Università degli Studi di Padova

Dipartimento di Fisica e Astronomia

"Galileo Galilei"

Corso di Dottorato di Ricerca in Fisica
XXXIII ciclo

# Study of the associated production of the Higgs boson with a leptonically decaying W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV



Supervisore:
Prof. Ugo Gasparini

Co-supervisore:
Dr. Alberto Zucchetta

Dottorando:
Siewyan Hoh

*"Three metamorphoses of the spirit have I designated to you: how the spirit became a camel, the camel a lion, and the lion at last a child."*

- Friedrich Nietzsche, Thus spoke Zarathustra

# *Abstract*

This thesis presents the study of the Standard Model (SM) Higgs boson production in association with a W boson decaying into a charged lepton and neutrino, followed by a Higgs boson decaying into a pair of W bosons, which subsequently decay into a pair of charged leptons, neutrinos and a pair of light quarks. The final state consists of two same charged leptons, moderate missing transverse energy and one or two jets. The full proton-proton collision dataset collected by the CMS detector in the years 2016-2018 (referred as "Run-II" dataset) at the center of mass energy, $\sqrt{s} = 13$ TeV, corresponding to the integrated luminosity of 137 fb$^{-1}$, is used in this study.

The pair of isolated leptons, selected to be of the same charge, greatly reduce the backgrounds from other physics processes having much larger production cross sections. Events are then categorized into 1 and 2-jets categories, and further sub-categorized into two muons (same flavor) or a pair of electron and muon (different flavor) to maximize the signal sensitivity.

The dominant backgrounds to the signal regions are WZ, non-prompt processes originating from W+jets production and heavy flavor charm or beauty decay, and W$\gamma^{\star}$ with $\gamma$ conversion giving two leptons in which one of the final state lepton is undetected. The WZ background normalization is estimated from data control region, and the non-prompt processes are modeled with a data-driven method. Other minor backgrounds are modeled from simulation. The instrumental background such as the electron charge mis-assignment is estimated from data and taken into account for backgrounds containing same charge electron and muon pair. The non-prompt background in the signal regions is reduced by using a multivariate technique, exploiting the kinematical difference between the non-prompt and prompt leptons.

The associated WH production cross section times the $H \to W^{+}W^{-}$ branching fraction, is $\mu_{wh} = 0.89^{+0.64}_{-0.62}(stat)^{+0.74}_{-0.78}(exp)$ times the SM expectation, corresponding to an observed significance of 0.9 $\sigma$. The study of the measurement is dominated by statistical and systematic uncertainties arising from the background modeling and theoretical model.

Additionally, using the same dataset, an inclusive measurement of the Higgs boson production in association with a leptonically decaying Z or W boson (leptonic VH, V=Z,W) is performed by using sample of $H \to W^{+}W^{-}$ event candidates in which at least one W boson decays leptonically. The final states considered in the measurement are the associated WH production resulting in final states with two same sign leptons, and three leptons, and the associated ZH production yielding three and four leptons.

The measured signal strength modifiers are consistent with SM expectation in general, with large deviation observed in the associated ZH production with three lepton final state due to the background fluctuation. The combined result for the leptonic VH production signal strength is $\mu_{VH} = 1.85^{+0.47}_{-0.44}$ times their respective SM expectation, with the observed significance of 4.7 $\sigma$. The precision of the measurement is limited by statistics, and systematic effects from the non-prompt backgrounds.

The measurement is also performed in the exclusive phase space bins in the so-called Simplified Template Cross Section (STXS) framework. The signal strength modifiers are determined for the two STXS bins, defined in different regions in the associated vector boson transverse momentum. The STXS measurement for the leptonic VH production signal strength modifier is $\mu_{p_{\mathrm{T}}^{\mathrm{V}} < 150GeV} = 2.65^{+0.72}_{-0.66}$ and $\mu_{p_{\mathrm{T}}^{\mathrm{V}} > 150GeV} = 1.56^{+0.97}_{-0.88}$, with an observed significance of 4.7 $\sigma$ and 1.8 $\sigma$.

# Contents

# Acronyms

**APDs** Avalanche Photodiodes

**ATLAS** A Toroidal Large hadron collider ApparatuS

**BDT** Boosted Decision Tree

**BSM** Beyond Standard Model

**CERN** Conseil Européen pour la Recherche Nucléaire

**CHS** Charged Hadron Subtraction

**CMS** Compact Muon Solenoid

**CP** Charge conjugation and Parity symmetry

**CR** Control Region

**CSC** Cathode Strip Chambers

**CSV** Combined Secondary Vertex

**CTF** Combinatorial Track Finder

**DAQ** Data Acquisition System

**DT** Drift Tubes Chambers

**EB** Event Builder

**ECAL** Electromegnetic Calorimeter

**FSR** Final State Radiation

**FU** Filter Unit

**GSF** Gaussian-Sum Filter

**HCAL** Hadronic Calorimeter

**HL** High Luminosity

**HLT** High Level Trigger

**HPD** Hybrid Photo-Diodes

**ID** Identification

**ISO** Isolation

**ISR** Initial State Radiation

**JEC** Jet Energy Correction

**JER** Jet Energy Resolution

**JES** Jet Energy Scale

**KF** Kalman Filter

**L1** Level-1

**LCT** Local Charged Tracks

**LHC** Large Hadron Collider

**LHCb** Large Hadron Collider beauty

**LO** Leading Order

**MC** Monte Carlo

**ME** Matrix Elements

**MVA** Multivariate Analysis

**NLO** Next-To-Leading-Order

**NNLO** Next-to-Next-to-Leading-Order

**PDF** Parton Distribution Function

**PDs** Primary Datasets

**PF** Particle Flow

**pp** Proton-Proton

**PS** Parton Shower

**PUPPI** Pile-Up Per Particle Identification

**QCD** Quantum Chromodynamics Dynamics

**QED** Quantum Electrodynamics Dynamics

**RPC** Resistive Plate Chambers

**RU** Readout Unit

**SC** Superclusters

**SM** Standard Model

**SR** Signal Region

**STXS** Simplified Template Cross Section

**TP** Trigger Primitives

**UE** Underlying Events

**VEV** vacuum expectation value

**WLCG** Worldwide LHC Computing Grid

**WLS** WaveLength Shifting

**WP** Working Point

# Chapter 1

# Introduction

From the discovery of fire around 1.7 million years (to date) [1] to the discovery of the Higgs Boson on 4th of July 2012 [2, 3], these two significant events have shared a common theme about mankind's nature, curiosity. It is a living proof that curiosity has been the main driving force in all the scientific endeavor embarked by mankind to understand the natural world we find ourselves in. Our latest and most throughout understanding of matter and interactions in the physical world is summarized in a theoretical framework called the SM. The theory's prediction withstood all experimental attempts to nullify or supersede it so far. One of these predictions is the existence of the neutral charged scalar boson, called Higgs boson, as a consequence of introducing the electroweak symmetry breaking mechanism into SM to explain the origin of all particle's mass. Ever since the triumphant discovery of the new boson, the Higgs boson has been used as a probe to detect any potential deviation from the SM, also known as Beyond Standard Model (BSM), and a guide to perform many precision measurements on different decay channels to reinforce its discovery's significance or increase the discovery potential of other decay channels. The two research activities have been the recurring theme of the physics program of CMS and ATLAS experiments at Conseil Européen pour la Recherche Nucléaire (CERN).

For a SM Higgs boson with a mass of approximately 125 GeV, where the natural width is only a few MeV, five relevant decay channels play a significant role for measurement and detection at the Large Hadron Collider (LHC), as summarized in Fig. 1.1. In the $H \to \gamma\gamma$ and $H \to ZZ \to 4l$ channels, all final state particles can be very precisely measured and the resolution on the reconstructed Higgs boson mass is excellent (about 1-2 %). The fully leptonic decay channel $H \to W^+W^- \to l^+l^-\nu_l\bar{\nu}_l$ has a poor mass resolution (about 20%) due to the presence of neutrinos; however, the decay channel has the second largest branching ratio (about 22%) and the final state has relatively low level of background.

After the Higgs boson discovery, the partial statistics of Run-II collected in the year 2016 at $\sqrt{s} = 13$ TeV with an integrated luminosity of 35.9 $\text{fb}^{-1}$ was used to further constrain its properties. In particular, the combined measurement on all Higgs production channels has established the first observation of the Higgs boson decay to W boson pair $H \to W^+W^-$,

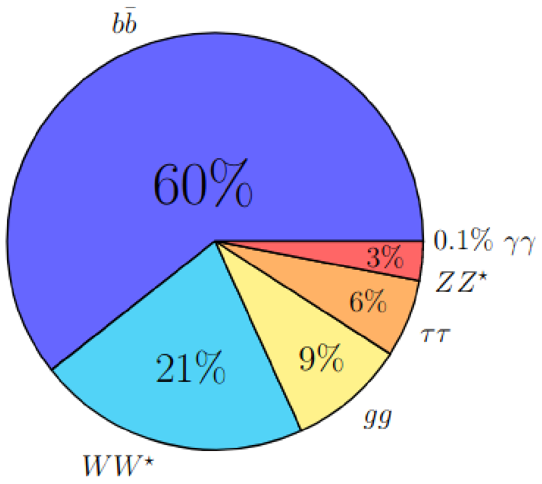



FIGURE 1.1: The breakdown of SM Higgs boson decay branching ratio [4]. The $\star$ on top of W and Z boson indicates off-shell, due to the Higgs mass threshold.

with an observed significance of 9.1 $\sigma$, to be compared with an expected significance of 7.1 $\sigma$ [5]. The main sensitivity of the measurement is driven by the dominant production mode, the gluon fusion production (ggF), and by the large branching fraction of W boson pair, which allow the study of production channels such as Higgs boson production via vector boson fusion (VBF) and associated production with vector boson (VH, V = W,Z boson).

The VH production channel, being the third largest production mode by cross section, has becoming more accessible particularly with the higher luminosity of the full Run-II dataset. An important final state to measure the VH production mechanism is the same charge or same sign[1] di-lepton final state, as it has small background contamination from dominant electroweak backgrounds. Specifically, the same sign di-lepton topology is sensitive to the Higgs boson associated production with a leptonically decaying W boson, followed by $H \to W^+W^-$ decay, where one of the W boson decays leptonically and the other hadronically. As a result, the final state consists of two same sign leptons where one of the leptons is originated from the associated W boson and the other from the Higgs decay, with a pair of hadronic jets and missing transverse energy due to the escaping neutrinos.

The thesis presents the study of the measurement on the Higgs boson associated production with a leptonically decaying W boson, followed by $H \to W^+W^-$ decay, using the same sign di-lepton final state for the first time. The result is combined with other sensitive final states according to the number and flavor of leptons and the number of jets, using event selection targeting the inclusive leptonic VH production in the $H \to W^+W^-$ decay channel. In addition, the combined measurement is also interpreted in STXS framework.

The analyses are performed on the full Run-II Proton-Proton (pp) collision dataset, corresponding to an integrated luminosity of 137 fb$^{-1}$ collected by the CMS detector at $\sqrt{s}$= 13 TeV.

[1]Throughout the thesis, same sign will be used to refer to same charge.

The thesis is structured as follow:

Chapter 2 provides an overview of the theoretical framework motivating the measurement.

Chapter 3 briefly outlines the technical aspects of the CMS detector.

Chapter 4 provides the information on the reconstructed physics objects and the event simulation common to the analyses.

Chapter 5 details the study of the associated WH production mode measurement using the same sign di-lepton final state.

Chapter 6 details the combined inclusive measurement, and the STXS measurement of the leptonic VH production mode.

Chapter 7 summarizes the thesis conclusion.

# Chapter 2

# Theoretical Overview

The SM is a theoretical asset of fundamental importance in high energy physics experiments, due to its extreme predictability on all the processes produced inside the particle collider. In this section, the theoretical framework of SM is presented, and the Higgs mechanism, which is responsible for generating the masses of the quarks and leptons, is discussed. The mechanism introduces a new field, called Higgs field, and the excited state of the Higgs field manifests as a new particle, the Higgs boson.

## 2.1 The Standard Model of particle physics

The SM is a well-established theoretical model formulated in quantum field theory describing all the elementary particles and fundamental interactions[1] in our Universe. The SM is based on the principle of symmetry, where different particles or fields of specific properties are described by different irreducible representations of a symmetry group. The fundamental interactions among the particles are uniquely determined by specifying their gauge symmetry group under a set of transformation laws. The local gauge symmetry group related to the strong, weak and the electromagnetic interactions in the SM is:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y \tag{2.1}$$

where the subscript $C$ is the color charge, $L$ indicates that the $SU(2)$ symmetry group is exclusive to the left-handed fields, and $Y$ is the hypercharge relating the electric charge and the weak isospin. The symmetry group $SU(3)_C$ represents the strong interaction described by Quantum Chromodynamics Dynamics (QCD), where 3 indicates the number of rows and columns in the unitary matrices forming the group, while the electromagnetic and the weak

[1] Gravitational interaction is not included in SM. Quantum gravity effects are not sensitive to the current accessible energy scale.

interactions are unified [6] as electroweak interaction under $SU(2)_L \times U(1)_Y$ symmetry group.

Elementary particles are divided in two major categories: the fermions, spin $\frac{1}{2}$ particle obeying the Pauli's exclusion principle, and the gauge bosons, spin 1 particles which can occupy and co-exist in the same quantum state with other bosons. This distinctive feature leads to a completely different role for these particles: the fermions constitute the matter; the bosons are the force carriers mediating the fundamental interactions in nature.

### 2.1.1 Elementary particles

There are 12 types of fermions: 6 types of quarks, having strong and electroweak interaction, and 6 types of leptons having only electroweak interacion and their corresponding anti-particles. Moreover, there are 13 type of bosons: the photon; the $W^+$, $W^-$ and Z bosons; 8 coloured gluons, and the Higgs boson. These elementary particles are summarized in Table 2.1.

TABLE 2.1: The elementary particles, fermions and bosons, described in the SM; the value for the masses are taken from Ref [7].

| | Name | Symbol | Spin ($\hbar$) | Mass ($MeV/c^2$) |
|---|---|---|---|---|
| Quarks | up | $u$ | $\frac{1}{2}$ | $2.16^{+0.49}_{-0.26}$ |
| | down | $d$ | $\frac{1}{2}$ | $4.67^{+0.48}_{-0.17}$ |
| | charm | $c$ | $\frac{1}{2}$ | $(1.27 \pm 0.02) \cdot 10^3$ |
| | strange | $s$ | $\frac{1}{2}$ | $93^{+11}_{-5}$ |
| | beauty | $b$ | $\frac{1}{2}$ | $(4.18^{+0.03}_{-0.02}) \cdot 10^3$ |
| | top | $t$ | $\frac{1}{2}$ | $(172.4 \pm 0.30) \cdot 10^3$ |
| Leptons | $e$ neutrino | $\nu_e$ | $\frac{1}{2}$ | $< 2 \cdot 10^{-6}$ |
| | electron | $e$ | $\frac{1}{2}$ | $0.5109989461 \pm 0.0000000031$ |
| | $\mu$ neutrino | $\nu_\mu$ | $\frac{1}{2}$ | $< 0.19$ |
| | muon | $\mu$ | $\frac{1}{2}$ | $105.6583745 \pm 0.0000024$ |
| | $\tau$ neutrino | $\nu_\tau$ | $\frac{1}{2}$ | $< 18.2$ |
| | tau | $\tau$ | $\frac{1}{2}$ | $1776.86 \pm 0.12$ |
| Gauge bosons | photon | $\gamma$ | 1 | 0 |
| | W boson | $W^\pm$ | 1 | $(80.379 \pm 0.012) \cdot 10^3$ |
| | Z boson | $Z^0$ | 1 | $(91.1876 \pm 0.0021) \cdot 10^3$ |
| | gluon | $g$ | 1 | 0 |
| | Higgs | $h$ | 0 | $(125.10 \pm 0.14) \cdot 10^3$ |

#### 2.1.1.1 Fermions

The elementary fermions are spin $\frac{1}{2}$ particles and described by an object $\Psi$, a spinor that transforms according to the $(\frac{1}{2},0)\oplus(0,\frac{1}{2})$ representation. Fermions are further categorized into quarks and leptons; the quarks participate in strong and weak interactions while the leptons interact only via the weak force, since they carry only the weak charge.

There are 6 types or "flavor" of leptons experimentally observed, arranged in the SM in 3 generations. Each generation consists of a charged lepton. The charged leptons are electron, muon and tau $(e,\mu,\tau)$ and the neutral leptons are the electron neutrino, muon neutrino and tau neutrino $(\nu_e,\nu_\mu,\nu_\tau)$. The electron is stable while muon and tau lepton decay via the weak interaction. In the SM, neutrinos are considered to be massless and they interact exclusively via the weak interaction.

Analogously, there are 6 different flavor of quarks grouped into 3 generations. Each generation consists of a pair of up-type quark (carrying electric charge $Q=+\frac{2}{3}$) and down-type quark ($Q=-\frac{1}{3}$). The up-type quarks are "up", "charm" and "top" quark $(u,c,t)$, while the down-type quarks are "down", "strange" and "bottom" quark $(d,s,b)$. Each quark carries non-zero colour charges in the three colour states (red,green,blue), weak charge and electric charge. Therefore, quarks are the only fundamental particles in the SM to experience all four[2] type of fundamental forces. In an isolated system, the strong and electromagnetic interactions conserve the number of different flavor of quarks, while the weak interaction can change the quark flavor. Quarks only exist in bound states with integer electromagnetic charge: quark-antiquark pairs (meson), or three quarks or anti-quarks (baryons). Mesons and baryons are subjected to strong interaction and are called hadrons.

#### 2.1.1.2 Gauge bosons

The gauge fields are introduced in the Lagrangian describing particles and their interaction as a consequence of gauge invariance. The quanta of the gauge fields are called gauge bosons, particles labeled as spin 1 and described by a vector $A$ that transforms according to the $(\frac{1}{2},\frac{1}{2})$ representation of $SU(2)$ symmetry group. Elementary particles whose interactions are described by a gauge theory, are interacting with each other through the exchange of gauge bosons.

The electromagnetic interaction is mediated through the exchange of photons between two electrically charged fermions. Since the photon is massless, the electromagnetic force has an infinite interaction range.

The force carriers for the weak interaction are the charged $W^+$, $W^-$ bosons and the neutral $Z^0$ boson, discovered at CERN in 1983 [8, 9]. The charged gauge bosons are named by the $W$ letter after weak force, while the neutral gauge boson name is $Z$ for its zero electric

[2] Including gravitational fource.

charge property. Since the $W^{\pm}$ and $Z^0$ bosons have a mass of the order of 100 GeV, the weak force acts on short range, in the order of $\approx 10^{-3}$ fm.

The quanta of the strong interaction field are called gluons, which are the force carrier of the strong force between colour charged particles. The gluons are massless and electrically neutral like the photon, but they have 8 different colour charge combinations. A single colour charge is assigned to quarks while gluons carry a colour and an anti-colour charge. Thus, there are 8 possible states for gluons, grouped into a colour octet and a singlet with three different colour charges. Unlike photons, gluons do carry colour charge by themselves and thus they can directly interact with one another. The strong force strength has a dynamic behavior, resulting in the fact that its strength becomes larger if the coloured charge particles are further apart.

Quantitatively speaking, on the scale of the size of a proton (radius of $\approx 10^{-15}$ m), the relative strength with respect to the strong force is $\frac{1}{100}$ for electromagnetic force and $\frac{1}{10000}$ for the weak force. To summarize, the fundamental forces and their respective gauge bosons are:

- The electromagnetic force, mediated by a massless $\gamma$,
- The weak force, mediated by two massive charged $W^+$,$W^-$ and a neutral $Z^0$ bosons,
- The strong force, mediated by 8 massless gluons,
- The gravitational force, which is hypothetically mediated by a spin 2 particle called graviton.

We observe that some bosons are massive and some are massless, and this intricate fact implies that there is something fundamental hidden in the SM. The gauge symmetry for the electroweak interaction $SU(2)_L \otimes U(1)_Y$ is broken as evident by the fact that its corresponding gauge bosons have different masses. This is only possible through the coupling to another fundamental boson called the Higgs boson, which is the quanta of the Higgs field introduced in the Lagrangian to give mass to the gauge bosons through the so called "Higgs mechanism", as will be described in Sec. 2.1.3. On the other hand, for the strong interaction which is based on the symmetry group of $SU(3)_C$, the symmetry is not broken and thus the corresponding spin-1 field quanta, gluons, are massless.

### 2.1.2 Fundamental interactions

All fundamental interactions between elementary particles are the consequence of gauge transformations under a certain gauge symmetry. In Quantum Electrodynamics Dynamics (QED), the Lagrangian for the electromagnetic interaction is defined by requiring the Lagrangian of a single fermion to be invariant under the $U(1)_{EM}$ gauge transformation.

The unified weak and electromagnetic interaction is the consequence of demanding the Lagrangian to be gauge invariant under the transformation of $SU(2)_L \otimes U(1)_Y$ symmetry group. The same goes for strong interaction, by imposing gauge invariance on the Lagrangian of a single quark under the transformation of $SU(3)_C$. Therefore, all three fundamental forces are described in the SM as its Lagrangian is formulated by invoking gauge invariance under the transformation of $SU(3) \otimes SU(2)_L \otimes U(1)_Y$ symmetry groups.

#### 2.1.2.1 Electromagnetic interaction

The free[3] Lorentz-invariant Lagrangian describing the relativistic motion of a fermion field $\psi$ is given by the Dirac Lagrangian:

$$\begin{aligned}\mathscr{L}_{Dirac} &= \bar{\psi}\left(i\gamma^{\mu}\partial_{\mu} - m\right)\psi \\ &= -m\bar{\psi}\psi + i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi\end{aligned} \tag{2.2}$$

The Lagrangian is invariant under the $U(1)$ global transformation. However, such transformation on the field at one point in space-time should not be correlated to any point in the whole Universe, otherwise this would violate special relativity, as no information can spread faster than light. Therefore a field should be transformed locally by taking into account of its continuous space and time coordinates $x_{\mu}$:

$$\psi \rightarrow \psi' = e^{ia(x)Q}\psi \tag{2.3}$$

where $a(x)$ is the arbitrary function dependent on space-time and $Q$ is the electromagnetic charge operator. Nonetheless, the transformed Lagrangian that followed is not locally $U(1)_{EM}$ invariant. In order to make a Lagrangian invariant under a local transformation, one needs to "gauge" the invariance by adding an extra term to the Lagrangian. The Lagrangian in Eq. (2.2) is made invariant by considering an interaction term between a spin $\frac{1}{2}$ field and a spin 1 field in the form of:

$$\mathscr{L}_{EM-int} = gA_{\mu}\bar{\psi}\gamma^{\mu}\psi \tag{2.4}$$

with an arbitrary coupling constant $g$ which determines the interaction strength of the two fields.

The gauge transformation[4] dictates that the spin 1 gauge field $A_{\mu}$ (the physical photon field) transforms as:

$$A_{\mu} \rightarrow A'_{\mu} = A_{\mu} + \partial_{\mu}a(x) \tag{2.5}$$

such that the equally local $U(1)_{EM}$ invariant Lagrangian is given by Maxwell Lagrangian:

$$\mathscr{L}_{Maxwell} = \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.6}$$

[3] Non-interacting, exists free of any external potential.

[4] The combined transformation of $\psi$, $\bar{\psi}$ and $A_{\mu}$ is called $U(1)$ gauge transformation.

and the field strength tensor $F_{\mu\nu}$ is given as:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.7}$$

Note that Maxwell Lagrangian is derived from the Proca Lagrangian[5] whose mass term $m^2 A_\mu A^\mu$ is neglected[6] as a consequence of the transformation in Eq. (2.5).

By adding all the $U(1)_{EM}$ gauge invariant Lagrangian terms from Eq. (2.2), Eq. (2.4) and Eq. (2.6), we finally arrive at the QED Lagrangian describing the electromagnetic interaction between the fermions and the photon field:

$$\begin{aligned}\mathscr{L}_{QED} &= \mathscr{L}_{Dirac} + \mathscr{L}_{EM-int} + \mathscr{L}_{Maxwell} \\ &= \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - g\bar{\psi}\gamma^\mu \psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\end{aligned} \tag{2.8}$$

To unclutter the notation, it is conventionally and mathematically motivated to introduce the covariant derivative to replace the ordinary derivative, which takes the form:

$$\partial^\mu \rightarrow D^\mu \equiv i\partial^\mu - igA^\mu \tag{2.9}$$

and finally Eq. (2.8) can be written as:

$$\mathscr{L}_{QED} = -m\bar{\psi}\psi + i\bar{\psi}\gamma_\mu D^\mu \psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.10}$$

Noether's theorem states that the transformation in Eq. (2.3) under the internal symmetry $U(1)_{EM}$ is connected to a conserved quantity. The Noether's current corresponds to an electric four-current:

$$j_\mu^{EM} \equiv -g\bar{\psi}\gamma^\mu \psi \tag{2.11}$$

and the conserved quantity, the electric charge $Q$ is determined by integrating $j_0^{EM}$[7]:

$$Q = -g\int d^3x \bar{\psi}\gamma^0 \psi \tag{2.12}$$

In the quantum framework, the object $\Psi$ is related to the probability amplitude and this interpretation requires the overall probability $\int d^3x\bar{\psi}\gamma^0\psi = 1$. By this assertion, Eq. (2.12) implies the coupling strength $|-g|^2$ is in fact the conserved quantity, which for electromagnetism is the electric charge.

[5]The Proca Lagrangain is given by $\mathscr{L}_{Proca} = \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + m^2 A_\mu A^\mu$.

[6]The absence of mass term for photon is consistent with the experimental observation.

[7]The zeroth component of $j_\mu^{EM}$ corresponding to electric charge density.

#### 2.1.2.2 Electroweak interaction

Inspired by the success of identifying a $U(1)$ internal symmetry in the Dirac Lagrangian, one would like to know if there exist any other internal symmetry in the Lagrangian. It turns out that we can find an internal symmetry for 2 massless spin $\frac{1}{2}$ fields, which is given by adding twice the Dirac Lagrangian shown in Eq. (2.2) without the mass term[8]:

$$\mathscr{L}'_{D1+D2} = i\bar{\Psi}\gamma_{\mu}\partial^{\mu}\Psi \tag{2.13}$$

and the Lagrangian acts on the doublet $\Psi$ consisting of 2 Dirac spinor:

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \tag{2.14}$$

Each spinor is a Lorentz invariant object constructed by a pair of left-handed and right-handed Weyl spinors[9], and the linear combination in the basis of left-chiral and right-chiral state is written as:

$$\psi = \begin{pmatrix} \chi_L \\ \zeta_R \end{pmatrix} = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \zeta_R \end{pmatrix} \tag{2.15}$$

The left-chiral and the right-chiral state can be projected out of the doublet $\Psi$ through the projection operator $P_L$, $P_R$:

$$P_L\psi = P_L \begin{pmatrix} \chi_L \\ \zeta_R \end{pmatrix} = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix} \equiv \psi_L \tag{2.16}$$

$$P_R\psi = P_R \begin{pmatrix} \chi_L \\ \zeta_R \end{pmatrix} = \begin{pmatrix} 0 \\ \zeta_R \end{pmatrix} \equiv \psi_R \tag{2.17}$$

such that $P_L$, $P_R$ are defined as:

$$P_{L/R} = \frac{I \mp \gamma^5}{2} \quad ; \quad \gamma^5 = i\gamma_0\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.18}$$

The term $I$ is the identity matrix, and $\gamma^5$ is called the chirality operator, because states of pure chirality are eigenstate of $\gamma^5$ with eigenvalue -1 and +1 respectively.

We learn from Madam Wu's experiment that only left-handed fermions are exclusively interacting via weak interaction [10], implying that the Lagrangian Eq. (2.13) is not invariant under parity transformation. To accommodate the idea of parity violation in SM, only the left-chiral fields are written as $SU(2)_L$ doublets as the fields are mixed and transformed into each other. Following the same line of thought, the right-chiral fields are described by $SU(2)$ singlets as they do not transform into each other. Therefore under the electroweak gauge symmetry $SU(2)_L \otimes U(1)_Y$, the left-handed fermion fields transform as a doublet and the

[8] Otherwise the Lagrangian will not be globally invariant under $SU(2)$ transformation.

[9] The Weyl spinor is a two component spinor encoding the spin up state and spin down state.

right-handed fermion fields transform as a singlet:

$$\Psi_L \to \Psi'_L = e^{i\vec{\alpha}(x)I_i + i\beta(x)Y}\Psi_L \qquad SU(2)_L \tag{2.19a}$$
$$\psi_R \to \psi'_R = e^{i\beta(x)Y}\psi_R \qquad U(1)_Y \tag{2.19b}$$

where sum over the index $i$ is implicitly assumed and $\vec{\alpha}(x) = (\alpha^1(x), \alpha^2(x), \alpha^3(x))$, $\beta(x)$ are arbitrary real functions of the space-time coordinates. There are 3 generators of the $SU(2)_L$ group[10], called isospin vector operators, given as $I_i = \frac{\sigma_i}{2}$ (i=1,2,3) and $\sigma_i$ are the familiar Pauli spin matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.20}$$

The parameter $Y$ in Eq. (2.19b) is the weak hypercharge operator relating the two conserved quantities $Q$ from Eq. (2.12) and $I_3$ according to the Gell-Mann-Nishijima relation [11]:

$$Y = 2(Q - I_3) \tag{2.21}$$

The weak hypercharge for fermion fields, given their electric charges and isospin charges, is summarized in Table 2.2. The weak isospin doublet is defined by the eigenstate of $\frac{1}{2}\sigma_3$ generator[11] and the corresponding eigenvalues are the isospin charge $+\frac{1}{2}$ and $-\frac{1}{2}$. Hence, the following notation is adopted to distinguish between the handedness for three generation:

1. The left-handed lepton doublet:

$$l^i_L = \begin{pmatrix} \nu_i \\ e_i \end{pmatrix}_L \quad , \quad i = 1, 2, 3 \tag{2.22}$$

2. The left-handed quark doublet:

$$q^i_L = \begin{pmatrix} u_i \\ d'_i \end{pmatrix}_L \quad , \quad i = 1, 2, 3 \tag{2.23}$$

3. The right-handed lepton singlet:

$$\text{up-type lepton} \quad \nu^i_R \quad , \quad i = 1, 2, 3 \tag{2.24}$$
$$\text{down-type lepton} \quad e^i_R \quad , \quad i = 1, 2, 3 \tag{2.25}$$

4. The right-handed quark singlet:

$$\text{up-type quark} \quad u^i_R \quad , \quad i = 1, 2, 3 \tag{2.26}$$
$$\text{down-type quark} \quad d^i_R \quad , \quad i = 1, 2, 3 \tag{2.27}$$

[10] The generator of $SU(2)$ is a 2 × 2 Hermitian traceless matrices consists of 3 degree of freedom.

[11] Only Pauli matrix $\sigma_3$ is diagonal, the other generators $\sigma_1$ and $\sigma_2$ are not eigenstates for the doublet.

with the subscript $i$ running over the different fermion families. The down-type quark mass eigenstate in the doublet Eq. (2.23) is expressed in terms of eigenstates of the weak interaction, $d_i' = \sum_j V_{ij} d_j$, where $V$ is the Cabbibo-Kobayashi-Maskawa mixing matrix [12], to account for the experimental observation of the quark mixing.

TABLE 2.2: The weak hypercharge for left-handed and right-handed fermion field.

| Field | $e_L$ | $\nu_{eL}$ | $u_L$ | $d_L$ | $e_R$ | $u_R$ | $d_R$ |
|---|---|---|---|---|---|---|---|
| Hypercharge Y | -1 | -1 | $\frac{1}{3}$ | $\frac{1}{3}$ | -2 | $\frac{4}{3}$ | -$\frac{2}{3}$ |

The Lagrangian Eq. (2.13) is made gauge invariant under the transformation in Eq. (2.19a) and Eq. (2.19b) by introducing gauge vector fields in covariant derivative form:

$$D_\mu = \partial^\mu - igB^\mu Y - ig'W^\mu \tag{2.28}$$

The $W^\mu$ gauge fields defined as $W^\mu \equiv \frac{\sigma_i}{2}(W^\mu)_i, i = 1,2,3$, are associated with $SU(2)_L$ group and the $B_\mu$ gauge field is associated with $U(1)_Y$ group; $g$, and $g'$ are their respective gauge coupling constants. The gauge fields transform itself as:

$$B_\mu \to B'_\mu = B_\mu - \frac{1}{g}\partial_\mu \beta(x) \qquad U(1)_{EM} \tag{2.29a}$$

$$W^i_\mu \to W^{i\prime}_\mu = W^i_\mu - \frac{1}{g'}\partial_\mu \alpha^i(x) - \epsilon_{ijk}\alpha^j(x)W^k_\mu \qquad SU(2)_L \tag{2.29b}$$

where $\epsilon_{ijk}$ is the structure constant of $SU(2)$ group given by $[\sigma^i, \sigma^j] = i\epsilon_{ijk}\sigma^k$. The respective transformation above resulted into local invariant Lagrangian for each gauge fields:

$$\mathscr{L}_{gauge} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}W^{\mu\nu} \tag{2.30}$$

with the field strength tensors $B_{\mu\nu}$ and $W_{\mu\nu}$ gauge fields defined as:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.31a}$$

$$W^{\mu\nu} = \partial^\mu W^\nu - \partial^\nu W^\mu - ig[W^\mu, W^\nu] \tag{2.31b}$$

The last term in Eq. (2.31b) arises from the non-Abelian nature of the group $SU(2)$, since the generators $\sigma^i (i = 1,2,3)$ do not commute with each other. By summing the contribution from the Lagrangians of Eq. (2.13) and Eq. (2.30), and replacing the ordinary derivative $\partial^\mu$ with covariant derivative Eq. (2.28), the Lagrangian can be written compactly as:

$$\mathscr{L} = i\bar{\Psi}_L \gamma_\mu D^\mu \Psi_L - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}W^{\mu\nu} \tag{2.32}$$

From the unification of two forces, we obtain the relation described in Eq. (2.21), implying that the $A_\mu$ field associated with the generator $Q$ of $U(1)_{EM}$ must be a combination of gauge fields $W_\mu^3$ and $B_\mu$, which are associated to the generators $I_3$ and $Y$ respectively. Consequently the mixing between the two gauge fields $W_\mu^3$ and $B_\mu$ defines the neutral weak boson $Z_\mu$ and the photon field $A_\mu$. On the other hand, the charged weak bosons $W^+$ and $W^-$ are given as a complex combination of the two gauge fields, $W_\mu^1$ and $W_\mu^2$. In summary, we have:

$$\text{the photon} \qquad A^\mu = \frac{1}{\sqrt{g^2+g'^2}}\left(gW_3^\mu + g'B^\mu\right) \tag{2.33a}$$

$$\text{the neutral weak boson} \qquad Z^\mu = \frac{1}{\sqrt{g^2+g'^2}}\left(g'W_3^\mu - gB^\mu\right) \tag{2.33b}$$

$$\text{the charged weak bosons} \qquad W_\pm^\mu = \frac{1}{\sqrt{2}}(W_1^\mu \mp iW_2^\mu) \tag{2.33c}$$

and the two coupling strengths $g$ and $g'$ are related by the weak mixing angle or Weinberg angle, $\theta_W = \tan^{-1}(g'/g)$.

The absence of mass terms for both boson and fermion fields in the form $\frac{1}{2}m^2W^\mu W_\mu$, and $\frac{1}{2}m^2(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)$ are justifiable as these terms would destroy the gauge invariance of the Lagrangian. Therefore, the Lagrangian Eq. (2.32) describes only a system of massless fermions and bosons. However, experimental results show that all gauge bosons of weak interactions are massive. To include the mass terms for $W^\pm$ and $Z$ without breaking the local gauge invariance, we need the "Higgs mechanism" to account for the mass for those bosons and subsequently for the masses of the other particles, as it will be discussed in Sec. 2.1.3.

#### 2.1.2.3 Strong interaction

The QCD is the quantum field theory of strong interactions describing the interaction among color charge carried particles. The theory is based on the $SU(3)$ gauge group of the quark color fields. Hence, the free Lorentz-invariant Lagrangian describing the dynamics of a system consisting of 3 massless spin $\frac{1}{2}$ fields is given as

$$\mathscr{L}_{Q1+Q2+Q3} = \bar{\Psi}_q(i\gamma^\mu\partial_\mu - m)\Psi_q \tag{2.34}$$

where the $\Psi_q$ is defined as a triplet of spin $\frac{1}{2}$ fields:

$$\Psi_q = \begin{pmatrix}\psi_{q1}\\ \psi_{q2}\\ \psi_{q3}\end{pmatrix} \tag{2.35}$$

This symmetry group describes the transformation property with $3 \times 3$ matrices with unit determinant, and requires those generators to be Hermitian and tracelss. The triplet transforms under the $SU(3)_C$ gauge group according to:

$$\psi_q \to \psi'_q = e^{i\alpha_A(x)T_A}\psi_q \tag{2.36}$$

where $\alpha_A(x)$ $(A = 1, ..., 8)$ are arbitrary real functions and $T_A$ are the 8 generators[12] of the $SU(3)$ group. Just as the generators of $SU(2)$ are connected to the Pauli matrices, these $SU(3)$ generators are connected to the Gell-Mann matrices, $T_A = \frac{1}{2}\lambda_A$. The Gell-Mann matrices $\lambda_A$ are:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \qquad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \tag{2.37}$$

Through a local $SU(3)$ transformation, one gets a total of 8 conserved quantities. The particles inside the triplet are labeled by the eigenvalue of the conserved quantities that belongs to the diagonal generator. The $SU(3)$ group has two such generators, $\frac{1}{2}\lambda_3$ and $\frac{1}{2}\lambda_8$:

$$\frac{1}{2}\lambda_3 = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad ; \quad \frac{1}{2}\lambda_8 = \begin{pmatrix} \frac{1}{2\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{2\sqrt{3}} & 0 \\ 0 & 0 & -\frac{1}{\sqrt{3}} \end{pmatrix} \tag{2.38}$$

and their eigenvalues are $+\frac{1}{2}, -\frac{1}{2}, 0$ and $\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, -\frac{1}{\sqrt{3}}$ respectively. Therefore, the strong interacting fermion carries two additional labels[13] in the triplet. A given quark $Q$ can be assigned with the following labels:

$$\text{colour red} \qquad \left(+\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \qquad \text{for} \quad q_r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} Q \tag{2.39a}$$

$$\text{colour blue} \qquad \left(-\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \qquad \text{for} \quad q_b = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} Q \tag{2.39b}$$

$$\text{colour green} \qquad \left(0, -\frac{1}{\sqrt{3}}\right) \qquad \text{for} \quad q_g = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} Q \tag{2.39c}$$

where the label uniquely defines the colour charge for each quark. Since leptons do not participate in the strong interaction, they transform under the $SU(3)$ gauge group as a singlet and have zero colour charge. In addition, one can use the eight-dimensional adjoint representation of $SU(3)$ to assign colour to the gauge fields $G_\mu$, the gluons.

[12] There are 3 × 3 traceless Hermitian matrices.

[13] In the basis spanned by the eigenvectors of the Cartan generators [13].

The $SU(3)$ gauge invariance is invoked, resulting into the addition of eight spin 1 vector gauge fields $G_\mu$ through replacing the partial derivative with covariant derivative:

$$D_\mu = \partial^\mu - i g_s G_\mu \tag{2.40}$$

into the Lagrangian. The $G_\mu$ gauge fields are defined as $G_\mu \equiv T^C (G_\mu)^C$, where $g_s$ is the gauge coupling constant. The gauge fields transform as:

$$G^A_\mu \rightarrow (G^A_\mu)' = G^A_\mu - \frac{1}{g_s}\partial_\mu \alpha(x)^A - f^{ABC}\alpha^B(x)G^C_\mu \tag{2.41}$$

where $f^{ABC}$ are the structure constants for $SU(3)$ and A,B,C can take on every value from 1 to 8, given by $[T_A, T_B] = i f^{ABC} T^C$. The field strength tensor $G_{\mu\nu}$ is defined as:

$$G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu - g_s[G_\mu, G_\nu] \tag{2.42}$$

and the dynamical contribution from the Lagrangian of $G_\mu$ gauge fields is given as:

$$\mathcal{L}_{gauge} = -\frac{1}{4}(G_{\mu\nu})^A (G^{\mu\nu})^A \tag{2.43}$$

Finally, the sum of Eq. (2.34) and Eq. (2.43) produces the QCD Lagrangian:

$$\begin{aligned}\mathcal{L}_{QCD} &= \mathcal{L}_{Q1+Q2+Q3} + \mathcal{L}_{gauge} \\ &= \bar{Q}(i\gamma^\mu D_\mu - m)Q - \frac{1}{4}(G_{\mu\nu})^A (G^{\mu\nu})^A \end{aligned} \tag{2.44a}$$

The $SU(3)$ symmetry is not broken because the mass term in the form of $\bar{Q}mQ$ is $SU(3)$ gauge invariant, as long as all particles in the triplet have equal mass. Hence, the objects inside a triplet describe the same quark in different colour, which indeed have equal mass.

The main feature of the strong force is that the interaction becomes stronger as the distance increases and vice versa. To study the dynamical behavior of the strong force, the QCD effective coupling constant defined as $\alpha_s = g_s^2/4\pi$ is expressed in terms of the momentum transfer $Q^2$[14] between the interacting quarks:

$$\alpha_s(Q^2) = \frac{1}{b_o ln(Q^2/\Lambda_{QCD})} \tag{2.45}$$

with $b_0$ being the leading term in the perturbative QCD expression [14], and $\Lambda_{QCD}$ defines the cut-off of QCD perturbative theory validity. The $\Lambda_{QCD}$ is found to be $\approx$ 214 MeV corresponding to a distance of $\approx$ 1 fm (the size of a nucleus). Below the $\Lambda_{QCD}$ scale, we have the non-perturbative regime where soft parton interaction physics is dominating. On the other hand, at the higher $Q^2$ scale relevant to the LHC collisions ($Q^2_{LHC} >> \Lambda_{QCD}$), the $\alpha_s$ becomes so small

[14] The four-vector energy-momentum transfer defined as $Q^2 = -q \cdot q$, where $q = (E_q, \vec{q}) = p_1 - p_2 = (E_1 - E_2, \vec{p}_1 - \vec{p}_2)$ for $p_1(p_2)$ are four momenta before and after interaction.

(colour charge decreased) that quarks and gluons are approximately treated as free particles and their interactions can be described by the perturbation theory. This important feature is known as "asymptotic freedom", which allows the application of perturbative methods when computing quantitative predictions on parton-parton[15] hard scattering processes.

In high energy physics experiments, observable particles like hadrons are produced from quarks and gluons via the showering process, called hadronization. When a parton is produced in an interaction, the QCD potential tries to keep the parton bounded until the energy is sufficient to create $q\bar{q}$ pairs along showered path seeded by the parton's projectile. The new partons produced by showering are approximately collinear with the original parton and combine into mesons or baryons, such that a jet of colorless particles is observed. As a consequence, no coloured objects like quarks are observed in the final state of a deep inelastic scattering process; this property of the QCD dynamics is known as "quark confinement".

During fragmentation, particle energies become increasingly smaller and perturbative QCD is no longer applicable. This makes most fragmentation processes non-perturbative and not analytically predictable. In order to describe jet properties, phenomenological models are usually applied and included in the Monte Carlo (MC) simulation of the jet kinematics, which will be discussed in Sec. 4.5.

### 2.1.3 The Higgs mechanism

The requirement of massless guage fields in electroweak Lagrangian Eq. (2.32) is gravely contradicting the experimental observation, that the weak interaction is mediated by massive $W^+$, $W^-$ and $Z$ gauge bosons. In order to retain gauge invariance, a new scalar potential called Higgs potential is introduced in the Lagrangian, and by breaking the symmetry spontaneously, the particle masses are generated naturally by interaction with a new scalar field called the Higgs field. This collective process generating the particle masses under the SM is termed as the "Higgs mechanism".

To begin with, we consider a locally $SU(2)$ and $U(1)$ invariant Lagrangian for spin-0 fields written as:

$$\mathscr{L}_{Higgs} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \tag{2.46}$$

such that the complex scalar doublet $\Phi$ needed to generate masses of the electroweak gauge bosons consists of a neutral scalar field $\phi^0$ and a charged field $\phi^+$:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.47}$$

[15]The colour-charged constituents of hadrons, such as quarks, anti-quarks, and gluons, are collectively called partons.

This minimal Higgs model has two complex scalar fields forming a weak isospin doublet where the upper and lower components of the doublet differ by one unit of charge. The weak isospin $(T, T^3)$ and hypercharge $(Y)$ quantum number of the Higgs doublet are summarized in Table 2.3. The doublet is transformed under the symmetry $SU(2) \otimes U(1)_Y$ as:

TABLE 2.3: The properties of the Higgs complex scalar doublet in the minimal Higgs model.

| Field | Isospin $T$ | Isospin $T^3$ | Hypercharge $Y$ |
|---|---|---|---|
| $\phi^+$ | $\frac{1}{2}$ | $\frac{1}{2}$ | 1 |
| $\phi^0$ | $\frac{1}{2}$ | $-\frac{1}{2}$ | 1 |

$$\Phi \rightarrow \Phi' = e^{ib_i(x)\frac{\sigma_i}{2}}\Phi \tag{2.48}$$

and the gauge fields $W^\mu$ and $B^\mu$ in Eq. (2.46) are transformed as shown in Eq. (2.29a) and Eq. (2.29b).

The Higgs potential, $V(\Phi)$ is an interaction energy characterized by two independent parameters $\mu$, $\lambda$:

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \tag{2.49}$$

where the two constant parameters are given as $\mu^2 < 0$ and $\lambda > 0$, so that it has a minimum value different than zero and it is bounded from below. The quadratic nature of the potential implies that the minimized $V(\Phi)$ corresponds to a non-zero infinite minimum field value for every value of $\varphi$:

$$\phi_{min} = \pm\frac{1}{\sqrt{2}}\nu e^{i\varphi} \quad ; \quad \nu = \sqrt{\frac{\mu^2}{\lambda}} \tag{2.50}$$

where $\nu$ is called the vacuum expectation value (VEV) of the Higgs field, approximated as 246 GeV [7]. All these minima lie on the circle with radius $\frac{1}{\sqrt{2}}\nu$, as shown in Fig. 2.1. By choosing a new vacuum state for the doublet out of infinite possible vacuum states, the Lagrangian is no longer invariant under the gauge symmetry group and thus the symmetry is broken[16]. An economical choice for the minimum is:

$$\Phi_{min} = \begin{pmatrix} \phi^+_{min} \\ \phi^0_{min} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu \end{pmatrix} \tag{2.51}$$

The fluctuation around the minimum energy are calculated using perturbation theory, by expanding the field $\phi(x)$ around the chosen vacuum state, Eq. (2.51). The excited state of the Higgs doublet is given by the first order in the series expansion:

$$\Phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} \phi_{1r}(x) + i\phi_{1c}(x) \\ \nu + \phi_{2r} + i\phi_{2c}(x) \end{pmatrix} \approx e^{ib_i(x)\cdot\frac{\sigma_i}{2}}\begin{pmatrix} 0 \\ \frac{\nu+h(x)}{\sqrt{2}} \end{pmatrix} \tag{2.52}$$

[16]This procedure is colloquially termed as spontaneous symmetry breaking.

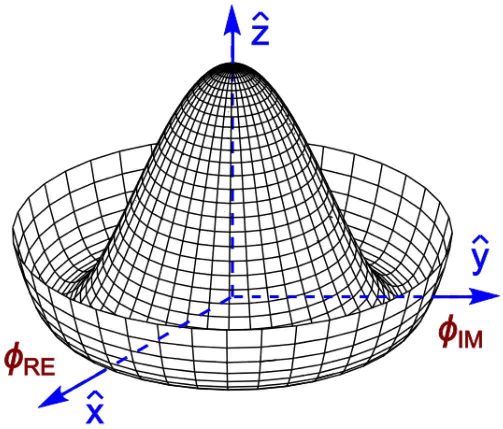



FIGURE 2.1: The Higgs potential for the complex scalar doublet field $\Phi$. The $\hat{z}$-axis corresponds to the interaction energy potential $V(\Phi)$, and $\phi_{RE}$, $\phi_{IM}$ denotes the real and imaginary part of the doublet [13].

parameterized by the four field components $\phi_{1r}$, $\phi_{1c}$, $\phi_{2r}$, $\phi_{2c}$. Due to gauge invariance, the exponent term in Eq. (2.52) can be eliminated according to the transformation, Eq. (2.48). By gauging away the three gauge fields[17] $\phi_{1r}$, $\phi_{1c}$, $\phi_{2c}$ and choosing only one gauge field, the excited state Eq. (2.52) is rewritten as:

$$\Phi_{un}(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.53}$$

The unitary gauge approach implies that out of four gauge fields in the complex scalar doublet, Eq. (2.52), three has been transformed into the longitudinal components of the gauge bosons $W^{\pm}$, $Z^0$, while the remaining component, which is neutral and scalar, becomes the physical field $h$ called the Higgs field. The excitation of the Higgs field which describes the particle state above the chosen vacuum state gives rise to a new particle, the Higgs boson, which is experimentally detectable to verify the theory.

Under the spontaneous symmetry breaking framework, the Higgs Lagrangian Eq. (2.46) can be expressed in terms of the VEV and the Higgs field $h(x)$, by adding a constant term $-\frac{1}{4}\mu^2 v^2$ to the Higgs potential Eq. (2.49) and inserting Eq. (2.53) in Eq. (2.46):

$$\begin{aligned} \mathscr{L}_{Higgs} = &\frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g'^2\left((W_1^\mu)^2 + (W_2^\mu)^2\right)(v+h)^2 \\ &+ \frac{1}{8}\left(g'^2 W_3^\mu - gB^\mu\right)^2 (v+h)^2 - V(h) \end{aligned} \tag{2.54}$$

[17] This is only possible because the $SU(2)$ symmetry group is local. In contrast to the global symmetry, the field components can not be gauged away and they are commonly interpreted as massless Goldstone bosons [15].

where the Higgs potential is given by:

$$V(\Phi) \rightarrow V(h) = -\mu^2 h^2 - \frac{\mu^2}{v} h^3 - \frac{\mu^2}{4v^2} h^4 \tag{2.55}$$

Note that choosing a particular state, Eq. (2.51), renders the Lagrangian no longer invariant under the $SU(2)_L \otimes U(1)_Y$ gauge transformation. However, the assignment of VEV to the neutral component of the Higgs doublet ensures the conservation of electric charge and the $U(1)_{EM}$ symmetry remains unbroken.

Finally, the full physics content of the Lagrangian is obtained by replacing the gauge fields with their respective physical fields via substitution of Eq. (2.33a), Eq. (2.33b), Eq. (2.33c) into Eq. (2.54), adding the dynamical contribution of the gauge field Eq. (2.30):

$$\begin{aligned}
\mathscr{L} &= \mathscr{L}_{Higgs} + \mathscr{L}_{gauge} \\
\mathscr{L} &= (D_\mu \Phi_{un})^\dagger (D^\mu \Phi_{un}) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W_{\mu\nu} W^{\mu\nu} \\
&= \frac{1}{2} \partial_\mu h \partial^\mu h + \mu^2 h^2 + \frac{\mu^2}{v} h^3 + \frac{\mu^2}{4v^2} h^4 \\
&- \frac{1}{2} (W^+)_{\mu\nu} (W^-)^{\mu\nu} + \frac{1}{4} g'^2 v^2 (W^+)_\mu (W^-)^\mu - \left( \frac{h^2}{4} + \frac{vh}{2} \right) g'^2 (W^+)_\mu (W^-)^\mu \\
&- \frac{1}{4} Z_{\mu\nu} Z^{\mu\nu} - \frac{1}{8} v^2 (g^2 + g'^2) Z_\mu Z^\mu - \left( \frac{h^2}{8} + \frac{vh}{4} \right) (g^2 + g'^2) Z_\mu Z^\mu \\
&- \frac{1}{4} A_{\mu\nu} A^{\mu\nu}
\end{aligned} \tag{2.56}$$

The Eq. (2.56) describes the Lagrangian for a free massive neutral scalar boson field $h(x)$, a free massive neutral vector boson field $Z_\mu(x)$, a pair of massive charged vector boson fields $W_\mu^\pm(x)$, and a massless photon field $A_\mu(x)$. From the expected mass term in the form of $\frac{1}{2} m^2 \phi^\dagger \phi$, their respective mass can be identified from Eq. (2.56) as:

$$m_h = v\sqrt{2\lambda} \tag{2.57}$$

$$m_W = \left( \frac{g' v}{2} \right) \tag{2.58}$$

$$m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2} \tag{2.59}$$

$$m_\gamma = 0 \tag{2.60}$$

The electroweak bosons acquire mass through the VEV of the Higgs field and their respective gauge interaction coupling constants $g$, $g'$. The combination corresponding to Z boson in Eq. (2.33b), which is associated to the neutral Goldstone boson of the broken symmetry, has acquired mass through the Higgs mechanism and the field corresponding to the photon has remained massless. Hence, the resulting mass $m_\gamma = 0$ is used as a consistency check for the VEV assignment in the Higgs doublet.

Concerning the Higgs mass $m_h$, the parameter $v$ is fixed but $\lambda$ is not experimentally constrained in the SM; the mass of the Higgs boson is not predicted by the theory. Therefore, at the fundamental level, the existence of the Higgs boson is indispensable to the foundation of SM.

### 2.1.4 Fermion mass and Yukawa interaction

The Lorentz invariant mass term is usually expressed in combination of left-handed and right-handed fields:

$$\bar{\psi}\psi = \bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L \tag{2.61}$$

However, Eq. (2.61) is not invariant according to their respective transformation properties Eq. (2.19a), Eq. (2.19a) under $SU(2)$ due to handedness. Nevertheless, a $SU(2)$, $U(1)$ and Lorentz invariant[18] term that couple to a spin 0 doublet and spin $\frac{1}{2}$ spinor together with the equally allowed Hermitian conjugate term can be constructed as:

$$\mathscr{L} = -\lambda\left(\bar{\Psi}_L\Phi\psi_R + \bar{\psi}_R\bar{\Phi}\Psi_L\right) \tag{2.62}$$

The coupling constant, $\lambda$ is called a Yukawa coupling describing not only the interaction between the fermions and the Higgs field, but also leading to a mass terms for the spin $\frac{1}{2}$ field after the symmetry breaking. Since there is a second field in the doublet $\Psi$, two representations of the Higgs field are needed to give masses to the down quarks and electrons, and to the up quarks.

Therefore, the non-invariance can be fixed by introducing two representations of the Higgs field into the Lagrangian Eq. (2.61), $\phi$ and its hermitian conjugated version given as $\tilde{\phi}_i = \epsilon_{ij}\phi_j^\star$:

$$\Phi = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix} \equiv \begin{pmatrix}0\\ \frac{v+h}{\sqrt{2}}\end{pmatrix} \quad with \quad Y(\Phi) = +\frac{1}{2} \tag{2.63}$$

$$\tilde{\Phi} = \begin{pmatrix}\phi^0\\ \phi^-\end{pmatrix} \equiv \begin{pmatrix}\frac{v+h}{\sqrt{2}}\\ 0\end{pmatrix} \quad with \quad Y(\tilde{\Phi}) = -\frac{1}{2} \tag{2.64}$$

As a result, by using the notation Eq. (2.22) for leptons and Eq. (2.23) for quarks, the Yukawa Lagrangian describing the interaction between the Higgs and fermion field is given by:

$$\mathscr{L}_{Yukawa} = \mathscr{L}^{lepton}_{Yukawa} + \mathscr{L}^{quark}_{Yukawa} \tag{2.65}$$

$$\mathscr{L}^{lepton}_{Yukawa} = -\lambda_1^i\left(\bar{l}_L^i\Phi e_R^i + \bar{e}_R^i\Phi^\dagger l_L^i\right) - \lambda_2^i\left(\bar{l}_L^i\tilde{\Phi}\nu_R^i + \bar{\nu}_R^i\tilde{\Phi}^\dagger l_L^i\right) \tag{2.66}$$

$$\mathscr{L}^{quark}_{Yukawa} = -\lambda_d^i\left(\bar{q}_L^i\Phi d_R^i + \bar{d}_R^i\Phi^\dagger q_L^i\right) - \lambda_u^i\left(\bar{q}_L^i\tilde{\Phi}u_R^i + \bar{u}_R^i\tilde{\Phi}^\dagger q_L^i\right) \tag{2.67}$$

[18]The spin 0 field does not transform at all under Lorentz transformation.

where the Yukawa couplings, $\lambda_i$ with $i = 1, 2, 3$ correspond to the flavor interaction eigenstates. Note that the neutrino has no right-handed partner in SM, it will not acquire mass term through the Yukawa coupling, and thus the terms are omitted. By choosing the VEV for spin 0 field (breaking the symmetry), which is equivalent to inserting Eq. (2.53) into the Lagrangian Eq. (2.65). After expanding the fields at VEV, the lepton sector reads:

$$\begin{aligned}\mathscr{L}^{lepton}_{Yukawa} &= -\lambda^i_1 \left(\bar{l}^i_L \Phi_{un} e^i_R\right) + h.c. \\ &= -\frac{\lambda^i_1 v}{\sqrt{2}} \left(\bar{e}^i_L e^i_R\right) - \frac{\lambda^i_1 h}{\sqrt{2}} \left(\bar{e}^i_L e^i_R\right) + h.c.\end{aligned} \tag{2.68}$$

Analogously, the quark sector becomes:

$$\begin{aligned}\mathscr{L}^{quark}_{Yukawa} &= -\lambda^i_d \left(\bar{q}^i_L \Phi d^i_R\right) - \lambda^i_u \left(\bar{q}^i_L \tilde{\Phi} u^i_R\right) + h.c. \\ &= -\frac{\lambda^i_d v}{\sqrt{2}} \left(\bar{d}^i_L d^i_R\right) - \frac{\lambda^i_u v}{\sqrt{2}} \left(\bar{u}^i_L u^i_R\right) - \frac{\lambda^i_d h}{\sqrt{2}} \left(\bar{d}^i_L d^i_R\right) - \frac{\lambda^i_u h}{\sqrt{2}} \left(\bar{u}^i_L u^i_R\right) + h.c.\end{aligned} \tag{2.69}$$

Remarkably, by the same fashion the lepton mass terms can be deduced from Eq. (2.68) as:

$$m_e = \frac{\lambda_e v}{\sqrt{2}} \quad ; \quad m_\mu = \frac{\lambda_\mu v}{\sqrt{2}} \quad ; \quad m_\tau = \frac{\lambda_\tau v}{\sqrt{2}}$$

where $f_e$, $f_\mu$, $f_\tau$ are dimensionless Yukawa coupling for electron, muon, and tau lepton. While for the 6 flavor quark mass terms are inferred from Eq. (2.69) as:

$$\begin{aligned} m_u &= \frac{\lambda_u v}{\sqrt{2}} \quad ; \quad m_c = \frac{\lambda_c v}{\sqrt{2}} \quad ; \quad m_t = \frac{\lambda_t v}{\sqrt{2}} \\ m_d &= \frac{\lambda_d v}{\sqrt{2}} \quad ; \quad m_s = \frac{\lambda_s v}{\sqrt{2}} \quad ; \quad m_b = \frac{\lambda_b v}{\sqrt{2}}\end{aligned}$$

Through the Yukawa Lagrangians Eq. (2.68) and Eq. (2.69) the charged leptons and quarks not only acquire mass from the VEV of the Higgs field; in addition, an interaction term is introduced between the fermions fields and the Higgs fields, with a coupling strength proportional to the fermion mass given as:

$$\lambda = \frac{\sqrt{2} m_f}{v} \tag{2.70}$$

In summary, the possible Higgs boson interaction with SM particles under the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ are characterized by the Feynman diagrams shown in Fig. 2.2.

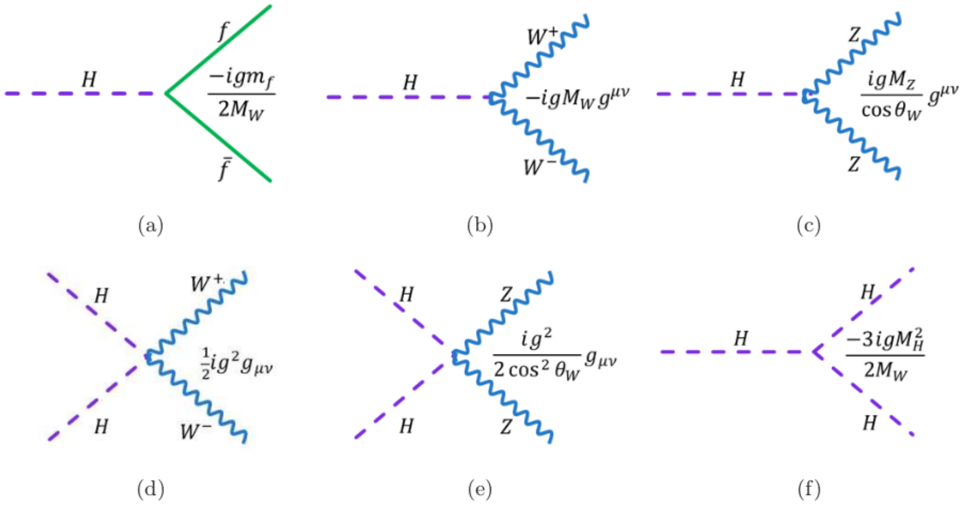



FIGURE 2.2: The Feynman diagrams for Higgs bosons interaction with itself, and to fermions and gauge bosons [16].

## 2.2 The Standard Model Higgs boson at Hadron Collider

The signatures for SM Higgs boson detection at collider experiments depend on the production process and the decay pattern. Both the production and the decay contribute to the specific kinematic features which can be used to distinguish signal from the background events. A multitude of experimental searches on Higgs boson were performed during the first run period of LHC with proton-proton collisions at $\sqrt{s} = 7, 8$ TeV. On 4 July 2012, the observation of a new particle at a mass of approximately 125 GeV with Higgs boson-like properties was finally reported by the A Toroidal Large hadron collider ApparatuS (ATLAS) [3] and CMS [2, 17] Collaborations with integrated luminosity of 5.1 (5.3) $\text{fb}^{-1}$ at 7 (8) TeV. Subsequent measurements of the new particle properties such as spin, parity, and coupling strength to SM particles are found to be consistent, within the experimental uncertainties, with the expectation for the SM Higgs boson [18, 19, 20, 21]. The full Run-II dataset collected during the LHC proton-proton collisions at a higher energy $\sqrt{s} = 13$ TeV provides more statistical power to further constrain the Higgs boson mass measurement and elucidate other production and decay modes.

### 2.2.1 Higgs boson production mode

The Higgs boson production in the pp collision occurs through four main production mechanisms, their Feynman diagrams corresponding to these production products are shown in Fig. 2.3 and their production cross section is presented in Fig. 2.4.

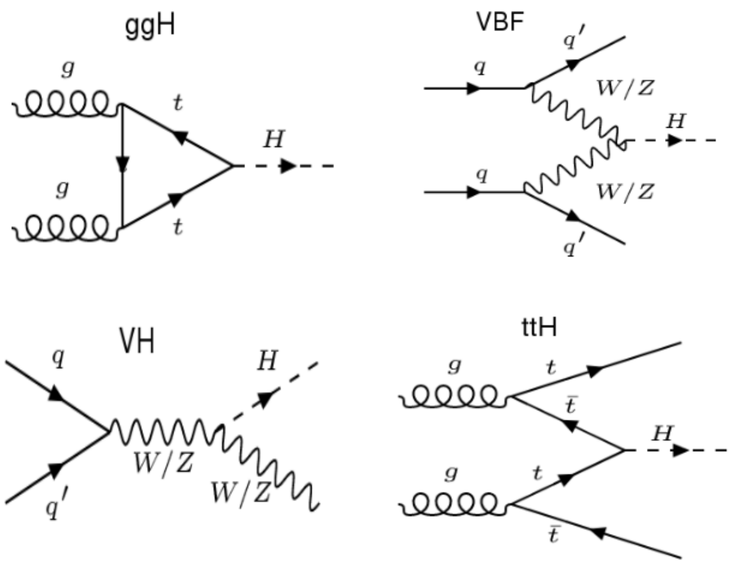



FIGURE 2.3: The leading order Feynman diagrams contributing to the Higgs boson production in pp interaction: gluon fusion (ggH), vector-boson fusion (VBF), vector boson associated production (VH), and top quark associated production (ttH).

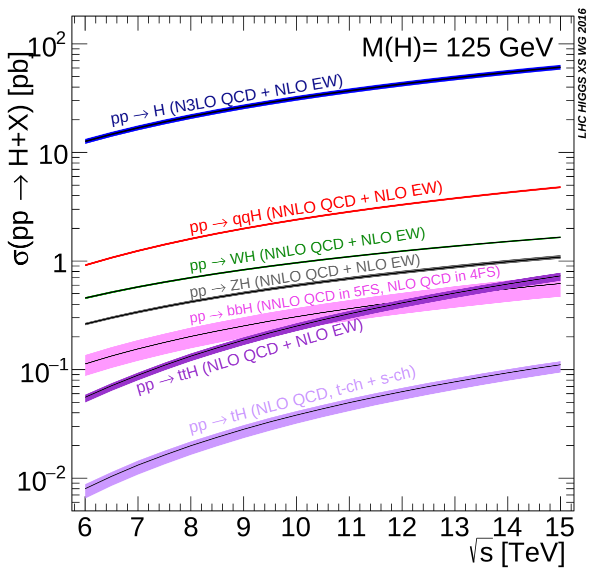



FIGURE 2.4: The theoretical prediction for SM Higgs boson production cross section as a function of the center of mass energy, $\sqrt{s}$, for proton-proton collisions. The major production cross section are gluon fusion, vector boson fusion, associated production $WH$, $ZH$, $b\bar{b}H$, $t\bar{t}H$ and $tH$ in descending order. The theoretical uncertainties are indicated as bands [7].

#### 2.2.1.1 Gluon fusion (ggH)

At tree level the Higgs boson hardly couples to light-flavor quarks as the coupling strength is proportional to their mass, and there is no coupling to gluons, since the $SU(3)_C$ symmetry of QCD is not broken. The ggH is the dominant Higgs production mechanism at LHC thanks to the abundance of gluons from the colliding protons, resulting in loop-induced process yielding a sizable production cross section. Thus, the top quark naturally becomes the candidate for the particle loop as its has strong coupling to gluons and the largest of all SM couplings to the Higgs boson [7]. The production mode is indicated as process $pp \rightarrow H$.

#### 2.2.1.2 Vector bosons fusion (VBF)

Since the SM Higgs boson has sizable couplings to the $W$, $Z$ bosons and to the top quark, it can be produced via their gauge boson couplings. In the VBF process, which has the second largest production cross section at the LHC, the two incoming quarks each radiate a $W$ or a $Z$ boson, which merge to form a Higgs boson. As the two outgoing quarks are not connected by any quark or gluon line, QCD radiation in the region between the two jets is suppressed, resulting in final states with two jets with high rapidity in the forward and backward regions of the detector. The production mode is indicated as process $qq \rightarrow qqH$.

#### 2.2.1.3 Associated production (VH)

The VH processes are characterized by smaller production cross section than VBF. This mode is experimentally feasible since the leptons and neutrinos from the $W$ or $Z$ decay could be a handle to select the signal events. The production mode is indicated as $qq \rightarrow VH$ with $V = W, Z$.

#### 2.2.1.4 Associated top quark production $t\bar{t}H$

This production mode provides a direct measurement of Yukawa coupling between top quark and Higgs boson. Despite of the low production cross section and the complex final state with large number of hadronic jets, it is possible to measure this channel thanks to the advancements of multivariate analysis and machine learning techniques. The channel was observed with a significance of $5.2\sigma$ (expected significant $4.2\sigma$) [22]. The production mode is indicated as $qq \rightarrow t\bar{t}H$.

### 2.2.2 Higgs boson decay modes

The Higgs boson is an unstable particle with very short lifetime[19]. Therefore, the presence of a Higgs boson is detected only through its decay products. Like other mediators, the Higgs boson is described by Breit-Wigner propagator for unstable particles. At tree level, all decay rates are determined by the Higgs coupling to SM particles which are constrained by the unitary condition. By definition the Higgs field couples to all particles[20], implying that different Higgs decay rates are proportional to the masses of decay products; therefore, the Higgs boson preferably decays into the heaviest state allowed by the phase space. The Fig. 2.5 shows the branching ratios for different decay modes as predicted by the SM.

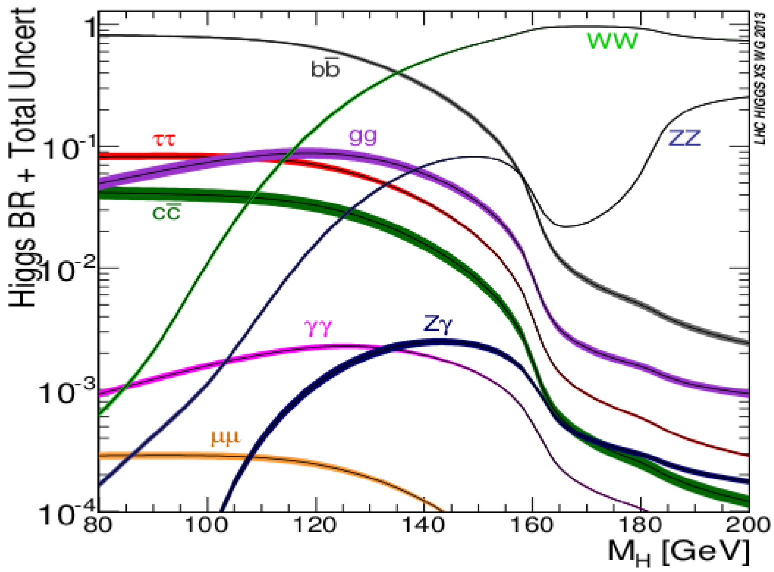



FIGURE 2.5: The Higgs boson branching ratios with uncertainty widths. The decay of the Higgs boson to a pair of W bosons has the second highest branching ratio at a pole mass of 125 GeV [23].

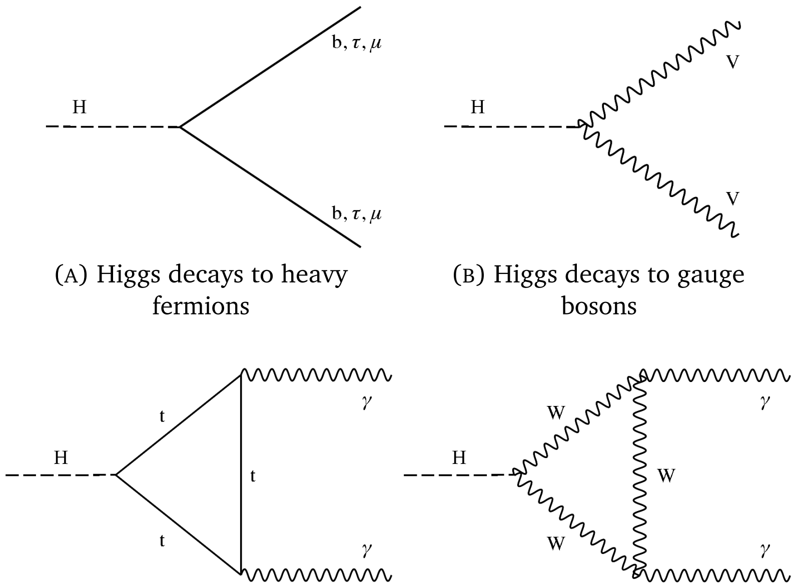
At leading order in the SM couplings, the Higgs boson can decay into pairs of heavy fermions through Yukawa interactions (Figure 2.6a), and into pairs of $W$ and $Z$ bosons through $SU(2)_L$ interactions (Figure 2.6b). The Yukawa decays have partial widths proportional to the square of the ratios between fermion mass and Higgs boson vacuum expectation value (Eq. (2.70)). Hence, for $m_H$ below the $t\bar{t}$ threshold, the decays are preferentially to $b\bar{b}$ pairs, with smaller contributions from $c\bar{c}$ pairs, $\tau^+\tau^-$ pairs, and negligible ones from the other fermion pairs. For $m_H$ above twice the mass of the $W$ or $Z$ bosons, the Higgs decays into a pair of on-shell gauge bosons becomes dominant.

The loop-induced decays into pairs of photons and gluons are also important. The decay into gluons is the mirror process of the gluon fusion production, happening through a quark loop

[19] $\tau \sim 1.6 \cdot 10^{-22} s$ [7].

[20] As previously discussed, Higgs boson does also couple to itself.

(Figure 2.3); the large Yukawa coupling of the top quark and the eight-fold color multiplicity of the final state allow this loop decay to be competitive to the tree-level decay into $\tau^+\tau^-$ pairs. Higgs boson decays into diphotons can happen both through a fermion loop (Figure 2.6c), dominated by the top quark contribution, and through a $W$ boson loop with two $WW\gamma$ vertices (Figure 2.6d). The partial width of the $H \to \gamma\gamma$ decay is about a factor 40 smaller than $H \to gg$, due to smaller electroweak couplings. Nevertheless this channel has several unique advantages, playing a major role both in the discovery of the Higgs boson and in the study of its properties.




(A) Higgs decays to heavy fermions

(B) Higgs decays to gauge bosons

(C) Higgs decays to 2 photons via top quark loop

(D) Higgs decays to 2 photons via gauge boson loop

FIGURE 2.6: Leading order Feynman diagrams for Higgs boson decays [24].

In LHC experiment, the searches and measurements on Higgs boson's properties are based on the Higgs decay mode. Those studies are relied on the sensitivity provided by the final state characterized by the decay products, reconstructed in the CMS detector. One of the measured properties of the Higgs boson is the signal strength modifier, $\mu = \sigma/\sigma_{SM}$, where $\sigma$ is the observed production cross section and $\sigma_{SM}$ is the value predicted by the SM for a given mass hypothesis[21].

The Higgs boson measurements at Run-I had covered the most sensitive bosonic channels, $H \to \gamma\gamma$, $H \to 4l$ and $H \to 2l2\nu$ from the ATLAS and CMS experiments independently; by combining the two results, the Higgs boson mass was measured at $m_H = 125.09 \pm 0.21_{stat} \pm 0.11_{syst}$ GeV [25], with the best-fit signal strength determination in agreement with SM expectation $\hat{\mu} = 1$ [20]. With the combined Run-I and partial Run-II datasets, the re-observation of Higgs boson with high significance at the expected mass in $H \to 4l$ and $H \to \gamma\gamma$ by either experiment was established [26, 27, 28, 29].

[21] The mass of the SM Higgs boson is 125.09 GeV [25].

The CMS experiment had used combination of decay channels to increase the observation of previously uncovered Higgs boson production mode and vice versa. For example, the $t\bar{t}H$ production mode has become accessible in combined Run-I and partial Run-II datasets, the sensitivity from the full spectrum of decay channels is optimized by using multivariate method [30] and deep neutral network [31], eventually the measurement established the observed (expected) significance of 5.2 $\sigma$ (4.2 $\sigma$) [22]. On the other hand, the evidence for $H \rightarrow \mu\mu$ decay mode has been established by a combined search in four exclusive categories sensitive to ggH, VBF, VH, and ttH production channels, reported an observed (expected) significance of 3.0 $\sigma$ (2.5 $\sigma$) [32].

Using the same combined dataset, the largest decay rate of Higgs boson decay channel, $H \rightarrow b\bar{b}$, has finally come to light, reported an observed (expected) significance of 5.6 $\sigma$ (5.5 $\sigma$) [33]. The measurement is possible only by using combination of deep neutral network and multivariate method to achieve high precision to control large QCD background in order to extract the signal.

Concerning the $H \rightarrow W^+W^-$ decay channel, the full Run-II dataset has not yet been completely analyzed by ATLAS nor CMS. Nevertheless, there have been analyses performed on partial dataset of full Run-II. ATLAS analyzed $WW \rightarrow e\nu\mu\nu$ decay channel in the ggH, VBF and $VH$ production modes with 2015 and 2016 datasets [34, 35]. With this limited dataset, the measured ggH signal strength yielded $\mu_{ggH} = 1.10 \pm 0.20$ [35], with the largest uncertainties being due to the experimental systematic uncertainties.

CMS performed a more complete analysis with the 2016 dataset, with most production channels covering both the opposite and same flavor final states of opposite charge leptons (electrons or muons), obtaining a combined inclusive signal strength of $\mu_{H\rightarrow WW} = 1.28$ $\pm$ 0.10 (stat) $\pm$ 0.11 (syst) $^{+0.10}_{-0.07}$ (theo) [5]. The same analysis also established the first observation of the Higgs boson decays into two W bosons with an observed (expected) global significance of 9.1$\sigma$ (7.1$\sigma$).

The VH production mode was measured with other dominant production mode in the global analysis. Particularly, it was measured by three event categories according to the number and/or flavor of the reconstructed jets and charged leptons in the final state. Those event categories were aimed to select leptonic decay of $H \rightarrow W^+W^-$, with 3-lepton WH-tagged and 4-lepton ZH-tagged categories targeted to associated produced leptonically decaying vector bosons; while 2-jet different flavor (DF) VH-tagged category targets hadronically decaying vector bosons. In the latter case, only DF phase space is considered to avoid Z+jets background contaminated phase space.

The combined inclusive signal strength for $\mu_{WH}$ was measured exclusively by 3 event categories, and reported about 1$\sigma$ deviation from SM expectation. The measurement was mainly driven by poor sensitivities from 2-jet DF VH-tagged and 3-lepton WH-tagged categories. The 2-jet DF VH-tagged category's sensitivity was reduced by large background contamination coming from t$\bar{t}$, non-resonant WW state and non-prompt background; Meanwhile, the 3-lepton WH-tagged category's suffered from non-prompt lepton backgrounds. Both

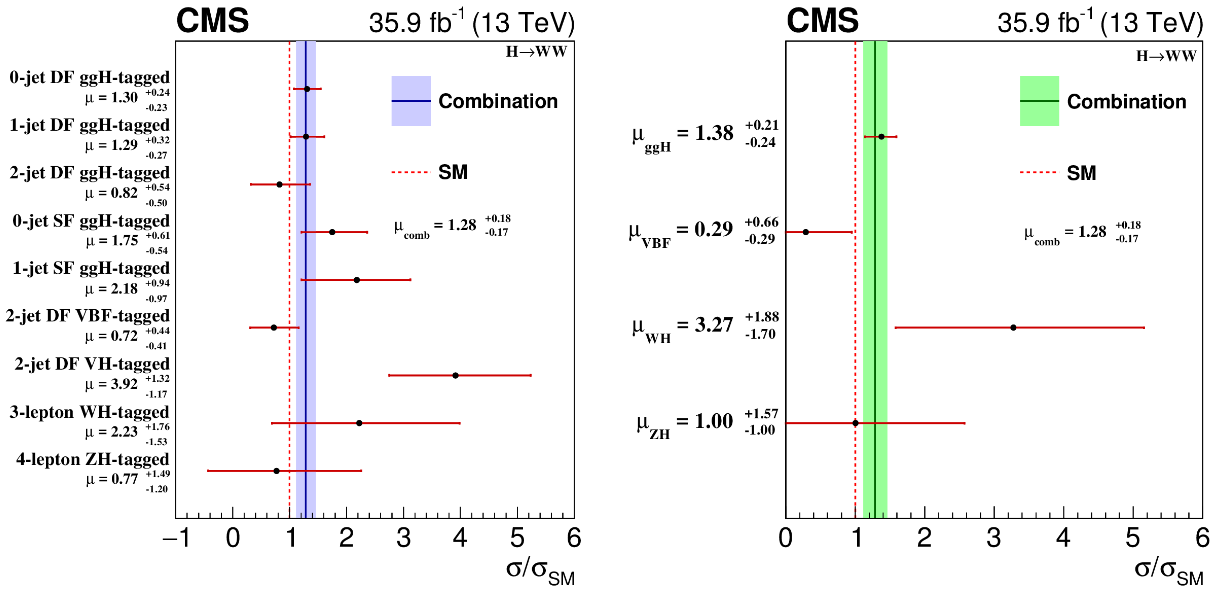



FIGURE 2.7: Observed signal strength modifiers for each category used in the combination (Left) and for the main SM Higgs boson production mechanisms (Right). The vertical continuous line represents the combined signal strength best fit value, while the horizontal bars and the filled area show the 68% confidence intervals. The vertical dashed line corresponds to the SM expectation [5].

categories were measured with limited statistics. The breakdown of each individual event category (production mode) contribution to the global signal strength $\mu_{comb}$ are shown in the left (right) plot in Fig. 2.7.

The $H \to W^+W^-$ analysis interpreted in STXS framework, using STXS "stage-0" [36] in bins of the vector boson transverse momentum, was reported in Fig. 2.8. At this stage, each main production mode has a single inclusive bin, with associated Higgs production separated into $q\bar{q} \to WH$, $q\bar{q} \to ZH$ and $gg \to ZH$ channels. Concerning the VH production mode, their respective cross section for $\sigma_{WHlep.}$ ($\sigma_{ZHlep.}$) are found to be $50^{+55}_{-46}$ ($4.1^{+9.6}_{-4.1}$) fb, and their signal strengths are compatible with SM expectation within its uncertainties [5].

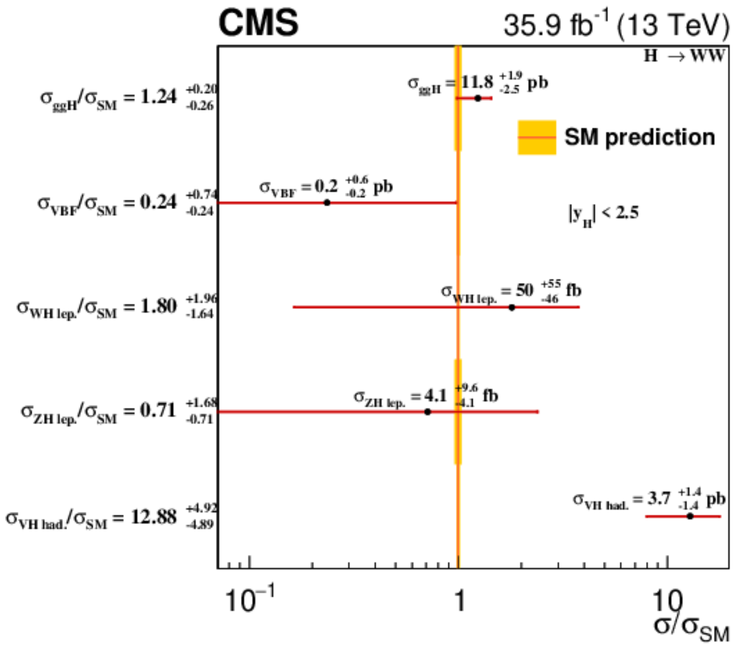



FIGURE 2.8: Observed cross sections for the main Higgs boson production modes, normalized to the SM predictions. Cross section ratios are measured in a simplified fiducial phase space defined in "stage-0" STXS. The vertical line and band correspond to the SM prediction and associated theoretical uncertainty [5].

# Chapter 3

# Experimental Setup

## 3.1 The Large Hadron Collider

The LHC is an accelerator and collider system located between Swiss and France border at the European Organization of Nuclear Research, or CERN. The LHC consists of a succession of machines that accelerate proton particles to increasingly higher energies to collide with each other, and in parallel condition the beams to maximize the effectiveness of the collision. Fig. 3.1 gives a schematic view of CERN's accelerator complex.

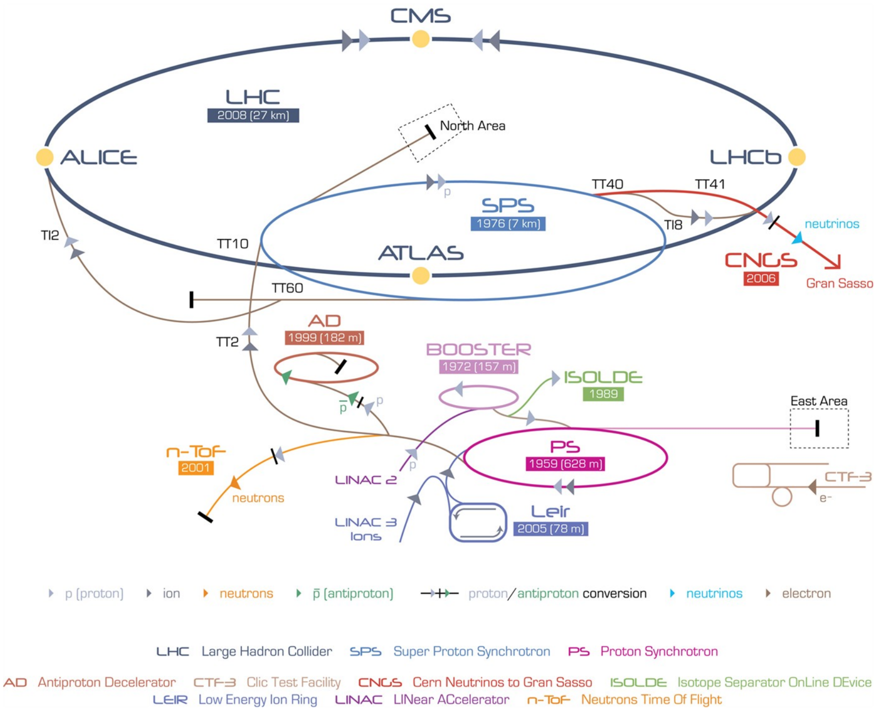



FIGURE 3.1: The CERN accelerator complex serving 4 major experiments [37].

The four major experiments generally named after the detector are instrumented at different points along the LHC where protons are made to collide. They are:

- A Toroidal Large hadron collider ApparatuS (ATLAS) [38]
- Compact Muon Solenoid (CMS) [39]
- Large Hadron Collider beauty (LHCb) [40]
- A Large Ion Collider Experiment (ALICE) [41]

The ATLAS and CMS detectors are the two general-purpose detectors deployed to explore much largest spectrum of physics. The use of two detectors provides an independent approach to cross-check the same physics process of interest. On the other hand, Large Hadron Collider beauty (LHCb) experiment focuses on Charge conjugation and Parity symmetry (CP) violation in electroweak interaction, in particular to study asymmetries between matter and anti-matter through rare decays of hadrons containing bottom quarks. Lastly, ALICE experiment designed to study heavy ion collisions, through the production of a new state of matter called quark-gluon plasma.

The LHC's figure of challenge is the center of mass energy and the luminosity. The latter is the important quantity use to indicate the collision rate at LHC. The LHC designed instantaneous luminosity $\mathscr{L}_{inst}$ is $10^{34}s^{-1}cm^{-2}$, translating into around 1 billion proton interaction per second. The $\mathscr{L}_{inst.}$ is dependent on the accelerator's construction parameters and the beams parameters, and it is given by the formula:

$$\mathscr{L}_{inst} = f\frac{\eta_1\eta_2}{4\pi\sigma_x\sigma_y} \tag{3.1}$$

where $\eta_1$ and $\eta_2$ are the number of particles contained in the two bunches colliding at a frequency $f$, while $\sigma_x$ and $\sigma_y$ are the beam sizes in the transverse plane. At the LHC collision frequency of $f = 40$ MHz, the transverse size of the beam is squeezed to around 15 $\mu m$.

It is conventional to quote the luminosity in a specific period, the integrated luminosity, defined by:

$$\mathscr{L} = \int \mathscr{L}_{inst} \cdot dt \tag{3.2}$$

For a given process, the number of observed events is related to the integrated luminosity $\mathscr{L}$ by:

$$N^{obs}_{events} = \sigma \times \epsilon \times \mathscr{L} \tag{3.3}$$

where $\sigma$ is the cross section of the process of interest, and $\epsilon$ is the efficiency of detection, optimized by the experimentalists.

So far, the LHC operated at three different p-p center of mass energy, $\sqrt{s} = 7$, 8 and 13 TeV, and in addition in a number of different heavy ion collision scenarios, involving lead or xenon nuclei. The 7 and 8 TeV pp operation began in 2010 and ended in 2012, a period that is referred to as Run-I; the 13 TeV operation began in 2015 and ended in 2018, a period referred to as Run-II. The Run-III period is scheduled to start in 2021, and the operation is anticipated to be at design luminosity, $\sqrt{s} = 14$ TeV. At the end of Run-III, a major upgrade, called the High Luminosity (HL) upgrade is planned to achieve the targeted integrated luminosity of 3000 $\text{fb}^{-1}$, with an increase by nearly one order of magnitude of the instantaneous luminosity with respect to Run-III. The timeline of the LHC programme between 2011 and 2037 year is described in Fig. 3.2

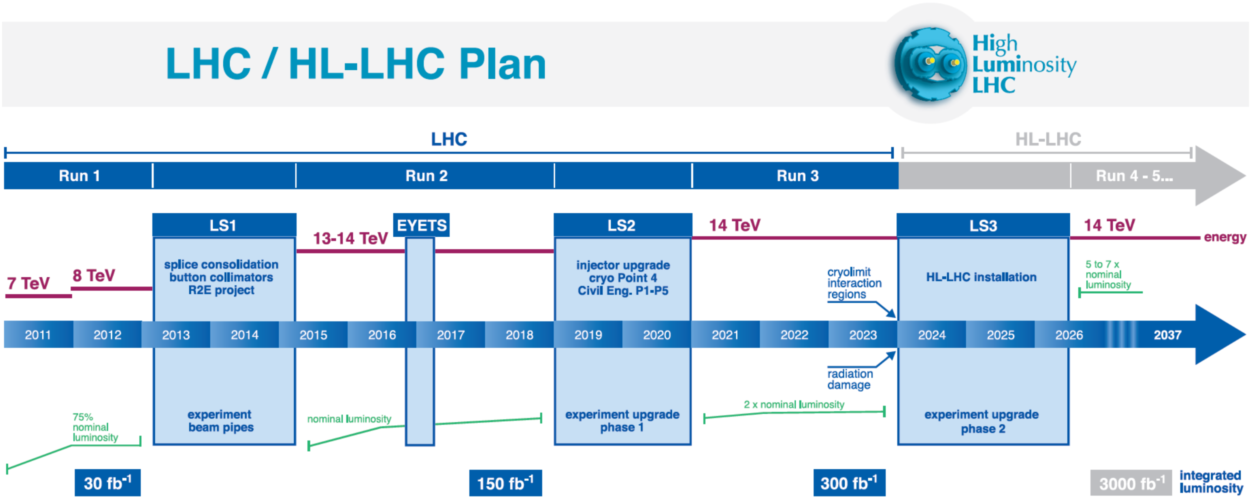



FIGURE 3.2: The roadmap sketch for LHC programme up to the high-luminosity LHC, or called HL-LHC [42].

At the present design luminosity, the collision produces $<\mu>\approx 25$ interactions per bunch crossing. Given a hard scattering process of interest which typically triggers the events, the multiple additional interactions from other protons in the same or nearby bunches are called pile-up[1] interactions, and are recorded together with the interaction of interest by the detector. There are two different types of pile-up: the in-time pile-up occurs when additional collisions happened in the same bunch-crossing containing the collision of interest; and the out-of-time pile-up, occurs when additional proton-proton collisions happened in the bunch crossing before or after the one containing the collision of interest. As the luminosity increases, the probability of multiple proton-proton interactions increases in a single bunch crossing. Therefore, it is important to mitigate this pile-up effect by improving the identification and reconstruction of a single primary collision, where the physics event of interest occurs. At the end of Run-II deliverance, the CMS detector recorded a dataset corresponding to an integrated luminosity of 150.26 $\text{fb}^{-1}$. The cumulative luminosity delivered by LHC and recorded by CMS for the Run-II data taking periods are shown in Fig. 3.3.

[1]These interactions are usually softer in the transferred momentum between colliding particles; these contribution "pile-up" on top of hard interaction event of interest.

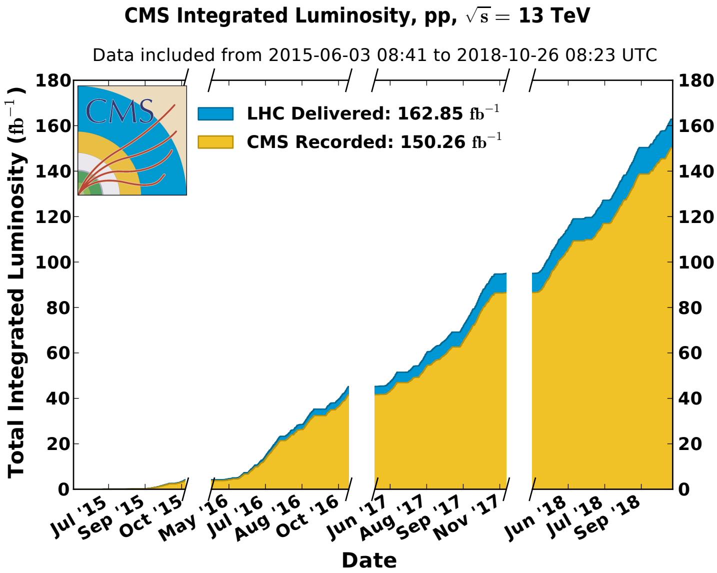



FIGURE 3.3: The cumulative delivered and recorded luminosity versus time for 2015-2018 (pp data only)

## 3.2 The CMS experiment

The CMS apparatus is a general purpose detector designed to probe a vast topology of physics processes, from the search of the Higgs bosons, to the SM measurements and BSM physics searches. The detector is employed to detect, identify and reconstruct almost all the physics objects produced in the pp collisions. The typical physics objects constituting an event are electrons, muons, photons and jets. In contrast to ATLAS's design, the CMS detector is designed to cater a single superconducting solenoid magnet, which is capable to generate a uniform 3.8 T magnetic fields along the beam axis. The detector is layered with multiple cylindrical detecting components, coaxial with the beam direction (barrel region), and hermetically closed at both ends with detecting disks (endcap region).

### 3.2.1 The detector coordinate convention

The coordinate system used by CMS is a right-handed Cartesian system, with the origin at the center of the detector, in the nominal beams collision point. In addition, spherical coordinates $(r, \theta, \phi)$ are often used, where $\theta$ is the angle formed with the beam direction (assumed to be the Z axis) and $\phi$, the azimuthal angle in the plane transverse to the beam; in this plane, the x axis is oriental towards the center of the LHC collider. The particle momentum is described either in terms of its cartesian components $(p_x, p_y, p_z)$, or in term of the $(p_{\mathrm{T}}, \eta, \phi)$ components. The transverse momentum $p_{\mathrm{T}}$ is given by the projection of the momentum in the plane transverse to the beam direction:

$$p_T^2 = p_x^2 + p_y^2 \tag{3.4}$$

The pseudorapidity $\eta$ is given by:

$$\eta = -ln(tan\frac{\theta}{2}) = \frac{1}{2}ln(\frac{|p| + p_z}{|p| - p_z}) = tanh^{-1}\frac{p_z}{|p|} \tag{3.5}$$

The choice of using $\eta$ over $\theta$ is due to the transformation properties of their separation between two particles, $\Delta\eta = |\eta_1 - \eta_2|$ and $\Delta\theta = |\theta_1 - \theta_2|$. For massless particles, $\Delta\eta$ is invariant under Lorentz boost along the z-axis, which is not the case for $\Delta\theta$. This property is crucial in hadron collisions, as the momenta of the interacting partonic constituents are unknown on an event-by-event basis. From $\Delta\eta$ and $\Delta\phi$, a distance parameter between two objects:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.6}$$

is defined as a measure of how separated the trajectories of two objects are. Another quantity that is commonly encountered is the rapidity $y$:

$$y = \frac{1}{2}ln(\frac{E + p_z}{E - p_z}) \tag{3.7}$$

As implied, $y$ reduces to $\eta$ for case of massless particles. $y$ is also invariant under Lorentz boost along the z-axis, but unlike $\Delta\eta$ this invariance holds also for massive particles.

## 3.3 The CMS detecor

The schematic view of the CMS detector, which has a length of 21.5 m, a diameter of 15 m and a weight of about 14000 tons, is shown in Fig. 3.4.

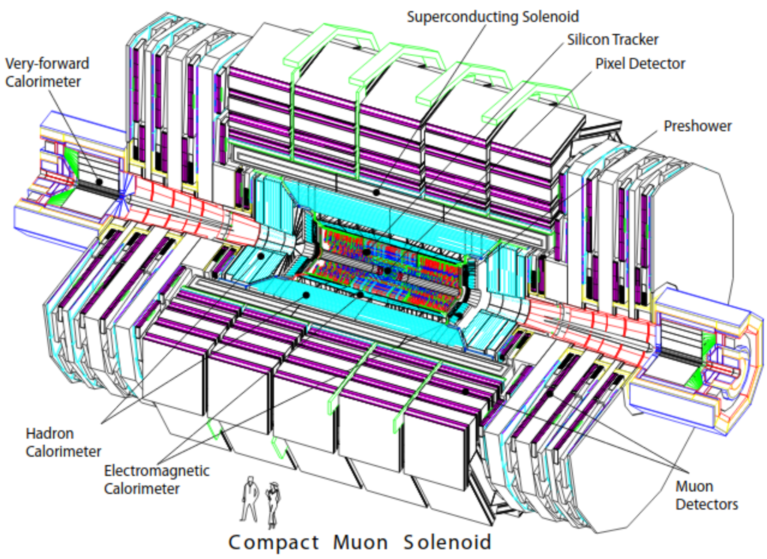



FIGURE 3.4: A perspective view of the CMS detector. [39]

From the inner region to the outer region, the various CMS sub-detectors and components are:

- **Silicon tracker**, consisting of an inner silicon pixel vertex detector and an array of silicon microstrip detectors with total active area of about 215 $m^2$, and situated in region of $|\eta| < 2.5$ and radial distance of $r < 1.2$ m. It is used to reconstruct charged particle tracks and vertices;

- **Electromegnetic Calorimeter (ECAL)**, is constructed with scintillating crystals of lead tungstate ($PbWO_4$), and occupies the region defined by $|\eta| < 3$ and 1.2 m $< r <$ 1.8 m. The ECAL delivers the measurement of the trajectory and energy deposited by the passing electrons and photons;

- **Hadronic Calorimeter (HCAL)**, is made up of brass layers alternated with plastic scintillators and it is placed in the region $|\eta| < 5$ and 1.8 m $< r <$ 2.9 m. It is used to measure the direction and energy doposited by the passing hadrons;

- **Superconducting solenoid magnet**, it generates an uniform magnetic field in its internal region necessary to bend the trajectory of charged particles. The magnetic lines are closed inside 21 m long iron yoke, with a diameter of 14 m. In this return yoke, a magnetic field of approximately 1.2 - 1.8 T is present, pointing at the opposite direction with respect to the internal field. The magnet is placed in the region $|\eta| < 1.5$ and 2.9 m $< r <$ 3.8 m. The measurement of charged particle momentum is allowed through the curvature observed in the tracking system;

- **Muon system**, is the outermost detection system, accommodated in the region $|\eta| <$ 2.4 and 4 m $< r <$ 7.4 m. It consists of Drift Tubes Chambers (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and endcaps, mainly for trigger purposes. The muon chambers are housed inside the iron structure of the return yoke. It is installed to reconstructing the tracks of transversing muons;

The proton-proton collision occurs in the beam pipe region, which is immediately encapsulated by the tracking system. The collision produces charged particles, which begin to transverse outward leaving signal or hits in the silicon tracker. The collection of hits traces the charged particle trajectories, reconstructed as tracks. Eventually, the interaction point where the physics interaction occurred is back-tracked from the collection of tracks, reconstructed as primary vertex.

Electrons and photons are absorbed by the ECAL, where their electromagnetic showers are detected in neighboring cells, from which the energy and direction of the particle can be determined. Charged and neutral hadrons may initiate a hadronic shower in the ECAL as well,which is subsequently fully absorbed in the HCAL. The corresponding clusters are used to estimate their energies and directions. Muons and neutrinos traverse the calorimeters with little or no interactions. While neutrinos escape undetected, muons produce hits in additional tracking layers inside the muon detectors, located outside the calorimeters. In Fig. 3.5 shows the sketch of different responses of the various CMS sub-detectors to the passage of different types of particles.

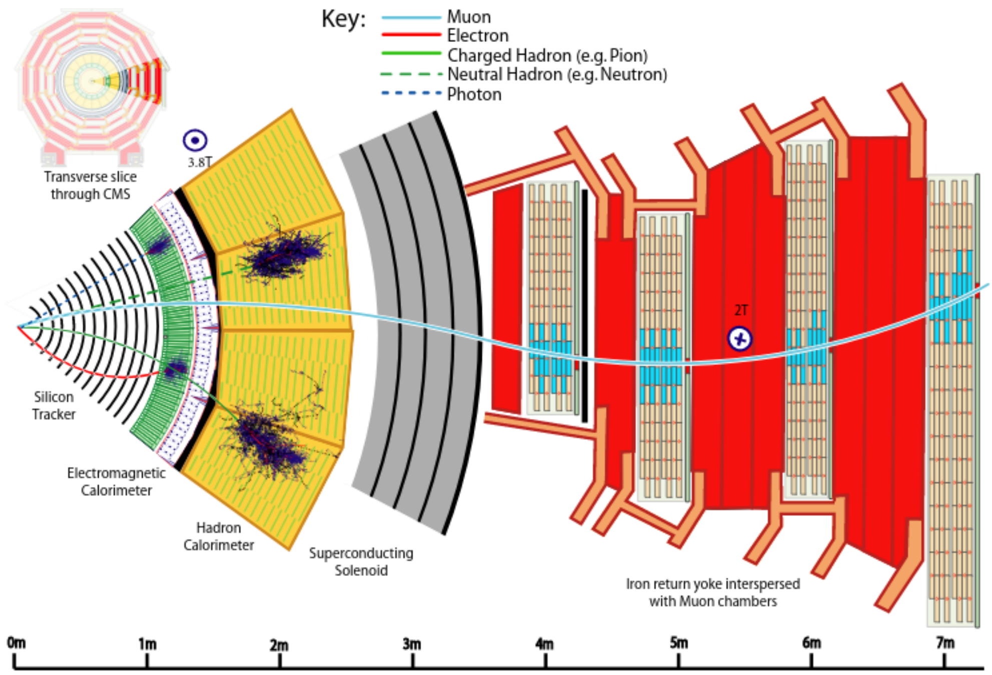



FIGURE 3.5: Sketch of a transverse slide of the CMS detector, showing the different types of particle interactions, from the beam interaction region to the muon detector. The muon and the charged pion shown are positively charged, while the electron is negatively charged. [43]

### 3.3.1 The solenoidal magnet

The CMS magnet [44] is the largest superconducting solenoid ever constructed in the world, with a diameter of 6 m and a length of 12.5 m, providing the bending power to the charged particle trajectories. The magnet houses the tracker, ECAL and the HCAL in the internal bore. In order to achieve momentum resolution of $\Delta p/p \sim 0.1$ at 1 TeV, the magnet needed to generate 3.8 T magnetic field to the sub-detector systems. The energy stored in the magnet is about 2.7 GJ at full current. The superconductor is made of 4 Niobium-Titanium layers and it is cooled down to about 4 K through a liquid Helium cooling plant. In the case of a quench, when the magnet loses its superconducting property, the energy is dumped to resistors within 200 ms. The solenoid is surrounded by iron yokes which serve as both the support structure for the detector, and to prevent leakage to the outside environment. As the muon chambers are interspersed between these yokes, muon momentum measurements benefit from the second bending provided by the returned field.

### 3.3.2 The tracker

The tracker [45] consists of different sub-detector modules, placed strategically in close proximity to the beams collision point. The primary goal is to achieve high resolution reconstruction of charged particle's trajectories produced from the interaction point, and to identify the secondary vertices position from particles with short life time. Track recognition and reconstruction are a taunting task[2] in CMS; the tracker must have the properties such as low occupancy and large redundancy of the measured points (hits) per track to meet the challenges. The former requirement is achieved by building a detector with high granularity[3]; the latter is achieved by having several detecting layers, which is necessary to reduce the ambiguity on the hits assignment for a given track.

In the central $\eta$ region, the pixel tracker consists of 3 co-axial barrel layers at radii between 4.4 cm and 10.2 cm. On the other hand, the strip trackers consists of 10 co-axial barrel layers extending outwards to radius of 110 cm. Both sub-detectors are enclosed by endcap disks on 2 side of the barrel, 2 disks in the pixal tracker, and 3 small + 9 large disks in the strip tracker, extending the acceptance of the tracker up to $|\eta| < 2.5$. The pictorial representation of a slide of the tracker is shown in Fig. 3.6, outlining the orientation of silicon strip detectors and silicon pixel detectors instrumented in the tracking system.

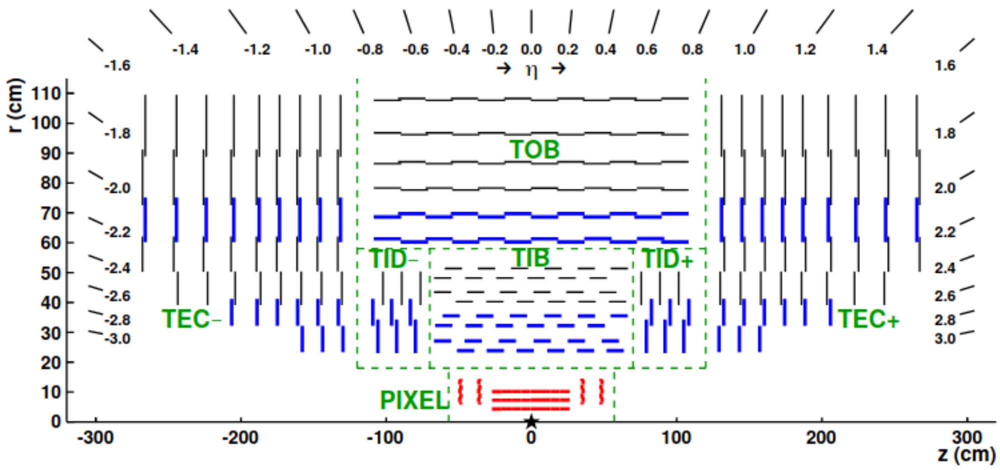



FIGURE 3.6: The CMS tracking system, showing the location and labeling of each components.

The pixel tracker was upgraded during the technical stop at the end of 2016, and the new configuration was used during 2017 and 2018. The "Phase-I" version of the pixel tracker barrel has 4 layers of pixels at radii between 2.9 cm and 16 cm, and 3 endcap disks [46]. The total silicon area of the CMS Phase-1 pixel detector is 1.9 $m^2$, while the total silicon area of the original pixel detector was 1.1 $m^2$, thus improving the pattern recognition and track reconstruction [47].

[2]at nominal $\mathscr{L}_{inst}$, an average of 20 pile-up events overlapping events of interest, leading to about 1000 tracks to be reconstructed for every 25 ns.

[3]The granularity of a detector is defined as the angular range ($\Delta\eta \times \Delta\phi$) that each individual element is resolved.

The tracking system has a cylindrical shape with length of 5.8 m and a diameter of 2.5 m, with the axis aligned to the beams direction. The average number of hits per track is 12-14, allowing high reconstruction efficiency and low rate of fake tracks. The hit reconstruction resolution is about 10-15$\mu$m in the barrel and 15$\mu$m in the endcaps. The material budget of the tracker is $|\eta|$-dependent with a maximum of 2 at $|\eta| \sim 1.6$ expressed in units of radiation length $X_0$. A more detailed discussion on the tracker and its performance can be found in [45].

### 3.3.3 The electromagnetic calorimeter

The ECAL [48] main task is to identify electrons and photons, and measure their respective energy. The detector consists of two parts, the ECAL barrel (EB) and 2 endcaps (EE), both consist of 61200 and 7324 scintillating crystals of lead tungstate ($PbWO_4$). The properties of $PbWO_4$ crystals made them a suitable candidate for detection during LHC operation. The high density ($\rho = 8.3\ g/cm^3$), short radiation length ($X_o = 0.89$ cm) and small Moliere radius (2.2 cm) are desirable properties to construct a compact and high granularity calorimeter. Another advantage is the radiation hardness and fast scintillation decay time ($\tau = 10$ ns), which permits about 80 % collection of the produced light within 25 ns interval between two consecutive bunch crossings.

The crystals are group into 5 × 5 matrices called towers. The barrel has an inner radius of 129 cm, a length of 630 cm and extends in the region $|\eta| < 1.479$. The crystal's dimension in the barrel is 22 × 22 $mm^2$ at the front face, 26 × 26 $cm^2$ at the rear face, and length 25.8 $X_0$, are constructed to provide a granularity of 0.0174 × 0.0174 in $\eta - \phi$ space. On the other hand, the crystals in the endcap, of dimension 28.6 × 28.6 $mm^2$ at the frontal area, and rear surface of 30 × 30 $mm^2$ are arranged in an x-y grid with coverage $|\eta| \in [1.479,3.0]$. Avalanche Photodiodes (APDs) are used as photo-detectors connected with the crystals in the barrel region and endcap region. In the range of $|\eta| \in [1.653,2.6]$, a preshower system (ES) is installed in front of EE to aid in the identification of neutral pions, photons and electrons. A slide representation of the ECAL detector is shown in Fig. 3.7.

The energy resolution of a homogeneous calorimeter is expressed by the sum in quadrature of 3 terms:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \tag{3.8}$$

where $a$ represents the stochastic term , $b$ is the noise term and $c$ is a the constant term which will be shortly discussed below. At low energies, the stochastic term $a$ dominates. It take into account the contribution of statistical fluctuations in the number of photo-electron collection, the amount of crystal's light emission, the light collection efficiency and photo-detector quantum efficiency[4]. The noise term $b$ includes the pile-up events contribution and

[4]The quantum efficiency is the ratio between the number of collected photo-electrons and the number of photons incident on the photo-detector.

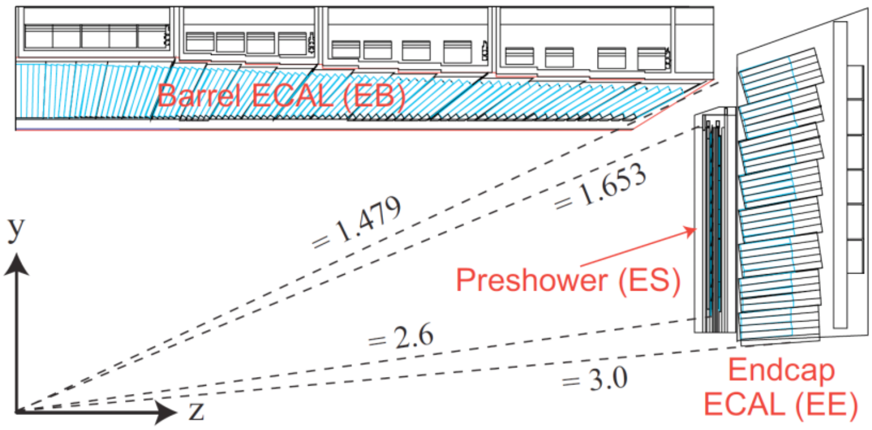



FIGURE 3.7: The ECAL in CMS detection system, the schematic diagram shows a quadrant of ECAL with $\eta$ coverage of EB, EE and ES.

electronic noise[5].The constant term $c$, is dominant at high energies. This term accounts for contributions such as non-uniformity longitudinal light collection, inter-calibration errors and the leakage of energy from the rear side of the crystal.

The EB resolution for electrons was measured using test beams to be:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%\ \text{GeV}^{1/2}}{\sqrt{E}}\right)^2 + \left(\frac{12\%\ \text{GeV}}{E}\right)^2 + (0.3\%)^2 \tag{3.9}$$

where the energy $E$ is expressed in GeV.

### 3.3.4 The hadronic calorimeter

The HCAL [49] is complementing the ECAL in jet energy and direction measurement. Also, the HCAL hermetic structure allows the measurement of energy imbalance in the transverse plane, $E_{\mathrm{T}}^{\mathrm{miss}}$, a typical signature of non-interacting particle such as neutrino. The HCAL is a sampling calorimeter comprising of 4 sub-detectors, and covering the region $|\eta| < 5$: the HB (Barrel Hadronic Calorimeter), is located in the barrel region inside the solenoid ($|\eta| < 1.4$); HE (Endcap Hadronic Calorimeter), is placed in the endcap region inside the magnet ($1.3 < |\eta| < 3$); HO (Outer Hadronic Calorimeter), is placed along the inner wall of the magnetic field return yoke; HF (Forward Hadronic Calorimeter), placed in the very forward region outside of the magnetic coil ($3 < |\eta| < 5$). The longitudinal view of the HCAL is shown in Fig. 3.8

Since HCAL is mostly installed inside the magnetic coil, it is crucial to use non-magnetic material as an absorber. The brass absorber layers with 5 cm are interleaved with 3.7 mm

[5] Both due to the photo-detector and pre-amplifier. These contributions depend on $\eta$ and the LHC operational luminosity.

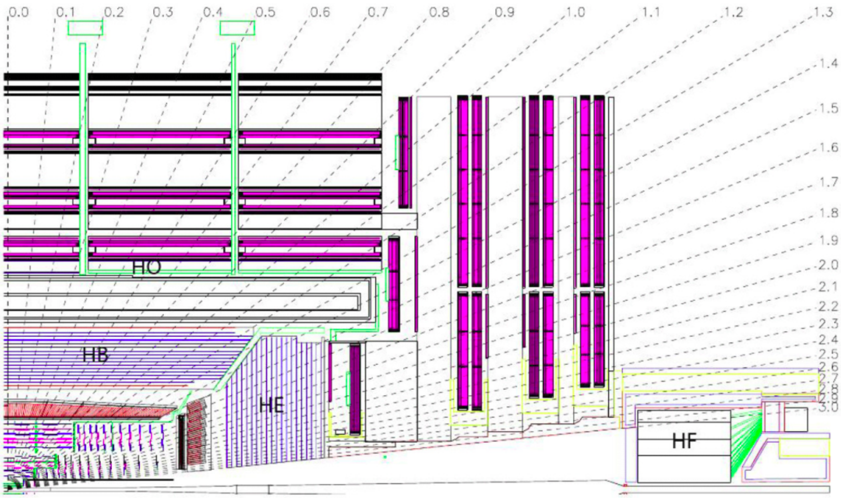



FIGURE 3.8: Longitudinal view of the CMS detector, with HCAL sub-detector components.

plastic scintillators in HB and HE. The scintillation light is collected by WaveLength Shifting (WLS) fibers and read out by Hybrid Photo-Diodes (HPD). The granularity of the calorimeter is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi \approx 0.17 \times 0.17$ for $|\eta| \geq 1.6$. HO is made of 5 rings installed in the wheel that compose the return yoke and is divided in 12 sectors. It consists of scintillating layers, with the same granularity as HB. The solenoid coil is used as an additional absorber to increase the effective depth of the calorimeter in the barrel region.

The energy resolution in the different regions of HCAL is parameterized by stochastic and constant terms, given in barrel and endcap:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{90\%\ \mathrm{GeV}^{1/2}}{\sqrt{E}}\right)^2 + (4.5\%)^2 \tag{3.10}$$

and in HF:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{172\%\ \mathrm{GeV}^{1/2}}{\sqrt{E}}\right)^2 + (9\%)^2 \tag{3.11}$$

where the energy $E$ is expressed in GeV.

### 3.3.5 The muon system

The muon system [50] is tasked to identify and measure high $p_\mathrm{T}$ muons, by combining information with the tracker. The muon system is installed outside the magnetic coil, and embedded in the return yoke to fully exploit the 1.8 T return flux. There are 3 different sub-detectors making up the muon system:

- DT: installed in the barrel region ($|\eta| < 1.2$), where the occupancy is relatively low ($< 10\ Hz/m^2$);

- CSC: which are installed in the endcaps, where occupancy is high (> 100 $Hz/m^2$) due to the small angle made with respect to the beam intensity;
- RPC: are placed both in the barrel and endcaps, complementing DT and CSC systems to ensure robustness and redundancy to the muon spectrometer.

The geometrical disposition of the CMS muon system is shown in Fig. 3.9.

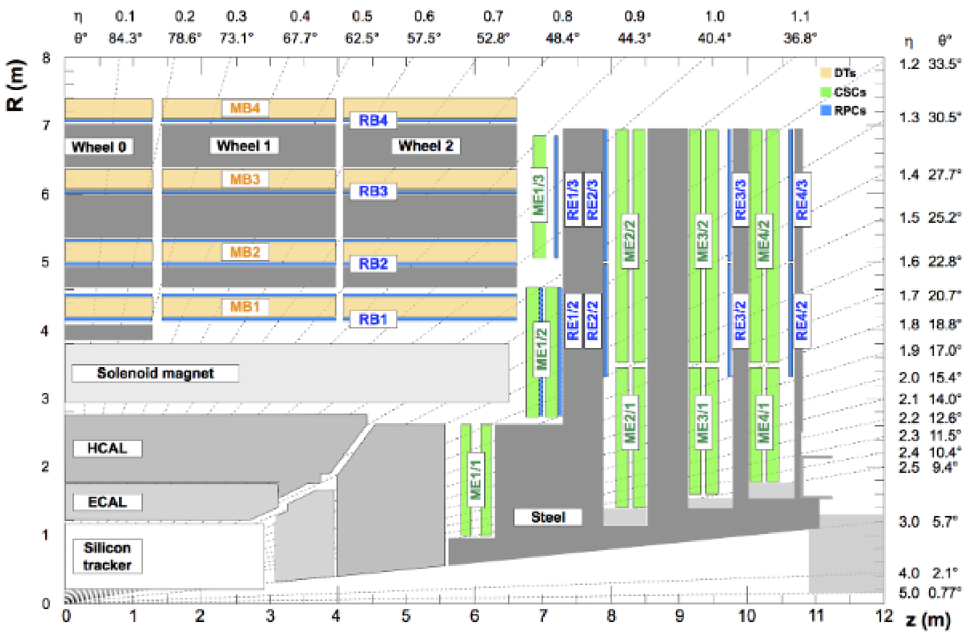



FIGURE 3.9: Schematic diagram of a quadrant of muon system.

Along the longitudinal direction, the barrel region is divided in 5 wheels, each wheel is sub-divided into 12 sectors covering $30^0$ azimuthal angle each. The wheels are composed of 4 concentric rings of chambers, called stations, and interspersed in the layers of the iron return yoke. Each station houses 12 DT chambers both in MB (Muon Barrel) and RB (Rear Barrel), in such a way that a high $p_T$ muon crosses at least three stations. The DT system is a rectangular drift tube cell, filled with gas mixture of $Ar$ (85%) and $CO_2$ (15%) and bundled together to form detection layer. Group of 4 layers form a superlayer. Each chamber consists of two superlayers with anode wires parallel to the beam axis and one perpendicular to it, providing two measurements of the $r-\phi$ coordinate and one measurement of the $z$ coordinate of the track hit position. The hit resolution is about 100 $\mu$m in both $r-\phi$ and $z$ direction.

In the endcap region ($|\eta| \in [1.2, 2.4]$), CSC is installed to cope with high occupancy as the sub-detector has faster response time. As in the MB region, there are four stations labeled ME (Muon Endcap) 1-4, overlapping in $\phi$ to prevent any gaps in acceptance. The CSCs are based on the multi-wire proportional counters [51], filled with gas mixture of $Ar$ (40%), $CO_2$ (50%) and $CF_4$ (10%), equipped with cathode strip readout with respect to the anode wires, enabling accurate hit position measurements radially and longitudinally. The typical spatial resolution is about 80 - 85 $\mu$m.

The RPCs are gaseous detectors characterized by a coarse spatial resolution, but able to perform precise time measurements. These chambers are made of 4 bakelite planes, with 2 mm gap between the plates filled with mixture of $C_2H_2F_4$ (94.5%) and Isobutane. The central part of the chamber is equipped with insulated aluminum strips, used to collect the signal generated by transversing particles. In the barrel, the strips are rectangularly segmented and run along the beam axis, whereas the endcaps are equipped with trapezoidal shaped strips.

A more detailed discussion on the muon chamber and identification performance can be found in Reference [50].

## 3.4 The trigger and data acquisition system

Based on the current instantaneous luminosity of LHC of the order of $10^{34}cm^{-2}s^{-1}$, given the p-p inelastic collision cross section $\sigma_{in} \approx 10$ mb, the LHC would deliver collisions at a rate about 0.1 GHz:

$$\mathscr{L}_{inst} \times 10 \text{ mb} = 10^8 \text{ Hz} = 0.1 \text{ GHz} \tag{3.12}$$

which implies 1 billion events per second. An event from LHC collision takes up about 1 $MB$ of storage space, therefore it is impossible to write all 1 billion events to disk for every second. By the benchmark value of 200 $MB/s$ the current electronics are capable of recording to disk, the 1 billion events must be reduced to about 200 events to be recorded and analyzed; this reduction is done through the triggering system, organized in two different levels.

The events filtering process is undertaken by the Level-1 (L1) trigger [52], based on hardware devices which reduce the event rate from the 40 MHz bunch crossing rate to 100 kHz. The L1 is instrumented with custom hardware installed physically with the detector to meet the challenging demand [53]. The L1 trigger decision has to be made for each bunch crossing within 3.2 $\mu s$.

The High Level Trigger (HLT) [54] is the second step in event filtering chain which further reduces the L1 output rate to 1 kHz for storage purpose. The HLT is a set of software algorithm running on a farm of commercial processors.

An event is recorded to mass storage if it passes L1 and HLT trigger. Given the extreme high rate of data readout from the detector, a robust and reliable online operation of Data Acquisition System (DAQ) is needed. The CMS DAQ system is shown in Fig. 3.10.

Event fragments are read out and stored in Readout Unit (RU) for each event accepted by the L1 trigger. The fragments are subsequently assembled into complete events by an Event Builder (EB) through a complex of switched networks into Event Buffer. The full event content is then handed to one of the HLT Filter Unit (FU). The FU execute a series of physics reconstruction and filter algorithms and events that are found to be sufficiently interesting for offline analysis are forwarded to the storage manager. The decoupling of physics algorithm execution from data flow allows each FU to continue operation in the case of a problematic event, recover the content of the problematic event, and forward it to be stored unprocessed.

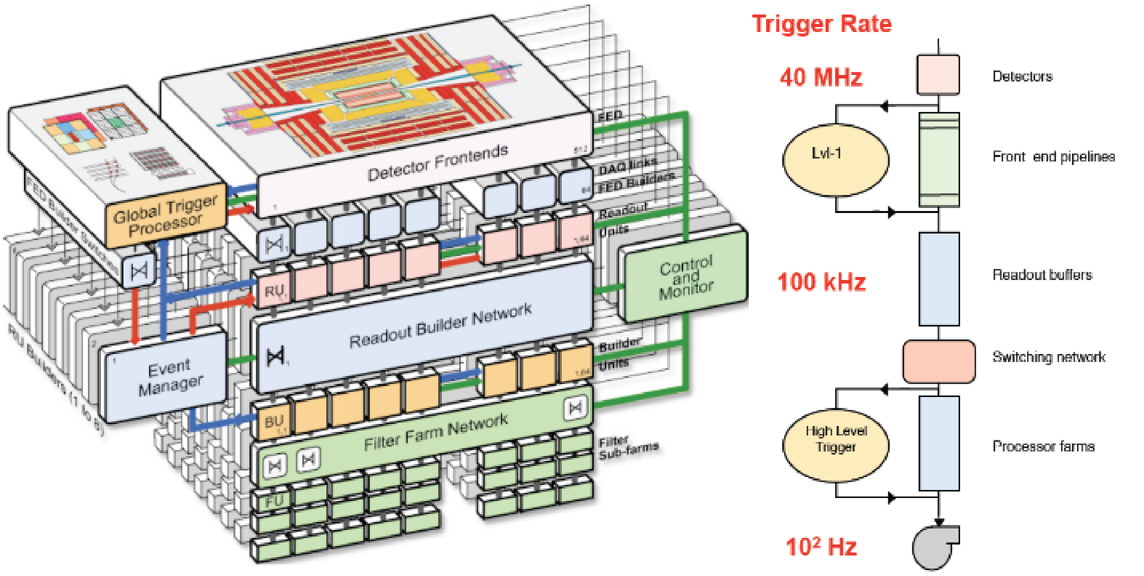



FIGURE 3.10: CMS DAQ architecture. The size of the event builder (72 RUs, 288 Builder Units) represents one slice; the system can be equipped with up to eight slices [55].

### 3.4.1 Level-1 trigger

The CMS L1 trigger system is constituted by 3 main sub-systems, there are:

- L1 Calorimeter Trigger
- L1 Muon Trigger
- Global Trigger

The Trigger Primitives (TP) from calorimetries (ECAL and HCAL), and from the muon detectors (DT,CSC,RPC) are processed in several steps before the combined event information is evaluated in the L1 Global Trigger and a decision is made whether to accept the event or not. The CMS L1 trigger schematic is summarized in Fig. 3.11.

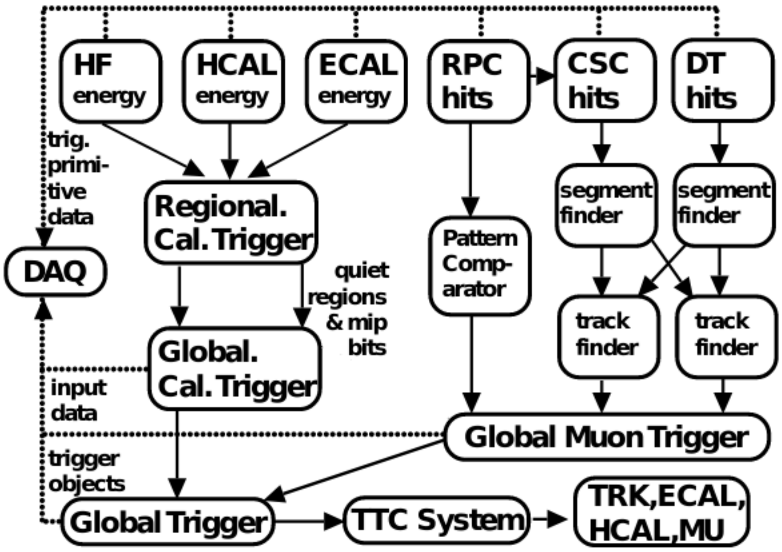



FIGURE 3.11: The overview of CMS L1 trigger system [52].

#### 3.4.1.1 The level-1 calorimeter trigger

The L1 Calorimeter Trigger are calorimeter towers, clustering signals collected from ECAL and HCAL. The signal is processed by the Regional Calorimeter Trigger, where $e/\gamma$, and transverse energy $E_{\mathrm{T}}$ candidates are identified. The Global Calorimeter Trigger sorting the $e/\gamma$ candidates further, finding jets (classified as central, forward, and tau) using the $E_{\mathrm{T}}$ sums, and calculating global quantities such as $E_{\mathrm{T}}^{\mathrm{miss}}$, which later send the combined information to the Global Trigger.

#### 3.4.1.2 The level-1 muon trigger

The L1 Muon Trigger comprises of RPC, CSC and DT sub-system. The trigger information from them are combined in the L1 Global Muon Trigger.

The RPC triggers provide information such as Track Segments and CSC logics to the Global Muon Trigger. Those information is vital for $p_{\mathrm{T}}$ estimation and resolves hit position ambiguities in case of two or more muon tracks crossing the same CSC chamber.

The CSC trigger build Local Charged Tracks (LCT), a track segments made from the cathode strip only, with $p_{\mathrm{T}}$ and quality flag assigned. The best three LCTs in each sector of nine CSC chambers are passed to the CSC Track Finder, which uses the full CSC information to build tracks, assigns them a $p_{\mathrm{T}}$ and a quality flag and sends them to the Global Muon Trigger.

The DTs find a groups of aligned hits in the four chambers of a superlayer, and try to combine it with other from the second superlayer in order to measure the $\phi$ angle. The sub-detector on-board DT Track Correlator is responsible for this task, and later the best 2 segments are sent to the DT Track Finder, which builds tracks and sends them to the Global Muon Trigger.

The Global Muon Trigger sorts the RPC, CSC and DT muon tracks and tries to combine them. The final set of muons is sorted according to the quality, and the best four tracks are passed to the L1 Global Trigger.

### 3.4.2 High level trigger

The HLT is designed to further reduce the L1 trigger output to a manageable rate for mass storage. Unlike L1 trigger, HLT runs on commercial processors, and tasked to perform event reconstruction using information of all the sub-detectors within the timing budget of $\mathcal{O}(100ms)$. Additionally, the HLT selection is performed at reconstructed level where the events are fully reconstructed by the same software used in offline analysis[6]. Those events passing the HLT are stored in local disks, or in CMS Tier 0[7].

Technically a full list of cuts used by the HLT is compiled into a HLT menu, which consists of the order of 100 number of paths. A path can impose only a simple set of cuts, and a cross paths can impose more stringent criteria on correlations between different triggering objects on triggering.

[6] Although with preliminary values of detector calibration and alignment constants

[7] The Worldwide LHC Computing Grid (WLCG) is composed of four levels, or "Tiers", identified with numbers 0, 1, 2 and 3. Each Tier is made up of several computer centres and provides a specific set of services; they process, store and analyse all the data from the LHC. Tier 0 is the CERN Data Centre. All of the data from the LHC pass through this central hub. Tier 0 distributes the raw data and the reconstructed output to Tier 1's, and reprocesses data when the LHC is not running [56].

# Chapter 4

# Event reconstruction and simulation

In order to measure the physical process of interest and compare the observation with the SM predictions, it is crucial to have them as accurate as possible. This requires the development of a robust event generation framework, simulating the whole process from the initial collision between the constituents of the protons, down to the particles that interact with the detector.

The electronic signals that are read out from the detectors are not directly interpretable. They must be processed by reconstruction algorithms to reproduce the final state physical objects in the event. At the analysis level, these objects are used to select a region of the phase space of the relevant kinematical variables, in which the signature from a specific interaction process is enhanced. The event simulation and the processing of the experimental data requires a large computing infrastructure and software framework.

## 4.1 Event simulation

An event is a collision between the protons, resulting into a collection of out-going particles which eventually registered as electronic signal readout by CMS sub-detectors for every bunch crossing (25 ns). The simulation of such events in LHC involves multiple steps. The various physical processes at parton level is simulated by MC technique using software packages describing the underlying theoretical model, as for instance a Matrix Elements (ME) generator, followed by a parton shower generator describing the non-pertubative processes in which final state particles are generated, and finally by the detector simulation.

### 4.1.1 The proton-proton collision

The parton model describes how the proton is composed of its valence quarks[1] bathed in a sea of quarks and gluons. Additionally, the number of quarks that can be found within the proton is not a constant, as it depends on the scale of the momentum exchanged between the protons during the scattering process.

When protons collide in the LHC, they usually undergo inelastic scattering processes. The collision with the largest momentum transfer, $Q^2$, is referred to as the hard scattering event, and the cross section calculation is factorized into long-distance and short-distance components according to the QCD factorization theorem [57]. The scale that separates the hard (short-distance) and soft (long-distance) part of the process $pp \to X$ is called the factorization scale $\mu_F$. Moreover, a renormalization scale $\mu_R$ is introduced to deal with the divergences at high energy, or short distances. The cross section calculated to all orders is independent on these scales. For fixed order calculations, they are usually set to the hard scale[2] in the process, and the uncertainties due to the limited order are usually estimated from varying the scales by a factor 2.

The total inclusive cross section of the reaction is given by:

$$\sigma(pp \to X) = \sum_{i,j} \int dx_i dx_j f_i(x_i, \mu_F^2) f_j(x_j, \mu_F^2) \times \hat{\sigma}_{ij\to X}(x_i, x_j s, \mu_F^2, \mu_R^2) \tag{4.1}$$

where $x_i$, $x_j$ are the four-momentum fractions carried by the two interacting partons, and $f_{i,j}(x_{i,j}, \mu_F)$ is their associated Parton Distribution Function (PDF) which is determined from the fits to experimental measurements [58, 59], and expresses the probability of finding the parton with momentum fraction $x_i$ at some scale $\mu_F$. The sum runs over all initial state partons $i$ and $j$. The term $\hat{\sigma}_{ij\to X}$ is the partonic cross section for the final state $X$ that depends on the partonic centre-of-mass energy $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$, with $s$ being the proton-proton centre-of-mass energy.

### 4.1.2 The matrix element generator

The partonic cross section $\hat{\sigma}_{ij\to X}$ is calculated within perturbative QCD from all the contributing diagrams evaluated at a given order. The Leading Order (LO) accuracy refers to the set of diagrams with the lowest number of the relevant vertices that gives a non-zero contribution to the cross section. The Next-To-Leading-Order (NLO) and Next-to-Next-to-Leading-Order (NNLO) are then the set of diagrams that involve increasingly higher number of vertices. Examples of dedicated software packages are POWHEGV2 [60, 61, 62] and MADGRAPH5_aMC@NLO [63].

[1]The 3 valence quarks are two up-quark and one down-quark.

[2]Hard scale refers to the short distance interaction scale.

### 4.1.3 Parton showering and hadronization

For a realistic simulation of pp collision events, the ME generator has to be interfaced with a Parton Shower (PS) program. The PS performs the Initial State Radiation (ISR) and Final State Radiation (FSR) of the partons generated by the ME program, and the backward and forward evolution from them to the initial protons and the final state particles [64]. As the partonic evolution series contains terms that are of similar magnitude at each order, it is necessary that the entire series is considered without any truncation to a given finite order. A technique known as resummation is used for this purpose up to some finite logarithmic accuracy [65]; the lowest logarithmic accuracy is known as the leading logarithm (LL), followed by next-to-LL and so on. The PS tool that is used in this work is the one in PYTHIA8 [66], which also handles the treatment of the interactions of spectator partons in the proton, known as the Underlying Events (UE). Since the ME can also be evaluated for processes where additional partons such as ISR and FSR are produced, a scheme has to be defined to correctly assign the contributions as belonging to the ME or to the PS. This is achieved by defining a boundary in the phase space between the two, where the PS handles only the emission of additional partons below a certain momentum scale. This procedure is known as the merging scheme. Due to the different subtleties involved when virtual higher-order corrections are included, different merging schemes are used at LO [67] and NLO [68].

### 4.1.4 Detector simulation

In the CMS experiment, the interactions of particles with the detector material are simulated with GEANT4 [69]. The output of this step is a collection of time-stamped energy deposits for every sensor. After this, the digitization step follows, by emulating the electronics response. The simulated event output is subsequently processed with the same workflow as the data collected in the real experiment [70].

## 4.2 The track reconstruction

The good performance of the tracking system has a key role in the reconstruction of the trajectory of charged particles, and is achieved by means of an iterative tracking strategy [45], based on the Kalman Filter (KF) algorithm [71]. The main goal of the tracking system is to reconstruct the trajectory of a charge particle transversing the tracker devices with large efficiency and the lowest possible fake rate, providing information on momentum and position of the particle.

The main standard algorithm used by CMS is the Combinatorial Track Finder (CTF) [72], and it consists of three stages:

- track seeding: triplets of pixel hits are used as a tracking seeds; they are required to be above a set of minimum criteria to be considered as a triplet of hits (track) for a charged particle.
- pattern finding: starting with the seeded parameters, the track trajectory is extrapolated to the neighboring tracker layers and compatible hits are assigned to the track. At each iteration the track parameters are updated with new hits and then assigned a quality, and the best quality tracks are kept for further processing. Possible ambiguities among tracks which share common hits are resolved in favour of the best quality trajectories;
- track fitting: the trajectory is fitted using a least-squares fit in two stages. A first forward fit, inside-out from the interaction region, removes the approximations and biases of the track seeding and pattern finding stages. A second outside-in smoother fit yields the final best estimates of the track parameters ($p_T, \eta, \phi$) at the origin vertex.

For prompt charged particles with $p_T > 0.9$ GeV, the average reconstruction efficiency of 94% was measured in the barrel region of the tracker, and 85% in the endcap region; the inefficiency is mainly caused by nuclear interaction in the tracker materials, and by the falsely reconstructed tracks due to the secondary particles production, which are at the few percent level (1.5%) [45].

## 4.3 The particle flow technique

The Particle Flow (PF) event reconstruction technique [43] is used for the reconstruction and identification of all the stable particles in the event, such as electrons, muons, photons, charged and neutral hadrons. The technique makes use of the combination of information from all CMS sub-detectors to determine the energy, direction and type of the observed particles. Subsequently, the collection of particles are used as an input for building jets, measuring missing transverse energy $E_T^{miss}$, reconstructing $\tau$ leptons from their decay products, quantify charged lepton isolation and perform b-tagging.

The first step of the PF technique consists in the reconstruction of the basic elements from the various sub-detectors, such as charged-particle tracks, calorimeter clusters and muon tracks. These elements, which are provided by the sub-detectors with high efficiency and low fake rate, are then connected together with a link algorithm [43].

The high granularity of ECAL is the additional key element for the feasibility of the PF technique, allowing the reconstruction of photons and electrons with high energy resolution in a dense environment. The high detection efficiency of the calorimeters is based on a specific calorimeter clustering algorithm, which is performed separately in each sub-detector. The algorithm is based on three steps: in the first step "cluster seeds" are identified as local calorimeter cells with an energy deposit above a given threshold. Then "topological clusters" are grown from the seeds by gathering cells with at least one side in common with a cell already in the cluster, and with an energy above a given threshold. A topological cluster usually gives rise to as many "particle flow clusters" as its seeds. Thus the energy of each cell is shared among the particle flow clusters according to the cell-cluster distance.

Subsequently, the link algorithm connects all those PF elements to each other by identifying blocks of elements, called the PF blocks, that are topologically compatible. From these PF blocks, PF candidates are identified as follows:

- Electrons: Electrons are characterized as tracks stopping at the ECAL, having energy losses by bremsstrahlung in the tracker layers. The link between a charged-particle track (refitted with the Gaussian-Sum Filter (GSF) [73]) and one or more ECAL clusters identifies a PF electron. The detail of electrons reconstruction is given in Sec. 4.4.2. After the identification, the corresponding tracks and clusters are removed from the PF block;
- Muons: muons are reconstructed as described in Sec. 4.4.4. A "global" muon gives rise to a PF muon if its combined $p_{\mathrm{T}}$ measurement is compatible within 3 standard deviations with the one provided by the standalone tracker. The corresponding track is removed from the PF block;
- Charged hadrons: the remaining tracks give rise to PF charged hadrons. Tracks can be linked to ECAL and HCAL clusters, and the energy is determined taking into account information from calorimeters;
- Photons and neutral hadrons: ECAL clusters not linked with tracks give rise to PF photons, while the remaining HCAL clusters are identified as PF neutral hadrons.

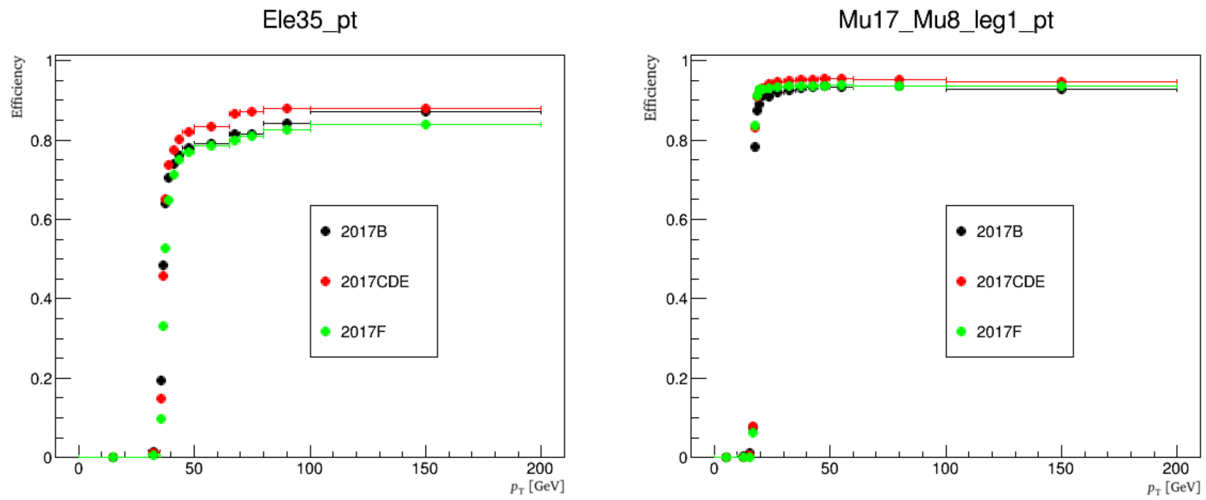
After the identification of all PF candidates in the event, PF jets are clustered as described in Sec. 4.5. The last step is the reconstruction of the missing transverse momentum PF $p_{\mathrm{T}}^{\mathrm{miss}}$, which is described in Sec. 4.7.

## 4.4 Lepton reconstruction and identification

### 4.4.1 Lepton trigger efficiency

Since the analysis involves two leptons in the final state, the interesting events are selected by using the single and double lepton triggers. The lepton triggers are characterized by $p_{\mathrm{T}}$ threshold above which the trigger efficiency reaches its maximum (plateau region). The trigger efficiency for single and double lepton triggers is calculated in bins of $\eta$ and $p_{\mathrm{T}}$ using a Tag and Probe technique as described in Sec. 4.4.8, separately for muons and electrons. For double lepton triggers, the efficiency is calculated separately for each leg of the trigger.

Fig. 4.1 shows the lepton trigger efficiencies, with their plateau reached above the respective $p_{\mathrm{T}}$ thresholds.




(A) Efficiency of the single electron trigger with $p_{\mathrm{T}}$ threshold of 35 GeV

(B) Efficiency of the double muon trigger with first leg $p_{\mathrm{T}}$ threshold of 17 GeV

FIGURE 4.1: The lepton trigger efficiencies in different data-taking periods, as a function of lepton $p_T$. The dataset used corresponds to 41.5 $\mathrm{fb}^{-1}$ collected in 2017.

Since the trigger efficiency of the lepton in term of $p_{\mathrm{T}}$ is not a step function but a steep sigmoid shaped function (turn-on region), thus the simulated samples need to be corrected to account for the trigger efficiency difference with the data. This can be achieved by including the HLT trigger in the event simulation or calculating the trigger efficiency in data and then apply it on the simulated events. The latter is motivated by the fact that the simulated HLT trigger information may not reflect exactly what will be used in data; the second approach is used in this analysis.

The efficiency for a di-lepton event to pass the single lepton trigger, $\epsilon_{S,2l}$ is given by either one of the two leptons passes the single lepton trigger, unless both fail to pass it:

$$\epsilon_{S,2l} = 1 - \left[\left(1 - \epsilon_{S,l1}\right) \cdot \left(1 - \epsilon_{S,l2}\right)\right] \tag{4.2}$$

where $\epsilon_{S,l1}$ and $\epsilon_{S,l2}$ are the efficiencies for leading and trailing lepton to pass the single lepton trigger. On the other hand, the efficiency of the double lepton triggers is given by the combined efficiency of the two trigger legs[3]; the efficiency for a di-lepton event to pass the double lepton triggers, $\epsilon_{D,2l}$ is given by:

$$\epsilon_{D,2l} = \epsilon^{leg1}_{D,l1} \cdot \epsilon^{leg2}_{D,l2} + \left(1 - \epsilon^{leg1}_{D,l1} \cdot \epsilon^{leg2}_{D,l2}\right) \cdot \epsilon^{leg2}_{D,l1} \cdot \epsilon^{leg1}_{D,l2} \tag{4.3}$$

[3]The two trigger legs are considered independent given the correlations are small.

where $\epsilon_{D,l1}^{leg1(2)}$ is the efficiency of the first lepton to pass the first(second) leg of the double lepton trigger, and $\epsilon_{D,l2}^{leg2(1)}$ is the efficiency of the second lepton to pass the second(first) leg of the double lepton trigger.

Hence, the combined trigger efficiency $\epsilon_{2l}$ used to reweight the di-lepton events passing either single or double lepton trigger in the simulation is given by:

$$\begin{aligned}\epsilon_{2l} = 1 - \Big[\big(1-\epsilon_{S,l1}\big)\cdot\big(1-\epsilon_{S,l2}\big)\Big] \\ + \big(1-\epsilon_{S,l1}\big)\cdot\big(1-\epsilon_{S,l2}\big) \\ \cdot\Big[\epsilon_{D,l1}^{leg1}\cdot\epsilon_{D,l2}^{leg2} + \Big(1-\epsilon_{D,l1}^{leg1}\cdot\epsilon_{D,l2}^{leg2}\Big)\cdot\epsilon_{D,l1}^{leg2}\cdot\epsilon_{D,l2}^{leg1}\Big]\end{aligned} \tag{4.4}$$

### 4.4.2 Electron and isolated photon reconstruction and identification

The Electrons are reconstructed by combining energy clusters in ECAL with the reconstructed tracks from the inner tracker. Due to the thickness of materials between the tracker and ECAL, the transversing electrons are emitting a non-negligible fraction of their energy via bremsstrahlung into photons before reaching ECAL. To determine the initial energy of the electron, the energy of the electron track and candidate bremsstrahlung photons are collected and grouped into a Superclusters (SC), reconstructed in a small window in $\eta$ and extended window in $\phi$ around the electron direction. The bremsstrahlung photons induce kinks in the electron track, which is not well described by the standard track reconstruction; instead, a more adaptive algorithm, GSF is used to perform the iterative tracking.

There are two types of seeding strategy in CMS experiment, namely the ECAL- and tracker-driven seeding [74]. The ECAL-driven seeding, which starts from the SC energy and position[4], is used to estimate the electron trajectory in the first layers of the tracker, selecting electron seeds from all the reconstructed track seeds; the tracker-based seeding, which relies on tracks that are reconstructed using the general algorithm for charged particles, extrapolated towards the ECAL and matched to the SC. The association of the cluster and the track is performed with inside-out and outside-in approaches depending on the electron $p_{\mathrm{T}}$. The ECAL-driven seeding is suited for reconstructing energetic electrons by taking advantage of the high granularity ECAL, while tracker-driven seeding is preferred for low energy electrons whose reconstructed trajectories are more effectively measured by the tracking system. The momentum resolution varies from 1.7 % to 4.5 %, as measured on electrons from Z boson decays using Run-I datasets [74].

The electron charge measurement is affected by bremsstrahlung photons, with the photons converting into an electron positron pair within the tracker volume, resulting in additional recorded hits which may be included in the electron track fit. The electron charge is estimated from three measurements:

[4] The estimation of the position performed by taking weighted mean of the crystal energies, with correction applied based on the shower depth [49].

- from the GSF track curvature,
- from the associated KF track curvature,
- from the difference in $\phi$ space between the SC and the first hit of the GSF track.

Isolated photon candidates are reconstructed only from calorimeter clusters whose shapes are compatible with those expected from a photon showers. Similar to electron reconstruction, dedicated efforts and strategies are undertaken to reduce contamination coming from photon conversion into electron-position pairs inside the tracker system. It is worth noting that the achieved energy resolution of photons is between 1 and 3% [75] providing a di-photon mass resolution of 1.1 GeV at the mass of 125.09 GeV, making the $H \to \gamma\gamma$ decay one of the golden decay channels for the Higgs boson discovery [2].

The electron Identification (ID) selection is driven by a set of variables providing discriminating power to identify prompt electrons from background sources. In order to achieve a good discrimination, several variables are used:

- $\Delta\eta(SC, track)$ and $\Delta\phi(SC, track)$: the spatial matching between the track and the SC in the $\eta$ and $\phi$ coordinates, respectively;
- $|1/E - 1/p|$: the difference of the inverse of energy $E$ measured in ECAL and the inverse of momentum $p$ measured in the tracker;
- $\sigma_\eta$: variable related to the calorimeter shower shape, measuring the width of the ECAL SC along the $\eta$ direction, computed from all the crystals in the $5 \times 5$ block of crystals, centered on the highest energy crystal of the SC seed [74];
- $H/E_{SC}$: the ratio between the energy deposited in the HCAL tower behind the ECAL seed and the SC seed energy;
- the number of missing hits in the back-propagation of the track to the interaction point;
- a photon conversion veto ($\gamma \to e^+e^-$) based on missing hits in the inner layers of the tracker;
- $d_{xy}$ and $d_z$: the transverse and longitudinal impact parameters with respect to the primary vertex;
- Isolation (ISO) variables: mis-identified jets or heavy-flavor quarks decaying into electrons have significant energy flow near their trajectory; a tight isolation selection will thus improve the signal purity.

A multiple set of selections cutting on the above ID variables is defined to target a particular signal efficiency and purity. The given set of selection is called a Working Point (WP), implemented either by a cut-based approach or a Multivariate Analysis (MVA)-based approach, which generally provides a better discriminating power compared to the former. Different WPs serve different purposes, depending on the signal purity of interest in a given study; they are:

- **Tight working point**: corresponds to a 70% selection efficiency for electrons with $p_T$ > 20 GeV, which is used in analyses which need to reject large yield backgrounds;
- **Medium working point**: corresponds to a 80% selection efficiency for electrons with $p_T > 20$ GeV. Usually this WP is selected for $W$ and $Z$ cross-section measurement studies;
- **Loose working point**: corresponds to a 90% selection efficiency. This WP is recommended for study involving final state with negligible hadronic backgrounds;
- **Veto working point**: corresponds to a 95% selection efficiency. This is not recommended for signal selections. However this WP is useful for extra lepton veto when counting electrons in the event.

During the 2016 data taking, the loose and tight WPs are defined by the so-called "HLT-safe", cut-based selection given in Tab. 4.1. This selection is designed to be insensitive to the identification and detector-based isolation requirements imposed by the online triggers. During the 2017 and 2018 run periods, the medium WP defined in Tab. 4.2 is used in the analysis.

TABLE 4.1: The loose and tight cut-based, HLT-safe electron ID selection for 2016 data taking, in barrel and endcap regions.

| ID variables | Barrel | | Endcap | |
|---|---|---|---|---|
| | Loose | Tight | Loose | Tight |
| $\Delta\eta$(SC,track) | $< 0.004$ | $< 0.00308$ | - | $< 0.00605$ |
| $\Delta\phi$(SC,track) | $< 0.020$ | $< 0.0816$ | - | $< 0.0394$ |
| $\|1 \setminus E_{SC} - 1 \setminus p\|$ | $< 0.013$ | $< 0.0129$ | $< 0.013$ | $< 0.0129$ |
| $\sigma_\eta$ | $< 0.011$ | $< 0.00998$ | $< 0.031$ | $< 0.0292$ |
| $H \setminus E_{SC}$ | $< 0.060$ | $< 0.0414$ | $< 0.065$ | $< 0.0641$ |
| Passing conversion veto | - | True | - | True |
| GSF track $\chi^2/dof$ | - | - | $< 3.0$ | - |
| Expected missing inner hits | - | $\leq 1$ | - | $\leq 1$ |
| $d_{xy}$(track,vertex) cm | - | $< 0.05$ | - | $< 0.1$ |
| $d_z$(track,vertex) cm | - | $< 0.10$ | - | $< 0.2$ |

The alternative MVA-based approach employs a single discriminator variable which is used to define a selection targeting a desired signal efficiency. The MVA variable is computed

TABLE 4.2: The medium cut-based electron ID selection for run periods of 2017 and 2018, in barrel and endcap regions.

| ID variables | Barrel | Endcap |
|---|---|---|
| $\Delta\eta$(SC,track) | $< 0.00365$ | $< 0.00625$ |
| $\Delta\phi$(SC,track) | $< 0.0588$ | $< 0.0355$ |
| $\|\|1 \setminus E_{SC} - 1 \setminus p\|$ | $< 0.0327$ | $< 0.0335$ |
| $\sigma_{\eta}$ | $< 0.0105$ | $< 0.0309$ |
| $H \setminus E_{SC}$ | $< 0.026$ | $< 0.026$ |
| Passing conversion veto | True | True |
| Expected missing inner hits | $\leq 1$ | $\leq 1$ |
| $d_{xy}$(track,vertex) cm | $< 0.05$ | $< 0.1$ |
| $d_{z}$(track,vertex) cm | $< 0.1$ | $< 0.2$ |

based on multiple variables of the electron object and provides the best separation between the signal and backgrounds using the Boosted Decision Tree (BDT) technique [76]. The BDT training is performed using electrons from Z decays, with prompt electrons as signal and non-prompt electrons as background[5]. Conventionally, the three WPs defined for MVA-based ID selection according to the BDT output values, are MVA-loose ID, MVA-90 ID and MVA-80 ID, corresponding to 98%, 90% and 80% signal efficiency respectively.

After the selection described above, in the region defined for the signal extraction in a given analysis there may be still a significant background retained for identified prompt electrons, due to mis-identified jets or non-prompt electrons originating from semi-leptonic decays of bottom or charm quarks. In both cases, the electron candidates are surrounded by a significant energy flow along their trajectory, and isolating them from such activity will greatly reduce these sources of background. The ISO selection variable using particle-based variables relative to the electron $p_{\mathrm{T}}$ ($p_T^e$), is defined as:

$$I_{\Delta R}^{rel} = \frac{1}{p_T^e}\left[\sum_{ChH} p_T + \max\left(0, \sum_{\gamma^0} p_T + \sum_{NH} p_T - \rho A_{eff}\right)\right] \tag{4.5}$$

where the sums run over the charged hadrons (ChH), photons ($\gamma^0$), and neutral hadrons (NH) reconstructed by the PF algorithm within the cone radial distance $\Delta R$ (Eq. (3.6)) of 0.3 or 0.4 around the trajectory of the electrons; $\rho$ is the energy density due to pile-up events, scaled by the effective area $A_{eff}$ of all components of the isolation computation. This $A_{eff}$ quantity is defined by the sum over all the components of the quantity $\Delta R^2 = \Delta\eta^2 + \Delta\phi^2$, where the pseudorapidity and the azimuthal distances are computed with respect to the electron track. If the hadronic contribution inside the cone is large, the electrons are likely produced from the decays of heavy fermion, yielding a high $I_{\Delta R}^{rel}$ value.

[5] Reconstructed electrons from $\tau$ decays are ignored.

### 4.4.3 Electron momentum scale and resolution

The combination of the tracker and the calibrated ECAL measurements determines the electron momentum. The energy response in each crystal is calibrated with a correction factor to take into account effects such as energy leakage or changes in the crystal transparency induced by radiation[6]. The SC energy is also corrected using an MVA technique, selecting Drell-Yan events in the data, and comparing to the simulation. A detailed description of the technique used to estimate the electron scale and resolution is given in [74].

### 4.4.4 Muon reconstruction and identification

Muons are the only charged particles transversing the CMS apparatus. The muon spectrometer residing at the outermost layers of the CMS detector allows muons to be identified with high efficiency up to pseudorapidity of 2.4. Given the different detection systems in CMS, the muon physics objects are reconstructed into three collections or types:

1. Standalone muon: a fitted muon trajectory using hits reconstructed from the muon chambers. Each of the muon sub-detectors, DTs, RPCs, and CSCs, performs a local reconstruction and its respective candidates are combined and fitted with a KF algorithm;

2. Tracker muon: is reconstructed from selected inner tracks with certain kinematic criteria and matched via track extrapolation with at least one muon segment built with CSC or DT chambers. The energy loss through the intervening materials are accounted in the track fitting process.

3. Global muon: is reconstructed performing a global fit starting from standalone muons and extrapolating its trajectory into the inner tracker, matched with compatible tracks hits.

Muons reconstructed only as standalone-muon tracks have worse momentum resolution due to the shorter track lengths and to the multiple scattering induced by the iron between the muon stations.

The global muon fit provides a better momentum resolution with respect to the tracker muon for $p_{\mathrm{T}}$ above 200 GeV; below this value the resolution is dominated by the tracker muon fit; about 99% of the muons produced within the geometrical acceptance of the muon system are reconstructed either as a global muon or a tracker muon and very often as both.

The muon momentum resolution for Run-II is measured to be 1% in the barrel and 3% in the endcap for momenta up to around 100 GeV [50]; this is achieved also thanks to the sophisticated alignment system [78].

[6] The continuous monitoring of the crystal transparency is achieved by a laser-monitoring system [77].

The muon ID is achieved by applying a specific set of selections on discriminating variables implemented through the definition of WPs, in both cut-based and MVA-based versions, similarly to the electron treatment. For all the three data taking years, the tight and loose WPs selection criteria are summarized for the cut-based version in Tab. 4.3.

TABLE 4.3: The loose and tight cut-based muon ID selection requirements.

| ID variables | Tight | Loose |
|---|---|---|
| Global muon | True | - |
| PF muon | True | - |
| Tracker muon | True | True |
| $\chi^2/dof$ of global muon track | $< 10$ | - |
| Muon chamber hits in global track fit | $> 0$ | - |
| Matched muon station segments | $> 1$ | $> 0$ |
| $d_{xy}$(track,vertex) cm | $< 0.2$ | $< 0.3$ |
| $d_z$(track,vertex) cm | $< 0.5$ | $< 20$ |
| Pixel hits | $> 0$ | $> 0$ |
| Tracker layers with hits | $> 5$ | $> 5$ |

### 4.4.5 Muon isolation

Similarly to the electron case, in a typical signal region for a given physics analysis, there may be a considerable amount of non-prompt muons due to jet mis-identification and semi-leptonic decays of heavy flavor quarks; moreover, the presence of punch-through hits in the muon chambers from hadron shower remnants can lead to the reconstruction of non-prompt muons. Therefore, these muons are expected to be reduced from selection criteria looking for nearby hadronic activities, as prompt muons are expected to be isolated in the event.

Isolated prompt muons are required to pass requirement with a pile-up mitigation correction called $\Delta\beta$ correction. The ISO selection uses the isolation variable relative to the muon $p_{\mathrm{T}}$ ($p_T^{\mu}$), or PF relative isolation, defined as follows:

$$I_{\Delta\beta}^{rel} = \frac{1}{p_T^{\mu}} \left[ \sum_{ChH} p_T + \max\left( 0, \sum_{NH} p_T + \sum_{\gamma^0} p_T - 0.5 \sum_{ChH-PU} p_T \right) \right] \tag{4.6}$$

The subscript definition is similar to the electron's isolation variable Eq. (4.5); the new term $ChH - PU$ refers to the charged hadrons contribution that does not originate from the primary vertex. The sums are performed in a cone of radius $\Delta R$ of 0.4 around the direction of muon.

### 4.4.6 Muon momentum scale and resolution

The measurement of the muon $p_T$ is sensitive to the alignment between the tracker and the muon chambers, to the material composition and distribution inside the detector and to the magnetic field produced by the solenoid. The imperfect knowledge of the magnetic field and the effect of the material distribution introduce a bias in the muon $p_T$ that is generally independent on the $p_T$ itself, while the effect of the alignment produces a bias that increases linearly with the $p_T$. Different methods are used to estimate the muon $p_T$ scale and resolution effects, and to determine the corresponding uncertainties depending on the $p_T$ range. At low and intermediate muons $p_T$ (below 100 GeV), the di-muon events arising from the $J/\psi$ and $Z$ resonance decays are used to correct the $p_T$ scale and to measure the $p_T$ resolution. In the high $p_T$ regime, the muon $p_T$ scale and resolution are instead measured using cosmic ray muons passing the whole CMS detector, from the uppermost to the lowermost muon chambers.

### 4.4.7 Multivariate lepton identification

In addition to the cut-based ID particle selection mentioned in Sec. 4.4.2 and Sec. 4.4.4, and to the MVA-based electron selection, a physics analysis-oriented MVA-based ID variable using a BDT algorithm was developed [79] to separate prompt leptons from non-prompt leptons. The BDTs are trained separately for electrons and muons on simulated samples and for different data taking years. Leptons decays from known electroweak process such as Z+jets and W+jets processes, and leptons produced from leptonic $\tau$ decays in $t\bar{t}H$ events are used as signal samples, while non-prompt leptons produced from semi-leptonic $\mathrm{t\bar{t}}$ +jets events are used as background sample.

The BDT takes a set of discriminating variables, such as the lepton kinematics, lepton isolation, properties of the lepton's nearest jet, tracks impact parameters and other variables used in lepton ID criteria as inputs. The respective output score from the training will be referred as electron (muon) `ttHMVA` score, and their distributions are shown in Fig. 4.2. The shape difference between the prompt and non-prompt leptons provides an excellent discriminating power to distinguish the two processes by cutting at certain value corresponding to the desired signal and background efficiency.

The prompt lepton `ttHMVA` discriminant has similar performance for 2017 and 2018 data taking years, as shown in Fig. 4.3, while for 2016 the performance is slightly worse, as shown in Fig. 4.4.

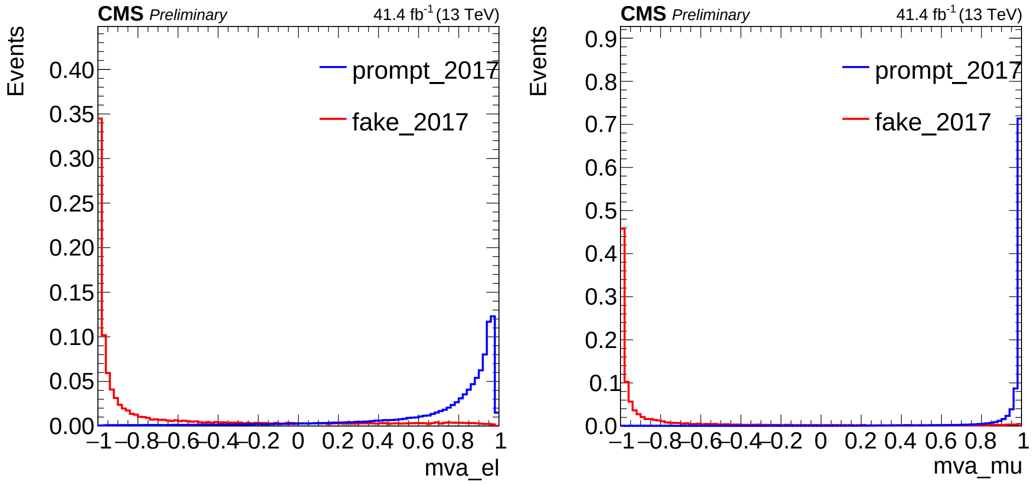



FIGURE 4.2: The output ttHMVA distribution both for electrons (left) and muons (right). The prompt (blue) and fake or non-prompt (red) leptons distributions are shown for the 2017 data taking year.

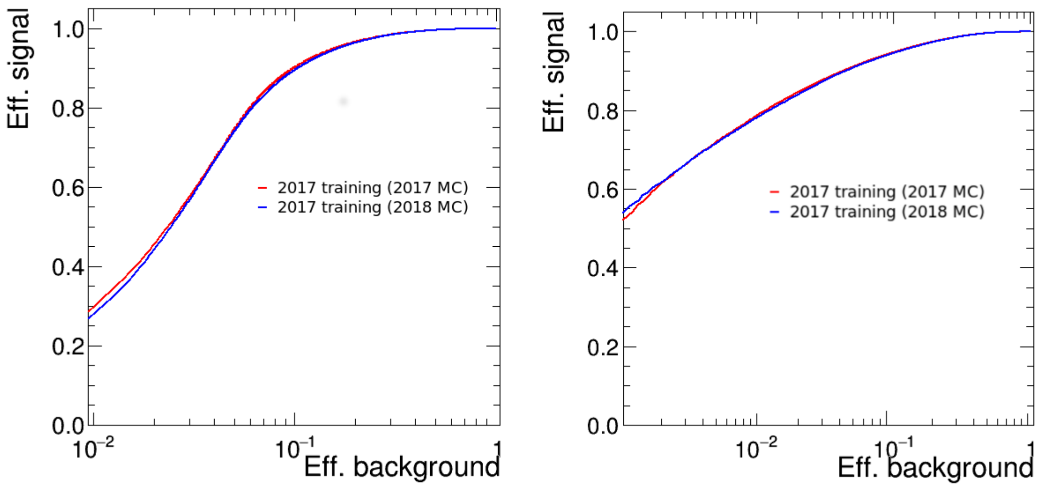



FIGURE 4.3: The performance of ttHMVA BDT output for electrons (left) and muons (right).

### 4.4.8 Study of the lepton identification and isolation efficiency

The ID and ISO selections in data and simulation have different selection efficiencies and the simulated events are corrected for the observed differences by a scale factor, defined as the ratio between the two efficiencies.

The Tag and Probe technique [80] is a method to estimate the efficiency of a selection on the data or simulation, based on a known resonance such as the $J/\psi$, or the Z boson. In general, the tag is an object passing a very tight selection criteria to isolate the required particle type; in this way the non-prompt lepton passing the tag selection criteria should be negligibly small. On the other hand, the probe object is selected by pairing the probe to the tag object

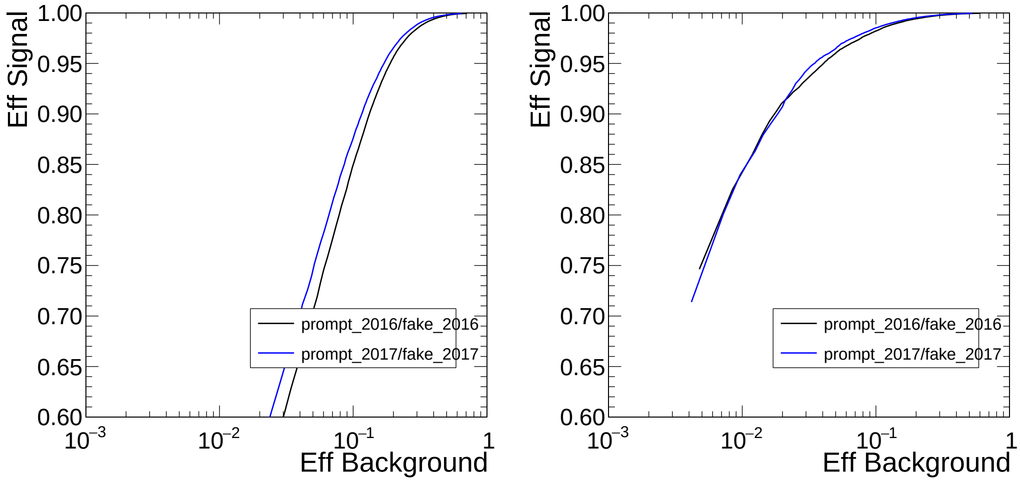



FIGURE 4.4: Performance of the `ttHMVA` BDT output for electrons (left) and muons (right). The `ttHMVA` score is trained and evaluated in 2016 (black) and 2017 (blue) simulated samples respectively.

so that the invariant mass of the pair is consistent with the mass of the resonance. The combinatorial backgrounds may be eliminated through background subtraction methods such as fitting, or side-band subtraction [81]. The efficiency $\epsilon$ of the selection criterion being examined, as a function of $p_{\mathrm{T}}$ and $\eta$, is given by the expression:

$$\epsilon(p_T, \eta) = \frac{N_{pass}^{probes}}{N_{pass}^{probe} + N_{fail}^{probe}} \tag{4.7}$$

where $N_{pass}^{probes}$ and $N_{fail}^{probes}$ are the number of events in which the probe object passes or fails the selection.

For the estimation of the electron (muon) ID selection efficiency, the tag is chosen to be a well identified and isolated electron (muon), while the probe is chosen as an electron (muon) identified with loose selections. The invariant mass of the tag-probe pair is required to be within a window around the Z boson mass (the effect of changing the Z mass window is included as a systematic uncertainty). After that, the probe is required to pass the analysis ID selections discussed in Sec. 5.5 for electrons and Sec. 5.6 for muons, and the efficiency is computed both in data and simulation.

Both the lepton ID and the lepton ISO selection efficiencies are measured by the fitting method [80], to take into account the combinatorial background below the resonance peak. The signal plus background fit to the invariant mass distribution is performed simultaneously in two categories, corresponding to events in which the probe lepton passes or fails the identification requirements, and separately in bins of $p_{\mathrm{T}}$ and $\eta$.

At this stage of selection, the electron and muon ID and ISO efficiencies are close to 90-95%, slightly depending on the pseudorapidity. The global efficiency relevant for the analysis will be discussed in detail in the next chapter in Sec. 5.4.1, after the additional selection discussed there.

## 4.5 Jet reconstruction and identification

Due to the QCD interactions occurring after the hard scattering process, the outgoing partons manifest themselves as sprays of final state particles. As the process of interest involves quarks in the final state, the momenta of the quarks can be studied by clustering the particles in the vicinity of the quark, reconstructing a single object called jet.

Most physics analysis make use of PF jets which are reconstructed by clustering all particles reconstructed via the PF algorithm discussed in Sec. 4.3 without any distinction of type and energy threshold. This method provides an excellent improvement in the jet momentum and spatial resolution compared to the reconstructed jets using only calorimeters information.

The jet object is defined through sequential, iterative clustering algorithm combining four-momenta of input particles until a certain set of conditions is satisfied [82]. The standard jet clustering algorithm used by CMS is the anti-$k_t$ algorithm [83]. The anti-$k_t$ algorithm is a sequential recombination type algorithm, with well defined attributes such as infrared and collinear safety [84]. It makes use of two distance parameters: $d_{ij}$, the distance between the two particles $i$ and $j$, given by Eq. (4.8); and $d_{iB}$, the distance between the particle $i$ and beam jets[7], given by Eq. (4.9):

$$d_{i,j} = min(k_{t,i}^{-2}, k_{t,j}^{-2})\frac{\Delta_{i,j}^2}{R^2} \tag{4.8}$$

$$d_{i,B} = k_{t,i}^{-2} \tag{4.9}$$

where $\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_2)^2$ with $y_i$, $\phi_i$ and $k_{t,i}$ are the rapidity, azimuthal angle and the transverse momentum of particle $i$, respectively. The parameter $R$ is defined as the radius of the cone, and only particles lie within the base area of the cone ($A_{cone} = \pi R^2$) are considered. The algortihm proceeds as follows:

- the distances $d_{ij}$ are calculated for all pairs of particles $i, j$ and the distances $d_{iB}$ are calculated for each particle $i$ accordingly;
- if the smallest distance is $d_{ij}$, particles $i$ and $j$ are combined into a single new particle summing their four-momenta, and the algorithm returns to the first step;
- if the smallest distance is $d_{iB}$, particle $i$ is declared as a final state jet and removed from the event, and the algorithm returns to the first step;

[7] The beam jets are produced with small momenta particles transverse to the beam axis from the UE and ISR of a hadron collision [85].

- the procedure is repeated until all particles in the event are assigned to a jet.

Jets clustered with the anti-$k_t$ algorithm with cone radius of $R = 0.4$ are conventionally called AK4 jets. To improve the purity of the clustered jets, the anti-$k_t$ algorithm can be improved with pile-up subtraction methods such as the Charged Hadron Subtraction (CHS) method [43]. The method removes all charged particles not associated to the primary vertex during the algorithm iteration.

### 4.5.1 Calibration of the jet energy response

Jet energy corrections are employed to account for the difference between measured jet energy and the true energy of the particles in the final state. To achieve this, the jet energy is simulated by clustering all the simulated stable particles[8] produced in the event excluding neutrinos via the same jet algorithm. The mismatch with the jet energy from reconstructed particles originates from the non-uniform and non-linear response of the CMS calorimeters, detector inefficiencies, electronics noise and pile-up effects. The mismatch is addressed as a correction factor, which is derived from a sequential procedure called Jet Energy Correction (JEC) [86], schematically shown in Fig. 4.5.

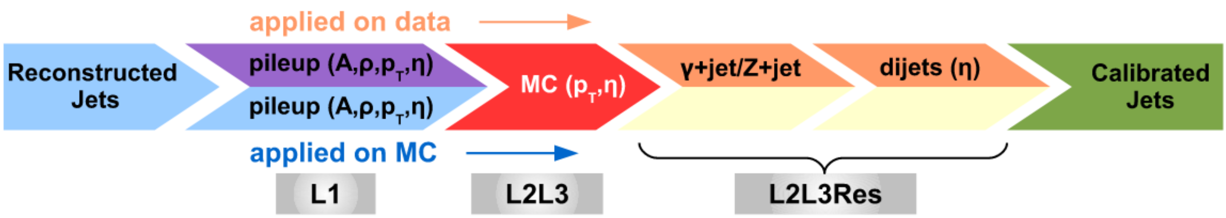



FIGURE 4.5: The sequential approach of the jet energy correction applied both on simulation and data.

In the first step, a L1 offset correction is applied to estimate and subtract the remaining energy contribution that is not associated with the hard scattering in the event. The estimation approach called the jet area method [87] is used before the correction. In this method, soft jet activities such as contribution from underlying events, instrumental and pile-up effects, are represented by the energy density per unit area, $\rho$. It is estimated event by event using the median of the distribution variable $p_T^j/A_j$, where $j$ runs over all jets in the event and $A_j$ is the jet effective area. The instrumental and pile-up effects are contaminating the hard jet energy measurement, and need to be subtracted from the momentum measurement. At detector level, the measured density is the convolution of the true particle-level activity (underlying event, pile-up) with the detector response to the various particle types. Therefore, the

[8]With $c\tau > 1$ cm.

event-by-event and jet-by-jet offset correction can be defined as:

$$C_{offset}(p_T^{raw}, A_j, \rho) = 1 - \frac{(\rho - <\rho_{UE}>) \cdot A_j}{p_T^{raw}} \tag{4.10}$$

where $<\rho_{UE}>$ represents the average energy density component due to underlying event and electronics noise, measured in events with exactly one reconstructed primary vertex, and $p_T^{raw}$ is the raw jet four-momentum.

In a second step, the L2L3 MC calibration correction is applied to correct for non-linear response of the reconstructed jet energy, as a function of $p_{\mathrm{T}}$ and $\eta$. The correction is evaluated from simulated QCD events processed through the same jet reconstruction algorithm as data samples. Each reconstructed jet is spatially matched to a generated jet with $\Delta R <$ 0.25. For each generated jet transverse momentum $p_T^{gen}$, the response variable is defined as $R = p_T^{reco}/p_T^{gen}$, where $p_T^{reco}$ is matched reconstructed jet transverse momentum. Hence, the average correction applied in each $p_T^{reco}$ bin is derived as a function of $<p_T^{reco}>$:

$$C_{MC}(p_T^{reco}) = \frac{1}{<R>} \tag{4.11}$$

Finally, the small residual corrections (L2L3Res) between data and simulation are calculated by using real data. The absolute (relative) jet energy scale $C_{abs}(p_T)$ ($C_{rel}(\eta)$) corrections are introduced to ensure that the jet response is independent of $p_{\mathrm{T}}$ and $\eta$. The former correction is estimated by missing transverse energy projection fraction (MPF) method [88], which uses $\gamma$+jets and Z+jets events[9] data samples; the latter is estimated by tag and probe technique by using di-jet events data samples.

The different correction terms are applied in sequence as described in Fig. 4.5:

$$C_{tot} = C_{offset}(p_T^{raw}) \cdot C_{MC}(p'_T, \eta) \cdot C_{rel}(\eta) \cdot C_{abs}(p''_T) \tag{4.12}$$

where $p'_T$ is the jet $p_{\mathrm{T}}$ after the L1 offset correction is applied and $p''_T$ is the jet $p_{\mathrm{T}}$ after the L2L3 correction.

### 4.5.2 Jet energy resolution

Usually the Jet Energy Resolution (JER) in data is worse than in simulation. Hence, a smearing procedure is introduced to the simulated jets in order to better describe the data.

An independent study on JER as a function of pile-up, jet size (cone radial parameter $R$) and jet flavor is carried out both in data and simulation by using di-jet and $\gamma$+jet events. The

[9]These events have no intrinsic missing transverse energy. At parton level, the $\gamma$ and Z bosons are purely balanced by the hadronic recoil in the transverse plane.

typical JER in the central region are 15-20% at 30 GeV, about 10% at 100 GeV, and 5% at 1 TeV [89].

The reconstructed jet $p_{\mathrm{T}}$ in simulated events are corrected for the JER by scaling with the observed $p_{\mathrm{T}}$ difference between reconstructed and generated jets, matched using $\Delta R$ and $\Delta p_T$ requirements. Later, a Gaussian smearing of the $p_{\mathrm{T}}$ distribution of the reconstructed jet is applied in order to obtain the desired resolution. The Gaussian smearing procedure is also applied to those reconstructed jets not matching any generated jets.

### 4.5.3 Jet identification

The quality of the PF jets are controlled by a set of basic quality criteria to reduce contamination of mis-reconstructed jets originated from instrumental backgrounds such as noisy calorimetric cells. These criteria are collectively called jet ID, which are used to reject poor-reconstructed jets while maintaining a large fraction (about 99%) of real jets.

The tight jet ID is defined by a set of selection, using jet variables based on the electromagnetic and hadronic activities, such as the number of charged and neutral particles inside the reconstructed jets, as a function of $\eta$. For AK4 jets in the region $|\eta| \leq 2.7$, the jet ID requirements are summarized in Tab. 4.4. In the table, the neutral hadron (electromagnetic) fraction referes to the ratio between the hadronic (electromagentic) energy not associated to charged particles and the total energy of the jet.

TABLE 4.4: The tight jet ID criteria based on PF jets with $|\eta| < 2.7$. The selection using neutral component is defined for jets in $|\eta| < 2.7$ region; for 2016 and 2017 years, the selection using charged components is only applicable for jet in $|\eta| < 2.4$ region.

| Jet ID variables | 2016 | 2017 | 2018 |
|---|---|---|---|
| Number of constituents | $> 1$ | $> 1$ | $> 1$ |
| Neutral hadron fraction | $< 0.99$ | $< 0.90$ | $< 0.90$ |
| Neutral electromagnetic fraction | $< 0.90$ | $> 0.90$ | $< 0.99$ |
| Charged multiplicity | $> 0$ | $> 0$ | $> 0$ |
| Charged hadron fraction | $> 0$ | $> 0$ | - |
| Charged electromagnetic fraction | $< 0.90$ | - | - |

In the region of $2.7 < |\eta| \leq 3.0$, the jet ID selection is tighter: the AK4 jets are required to contain more than two neutral particles in the jets, with neutral electromagnetic fraction to be above 0.01 for 2016, and between 0.02 and 0.99 for 2017 and 2018. In addition, the neutral hadron fraction must be below 0.98 for 2016. For AK4 jets in the region $2.7 < |\eta| \leq 3.0$, the jet ID requirements are summarized in Tab. 4.5.

For AK4 jets in the region $3.0 < |\eta| < 5.0$, the tight jet ID selection requires the neutral electromagentic fraction to be below 0.90 and number of neutral particles contained in the jet to be above 10 for all run periods; the neutral hadron fraction is required to be above

TABLE 4.5: The tight jet ID criteria based on PF jets with $2.7 < |\eta| \leq 3.0$.

| Jet ID variables | 2016 | 2017 | 2018 |
|---|---|---|---|
| Number of neutral particles | > 2 | > 2 | > 2 |
| Neutral hadron fraction | < 0.98 | - | < - |
| Neutral electromagnetic fraction | > 0.01 | (0.02, 0.99) | (0.02, 0.99) |

0.02 and 0.2 for 2017 and 2018. The jet ID requirements are summarized in Tab. 4.6 for this pseudorapidity region.

TABLE 4.6: The tight jet ID criteria based on PF jets with $3.0 < |\eta| < 5.0$.

| Jet ID variables | 2016 | 2017 | 2018 |
|---|---|---|---|
| Number of neutral particles | > 10 | > 10 | > 10 |
| Neutral hadron fraction | < - | 0.02 | < 0.2 |
| Neutral electromagnetic fraction | < 0.90 | < 0.99 | < 0.99 |

## 4.6 The b-tagging

Given the fact that top quarks and Higgs boson decay into bottom quarks frequently, the identification of bottom quark-initiated jets provides a strong background suppression or signal enrichment capability for many analyses.

The b-tagging is a reconstruction technique that takes advantage of the characteristic properties of b hadrons and assigns to each jet a likelihood to contain a b hadron. In particular, the technique takes advantage of the fact that the b hadrons tend to decay a few millimeter away from the primary vertex, producing tracks within the jet with high impact parameters, or with the presence of secondary vertex. Additionally, on average, the bottom quark-initiated jets tend to contain more tracks compared to other flavors.

A variety of b-tagging algorithms based on different input information has been developed in CMS experiment [90]. A common feature is that all these algorithms produce an output discriminator which describes how likely it is that a given jet is actually a b-jet[10].

Given the large amount of observables available, b-taggers make use of machine learning algorithms to combine them into one powerful discriminator. The identification can also be improved by using the progresses in the field of deep machine learning [31]. The improved version of Combined Secondary Vertex (CSV) [91], called the `DeepCSV` algorithm, is trained by using the jet and track properties, track secondary vertex and soft lepton information. The `DeepCSV` outperforms all the other b-jet identification algorithms, when discriminating against charm jets or light-flavor jets, except for b-jet identification efficiencies above 70%, as shown in Fig. 4.6.

[10] The higher is the discriminator value, the higher is the probability that the jet is a real b-jet.

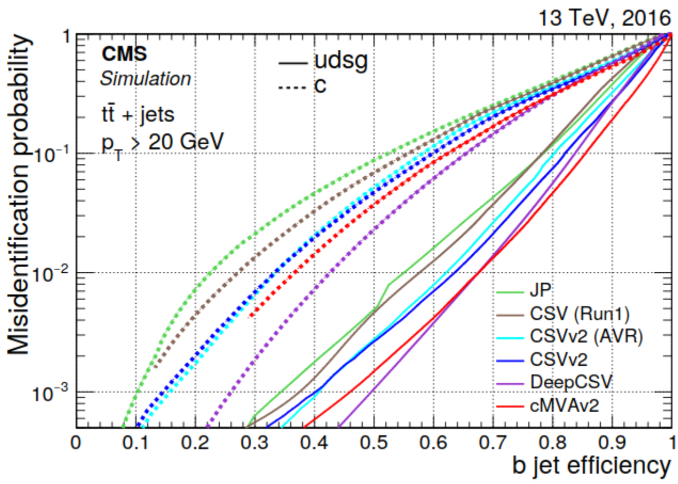



FIGURE 4.6: The performance of different b-tagging algorithms, showing the mis-identification probability for charm and light-flavour jets versus the probability of identify the b-jets. The curves are obtained from simulated $t\bar{t}$ events using anti-$k_T$ jets clustered with R=0.4 and requiring $p_{\mathrm{T}} > 30$ GeV.

Three working points, "loose" (L), "medium" (M), and "tight" (T) are defined, corresponding to thresholds on the DeepCSV discriminator after which the mis-identification probability for light-flavour jets is about 10%,1%, and 0.1%.

The DeepCSV output discriminant distribution is then corrected for the difference between simulation and data by scale factors obtained via an iterative fitting technique [92]. The scale factors for both b and light flavor (udsg) jets are separately derived as a function of the discriminator value, jet $p_{\mathrm{T}}$ and $\eta$. In particular, simulated event samples enriched in $t\bar{t}$ process are used for b-jet scale factor measurement, while the light flavor jets are measured in 99% pure Z+jets events [90]. In the analysis, the simulated events are reweighted using these scale factors, with the event weights $w$ computed as:

$$w = \prod_{i}^{N_{jets}} SF_{jet_i}\left(p_{Ti}, \eta_i, D_i\right) \tag{4.13}$$

where $D_i$ is the discriminator value and $i$ runs over all the jets in the event with $|\eta_i| < 2.4$ and $p_{Ti} > 20$ GeV, for which b-tagging is applied.

## 4.7 Missing transverse momentum

In a hadron collider, the exact incoming partons longitudinal momentum is not known, and the initial transverse momentum is expected to be practically zero. Hence, the conservation of momentum implies that the sum of the transverse momentum of all the particles in the final state of the collisions should be zero. The missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ is a vector quantity defined as the negative vectorial sum of transverse momenta of all reconstructed PF objects:

$$\vec{p}_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{i \in PF-obj} \vec{p_{T,i}} \tag{4.14}$$

where the sum runs over all the PF objects in an event. The modulus of Eq. (4.14) is called missing transverse energy:

$$E_{\mathrm{T}}^{\mathrm{miss}} = \left| \vec{p}_{\mathrm{T}}^{\mathrm{miss}} \right| \tag{4.15}$$

The presence of $E_{\mathrm{T}}^{\mathrm{miss}}$ not compatible with zero is a signature of particles which have escaped the detector without interaction, such as neutrinos or particles predicted by BSM models. However, it could also originate from detector inefficiencies.

The $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm produces PF $p_{\mathrm{T}}^{\mathrm{miss}}$ according to the Eq. (4.14) using all the PF candidates in the event. Alternatively, a second algorithm has been developed to further reduce the pile-up contribution in the computing $p_{\mathrm{T}}^{\mathrm{miss}}$. This algorithm relies on the Pile-Up Per Particle Identification (PUPPI) method [93], using both global information from the event, and local information to quantify the pile-up effect at the particle level, by applying a weight to re-scale the four momentum of particles in a jet and correct jet shapes in an observable-independent way. Ideally, particles coming from pile-up would get a weight of zero and particles coming from the hard scattering interaction would get a weight of one. This leads to a pile-up corrected event, which is input to the jet finding algorithm without the need for further pile-up corrections.

Instrumental effects such as spurious deposits may occur in the ECAL and HCAL due to direct particle interaction with sensors such as photodetectors or photomultiplier, particles shower with non-collision origins[11], calorimeter dead cells, and noise from readout electronics. A cleaning procedure is applied to the data in order to remove the affected events.

### 4.7.1 The $E_{\mathrm{T}}^{\mathrm{miss}}$ scale and resolution measurement

The performance (scale and resolution) of the $E_{\mathrm{T}}^{\mathrm{miss}}$ can be studied in events with an identified Z boson or an isolated photon. Momenta of leptons and photons can be reconstructed with good resolution of about 1-6% [94], while momenta of jets are reconstructed with less precision, with typical resolutions of 5-15% [89]. Therefore, the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution in Z or $\gamma$ + jets events is dominated by the hadronic activity in the event.

The comparison of the momenta of the vector boson with respect to the hadronic recoil system is used to measure the $E_{\mathrm{T}}^{\mathrm{miss}}$ performance. Fig. 4.7 shows schematically the vector boson momentum in the transverse plan, $\vec{q}_T$, and the transverse momentum $\vec{u}_T$ of the hadronic recoil, defined as the vectorial sum of the transverse momenta of all particles

[11] Caused by beam halo particles, a machine-induced background mainly originating from interaction of the beam protons with collimators before the CMS detector.

except the vector boson (or its decay products, in the case of Z bosons). The momentum conservation in the transverse plane implies that $\vec{q}_T + \vec{u}_T + \vec{p}_{\mathrm{T}}^{\mathrm{miss}} = 0$.

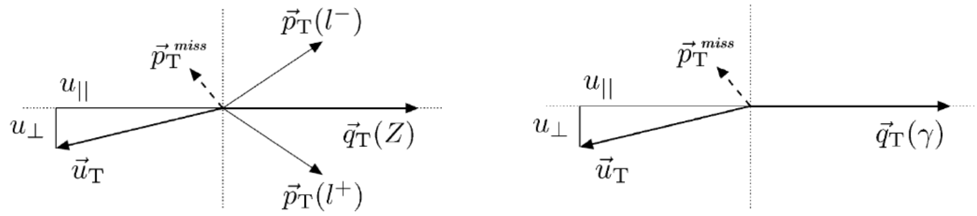


FIGURE 4.7: Illustration of the $Z \to l^+l^-$ (left) and the photon (right) event kinematics in the transverse plane. The vector $\vec{u}_T$ denotes the vectorial sum of all particles reconstructed in the event except for the two leptons from the Z decay (left) or the photon (right).

The $E_{\mathrm{T}}^{\mathrm{miss}}$ properties are evaluated using two components of $\vec{u}_T$, one parallel ($u_{\|}$) and one perpendicular ($u_{\perp}$) to the axis defined by $\vec{q}_T$. The distributions of these variables are parametrized using a convolution of a Breit-Wigner and a Gaussian function, which is found to provide a good description of the observables and is used to measure the resolution in $u_{\|}$, $u_{\perp}$, $\sigma(u_{\|})$ and $\sigma(u_{\perp})$, respectively. The resolutions obtained using early data at 13 TeV are shown in Fig. 4.8 and compared to simulation of $Z \to \mu^+\mu^-$ events for PF $E_{\mathrm{T}}^{\mathrm{miss}}$ and PUPPI $E_{\mathrm{T}}^{\mathrm{miss}}$ observables. The resolution measured for PUPPI $E_{\mathrm{T}}^{\mathrm{miss}}$ shows less degradation in higher pile-up scenarios compared to PF $E_{\mathrm{T}}^{\mathrm{miss}}$.

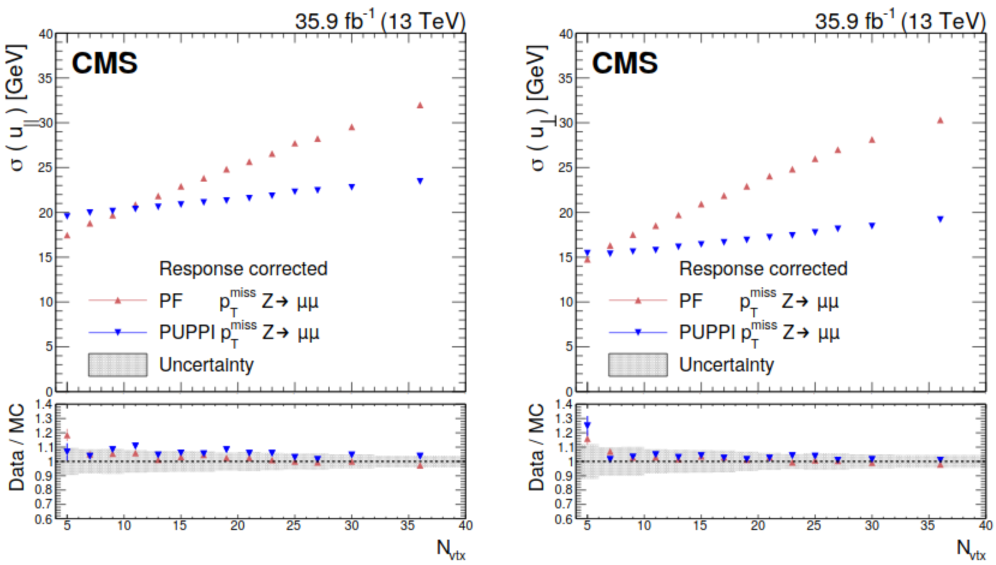



FIGURE 4.8: The PUPPI and PF $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution of $u_{\|}$ (left) and $u_{\perp}$ (right) components of the hadronic recoil, as a function of number of vertex ($N_{vtx}$), in $Z \to \mu^+\mu^-$ events. The upper frame shows the resolution in data; the lower frame shows the ratio of data to simulation with the error band displaying the systematic uncertainty of the simulation [95].

# Chapter 5

# Measurement of the Higgs boson production in association with a W boson in final states with two same sign leptons and light quarks

The measurement of the Higgs boson produced in association to a W boson and decaying to a pair of W bosons in the same-sign lepton topology has never been explored by the CMS experiment to date; its first measurement is presented in this chapter.

This channel is free from most of the dominant electroweak backgrounds thanks to the same sign requirement of the two leptons. Moreover, the different flavor leptonic decay mode has a large branching fraction, and therefore is expected to lead to a good sensitivity; in fact, the final state benefits from the large branching ratio of $H \rightarrow W^{+}W^{-}$ decay channel, making this channel a good candidate for the study, especially in view of the large integrated luminosity, 137 $\text{fb}^{-1}$, collected in the full Run-II dataset collected at $\sqrt{s} = 13$ TeV.

## 5.1 Signal characterization

In the inelastic proton-proton interaction, a highly energetic off-shell W boson may be produced via a s-channel diagram, initiated by two incoming light quarks from the proton (see Fig. 5.1). Subsequently, the intermediate W boson radiates an on-shell Higgs boson. This process is also known as Higgs-strahlung. In their reference frame, the two bosons recoil towards each other, resulting in a back-to-back topology as shown in the plot of Fig. 5.2a, and the $p_{\text{T}}$ of the two bosons are strongly correlated, as shown in Fig. 5.2b.

The associated on-shell W boson decays into a charged lepton and a neutrino, while the Higgs boson decays to $H \rightarrow W^{+}W^{-}$, and the two W bosons subsequently decay to a charged lepton and neutrino, and a pair of light quarks. Since the mass of the Higgs boson is smaller than the sum of the two W boson masses, either one of the W bosons is off-shell. Hence, the off-shell W decay products are expected to be emitted closer to each other, as shown by the distributions in Fig. 5.2c and Fig. 5.2d.

The scalar nature of the Higgs boson implies that the two W bosons must have opposite polarization[1]. From the V-A[2] nature of the W bosons, it follows that the lepton and the system of the light quarks in the event are preferentially emitted at small azimuthal angle with respect to each other, as shown in Fig. 5.2e. The lepton from $H \rightarrow W^+W^-$ decay channel and the lepton from the associated W boson have a rather large pseudorapidity separation as shown in Fig. 5.2f.

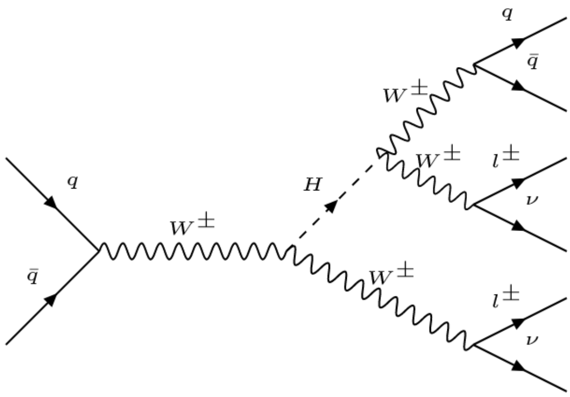
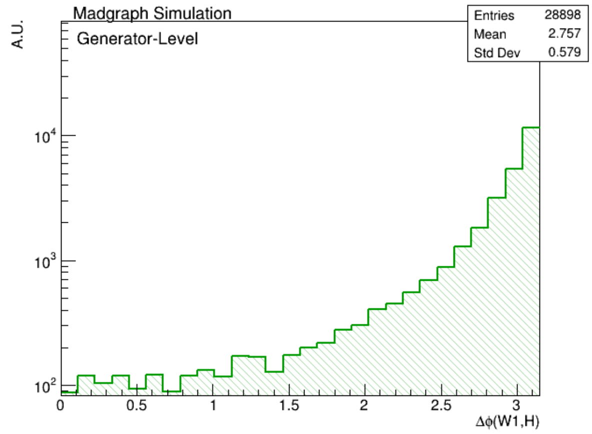



FIGURE 5.1: The Feynman diagrams for associated production of a Higgs boson with a leptonically decaying W bosons in the same sign final state.

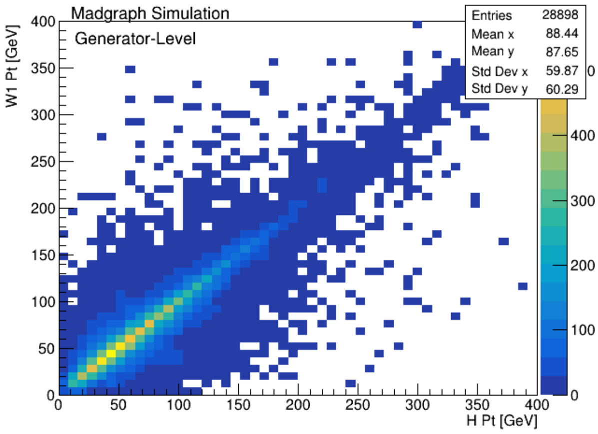
At reconstruction level, the same sign di-lepton final state has the following properties:

- the pair of same sign leptons are widely separated in azimuthal angle;
- the leading lepton is expected to trigger the event, and most of the times (∼95%) originates from the leptonically decaying associated W boson, while the sub-leading $p_{\mathrm{T}}$, or trailing lepton, originates from the Higgs decay;
- in events with off-shell leptonic W boson decay, the corresponding neutrino contributes negligibly to the total missing transverse momentum;
- events with off-shell hadronic W boson decay may result into a pair of jets which are too close each others, so that the CMS standard jet algorithm [83] may cluster them together, resulting into 1 instead of 2 reconstructed jets in the final state;
- the trailing lepton is expected to be closer in azimuthal angle to the reconstructed jet(s) system;

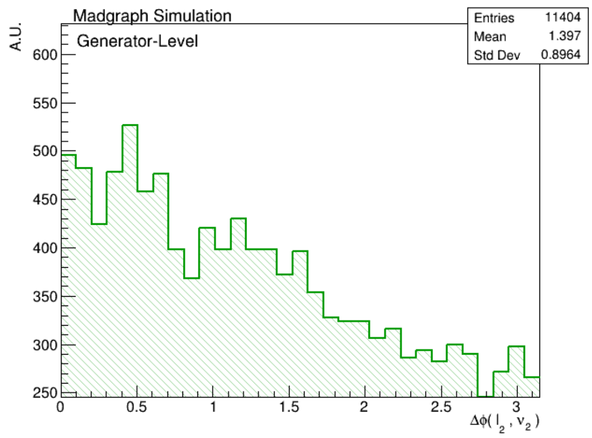
[1]The W boson only couples to left handed fermions.
[2]Refers to Vector-Axial Vector structure.



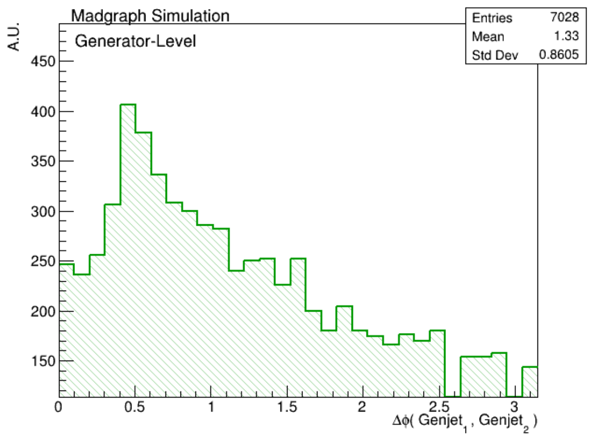

(A) The azimuthal angle between the associated W boson and the Higgs boson.



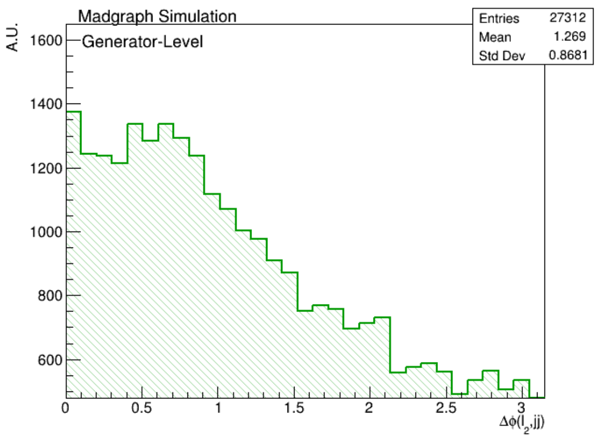

(B) The $p_T$ correlation between the associated W boson and the Higgs boson.



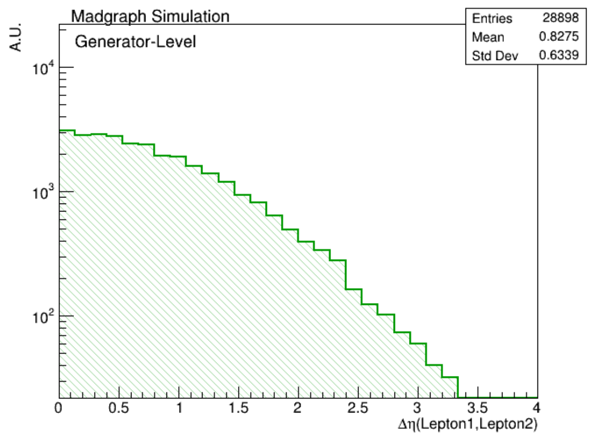

(C) The azimuthal angle between the leptonic off-shell W decay products.




(D) The azimuthal angle between the hadronic off-shell W decay products.




(E) The azimuthal angle between the lepton and jets, from the Higgs decay.




(F) The pseudorapidity between the pair of same sign di-lepton.

FIGURE 5.2: The generator level kinematic distribution of the W bosons and its decay products. The setup for the generator level study can be found in Appendix A.

The relevant backgrounds having the same or similar final state as the signal process, in descending order of importance, are summarized below:

- **WZ** : the fully leptonic decay of the di-boson production process yields two same sign di-lepton final state when one of the three leptons escapes detection due to the tracker acceptance or inefficiency.
- **Non-prompt background** : this background contains events where jets or photons, mainly originated from the W+jets production process, are mis-identified as leptons, or when genuine leptons from semi-leptonic charm and beauty decays are identified as isolated, due to poorly reconstructed jets; the main sources of these non-prompt leptons are from W+jets, semi-leptonic $t\bar{t}$, QCD and hadronic top processes.
- **W$\gamma^{\star}$** : the W$\gamma^{\star}$ background enters the signal region when a highly energetic off-shell photon converts into a pair of electrons in the detector material, and is reconstructed as a single electron.
- **Z+jets** : it is a common dominant electroweak background for final states with two opposite sign leptons; this background is relevant to same sign different flavor final state via $Z \rightarrow \tau^{+}\tau^{-} \rightarrow \mu^{\pm}\nu e^{\mp}\nu$ when the charge of the electron from the $\tau$ decay is misidentified (charge flipping).
- **Fully leptonic $t\bar{t}$ and $W^{+}W^{-}$** : these background processes can originate two isolated leptons with different flavor in the final state, and the contribution to the signal region is due to the electron charge flipping.
- **$W^{\pm}W^{\pm}$+jets** : this background has the same final state with the signal process; the two same sign leptons and a pair of jets are originated from the scattering of W bosons with the same charge, radiated from the incoming partons during the collision; the recoiling partons are observed as jets with a large rapidity gap.
- **VVV** : this background consists of tri-bosons processes, WWW, WWZ, WZZ, ZZZ. The most relevant background to the signal region comes from the WWW process, when two same charge W bosons decay leptonically and the other hadronically.

## 5.2 Analysis strategy

The analysis is looking for the signal events consisting of two isolated same sign leptons separated with large pseudorapidity, one or two jet(s) and moderate missing transverse energy, dominated by the neutrino from the decay of the associated W boson.

The measurement makes use of the full Run-II dataset collected at 13 TeV from the pp collisions. The signal event is triggered by using a variety of leptonic trigger, requiring at least one well identified and isolated lepton passing the respective trigger threshold.

Different analysis strategies were attempted, including multivariate and cut-and-count methods. Finally, the best results were obtained by using a shape based analysis.

The signal region is pre-selected according to the same sign di-lepton topology, and the sensitivity of the signal process is maximized by an event categorization based on the di-leptons flavor composition and the number of jets. The same flavor pair consisting of di-electrons ($e^{\pm}e^{\pm}$) is found to be marginal in terms of signal sensitivity and not considered in the analysis. The event categories used in the signal region are summarized in Tab. 5.1. The details of the signal region definition will be discussed in Sec. 5.5 and summarized in Tab. 5.7.

TABLE 5.1: The event categories based on the flavor of two same sign leptons, and the number of jet in the final state.

| Flavor composition | 1-jet | 2-jets |
|---|---|---|
| $\mu^{\pm}$-$\mu^{\pm}$ | ✓ | ✓ |
| $e^{\pm}$-$\mu^{\pm}$ | ✓ | ✓ |

The analysis sensitivity is expected to be limited by the presence of non-prompt backgrounds in the signal region. The `ttHMVA` multivariate variable, based on the BDT algorithm (see Sec. 4.4.7) used to discriminate non-prompt from prompt leptons, has been developed to optimize the lepton identification selection. The optimized lepton selection reduces drastically the non-prompt background and the signal yield remains roughly the same, thus improving the overall signal significance.

The systematic uncertainties related to the most important background processes are reduced using data-driven methods which improve the background modeling. In particular, the non-prompt background is estimated using dedicated, signal free data samples via the so-called fakeable object method [96], used to extrapolate the non-prompt leptons contribution in the signal region.

The effect of charge mis-assignment in opposite sign two leptons backgrounds such as Z+jets, fully leptonic $t\bar{t}$ and non-resonant $W^{+}W^{-}$ is accounted by the simulation, corrected by a scale factor measured on the data from a dedicated study. The charge mis-assignment mainly affects the electron, as the track information is only reconstructed within the tracker. For a muon, the charge measurement can be precisely determined by combining extra information from the track reconstructed outside the tracker, in the muon station. Hence, this background is only relevant to final state with same sign different flavor.

The remaining backgrounds and the signal processes are modeled from MC simulation; the WZ background normalization is estimated in a dedicated data control region, through a simultaneous fit to the data in the signal and control regions.

To improve discrimination between signal and background, a mass variable is defined as a proxy for the reconstructed Higgs boson mass. The definition of the mass variable will be discussed in Sec. 5.5. The mass variable is then fitted by using the maximum likelihood method. The signal strength, the WZ normalization and a set of nuisance parameters are simultaneously measured and extracted by maximizing the likelihood function.

## 5.3 Data and Monte Carlo samples

In this section, the data samples, the set of triggers, and the simulated samples used in the analysis are described.

### 5.3.1 Data samples

The data sample used in this study is the full Run-II dataset collected from pp collisions during the three data taking years from 2016 to 2018. Events passing trigger selections are recorded in Primary Datasets (PDs), each one based on a list of similar triggers and containing events passing at least one of those triggers. The PDs used in the analysis are:

- **SingleMuon**: at least one identified and reconstructed muon in the online processing.
- **SingleElectron**: at least one online identified and reconstructed electron.
- **DoubleMuon**: at least two online identified and reconstructed muons.
- **DoubleEG**: at least two online identified and reconstructed electrons.
- **MuonEG**: at least one electron and one muon identified and reconstructed online.

The dataset was certified as free from instrumental pathological noises from all subsystems and ready for being analyzed. Additionally, dedicated noise filters are applied on data and simulated samples to reduce noise-dominated events. The dataset corresponds to the integrated luminosity of 137 $\text{fb}^{-1}$, amounting to about 91% of the total luminosity delivered by LHC.

### 5.3.2 Triggers

A signal event is expected to be triggered by the leading $p_{\mathrm{T}}$ lepton originated from the associated W boson, while the trailing lepton is most of the times originated from the Higgs boson's decay. Events are required to contain at least two leptons, to either fire the lowest thresholds unprescaled single electron, single muon or the muon-electron triggers. Due to the high LHC instantaneous luminosity, the single lepton trigger is designed to have high HLT $p_{\mathrm{T}}$ threshold in order to sustain the trigger rate. On the other hand, the double-lepton triggers allow lower $p_{\mathrm{T}}$ thresholds. Hence, a combination of single and double lepton triggers is employed to maximize the sensitivity to the Higgs boson signal. A procedure to avoid double counting is adopted for events appearing in more than one PDs. The leptonic trigger $p_{\mathrm{T}}$ thresholds for each trigger path are summarized in Tab. 5.2.

TABLE 5.2: The trigger threshold in $p_{\mathrm{T}}$ (GeV) applied in each lepton trigger at the HLT level, for each data-taking year. The value in brackets indicates the second lepton $p_{\mathrm{T}}$ threshold.

| Trigger path | 2016 | 2017 | 2018 |
|---|---|---|---|
| SingleMuon | 24 | 27 | 24 |
| SingleElectron | 27 | 35 | 35 |
| DoubleMuon | 17 (8) | 17 (8) | 17 (8) |
| DoubleEG | 23 (12) | 23 (12) | 23 (12) |
| MuonEG | 23 (12) | 23 (12) | 23 (12) |

The combined trigger efficiency is estimated from data as described in Sec. 4.4.1, and applied as a weight to the simulated events.

### 5.3.3 Simulated samples

MC simulation is extensively employed for signal modeling and background estimation. The signal and background processes are simulated with different ME generators depending on the simulated hard-scattering process. For samples generated by MADGRAPH5_aMC@NLO, the matching between the ME to PS is performed with MLM scheme [67] at LO and FxFx scheme [68] at NLO. The effects of PS, hadronization and multi-parton interactions is commonly simulated by PYTHIA8, unless otherwise stated.

All the signal samples are generated assuming a Higgs boson mass of 125.09 GeV [25]. The Higgs production modes, including gluon fusion (ggH), vector-boson fusion (VBF) and vector boson associated production (VH) are generated with POWHEGV2, designed to describe the properties of these processes at NLO accuracy; the top quark pair associated production ($\mathrm{t\bar{t}H}$), on the other hand, is generated with MADGRAPH5_aMC@NLO. The decay modes, $H \rightarrow W^{+}W^{-}$ and $H \rightarrow \tau^{+}\tau^{-}$ are considered and simulated with PYTHIA8 for all signal samples. For associated WH production, the positively and negatively charged associated W boson samples are generated and simulated separately.

The simulated signal samples are normalized using the appropriate cross sections [97], and branching fractions [98] provided by the LHC Higgs Cross Section Working Group. The simulated Higgs boson samples used for signal samples are summarized in Tab. 5.3.

Concerning the background processes, the Drell-Yan production of $\mathrm{Z}/\gamma^{\star}$ is generated using MADGRAPH5_aMC@NLO at LO with up to 4 additional jets, and the $\mathrm{Z}/\gamma^{\star}$ $p_{\mathrm{T}}$ distribution is reweighted to match the distribution observed in di-muon events from data. The top quark

TABLE 5.3: Simulated samples and cross sections for signal processes.

| Sample | Production mode | Decay mode | $\sigma \times BR$ (pb) |
|---|---|---|---|
| ggH_HWW_m125 | ggH | $H \to W^+W^-$ | 0.9913 |
| qqH_HWW_m125 | VBF | $H \to W^+W^-$ | 0.0896 |
| WplusH_HWW_m125 | VH | $H \to W^+W^-$ | 0.1810 |
| WminusH_HWW_m125 | VH | $H \to W^+W^-$ | 0.1160 |
| ZH_HWW_m125 | VH | $H \to W^+W^-$ | 0.1870 |
| ggZH_HWW_m125 | VH | $H \to W^+W^-$ | 0.1057 |
| ttH_Hww_m125 | ttH | $H \to W^+W^-$ | 0.2120 |
| ggH_Htt_m125 | ggH | $H \to \tau^+\tau^-$ | 2.7757 |
| qqH_Htt_m125 | VBF | $H \to \tau^+\tau^-$ | 0.2370 |
| WplusH_Htt_m125 | VH | $H \to \tau^+\tau^-$ | 0.0532 |
| WminusH_Htt_m125 | VH | $H \to \tau^+\tau^-$ | 0.0341 |
| ZH_Htt_m125 | VH | $H \to \tau^+\tau^-$ | 0.0550 |

processes such as $t\bar{t}$, single top (t-channel and s-channel production), tW and $t\bar{t}V$ (with V=W,Z bosons) are generated either by POWHEGV2 or MADGRAPH5_aMC@NLO, with full spin correlation simulated with MADSPIN [99]. The cross section of different single top quark processes are estimated at NLO accuracy [100], while the $t\bar{t}$ cross section is computed at NNLO accuracy [101].

The non-resonant $q\bar{q} \to W^+W^-$ process is generated by POWHEGV2 at NLO accuracy, and normalized to the cross section calculated at NNLO accuracy [102]. The LO cross section for the gluon induced $W^+W^-$ process, $gg \to W^+W^-$ is obtained directly from MCFM [103], and a factor of 1.4 is applied to correct for the difference between the LO and NLO theoretical calculation [104]. The contribution of the interference between the $gg \to W^+W^-$ and $gg \to H \to W^+W^-$ processes is evaluated using MCFM and is found to be negligible compared to the signal contribution.

The W$\gamma$ and Z$\gamma$ samples are generated by MADGRAPH5_aMC@NLO at LO and NLO accuracy respectively. The sample W$\gamma^\star$ is defined by events with generator-level invariant mass of two leptons (from $\gamma^\star$ decay) between 0.1 GeV and 4 GeV from W$\gamma$ and WZ samples. In addition, a scale factor of 0.94, is derived from an independent study using control region enriched with W$\gamma^\star$ and WZ processes, and applied to the W$\gamma^\star$ samples. On the other hand, the sample Z$\gamma^\star$ is defined by events with non-zero invariant mass of two leptons, at generator-level, from Z$\gamma$ sample.

The di-boson processes such as WZ, ZZ are generated using MADGRAPH5_aMC@NLO at NLO with 1 jet. The normalization of WZ background is eventually taken from a fit to the data and the NLO cross section is used as a prior value in the fit. Other multi-boson processes, such as WZ, ZZ, and VVV (V=W/Z), are generated with MADGRAPH5_aMC@NLO and normalized to the cross section calculated at NLO accuracy. The simulated samples used for background processes are summarized in Tab. 5.4.

TABLE 5.4: Simulated samples and cross sections for background processes. The Single top quark production includes tW process, s- and t-channels production; the tri-boson production, VVV, includes WWW, WWZ, WZZ, ZZZ.

| Process | $\sigma \times BR$ (pb) |
|---|---|
| Z+jets (10 GeV $< m_{ll} <$ 50 GeV) | 18610.0 |
| Z+jets ($m_{ll} >$ 50 GeV) | 6025.2 |
| $t\bar{t} \rightarrow WWb\bar{b} \rightarrow 2l2\nu b\bar{b}$ | 87.310 |
| Single top | 199.84 |
| $q\bar{q} \rightarrow W^+W^-$ | 12.178 |
| $gg \rightarrow W^+W^-$ | 0.5905 |
| $W^+W^- + jets$ | 0.08875 |
| WZ | 64.185 |
| ZZ | 4.996 |
| W$\gamma$ | 405.271 |
| Z$\gamma$ | 131.3 |
| VVV | 0.4180 |

For 2016 data taking year, all processes are generated using NNPDF 3.0 [105] PDF and the set of UE tune, consisting of underlying event parameters is CUETP8M1 [106]. For the 2017 and 2018 run period, the processes are generated using NNPDF3.1 [107] PDF and the UE tune is CP5 [108].

For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package. The distribution of the number of pileup interactions in the simulation is adjusted to match the one observed in data. The average number of pileup interactions was 23 (32) in 2016 (2017 and 2018).

## 5.4 Event selection

Event selection is performed at particle and global event level. The former selection is carried out to define well identified and reconstructed physics objects in order to provide a non ambiguous description for the final state as efficiently as possible, and to reduce any possible source of systematic uncertainties; the latter selection is mainly driving the performance of the analysis by increasing the significance of the signal, and reducing the contamination of the background in the signal region.

The physics objects used in the analysis are defined as described in Sec. 4.4 for leptons, Sec. 4.5 for jets, and Sec. 4.7 for the transverse missing energy. Additional selections are applied to further enhance the purity of the signal samples and to perform specific data driven background estimation.

### 5.4.1 Leptons

The analysis makes use of two WPs to identify leptons for different application:

- the lepton passing tight WP is called **tight lepton**, and is used for analysis-level selection;
- the lepton passing loose WP is called **loose lepton**, and is used for non-prompt background estimation.

For 2016 data taking, both the loose and tight electrons must satisfy the HLT-safe ID selection given in Tab. 4.1. For tight electrons, the requirement to have no missing hits in the reconstructed track further reduces the photon conversion background. For 2017 and 2018 data taking, both the loose and tight electrons must satisfy the cut-based medium ID selection defined in Tab. 4.2, with additional requirements on calorimetric observables ($\sigma_\eta$ and $|1/E_{SC} - 1/p|$) to further reduce photon conversion backgrounds. For the tight electrons, for all the three years data taking, the ISO parameter described in Sec. 4.4.2, computed with cone size $\Delta R = 0.3$, is used to isolate electrons from non-prompt electron backgrounds processes, requiring $I^{rel}_{\Delta R=0.3} < 0.06$. Additionally, the MVA-based ID selection targeting at 90% signal efficiency, MVA-90 ID, which is found to be effective in rejecting background without loosing too much signal efficiency, is used to select tight electron as well.

The charge of the electron mis-assignment is reduced by the requirement that the charge measured by the three methods discussed in Sec. 4.4.2, are in agreement.

Finally, for all three data taking years, the selection for tight electrons is made more stringent by selecting the BDT discriminant `ttHMVA` to be larger than 0.7 (Fig. 4.2), to further reject non-prompt electrons contribution. The electron object definition is summarized in Tab. 5.5.

The tight and loose muons are defined according to the ID selection given in Tab. 4.3. Most non-prompt muon are dominated by heavy-flavor contribution above 10 GeV, which can be easily rejected by a requirement on $I^{rel}_{\Delta\beta}$ (see Eq. (4.6)) with cone size of $\Delta R = 0.4$. The ISO selection $I^{rel}_{\Delta\beta} < 0.15$ is used for tight muon, while a looser selection, $I^{rel}_{\Delta\beta} < 0.4$, is applied for loose muons, which are used in the data-driven fake rate measurement described in Sec. 5.6.1.2. For tight muons, additional impact parameter selections are used to further reduce the background due to the non-prompt muon and photon conversion processes. In particular, a tighter selection on $d_{xy}$ is used for muon $p_{\mathrm{T}}$ below 20 GeV, as this region is more affected by this background.

Similarly to the electron case, the tight muon is further required to pass the selection `ttHMVA` to be larger than 0.8 (see Fig. 4.2). The muon object definition used is summarized in Tab. 5.6.

TABLE 5.5: The object definition for both tight and loose electron ID selection. The isolation and impact parameter values for electron with(without) bracket indicates the cut value for barrel(endcap) region.

| Working point | 2016 | 2017, 2018 |
|---|---|---|
| | $p_T > 10$ GeV<br>$\|\eta\| < 2.5$ | |
| Tight | Tight HLT-safe ID<br>MVA-90 ID<br>$I^{rel}_{\Delta R=0.3} < 0.05880$ (0.0571)<br>ttHMVA $> 0.7$<br>Missing hits $= 0$ (0) | Medium ID<br>MVA-90 ID<br>$I^{rel}_{\Delta R=0.3} < 0.06$ (0.06)<br>ttHMVA $> 0.7$<br>$\sigma_\eta <$ - (0.03)<br>$\|1/E_{SC} - 1/p\| <$ - (0.014) |
| Loose | Loose HLT-safe ID | Medium ID<br>$\sigma_\eta <$ - (0.03)<br>$\|1/E_{SC} - 1/p\| <$ - (0.014) |

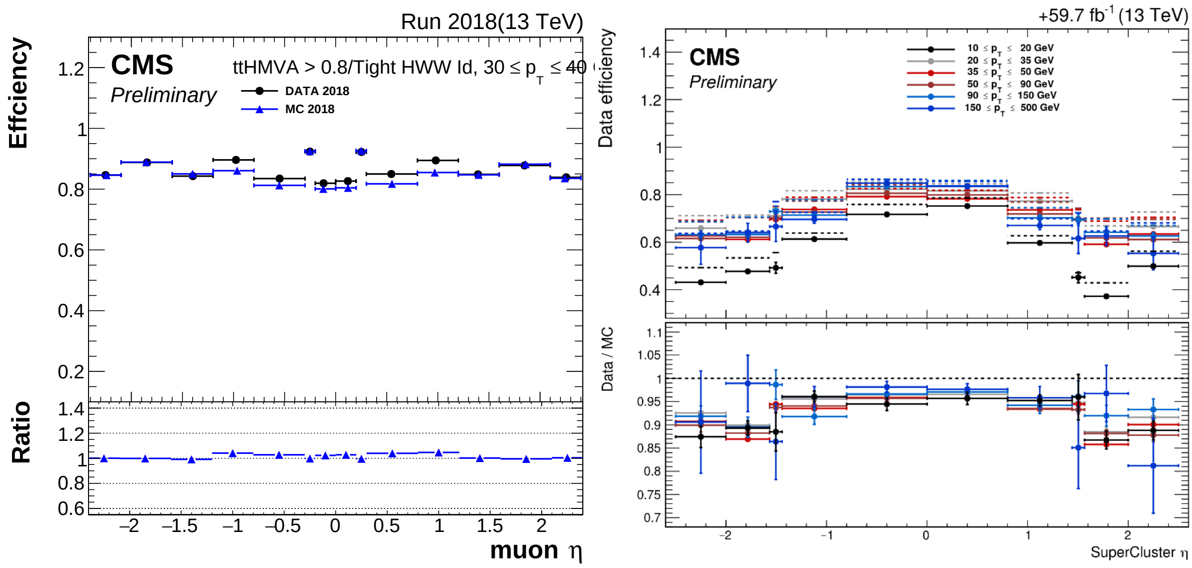
TABLE 5.6: The object selection for Tight and Loose ID muons. The impact parameter value with(without) bracket indicates the cut value for muon $p_T$ <20 ($p_T$ >20) GeV.

| Working point | Full Run-II |
|---|---|
| | $p_T > 10$ GeV<br>$\|\eta\| < 2.4$ |
| Tight | Tight ID<br>$I^{rel}_{\Delta\beta} < 0.15$<br>ttHMVA $> 0.8$<br>$\|d_{xy}\| < 0.01$ (0.02) cm<br>$\|d_z\| < 0.1$ cm |
| Loose | Soft ID<br>$I^{rel}_{\Delta\beta} < 0.40$ |

The ID selection on both electron and muon resulted in different efficiency in data and simulation. These differences are corrected by a scale factor defined as a function of lepton $p_T$ and $\eta$ respectively, shown in Fig. 5.3. The methodology used to derive the lepton reconstruction and identification scale factor is described in Sec. 4.4.8.

### 5.4.2 Jets, b-tagging and missing transverse energy

Jets considered in the analysis are reconstructed by clustering PF-objects using anti-$k_T$ algorithm, with a distance parameter of 0.4 (AK4 PF-jets). The CHS method and JEC are applied as described in Sec. 4.5. The jets are further selected by requiring the tight ID selection discussed in Sec. 4.5. The jet counting event categorization refers to jets with $p_T >$ 30 GeV.




(A) Muon $p_T$ between 30 GeV and 40 GeV

(B) Electron $p_T$ in various range

FIGURE 5.3: Lepton identification and isolation efficiencies in data and simulation (top panels) and data/simulation scale factor (bottom panels), as a function of the lepton $\eta$ in different $p_T$ bins. The data set used for this comparison corresponds to 59.7 fb$^{-1}$ collected at 13 TeV in 2018.

The b-tagging uses the medium WP of DeepCSV algorithm (see Sec. 4.6), corresponding to a b-tagging efficiency of 68% with 1% light-flavor jet mis-identification. The corresponding threshold values of the DeepCSV discriminant are 0.2217 for 2016, 0.1522 for 2017, and 0.1241 for 2018 datasets respectively. The efficiency difference between data and simulation is corrected with the scale factor computed via Eq. 4.13, and applied to the simulated event.

The missing transverse energy is reconstructed as the negative vectorial sum of all PF-based reconstructed particle in an event. The pile-up mitigated, PUPPI $E_T^{miss}$ variable is used, as it provides better resolution (see Fig. 4.8) compared to the $E_T^{miss}$ computed from standard (without pile-up mitigation) PF-jet collection. The spurious contribution to the PUPPI $E_T^{miss}$ is suppressed by applying several filters as stated in Sec. 4.7. Throughout the thesis, the term $E_T^{miss}$ is implicitly referred to the PUPPI $E_T^{miss}$.

## 5.5 Signal region event selection and categorization

The signal region is defined from an optimal event selection chosen to isolate signal candidate events from background events. The signal sensitivity is maximized by sub-categorizing the signal region according to the di-lepton flavor combination and the number of jet, as given Tab. 5.1.

The signal candidate events must contain two leptons passing the tight ID selection, with $p_T > 25$ and $p_T > 20$ GeV respectively, and $|\eta| < 2.5(2.4)$ for electrons(muons). Events containing additional leptons with $p_T > 10$ GeV are rejected. The di-lepton invariant mass is required to be greater than 12 GeV to remove contributions from low-mass resonances and the two leptons must be separated by less than 2 units in pseudorapidity. Events containing a medium b-tagged jet with $p_T > 20$ GeV are removed, to suppress background such as $t\bar{t}$ production. A minimum $E_T^{miss}$ is required to further reduce QCD backgrounds (by about 40%).

Events in 1-jet signal region are required to contain exactly one jet with $p_T > 30$ GeV and $|\eta| < 4.7$. Events in the 2-jets category are required to contain at least two jets with $p_T > 30$ GeV and $|\eta| < 4.7$. For events containing more than two jets, only the two jets with highest $p_T$ are considered. The di-jet jet invariant mass is required to be less than 100 GeV, to reject background processes such as $W^{\pm}W^{\pm}$ vector boson scattering.

The signal regions are further divided into same sign $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ categories. Since the same sign same flavor category may contain contamination from Z+jets background, additional selection is applied in this case by removing events with di-lepton invariant mass within the Z boson mass window of ±15 GeV.

A discriminating variable, called "mass variable" $m_{lljj}$, is computed as the invariant mass of the system consisting of the jet (either the single jet or the di-jet pair) and twice the four-momentum of the trailing lepton, where this double-counting is used to recover the missing momentum from the neutrino in the Higgs boson decay chain. If an event in the 1-jet category contains a second jet with $20 < p_T < 30$ GeV, the jet is included in the variable computation, otherwise the four-momentum of the single jet is used. Finally, the events in all categories are required to have $m_{lljj} > 50$ GeV. The selection criteria are summarized in Tab. 5.7.

TABLE 5.7: Event selection used to define the Signal Region (SR), in different event categories.

| | Preselection | | | |
|---|---|---|---|---|
| Leading and trailing lepton $p_T$ (GeV) | $> 25, 20$ | | | |
| Third lepton $p_T$ (GeV) | $< 10$ | | | |
| Di-lepton invariant mass $m_{ll}$ (GeV) | $> 12$ | | | |
| $\lvert\Delta\eta(l_1, l_2)\rvert$ | $< 2.0$ | | | |
| b-jet veto | DeepCSV medium WP applied to all jets with $p_T > 20$ GeV | | | |
| $E_T^{miss}$ (GeV) | $> 30$ | | | |
| Mass variable $m_{lljj}$ (GeV) | $> 50$ | | | |
| | 1-jet $e\mu$ SR | 2-jet $e\mu$ SR | 1-jet $\mu\mu$ SR | 2-jet $\mu\mu$ SR |
| Jets with $p_T > 30$ GeV | =1 | $\geq 2$ | =1 | $\geq 2$ |
| $\lvert m_{ll} - m_Z\rvert$ (GeV) | - | - | $> 15$ | $> 15$ |
| Di-jet invariant mass $m_{jj}$ (GeV) | - | $< 100$ | - | $< 100$ |

The mass variable distributions after the full selection are shown in Fig. 5.4; the signal processes are scaled by a factor of 30, to emphasize the shape difference between the signal and background samples, revealing the better signal sensitivity in the low mass region for both 1-jet and 2-jets categories. The mass variable binning choice is the same as the one used in the fit, discussed in Sec. 5.8.

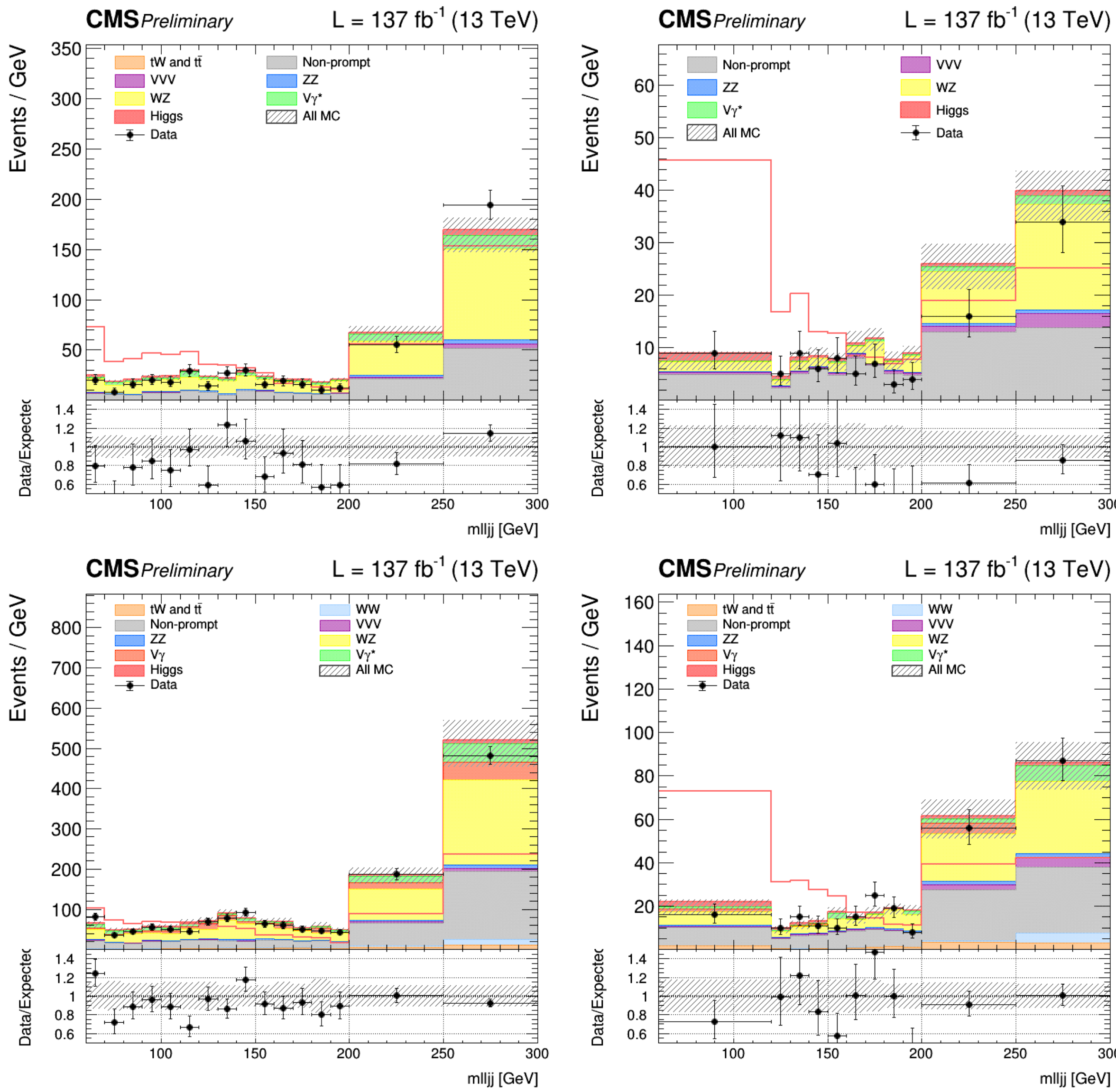



FIGURE 5.4: The mass variable distributions for same flavor $\mu\mu$ (top row) and different flavor $e\mu$ (bottom row) 1-jet (left column) and 2-jet (right column) event category. The signal contribution is scaled by 30 to highlight the shape discriminant feature.

The normalization of the most important background, such as the WZ background, is estimated in the dedicated WZ control region described in Sec. 5.6.3, while the non-prompt backgrounds and and mis-assignment background are estimated by data-driven methods. The details of the background estimation study will be discussed in the next section.

## 5.6 Background estimation

Background events could be directly modeled using simulated samples; however, the rate of object mis-identification may not be accurately predicted by the MC simulation. For example, the non-prompt lepton background where the mis-reconstruction occurs at the tail of the isolation variable distribution, involving jets that pass the lepton isolation requirement: an inadequate modeling of the underlying processes by which jets are mis-identified as leptons, or even a small mis-modeling of the events pile-up can introduce non-negligible differences between simulation and data. Another example of object mis-reconstruction is the electron charge mis-assignment; if the electron charge in events with an opposite sign electron and muon pair is mis-measured, this results into a same sign di-lepton topology.

The background yield and kinematics distribution are estimated using control data samples. In a similar way, the correction factor of electron charge mis-assignment is inferred using data samples and used as a weight to the simulated events.

### 5.6.1 Non-prompt and fake leptons background

The reconstructed leptons in the CMS detector may or may not be real leptons. Those leptons could be reconstructed from jets, photons or from tracks that are matched to unrelated hits in the ECAL or muon chambers. Those non-genuine leptons are called fake leptons; in other cases, there are real leptons produced by physics processes involving heavy quark decays which pass the prompt lepton selection; they are considered as fake leptons as they are not part of the prompt decay of the signal process of interest.

There are different sources of fake leptons, and they are dependent on the lepton type. For electrons, the fake leptons can originate from charged hadrons, photon conversion, and semi-leptonic heavy-flavor decays. For muons, almost all fake muons originate either from semi-leptonic heavy-flavor decays or meson decaying in the flight. In general, the contribution of fake leptons can be controlled by the ISO and ID selection requirements (see Sec. 4.4.5 for muon and Sec. 4.4.2 for electron).

When a jet is mis-identified as a lepton (in particular this is important in the W+jets), an event enters the signal regions with two identified leptons, missing transverse energy, jets and no other significant event characteristics. As a result, these events cannot be suppressed by event selection. In this case, it is necessary to use a data-driven approach to estimate the non-prompt background; this is based on the fakeable object method.

#### 5.6.1.1 The fakeable object method

The fakeable object method [96] is a general technique in which particle level selection criteria are used to suppress background. The method uses loose leptons, which pass or fail tight criteria (the criteria for signal selection) to determine an extrapolation factor relating the background events in the control sample to the background in the signal region. The loose lepton selection is defined in such a way that the amount of non-prompt and fake leptons in a dedicated background control sample (designed to be essentially signal free) is significant, while the tight lepton selection is the same used for the signal extraction in the signal region.

The number of events with two leptons passing the loose selection $N_{ll}$ is given by the sum of $N_{pp}$ events with both prompt leptons, $N_{fp}$ events with one prompt and one fake lepton, and $N_{ff}$ events where both leptons are fake. These numbers are related to the measurable number of these events $N_{t,x}(x = 0, 1, 2)$ with 0, 1, or 2 leptons passing the tight selection:

$$N_{ll} = N_{pp} + N_{fp} + N_{ff} = N_{t,2} + N_{t,1} + N_{t,0} \tag{5.1}$$

$$N_{t,0} = (1-p)^2 N_{pp} + (1-p)(1-f)N_{fp} + (1-f)^2 N_{ff} \tag{5.2}$$

$$N_{t,1} = 2p(1-p)N_{pp} + [f(1-p) + p(1-f)]N_{fp} + 2f(1-f)N_{ff} \tag{5.3}$$

$$N_{t,2} = p^2 N_{pp} + pf N_{fp} + f^2 N_{ff} \tag{5.4}$$

The fake rate, $f$ is defined as the probability that a jet passing loose lepton selection, referred as "fakeable object", also passes tight selection. The fake rate $f$ is measured by counting how many fakeable objects in the control sample pass the tight lepton selection, parametrized as a function of flavor, $p_{\mathrm{T}}$ and $\eta$ of the fakeable object:

$$f = \frac{N_{tight}(\text{flavor}, p_{\mathrm{T}}, \eta)}{N_{loose}(\text{flavor}, p_{\mathrm{T}}, \eta)} \tag{5.5}$$

where the $N_{tight}$, $N_{loose}$ are respectively the number of tight and loose leptons in the event sample. The remaining small contamination from prompt leptons is corrected by subtracting its estimate from simulation both in the numerator and denominator. The fake rate is also sensitive to the flavor and $p_{\mathrm{T}}$ of the quark from which the non-prompt or fake lepton originates and to the hadronic energy in the event. Differences in these variables between the measurement region and the signal region can introduce a bias in the non-prompt estimate and care has to be taken in defining a loose lepton object so that the fake rate is not strongly dependent on them.

The method also needs the measurement of prompt ratio, $p$, defined as the probability that a prompt lepton passing the loose selection, also passes the tight selection. The ratio is measured using a Tag and Probe technique, by counting how many loose passing probes in the Z+jets control sample[3], passes the tight selection.

The set of linear equation, Eq. (5.2), Eq. (5.3), and Eq. (5.4) are inverted to extract the fake lepton contribution, relating the number of prompt and fake leptons with the number of

[3] The control sample consists of Z candidate events, with tag defined as the leading lepton passing the tight selection, and the probe being the loose passing lepton.

tight and loose leptons:

$$\begin{pmatrix} N_{pp} \\ N_{fp} \\ N_{ff} \end{pmatrix} = \frac{1}{(p-f)^2} \begin{pmatrix} f^2 & -f(1-f) & (1-f)^2 \\ -2fp & p(1-f)+f(1-p) & -2(1-p)(1-f) \\ p^2 & -p(1-p) & (1-p)^2 \end{pmatrix} \begin{pmatrix} N_{t,0} \\ N_{t,1} \\ N_{t,2} \end{pmatrix} \tag{5.6}$$

The number of events with two prompt leptons passing both loose and tight selection is thus:

$$N_{pp} = \frac{1}{(p-f)^2}\left[(1-f)^2 N_{t,2} - f(1-f)N_{t,1} + f^2 N_{t,0}\right] \tag{5.7}$$

where every lepton passing the tight selection is weighted by a factor $(1-f)$, while the failing leptons are weighted by $f$. Hence, the number of signal events can be estimated by the first term in Eq. (5.4), $N_{signal} = p^2 N_{pp}$.

In a similar way, both the yield and kinematics distribution of the non-prompt leptons background are estimated by weighting event numbers $N_{t,0}$, $N_{t,1}$ and $N_{t,2}$ respectively according to the Eq. (5.6), given by the second and third terms of Eq. (5.4):

$$N_{non-prompt} = pfN_{fp} + f^2 N_{ff} \tag{5.8}$$

The lepton fake rate measurement is described in Sec. 5.6.1.2; the prediction given by Eq. (5.8) will be used in the validation test of the method discussed in Sec. 5.6.1.3.

#### 5.6.1.2 Lepton fake rate measurement

According to Eq. (5.5), the measurement of lepton fake rate is performed by counting how many fakeable objects in a control region enriched with lepton-like objects pass the tight lepton selection.

A di-jet control region dominated by QCD production processes is defined by selecting events containing one loose lepton and jets, both in simulated and data samples. The single leptonic trigger is used to collect the data events. The loose lepton definition is flavor dependent, and is given respectively in Tab. 5.5 for electrons and Tab. 5.6 for muons, while the jet definition is the one used in the analysis, discussed in Sec 4.5.3.

This QCD enriched sample may contain real leptons originating from W or Z boson leptonic decays, which can bias the lepton fake rate. Leptons from W boson are greatly reduced by selecting events with $E_{\mathrm{T}}^{\mathrm{miss}} < 20$ GeV. The transverse mass variable of W boson $m_{\mathrm{T}}^{\mathrm{W}}$, given by:

$$m_{\mathrm{T}}^{\mathrm{W}} = \sqrt{2p_T^l E_T^{miss}\left(1-\cos\left(\phi_l - \phi_{E_T^{miss}}\right)\right)} \tag{5.9}$$

where $p_T^l$ is the lepton $p_{\mathrm{T}}$, and the azimuthal angle for lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$ are $\phi_l$ and $\phi_{E_T^{miss}}$ respectively, is required to be lower than 20 GeV as well. The contamination from Z+jets process is reduced by requiring exactly one lepton in the event.

The $p_{\mathrm{T}}$ of the jet being mis-identified, referred as near-side jet (the closer to the lepton), is varied by a selection on the $p_{\mathrm{T}}$ of the away-side jet (the recoiling jet). The assumption is that the $p_{\mathrm{T}}$ of the near-side jet is correlated to the $p_{\mathrm{T}}$ of the away-side jet. Therefore, by using the away-side jet as a proxy for the near-side jet, the measurement is unbiased by the presence of the associated reconstructed lepton.

The away-side jet is defined as the jet with the highest $E_{\mathrm{T}}$ separated from the loose lepton with $\Delta R(l, j) > 1$. The jet $E_{\mathrm{T}}$ threshold used for electron fake rate measurement is $E_T > 35$ GeV, while $E_T > 25$ GeV is used for the muon fake rate.

The event selection defining the di-jet control region is summarized in Tab. 5.8.

TABLE 5.8: Event selection used to define di-jet control region.

| Event selection |
|---|
| exactly 1 loose e ($\mu$) with $p_{\mathrm{T}} > 25$ (20) GeV and $\lvert\eta\rvert < 2.5$ (2.4) |
| 1 away-jet jet with $p_{\mathrm{T}} > 10$ GeV and $\lvert\eta\rvert < 2.5$ |
| $\Delta R(l, j) > 1$ with e ($\mu$) with jet $E_{\mathrm{T}} > 35$ (25) GeV |
| $E_{\mathrm{T}}^{\mathrm{miss}} < 20$ GeV |
| $m_{\mathrm{T}}^{\mathrm{W}} < 20$ GeV |

To further reduce the effect from prompt leptons originating from W/Z+jets events, the residual electroweak background contamination is estimated from the simulated samples, and subtracted from both denominator and numerator in Eq. (5.5). The fake rate for electron and muon for each data taking year are shown in Fig. 5.5 and Fig. 5.6 respectively as a function of the lepton $p_{\mathrm{T}}$ and $\eta$. The electroweak contribution subtraction shows a significant effect on the fake rate for both leptons, for $p_{\mathrm{T}} > 20$ GeV.

#### 5.6.1.3 Non-prompt background validation

As a closure test to validate the data-driven method, the estimated non-prompt background predicted by Eq. (5.8) is compared with the observed data in a same sign di-lepton control region. This control region is obtained by the selection shown in Tab. 5.9. In particular, it consists of 0-, 1-, 2-jets categories in the $e^{\pm}\mu^{\mp}$ topology, and the events must satisfy the selection in the trailing lepton $p_{\mathrm{T}}$ orthogonal to the signal region selection given in Tab. 5.7.

The validation is performed in different categories to cover all possible non-prompt background contribution to the analysis. Fig. 5.7 and Fig. 5.8 show the distributions of various kinematical quantities together with the event yields (shown in legend) for each background processes. The transverse mass variable of the Higgs boson $m_{\mathrm{T}}^{\mathrm{H}}$, is defined using Eq. (5.9) with the di-lepton $p_{\mathrm{T}}$ instead. The non-prompt background estimation is in agreement with the observed data within the statistical uncertainties.

The primary source of background from lepton mis-identification (gray color) in the 0-jet category comes from mis-identified W+jets events while for the 1- and 2-jets categories,

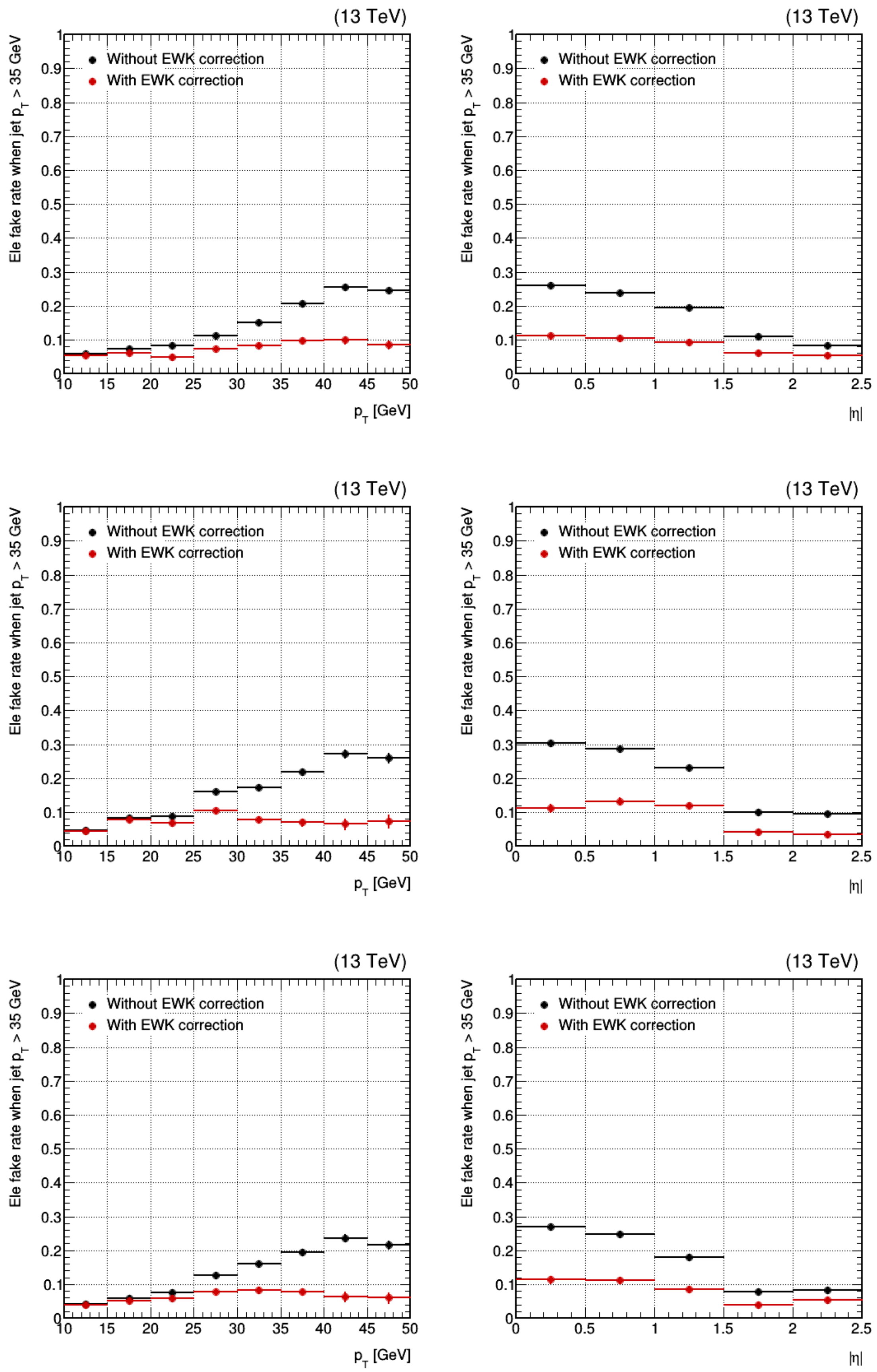



FIGURE 5.5: Electron fake rate as a function of the loose lepton $p_T$ (left) and $\eta$ (right), before and after electroweak correction (in black and red, respectively), from a control region sample obtained with a jet threshold of 35 GeV, and for the full Run-II: 2016 (top), 2017 (middle) and 2018 (bottom).

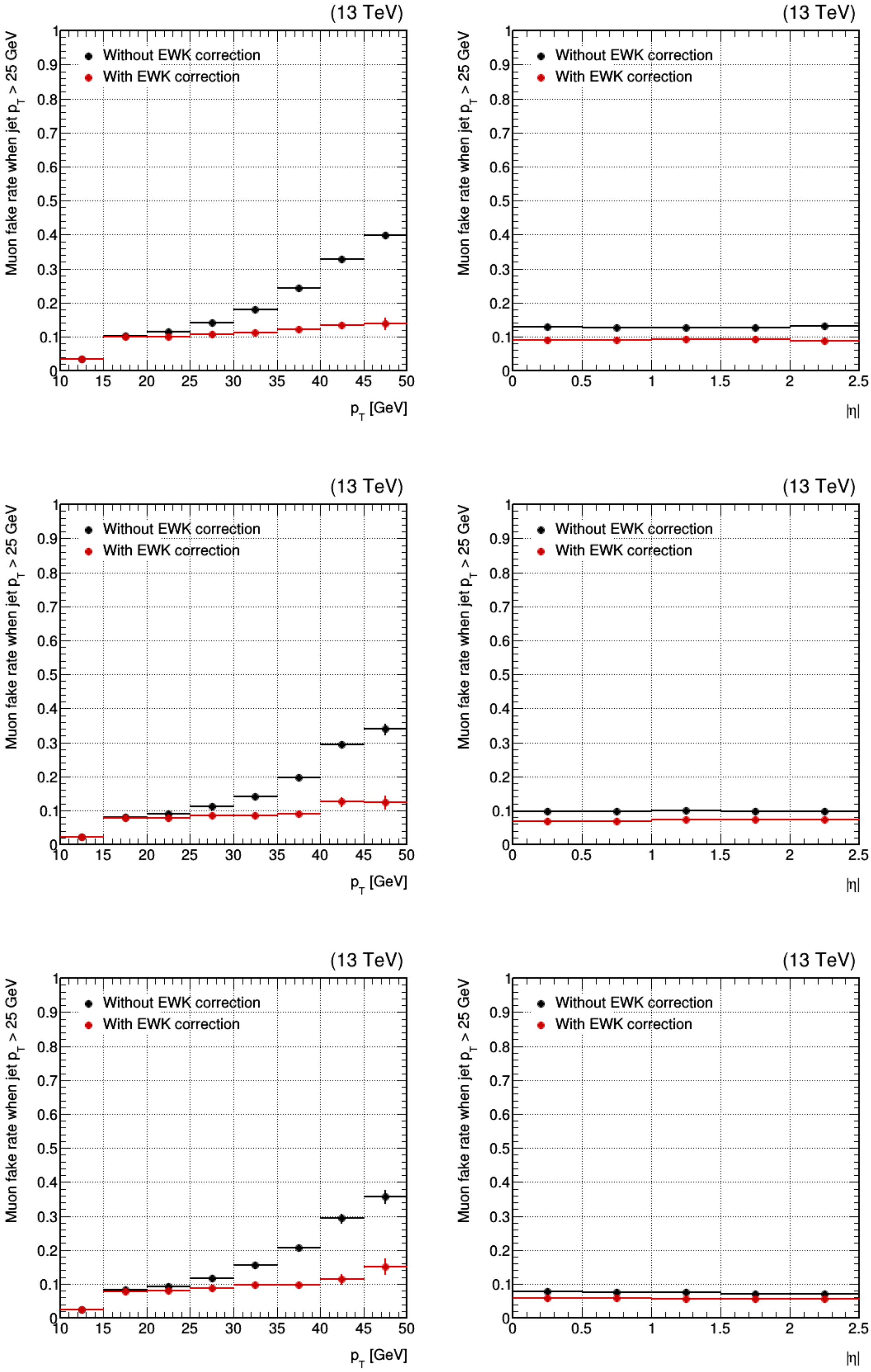



FIGURE 5.6: Muon fake rate as a function of the loose lepton $p_T$ (left) and $\eta$ (right), before and after electroweak correction (in black and red, respectively), from a control region sample obtained with a jet threshold of 25 GeV, and for the full Run-II: 2016 (top), 2017 (middle) and 2018 (bottom).

TABLE 5.9: The event selection used for defining the same sign di-lepton control region for non-prompt background validation.

| Event selection |
|---|
| 2 leptons, veto third lepton with $p_T > 10$ GeV |
| same sign different flavor di-lepton ($e^{\pm}\mu^{\pm}$) |
| leading (trailing) lepton $p_{\mathrm{T}} > 20$ (13) GeV with $\|\eta\| < 2.5$ |
| di-lepton invariant mass > 12 GeV |
| di-lepton $p_{\mathrm{T}} > 30$ GeV |
| transverse mass of W boson $m_{\mathrm{T}}^{\mathrm{W}} > 60$ GeV |
| $E_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV |
| trailing lepton $p_{\mathrm{T}} < 20$ GeV |

semi-leptonic $\mathrm{t\bar{t}}$ decays are the main source of background; sub-dominant backgrounds such as QCD multijet and hadronic top (both processes with two fake leptons) are present as well, but at a much smaller level. Contrary to the non-prompt backgrounds, the fully-leptonic $\mathrm{t\bar{t}}$, tW (yellow color) and Drell-Yan (green color) backgrounds are the source of electron charge mis-assignment, where the electron's charge in the pair of prompt leptons is mis-identified. The details of the electron charge mis-assignment will be discussed in the next section.

### 5.6.2 Electron charge mis-assignment background

The next relevant background to the same sign di-lepton topology originates from the electron charge mis-assignment. The probability of charge mis-identification occurs with values between $10^{-3}$ and $10^{-4}$, depending on $p_{\mathrm{T}}$ and $\eta$ of the electrons. Background processes consisting of a pair of opposite sign electron muon pair ($e^{\pm}\mu^{\mp}$), such as fully leptonic $\mathrm{t\bar{t}}$ decay, non-resonant $W^+W^-$ and leptonic $\tau$ decay of Z boson are susceptible to charge mis-assignment; as a result, these background processes enter into the same sign signal region. To account for potential mis-modeling of charge mis-assignment rate in the MC simulation, the charge flip probability is measured in data and simulation in a sub-sample enriched with Z+jets events, in order to derive a scale factor correcting the simulation prediction for this source of background.

An opposite sign (OS) control region of $Z \rightarrow e^+e^-$ process is defined by requiring two tight electrons, according to the selection reported in Tab. 5.5, with the di-electron invariant mass within 15 GeV around the Z boson mass. In addition, a third lepton is removed if the lepton $p_{\mathrm{T}}$ is above 10 GeV. For the data sample, a double electron trigger is used, and the leading and trailing electron $p_{\mathrm{T}}$ are required to be 5 GeV above the designed trigger threshold (See Sec. 5.3.2), to avoid the trigger turn-on $p_{\mathrm{T}}$ region.

The same sign (SS) control region enriched in events with mis-assigned electron charge is defined by inverting the requirement of the charge lepton pair, keeping the same kinematical selection as for the OS control region. The event selection defining the control regions is

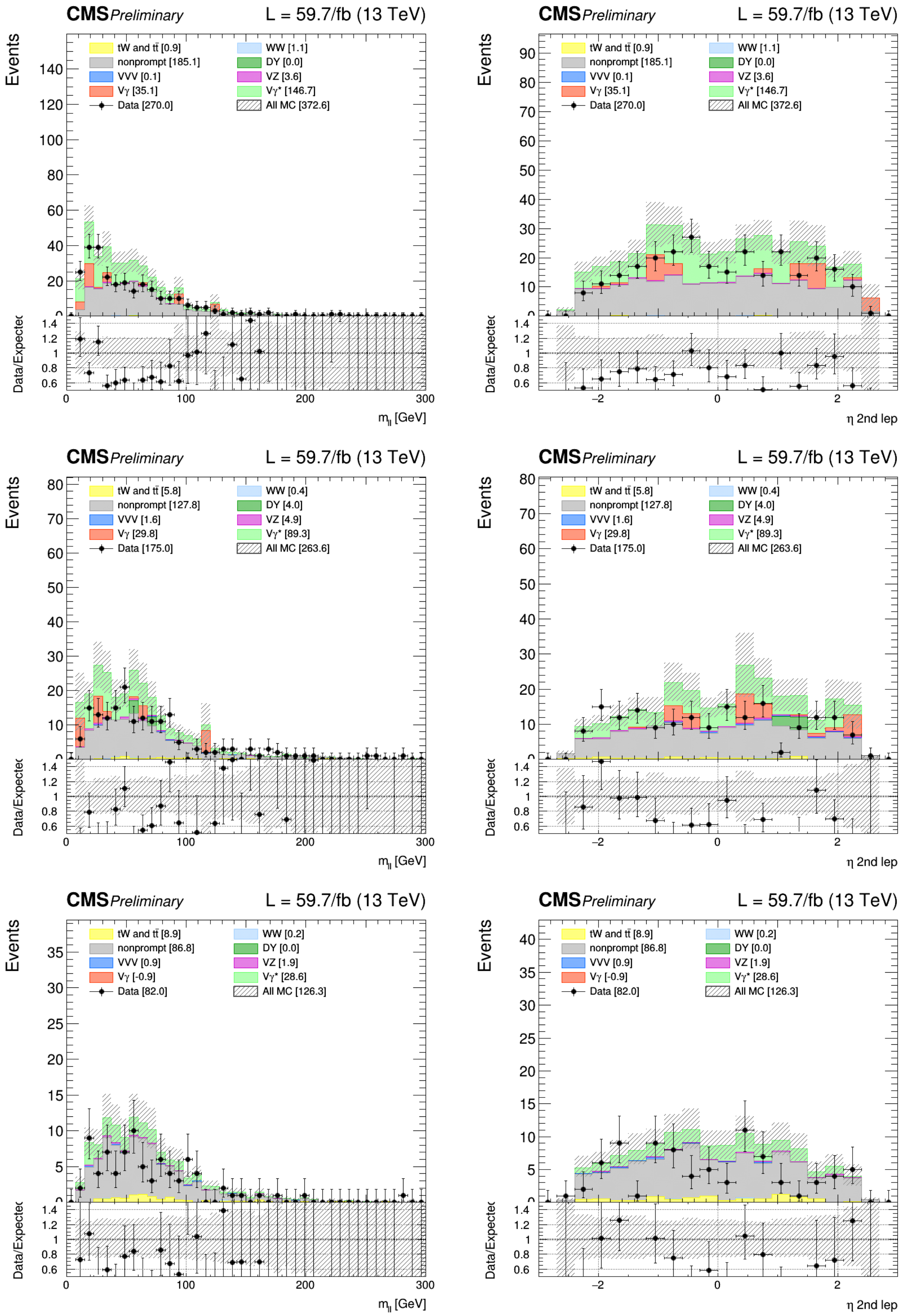



FIGURE 5.7: The invariant mass of dilepton (left) and trailing lepton $\eta$ (right) distributions in the same sign e$\mu$ control region for the data taking year 2018, in three different jet categories (top: 0j, middle: 1j, bottom: $\geq$ 2j). Errors are statistical only.

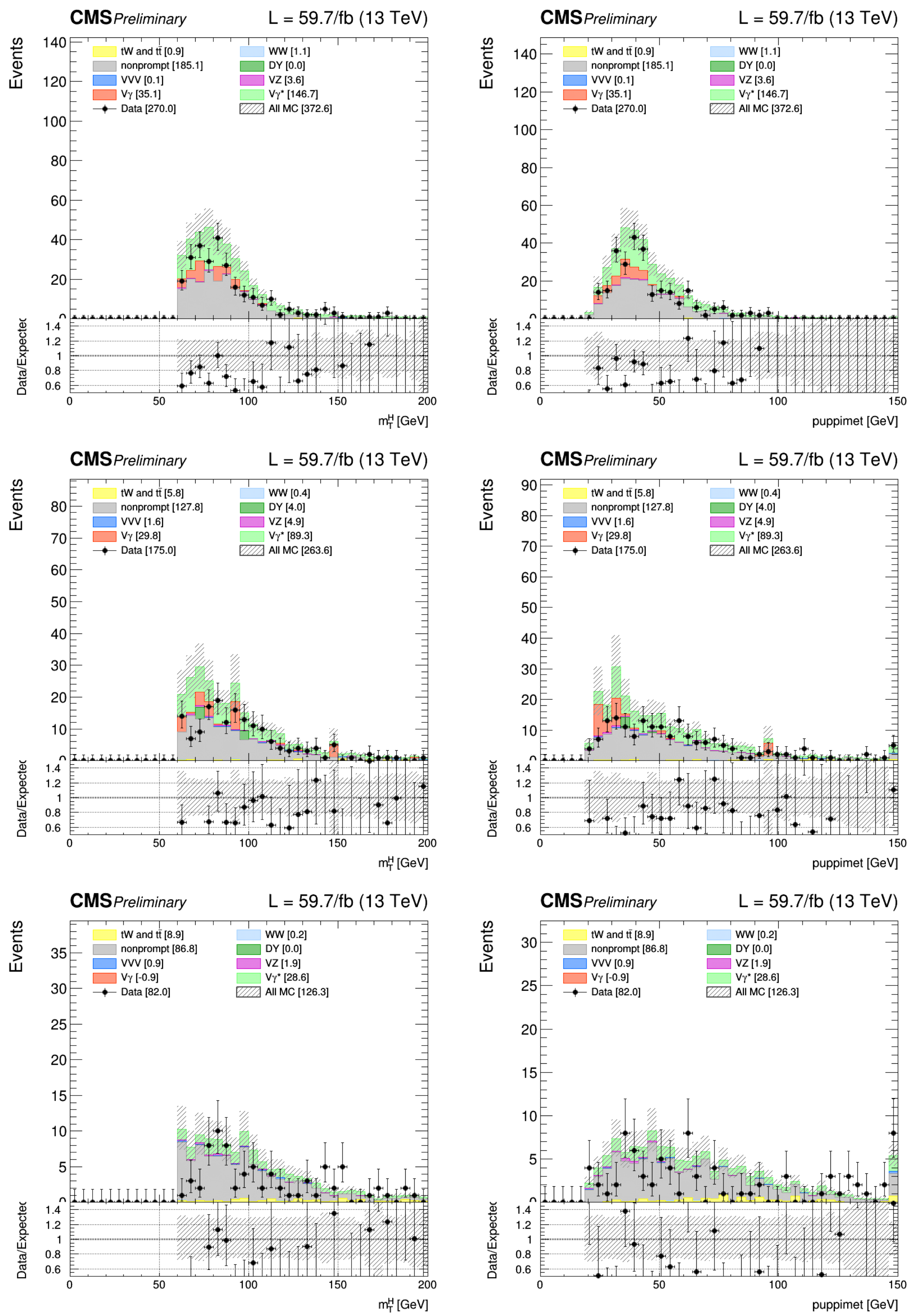



FIGURE 5.8: The transverse mass of the Higgs boson (left) and Puppi $E_T^{\text{miss}}$ (right) distributions in the same sign e$\mu$ control region for the data taking year 2018, in three different jet categories (top: 0j, middle: 1j, bottom: $\geq$ 2j). Errors are statistical only.

summarized in Tab. 5.10; the pseudorapidity distribution of the two electrons in OS and SS samples, after the selection, is shown in Fig. 5.9 and compared with the simulation expectation for the Drell-Yan+jets process. Contributions from other processes are very small in both regions and thus neglected. The discrepancy between data and simulation observed in the SS sample is due to the fact that the charge mis-assignment effect is not properly simulated in the MC. A measurable quantity, called the mis-charge rate $\epsilon^{pair}_{mis-charge}$,

TABLE 5.10: The event selection for SS and OS control regions.

| Event selection |
|---|
| 2 electrons, remove third lepton with $p_T > 10$ GeV |
| 2 electrons with $p_{\mathrm{T}}$ selected at 5 GeV above HLT $p_{\mathrm{T}}$ threshold |
| $\left\|m_{ee} - 91.2\right\| < 15$ GeV |
| OS or SS requirement on di-electron |

defined by the ratio between the number of events $N_{OS}$ and $N_{SS}$ respectively from OS and SS sub-samples, as a function of $p_{\mathrm{T}}$ and $\eta$ of the di-electrons pair:

$$\epsilon^{pair}_{mis-charge} = \frac{N_{SS}(p_T, |\eta|)}{N_{OS}(p_T, |\eta|)} \tag{5.10}$$

is related to the probability for the mis-charge measurement of a single electron according to the following equation:

$$\frac{N_{SS}(p_T, |\eta|)}{N_{OS}(p_T, |\eta|)} = \frac{\epsilon_{cf}(p_T^{(1)}, |\eta^{(1)}|) \cdot \left(1 - \epsilon_{cf}(p_T^{(2)}, |\eta^{(2)}|)\right) + \epsilon_{cf}(p_T^{(2)}, |\eta^{(2)}|) \cdot \left(1 - \epsilon_{cf}(p_T^{(1)}, |\eta^{(1)}|)\right)}{1 - \left[\epsilon_{cf}(p_T^{(1)}, |\eta^{(1)}|) \cdot \left(1 - \epsilon_{cf}(p_T^{(2)}, |\eta^{(2)}|)\right) + \epsilon_{cf}(p_T^{(2)}, |\eta^{(2)}|) \cdot \left(1 - \epsilon_{cf}(p_T^{(1)}, |\eta^{(1)}|)\right)\right]} \tag{5.11}$$

where $\epsilon_{cf}$ is the electron charge flip probability for the first and second electron pair in the region, indicated by the superscript (1) and (2) on their kinematics variables. The numerator represents the total probability of two mutually exclusive events, accounting for the probability for the first electron's charge is flipped but the second's is not flipped, and vice versa. The denominator is defined as the complementary probability for the occurrence of OS events.

The measurement is performed in two different $p_{\mathrm{T}}$ and $\eta$ bins of the two electrons:

- electron $|\eta| \in \{ [\ 0\ ,\ 1.4\ ], [\ 1.4\ ,\ 2.5\ ] \}$;
- electron $p_{\mathrm{T}} \in \{ [\ 17\ ,\ 20\ ], [\ 20\ ,\ 200\ ] \}$

and the lower bound of electron $p_{\mathrm{T}}$ is determined by the double lepton trigger $p_{\mathrm{T}}$ threshold at the plateau region. Hence, there are in total 16 equations corresponding to Eq. (5.11), one for each combination of $p_{\mathrm{T}}$ and $\eta$ of the two electrons.

The mis-charge rate may be biased due to the presence of other small backgrounds under the Z boson mass resonance. The residual electroweak background is eliminated by fitting

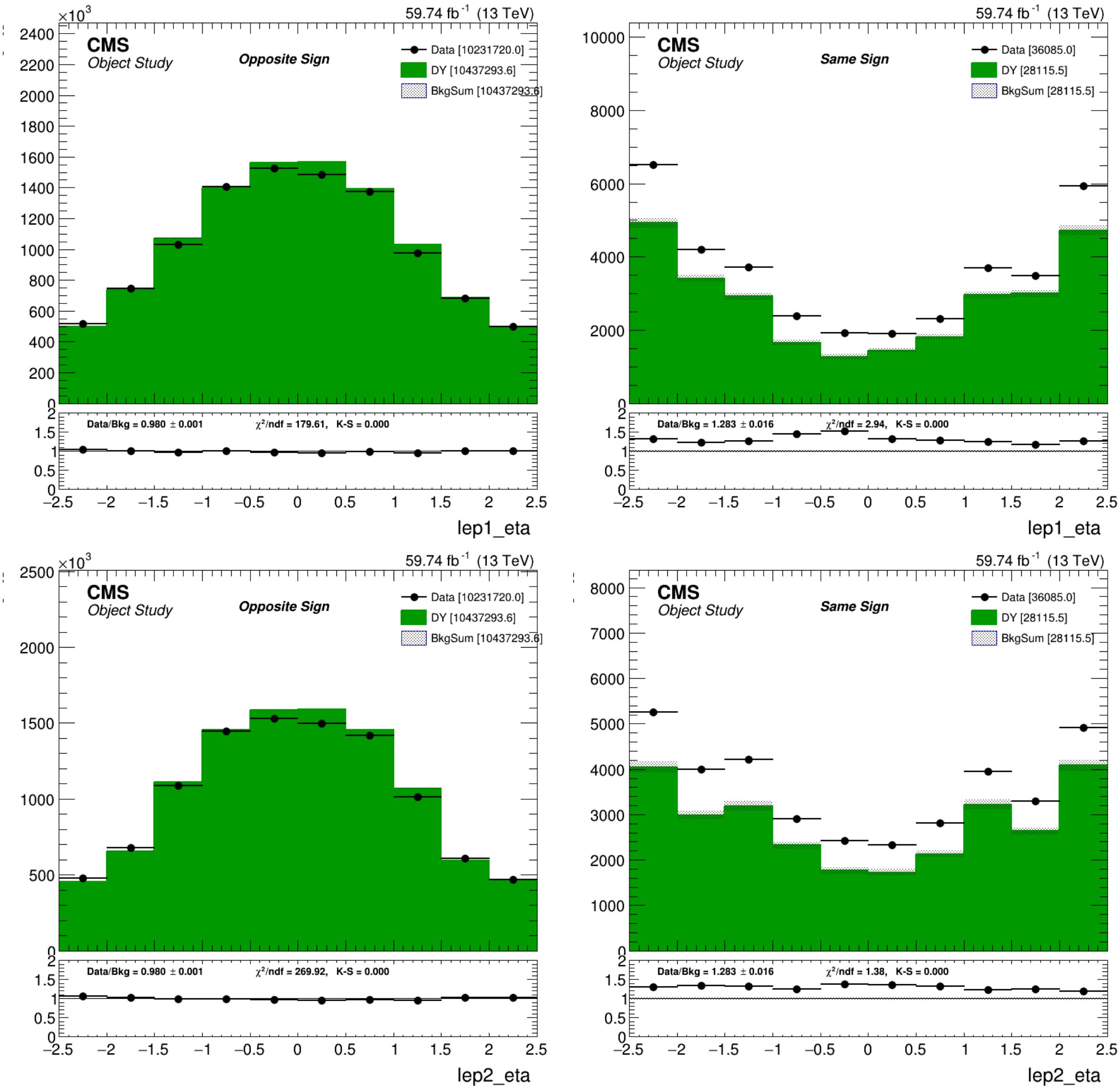



FIGURE 5.9: The leading (top row) and trailing (bottom row) electron pseudorapidity distribution in the OS (left column) and SS (right column) di-electron control samples, after the event selection. Errors are statistical only.

the di-lepton invariant mass to the Z boson mass resonance for both SS and OS regions, and the extracted number of signal OS and SS events is used for the computation of $\epsilon^{pair}_{mis-charge}$. The fitting function is given by a signal Probability Distribution Function (pdf) convoluting a Breit-Wigner and a Gaussian function, superimposed to a background pdf given by an exponential function. The fits to the invariant mass distributions of di-electron for OS and SS regions are shown in Fig. 5.10 for the 2018 data taking.

The quantities $\epsilon^{pair}_{mis-charge}$ are measured by fitting the mass distribution in the different $p_{\mathrm{T}}$, $\eta$ bins using the least square fit method, and the individual charge flip probabilities are extracted using Eq. (5.11). These probabilities are measured both in data and simulation, and separately for the three data taking years; the ratio between the two defines the scale

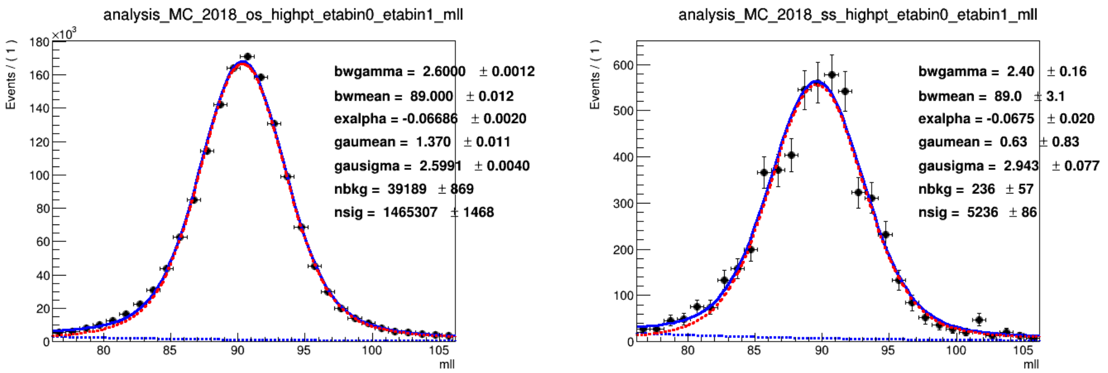



FIGURE 5.10: The fit to the invariant mass distribution of the di-electron for OS (left) and SS (right) samples.

factor used to account for the difference between data and simulation. The results are reported for electrons with $p_{\mathrm{T}}$ between 20 and 200 GeV in Tab. 5.11.

TABLE 5.11: The charge flip probabilities for data and simulation, for both leading and trailing electron with 20 GeV $< p_{\mathrm{T}} <$ 200 GeV, for different $\eta$ regions.

| $\lvert\eta\rvert$ range | $\epsilon_{cf}$ (DATA) | $\epsilon_{cf}$ (MC) | SF |
|---|---|---|---|
| 2016 dataset | | | |
| $0 < \lvert\eta\rvert < 1.5$ | $(5.60 \pm 0.11) \times 10^{-4}$ | $(5.82 \pm 0.11) \times 10^{-4}$ | $0.96 \pm 0.03$ |
| $1.5 < \lvert\eta\rvert < 2.5$ | $(8.44 \pm 0.01) \times 10^{-3}$ | $(7.52 \pm 0.07) \times 10^{-3}$ | $1.12 \pm 0.02$ |
| 2017 dataset | | | |
| $0 < \lvert\eta\rvert < 1.5$ | $(5.29 \pm 0.10) \times 10^{-4}$ | $(3.68 \pm 0.11) \times 10^{-4}$ | $1.44 \pm 0.04$ |
| $1.5 < \lvert\eta\rvert < 2.5$ | $(4.52 \pm 0.05) \times 10^{-3}$ | $(3.32 \pm 0.05) \times 10^{-3}$ | $1.36 \pm 0.02$ |
| 2018 dataset | | | |
| $0 < \lvert\eta\rvert < 1.5$ | $(5.63 \pm 0.09) \times 10^{-4}$ | $(3.86 \pm 0.07) \times 10^{-4}$ | $1.46 \pm 0.02$ |
| $1.5 < \lvert\eta\rvert < 2.5$ | $(4.85 \pm 0.04) \times 10^{-3}$ | $(3.77 \pm 0.04) \times 10^{-3}$ | $1.29 \pm 0.01$ |

As a consistency check, the derived scale factors are applied as per event weight on the same sign simulated sample. The resulting pseudorapidity distributions for leading and trailing electrons are shown in Fig. 5.11.

### 5.6.3 WZ control region and background normalization

The WZ production process, which is by far the most important source of irreducible background, may not be correctly predicted by the simulation. Moreover, the simulated background originating from $W\gamma^{\star}$ production process may not be able to cover the low di-lepton mass region, since there is a generator level requirement of $m_{\gamma^{\star}} > 4$ GeV (see Sec. 5.3.3). Hence, there could be some appreciable rate of events below this threshold passing the selection criteria of the analysis. It is then crucial to have a reliable prediction of these

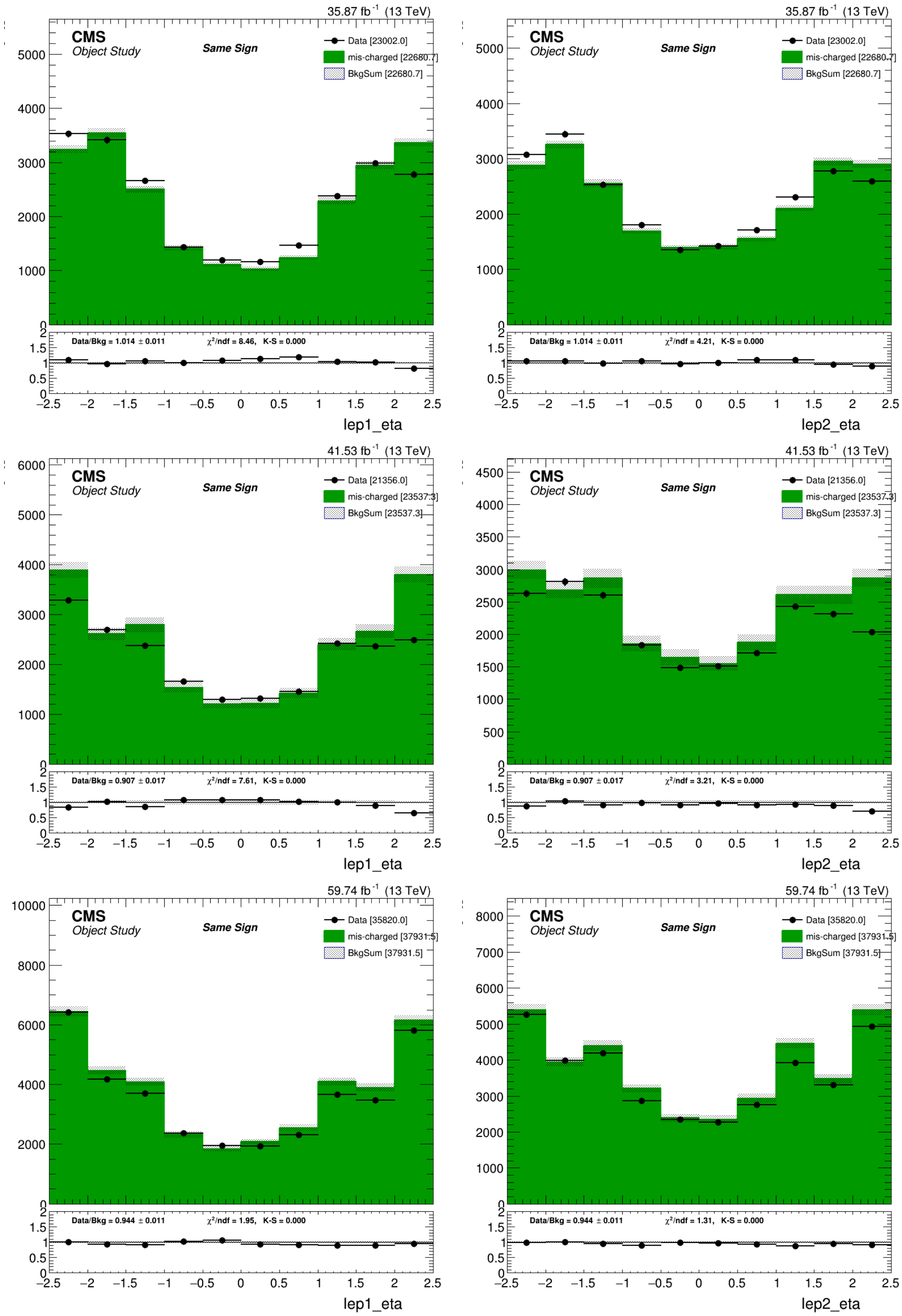



FIGURE 5.11: The leading (left column) and trailing (right column) electron pseudorapidity distribution in the SS di-electron control regions for 2016 (top row), 2017 (middle row) and 2018 (bottom row) data taking year. Errors are only statistical in nature.

backgrounds cross section, validating the simulation by using data in a dedicated control region.

The control regions enriched with WZ process requires at least three leptons identified by the tight ID selection given in Tab. 5.5 and Tab. 5.6. A fourth lepton is allowed if the lepton $p_T$ is smaller than 10 GeV. In addition, the invariant mass of the tri-lepton is required to be above 100 GeV and the absolute value of the total charge must be 1. In addition, the pair of oppositely charged di-lepton is selected to be within the Z boson mass window. The control region is categorized by the number of jets as for the same sign signal region categorization. The event selection is summarized in Tab. 5.12.

TABLE 5.12: The event selection defining the WZ control region.

| Event selection |
| --- |
| 3 leptons, with third lepton $p_T > 15$ GeV |
| veto fourth lepton with $p_T > 10$ GeV |
| tri-lepton invariant mass, $m_{lll} > 100$ GeV |
| total charge of 3 leptons, $ch_{lll} = \pm 1$ |
| Z mass window, ±25 GeV |
| categorized into 1-, 2-jets |

The invariant mass of Z boson is reconstructed by selecting events consisting of two opposite sign leptons, as shown in Fig. 5.12. The kinematical properties of the WZ background are well reproduced by the simulation; the WZ events yield is fitted simultaneously with the signal yield in the signal region, as discussed in Sec. 5.8.

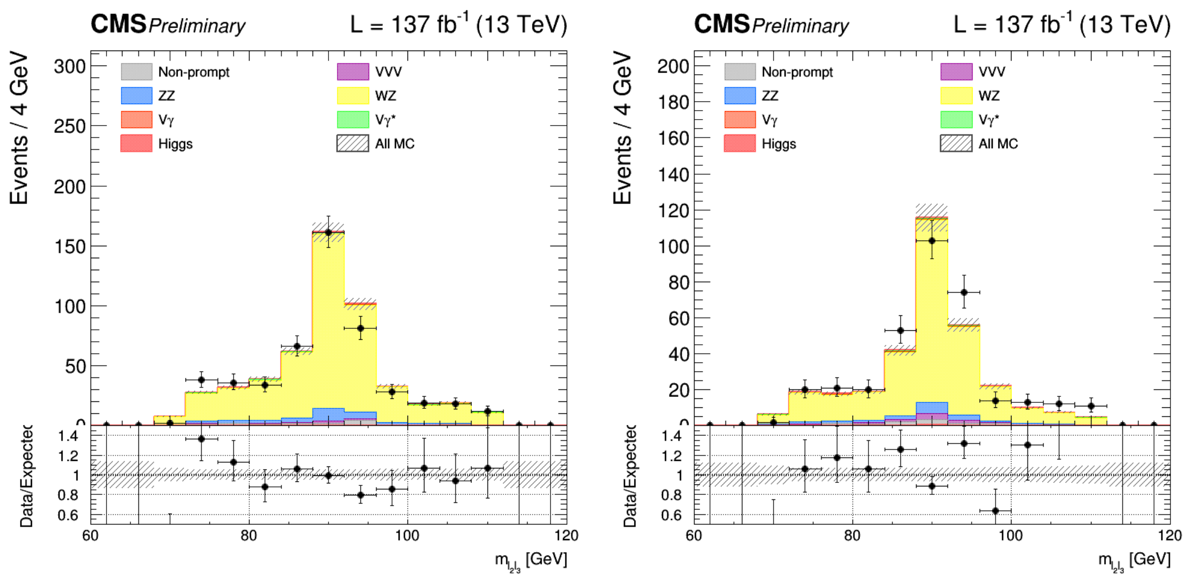



FIGURE 5.12: Invariant mass distributions for the opposite sign di-lepton in the WZ control region, in 1-jet (left) and 2-jets (right) category after the fit to data. The signal is overlaid for shape comparison. The uncertainty band is determined by sampling the posterior nuisance parameter covariance matrix.

## 5.7 Systematic uncertainties

The systematic uncertainties in the analysis play an important role, because there is no expected pronounced Higgs mass peak structure in the $m_{lljj}$ distributions. The systematic uncertainties arise from three sources: background estimation, experimental measurements, and theoretical uncertainties. The systematic uncertainties are implemented as nuisance parameters in the fit to extract the signal strength. Each uncertainty may be modeled by one or several nuisance parameters, corresponding to specific samples.

The systematic uncertainties may affect the measured signal strength and the background cross section in number of different ways, such as the normalization, the shape of the distributions or both. The normalization uncertainties are represented by a constant factor, while the shape uncertainties are described with a set of histograms corresponding to the nominal shape template, $+1\sigma$ and $-1\sigma$ variation value, with the same normalization. On the other hand, systematic effects due to experimental uncertainties are studied by applying a scaling and smearing of the variables related to physics objects, followed by a subsequent recalculation of all the correlated variables.

All experimental sources will be discussed in Sec 5.7.2. For backgrounds with a data driven normalization estimation, only the shape uncertainty is considered. A brief description of each source of systematic uncertainty is discussed in the following sections.

### 5.7.1 Uncertainties on the background modeling

The main systematic uncertainties affecting the backgrounds are:

- **non-prompt background**: the non-prompt background is modeled by applying weights (fake scale factors) to events passing a loose lepton selection. Systematic uncertainties in fake scale factors arise from the limited size of the data samples and the difference in the flavor composition of the jets faking the leptons between the measurement sample and the signal region. The maximum deviation of the fake scale factors from the nominal values is between 5 to 10%. To account for the differences related to the flavor of the non-prompt lepton sources, a conservative 30% uncertainty in the background normalization is included, which is correlated among the three years. Uncertainties on the fake scale factors are uncorrelated among the datasets from different years, but correlated between categories.
- **electron charge mis-identification in Drell-Yan, WW, $\mathrm{t\bar{t}}$, tW backgrounds**: these backgrounds contains a pair of opposite sign leptons with at least one electron in the final state, and mis-charge weights are applied to account for the electron charge flip probability. Uncertainties in the charge mis-measurement is estimated by varying the nominal value by $\pm 1$ standard deviation.

- **top quark and $W\gamma^{\star}$ backgrounds**: these backgrounds receive specific weights to better match the simulation prediction with the observation from the collision data:
  - the $\mathrm{t\bar{t}}$ and single top processes are treated as one single background in the fit, the uncertainty in the composition of the top background is given by the ratio of respective cross sections, which is of the order of 8%;
  - the $p_{\mathrm{T}}$ of $\mathrm{t\bar{t}}$ events is reweighted to NNLO precision, the uncertainty is computed by the difference between before and after applying the reweighting;
  - the $\mathrm{W}\gamma^{\star}$ background is derived from WZ simulated sample, and weighted with a scale factor that carries an uncertainty of 25%.

  All these uncertainties are correlated among the three years.

### 5.7.2 Experimental uncertainties

This section describes the uncertainties in MC modeling. All uncertainties under this category are uncorrelated among the three years unless otherwise stated. The following experimental systematic sources have been taken into account:

- **Integrated luminosity**: the uncertainty in the measured luminosity is determined from the luminosity calibration obtained from the van der Meer scans [109]; it amounts to 2.5%, 2.3%, and 2.5% for 2016 [110], 2017 [111], and 2018 [112] datasets, respectively. This uncertainty is partially correlated among the three years.
- **Trigger efficiency**: the efficiencies of the used triggers (see Tab. 5.2) were computed using Tag and Probe method selecting Z boson decay events (see Sec. 4.4.1). For both electron and muon, the associated uncertainties are obtained by varying the "tag" selection and the Z boson mass window and are uncorrelated among the three years; they are estimated as less than 1% for each.
- **Lepton reconstruction and identification efficiency**: the lepton reconstruction and identification efficiencies are measured with the Tag and Probe method on data. To correct for the difference in the lepton identification efficiencies between data and MC, the scale factors dependent on $p_{\mathrm{T}}$ and $\eta$ are applied to the MC simulation (see Sec. 5.4.1). The resulting uncertainty is $\sim$1% for electrons and $\sim$2% for muons.
- **Muon and electron momentum scale and resolution**: these uncertainties arise due to limited knowledge of the detector material and magnetic field and are $p_{\mathrm{T}}$ and $\eta$-dependent. The associated uncertainties on both the scale and resolution individually amount to $\sim$0.6-1% for electrons and $\sim$0.2% for muons.
- **Jet energy scale uncertainties**: the effect of the Jet Energy Scale (JES) uncertainties are propagated to the final $m_{lljj}$ distributions by shifting the JES, and accordingly the

$E_{\mathrm{T}}^{\mathrm{miss}}$, by one standard deviation. The fit templates are modified both in shape and normalization as a result of this variation. The JES uncertainties affect the background normalization in the signal region at the level of ∼10%. The JES uncertainties are partially correlated between years.

- **Residual $E_{\mathrm{T}}^{\mathrm{miss}}$ uncertainties**: uncertainties in $E_{\mathrm{T}}^{\mathrm{miss}}$ that are not clustered in jets or identified as leptons are assessed by varying the momentum of these PF candidates. The effect on the rates in the signal region is of the order of 1% to 10% depending on the background.
- **b-tagging scale factors**: uncertainties on the b-tagging scale factor related to theoretical modeling from the scale factor measurement are correlated among the datasets, while the remainder, arising from statistical uncertainties, are uncorrelated. The effect amounts to ∼1-2% depending on the process.
- **Pileup**: uncertainties on the amount of simulated number of additional pp interaction are assessed by varying the inelastic cross section used to generate the pileup distribution by 5% [113].
- **Prefiring**: Inefficiencies due to the L1 trigger prefiring in 2016 and 2017 were corrected in MC simulation using additional trigger efficiency scale factors. The corresponding uncertainties are uncorrelated between the two years.

### 5.7.3 Theoretical uncertainties

Theoretical uncertainties on the cross section and kinematical properties of relevant background processes generally arise from missing higher-order QCD corrections and PDF uncertainties. These uncertainties affect both the event yields and the mass variable shape of the background predictions, and the signal shape. All these uncertainties are correlated among the three years unless otherwise stated.

- **Parton distribution function**: the PDF affects the shapes and the overall cross sections of the backgrounds. Since the shape variation within the signal region is found to be small, the former is translated into effects on the geometrical acceptance. The variations of the PDF, computed following the prescription of the LHC Higgs cross section working group [114], are used to evaluate the PDF uncertainties of the Higgs boson signal. The PDF uncertainties for the background processes are evaluated by reweighting the MC simulation with different PDF sets based on parton-level kinematics, and propagating the effect on the normalization. The overall rate differences of these uncertainties is not considered for backgrounds whose normalization is constrained in a control region.
- **Renormalization and factorization scales**: missing higher-order terms in the perturbation series will also affect the shapes and the overall cross sections. For the

background processes, MC simulation events are reweighted by varying the renormalization and factorization scales up and down by a factor of two. The uncertainty is taken from the envelope of the varied distributions as one standard deviation variations. Uncertainties on Higgs boson production processes are split into cross section and acceptance. The former is provided by the LHC Higgs cross section working group [114], the latter is calculated by means of event reweighting. The overall rate effect of these uncertainties is not considered for backgrounds whose normalization is constrained in control region.

- **Parton shower modeling**: uncertainty in PS modeling primarily affects the jet multiplicity. The PS uncertainty is parametrized as a function of the number of quarks in the hard process, using the weights computed in PYTHIA8.
- **Underlying event modeling**: uncertainty associated to the UE model is estimated by shifting the respective PYTHIA8 UE tunes (see Sec. 5.3.3) by one standard deviation. This uncertainty is 1.5% and it is uncorrelated between 2016 and the other two years, but correlated between 2017 and 2018.

### 5.7.4 Statistical uncertainty

The uncertainties due to the finite statistics of the simulated samples are estimated by using the Barlow and Beeston approach [115]; a single nuisance parameter is used to scale the sum of the process yields in each bin, and varied within its total statistical uncertainty. The statistical uncertainty is uncorrelated among each bins, backgrounds and categories for all three years.

## 5.8 Statistical fitting method

The analysis uses a template method [116] that exploits the shape difference in the distributions of a given observable, called the fit variable, between the background and signal. The templates are derived from MC simulations or from control regions of the data. The background and signal templates are then fit to the observed data.

In this analysis, the mass variable defined in Sec. 5.5 is used as fit variable. Finer binning is used for 1-jet category to optimize the signal to background ratio, while a coarser binning is used in the 2-jet category to reduce the statistical fluctuation. The binning for the templates is given by:

- 1-jet, $mlljj \in [60,70,80,90,100,110,120,130,140,150,160,170,180,190,200,250,300]$ GeV

- 2-jet, $mlljj \in [60,120,130,140,150,160,170,180,190,200,250,300]$ GeV

A linear combination of signal and background templates is used to describe the total expected yield for a given $i^{th}$ bin in a histogram:

$$\nu_i = \mu_{WH} \cdot s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}) \tag{5.12}$$

where $s_i$ and $b_i$ are the number of expected events for signal and background components, $\mu_{wh} = \sigma_{wh}/\sigma_{SM}$ is the signal strength multiplier, the ratio between the measured associated WH production cross section and its SM expectation, and $\boldsymbol{\theta}$ is the set of nuisance parameters used to incorporate the systematic uncertainties affecting the signal and background processes.

The signal template is given by the sum of the contributions of the different Higgs production mechanisms described in Tab. 5.3, as predicted by the SM; the background template is defined by the total contribution of each background components described in Tab. 5.4, after the full analysis selection. Each source of systematic uncertainty outlined in Sec. 5.7 are considered as nuisance parameters in the fit.

The likelihood function in the $i^{th}$ bin of a $N_{bin}$ distribution is based on the Poisson probability:

$$L\left(data|\nu(\boldsymbol{\theta})\right) = \prod_{i \in N_{bin}}^{N_{bin}} \frac{(\mu_{WH,i} \cdot s_i + b_i)^{n_i}}{n_i!} e^{-(\mu_{WH,i} \cdot s_i + b_i)} \tag{5.13}$$

where $\nu_i$ is the total expected events yield given by Eq. (5.12) and $n_i$ is the number of observed event in the $i^{th}$ bin. The likelihood $L_{tot}$ is constructed by taking the product of all bins $N_{bin}$ from all event categories $N_{category}$ and for each year $N_{year}$:

$$L_{tot}\left(data|\mu_{wh}, \boldsymbol{\theta}\right) = \prod_{k \in N_{year}}^{N_{year}} \prod_{j \in N_{category}}^{N_{category}} \prod_{i \in N_{bin}}^{N_{bin}} L_{i,j,k}\left(data|\nu_{i,j,k}(\boldsymbol{\theta})\right) \cdot p(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}) \tag{5.14}$$

where $p(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})$ are the prior PDFs of the nuisance parameters $\boldsymbol{\theta}$ in the fit with the corresponding set of default values $\tilde{\boldsymbol{\theta}}$ [117]. For most nuisance parameters, a log-normal prior distribution is used, with a standard deviation corresponding to the given systematic uncertainty; some of the experimental uncertainties related to the shape of signal and background processes are modeled by means of shape variations as explained in Sec. 5.7.2.

## 5.9 Results

The signal and background event yields, and the number of observed events in data, for each channel after the full event selection are shown in Tab. 5.13. The uncertainties in the expected yields are obtained with the initial value of the systematic uncertainties, whereas the posterior ones are obtained with the fitted values of the nuisance parameters.

The signal strength modifier $\mu_{wh}$, is extracted through a maximum likelihood fit to the observed data, where the contributions of the different signal samples are multiplied by the same signal strength modifier, using all the production modes and both the $H \rightarrow W^+W^-$ and $H \rightarrow \tau^+\tau^-$ decays. The normalization of the WZ background is unconstrained in the fit, and measured in the data control regions. The measured signal strength combining all categories, for the three years combined is:

$$\mu_{wh} = 0.89^{+0.64}_{-0.62}(stat)^{+0.74}_{-0.78}(exp)$$

corresponding to an observed significance of 0.9 $\sigma$ under the background-only hypothesis. The total likelihood is profiled as a function of signal strength modifier by maximizing, or specifically minimizing the log-likelihood function, $-\ln L(\mu, \boldsymbol{\theta})$. The 68% and 95% confidence intervals are given by the condition $-2\Delta lnL = 1(4)$, which can be observed graphically in Fig. 5.13.

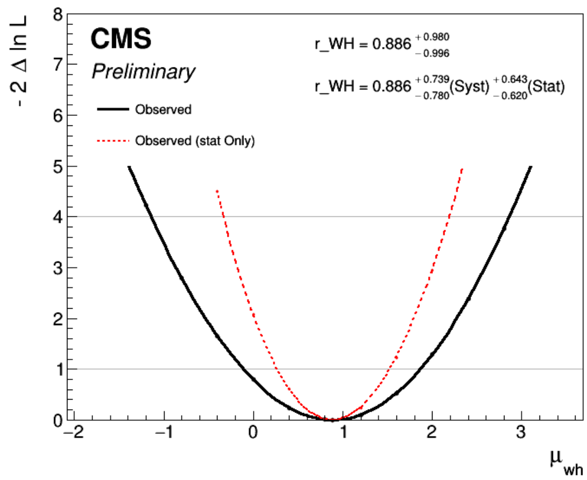



FIGURE 5.13: Observed likelihood profiles as a function of the signal strength. Red dashed curves include only statistical uncertainties. The 68% and 95% confidence intervals defined by $-2\Delta lnL = 1$ and 4 respectively are shown for reference.

A simultaneous fit is also performed in which separate signal strength modifiers are applied for each years, and left unconstrained. The expected and observed results for each year are shown in Tab. 5.14. The expected result is computed from a fit to the Asimov dataset [118], which provides an estimation of the expected signal sensitivity. The mass variable distributions after the fit to the data for full Run-II dataset are shown in Fig. 5.14, and separately for each year in Fig. 5.15 for different flavor and in Fig. 5.16 for same flavor.

TABLE 5.13: The predicted signal and background yields compared to number of observed events in different event categories after the full event selection. The yields are given in the format posterior (a priori) (± posterior uncertainty), where the posterior values come from a combined fit to data. The individual signal yields are given for different production mechanisms. Correlations between uncertainties are accounted for in the total uncertainty shown.

| Process | $e^{\pm}\mu^{\pm}$ 1-jet | $e^{\pm}\mu^{\pm}$ 2-jet | $\mu^{\pm}\mu^{\pm}$ 1-jet | $\mu^{\pm}\mu^{\pm}$ 2-jet |
|---|---|---|---|---|
| ggH($H \rightarrow W^+W^-$) | 0.46 (0.47) (± 0.06) | – | – | – |
| qqH($H \rightarrow W^+W^-$) | 0.04 (0.04) (± 0.01) | – | – | – |
| WH($H \rightarrow W^+W^-$) | 21.54 (24.37) (± 14.25) | 7.46 (8.43) (± 4.79) | 14.63 (16.45) (± 9.45) | 4.32 (4.83) (± 2.83) |
| ZH($H \rightarrow W^+W^-$) | 1.82 (1.83) (± 0.15) | 0.37 (0.37) (± 0.04) | 0.74 (0.75) (± 0.52) | 0.16 (0.16) (± 0.03) |
| ggZH($H \rightarrow W^+W^-$) | 0.17 (0.18) (± 0.06) | 0.08 (0.08) (± 0.02) | 0.09 (0.09) (± 0.02) | 0.01 (0.01) (± 0.00) |
| ttH($H \rightarrow W^+W^-$) | 0.13 (0.14) (± 0.02) | 0.49 (0.50) (± 0.07) | 0.08 (0.08) (± 0.03) | 0.32 (0.33) (± 0.05) |
| ggH($H \rightarrow \tau^+\tau^-$) | 0.06 (0.06) (± 0.03) | 0.02 (0.02) (± 0.03) | – | – |
| qqH($H \rightarrow \tau^+\tau^-$) | 0.03 (0.03) (± 0.00) | – | – | – |
| WH($H \rightarrow \tau^+\tau^-$) | 7.65 (8.63) (± 4.85) | 1.08 (1.22) (± 0.70) | 4.73 (5.34) (± 3.01) | 0.66 (0.74) (± 0.43) |
| ZH($H \rightarrow \tau^+\tau^-$) | 0.68 (0.68) (± 0.04) | 0.19 (0.19) (± 0.01) | 0.34 (0.34) (± 0.03) | 0.07 (0.07) (± 0.01) |
| Total Signal | 32.57 (36.42) (± 19.08) | 9.69 (10.80) (± 5.49) | 20.60 (23.05) (± 12.46) | 5.55 (6.13) (± 3.25) |
| WW | 31.97 (32.02) (± 3.57) | 5.73 (5.90) (± 0.72) | – | – |
| WZ | 609.83 (650.93) (± 21.42) | 114.38 (102.94) (± 7.42) | 277.32 (282.93) (± 12.56) | 52.43 (48.46) (± 4.43) |
| ZZ | 41.02 (40.72) (± 1.50) | 7.23 (7.12) (± 0.44) | 16.38 (16.16) (± 0.71) | 2.31 (2.26) (± 0.26) |
| W$\gamma$ | 116.21 (129.45) (± 17.32) | 3.50 (5.59) (± 2.42) | – | – |
| W$\gamma^\star$ | 113.78 (125.96) (± 17.53) | 12.36 (14.63) (± 2.48) | 45.33 (47.95) (± 6.95) | 4.13 (4.29) (± 0.93) |
| Z$\gamma$ | 19.46 (20.88) (± 4.67) | 1.78 (1.74) (± 2.64) | – | – |
| Z$\gamma^\star$ | 15.70 (15.39) (± 2.09) | 1.89 (1.87) (± 1.66) | 0.92 (0.93) (± 0.42) | – |
| VVV | 15.93 (16.12) (± 0.99) | 11.15 (11.22) (± 0.62) | 11.07 (11.10) (± 0.48) | 7.31 (7.33) (± 0.42) |
| Top | 45.21 (46.87) (± 5.03) | 13.49 (13.77) (± 2.26) | 0.10 (0.10) (± 0.03) | – |
| Non-prompt | 428.56 (473.25) (± 42.07) | 100.18 (115.87) (± 11.31) | 119.10 (171.68) (± 14.36) | 49.03 (74.56) (± 7.34) |
| Total Background | 1437.66 (1551.59) (± 43.21) | 271.69 (280.65) (± 14.75) | 470.23 (530.85) (± 18.24) | 115.21 (136.90) (± 8.24) |
| Total Prediction | 1470.24 (1588.01) (± 39.80) | 281.38 (291.46) (± 13.54) | 490.84 (553.90) (± 15.67) | 120.76 (143.03) (± 7.37) |
| Data | 1482.00 (1482.00) (± –) | 272.00 (272.00) (± –) | 504.00 (504.00) (± –) | 106.00 (106.00) (± –) |

TABLE 5.14: The expected (observed) significance and signal strength for the SM Higgs boson production via the associated WH production mechanism.

| Dataset | Expected significance | Observed significance | $\mu_{WH}$ |
|---|---|---|---|
| 2016 | 0.6 | 2.2 | $3.7^{+1.8}_{-1.7}$ |
| 2017 | 0.7 | 0.0 | $-1.4^{+1.5}_{-1.4}$ |
| 2018 | 0.7 | 0.6 | $0.8^{+1.3}_{-1.3}$ |
| Combination | 1.0 | 0.9 | $0.9^{+1.0}_{-1.0}$ |

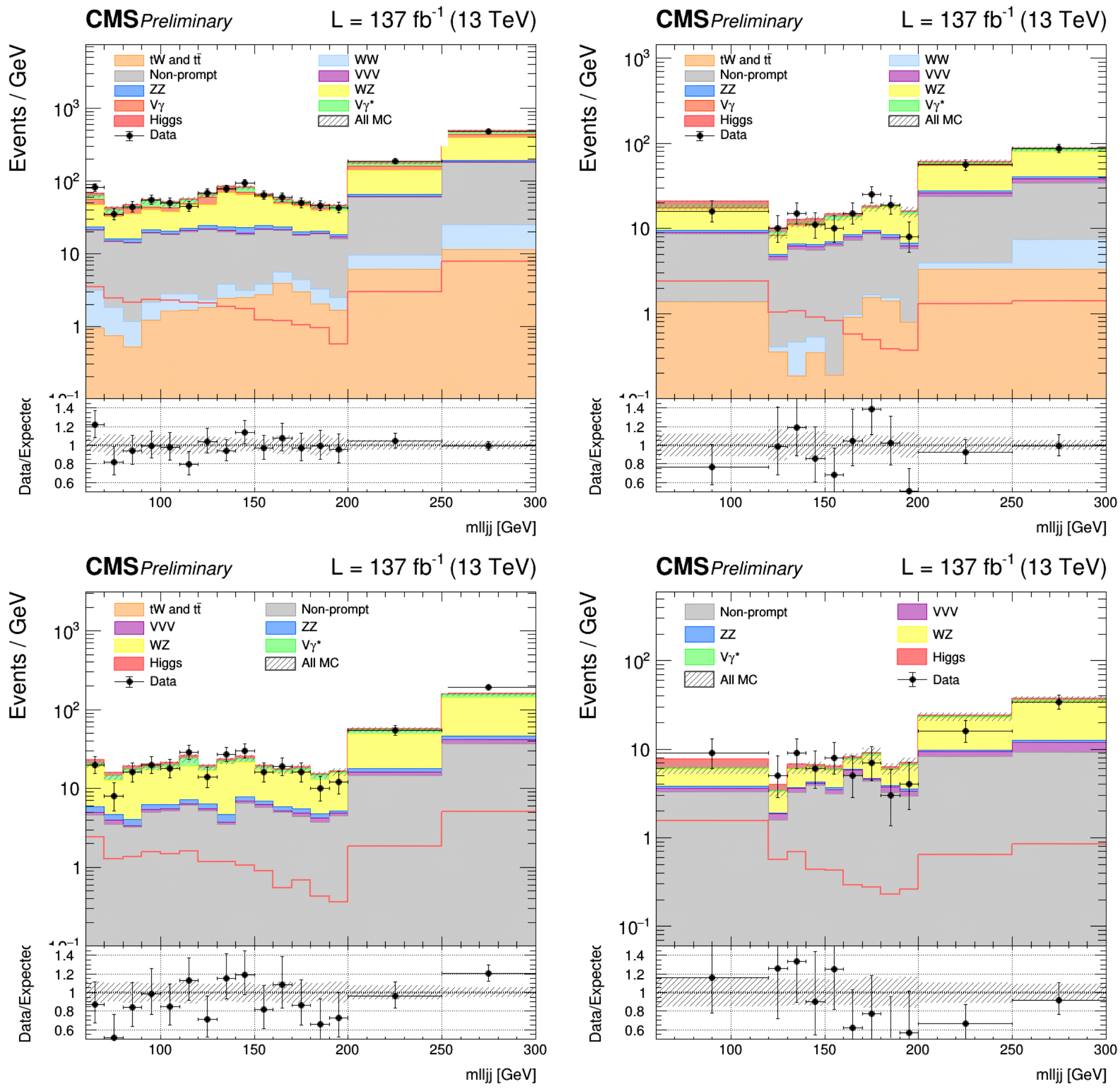



FIGURE 5.14: The post-fit mass variable distribution for different flavor $e\mu$ (top row) and same flavor $\mu\mu$ (bottom row), in 1-jet (left column) and 2-jets (right column) category for full Run-II dataset. The total uncertainty is taken as sum in quadrature of each uncertainty from all three datasets.

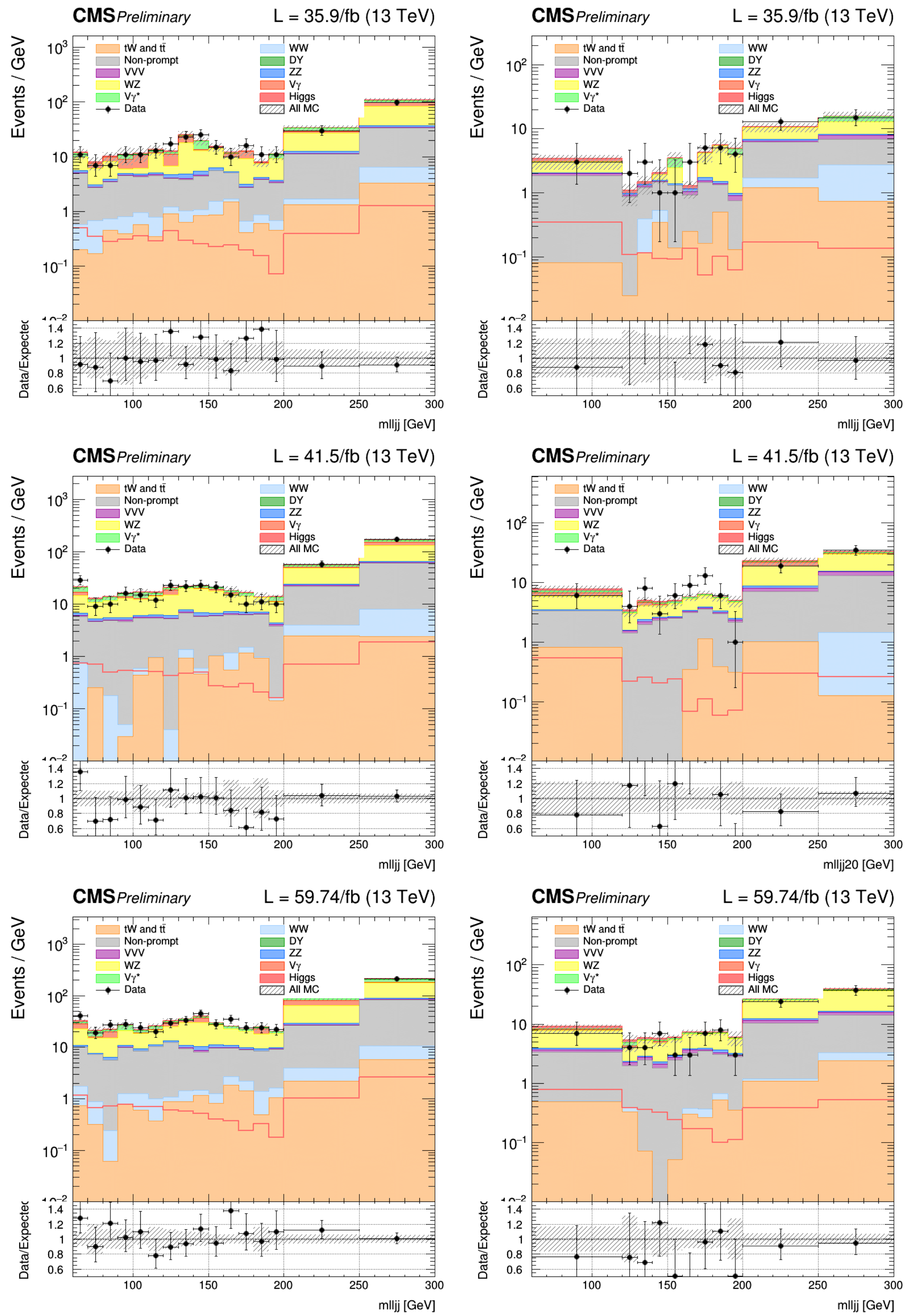



FIGURE 5.15: The post-fit mass variable distribution for different flavor $e\mu$ category for all three year datasets. The left column shows the distribution in 1-jet category, the right column shows the distribution in 2-jet category.

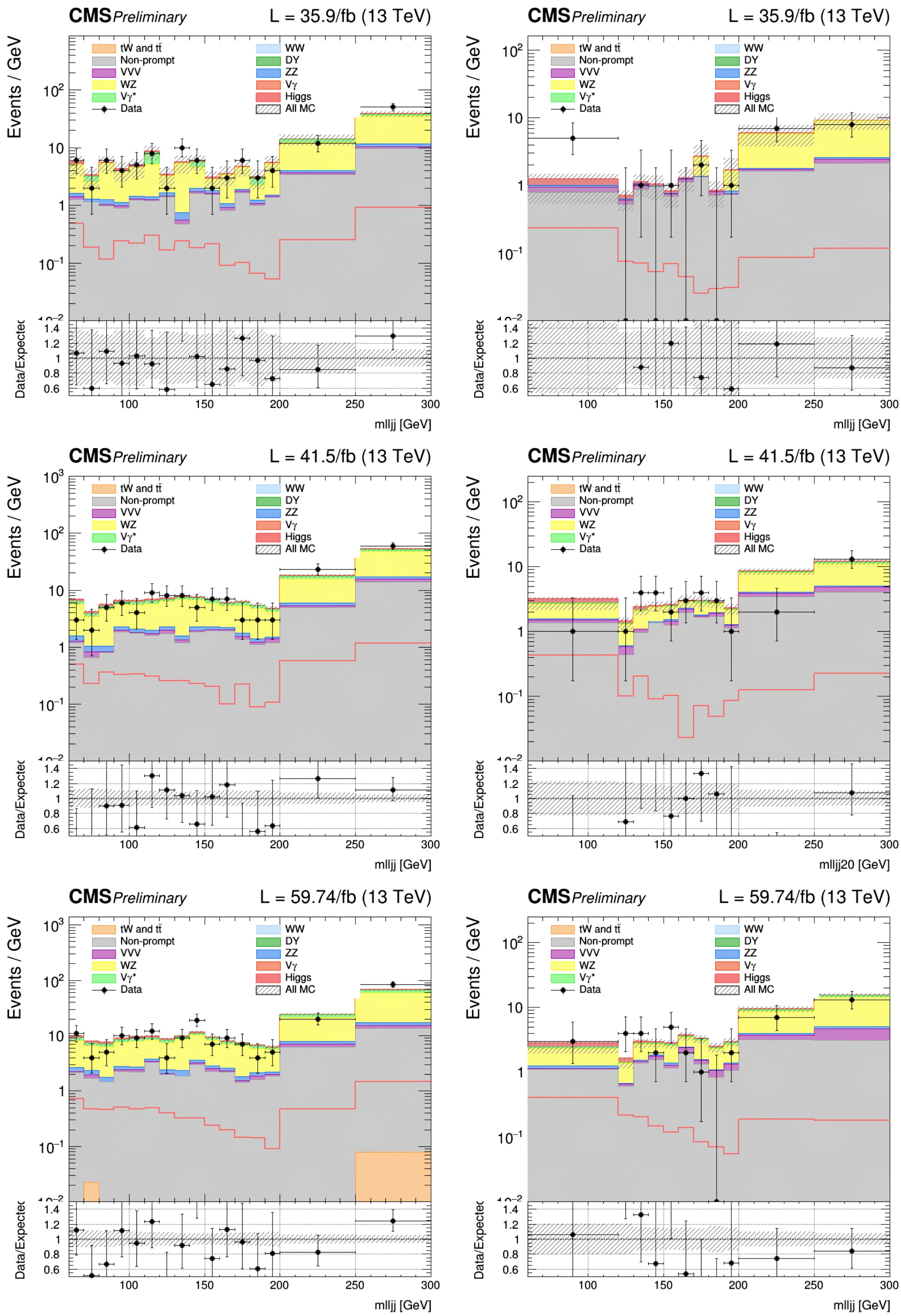



FIGURE 5.16: The post-fit mass variable distribution for same flavor $\mu\mu$ category for all three year datasets. The left column shows the distribution in 1-jet category, the right column shows the distribution in 2-jet category.

The variation or impact $\Delta\mu^{\pm}$ on the parameter of interest $\mu$ induced by changing the value of a nuisance parameter $\theta$ in the fit, is quantified by the formula:

$$\Delta\mu^{\pm} = \hat{\hat{\mu}}_{\theta_0 \pm \sigma_\theta} - \hat{\mu} \tag{5.15}$$

where $\hat{\mu}$ is the value maximizing the likelihood, and $\hat{\hat{\mu}}_{\theta_0 \pm \sigma_\theta}$ is the value of the parameter extracted from an unconstrained fit, when the expected value of the nuisance parameter $\theta_0$, is shifted by one standard deviation from its posterior value.

The measurement is dominated by the statistical uncertainty in data. As seen in Tab. 5.15, the main systematic contributions affecting the uncertainty in the signal strength arise from the non-prompt background normalization, fake lepton systematic uncertainty, background modeling such as the WZ background normalization, W$\gamma^\star$ and QCD scale.

TABLE 5.15: The main systematic sources and their contribution or impact to the signal strength uncertainty.

| Systematic uncertainty | $\Delta\mu/\mu$ |
|---|---|
| Non-prompt background ($\mu\mu$) normalization | 43 % |
| Non-prompt background ($e\mu$) normalization | 34 % |
| Fake lepton systematic (2016) | 18 % |
| Fake lepton systematic (2018) | 12 % |
| W$\gamma^\star$ scale factor | 12 % |
| Fake lepton systematic (2017) | 11 % |
| WZ data driven background normalization | 10 % |
| QCD scale | 6 % |
| JES scale | 6 % |

# Chapter 6

# Measurement of the Higgs boson production in association with a leptonic W or Z boson in the H → WW decay channel

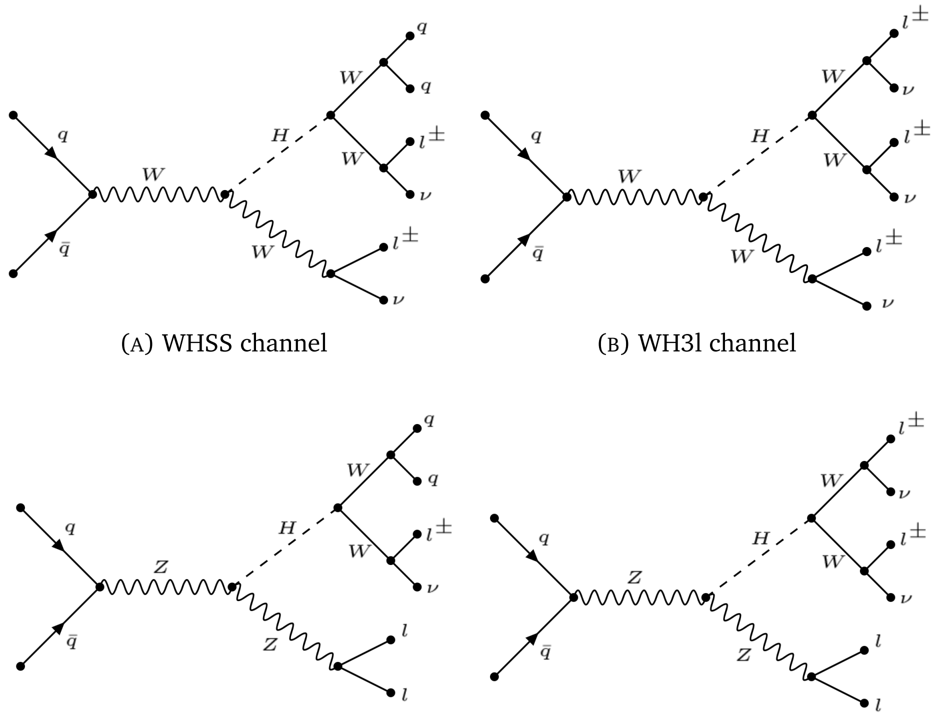
The chapter presents the measurement of the Higgs boson production in association with a leptonically decaying vector boson, followed by the Higgs boson branching fraction to a pair of W bosons. The full Run-II dataset corresponding to a total integrated luminosity of 137 $\text{fb}^{-1}$ at a center-of-mass energy of 13 TeV is used in the combined measurement.

The same sign di-lepton final state analysis discussed in Chapter 5 is considered in the combined measurement, together with other final states targeting the leptonic VH production mode, with V being the associated W or Z boson. In addition, the production cross sections are measured with respect to the vector boson transverse momentum, according to the so-called STXS framework, which will be discussed in Sec. 6.4.

## 6.1 Analysis strategy

The combined measurement considers $H \rightarrow W^{+}W^{-}$ events in which either one W boson decays leptonically and one hadronically, or both W bosons decay leptonically. Four final state topologies are considered, each one providing different sensitivity to the measurement, depending on the production mode, selection efficiency and the level of background in the final selected sample.

The fully leptonic decay of the W bosons pair is targeted in the WH3l and ZH4l channels, whose final states are purely leptonic, including the lepton from the associated W or Z boson; the hadronic and leptonic W bosons decays are relevant in the WHSS and ZH3l




(A) WHSS channel (B) WH3l channel

(C) ZH3l channel (D) ZH4l channel

FIGURE 6.1: Feynman diagrams for the signal processes considered in the combined measurement.

channels, leading to final states consisting of jets and leptons. The channels considered for the measurement are summarized in Tab. 6.1, and the corresponding Feynman diagrams are shown in Fig. 6.1.

TABLE 6.1: The channels considered for the combined measurement. The final state refers to the reconstructed objects in the signal events.

| Channel | Production mode | $H \to W^+W^-$ decay mode | Final state |
|---|---|---|---|
| WHSS | WH | leptonic and hadronic | 2 same sign leptons, $E_T^{miss}$, 1/2 jet(s) |
| WH3l | WH | leptonic | 3 leptons, $E_T^{miss}$ |
| ZH3l | ZH | leptonic and hadronic | 3 leptons, $E_T^{miss}$, 1/2 jet(s) |
| ZH4l | ZH | leptonic | 4 leptons, $E_T^{miss}$ |

The combined measurement is performed adopting two approaches:

- inclusive signal strength measurement, in which the inclusive VH leptonic production cross section, expressed as a single signal strength modifier for all production processes, as postulated by the SM;
- STXS measurement, where the STXS cross section is defined in terms of the Higgs boson production mode, expressed as multiple signal strength modifiers, as a function

of the associated vector boson kinematics defined at generator level (see Sec. 6.4). The set of signal strength modifiers are measured using a dedicated approach in each search channel.

The two approaches follow the same fit strategy, which assumes that the relative yields of the different production mechanisms are consistent to the SM expectation. The leptonic VH production cross section times the branching fraction of $H \to W^+W^-$ is extracted by simultaneously fitting to the observable offering strong discrimination between the signal and the background, in all four channels and for all data taking years.

## 6.2 Event selection and categorization

The PDs used are given in Sec. 5.3.1. Events are selected according to single and double lepton triggers, using the same set of triggers and $p_T$ threshold given in Sec. 5.3.2. The trigger efficiencies are measured in the data and applied to the simulated events as described in Sec. 4.4.1. The MC simulated set of background and signal samples used to estimate the processes are described in Sec. 5.3.3. The electrons, muons, jets, b-tagging and $E_T^{miss}$ definitions are common among all channels, described in Sec. 5.4.

Dedicated pre-selections are applied to individual channels to obtain signal-like events by reducing the background contribution. The event categorization specific to the channel is defined to maximize the sensitivity to the signal and to provide a better description of important background via a data-driven method. Details on the event selection, categorization, and the discriminant variables used in the combined fit in each channel are described in the following sections. The systematic uncertainties considered in the combined fit is discussed in Sec. 6.3.

### 6.2.1 WHSS channel

As extensively discussed in Chapter 5, the WHSS channel targets the associated WH production mode, followed by the $H \to W^+W^-$ decay. The final state contains two same sign leptons, $E_T^{miss}$, and jets. The channel provides sensitivity to the signal process by exploiting the same sign topology, and the event categorization discussed in detail in Sec. 5.5. The Fig. 5.14 shows the distribution of the relevant mass variable introduced there, after the full analysis selection summarized in Tab. 5.7.

The WZ background normalization is constrained in 1-jet and 2-jets control regions defined in ZH3l channel, discussed in Sec. 6.2.3. This control region, which is used in common with the other channels of the combined analysis, is different from the WZ control region discussed in Chapter 5 (see Tab. 5.12). These large statistical control regions provide a better estimation of the WZ background normalization and stability to the combined fit. The Fig. 6.2 shows the mass variable distribution after the fit to data in the WZ control regions.

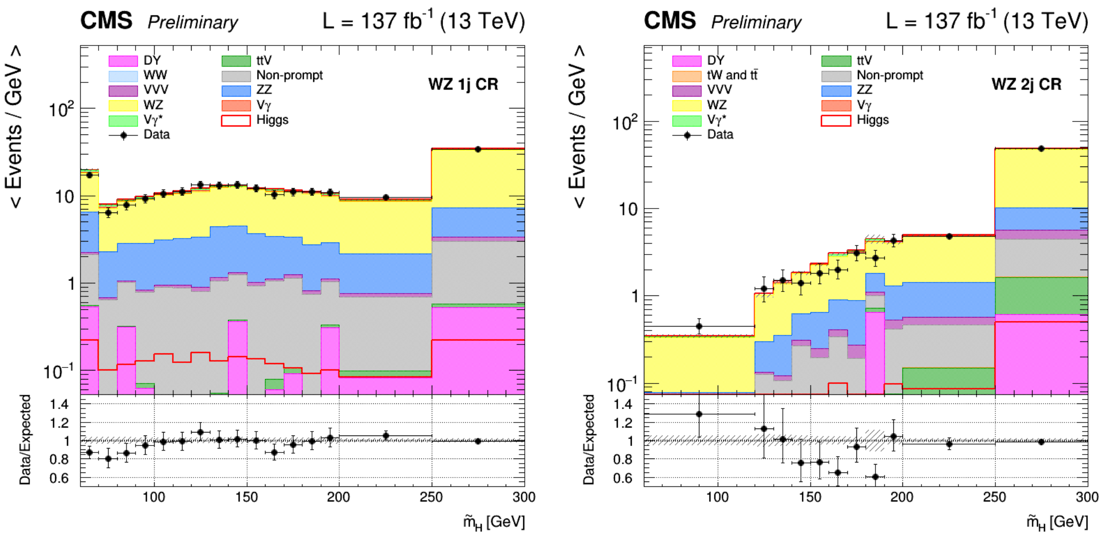



FIGURE 6.2: The mass variable distributions in the 1-jet (left) and 2-jets (right) WZ control regions after the fit to data. The bin contents are normalized by the bin width. The signal is shown as the topmost contribution to the stacked plot, and overlaid for shape comparison. The uncertainty band is determined by sampling the posterior nuisance parameter from the covariance matrix.

The main systematic uncertainties to this channel is similar to the results obtained in Sec. 5.9, except the WZ background normalization uncertainties, which is reduced thanks to the large amount of data in the WZ control regions.

### 6.2.2 WH3l channel

As the WHSS channel, also the WH3l channel targets the associated WH production mode, followed by the $H \to W^+W^-$ decay. The analysis selects events containing three leptons with $p_T > 25, 20, 15$ GeV and total charge $\pm 1$. Events containing a fourth lepton with $p_T > 10$ GeV are rejected. The invariant mass of any di-lepton pair, $m_{ll}$, is required to be greater than 12 GeV to remove events with low-mass resonances. In addition, events are rejected if they contain a jet with $p_T > 30$ GeV, or a b-tagged jet with $p_T > 20$ GeV, using the DeepCSV loose WP.

In this signal region, events are categorized based on the number of same sign same flavor di-lepton pairs. If one of these is present, the event is placed in the same sign same flavor (SSSF) category, while all other events are placed in the opposite sign same flavor (OSSF) category. To reject backgrounds containing Z bosons, events in the OSSF signal region must pass a Z boson veto, where all OSSF lepton pairs must be outside of the Z boson mass window by ±20 GeV. Events in the OSSF signal region must also satisfy $E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV to avoid overlapping dedicated control regions.

The dominant backgrounds are WZ, ZZ, $Z\gamma^{\star}$, and backgrounds containing non-prompt leptons coming from $\mathrm{t\bar{t}}$, tW, and Z+jets processes, which are estimated from data as described in Sec 5.6.1. The remaining backgrounds are estimated from simulated samples, with WZ and $Z\gamma$ backgrounds normalized using dedicated control regions.

The control regions for WZ and Z$\gamma$ have the same selection as the OSSF signal region, with the exception of an inverted Z boson mass selection and a different $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement. In particular, events in the WZ control region must satisfy $E_{\mathrm{T}}^{\mathrm{miss}} > 45$ GeV and the tri-lepton invariant mass to be above 100 GeV, while events in the Z$\gamma$ control region must satisfy $E_{\mathrm{T}}^{\mathrm{miss}}$ $<$ 40 GeV and the tri-lepton invariant mass is in between 80 and 100 GeV.

A summary of the event selection and categorization is shown in Tab. 6.2.

TABLE 6.2: Event selection used to define the SR, Control Region (CR) and categorization in the WH3l channel.

| | Preselection | | | |
|---|---|---|---|---|
| Lepton $p_{\mathrm{T}}$ (GeV) | $> 25, 20, 15$ | | | |
| Fourth lepton $p_{\mathrm{T}}$ (GeV) | $< 10$ | | | |
| Total tri-lepton charges, $ch_{lll}$ | $\pm 1$ | | | |
| $Min(m_{ll})$ (GeV) | $> 12$ | | | |
| Jets with $p_{\mathrm{T}} > 30$ GeV | 0 | | | |
| b-jet veto | DeepCSV loose WP applied to all jets with $p_{\mathrm{T}} > 20$ GeV | | | |
| | OSSF SR | SSSF SR | WZ CR | $Z\gamma$ CR |
| OSSF lepton pair | Yes | No | Yes | Yes |
| Z boson mass veto, $\lvert m_{ll} - m_{Z}\rvert$ (GeV) | $> 20$ | - | $< 20$ | $< 20$ |
| $E_{\mathrm{T}}^{\mathrm{miss}}$ (GeV) | $> 40$ | - | $> 45$ | $< 40$ |
| tri-lepton mass, $m_{lll}$ (GeV) | - | - | $> 100$ | $[80, 100]$ |

To discriminate between signal and background, a multivariate discriminator based on a BDT algorithm is used. The BDT is trained on events passing the OSSF and SSSF signal region selections without the Z boson mass veto requirement. The variables used in the BDT training are:

- the transverse momentum of each lepton in the event;
- the missing transverse momentum in the event;
- the Deep b-tagging discriminant of the leading and sub-leading jet in the event;
- the angular difference ($\Delta R$) and transverse mass between each lepton and missing transverse momentum;

- the invariant mass, transverse momentum of the tri-lepton system, and the azimuthal angle of the tri-lepton system and the transverse missing energy;
- the Z mass veto, minimum invariant mass difference to Z boson, invariant mass of di-lepton, transverse momentum, and angular difference ($\Delta R$) among all OSSF lepton pairs.

To extract the Higgs boson production cross section, a fit is performed to the BDT discriminant. The figure 6.3 shows the BDT discriminant distributions after the fit to data. The Fig. 6.4 shows the BDT discriminant distributions after the fit to data in WZ and Z$\gamma$ control regions.

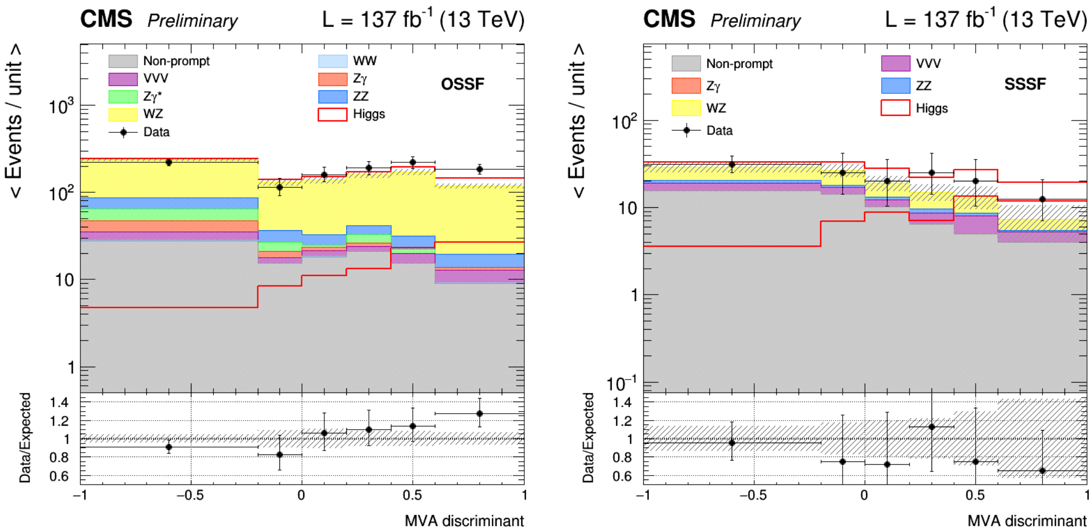



FIGURE 6.3: The BDT discriminant distributions in the OSSF (left) and SSSF (right) signal regions after the fit to data. The bin contents are normalized to the bin width. The signal is shown as the topmost contribution to the stacked plot, and overlaid for shape comparison.

The main systematic uncertainties arise from the non-prompt background normalization and statistical uncertainties.

### 6.2.3 ZH3l channel

The ZH3l channel targets the associated ZH production mode, followed by the $H \to W^+W^-$ decay. The final state contains three leptons, $E_T^{\text{miss}}$, and jets. The analysis selects events containing three leptons with $p_T > 25$, 20, 15 GeV and total charge $\pm1$. Events containing a fourth lepton with $p_T > 10$ GeV are rejected. The invariant mass of any di-lepton pair $m_{ll}$, is required to be greater than 12 GeV to reject low mass resonances. The events must contain an opposite sign same flavor di-lepton pair with an invariant mass difference with respect to the $m_Z$ below 25 GeV. Events are rejected if any b-tagged jet with $p_T > 20$ GeV passing the DeepCSV medium WP is found.

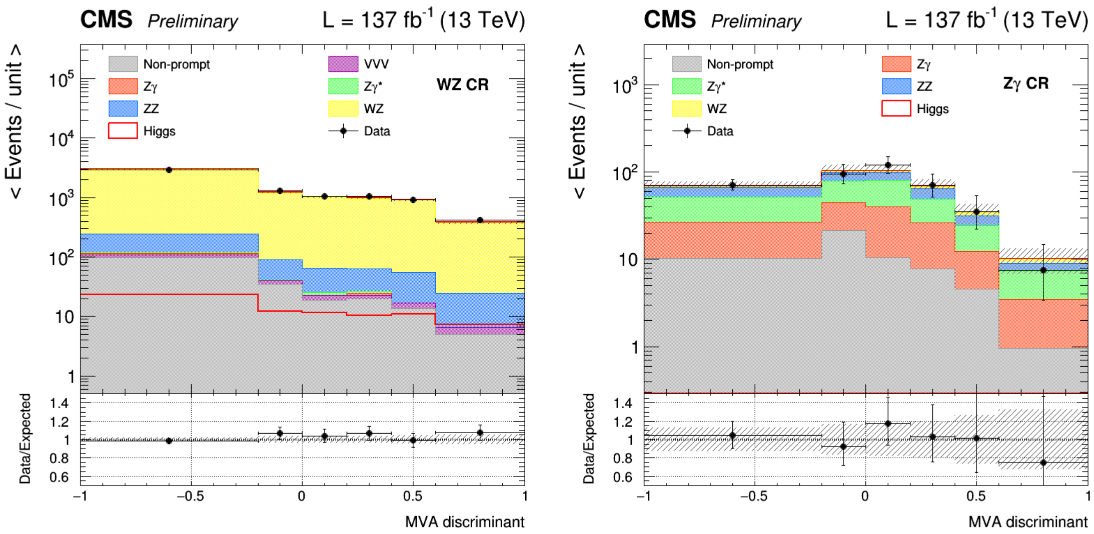



FIGURE 6.4: BDT distributions in the WZ (left) and Z$\gamma$ (right) control regions after the fit to data. The bin contents are normalized by the bin width. The signal is shown as the topmost distribution to the stacked plot, and overlaid for shape comparison. The uncertainty band is determined by sampling the posterior nuisance parameter from the covariance matrix.

Events are categorized based on the number of jets in the event. Events in the 1-jet category contain exactly one jet with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 4.7$, while events in the 2-jet category contain at least two jets passing these requirements. The signal region events must also have a W boson azimuthal separation, $\Delta\phi(lp_{\mathrm{T}}^{\mathrm{miss}}, j(j))$ smaller than $\pi/2$, and pass a Z internal conversion veto,$|m_{lll} - m_Z| > 20$ GeV.

The main backgrounds in the ZH3l analysis are WZ, ZZ, and Z+jets events. $Z\gamma/\gamma^*$, VVV, and $\mathrm{t\bar{t}}$ +jets processes also contribute to this final state. The Z+jets events are entering the signal region when a non-prompt lepton passes the lepton selection, and this non-prompt background is estimated from data, as described in Sec. 5.6. The remaining backgrounds are modeled using MC simulation. The WZ normalization is extracted from data control regions, which are categorized by the number of jets in the same way as the signal regions. The WZ control region is defined by inverting the W boson azimuthal angle requirement, $\Delta\phi(lp_{\mathrm{T}}^{\mathrm{miss}}, j(j))$ to ensure orthogonality between control and signal regions. The WZ control regions are also used to constrain the WZ background normalization in the WHSS channel 1-jet and 2-jet signal regions.

A summary of the event selection and categorization is shown in Tab. 6.3.

The shape discriminating variable used in the ZH3l channel is the transverse mass variable $m_T^H$:

$$m_T^H = \sqrt{\left(\sum_{i\in Obj} p_{T,i}\right)^2 - \left(\sum_{i\in Obj} p_{x,i}\right)^2 - \left(\sum_{i\in Obj} p_{y,i}\right)^2} \tag{6.1}$$

TABLE 6.3: Event selection used to define the SR, CR and categorization in the ZH3l channel.

| | Preselection | | | |
|---|---|---|---|---|
| Lepton $p_{\mathrm{T}}$ (GeV) | $> 25, 20, 15$ | | | |
| Fourth lepton $p_{\mathrm{T}}$ (GeV) | $< 10$ | | | |
| Total tri-lepton charge, $ch_{lll}$ | $\pm 1$ | | | |
| $Min(m_{ll})$ (GeV) | $> 12$ | | | |
| b-jet veto | DeepCSV medium WP applied to all jets with $p_{\mathrm{T}} > 20$ GeV | | | |
| $\lvert m_{ll} - m_Z \rvert$ (GeV) | $< 25$ | | | |
| $\lvert m_{lll} - m_Z \rvert$ (GeV) | $> 20$ | | | |
| | 1-jet SR | 2-jets SR | 1-jet WZ CR | 2-jets WZ CR |
| Jets with $p_{\mathrm{T}} > 30$ GeV | $= 1$ | $\geq 2$ | $= 1$ | $\geq 2$ |
| $\Delta\phi(lE_{\mathrm{T}}^{\mathrm{miss}}, j(j))$ | $< \pi/2$ | $< \pi/2$ | $> \pi/2$ | $> \pi/2$ |

where $p_{x,i}$, $p_{y,i}$ are the x- and y-component of the four-momentum of the lepton, $E_{\mathrm{T}}^{\mathrm{miss}}$ and the jet(s) from the Higgs decay. Fig. 6.5 shows the $m_T^H$ distributions after the fit to data. The Fig. 6.6 shows the $m_{\mathrm{T}}^{\mathrm{H}}$ distributions after the fit to the data in the WZ control regions.

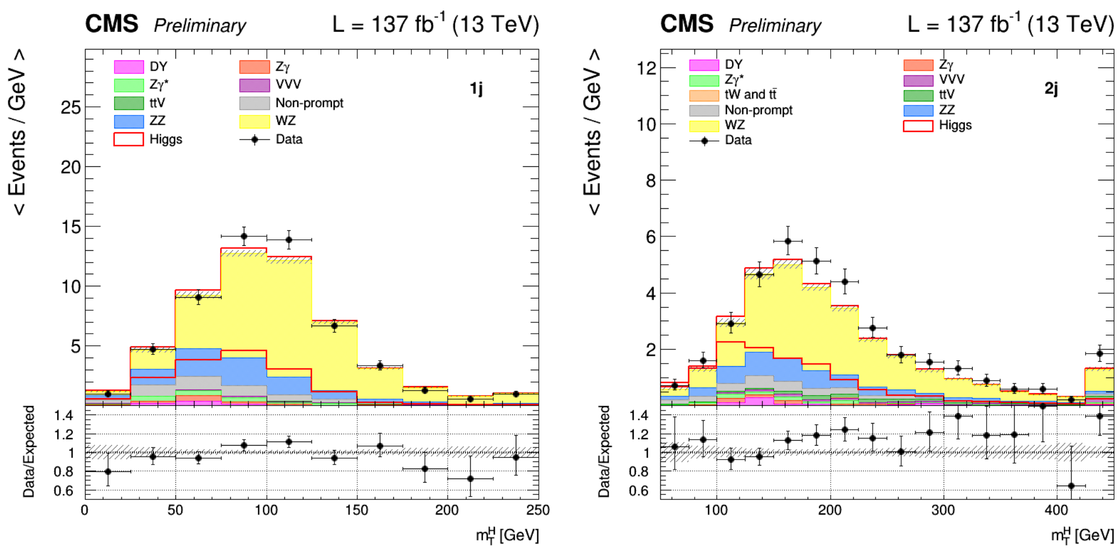



FIGURE 6.5: The $m_T^H$ distributions for ZH3l channel, in the 1-jet category (left) and 2-jet category (right) signal regions after the fit to data. The bin contents are normalized to the bin width. The signal is shown as the topmost contribution to the stacked plot, and overlaid for shape comparison. The overlaid signal is scaled by a factor 10 for visibility.

The sensitivity of this channel is affected by various sources of systematic uncertainty, including the non-prompt background normalization, JES, WZ background normalization and statistical uncertainties. The uncertainty on the 0-jet WZ (defined in WH3l channel) and $Z\gamma$ background normalizations also has a large impact on the ZH3l channel's signal strength despite these backgrounds do not contribute directly to the ZH3l signal region. This is due

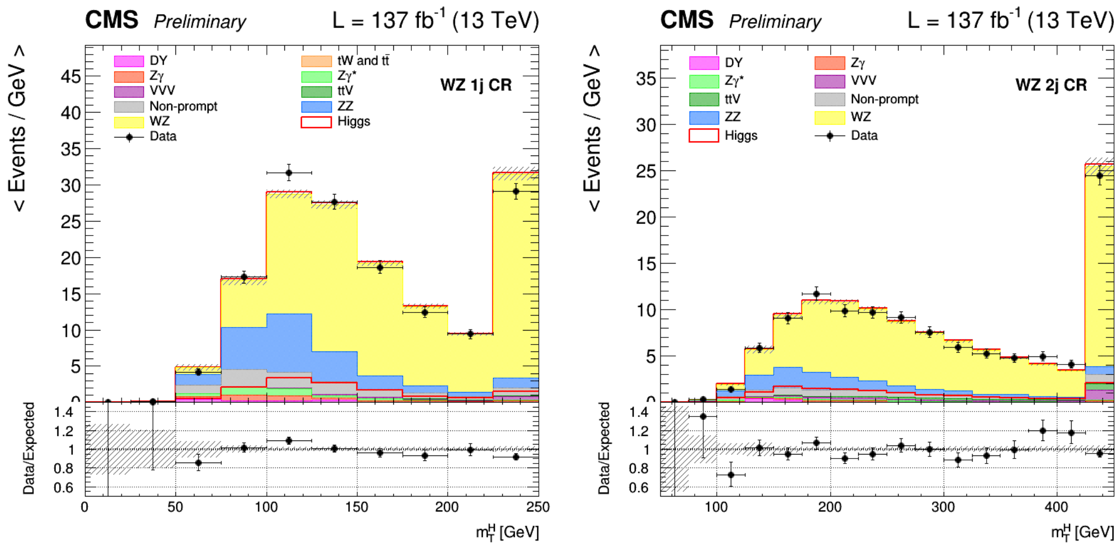



FIGURE 6.6: The $m_T^H$ distributions for ZH3l channel, in the 1-jet category (left) and 2-jet category (right) WZ control region after the fit to data. The bin contents are normalized to the bin width. The signal is shown as the topmost contribution to the stacked plot, and overlaid for shape comparison. The overlaid signal is scaled by a factor 10 for visibility. The uncertainty band is determined by sampling the posterior nuisance parameter from the covariance matrix.

to the correlations between these nuisance parameters and nuisance parameters affecting the overall WZ and Z$\gamma$ normalizations (see left plot in Fig. 6.12). The uncorrelated fake systematic uncertainties from WHSS channel also impact this channel sensitivity, as both channels share the common WZ control regions.

### 6.2.4 ZH4l channel

As the ZH3l channel, also the ZH4l channel targets the associated ZH production mode, followed by the $H \to W^+W^-$ decay. The selected events contain four leptons with $p_T > 25$, 15, 10, 10 GeV and zero total charge. Events containing a fifth lepton with $p_T > 10$ GeV are rejected. The invariant mass of any di-lepton pair $m_{ll}$ is required to be larger than 12 GeV to reject low mass resonances. The opposite sign same flavor lepton pair with invariant mass closest to the Z boson mass is considered as the Z boson candidate, while the remaining lepton pair, supposed to originate from the Higgs boson, is referred to as the $X$ candidate. The Z boson candidate mass is required to be within ±15 GeV of the Z boson mass window. Events are rejected if they contain any b-tagged jet with $p_T > 20$ GeV passing the DeepCSV loose WP.

The events are categorized based on the flavor of the $X$ lepton pair. Events in the XSF category have a same flavor $X$ lepton pair, while events in the XDF category have a different flavor $X$ lepton pair. In the XSF category, events are required to satisfy $m_{llll} > 140$ GeV, 10

$< m^X_{ll} < 65$ GeV, and $E_T^{miss} > 35$ GeV. Events in the XDF category must have $10 < m^X_{ll} < 70$ GeV and $p_T^{ll} > 20$ GeV.

The ZZ production is the main background for this channel. Additional contributions arise from $t\bar{t}Z$, VVV, and $V\gamma$ processes. These backgrounds are modeled with MC simulation. The ZZ normalization is extracted from a control region, defined by requiring $75 < m^X < 105$ GeV and $E_T^{miss} < 35$ GeV.

The event selection and categorization in the ZH4l channel is summarized in Tab. 6.4.

TABLE 6.4: Event selection used to define the SR, CR and categorization in the ZH4l channel.

| | Preselection | | |
|---|---|---|---|
| Lepton $p_T$ (GeV) | $> 25, 15, 10, 10$ | | |
| Fifth lepton $p_T$ (GeV) | $< 10$ | | |
| Total four lepton charge, $ch_{llll}$ | 0 | | |
| $Min(m_{ll})$ (GeV) | $> 12$ | | |
| Z mass window, $\|m_{ll} - m_Z\|$ (GeV) | $< 15$ | | |
| B-jet veto | DeepCSV loose WP applied to all jets with $p_T > 20$ GeV | | |
| | XSF SR | XDF SR | ZZ CR |
| X pair flavor | Same | Different | - |
| Tetra-lepton mass, $m_{llll}$ (GeV) | $> 140$ | - | - |
| Invariant mass of X pair, $m^X_{ll}$ (GeV) | [10,65] | [10,70] | [75,105] |
| $E_T^{miss}$ (GeV) | $> 35$ | $> 20$ | $< 35$ |

A multivariate discriminator based on a BDT is used to discriminate between signal and background. The BDT is trained on events passing the pre-selection, with $p_T^{miss} > 20$ GeV and $10 < m^X_{ll} < 70$ GeV. The variables used in the training are:

- the angular distance ($\Delta R$) between the two leptons decayed from Higgs and Z boson;
- the transverse mass of the leading and trailing leptons and missing transverse momentum;
- the transverse mass of the leptons constituting the X leptons and the missing transverse momentum system;
- the azimuthal angle separation ($\Delta\phi$) and the invariant mass of X lepton pair;
- the missing transverse momentum.

Fig. 6.7 shows the BDT discriminant distributions after the fit to data, while Fig. 6.8 shows the BDT discriminant distribution after the fit to the data in the ZZ control region.

The channel sensitivity is affected by the QCD scale, $E_T^{miss}$ scale, electron efficiency and statistical uncertainties.

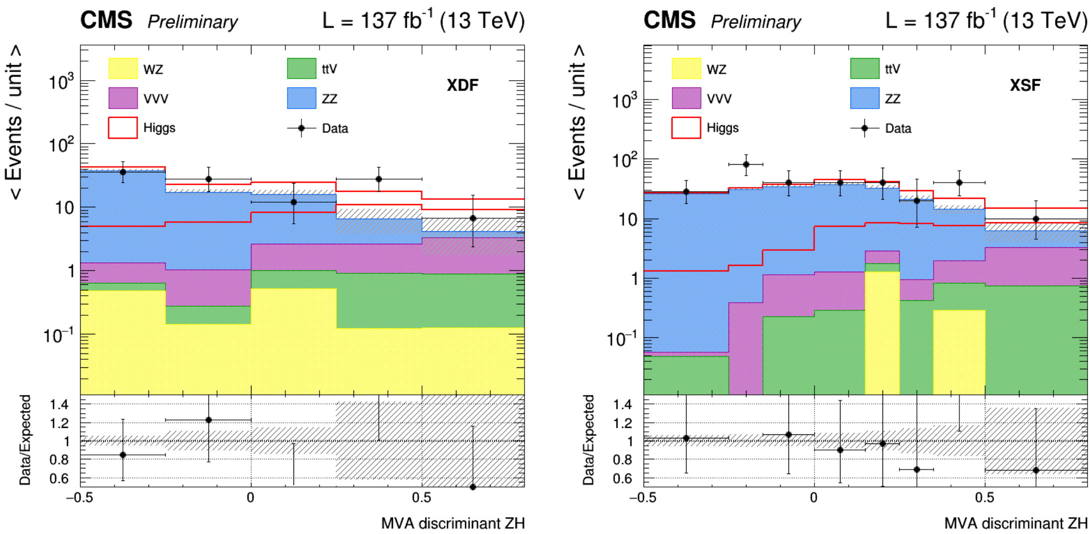



FIGURE 6.7: BDT discriminant distributions in the XDF (left) and XSF (right) signal regions after the fit to data. The bin contents are normalized by the bin width. The signal is shown as the topmost contribution to the stacked plot, and overlaid for shape comparison.

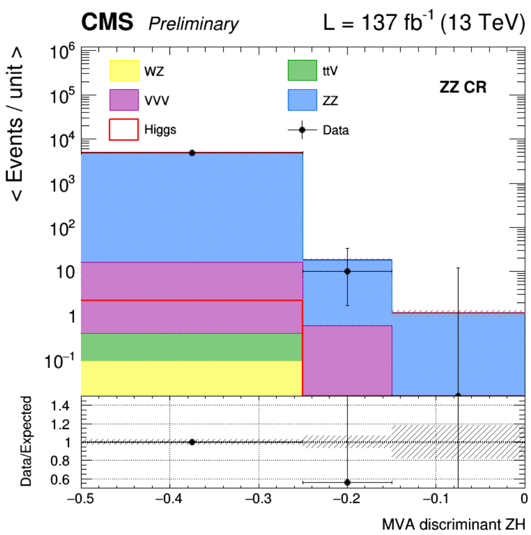



FIGURE 6.8: BDT discriminant distributions in the ZZ control region after the fit to data. The bin contents are normalized by the bin width. The signal is shown as the topmost contribution to the stacked plot, and overlaid for shape comparison. The uncertainty band is determined by sampling the posterior nuisance parameter covariance matrix.

## 6.3 Inclusive signal strength measurement

The statistical procedure used for each channel is the same as discussed in Sec. 5.8, and the combination between different decay channels follows the methodology developed by the ATLAS and CMS experiments [117, 119]. The probability distribution of the number of observed events in each category and in each bin of the discriminant distributions used to extract the signal is expected to be described by the Poisson distribution, with mean value given by the sum of the contributions from the processes under consideration. For each channel, the binning adopted for the discriminant distributions is chosen to maximize the signal sensitivity and to reduce the statistical fluctuations.

The uncertainties affecting the overall normalization of the signal and background processes, and the shape of the predictions across the distributions of the observables, are mostly common in each channel. Therefore, the sources and the treatment of uncertainties for the inclusive analysis are essentially the same ones described in Sec. 5.7. The uncertainties specific to the background modeling are correlated among final states except the non-prompt background normalization systematic uncertainties from different fake lepton sources. The systematic uncertainties from the experiment are uncorrelated among years, but correlated among final states; theoretical uncertainties are correlated among final states.

The control regions defined in different channels as described in Sec. 6.2 are independent from each others, except for the WHSS and ZH3l channels where the WZ background normalization from the respective signal regions is constrained in the same WZ control regions (See Sec. 6.2.3).

### 6.3.1 Results

A summary of the expected contribution from the associated WH and ZH production mechanisms, and the total number of expected events in each channel, after the full event selection is shown in Fig. 6.9. As expected, the dominant production mechanism in each channel is given by the targeted production mode. The $H \to \tau\tau$ decay modes have a significant, although not dominant, contribution to the Higgs boson signal, and are accounted for in the fit.

The signal and background event yields, and the number of observed events in data, for each channel after the full event selection are shown in Tab. 6.5. The uncertainties in the expected yields are obtained with the initial value of the nuisance parameters, whereas the posterior ones are obtained with the fitted values.

The signal strength modifier $\mu_{comb}$ is defined as the ratio between the leptonic VH production cross section and the SM expectation, and is extracted by performing a binned maximum likelihood fit simultaneously to all channels, assuming that the relative contributions of the different production mechanisms are as the SM ones. The normalization of the major

TABLE 6.5: The signal and background yields compared to number of observed events in all channels after the full event selection. The yields are given in the format posterior (a priori) (± posterior uncertainty), where the posterior values come from a combined fit to data. The individual signal yields are given for different production mechanisms. Correlations between uncertainties are accounted for in the total uncertainty shown.

| Process | WHSS | WH3l | ZH3l | ZH4l | Combination |
|---|---|---|---|---|---|
| ggH(H→ $W^+W^-$) | 0.5 (0.5) (± 0.0) | – | – | – | 0.5 (0.5) (± 0.0) |
| WH(H→ $W^+W^-$) | 53.1 (53.7) (± 19.8) | 37.2 (17.3) (± 5.8) | 3.2 (0.7) (± 0.7) | – | 135.7 (71.7) (± 9.6) |
| ZH(H→ $W^+W^-$) | 3.0 (3.1) (± 1.5) | 2.3 (1.1) (± 0.4) | 99.7 (23.7) (± 15.8) | 10.9 (6.2) (± 2.1) | 64.2 (34.1) (± 4.8) |
| ggZH(H→ $W^+W^-$) | 0.3 (0.4) (± 0.2) | 0.4 (0.2) (± 0.2) | 30.4 (7.2) (± 5.9) | 2.7 (1.5) (± 0.5) | 17.6 (9.2) (± 1.8) |
| ttH(H→ $W^+W^-$) | 0.9 (1.0) (± 0.1) | 0.1 (0.1) (± 0.0) | 0.8 (0.8) (± 0.1) | – | 1.8 (1.9) (± 0.1) |
| ggH(H→ $\tau^+\tau^-$) | 0.1 (0.1) (± 0.0) | – | – | – | 0.1 (0.1) (± 0.0) |
| WH(H→ $\tau^+\tau^-$) | 15.4 (15.8) (± 6.2) | 6.9 (3.2) (± 1.0) | 1.0 (0.2) (± 0.2) | – | 36.3 (19.3) (± 2.9) |
| ZH(H→ $\tau^+\tau^-$) | 1.3 (1.3) (± 0.5) | 0.7 (0.3) (± 0.1) | 23.4 (5.6) (± 3.9) | 2.8 (1.6) (± 0.5) | 16.6 (8.8) (± 1.2) |
| Total Signal | 74.6 (75.9) (± 27.9) | 47.5 (22.1) (± 7.3) | 158.5 (38.1) (± 25.2) | 16.4 (9.4) (± 3.1) | 272.9 (145.6) (± 15.1) |
| Z+jets | – | – | 35.0 (30.7) (± 7.2) | – | 36.8 (30.7) (± 7.3) |
| WW | 35.6 (31.1) (± 2.2) | 0.4 (0.3) (± 0.0) | – | – | 35.7 (31.5) (± 2.3) |
| WZ | 1039.1 (1082.2) (± 20.7) | 263.6 (231.6) (± 8.5) | 1377.7 (1357.3) (± 23.7) | 0.5 (0.5) (± 0.1) | 2710.9 (2671.6) (± 31.3) |
| ZZ | 70.2 (66.0) (± 1.9) | 28.1 (28.0) (± 1.7) | 342.7 (334.4) (± 8.5) | 45.5 (39.7) (± 1.2) | 491.1 (468.1) (± 9.7) |
| $W\gamma$ | 118.2 (134.7) (± 16.4) | – | – | – | 119.5 (134.7) (± 16.1) |
| $W\gamma^*$ | 183.3 (190.1) (± 18.4) | – | – | – | 178.2 (190.1) (± 16.8) |
| $Z\gamma$ | 30.4 (27.5) (± 5.3) | 11.7 (18.6) (± 1.9) | 43.3 (42.2) (± 5.0) | – | 86.4 (88.2) (± 7.4) |
| $Z\gamma^*$ | 26.0 (27.1) (± 2.2) | 17.7 (23.4) (± 3.6) | 57.1 (57.8) (± 6.2) | – | 102.3 (108.3) (± 7.5) |
| VVV | 44.1 (45.6) (± 0.9) | 15.0 (15.5) (± 0.5) | 31.0 (31.1) (± 0.8) | 3.3 (3.5) (± 0.3) | 94.0 (95.7) (± 1.3) |
| Top | 49.1 (48.9) (± 3.3) | – | 0.3 (0.2) (± 0.1) | – | 49.6 (49.1) (± 3.2) |
| ttV | – | – | 29.8 (29.3) (± 2.0) | 1.1 (1.1) (± 0.1) | 31.1 (30.3) (± 1.9) |
| Nonprompt | 685.9 (835.6) (± 51.1) | 59.2 (88.7) (± 6.6) | 160.7 (224.5) (± 16.1) | – | 865.0 (1148.7) (± 44.3) |
| Total Background | 2281.8 (2488.9) (± 48.4) | 395.6 (406.1) (± 12.1) | 2077.5 (2107.4) (± 28.9) | 50.4 (44.8) (± 1.3) | 4800.7 (5047.2) (± 49.0) |
| Total Yield | 2356.4 (2564.8) (± 41.0) | 443.2 (428.3) (± 12.5) | 2236.0 (2145.5) (± 28.1) | 66.8 (54.2) (± 3.3) | 5073.6 (5192.8) (± 48.2) |
| Data | 2364 | 435 | 2312 | 70 | 5181 |

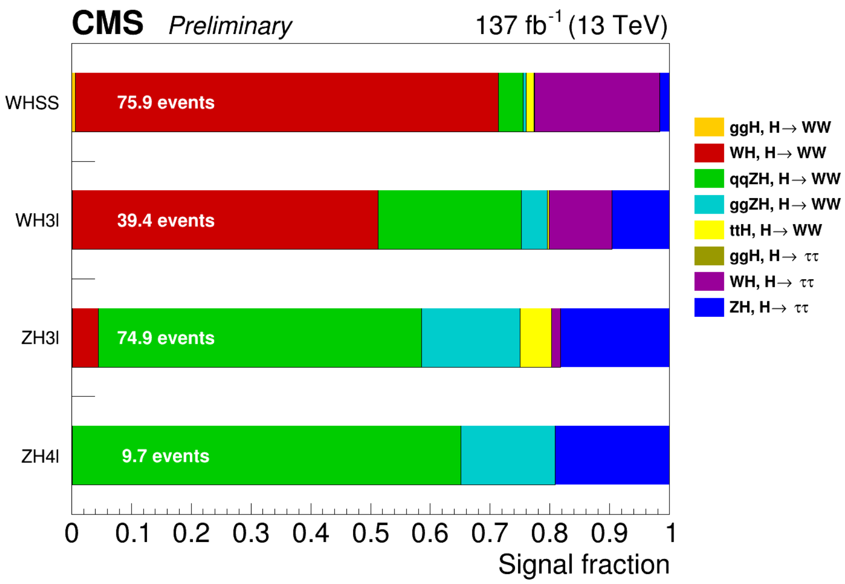



FIGURE 6.9: Expected relative fraction of associated WH and ZH production in the different final states considered in the combined fit, together with the expected signal yields.

backgrounds is estimated directly from data control samples that are enriched in events from each specific background. The contribution from all the signal samples are multiplied with the same signal strength, using all the production modes and both the $H \rightarrow W^+W^-$ and $H \rightarrow \tau^+\tau^-$ decay. The measured signal strength modifier is:

$$\mu_{comb} = 1.85^{+0.33}_{-0.32}(stat)^{+0.27}_{-0.25}(exp)^{+0.10}_{-0.07}(theo) \tag{6.2}$$

where the statistical, experimental, and theoretical uncertainties are reported separately. The expected and observed likelihood profiles as a function of the signal strength modifier with 68% and 95% likelihood intervals are shown in Fig. 6.10. The probability of observing a fluctuation, under the background-only hypothesis, corresponds to an observed significance of 4.7 $\sigma$, to be compared with the expected significance of 2.8 $\sigma$.

The compatibility of the observed signal with the SM predictions in each channel is assessed by performing a simultaneous fit in which separate signal strength modifiers are applied for each channel and each year, and left unconstrained. The expected and observed significances for each channel are shown in Tab. 6.6, and the contribution from different channels and years to the overall signal strength are summarized in Fig. 6.11.

The fitted discriminant distributions and the corresponding systematic effects on the signal strength are described for each channel respectively in Sec. 6.2. The three year combined result for each channel is consistent with the SM expectation within 1.5 $\sigma$, with the largest deviation observed in the ZH3l channel, corresponding to an asymptotic $p$-value of 0.23, which is driven by the excess of events observed in 2017 (see Fig. 6.5). The ZH4l channel is the most sensitive final state, followed by the WH3l channel, as they are benefiting from the cleaner signatures and higher purity; the sensitivity for both WHSS and ZH3l channels in

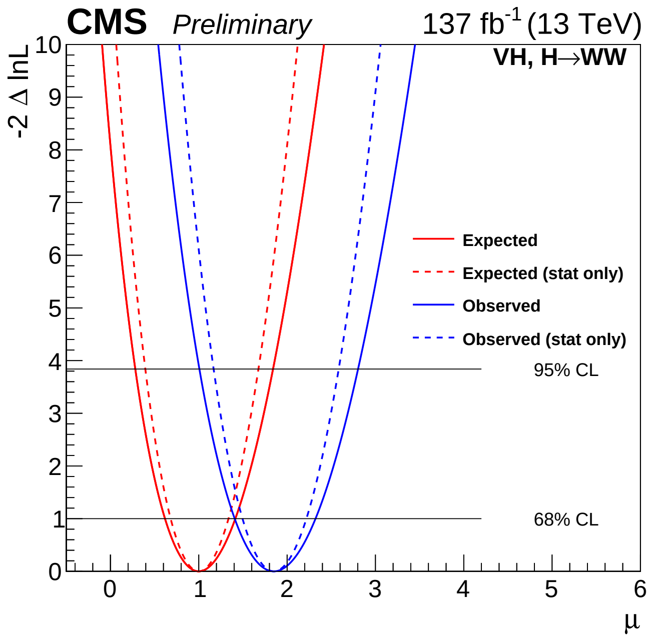



FIGURE 6.10: Expected and observed likelihood profiles as a function of the signal strength. Dashed curves include only statistical uncertainties. The 68% and 95% likelihood intervals are constructed observing where the likelihood profile crosses 1.0 and 3.84, respectively.

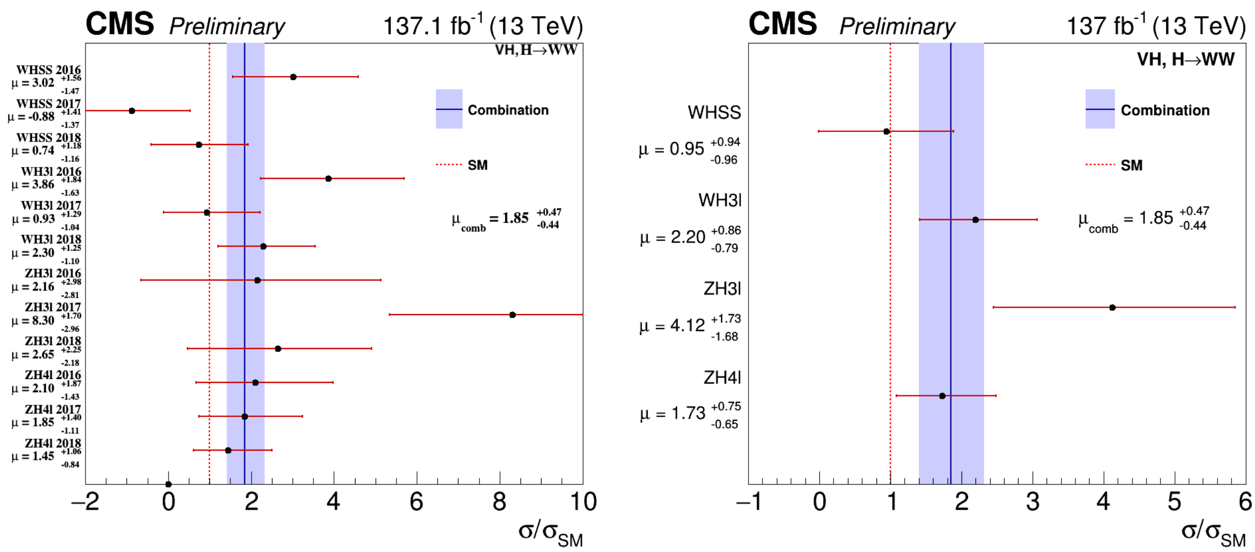



FIGURE 6.11: (Left) The observed signal strength modifiers for different years and channels; (Right) the three years combined observed signal strength modifiers for different channels. The vertical solid line represents the combined signal strength best fit value, while the horizontal bars and the filled area show the 68% confidence intervals. The vertical dashed line corresponds to the SM expectation.

TABLE 6.6: The expected and observed significance for different channels and the combination.

| Channel | Expected | Observed |
|---|---|---|
| WHSS | 1.1 | 1.0 |
| WH3l | 1.6 | 3.0 |
| ZH3l | 0.6 | 2.5 |
| ZH4l | 2.1 | 3.1 |
| Combination | 2.8 | 4.7 |

2017 are limited by the statistical fluctuation, but the former is more sensitive compared to the latter in the other years as there are less electroweak backgrounds in the signal region thanks to the same sign di-lepton selection.

The main impacts on the overall signal strength come from the non-prompt background normalization uncertainties, which do not benefit from combining channels as there are different non-prompt backgrounds with independent uncertainties in each channels: W+jets for WHSS, Z+jets, tW and $t\bar{t}$ for WH3l and Z+jets for ZH3l channel. However, the systematic uncertainties related to the fake rate derivation do benefit from combining channels since the same fake rate derivation is correlated in all channels.

The statistical uncertainties in the most significant bin of the fit template in the WH3l channel also have a noticeable impact on the signal strength. The WZ and Z$\gamma$ background normalization uncertainties in the WH3l channel also have a dominant impact as they are originated from the high sensitivity of the WH3l channel. The ZH4l channel's sensitivity causes the electron efficiency, $E_{\mathrm{T}}^{\mathrm{miss}}$ scale, and QCD scale uncertainty to also have large impact. A summary of the main systematic uncertainties and their effect on the signal strength is shown in Tab. 6.7.

The correlations between nuisance parameters and signal strengths for the fit applying a single signal strength to all signal processes and separate signal strengths in each channel are shown in Fig. 6.12. In general, correlation between the signal strength and nuisance parameters is lower than the correlation between nuisance parameters.

In the case of single signal strength fit, the nuisance parameters affecting the WZ background normalization are highly (anti)correlated. This reflects the significant contribution from WZ background in all channels. In particular, the WZ PDF uncertainties (pdf_qqbar) are strongly anti-correlated with the WZ normalization in different control regions ( CMS_hww_WZ3lnorm, CMS_hww_WZ3l1jnorm, and CMS_hww_WZ3l2jnorm ), thus driving a strong correlation between the three rate uncertainties, despite them describing different regions of phase space. Other notable (anti)correlations between nuisance parameters generally occur between the electron and muon efficiencies in a given year, or fake lepton systematic uncertainties affecting the same non-prompt background.

TABLE 6.7: The main systematic sources and their contribution or impact to the overall signal strength uncertainty.

| Systematic uncertainty | $\Delta\mu/\mu$ (%) |
|---|---|
| Renormalization and factorization scale | 3 |
| Parton distribution function | 2 |
| Parton shower, underlying event | 2 |
| Non-prompt background normalization | 9 |
| Sample size of simulation data | 8 |
| Electron | 3 |
| b-tag | 3 |
| Jet | 2 |
| Luminosity | 2 |
| WZ background normalization | 2 |
| $Z\gamma$ background normalization | 2 |
| ZZ background normalization | 1 |
| Muon | 1 |

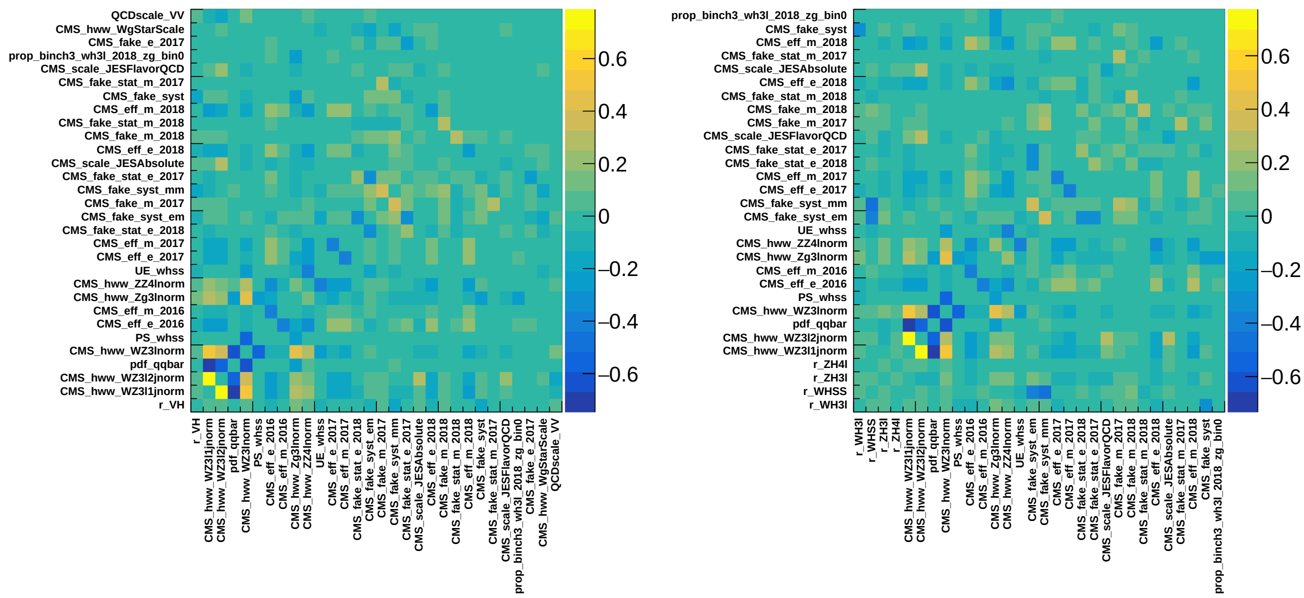



FIGURE 6.12: The correlations between nuisance parameters in the fit applying (Left) single signal strength to all signal processes and (Right) separate signal strengths in each channel. The correlations are shown between single(multiple) signal strength(s) and the 29(26) nuisance parameters with highest correlation.

For the case of separate signal strength fit, the observed correlations follow a similar pattern, with a strong correlations observed between the WHSS and WH3l signal strengths and the non-prompt background uncertainties in respective channels, while the ZH4l signal strength shows the lowest overall correlation.

## 6.4 Simplified Template Cross Section measurement

Given the fact that many Higgs boson measurements were performed by interpreting the result in signal strength and multiplicative coupling modifier in Run-I and early Run-II data analysis, a more complex measurement of the signal strength is experimentally motivated to extract more information to further test the SM compatibility, and study the Higgs boson properties. The so-called STXS framework was developed to improve these measurements in Run-II by scrutinizing in more detail different kinematic regions; it is also designed to separate measurement and interpretation steps in order to reduce the dependencies that are contained in the measurements, including the dependence on the underlying physics model (SM or BSM).

The primary goals of the STXS framework are to maximize the sensitivity of the measurements while at the same time to minimize their theory dependence. Since these are competing requirements, a nontrivial compromise has to be achieved. Some of their defining features are:

- measure signal strength modifiers and cross sections in mutually exclusive regions of phase space and separated into production modes, which allows the isolation of possible BSM effects;
- analyses can use different optimized selection criteria for the definition of the STXS categories at the reconstruction level;
- the kinematic regions are inclusive over the Higgs decays, which allows a global combination of the measurements in all decay channels;
- if an analysis has enough sensitivity, a finer binning is suggested in the framework.

The measured exclusive regions of phase space, called STXS bins, are specific to the different production modes; they are defined unambiguously by using generator-level information in the MC simulation about the Higgs boson and the jets in the event. All the STXS bins require a Higgs boson rapidity value to be less than 2.5, corresponding to the central region; higher Higgs boson rapidity region of the phase space is not accessible due to the limited coverage of the detector in the forward region. Generator-level jets are defined as anti-$k_T$ jets with a distance parameter of 0.4 and a minimum $p_{\mathrm{T}}$ of 30 GeV, clustered from all stable particles not coming from the Higgs boson decay. These generator-level jets also include the charged leptons, photons and neutrinos.

The STXS bins have been progressively defined in stages corresponding to increasingly fine granularity given the availability of larger data samples. Concerning the leptonic VH production mechanism (V=W,Z boson), the so-called stage 1.2 STXS was defined [120], where measurements in each of the production modes ($q\bar{q} \to WH$, $q\bar{q} \to ZH$ and $gg \to ZH$) were performed in bins divided by kinematic variables at generator-level such as the number of jets or the $p_{\mathrm{T}}$ of the associated vector boson, $p_{\mathrm{T}}^{\mathrm{V}}$.

Due to the statistical limitation, not all STXS bins are sensitive to the associated VH production signal strength. To provide an optimal measurement to the full spectrum of $p_{\mathrm{T}}^{\mathrm{V}}$, the signal strength is measured in two merged bins, defined by requiring $p_{\mathrm{T}}^{\mathrm{V}}$ to be less than or greater than 150 GeV; the STXS bins specific to $q\bar{q} \to ZH$ and $gg \to ZH$ are merged together to improve signal sensitivity (see Fig. 6.14), resulting into a total of four STXS bins, summarized in Fig. 6.13.

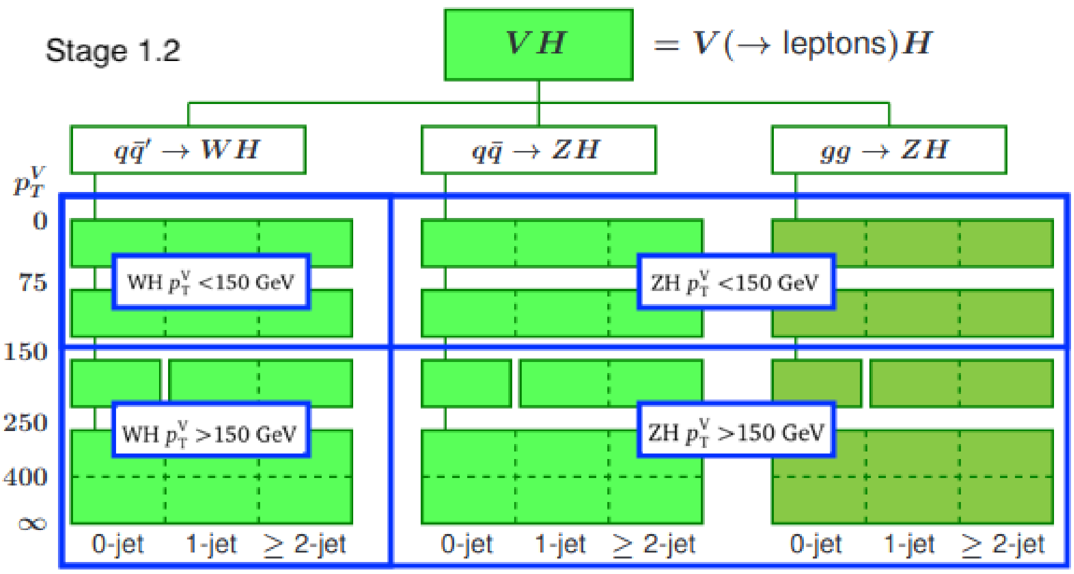



FIGURE 6.13: The four merged STXS bins are shown in blue boxes, based on the STXS stage 1.2 categorization for leptonic VH production. For bins shown in green boxes, the solid lines divide different bins and the dashed boundaries represent additional bins that could be used if there was enough sensitivity.

### 6.4.1 Event selection and categorization in VH STXS

The STXS analysis is using the same event selection criteria and categorization of the inclusive analysis from each channel described in Sec. 6.2, with additional selection on the reconstruction-level $p_{\mathrm{T}}^{\mathrm{V}}$ corresponding to the generator-level $p_{\mathrm{T}}^{\mathrm{V}}$ threshold described in Fig. 6.13.

The reconstruction-level $p_{\mathrm{T}}^{\mathrm{V}}$ is defined differently depending on the associated vector boson type and decay channel. In the WHSS and WH3l channels, the W $p_{\mathrm{T}}$ cannot be fully reconstructed due to the unobserved neutrinos, hence dedicated proxy variables are defined respectively for both channels (see Appendix B); for the ZH3l and ZH4l channels, the reconstruction-level Z $p_{\mathrm{T}}$ is defined as the $p_{\mathrm{T}}$ of the opposite sign same flavor di-lepton pair whose invariant mass is closest to the nominal Z boson mass. Some reconstruction-level bins are merged to maximize the signal sensitivity in the corresponding STXS bins. For example, the OSSF and SSSF categories from WH3l channel are merged into one bin for $p_{\mathrm{T}}^{\mathrm{V}}$ above 150 GeV (Merged $p_{\mathrm{T}}^{\mathrm{V}} > 150$ GeV) to avoid large statistical uncertainties.

Prior to the signal extraction, the level of compatibility between the STXS bins (generator-level bins) and the corresponding reconstruction-level bins for each channel is scrutinized by evaluating the signal purity in the bins. The purity provides an estimate on the migration effects between adjacent bins because of the limited resolution of the reconstructed $p_{\mathrm{T}}^{\mathrm{V}}$. The response matrix is defined as a two-dimensional histogram, with the generator-level bins specific to different production modes according to $p_{\mathrm{T}}^{\mathrm{V}}$ on the y-axis, and the corresponding reconstruction-level bins on the x-axis. In each bin is indicated the corresponding purity $P_i$ of the signal samples, given as:

$$P_i = \frac{N_i^{Gen,Reco}}{N_i^{Reco}} \tag{6.3}$$

where $N_i^{Gen,Reco}$ is the number of events originating from the generator- and reconstruction-level in the $i^{th}$ bin, divided by the number of events originating from the same reconstruction-level bin, $N_i^{Reco}$. The response matrices are shown in Fig. 6.14 separately for each data taking year; each column is normalized to one, meaning that all the generator-level bin contribution to a given reconstruction-level bin add up to one.

The migration effects observed between the two different generator-level WH $p_{\mathrm{T}}^{\mathrm{V}}$ bins, from the $e^{\pm}\mu^{\pm}$ and $\mu^{\pm}\mu^{\pm}$ 1-jet $p_{\mathrm{T}}^{\mathrm{V}}$ >150 GeV bins, are due to the proxy variable over-estimating the generator-level $p_{\mathrm{T}}^{\mathrm{V}}$ bin as most of the signal process is populated below $p_{\mathrm{T}}^{\mathrm{V}}$ <150 GeV (See Fig. B.1). Apart from these two bins, in general there are limited amount of bin-by-bin migrations observed between the generator-level bins; all the reconstruction-level bins targeting specific production mode are sensitive to the corresponding generator-level bins.

### 6.4.2 Results

The signal strength measurement on STXS bins is performed by using the signal region categorization described in Sec. 6.4.1, and the same background control regions, signal-discriminating kinematic observables, and sources of uncertainties as the inclusive signal strength measurements. However, the theoretical uncertainties such as the PDF and QCD scale affecting the normalization of the expected signal process is not considered in the measurement since cross sections are measured in place of signal strengths.

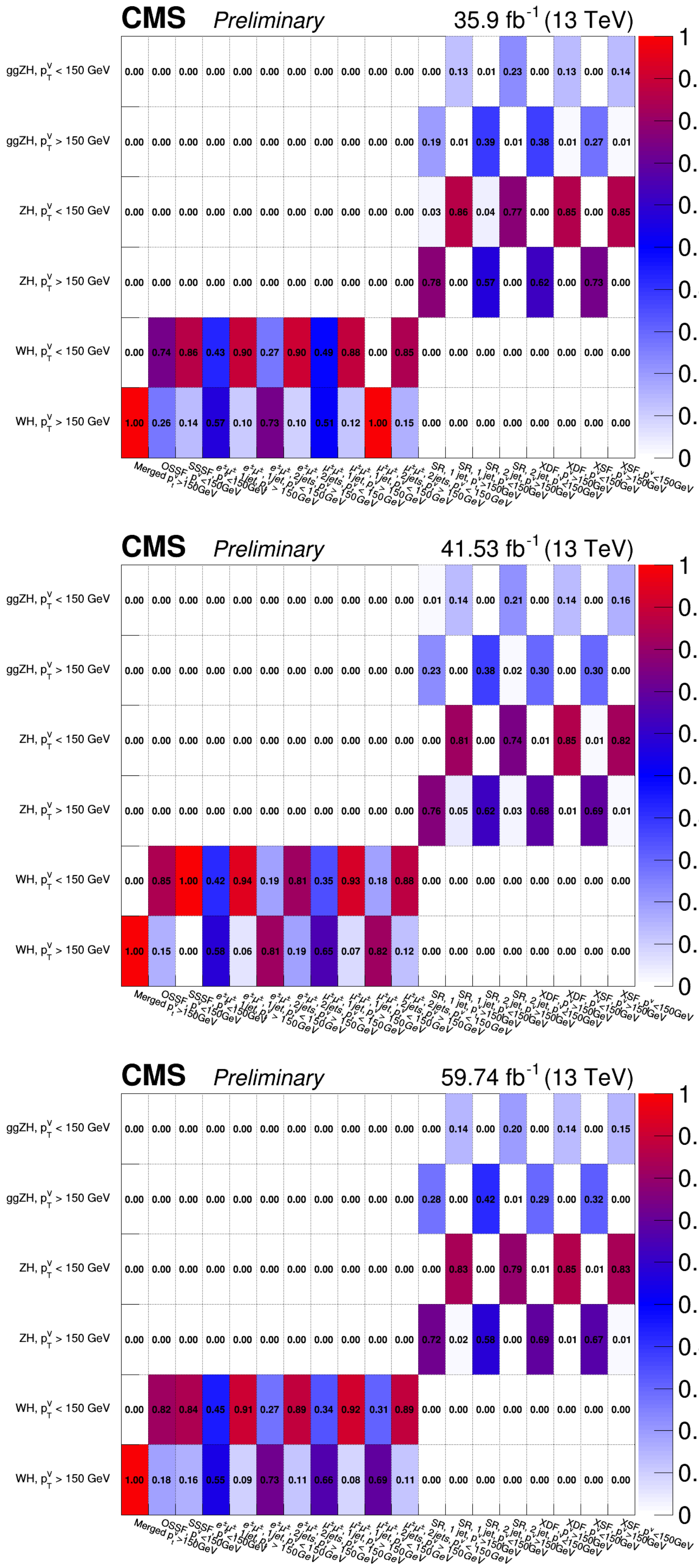
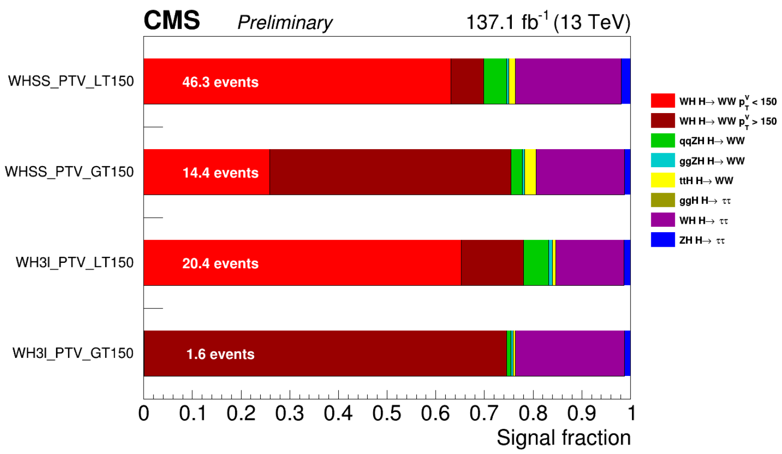


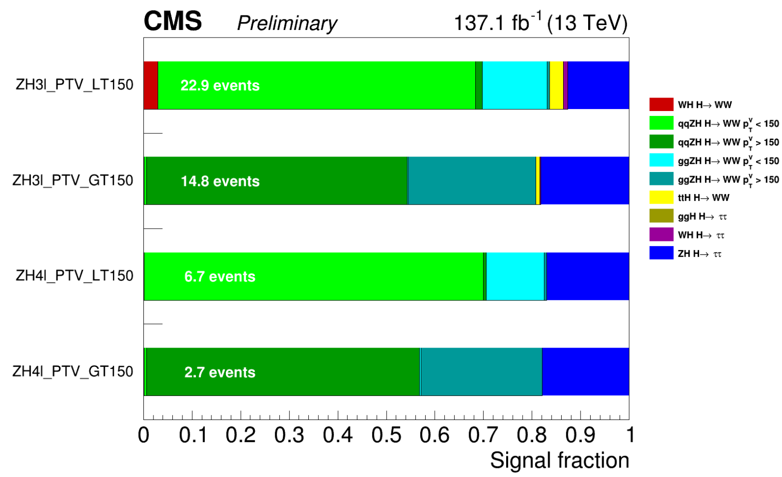

FIGURE 6.14: Response matrices for V(→ leptons)H leptonic channel with V = W and Z bosons, mapping the STXS signal binned in different production modes and $p_T^V$ (y-axis) to the signal regions binned in respective reconstructed $p_\mathrm{T}^\mathrm{V}$ in different event categories (x-axis), for all three years.

In this analysis, only the VH signal samples with $H \rightarrow W^+W^-$ decay are considered as signal processes in different kinematic regions, this is different with respect to the inclusive analysis described in Sec. 6.3, where a single signal strength was obtained using other Higgs boson production modes and decays.

The expected number of signal events in each reconstructed bin specific to the production modes defined in different $p_T^V$ is shown in Fig. 6.15. The WH reconstruction-level bins have about 70% of the signal events coming from the corresponding generator-level bins, with negligible ZH generator-level bins contribution (< 5%); the ZH reconstruction-level bins have about 80% signal events from ZH generator-level bins, including the gluon-initiated associated ZH production. There are events from $H \rightarrow \tau^+\tau^-$ decay (about 40%) contributing to both WH and ZH reconstruction-level bins. Those events and other non-STXS binned signal production events are constrained to their SM expectation in the fit.




(A) Associated WH production mechanism




(B) Associated ZH production mechanism

FIGURE 6.15: The expected signal fraction for $p_T^V < (>)150$ GeV reconstruction-level bins for different production modes.

The signal strength modifier for each STXS bin is extracted by performing a binned maximum likelihood fit simultaneously to all categories and all years, and the measured signal strength modifiers are summarized in Tab. 6.8.

TABLE 6.8: The measured signal strength modifiers and expected (observed) significance for STXS bins is given by the Fig. 6.13.

| STXS bin | Expected significance | Observed significance | $\sigma/\sigma_{SM}$ |
|---|---|---|---|
| WH $p_T^V$ <150 GeV | 1.24 | 1.64 | $1.5^{+1.0}_{-0.9}$ |
| WH $p_T^V$ >150 GeV | 0.83 | 2.23 | $3.6^{+1.8}_{-1.6}$ |
| ZH $p_T^V$ <150 GeV | 1.59 | 4.37 | $3.4^{+1.1}_{-1.0}$ |
| ZH $p_T^V$ >150 GeV | 1.18 | 0.83 | $0.8^{+1.2}_{-0.9}$ |

The signal strength for leptonic VH production is obtained by applying a signal strength modifier to each $p_T^V$ STXS bin. For the $p_T^V$ <150 GeV bin, the measured signal strength modifier, with the uncertainty separately reported in term of theoretical systematic, experimental systematic and statistical uncertainties is:

$$\mu_{p_T^V\,<\,150GeV} = 2.65^{+0.57}_{-0.55}(stat)^{+0.38}_{-0.32}(exp)^{+0.08}_{-0.07}(theo)$$

with an observed significance of 4.7 $\sigma$ to be compared with expected significance of 2.0 $\sigma$; while for the $p_T^V > 150$ GeV bin, the measured signal strength modifier is

$$\mu_{p_T^V\,>\,150GeV} = 1.56^{+0.85}_{-0.77}(stat)^{+0.43}_{-0.40}(exp)^{+0.11}_{-0.09}(theo)$$

with an observed significance of 1.8 $\sigma$ to be compared with the expected value of 1.5 $\sigma$. Finally, a combined signal strength of $2.11^{+0.46}_{-0.43}$ is measured applying a single signal strength in all STXS bins, corresponding to an observed (expected) significance of 5.3 (2.7) $\sigma$. While the combined signal strength differs from the inclusive result due to the STXS event categorization, the two results agree within the uncertainties. The inclusive production mode, split production mode, and combined signal strengths are summarized graphically in Fig. 6.16.

The dominant impacts on the WH signal strength, in particular, the $p_T^V$ <150 GeV bin are the non-prompt background normalization, WZ and Z$\gamma$ background normalization, and statistical uncertainties from the most sensitive bins of the WH3l fit template. On the other hand, the statistical uncertainties have a large impact on the $p_T^V$ >150 GeV signal strength, along with JES and b-tagging efficiency uncertainties, driven by both WHSS and WH3l channels.

Concerning the ZH signal strength, the electron efficiency and WZ background normalization uncertainties have the largest impact on the $p_T^V$ <150 GeV bin, while the QCD scale, JES,

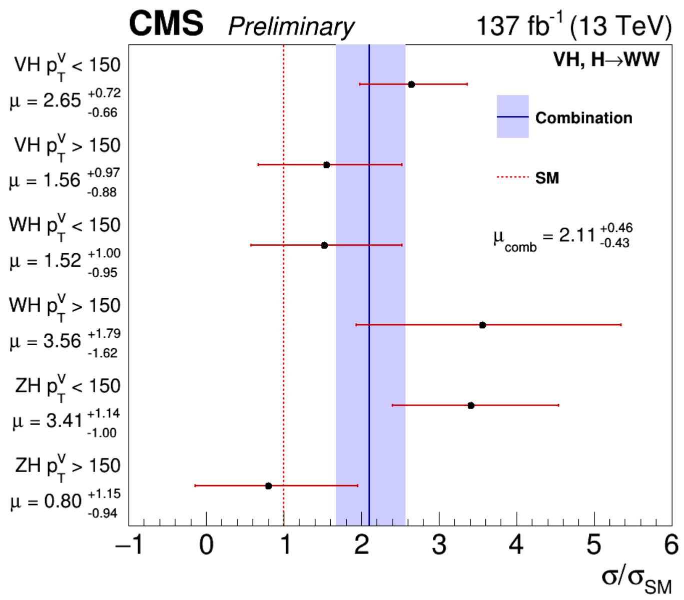



FIGURE 6.16: Comparison between the combined signal strength, and the signal strengths for each $p_T^V$ bin and production mode (WH, ZH and VH). The vertical dashed line corresponds to the SM prediction and associated theoretical uncertainty.

WZ and ZZ background normalization, and electron efficiency uncertainties have the largest impact on the $p_T^V$ >150 GeV bin.

The dominant impacts on the VH signal strength is similar to the WH signal strength's impacts for the $p_T^V$ <150 GeV bin; for $p_T^V$ >150 GeV, the main impacts are mostly statistical uncertainties, followed by JES, and QCD scale.

The correlations between nuisance parameters and signal strengths for the fit applying a single signal strength and separate signal strengths in each channel are shown in Fig. 6.17. For both cases, the correlations between nuisance parameters follow similar trends to the inclusive fit. In the case of using separate signal strength fit, the WH signal strengths in both $p_T^V$ <(>)150 GeV bins are fairly anti-correlated, because of the migration effects caused by the W $p_T$ proxy variable; the ZH signal strengths in both $p_T^V$ <(>)150 GeV bins are uncorrelated, showing that the Z $p_T$ proxy provides good discrimination.

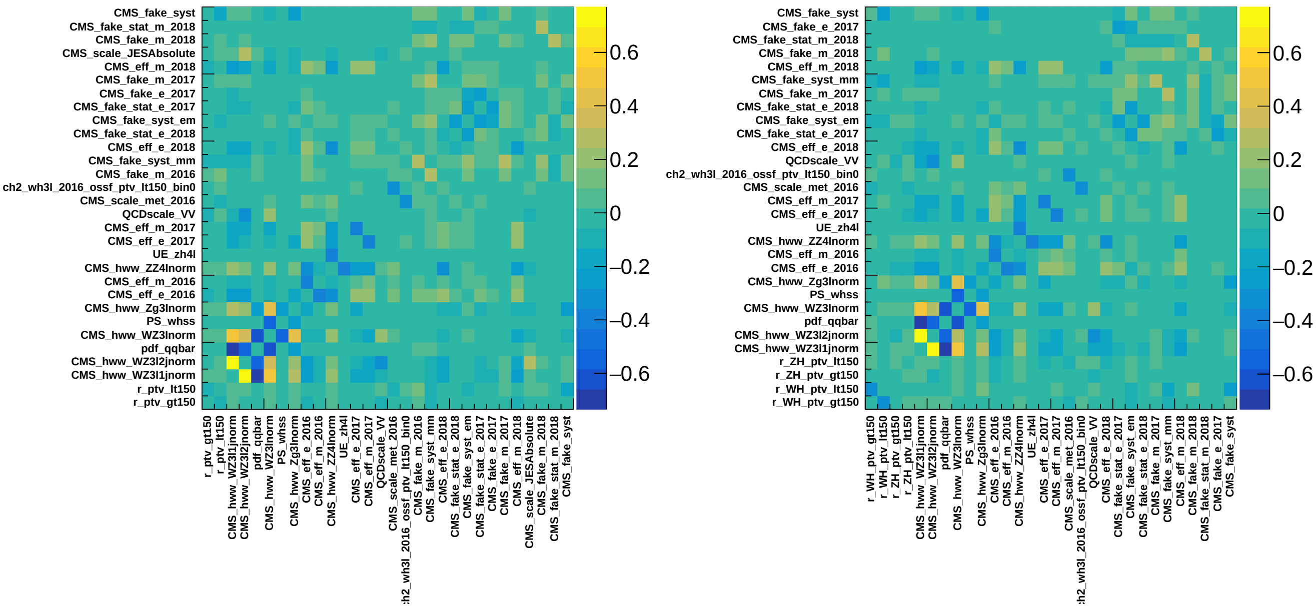



FIGURE 6.17: The correlations between nuisance parameters for the fit applying (Left) single signal strength to all STXS binned signal processes and (Right) separate signal strengths in each production mode. The correlations are shown between single(multiple) signal strength(s) each for $p_T^V >(<) 150$ GeV bin and the 29(26) nuisance parameters with highest correlation.

# Chapter 7

# Summary

The measurement of all known Higgs boson final states, in particular those with lower expected yield and complex experimental signature, has become the most relevant goal in the near future LHC data analysis program. One of these final states is the same sign di-lepton final state. It consists of two same sign leptons, a moderate missing transverse energy and a pair of jets, and it is directly sensitive to the associated WH production channel with a leptonically decaying W boson, followed by the $H \rightarrow W^+W^-$ decay.

In this thesis, the study of the Higgs boson production measurement using the same sign di-lepton final state is carried out for the first time, on the full Run-II dataset collected at center of mass energy of 13 TeV, with an integrated luminosity of 137 $\text{fb}^{-1}$.

The same sign di-lepton topology provides sensitivity to the measurement as the final state is less prone to the dominant electroweak backgrounds; the process is also produced at larger rate via the $H \rightarrow W^+W^-$ decay channel. The sensitivity of the analysis is improved by using a multivariate technique in separating non-prompt from prompt leptons; a Higgs mass proxy is used to improve the discrimination between the signal and background. The most important background such as WZ background is normalized in dedicated data control sample; the non-prompt background is modeled directly from the data sample via the fakeable objects method; and the charge mis-identification on the backgrounds containing same charge electron and muon pair and W$\gamma^\star$ are also estimated from data. The study on this final state reports a measured signal strength:

$$\mu_{wh} = 0.89^{+0.64}_{-0.62}(stat)^{+0.74}_{-0.78}(exp) \tag{7.1}$$

with an observed significance of 0.9 $\sigma$. The result is consistent with the MC expectation. The measurement is limited by the limited statistics, and the main systematic uncertainties come from the fake lepton systematic, background modeling such as non-prompt, WZ and W$\gamma^\star$ backgrounds, QCD scale, and JES scale.

An inclusive measurement targeting the associated leptonic VH production modes is performed by using events with the Higgs boson decaying to a pair of W bosons or tau leptons.

The combined measurement profits from the clean signatures and high signal purity from the leptonic final states such as the WH3l, the ZH4l, the WHSS and the ZH3l channels.

Using the same dataset, the measured signal strength modifier for inclusive leptonic VH production is:

$$\mu_{comb} = 1.85^{+0.33}_{-0.32}(stat)^{+0.27}_{-0.25}(exp)^{+0.10}_{-0.07}(theo) \tag{7.2}$$

corresponding to an observed significance of 4.7 $\sigma$. The discrepancy between the observed and expected significance is mainly driven by the excess of data observed in the ZH3l channel due to background fluctuation.

In addition, the measurement is performed in the exclusive phase space regions within the Simplified Template Cross Section (STXS) framework. By using the stage 1.2 STXS definition, used for the first time for this analysis, the Higgs boson production cross section is quantified in terms of a signal strength modifier, defined as the ratio of the observed cross section over the prediction from the SM, measured with respect to the transverse momentum of the associated vector boson $p_{\mathrm{T}}^{\mathrm{V}}$. The measured signal strength modifier for $p_{\mathrm{T}}$ <150 GeV bin is

$$\mu_{p_{\mathrm{T}}^{\mathrm{V}} < 150GeV} = 2.65^{+0.57}_{-0.55}(stat)^{+0.38}_{-0.32}(exp)^{+0.08}_{-0.07}(theo)$$

with observed significance of 4.7 $\sigma$ while for the $p_{\mathrm{T}}^{\mathrm{V}} > 150$ GeV bin, the measured signal strength modifier is

$$\mu_{p_{\mathrm{T}}^{\mathrm{V}} > 150GeV} = 1.56^{+0.85}_{-0.77}(stat)^{+0.43}_{-0.40}(exp)^{+0.11}_{-0.09}(theo)$$

with an observed significance of 1.8 $\sigma$.

The inclusive and STXS measurements are compatible with each other within their uncertainties; the signal strength for different Higgs boson associated production processes are measured and found to be consistent with SM predictions within 1.5 $\sigma$.

Most of the measurements performed in the thesis are mainly limited by the statistical and systematic uncertainties arising from background modeling such as the non-prompt background. The precision of the measurement can be improved by performing the same analysis on a larger statistics dataset, which will be collected during the Run-III or HL-LHC period (see Fig. 3.2). Meanwhile, the fake lepton systematic uncertainty can be further reduced by using the combination of multivariate method and deep neural network technique to discriminate the prompt from non-prompt leptons.

A full description of the Higgs boson properties requires measuring both inclusive and differential cross sections in all production modes. The signal sensitivity in these measurements can be further improved by combining results from different decay channels across multiple final states. In the near future, the leptonic VH production results will be combined with the measurements targeting the main Higgs production modes such as ggH and VBF processes, to complete the global signal strength measurement and to contribute to the measurement of the Higgs boson properties.

# Appendix A

# Generator level signal study setup

This appendix presents the setup for the generator-level signal study in Sec. 5.1.

The signal process is the Higgs boson production in association with leptonic W boson, followed by Higgs boson decays into a pair of W bosons with at least one W boson decays leptonically and the other hadronically. The MADGRAPH5_aMC@NLO generator is used for the study, the simulated signal samples are generated separately according to the charge of the associated W boson, with Higgs boson mass of 125 GeV is assumed. The Higgs production process is generated up to 2 jets at LO accuracy; due to the mass threshold effect, one of the W boson from the Higgs decay is off-shell. The final state contains a pair of same charged leptons, two neutrinos, and four quarks including the extra quarks from ISR. An example for signal process run card is given below:

```
import model sm
define p = g u c d s u~ c~ d~ s~
define j = g u c d s u~ c~ d~ s~
define l+ = e+ mu+
define l- = e- mu-
define vl = ve vm vt
define vl~ = ve~ vm~ vt~
generate p p > w+ h , w+ > l+ vl , h > l+ vl j j @1
add process p p > w+ h j , w+ > l+ vl , h > l+ vl j j @2
add process p p > w+ h j j , w+ > l+ vl , h > l+ vl j j @3
output sshwplus012j_M125_4f_LO_MLM -nojpeg
```

LISTING A.1: The MADGRAPH5_aMC@NLO generator's process card defining the signal process with a positively charged associated W boson.

The signal process is generated with 20000 events each, with NNPDF 3.0 [105] PDF. The generated events are passed to PYTHIA8 [66] to simulate the effects of PS, hadronization and multi-parton interactions; the tune CUEP8M1 [106] is used and the matching between the ME to PS is performed with MLM scheme [67] at LO (with matching scale = 30 GeV). The detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [69]. The simulated samples are in NanoAOD format [121].

All generator-level physics objects are identified using MC particle numbering scheme (pdgId) [7], together with the particle history information provided in NanoAOD format. The leptons and neutrinos are identified as a stable particle (with status code = 1), and decayed promptly from the hard-scattering process.

The bosons are identified to be the last copy in the decay chain and the boson's pdgId, and their mother's pdgId. If there is no Higgs boson mother found in the boson's decay chain, the W boson is identified as the associated produced W boson, otherwise they are identified as W boson(s) originated from the Higgs decay.

The generator-level jets (GenJet), produced by the anti-$k_t$ jet algorithm [83], whose matched ($\Delta R < 0.4$) to the quarks originating from the hard-scattering process are considered in the study.

The identified objects are sorted in descending $p_T$ order, and used to validate the identification criteria. The identified leading (trailing) lepton and neutrino are used to reconstructing the W boson, a pronounced W mass peak is observed on the invariant mass distribution of the leading pair, while a W mass peak and an off-shell structure are observed on the invariant mass distribution of the trailing pair, as shown in Fig. A.1. The identification of generator-level objects is validated and used for signal characterization study.

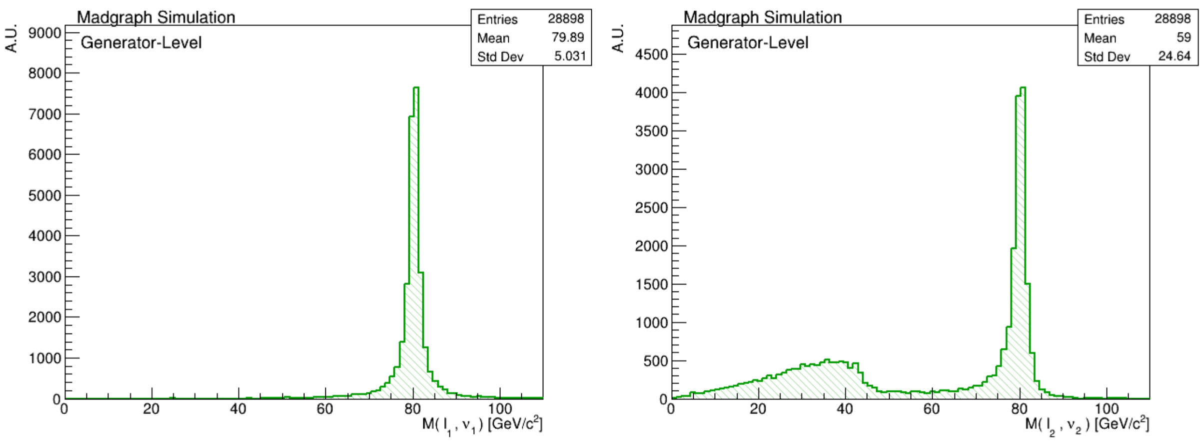



FIGURE A.1: The generator level invariant mass distributions for leading (left) and trailing (right) the leptons and neutrinos.

# Appendix B

# Proxy variables for associated WH production

This appendix presents the proxy variable definition used to define the reconstructed category in Sec. 6.4.1; its definition depends on the final state considered.

The WH3l channel W $p_{\mathrm{T}}$ is difficult to resolve given the ambiguity from the three neutrinos in the final state. Instead, the $l_W$ is defined as the lepton pointing away from the opposite sign di-lepton system with smallest angular separation ($\Delta R$ given by Eq. (3.6)), is used as a proxy for W $p_{\mathrm{T}}$.

The proxy variable for WHSS channel can be fairly constructed given the two neutrinos in the final state are separated far away from each other. For a given signal candidate event, the lepton and neutrino from the associated W decay are named $l_W$ and $\nu_W$, while the lepton and neutrino from the Higgs decay are named $l_H$ and $\nu_H$. The same sign lepton pair, $l_W$ and $l_H$ are identified as the lepton with the largest and smallest azimuthal angle separation with respect to the jet(s) system, which is given by the jet(s) with the closest relative distance ($\Delta R$ given by Eq. (3.6)) to each other, provided at least two jets are fully reconstructed in the event; else it is given by a jet for event with at least one jet.

Due to the presence of two neutrinos in the final state, the neutrino four-momenta cannot be precisely determined. However, it is possible to have an approximate estimation of these quantities by performing some assumptions:

- $\nu_H$ is collinear with $l_H$;
- the invariant mass of $\nu_H$, $l_H$, and the two light quarks, $qq$ should correspond to the Higgs boson mass;
- the $p_x$ and $p_y$ components of the $\nu_W$ should correspond to the $p_{\mathrm{T}}^{\mathrm{miss}}$'s x and y components;

- the $p_z$ component of the $\nu_W$ is recovered by imposing that the $\nu_W$ and $l_W$ invariant mass should correspond to the nominal W boson mass ($M_W = 80.4$ GeV).

The four-momentum of the neutrino $\nu_W$ is inferred from a global quantity, $p_{\mathrm{T}}^{\mathrm{miss}}$ by subtracting $\nu_H$; for events with at least two reconstructed jets, the $\nu_H$ approximation is improved by using the nominal Higgs boson mass as a constraint via an in-situ scaling factor $k$:

$$\nu_W = \begin{cases} p_{\mathrm{T}}^{\mathrm{miss}} - l_H & , N_{jet} = 1 \\ p_{\mathrm{T}}^{\mathrm{miss}} - k l_H & , N_{jet} \geq 2 \end{cases} \tag{B.1}$$

and $k$ is given by:

$$\begin{aligned} \left\| l_H + \nu_H + jj \right\| &= 125\,\mathrm{GeV} \\ \left\| l_H + \left(k l_H\right) + jj \right\| &= 125\,\mathrm{GeV} \\ \left\| (k+1)\, l_H + jj \right\| &= 125\,\mathrm{GeV} \\ \rightarrow k &= \frac{125\,\mathrm{GeV}}{\left\| l_H + jj \right\|} - 1 \end{aligned} \tag{B.2}$$

where $jj$ is the four-momentum of the jets system.

The $p_z$ component of the $\nu_W$ from Eq. (B.1) is recovered by solving the lepton-neutrino system [122]:

$$M_W^2 = p_{l_W}^2 + p_{\nu_W}^2 \tag{B.3}$$

and the solution[1] is used to calculate the neutrino energy:

$$E_{\nu_W} = \sqrt{p_x^2 + p_y^2 + p_z^2} \tag{B.4}$$

completing the kinematics of the massless neutrino. Finally, the proxy variable for reconstruction-level W $p_{\mathrm{T}}$ is obtained by summing the transverse momentum of $l_W$ and $\nu_W$.

The performance of the proxy variable is shown in Fig. B.1. The proxy variable is moderately correlating with the generator-level W boson $p_{\mathrm{T}}$ with a correlation factor of 0.64; the proxy variable over-estimates the generator-level $p_{\mathrm{T}}$ in the range of [50, 150] GeV and under-estimates $p_{\mathrm{T}}^{\mathrm{V}}$ outside of that range. Overall, the proxy variable is able to describe the generator-level W $p_{\mathrm{T}}$ within 1 $\sigma$ resolution.

[1]The $p_z$ is given by the minimal value of the positive solution, otherwise is given by the real part of the negative solution.

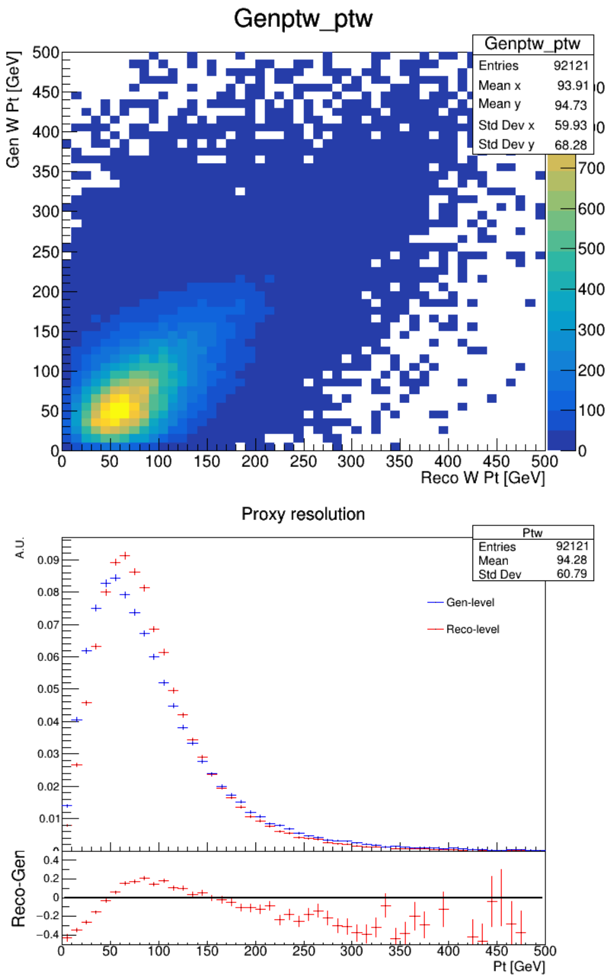



FIGURE B.1: The top plot shows the correlation between generator-level W $p_T$ (y-axis) and reconstruction-level W $p_T$ (x-axis), given by the proxy variable; the bottom plot shows the distribution of generator-level W $p_T$ overlaid with the proxy variable, the bottom plot shows the resolution of proxy variable $p_T$.

# Bibliography


[1] Steven R. James et al. “Hominid Use of Fire in the Lower and Middle Pleistocene: A Review of the Evidence [and Comments and Replies]”. In: *Current Anthropology* 30.1 (1989), pp. 1–26. DOI: 10.1086/203705. eprint: https://doi.org/10.1086/203705. URL: https://doi.org/10.1086/203705 (cit. on p. 1).

[2] Serguei Chatrchyan et al. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on pp. 1, 23, 54).

[3] Georges Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex] (cit. on pp. 1, 23).

[4] Matthew D. Schwartz. “TASI Lectures on Collider Physics”. In: *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016*. Ed. by Rouven Essig and Ian Low. 2018, pp. 65–100. DOI: 10.1142/9789813233348_0002. arXiv: 1709.04533 [hep-ph] (cit. on p. 2).

[5] Albert M. Sirunyan et al. “Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 791 (2019), p. 96. DOI: 10.1016/j.physletb.2018.12.073. arXiv: 1806.05246 [hep-ex] (cit. on pp. 2, 28–30).

[6] Sheldon L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579 –588. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: http://www.sciencedirect.com/science/article/pii/0029558261904692 (cit. on p. 6).

[7] M. Tanabashi, K. Hagiwara, and et.al. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001 (cit. on pp. 6, 18, 24–26, 140).

[8] G. Arnison et al. “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV”. In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2 (cit. on p. 7).

[9] M. Banner et al. “Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider”. In: *Phys. Lett. B* 122 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2 (cit. on p. 7).

[10] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413. URL: https://link.aps.org/doi/10.1103/PhysRev.105.1413 (cit. on p. 11).

[11] M. Gell-Mann. “The interpretation of the new particles as displaced charge multiplets”. In: *Nuovo Cim.* 4.S2 (1956), pp. 848–866. DOI: 10.1007/BF02748000 (cit. on p. 12).

[12] Makoto Kobayashi and Toshihide Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. DOI: 10.1143/PTP.49.652. eprint: https://academic.oup.com/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf. URL: https://doi.org/10.1143/PTP.49.652 (cit. on p. 13).

[13] Jakob Schwichtenberg. *Physics from Symmetry*. Undergraduate Lecture Notes in Physics. Cham: Springer International Publishing, 2018. DOI: 10.1007/978-3-319-66631-0 (cit. on pp. 15, 19).

[14] Tancredi Carli, Klaus Rabbertz, and Steffen Schumann. “Studies of Quantum Chromodynamics at the LHC”. In: *The Large Hadron Collider: Harvest of Run 1*. Ed. by Thomas Schörner-Sadenius. Cham: Springer International Publishing, 2015, pp. 139–194. DOI: 10.1007/978-3-319-15001-7_5. URL: https://doi.org/10.1007/978-3-319-15001-7_5 (cit. on p. 16).

[15] Michael E Peskin and Daniel V Schroeder. *An introduction to quantum field theory*. Includes exercises. Boulder, CO: Westview, 1995. URL: https://cds.cern.ch/record/257493 (cit. on p. 19).

[16] Huong Nguyen. “Search for the Standard Model Higgs Boson in Leptons plus Jets Final States”. In: (Jan. 2014). DOI: 10.2172/1155186. URL: https://www.osti.gov/biblio/1155186 (cit. on p. 23).

[17] Serguei Chatrchyan et al. “Observation of a New Boson with Mass Near 125 GeV in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV”. In: *JHEP* 06 (2013), p. 081. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571 [hep-ex] (cit. on p. 23).

[18] Georges Aad et al. “Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment”. In: *Eur. Phys. J. C* 76.1 (2016), p. 6. DOI: 10.1140/epjc/s10052-015-3769-y. arXiv: 1507.04548 [hep-ex] (cit. on p. 23).

[19] Georges Aad et al. "Evidence for the spin-0 nature of the Higgs boson using ATLAS data". In: *Phys. Lett. B* 726 (2013), pp. 120–144. DOI: 10.1016/j.physletb.2013.08.026. arXiv: 1307.1432 [hep-ex] (cit. on p. 23).

[20] Vardan Khachatryan et al. "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV". In: *Eur. Phys. J. C* 75.5 (2015), p. 212. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex] (cit. on pp. 23, 27).

[21] Vardan Khachatryan et al. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". In: *Phys. Rev. D* 92.1 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441 [hep-ex] (cit. on p. 23).

[22] Albert M Sirunyan et al. "Observation of $t\bar{t}$H production". In: *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex] (cit. on pp. 25, 28).

[23] CMS. *Website of the LHC Higgs Cross-section Working Group with public cross section results*. 2021. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG?redirectedfrom=LHCPhysics.LHCHXSWG (visited on 03/22/2021) (cit. on p. 26).

[24] Albert M Sirunyan et al. "Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.5 (2019), p. 421. DOI: 10.1140/epjc/s10052-019-6909-y. arXiv: 1809.10733 [hep-ex] (cit. on p. 27).

[25] Georges Aad et al. "Combined Measurement of the Higgs Boson Mass in $pp$ Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex] (cit. on pp. 27, 77).

[26] "Study of the Higgs boson properties and search for high-mass scalar resonances in the $H \to ZZ^* \to 4\ell$ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: (Aug. 2016) (cit. on p. 27).

[27] Serguei Chatrchyan et al. "Measurement of the Properties of a Higgs Boson in the Four-Lepton Final State". In: *Phys. Rev. D* 89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex] (cit. on p. 27).

[28] "Measurement of fiducial, differential and production cross sections in the $H \to \gamma\gamma$ decay channel with 13.3 $\text{fb}^{-1}$ of 13 TeV proton-proton collision data with the ATLAS detector". In: (Aug. 2016) (cit. on p. 27).

[29] "Measurement of the associated production of a Higgs boson and a pair of top-antitop quarks with the Higgs boson decaying to two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (2018) (cit. on p. 27).

[30] Pushpalatha C. Bhat. "Multivariate Analysis Methods in Particle Physics". In: *Ann. Rev. Nucl. Part. Sci.* 61 (2011), pp. 281–309. DOI: 10.1146/annurev.nucl.012809.104427 (cit. on p. 28).

[31] Daniel Guest et al. "Jet Flavor Classification in High-Energy Physics with Deep Neural Networks". In: *Phys. Rev. D* 94.11 (2016), p. 112002. DOI: 10.1103/PhysRevD.94.112002. arXiv: 1607.08633 [hep-ex] (cit. on pp. 28, 66).

[32] Albert M Sirunyan et al. "Evidence for Higgs boson decay to a pair of muons". In: *JHEP* 01 (2021), p. 148. DOI: 10.1007/JHEP01(2021)148. arXiv: 2009.04363 [hep-ex] (cit. on p. 28).

[33] A. M. Sirunyan et al. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex] (cit. on p. 28).

[34] *Measurements of the Higgs boson production cross section via Vector Boson Fusion and associated WH production in the* $WW^* \to \ell\nu\ell\nu$ *decay mode with the ATLAS detector at* $\sqrt{s} = 13$ *TeV*. Tech. rep. ATLAS-CONF-2016-112. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2231811 (cit. on p. 28).

[35] Morad Aaboud et al. "Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in $pp$ collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 789 (2019), pp. 508–529. DOI: 10.1016/j.physletb.2018.11.064. arXiv: 1808.09054 [hep-ex] (cit. on p. 28).

[36] Frank Tackmann et al. "Simplified template cross sections". In: (2016). URL: https://cds.cern.ch/record/2138079 (cit. on p. 29).

[37] CERN Bulletin. "The particle suppliers. Les fournisseurs de particules". In: BUL-NA-2010-077. 14/2010 (2010), p. 03. URL: https://cds.cern.ch/record/1255151 (cit. on p. 31).

[38] I Efthymiopoulos. *Overview of the ATLAS detector at LHC*. Tech. rep. ATL-CONF-99-002. 7. version revised com-conf-99-002 version1 on Tue Jun 22 21:18:36 MET DST 1999. Geneva: CERN, 1999. URL: http://cds.cern.ch/record/409257 (cit. on p. 32).

[39] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004 (cit. on pp. 32, 35).

[40] Jr. Alves A.Augusto et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005 (cit. on p. 32).

[41] K. Aamodt et al. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002 (cit. on p. 32).

[42] Andreas Hoecker. "Physics at the LHC Run-2 and Beyond. Physics at the LHC Run-2 and Beyond". In: arXiv:1611.07864 (2016). Lecture notes from the 2016 European School of High-Energy Physics, 15-28 June 2016, Skeikampen, Norway (61 pages, 56 figures), 153–212. 61 p. DOI: 10.23730/CYRSP-2017-005.153. URL: http://cds.cern.ch/record/2236645 (cit. on p. 33).

[43] A.M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det] (cit. on pp. 37, 50, 63).

[44] D Campi et al. "Commissioning of the CMS Magnet". In: *IEEE Trans. Appl. Supercond.* 17.2 (2007), pp. 1185–1190. DOI: 10.1109/TASC.2007.897754. URL: http://cds.cern.ch/record/1102022 (cit. on p. 37).

[45] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det] (cit. on pp. 38, 39, 49, 50).

[46] W. Adam et al. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16.02 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304 [physics.ins-det] (cit. on p. 38).

[47] "2017 tracking performance plots". In: (2017). URL: https://cds.cern.ch/record/2290524 (cit. on p. 38).

[48] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375 (cit. on p. 39).

[49] G. L. Bayatian et al. "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". In: (2006) (cit. on pp. 40, 53).

[50] A.M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det] (cit. on pp. 41, 43, 57).

[51] G. Charpak et al. "High-accuracy localization of minimum ionizing particles using the cathode-induced charge centre-of-gravity read-out". In: *Nuclear Instruments and Methods* 167.3 (1979), pp. 455 –464. DOI: https://doi.org/10.1016/0029-554X(79)90227-1. URL: http://www.sciencedirect.com/science/article/pii/0029554X79902271 (cit. on p. 42).

[52] Vardan Khachatryan et al. "The CMS trigger system". In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det] (cit. on pp. 43, 45).

[53] S. Dasu et al. "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems". In: (Dec. 2000) (cit. on p. 43).

[54] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. Geneva: CERN, 2002. URL: http://cds.cern.ch/record/578006 (cit. on p. 43).

[55] V. Halyo et al. “GPU Enhancement of the Trigger to Extend Physics Reach at the LHC”. In: *JINST* 8 (2013), P10005. DOI: 10.1088/1748-0221/8/10/P10005. arXiv: 1305.4855 [physics.ins-det] (cit. on p. 44).

[56] “The Grid: A system of tiers”. In: (2012). URL: https://cds.cern.ch/record/1997396 (cit. on p. 46).

[57] John C. Collins, Davison E. Soper, and George F. Sterman. “Factorization of Hard Processes in QCD”. In: vol. 5. 1989, pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313 (cit. on p. 48).

[58] Stefano Forte and Graeme Watt. “Progress in the Determination of the Partonic Structure of the Proton”. In: *Ann. Rev. Nucl. Part. Sci.* 63 (2013), pp. 291–328. DOI: 10.1146/annurev-nucl-102212-170607. arXiv: 1301.6754 [hep-ph] (cit. on p. 48).

[59] John M. Campbell, J.W. Huston, and W.J. Stirling. “Hard Interactions of Quarks and Gluons: A Primer for LHC Physics”. In: *Rept. Prog. Phys.* 70 (2007), p. 89. DOI: 10.1088/0034-4885/70/1/R02. arXiv: hep-ph/0611148 (cit. on p. 48).

[60] Paolo Nason. “A New method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 (cit. on p. 48).

[61] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph] (cit. on p. 48).

[62] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph] (cit. on p. 48).

[63] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph] (cit. on p. 48).

[64] Zoltan Nagy and Davison E. Soper. “What is a parton shower?” In: *Phys. Rev. D* 98.1 (2018), p. 014034. DOI: 10.1103/PhysRevD.98.014034. arXiv: 1705.08093 [hep-ph] (cit. on p. 49).

[65] S. Catani and L. Trentadue. “Resummation of the QCD perturbative series for hard processes”. In: *Nuclear Physics B* 327.2 (1989), pp. 323 –352. DOI: https://doi.org/10.1016/0550-3213(89)90273-3. URL: http://www.sciencedirect.com/science/article/pii/0550321389902733 (cit. on p. 49).

[66] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on pp. 49, 139).

[67] Johan Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *Eur. Phys. J. C* 53 (2008), pp. 473–500. DOI: 10.1140/epjc/s10052-007-0490-5. arXiv: 0706.2569 [hep-ph] (cit. on pp. 49, 77, 139).

[68] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph] (cit. on pp. 49, 77).

[69] V.N. Ivanchenko. "Geant4 toolkit for simulation of HEP experiments". In: *Nucl. Instrum. Meth. A* 502 (2003). Ed. by V.A. Ilyin, V.V. Korenkov, and D. Perret-Gallix, pp. 666–668. DOI: 10.1016/S0168-9002(03)00538-2 (cit. on pp. 49, 139).

[70] G. Boudoul et al. "Monte Carlo Production Management at CMS". In: *J. Phys. Conf. Ser.* 664.7 (2015), p. 072018. DOI: 10.1088/1742-6596/664/7/072018 (cit. on p. 49).

[71] Pierre Billoir and S. Qian. "Simultaneous pattern recognition and track fitting by the Kalman filtering method". In: *Nucl. Instrum. Meth. A* 294 (1990), pp. 219–228. DOI: 10.1016/0168-9002(90)91835-Y (cit. on p. 49).

[72] Wolfgang Adam et al. *Track Reconstruction in the CMS tracker*. Tech. rep. CMS-NOTE-2006-041. Geneva: CERN, 2006. URL: http://cds.cern.ch/record/934067 (cit. on p. 49).

[73] W. Adam et al. "RESEARCH NOTE FROM COLLABORATION: Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC". In: *Journal of Physics G Nuclear Physics* 31.9 (Sept. 2005), N9–N20. DOI: 10.1088/0954-3899/31/9/N01. arXiv: physics/0306087 [physics.data-an] (cit. on p. 51).

[74] Vardan Khachatryan et al. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV". In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det] (cit. on pp. 53, 54, 57).

[75] Vardan Khachatryan et al. "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV". In: *JINST* 10.08 (2015), P08010. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702 [physics.ins-det] (cit. on p. 54).

[76] A. Hoecker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: *arXiv e-prints*, physics/0703039 (Mar. 2007), physics/0703039. arXiv: physics/0703039 [physics.data-an] (cit. on p. 56).

[77] Kejun Zhu. "The CMS ECAL laser monitoring system". In: *2007 IEEE Nuclear Science Symposium and Medical Imaging Conference*. Vol. 1. 2007, pp. 153–158. DOI: 10.1109/NSSMIC.2007.4436306 (cit. on p. 57).

[78] Serguei Chatrchyan et al. "Alignment of the CMS tracker with LHC and cosmic ray data". In: *JINST* 9 (2014), P06009. DOI: 10.1088/1748-0221/9/06/P06009. arXiv: 1403.2286 [physics.ins-det] (cit. on p. 57).

[79] "Higgs boson production in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV". In: (Aug. 2020) (cit. on p. 59).

[80] Olaf Behnke et al. *Data analysis in high energy physics: a practical guide to statistical methods*. Weinheim: Wiley-VCH, 2013. DOI: 10.1002/9783527653416. URL: https://cds.cern.ch/record/1517556 (cit. on pp. 60, 61).

[81] G. Cowan. *Statistical data analysis*. 1998 (cit. on p. 61).

[82] Gavin P. Salam. "Towards Jetography". In: *Eur. Phys. J. C* 67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph] (cit. on p. 62).

[83] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-$k_t$ jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph] (cit. on pp. 62, 72, 140).

[84] Gerald C. Blazey et al. "Run II jet physics". In: *Physics at Run II: QCD and Weak Boson Physics Workshop: Final General Meeting*. May 2000, pp. 47–77. arXiv: hep-ex/0005012 (cit. on p. 62).

[85] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev. D* 48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 (cit. on p. 62).

[86] CMS Collaboration. "Determination of jet energy calibration and transverse momentum resolution in CMS". In: *Journal of Instrumentation* 6.11 (Nov. 2011), p. 11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det] (cit. on p. 63).

[87] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett. B* 659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph] (cit. on p. 63).

[88] B. Abbott et al. "Determination of the absolute jet energy scale in the D0 calorimeters". In: *Nucl. Instrum. Meth. A* 424 (1999), pp. 352–394. DOI: 10.1016/S0168-9002(98)01368-0. arXiv: hep-ex/9805009 (cit. on p. 64).

[89] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex] (cit. on pp. 65, 68).

[90] A.M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det] (cit. on pp. 66, 67).

[91] Serguei Chatrchyan et al. "Identification of b-Quark Jets with the CMS Experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex] (cit. on p. 66).

[92] Vardan Khachatryan et al. "Search for the associated production of the Higgs boson with a top-quark pair". In: *JHEP* 09 (2014). [Erratum: JHEP 10, 106 (2014)], p. 087. DOI: 10.1007/JHEP09(2014)087. arXiv: 1408.1682 [hep-ex] (cit. on p. 67).

[93] Daniele Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph] (cit. on p. 68).

[94] Serguei Chatrchyan et al. "Performance of CMS Muon Reconstruction in $pp$ Collision Events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det] (cit. on p. 68).

[95] Albert M Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". In: *JINST* 14.07 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex] (cit. on p. 69).

[96] H. Bakhshian et al. "Computing the contamination from fakes in leptonic final states". In: *CMS Physics Analysis Note* CMS-AN-10-261 (2010) (cit. on pp. 75, 85).

[97] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: 2/2017 (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph] (cit. on p. 77).

[98] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties". In: (July 2013). Ed. by S Heinemeyer et al. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph] (cit. on p. 77).

[99] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *JHEP* 03 (2013), p. 015. DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph] (cit. on p. 78).

[100] P. Kant et al. "HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions". In: *Comput. Phys. Commun.* 191 (2015), pp. 74–89. DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403 [hep-ph] (cit. on p. 78).

[101] Michal Czakon and Alexander Mitov. "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders". In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph] (cit. on p. 78).

[102] T. Gehrmann et al. "$W^+W^-$ Production at Hadron Colliders in Next to Next to Leading Order QCD". In: *Phys. Rev. Lett.* 113.21 (2014), p. 212001. DOI: 10.1103/PhysRevLett.113.212001. arXiv: 1408.5243 [hep-ph] (cit. on p. 78).

[103] John Campbell and Tobias Neumann. "Precision Phenomenology with MCFM". In: *JHEP* 12 (2019), p. 034. DOI: 10.1007/JHEP12(2019)034. arXiv: 1909.09117 [hep-ph] (cit. on p. 78).

[104] Fabrizio Caola et al. "QCD corrections to $W^+W^-$ production through gluon fusion". In: *Phys. Lett. B* 754 (2016), pp. 275–280. DOI: 10.1016/j.physletb.2016.01.046. arXiv: 1511.08617 [hep-ph] (cit. on p. 78).

[105] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph] (cit. on pp. 79, 139).

[106] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *Eur. Phys. J. C* 76.3 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex] (cit. on pp. 79, 139).

[107] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph] (cit. on p. 79).

[108] Albert M Sirunyan et al. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". In: *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex] (cit. on p. 79).

[109] Paul Lujan. "Van der Meer calibration of the CMS luminosity detectors in 2017". In: *PoS* ICHEP2018 (2019), p. 860. DOI: 10.22323/1.340.0860 (cit. on p. 100).

[110] *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. CMS-PAS-LUM-17-001. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257069 (cit. on p. 100).

[111] *CMS luminosity measurement for the 2017 data-taking period at* $\sqrt{s} = 13$ TeV. Tech. rep. CMS-PAS-LUM-17-004. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2621960 (cit. on p. 100).

[112] *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13$ TeV. Tech. rep. CMS-PAS-LUM-18-002. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2676164 (cit. on p. 100).

[113] Albert M Sirunyan et al. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV". In: *JHEP* 07 (2018), p. 161. DOI: 10.1007/JHEP07(2018)161. arXiv: 1802.02613 [hep-ex] (cit. on p. 101).

[114] Stefano Forte et al. *Gluon-gluon fusion*. Tech. rep. LHCHXSWG-DRAFT-INT-2016-011. 2016. URL: https://cds.cern.ch/record/2194224 (cit. on pp. 101, 102).

[115] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W (cit. on p. 102).

[116] Olaf Behnke et al., eds. *Data analysis in high energy physics: A practical guide to statistical methods*. Weinheim, Germany: Wiley-VCH, 2013 (cit. on p. 102).

[117] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1379837 (cit. on pp. 103, 122).

[118] I. Asimov. *The Complete Stories*. A Foundation book v. 1. Doubleday, 1990. URL: https://books.google.it/books?id=oyu-WOwT5wQC (cit. on p. 104).

[119] J.S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra". In: arXiv:1103.0354 (2011). Comments: Presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report, 115–120. 6 p. DOI: 10.5170/CERN-2011-006.115. URL: https://cds.cern.ch/record/1333496 (cit. on p. 122).

[120] Nicolas Berger et al. "Simplified Template Cross Sections - Stage 1.1". In: (2019). arXiv: 1906.02754 [hep-ph] (cit. on p. 129).

[121] Marco Peruzzi, Giovanni Petrucciani, and Andrea Rizzi. "The NanoAOD event data format in CMS". In: *J. Phys. Conf. Ser.* 1525.1 (2020), p. 012038. DOI: 10.1088/1742-6596/1525/1/012038 (cit. on p. 139).

[122] E. Richter-Was and Z. Was. "W production at LHC: lepton angular distributions and reference frames for probing hard QCD". In: *Eur. Phys. J. C* 77.2 (2017), p. 111. DOI: 10.1140/epjc/s10052-017-4649-4. arXiv: 1609.02536 [hep-ph] (cit. on p. 142).