist approach, for confidence level for claiming the presence of signal, CLs [159] is defined as

$$\text{CLs}(\mu) = \frac{p_\mu}{1 - p_b} \tag{B.7}$$

A signal is considered to be excluded at a confidence level (CL) of α, only if the calculated value of CLs(μ) is less than or equal to the value of 1 - α. Hence to exclude a signal with a signal strength μ at 95% CL, we need to set CLs (μ) ≤ 0.05.

B.3   Evidence and discovery of a signal

To quantify the excess of events in the observation, it is important to establish the presence of a signal with a certain confidence level above a background. "The discovery test statistics" for a positive signal can be expressed in terms of the background only hypothesis as,

$$\hat{q}_0 = -2\ln\frac{L\left(n, \bar{\theta}|0, \hat{\theta}_\mu\right)}{L\left(n, \bar{\theta}|\hat{\mu}, \hat{\theta}\right)} \qquad (B.8)$$

Larger values of $\hat{q}_0$ corresponds to the larger incompatibility of the background only hypothesis with the data, which hints that the presence of the signal-plus-background only hypothesis instead of only background. This incompatibility can be interpreted in terms of the p-value.

$$p = P(\hat{q}_0 \geq \hat{q}_0^{obs}|H_0) \qquad (B.9)$$

This p-value can be converted in terms of significance (Z) and can be expressed through a one sided Gaussian integral as:

$$p = \int_Z^\infty \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx \qquad (B.10)$$

therefore Z can be derived from the observed p-value using the above expression. By a convention in high energy physics, a significance of 3σ ($Z = 3$, $p = 1.3\times10^{-3}$) signifies the evidence of a signal process and further more a 5σ significance ($Z = 5$, $p = 2.8\times10^{-7}$) is accepted norm for the discovery of that process.