

---

# Anomalous couplings of the Higgs boson in the $WW$ decay channel with the CMS experiment at the LHC

---

ACOPLAMIENTOS ANÓMALOS DEL BOSÓN DE HIGGS  
EN EL CANAL DE DESINTEGRACIÓN  $WW$  CON EL  
EXPERIMENTO CMS DEL LHC.

Tesis doctoral realizada por:  
**Lourdes María Urda Gómez**

Dirigida por:  
**Dr. María Luisa Cepeda Hermida**  
**Dr. Dermot Anthony Moran**



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

Facultad de Ciencias Físicas  
Madrid, julio 2023





---

# Anomalous couplings of the Higgs boson in the $WW$ decay channel with the CMS experiment at the LHC

---

ACOPLAMIENTOS ANÓMALOS DEL BOSÓN DE HIGGS  
EN EL CANAL DE DESINTEGRACIÓN  $WW$  CON EL  
EXPERIMENTO CMS DEL LHC.

Tesis doctoral realizada por:  
**Lourdes María Urda Gómez**

Dirigida por:  
**Dr. María Luisa Cepeda Hermida**  
**Dr. Dermot Anthony Moran**



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

**CIEMAT**

Facultad de Ciencias Físicas  
**Programa de Doctorado en Física**  
Madrid, julio 2023







U N I V E R S I D A D  
**COMPLUTENSE**  
M A D R I D

**DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS  
PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR**

D./Dña. LOURDES MARIA URDA GOMEZ,  
estudiante en el Programa de Doctorado EN FÍSICA,  
de la Facultad de Ciencias Físicas de la Universidad Complutense de  
Madrid, como autor/a de la tesis presentada para la obtención del título de Doctor y  
titulada:

"ACOPLAMIENTOS ANOMALOS DEL BOSON DE HIGGS EN EL CANAL DE DESINTEGRACION WW CON EL EXPERIMENTO CMS DEL LHC"  
ANOMALOUS COUPLINGS OF THE HIGGS BOSON IN THE WW DECAY CHANNEL WITH THE CMS EXPERIMENT AT THE LHC"

y dirigida por: MARIA LUISA CEPEDA HERMIDA Y DERMOT ANTHONY MORAN

**DECLARO QUE:**

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita.

Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada de conformidad con el ordenamiento jurídico vigente.

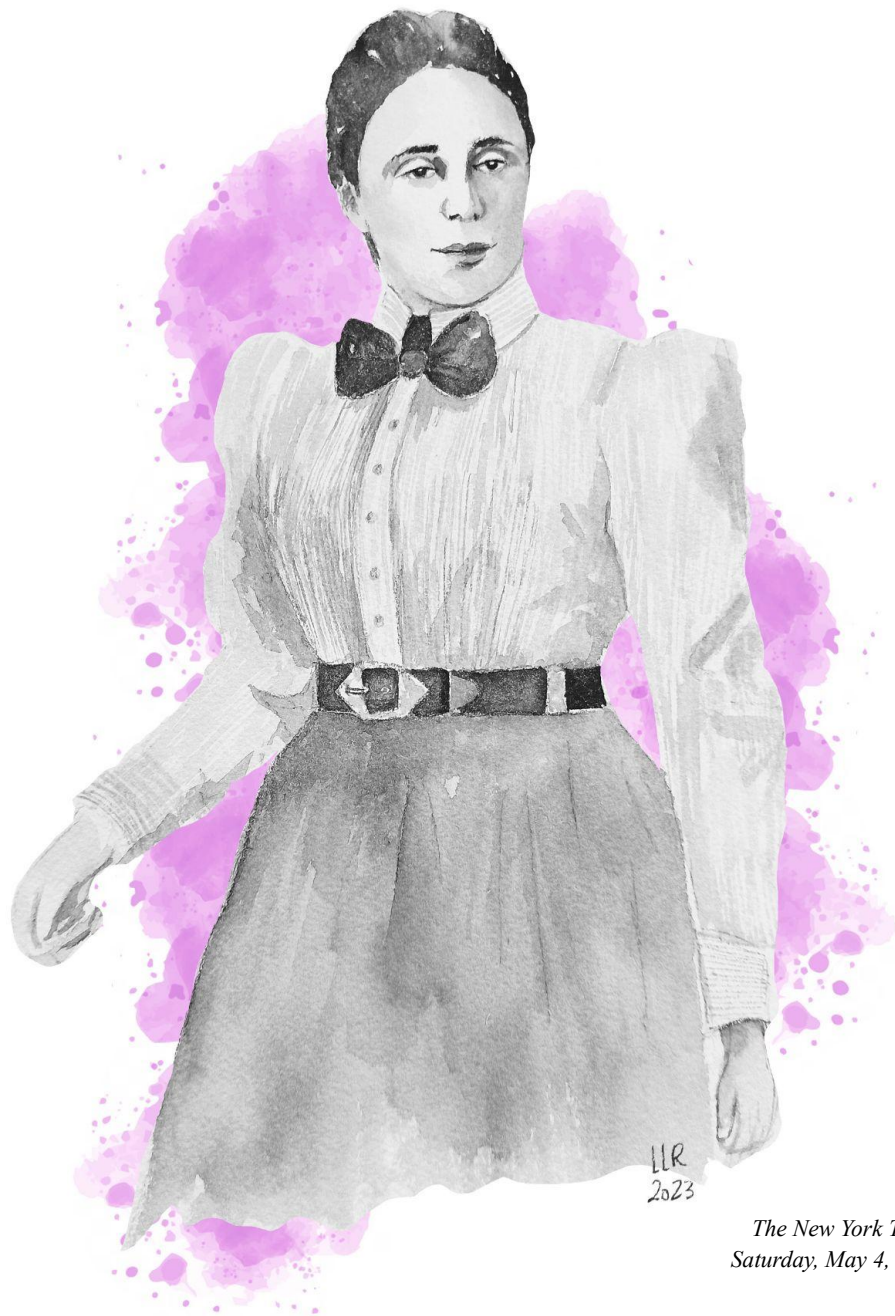
En Madrid, a 14 de junio de 2023

**URDA GOMEZ**  
**LOURDES**  
**MARIA - DNI**  
Fdo.: 25603042V

Firmado digitalmente  
por URDA GOMEZ  
LOURDES MARIA -  
DNI 25603042V  
Fecha: 2023.06.14  
11:57:49 +02'00'

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en  
la primera página de la tesis presentada para la obtención del título de Doctor.





*The New York Times*  
Saturday, May 4, 1935

*...“Fraeulein Noether was the most significant creative mathematical genius  
thus far produced since the higher education of women began.”...*

*Albert Einstein*  
Princeton University, May 1, 1935



*A Catalina Fernández Escalona*



# Contents

|  |           |
|--|-----------|
| <b>Abstract</b>  | <b>5</b>  |
| <b>Introduction</b>  | <b>9</b>  |
| <b>1 Standard model Higgs physics at the LHC</b>           | <b>13</b> |
| 1.1 The standard model of particles . . . . .              | 14        |
| 1.1.1 Particles and symmetries . . . . .                   | 14        |
| 1.1.2 The electroweak interaction . . . . .                | 16        |
| 1.1.3 Spontaneous symmetry breaking mechanism . . . . .    | 18        |
| 1.2 Phenomenology of proton-proton collisions . . . . .    | 22        |
| 1.2.1 Quantum Chromodynamics . . . . .                     | 22        |
| 1.2.2 Parton distribution functions . . . . .              | 23        |
| 1.3 The Higgs boson at the LHC . . . . .                   | 25        |
| 1.3.1 State of the art in Higgs physics . . . . .          | 25        |
| 1.3.2 Future prospects . . . . .                           | 33        |
| 1.4 SM extension using Effective Field Theory . . . . .    | 34        |
| 1.4.1 Building blocks . . . . .                            | 35        |
| 1.4.2 Dimension-6 Operators . . . . .                      | 36        |
| 1.4.3 Basis of Dimension 6 Operators . . . . .             | 38        |
| 1.5 Detector-level Higgs AC analysis . . . . .             | 39        |
| 1.5.1 Overview of Run 2 results . . . . .                  | 42        |
| <b>2 The CMS experiment and the LHC precision frontier</b> | <b>47</b> |
| 2.1 The Large Hadron Collider . . . . .                    | 48        |
| 2.1.1 Accelerator complex . . . . .                        | 48        |
| 2.1.2 Design parameters . . . . .                          | 50        |
| 2.1.3 LHC operational runs . . . . .                       | 52        |
| 2.2 The CMS detector . . . . .                             | 54        |
| 2.2.1 Coordinate system . . . . .                          | 56        |
| 2.2.2 Subdetectors . . . . .                               | 57        |
| 2.2.3 Trigger and DAQ . . . . .                            | 68        |
| 2.3 Objects reconstruction . . . . .                       | 71        |
| 2.3.1 Particle Flow algorithm . . . . .                    | 71        |
| 2.3.2 Missing transverse energy . . . . .                  | 79        |

|          |   |            |
|----------|---|------------|
| <b>3</b> | <b>CMS Muon Drift Tubes upgrade for the HL-LHC</b>        | <b>81</b>  |
| 3.1      | Segment reconstruction with the DT system . . . . .       | 82         |
| 3.1.1    | Hit efficiency definition . . . . .                       | 84         |
| 3.2      | The Slice Test setup and goals . . . . .                  | 84         |
| 3.2.1    | Phase-2 Slice Test On-Board DT electronics . . . . .      | 85         |
| 3.3      | Phase-2 vs Phase-1 readout systems . . . . .              | 86         |
| 3.4      | Phase-2 system time synchronization . . . . .             | 89         |
| 3.4.1    | Test Pulses run . . . . .                                 | 89         |
| 3.4.2    | Data taking in the Slice Test . . . . .                   | 91         |
| 3.4.3    | Calibration results . . . . .                             | 91         |
| <b>4</b> | <b>Anomalous couplings analysis in the HWW channel</b>    | <b>95</b>  |
| 4.1      | The HWW channel . . . . .                                 | 96         |
| 4.1.1    | Measurements of HWW cross-sections . . . . .              | 96         |
| 4.1.2    | Signal signatures . . . . .                               | 97         |
| 4.1.3    | Sources of background . . . . .                           | 102        |
| 4.2      | Data and simulated samples . . . . .                      | 105        |
| 4.2.1    | Data samples . . . . .                                    | 105        |
| 4.2.2    | Simulated samples . . . . .                               | 106        |
| 4.3      | Object identification and corrections . . . . .           | 115        |
| 4.3.1    | Muons and Electrons . . . . .                             | 115        |
| 4.3.2    | Jets . . . . .  | 119        |
| 4.3.3    | Boosted V-jets . . . . .                                  | 120        |
| 4.3.4    | MET . . . . .   | 124        |
| 4.4      | Event selection . . . . .                                 | 125        |
| 4.4.1    | Event categorization . . . . .                            | 125        |
| 4.5      | Observables and kinematic discriminants . . . . .         | 129        |
| 4.5.1    | HVV: Kinematics in VBF and VH productions . . . . .       | 131        |
| 4.5.2    | Hgg: Kinematics in 2-jet ggH production . . . . .         | 137        |
| 4.5.3    | Kinematics of HWW decay products for HVV vertex . . . . . | 138        |
| 4.5.4    | Validation in control regions . . . . .                   | 140        |
| 4.5.5    | Observed distributions in signal region . . . . .         | 144        |
| 4.6      | Multi-dimensional kinematic discriminants . . . . .       | 146        |
| 4.7      | Prefit distributions . . . . .                            | 153        |
| 4.8      | Signal Modelling . . . . .                                | 158        |
| <b>5</b> | <b>Signal extraction</b>                                  | <b>169</b> |
| 5.1      | Maximum Likelihood Fit . . . . .                          | 170        |
| 5.2      | Systematic uncertainties . . . . .                        | 172        |
| 5.3      | Data-driven background normalization . . . . .            | 174        |
| 5.4      | Sources of systematic uncertainties . . . . .             | 174        |
| 5.5      | Impact of the uncertainties on the results . . . . .      | 178        |
| 5.6      | Goodness of fit . . . . .                                 | 178        |
| 5.7      | Post-fit distributions . . . . .                          | 180        |



---

|          |   |            |
|----------|---|------------|
| <b>6</b> | <b>Results</b>  | <b>187</b> |
| 6.1      | AC interpretation . . . . .                             | 188        |
| 6.1.1    | HVV couplings in VBF and VH channels . . . . .          | 189        |
| 6.1.2    | HVV couplings in 0-jet and 1-jet ggH channels . . . . . | 192        |
| 6.1.3    | Full Run 2 combination for HVV couplings . . . . .      | 193        |
| 6.1.4    | Full Run 2 Combination for Hgg couplings . . . . .      | 195        |
| 6.2      | SMEFT interpretation . . . . .                          | 195        |
| 6.2.1    | Full Run 2 Combination for HVV couplings . . . . .      | 196        |
| 6.2.2    | Constraints in terms of EFT couplings . . . . .         | 199        |
| 6.3      | Summary . . . . .                                       | 200        |
| <b>7</b> | <b>Conclusions</b>                                      | <b>201</b> |
| <b>A</b> | <b>Non-prompt background estimation method</b>          | <b>207</b> |
| <b>B</b> | <b>Control region plots</b>                             | <b>211</b> |
| <b>C</b> | <b>Impacts</b>  | <b>221</b> |
|          | <b>Bibliography</b>                                     | <b>231</b> |



# Abstract

---

The thesis herein constitutes a comprehensive investigation of anomalous couplings between the Higgs boson (H) and the bosons W, Z, and g, including the effects of Charge-Parity (CP) violation in the context of the direct searches of anomalous couplings. The study employs various Higgs production mechanisms, such as gluon-gluon fusion, vector boson fusion, and the associated production with a W or Z boson, focusing on the Higgs boson decay to a pair of W bosons in the different flavor dilepton ( $e\mu$ ) final state. The data analyzed was collected by the CMS detector at the LHC from 2016 to 2018, with a center-of-mass energy of 13 TeV and an integrated luminosity of  $138 \text{ fb}^{-1}$ . To increase sensitivity to anomalous effects at the production vertex, the kinematic information from associated jets in the Higgs production was combined using matrix element techniques. The study also utilized Monte Carlo simulation and matrix-element reweighting to model all kinematic features in the production and decay of the Higgs boson, along with a full simulation of detector effects. The analysis resulted in a simultaneous measurement of four Higgs boson couplings to electroweak vector bosons within the framework of a standard model effective field theory. The measured values were found to be consistent with the standard model expectations for the Higgs boson, and constraints were set on the fractional contribution of the anomalous couplings to the Higgs boson cross-section. These results significantly surpass the precision and coverage of the last H to WW anomalous coupling analysis. Additionally, a hardware analysis is included, focusing on the Drift Tubes Phase-2 upgrades of the CMS detector to accommodate higher luminosities expected for future operations. Through precision measurements and advanced analysis techniques, this research contributes to our understanding of the fundamental nature of the universe, seeking to uncover new physics phenomena beyond the standard model.

# Resumen

---

La tesis presentada constituye una investigación exhaustiva de los acoplamientos anómalos entre el bosón de Higgs (H) y los bosones W, Z y g, incluyendo los efectos de violación de la conservación de Carga y Paridad (CP) en el contexto de las búsquedas directas de acoplamientos anómalos. El estudio emplea varios mecanismos de producción del bosón de Higgs, como la fusión de gluones, la fusión de bosones vectoriales y la producción asociada con un bosón W o Z, y se centra en la desintegración del bosón de Higgs en un par de bosones W y estado final  $e\mu$  y neutrinos. Los datos analizados fueron recolectados por el detector CMS en el LHC desde 2016 hasta 2018, con una energía de centro de masa de 13 TeV y una luminosidad integrada de  $138 \text{ fb}^{-1}$ . Para aumentar la sensibilidad a los efectos anómalos en el vértice de producción, se combinó la información cinemática de los jets asociados en la producción del Higgs utilizando técnicas de elementos de matriz. El estudio también utilizó simulación de Monte Carlo y repesado con elementos de matriz para modelar todas las características cinemáticas en la producción y desintegración del bosón de Higgs, junto con una simulación completa de los efectos del detector. El análisis resultó en una medición simultánea de cuatro acoplamientos del bosón de Higgs a bosones vectoriales electrodébiles dentro del marco de una teoría efectiva del campo modelo estándar. Se encontró que los valores medidos son consistentes con las expectativas del modelo estándar para el bosón de Higgs, y se establecieron limitaciones sobre la contribución fraccional de los acoplamientos anómalos a la sección eficaz del bosón de Higgs. Estos resultados superan significativamente la precisión y cobertura del último análisis de acoplamiento anómalo H a WW. Además, se incluye un análisis de hardware, centrándose en las actualizaciones de la electrónica de las cámaras de deriva del detector CMS para acomodar luminosidades más altas esperadas para la fase de alta luminosidad del LHC. A través de mediciones de precisión y técnicas avanzadas de análisis, esta investigación contribuye a nuestra comprensión de la naturaleza fundamental del universo, buscando descubrir nuevos fenómenos de física más allá del modelo estándar.

**From the atom to the standard model: A Journey Spanning Millennia.**

The field of modern particle physics originated in the early 20th century with investigations into the structure of the atom. Democritus first coined the term “atom” around 460-370 BC and proposed that everything is made up of invisible, indestructible, and indivisible atoms. In 1804, Dalton expanded on this concept, suggesting that matter consists of unbreakable atoms with different masses, giving rise to chemical elements. Mendeleev further advanced this understanding by developing the periodic table of elements in 1869. The late 19th century saw significant discoveries in physics, such as X-rays by Wilhelm Röntgen in 1896, radioactivity by Marie and Pierre Curie in 1898, and  $\beta$ -rays by Joseph Thomson in 1897. Thomson identified these rays as negatively charged particles, later named electrons, and discovered that atoms were divisible, marking the discovery of the first elementary particle. The early 20th century witnessed groundbreaking advancements. In 1909, Geiger and Marsden described the scattering of  $\alpha$ -particles by gold foil, proposing the existence of a minor, dense, positively charged nucleus within the atom, supervised by Rutherford. In 1911, Hess demonstrated that atmospheric ionization increased with altitude, suggesting that radiation originated from outer space. In 1913, Bohr formulated his atomic structure theory based on quantum physics, proposing three postulates that explained the behavior of electrons in atoms. In 1919, Rutherford discovered evidence of protons, and in 1931, Chadwick discovered the neutron. In parallel, Einstein identified X-rays as photons based on Planck’s postulates and proposed them as carriers of the electromagnetic force. By 1932, an atomic model with electrons surrounding a nucleus composed of protons and neutrons was proposed, but the force binding these sub-particles together remained unknown. In 1934, Pauli postulated the existence of the neutrino, an uncharged and massless particle that participated in energy exchanges during  $\beta$ -decay. Yukawa also proposed the concept of the strong nuclear force, which binds the components of the nucleus together and is related to a particle 200 times heavier than the electron. This particle was later discovered as the muon, not carrying the strong nuclear force. The particle proposed by Yukawa, known as the pion, was found a decade later. Further studies on cosmic rays led to the discovery of new particles. Efforts to classify all of them and understand their interactions gave rise to the development of the standard model of particle physics, which evolved throughout the latter half of the 20th century. Despite its remarkable achievements, the theory still leaves room for discovering new physics in present times.



# Introduction

Understanding the mechanisms of the universe and especially its creation has always been one of the most genuine goals of humanity. In the wake of the Second World War, CERN emerged as a beacon of international scientific collaboration, uniting nations and advancing our knowledge of the world we inhabit. Established in 1954, CERN has provided invaluable tools, such as the World Wide Web (WWW) in 1989, contributing to the evolution of our rational species. Throughout its existence, CERN has borne witness to remarkable discoveries, with the observation of the Higgs boson in 2012 by the ATLAS and CMS experiments standing as a pinnacle achievement in particle physics, completing the standard model (SM) theory.

The presented doctoral thesis focuses on the Higgs boson, a centerpiece in our comprehension of Nature and a fundamental component coupling to the other particles of the SM to provide masses. In particular, we seek to investigate the potential existence of anomalous couplings of the Higgs boson. The SM values of these couplings are not yet accessible experimentally, but new Beyond the SM (BSM) contributions can make them larger. Anomalous couplings raise intriguing questions: Is it possible for the Higgs boson to exhibit interactions with other particles that differ from our current understanding? Could the behavior of the Higgs field deviate from our existing knowledge? Is there a possibility for the Higgs boson to form couplings with unobserved particles? These questions create an avenue for investigating new BSM physics, particularly through precise studies of the Higgs boson coupling to other SM particles. By harnessing the energy and statistical power of the Large Hadron Collider (LHC) project, this thesis seeks to explore anomalous couplings in the H to WW (HWW) channel, utilizing the CMS detector during the LHC Run 2. The thesis builds upon the precision era of Higgs physics that began after the discovery of the Higgs boson using Run 1 data, confirming its compatibility with the SM expectation. To investigate anomalous

couplings in the HWW channel, we explore events where the WW decay leads to a final state composed of one muon, one electron, and two neutrinos. The analysis considers different Higgs production mechanisms, including gluon fusion, vector boson fusion, and Higgs *strahlung*. Kinematic information from the production and decay processes enables the exploration of anomalous couplings at the Higgs boson production and decay vertices. Notably, the thesis examines the boosted regime in Higgs production in association with vector bosons, leveraging AK8 V-jets with specific substructure reconstruction to access high-energy phase spaces where BSM effects are expected to be amplified.

To study the effects of anomalous couplings, a set of Monte Carlo (MC) simulated samples are utilized and reweighted using the MELA package to account for various signal hypotheses, both CP even and CP odd. The analysis builds upon a generic parametrization of the amplitude scattering formula of the Higgs coupling to vector bosons (HVV) where the potential BSM contributions are regulated by independent coefficients. Constraining such coefficients (referred to as couplings from now on) is the central goal of the presented analysis. The presence of anomalous effects can have an impact on the kinematic features of Higgs boson production and decay processes. Consequently, in order to accurately measure the couplings involved, we will employ kinematic discriminants. These discriminants provide us with valuable information and allow us to distinguish between different scenarios or hypotheses. The anomalous contributions included in the scattering amplitude are equivalent to higher-dimensional operators in the effective field theory (EFT) formulation. For the EFT approach, the SM Lagrangian is extended by a set of interactions representing the effects of new heavy particles. Technically, these new interactions correspond to dimension 6 operators in the extended lagrangian. These new interactions/operators may modify kinematic distributions of the Higgs boson decay products and of the particles produced in association.

The thesis pursues examining the compatibility of anomalous coupling effects in the MC simulations with the LHC Run 2 dataset, collected by the CMS experiment at a center-of-mass energy of  $\sqrt{s} = 13$  TeV between 2016 and 2018, accumulating a total integrated luminosity of  $138 \text{ fb}^{-1}$ . The analysis presented in this thesis will offer two different interpretations: anomalous couplings (AC) and the Standard Model Effective Field Theory (SMEFT). The AC interpretation considers the Higgs boson coupling to W and Z bosons to be equal, leading to four independent anomalous couplings associated with different Higgs models. In contrast,



the SMEFT interpretation assumes the  $SU(2) \times U(1)$  symmetry and aims to measure three independent anomalous couplings. The AC interpretation studies one anomalous coupling at a time, while the SMEFT interpretation studies all couplings simultaneously. Similar AC analyses have been conducted in the HZZ and  $H\tau\tau$  channels. Furthermore, ongoing efforts are underway to develop AC analyses in additional Higgs decay channels, such as  $Hbb$  and  $H\gamma\gamma$ . From a broader perspective, EFT analyses can be categorized into two types: direct searches and interpretations. Direct searches are often referred to as detector-level analyses as they propagate EFT effects to the detector and, thus, represent the method selected to perform the AC HWW analysis presented here. In contrast, interpretation analyses in the Higgs field typically use STXS (Simplified Templates Cross Section) differential measurements.

Additionally, this thesis incorporates a self-contained hardware study. The CMS experiment requires upgrades to accommodate the higher luminosities anticipated for the High-Luminosity LHC (HL-LHC) phase, expected to commence data collection around 2029. The Phase-2 upgrades of the LHC define the parameters for such operations. Specifically, a “Slice Test” was conducted during the LHC Shut Down 2, wherein a sector of the CMS Drift Tube (DT) detector was equipped with the newly developed DT electronics to test it for LHC Run 4 (Phase 2). The accurate calibration of the produced signals is essential to ensure optimal time resolution for future Local triggers. The calibration process involves inter-channel equalization within each DT chamber using dedicated Test-Pulse runs, where pulses are simultaneously injected into all wires, and the resulting Phase 2 signals are recorded. The analysis presented here demonstrates the stability of the new system over time, with average differences between different runs amounting to a maximum of 0.5 ns. This procedure will also serve as an effective monitoring tool.

The manuscript begins by providing an overview of the theoretical framework, specifically the SM and the state of the art of Higgs boson physics at the time of writing this thesis. Subsequently, the experimental setup is described, which includes the LHC and the CMS detector. The first part of the thesis focuses on the mentioned hardware study, introducing the methodologies employed in such studies. The second part of the thesis will explore anomalous couplings of the Higgs boson within the HWW channel. Ultimately, this research aims to contribute to the ongoing advancement of our understanding of the fundamental mechanisms of the universe.



# Chapter 1

## Standard model Higgs physics at the LHC

The opening chapter of this thesis centers on the study of Higgs physics at the LHC. It commences with an overview of the SM structure, highlighting the concept of spontaneous breaking of the electroweak symmetry. The discussion then shifts to the phenomenology of proton-proton (pp) collisions, encompassing the fundamental theory of Quantum Chromodynamics (QCD) and the significance of parton distribution functions (PDFs).

Subsequently, the chapter delves into the exploration of the Higgs boson at the LHC, presenting the current state-of-the-art in Higgs physics as well as discussing future prospects. To extend the reach of the SM, the concept of EFT is introduced along with an explanation of its building blocks, the dimension-6 operators. The Higgs and Warsaw bases of these operators are also comprehensively examined.

Lastly, the chapter addresses the phenomenology of AC involving the Higgs boson, specifically with explicit emphasis on the Run 2 detector-level analyses. In essence, this chapter provides a thorough introduction to the subjects addressed in this thesis, establishing the foundation for further exploration of Higgs physics at the LHC.

## 1.1 The standard model of particles

The standard model of particles and interactions is presently the best-describing theory of the structure of matter and three of the fundamental forces: electromagnetic, strong nuclear, and weak nuclear. Particle accelerators have become our best brute-force approach to testing the SM theory and identifying phenomena that lie beyond its scope. One of the most remarkable and recent discoveries was the observation of a new particle consistent with a SM Higgs boson. This was announced in July 2012 by the ATLAS and CMS Collaborations at CERN LHC [1, 2]. In this context, the following section introduces the theoretical aspects of the SM, with a particular emphasis on the Higgs boson significance as a fundamental component in the generation of particle masses and its potential to unveil new frontiers beyond the SM. In particular, the search for new physics in the Higgs sector presented in this work is based on EFTs that extend the model validity to a wider energy regime.

### 1.1.1 Particles and symmetries

The SM is built upon particles and symmetries, which form its foundational principles. In the model, the elementary particles correspond to excited states of quantum fields, which are mathematically described by Quantum Field Theory (QFT), which applies the principles of quantum mechanics to continuum field systems [3]. The SM itself is a renormalizable QFT, meaning that divergent terms arising in calculations are absorbed into a finite number of parameters. The primary formulation of the SM is in the Lagrangian description, which entirely determines the propagation of fields and their interactions. Specifically, the Lagrangian density is defined as a scalar function denoted by  $\mathcal{L}(\psi_i, \partial_\mu \psi_i, x^\mu)$  depending on the fields, their derivatives, and the spacetime coordinates  $x^\mu = (t, x, y, z)$ . According to the SM theory, it predicts a dynamic field that causes the action (S) to be stationary at its minimum. The action is defined as the integral over spacetime of a Lagrangian density, which, in natural units where  $c = \hbar = 1$ , is expressed as the following quantity:

$$S = \int d^4x \mathcal{L}(\psi_i, \partial_\mu \psi_i, x^\mu). \quad (1.1)$$

Regarding the symmetries, the SM Lagrangian must be invariant under a change of reference frames, such as spacetime translations, rotations, and Lorentz boost

transformations, all embedded in the Poincaré group. The Lagrangian also follows gauge symmetry, making it invariant under local transformations in the internal degrees of freedom of the fields. Gauge symmetry is the essential property of the SM as it naturally gives rise to the dynamics and interactions between the different fields. Additionally, due to Noether's theorem, local gauge symmetries enforce the conservation of charges associated with the interactions. The SM is also a relativistic theory; thus, the Lorentz group represents the fields. Spinors represent fermionic fields, vector fields represent interaction particles, and the Higgs boson is a scalar field. A general classification of SM particles appears in Fig. 1.1.

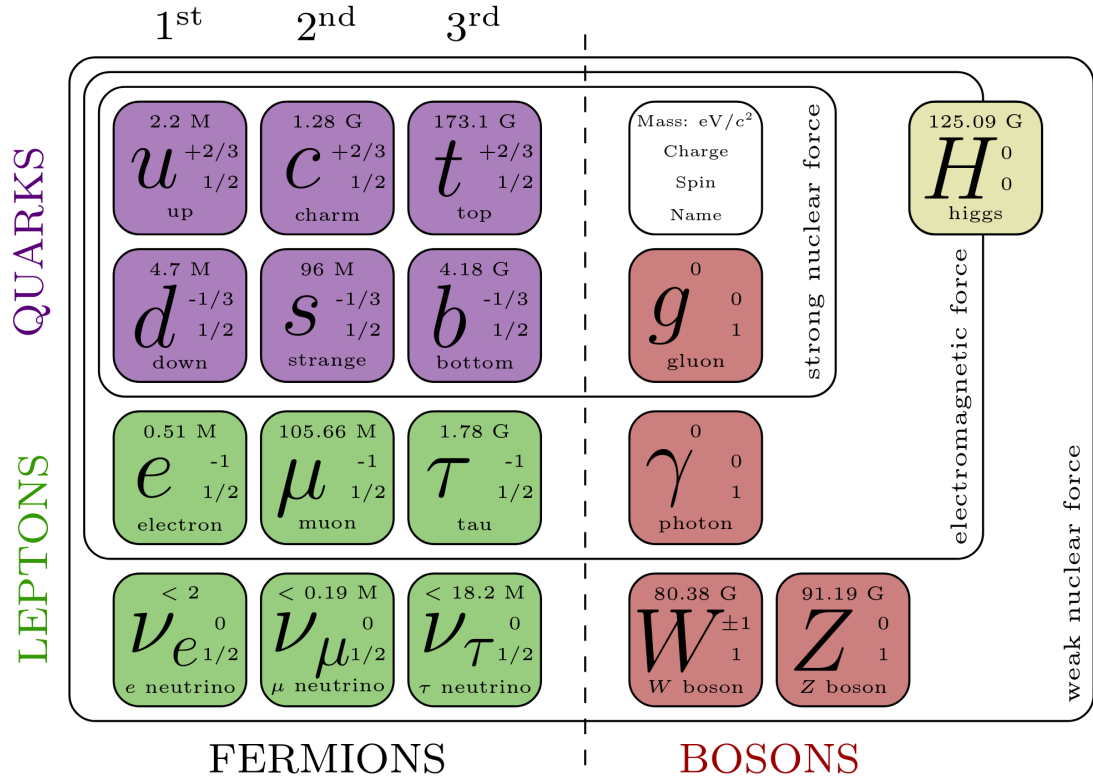


FIGURE 1.1: Particles of the SM depicted in squares with information on their mass, charge, spin, and name provided.

Within the SM, fermions have a spin value of  $1/2$  and are categorized into three generations. Each generation comprises two types of quarks (up/down, top/bottom, charm/strange) and two types of leptons (electron/ $\nu_e$ , muon/ $\nu_\mu$ , tau/ $\nu_\tau$ ). In the realm of quantum mechanics, spin represents the inherent form of angular momentum carried by elementary particles. As aforementioned, the SM describes three fundamental interactions: the strong interaction, mediated by gluons, the

electromagnetic interaction, mediated by photons, and the weak interaction, mediated by the  $Z$  and  $W^\pm$  bosons. All of these intermediary particles have a spin value of 1. Notably, the Higgs boson stands apart with its unique spin value of 0.

In addition to the organization of fermions and the description of interactions within the SM, it is important to mention the mass hierarchy and the masses of neutrinos. Within the three generations of fermions, the mass of particles increases from the first to the third generation. This is particularly evident in the quark sector, where the up and down quarks are significantly lighter than the top and bottom quarks, while the charm and strange quarks lie in between. When it comes to neutrinos, the SM originally assumed them to be massless. However, the Super-Kamiokande Collaboration has reported evidence of neutrino oscillations [4], and they were found to have mass. Neutrino masses are currently not accounted for in the SM and represent an active area of research in particle physics such as T2K and NO $\nu$ A [5, 6] experiments. However, BSM physics could provide an explanation of how neutrinos acquire their masses and why they are significantly lighter compared to the other particles in the SM.

Mathematically, the SM is described as a gauge theory, wherein quantum fields exhibit invariance under  $\mathbf{SU(3)}_C \times \mathbf{SU(2)}_L \times \mathbf{U(1)}_Y$  transformations. These symmetry groups govern the interactions between particles within the SM framework. On one hand, the strong force described by QCD is associated with the  $\mathbf{SU(3)}_C$  gauge group and describes the interactions among quarks and gluons, which are the fundamental constituents of hadrons such as protons and neutrons. On the other hand, the electromagnetic and weak interactions are unified under the Electroweak (EWK) force, which is described by the  $\mathbf{SU(2)}_L \times \mathbf{U(1)}_Y$  gauge group. By incorporating these symmetry groups and their associated gauge theories, the SM successfully captures the fundamental forces and their interactions within the realm of particle physics.

### 1.1.2 The electroweak interaction

The theory that unifies electromagnetic and weak interactions is a gauge theory characterized by the invariance under transformations of the  $\mathbf{SU(2)}_L \times \mathbf{U(1)}_Y$  symmetry group. The  $\mathbf{SU(2)}_L$  describes the weak interaction, while  $\mathbf{U(1)}_Y$  describes electromagnetism. This theory, independently developed by Glashow, Weinberg, and Salam, was recognized with the 1979 Nobel Prize in Physics [7]. The

$SU(2)_L \times U(1)_Y$  symmetry group representation entails that fermions are expressed as left-handed (L) doublets and right-handed (R) singlets. This arrangement results in distinct coupling strengths within the EWK interaction. The projection of matter fields to their respective chiral components is written as follows:

$$\psi_L = \frac{1}{2} (1 - \gamma^5) \psi, \quad \psi_R = \frac{1}{2} (1 + \gamma^5) \psi. \quad (1.2)$$

Both left-handed and right-handed components of fermions are affected by the  $U(1)_Y$  symmetry group. However, they are set apart via their representation within the  $SU(2)_L$  group. As an example, Eq. 1.3 shows such representations for the up and down quarks:

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad u_R, \quad d_R \quad (1.3)$$

A similar structure applies to other quarks and leptons, except for neutrinos which lack right-handed components within the model. As extensions of the model include right-handed components for neutrinos which have a negligible effect on the interactions at reachable energies ( $\sim$ LHC energies). Conversely, antineutrinos possess opposite quantum numbers to neutrinos and solely possess right-handed components in the SM description.

Within the  $SU(2)_L$  group, the conserved charge is known as weak isospin ( $I$ ), conventionally set to  $+\frac{1}{2}$  for up-type fermions and  $-\frac{1}{2}$  for down-type fermions. In turn, the right-handed singlets do not take part in the weak interaction and thus possess a weak isospin value of 0. The generators of the  $SU(2)_L$  group are mathematically represented by Pauli's matrices ( $\sigma_i$ ), which correspond to the three bosons  $W^i$  introduced through gauge invariance. In the context of Lie algebra [8], generators are represented by either differential operators or matrices which serve to represent and perform transformations within the group, generating its elements. These generators satisfy algebraic properties that capture the structure and symmetries of the group, inferring its associated conservation laws and determining its dimension. Together, these components form the essential elements necessary to formulate the  $SU(2)_L$  invariant Lagrangian term within the SM:

$$\mathcal{L}_{SU(2)_L} = -\frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} + i\bar{u}_R \gamma^\mu \partial_\mu u_R + i\bar{d}_R \gamma^\mu \partial_\mu d_R + i\bar{Q}_L \gamma^\mu D_\mu Q_L, \quad (1.4)$$

where the strength tensor  $W_{\mu\nu}^i$  is given by  $W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_W \epsilon^{ijk} W_\mu^j W_\nu^k$ , and

the covariant derivative is defined as  $D_\mu = \partial_\mu + ig_W \frac{\sigma_i}{2} W_\mu^i$ . It is important to emphasize that Eq. 1.4 cannot include any mass term, such as  $m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R)$  (for fermions) or  $m^2 V^\mu V_\mu$  (for bosons). This restriction is because such terms are not invariant under  $SU(2)_L$ . Consequently, to accommodate the masses of gauge bosons and fermions in the mathematical description of nature, an additional mechanism is necessary. In turn, the  $U(1)_Y$  group does admit mass terms, and SM lagrangian is indeed invariant under its symmetry. In the context of EWK interactions, the conserved current is known as the hypercharge  $Y$ . It is related to the isospin  $I$  through the electric charge  $Q$  by the equation  $Y = 2(Q - I)$ . The Lagrangian associated to  $U(1)_Y$  is:

$$\mathcal{L}_{U(1)_Y} = \sum_{\psi=u_R, d_R, L} \bar{\psi}(i\gamma_\mu \partial^\mu - Y g_Y B_\mu)\psi - \frac{1}{4} B^{\mu\nu} B_{\mu\nu}. \quad (1.5)$$

In summary, the underlying symmetry of the SM does not allow for the existence of massive fermions, as their presence would break the invariance under the symmetry group  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . However, experimental observations confirm the existence of massive particles, necessitating a mechanism to confer mass upon the particles within the framework of the SM. This mechanism is known as the Higgs mechanism, which relies on the phenomenon of spontaneous breaking of the EWK symmetry.

### 1.1.3 Spontaneous symmetry breaking mechanism

The Brout-Englert-Higgs (BEH) mechanism [9], originally proposed in 1964, describes the spontaneous symmetry breaking of the electroweak interaction (SSB), enabling the generation of masses for fermions and bosons in the SM. The SSB mechanism separates the electromagnetic and weak forces, and it served as a key inspiration for the unification of these forces in 1967 by Weinberg and Salam, building upon earlier work by Glashow in 1961. In the SM Lagrangian, they assumed the existence of electroweak symmetry in nature, which is spontaneously broken by a  $SU(2)_L$  doublet of complex scalar fields denoted as:

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (1.6)$$



endowing the W and Z bosons with their experimentally observed masses. The given Lagrangian describes the interaction and propagation of the H scalar field:

$$\mathcal{L}_H = (D_\mu H)^\dagger (D^\mu H) + \mu^2 H^2 + \lambda H^4, \quad (1.7)$$

where  $D_\mu = \partial_\mu + igT^i W_\mu^i + i\frac{1}{2}g'B_\mu$ . The Lagrangian includes terms for the Higgs self-coupling denoted by  $\lambda$  and the Higgs potential represented by  $\mu^2$ . Moreover, the model incorporates a mechanism for providing masses to fermions through Yukawa interactions between the fermions and the Higgs field.

$$\mathcal{L}_{Yukawa} = -y_u(\bar{Q}_L H u_R + \bar{u}_R H Q_L) - y_d(\bar{Q}_L H d_R + \bar{d}_R H Q_L). \quad (1.8)$$

In the SM, only the first term of Eq. 1.8 is present for leptonic interactions due to the absence of right-handed neutrinos, as explained earlier. Consequently, neutrinos are explicitly set to have zero mass in this model.

By incorporating the terms from Eq. 1.7 and Eq. 1.8 into the SM Lagrangian, the theory maintains its invariance under  $SU(2)_L \times U(1)_Y$  when  $\mu^2 > 0$ . However, it is essential to enforce the condition  $\mu^2 < 0$ . In this case, the Higgs potential exhibits an infinite number of degenerate minimum energy states that do not preserve  $SU(2)_L$  symmetry. This phenomenon is commonly referred to as the *Mexican hat* potential, depicted in Fig. 1.2.

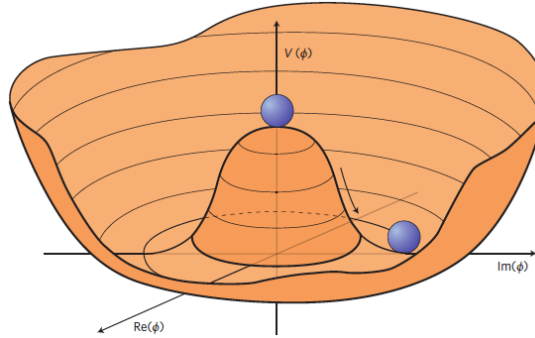


FIGURE 1.2: Illustration of the Higgs potential when  $\mu^2 < 0$ , resulting in the minimum being located at  $|\phi|^2 = -\mu^2/(2\lambda)$ . In this scenario, choosing any point at the bottom of the potential leads to the spontaneous breaking of the symmetry.

This requirement causes the Higgs field, denoted as  $H$ , to acquire a vacuum expectation value (VEV) that breaks the symmetry.

$$H = \begin{pmatrix} 0 \\ \sqrt{\frac{-\mu^2}{\sqrt{2}\lambda}} + h \end{pmatrix} = \begin{pmatrix} 0 \\ v + h \end{pmatrix}. \quad (1.9)$$

The scalar field is the Higgs boson and is denoted as  $h$ , which retains the remaining degree of freedom after symmetry breaking, with a VEV denoted as  $v$ . The VEV, measured to be 246.22 GeV [10], is related to the reduced Fermi constant  $G_F^0$  through the equation  $v^2 = 1/\sqrt{2}G_F^0$ . The final four degrees of freedom introduced by the local gauge symmetry  $SU(2)_L \times U(1)_Y$  give rise to the mass eigenstates associated with the physical bosons  $W^+$ ,  $W^-$ ,  $Z$ , and  $\gamma$ .

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2), \quad \begin{pmatrix} Z \\ \gamma \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W^0 \\ B \end{pmatrix}. \quad (1.10)$$

The weak-mixing angle, denoted as  $\theta_W$ , is determined solely by the coupling constants  $g_W$  and  $g_Y$  from Eq. 1.4 and Eq. 1.5, respectively. Following the process of spontaneous symmetry breaking, the  $W$  and  $Z$  bosons obtain mass values given by:

$$m_W = \frac{g_W v}{2} \text{ and } m_Z = \frac{v \sqrt{g_W^2 + g_Y^2}}{2}. \quad (1.11)$$

Upon implementing the BEH mechanism, the interaction terms of the  $W$  and  $Z$  bosons with the Higgs field become proportional to the square of the masses of the vector bosons, given by:

$$\mathcal{L}_H \xrightarrow{BEH} \frac{2}{v} m_W^2 (h W^{+\mu} W_\mu^-) + \frac{1}{v} m_Z^2 (h Z^\mu Z_\mu). \quad (1.12)$$

In this scenario, the Higgs boson acquires a mass given by:

$$m_H = \sqrt{2}|\mu|. \quad (1.13)$$

The VEV is calculated using the known value of  $m_W$ , but the mass of the Higgs boson, denoted as  $m_H$ , is a free parameter of the model that is only proportional to  $\mu$ , and therefore, its mass cannot be predicted and must be experimentally measured. As of the time of writing this thesis, the CMS experiment reported the most precise measurement of  $m_H$  with a value of  $125.38 \pm 0.14$  GeV [11].

In addition to the BEH mechanism, another essential ingredient is required to

incorporate mass terms for fermions. These mass terms are introduced in a gauge-invariant manner through Yukawa couplings [12], leading to the following expression:

$$m_\psi = \frac{y_\psi v}{\sqrt{2}}. \quad (1.14)$$

Introducing a coupling constant for the interaction between the Higgs boson and a fermion, which is proportional to the mass of the fermion and serves as a free parameter in the model. It is worth mentioning two additional comments in this regard: (1) neutrinos are considered massless in this thesis because their masses have a negligible effect on the performed measurements; and (2) Yukawa interactions are not incorporated as gauge interactions mediated by vector bosons. Furthermore, it should be noted that the discussion so far has only considered one generation of fermions, whereas nature has three generations. The Yukawa interaction for quarks may be expanded as:

$$\mathcal{L}_{Yukawa} = -Y_{ij}^d \bar{Q}_{Li} H d_{Rj} - Y_{ij}^u \bar{Q}_{Li} \epsilon H^* u_{Rj}. \quad (1.15)$$

The Yukawa couplings, denoted as  $Y^{u,d}$ , are described by 3x3 non-diagonal complex matrices. These matrices are related to the 2x2 antisymmetric tensor  $\epsilon$ . As a consequence, the mass eigenstates do not necessarily align with the gauge eigenstates, resulting in off-diagonal elements in the Yukawa matrices [13]. Diagonalizing these matrices leads to the W bosons mediating interactions exclusively between fermions of different flavors. The strength of these interactions between different generations is determined by the unitary nature of the Cabibbo-Kobayashi-Maskawa (CKM) matrix [14].

The measurement of the CKM elements and the theoretical prediction of the BEH mechanism has allowed experimental particle physicists to test the consistency of the SM. Moreover, since the discovery of the W and Z bosons by the UA1 and UA2 experiments [15, 16] at CERN in 1983, the search for the Higgs boson has become a central motivation in particle physics, motivating research at facilities ranging from LEP [17] and Tevatron [18] to contemporary studies at the LHC collider, where experiments such as ATLAS and CMS continue to investigate this particle.

## 1.2 Phenomenology of proton-proton collisions

### 1.2.1 Quantum Chromodynamics

QCD, or Quantum Chromodynamics, is a prominent example of a gauge theory that plays a crucial role in pp collisions. It is invariant under the 3-dimensional special unitary group  $SU(3)_C$ , with generators represented by the Gell-Mann matrices  $T^a$ , which are associated with the eight gluons, explaining the phenomenon of strong interactions. The conserved quantity associated with this symmetry is the “color charge”, denoted as “C”, which is carried only by quarks and gluons, specifically in color triplets and octets, respectively.

The Lagrangian for QCD is:

$$\mathcal{L}_{QCD} = \bar{\psi} (i\gamma^\mu \partial_\mu - g_s \gamma^\mu T_a G_\mu^a - m) \psi - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}, \quad (1.16)$$

where  $\gamma^\mu$  represents Dirac’s matrices, and  $G_{\mu\nu}^a$  is the field strength tensor running for the eight possible charges of a gluon. The last term in the Lagrangian incorporates self-interactions of gluons, accounting for the non-abelian nature of the group generators. The coupling constant  $g_s$  is often redefined as the strong coupling constant  $\alpha_s = \frac{g_s^2}{4\pi}$  [19]. At the one-loop Feynman diagram level, the renormalization of the group predicts a scaling behavior of  $\alpha_s$  as:

$$\frac{d\alpha_s}{dQ^2} = -\frac{\alpha_s^2}{12\pi Q^2} (33 - 2n_f), \quad (1.17)$$

where  $n_f$  is the number of accessible flavors, and  $Q$  is the scale of the interaction. When  $n_f < 17$ , the value of the strong coupling constant,  $\alpha_s$ , decreases as  $Q^2$  increases. This behavior signifies that the strong interaction becomes stronger at lower energy scales. This phenomenon is fundamental to the concept of *color confinement* [20], where quarks and gluons are bound together, forming color-neutral states called hadrons. Hadrons take the form of mesons (consisting of quark-antiquark pairs) or baryons (consisting of three quarks). The diverse range of hadrons observed in Nature arises from the various combinations of quark flavors and their different arrangements. In the early stages of the Universe, quarks and gluons existed in a state known as quark-gluon plasma (QGP), which is a highly energetic and dense medium. In the QGP, the quarks are not confined, and as it cools down, strong force grows, and eventually, the quarks become confined in

hadrons. This process of hadronization also occurs during high-energy particle interactions, such as those taking place in particle accelerators or cosmic ray collisions. According to QCD, the strength of the force between quarks increases as the distance between them grows. Consequently, isolated quarks cannot be directly studied, except for the top quark, which decays prior to hadronization due to its exceptionally high mass of approximately 173 GeV. On the other hand, at higher energies (or short distances), the quarks behave almost like free particles due to the low value of  $\alpha_s$ , leading to the phenomenon of *asymptotic freedom* [21]. This scenario governing pp collisions was discovered in 1973 by David Gross, Frank Wilczek, and David Politzer, who received the Nobel Prize in Physics in 2004.

### 1.2.2 Parton distribution functions

As discussed in the previous section, the Higgs boson exhibits a stronger interaction with more massive particles, leading to a hierarchy of interactions. This hierarchy is evident in both the production and decay of the Higgs boson in the environment of the LHC. In addition to the physics governing the scalar sector of the SM, the production of the Higgs boson in pp collisions is also influenced by the proton internal structure, which becomes resolved at the energy scale achieved by the LHC. In this context, the collision is understood as a scattering process between the constituents of the protons, i.e., quarks and/or gluons (generally referred to as *partons*). Each colliding parton carries a fraction ( $x_1, x_2$ ) of the 4-momentum of the proton, parametrized by Bjorken's variable  $x$  [22].

$$p_{parton} = xp_{proton} \implies \sqrt{s}_{parton} = \sqrt{x_1 x_2 s}. \quad (1.18)$$

The variable  $x$  is defined to be bounded between 0 and 1, which implies that the center-of-mass energy  $\sqrt{s}$  of the collision is lower than the original energy of the protons. The stochastic variables  $x_1$  and  $x_2$ , which represent the fractions of the 4-momentum of the proton carried by the colliding partons, are described by PDFs denoted as  $f_i(x)$ , where  $i$  represents each type of parton. These PDFs naturally satisfy the condition:

$$\sum_i \int_0^1 x f_i(x) dx = 1. \quad (1.19)$$

In the framework of *collinear factorization*, the PDFs determine the probability density of finding a parton with a specific longitudinal momentum fraction  $x$  at a

resolution scale of  $Q^2$ . The variation of parton density with  $Q^2$  has been observed to be in good agreement with experimental results [23].

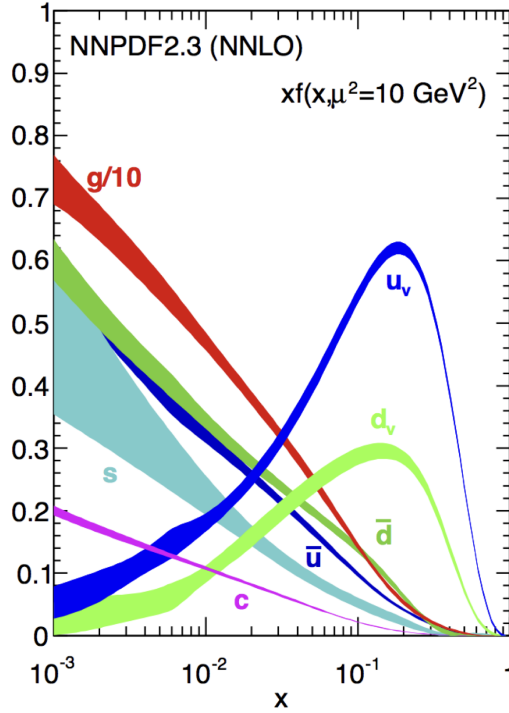


FIGURE 1.3: PDFs evaluated at the factorization scale of  $\mu_F^2 = 10 \text{ GeV}^2$ .

Mathematically, the formalism of factorization in QCD enables the cross-section of a proton-proton collision to be expressed as:

$$\sigma(pp \rightarrow F) = \sum_{a,b} \int dx_1 dx_2 f_a(x_1, \mu_F^2) f_b(x_2, \mu_F^2) \hat{\sigma}_{a,b \rightarrow F}, \quad (1.20)$$

representing the cross-section of a pp collision,  $a$  and  $b$  being parton flavors,  $\mu_F$  denotes the factorization scale, and  $\hat{\sigma}_{a,b \rightarrow F}$  the parton-level cross-section for the process of interest.

Due to low  $Q^2$  effects (when  $\alpha_s$  is large), the proton structure cannot be calculated using perturbative QCD. Instead, the internal proton structure is modeled using PDFs, which are obtained from fits to experimental data. The PDFs describe the longitudinal momentum distribution of the partons inside the proton at a certain scale. It is possible to predict the  $Q^2$  dependence of the proton structure perturbatively using the DGLAP equations. Conversely, at high energy scales, the strong coupling constant  $\alpha_s$  becomes small, allowing us to employ perturbative calculations of the cross-section in QCD. Such calculations involve expanding the

cross-section in terms of a series of higher-order terms, where each term represents the contributions from additional quark-gluon interactions.

Figure 1.3 illustrates an example of the PDFs used in this work, which are determined by combining data from the LHC, HERA [24], and Tevatron [18] colliders. The PDFs exhibit the dominance of u and d quarks at high  $x$  values, while gluons dominate at low  $x$  values. Since its discovery, this has had significant implications for the phenomenology of Higgs boson production at the LHC and allows for detailed studies of this particle.

## 1.3 The Higgs boson at the LHC

The search for the Higgs boson was challenging for previous colliders like LEP [17] at CERN and Tevatron [18] at Fermilab. For LEP, the achieved center of mass energy 209 GeV was not enough to produce the Higgs boson. Furthermore,  $e^+e^-$  colliders do not have the highest cross-section for Higgs production. Merely producing the Higgs boson is insufficient; it is necessary to first generate heavy particles that subsequently couple to the Higgs in order to observe it. On the other hand, the CDF [25] and D0 [26] Collaborations presented evidence of the Higgs boson using the Tevatron in a combined measurement [27].

The LHC design provided enough integrated luminosity to observe the Higgs boson for the first time in 2012. Over the last decade, the HEP community has obtained the most comprehensive understanding of the Higgs boson to date, as summarized in this section. Luminosity, which represents the amount of data collected and analyzed by the LHC, will be introduced later. This chapter presents a summary of our current knowledge about the Higgs boson based on LHC data until the end of Run 2 in 2018.

### 1.3.1 State of the art in Higgs physics

Over the past decade, studies by the CMS and ATLAS experiments have led to a significant advancement in our understanding of the scalar sector of the SM. During the LHC Run 2 period (2015-2018), approximately 8 million Higgs bosons were produced in pp collisions at  $\sqrt{s} = 13$  TeV at both CMS and ATLAS. The LHC

experiments have meticulously studied various production modes and decay channels, determining their cross-sections and branching fractions with high precision. This section is inspired by the paper from the CMS Collaboration published in Nature in July 2022 [28]. Figures and values presented in this section are sourced from this reference unless stated otherwise.

## Production modes

The three primary mechanisms for producing a Higgs boson in a pp collision are (a) gluon fusion (ggH), (b) vector boson fusion (VBF), and (c) associated production with a vector boson (VH), commonly known as Higgstrahlung, where “V” corresponds to either Z or W bosons. The leading order (LO) Feynman diagrams illustrating these mechanisms are shown in Fig. 1.4. While these are the main production modes considered in the presented analysis, other ways to produce a Higgs boson include those (d) in association with a top or bottom quark pair ( $t\bar{t}H$  or  $b\bar{b}H$ ), or (e, f) in association with a single top quark ( $tH$ ). The sector graph in Fig. 1.6 (bottom-left) illustrates the relative contributions of

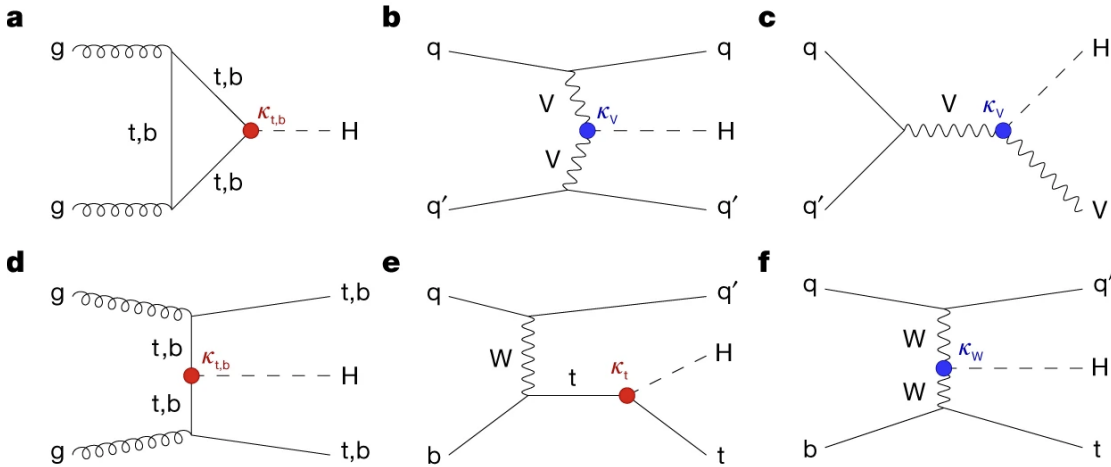


FIGURE 1.4: Feynman diagrams for various Higgs boson production mechanisms including (a) gluon fusion (ggH), (b) vector boson fusion (VBF), (c) associated production with a vector boson (VH), (d) production in association with a top or bottom quark pair ( $t\bar{t}H$  or  $b\bar{b}H$ ), and (e,f) production in association with a single top quark ( $tH$ ).

each production mechanism, or production channel, to the total cross-section. For instance, the ggH production dominates with an occurrence rate of 87%, followed by VBF with 7%, VH with 4%,  $t\bar{t}H$  with 1%, and so on. The table in the upper-left part of Fig. 1.6 presents the production cross-section ( $\sigma$ ) values in pb for a



Higgs boson with the most precise measured mass of 125.38 GeV at CMS [28]. Higgs events are commonly classified based on how they were produced, as those channels may be associated with different topologies.

### Decay modes

After its production, the Higgs boson rapidly decays into lighter particles following the SM lifetime of  $\tau_H \approx 1.6 \times 10^{-22}$  s, which is associated with a natural width of  $\Gamma = \hbar/\tau_H = 4.14 \pm 0.02$  MeV. The dominant decay modes of the Higgs boson are (g) heavy vector boson pairs, (h) fermion-antifermion pairs, and (i,j) photon pairs or  $Z\gamma$ , as depicted by the Feynman diagrams in Fig. 1.5. Each sector depicted

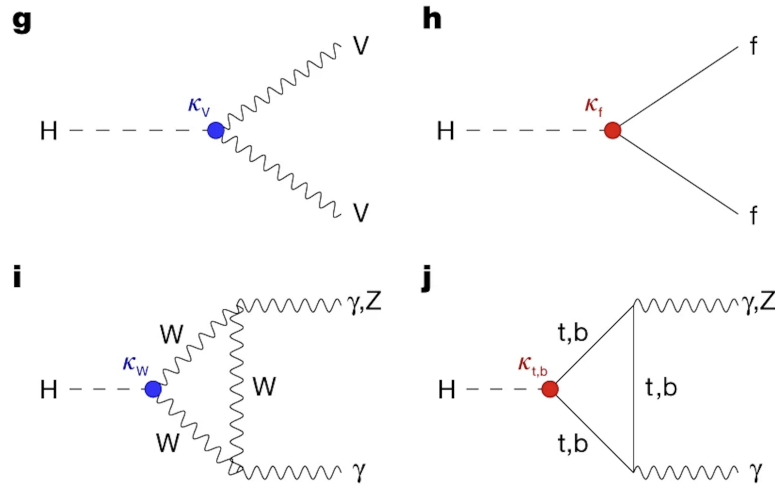


FIGURE 1.5: Feynman diagrams illustrating several Higgs boson decay modes, including (g) decay into heavy vector boson pairs, (h) decay into fermion-antifermion pairs, and (i,j) decay into photon pairs or  $Z\gamma$ .

in Fig. 1.6 (bottom-right) represents the relative branching ratio (BR) associated with the decay channels of the Higgs boson. The  $bb$  channel is the most probable final state, accounting for approximately 60% of Higgs boson decays. The experimental values for these branching fractions are compiled in the upper-right table of the same figure. The CMS Collaboration mainly categorizes Higgs physics analyses based on the decay channels, considering the potential overlap of events among different channels.

While Fig. 1.6 shows the theoretical values for the indicated Higgs mass, the experimental ones are derived for each analysis where a dedicated event selection is applied to target certain production modes and decay channels. A comprehensive

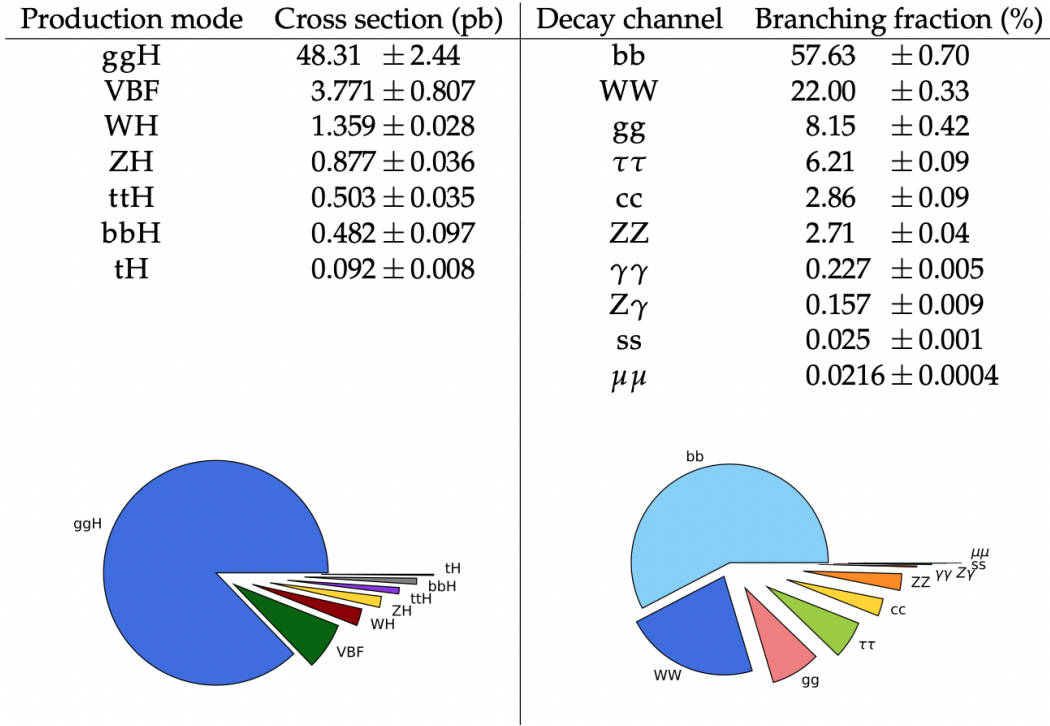


FIGURE 1.6: Top-left: Theoretical cross-section values in pb per Higgs production mode and their corresponding relative uncertainties are presented [29, 30]. Top-right: Theoretical branching ratio values associated with each Higgs boson decay channel and their relative uncertainties are displayed. Bottom-left: The sector graph illustrates the relative occurrence of different Higgs production modes. Bottom-right: The sector graph depicts the relative occurrence of different decay channels of the Higgs boson. Note that the theoretical references are found in Ref [28].

overview of the event selection features for various channels can be found in [28]. This thesis focuses on the Higgs boson decay channel to WW (HWW), where the final state consists of two different flavor leptons and two neutrinos. Despite benefiting from one of the largest branching fractions, the presence of neutrinos in the final state degrades the resolution of the obtained measurements. The approach to tackle this channel is described in section 4.1.

### The $\mu$ - and $\kappa$ -frameworks

The SM prediction for the Higgs boson couplings to bosons and fermions, given by the BEH mechanism, is expected to be proportional to  $m_V^2$  and  $m_\psi$ , respectively. Notable deviations from the expected production rates and decay branching fractions of the Higgs boson may suggest the presence of novel physics phenomena. To

assess the validity of the SM in this context, LHC experiments commonly employ two approaches.

To begin with, the  $\mu$  framework serves to quantify the level of agreement between observed data and the expectations derived from the SM. This is accomplished by fitting the data using a model that is parameterized by the signal strength, denoted as  $\mu$ . In this context, the signal strength represents the observed number of Higgs boson events in comparison to the rate predicted by the SM. This parameter establishes the relationship between the observed and SM-expected yields without altering the distribution of events. By combining the results from Run 2 data, a value of  $\mu = 1.002 \pm 0.057$  [28] is obtained, representing an enhancement in the precision of the measurement by a factor of 4.5 when compared to the value of  $\mu = 0.83 \pm 0.23$  [31] at the time of the discovery of the Higgs boson [1, 2]. Figure 1.7 shows the signal strength parameters for each specific production and decay channel mentioned earlier. Furthermore, the p-value derived from the comprehensive analysis of the Run 2 dataset signifies a 5.8% agreement with the predictions of the SM.

The  $\kappa$ -framework quantifies the potential impact of BSM physics on the physics of the Higgs production and decay rates. We may break down the cross-section by BR as follows [32]:

$$(\sigma \cdot \text{BR})(i \rightarrow H \rightarrow f) = \frac{\sigma_i \cdot \Gamma_f}{\Gamma_H}. \quad (1.21)$$

The production cross-section through the initial state  $i$  is represented by  $\sigma_i$ , while the partial decay width into the final state  $f$  is denoted by  $\Gamma_f$ , and the total width of the Higgs boson is expressed as  $\Gamma_H$ . To account for the coupling strength modifier at leading order, the  $\kappa$  parameters are introduced by expressing each component of Eq. 1.21 as the SM expectation multiplied by the square of the corresponding process as a coupling strength modifier.

$$(\sigma \cdot \text{BR})(i \rightarrow H \rightarrow j) = \frac{\sigma_i^{SM} \kappa_i^2 \cdot \Gamma_f^{SM} \kappa_f^2}{\Gamma_H^{SM} \kappa_H^2} \rightarrow \mu_i^f \equiv \frac{\sigma \cdot \text{BR}}{\sigma_{SM} \cdot \text{BR}_{SM}} = \frac{\kappa_i^2 \cdot \kappa_f^2}{\kappa_H^2}, \quad (1.22)$$

where the relative rate to the SM expectation is represented by  $\mu_i^f$ , while the

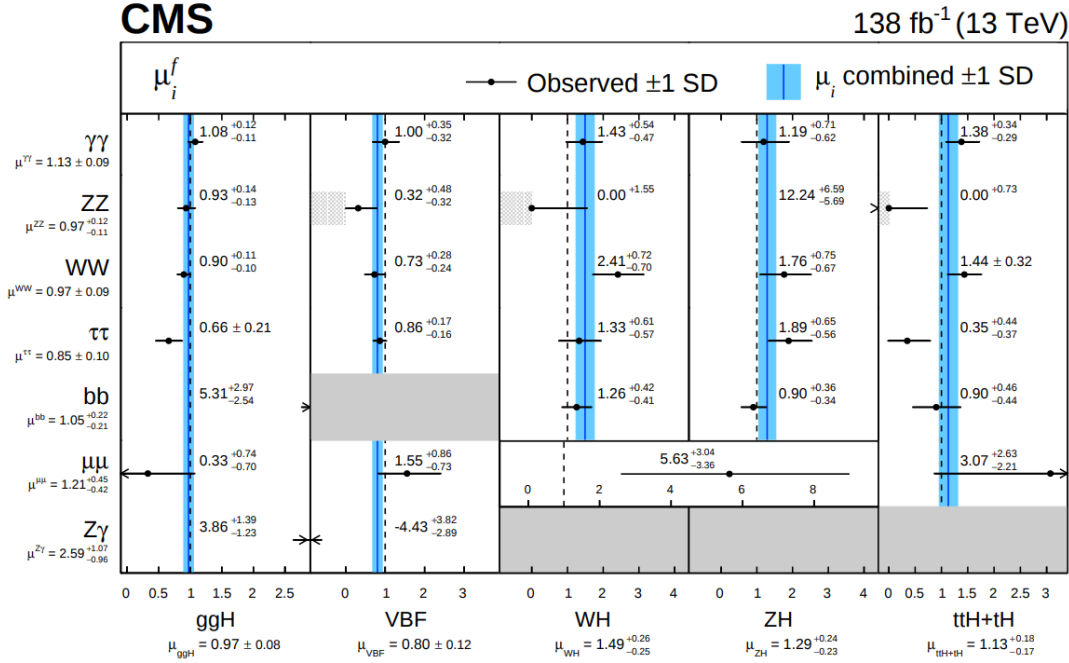


FIGURE 1.7: Signal strength parameters for each individual production mode and decay channel are denoted as  $\mu_i^f$ , while the combined signal strength per production mode and decay channel is represented by  $\mu_i$  and  $\mu^f$ , respectively. The dashed vertical lines at 1 indicate the expectation value of the SM, with light grey shading indicating that  $\mu$  is constrained to be positive and dark grey shading indicating an absence of measurements. The p-value, indicating the agreement with the SM prediction, is 5.8%

expression  $\kappa_H^2$  adjusts the SM Higgs boson mass width to account for the modifications of the SM Higgs boson coupling strengths, denoted by  $\kappa_i$ .

$$\kappa_H^2 \equiv \sum_j \frac{\kappa_j^2 \Gamma_j^{SM}}{\Gamma_H^{SM}}. \quad (1.23)$$

To account for the coupling of the Higgs boson to massive gauge bosons and fermions, we perform fits to data using signals models, which include the coupling modifiers  $\kappa_V$  and  $\kappa_f$ . Both scaling parameters,  $\kappa_V$ , and  $\kappa_F$ , have been found to be consistent with those expected for the SM Higgs boson, as illustrated in the left panel of Fig. 1.8. This plot also shows the evolution of the measurements of  $\kappa_V$  and  $\kappa_F$  with the accumulation of more data during Run 1 and Run 2, resulting in the most accurate determination to date, with an uncertainty of about 10%, and in agreement with the SM expectation (as those are not part of the SM). The right panel of the figure illustrates the hierarchy of Higgs boson coupling to the SM particles as a function of their masses.

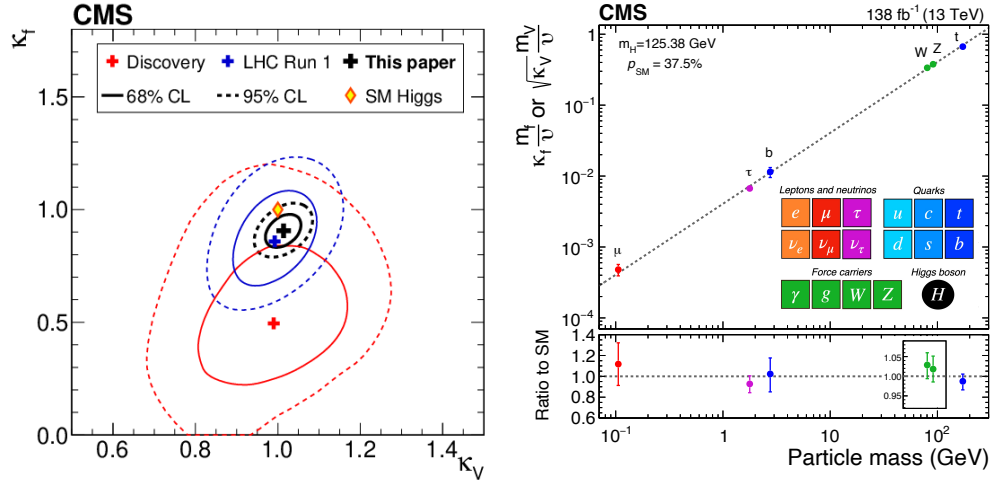


FIGURE 1.8: (Left) Constraints on the coupling modifiers,  $k_f$  for fermions and  $k_V$  for heavy gauge bosons, of the Higgs boson. The constraints are derived from different datasets: discovery, Run 1, and Run 2 [28]. The SM prediction (yellow diamond marker) corresponds to  $k_V = k_f = 1$ . (Right) Hierarchy of Higgs boson coupling to SM particles as a function of their masses.

Instead of only fitting two parameters, a more comprehensive analysis can be performed, taking into account not only heavy gauge bosons and fermions but also loop-induced processes such as Higgs boson couplings to gluons, photons, and  $Z\gamma$ . The outcome of this analysis is summarized in Fig. 1.9, surpassing the precision of the discovery data set by a factor of approximately 5.

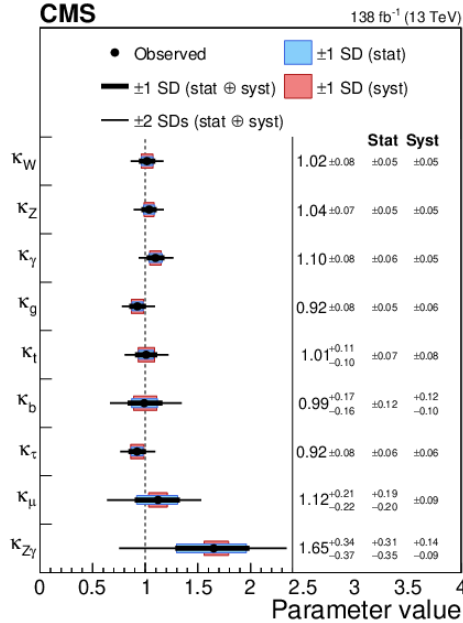


FIGURE 1.9: Coupling modifiers obtained from the fit show a p-value of 28% with respect to the SM prediction.

It is important to identify the limitations of the  $\kappa$ -framework, which relies on inclusive production and decay rates, which means it is unable to assess discrepancies in differential distributions and overlooks potential shape effects. Additionally, the  $\kappa$ -framework assumes contributions solely from SM particles without considering the possibility of BSM physics. When deviations from the expected behavior are observed, the  $\kappa$ -framework lacks guidance in terms of BSM physics. Taking all these factors into consideration, these limitations serve as motivation for exploring EFT approaches.

As an additional note in this regard, the studies of di-Higgs events, which involve the production and decay of two Higgs bosons, provide insights into the self-interaction parameter  $\lambda$  and reveal details about the Higgs potential. Although this topic is beyond the scope of this thesis, several CMS analyses investigate this topology, such as the  $HH \rightarrow b\bar{b}\tau\tau$  channel [28, 33]. Di-Higgs analyses at the LHC are statistically limited, but the future data sample from the HL-LHC holds promising prospects for these studies. The HL-LHC prospects for constraining the Higgs boson potential shape, combining efforts from ATLAS and CMS in various di-Higgs channels under investigation, are illustrated in Fig. 1.10. Finally, we may

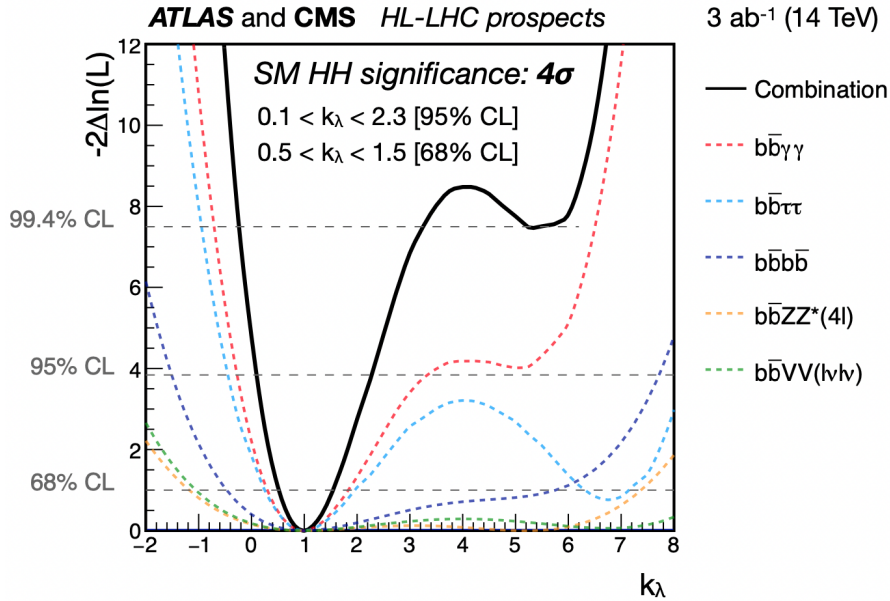


FIGURE 1.10: Combined ATLAS and CMS expected likelihood of the  $\kappa_\lambda$  parameter (Higgs self-interaction) for di-Higgs production searches at the HL-LHC, with the channels used in the combination indicated. Adopted from [34].

also consider invisible decays, where the Higgs boson may decay into long-lived neutral particles or dark matter particles that do not leave detectable traces in the CMS detector. This thesis does not cover but these studies, and references to

dedicated searches can be found in [35, 36]. The current experimental constraints on the branching fraction for invisible Higgs decays are  $\mathcal{B}_{Inv} < 0.16$  at 95% CL.

### 1.3.2 Future prospects

Since its discovery in 2012, the ATLAS and CMS experiments have extensively studied the properties of the Higgs boson. So far, all measurements have been consistent with the SM expectations. However, the associated uncertainties in these measurements leave room for discovering new physics beyond the SM. In the upcoming phases of the LHC, such as Run 3 and HL-LHC, we aim to reduce uncertainties further and enhance our comprehension of the Higgs boson and the SM. This will allow us to rigorously test the validity of the SM as a low-energy approximation of a more comprehensive theory and actively search for indications of new physics. As depicted in Fig. 1.7, the analysis of the full Run 2 dataset, which corresponds to an integrated luminosity of  $138 \text{ fb}^{-1}$ , reveals a favorable agreement between the observed data and the SM predictions, with a deviation within the range of  $1\sigma$  when considering the combined signal strength parameter,  $\mu$ . As additional collision data is accumulated, it is expected that the corresponding error bars will significantly diminish, leading to a notable improvement in the precision of the measurements. Regarding this, Fig. 1.11 provides an illustration of the

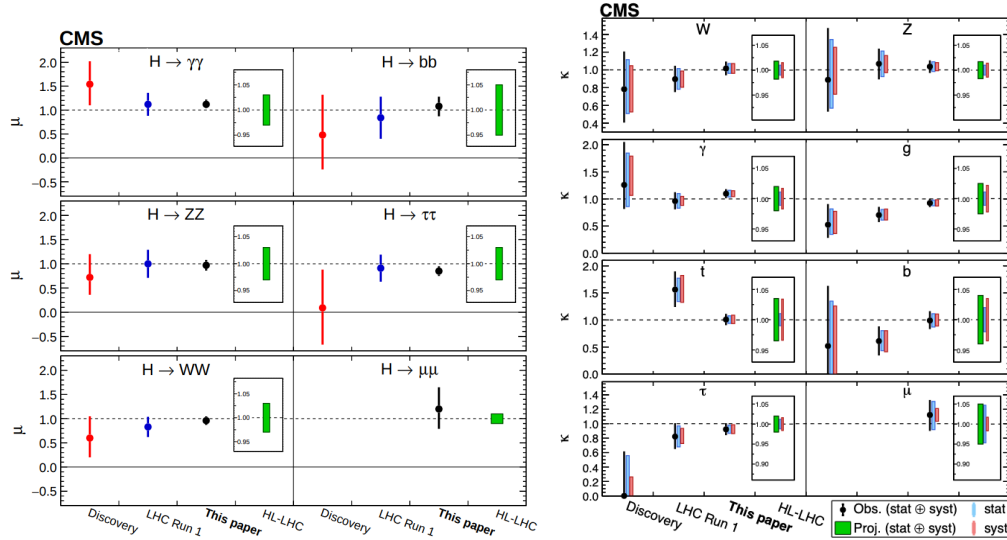


FIGURE 1.11: Observed and projected values obtained from fits in the  $\kappa$ -framework and the  $\mu$ -framework using various data sets and the HL-LHC projection. Note that measurements for  $H \rightarrow \mu\mu$  were not available in earlier data sets due to limited sensitivity.

reduction in uncertainty across the three datasets under consideration (discovery,

Run 1, Run 2). Additionally, the plots present a projection for the measurements of the signal strength parameter,  $\mu$  (left panel), and the  $\kappa$  parameter (right panel), extending to the HL-LHC scenario. The graph shows how the uncertainty progressively decreases as more data is collected, highlighting the potential for enhanced precision in these measurements at the HL-LHC.

With increasing accuracy in our measurements, the constraints on new physics (NP) become more stringent. Unresolved issues with the SM include the naturalness problem, the matter-antimatter asymmetry of the universe, the possible metastability of the vacuum state, the nature of dark matter, cosmic inflation, the origin of neutrino masses, CP violation, unification problems, and various related topics. Emphasizing its sensitivity to NP, the Higgs boson is widely regarded as a key component of current particle physics research [37]. Consequently, conducting in-depth studies of the Higgs boson and gaining a comprehensive understanding of its properties is of great importance. Such efforts are necessary to advance our understanding of fundamental physics and shed light on the nature of the universe at the smallest scales.

The analysis presented in this thesis focuses on finding NP by searching for anomalous couplings of the Higgs boson in the HWW channel, focusing on the fully leptonic different flavor final state. This search is conducted within the EFT framework. Although the SM has been regarded as a complete theory thus far, it is widely acknowledged that it is not the ultimate theory and is merely an effective description of a fraction of fundamental physics. The fact that we found no new particles, together with some concepts proposing the existence of new particles (which could couple to the Higgs boson) that have not been observed at the TeV scale, indicate a large energy scale difference between the SM and NP. This energy scale hierarchy provides a rationale for the application of EFTs.

## 1.4 SM extension using Effective Field Theory

EFT is a widely employed method in physics that simplifies complex physical phenomena by describing them within specific energy or length scales. In the field of high-energy physics, EFTs are employed to account for the effects of a phenomenon at a specific energy scale, denoted as  $\Lambda$ , without delving into its underlying causes, being the remarkable benefit of EFT its capacity to make model-independent statements. EFT is conceptualized as a tool to expand the SM by acknowledging that



the behavior of physics differs across varying energy scales, such that phenomena at one scale may have negligible impact on those at another scale. The energy hierarchy may be visualized as a ladder of energy scales, as illustrated in Fig. 1.12, with the lowest energy at the bottom rung and the highest energy at the top. Researchers can build EFTs from the bottom or top. The presented analysis discusses the essential components for constructing an EFT within the framework of the SM, also known as SMEFT. In this chapter which draws upon Ref. [38], we adopt the perspective of the SM as a low-energy EFT arising from an unknown underlying theory at the energy scale  $\Lambda$ . We employ the general principles of EFT to investigate the physics of the Higgs at CMS.

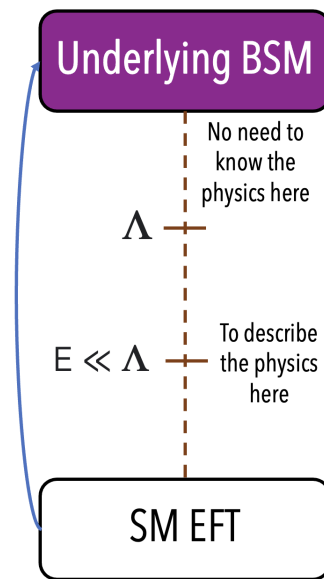


FIGURE 1.12: The diagram illustrates the energy relationship between the SM EFT and the underlying BSM theory. The energy cut-off  $\Lambda$  denotes the upper limit of validity for the EFT theory, while  $E$  represents the energy scale under consideration. The blue arrow depicts the process of matching the SM EFT to the underlying BSM theory, which are connected through coefficients.

### 1.4.1 Building blocks

All physical theories may be expressed as a sequence of EFTs. Even without knowing about the underlying model, it is possible to construct an EFT based on a few general assumptions. Taking Fermi's theory as an illustration of a low-energy EFT of the SM that serves as a representation of the complete underlying theory. In Fig. 1.13, the Feynman diagram on the left side depicts the process in the full theory, whereas the second diagram (right side) represents the effective interaction

within the low-energy EFT, in which the W boson propagator  $\frac{i}{p^2 - M^2}$  simplifies to a contact interaction  $\frac{ig^2}{M^2}$ , accurately capturing the physics when the momentum squared,  $p^2$ , is much smaller than the squared mass of the W boson,  $M^2$ . EFTs are

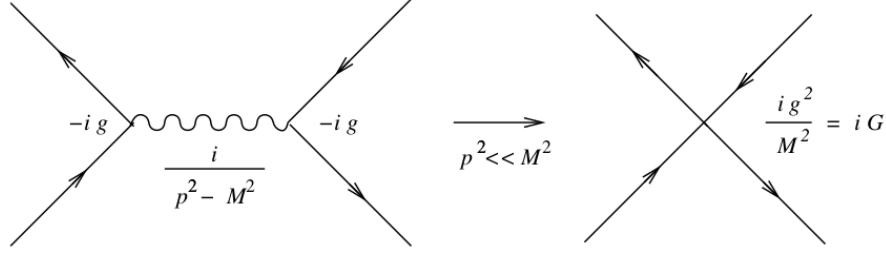


FIGURE 1.13: Effective theory of W boson interaction when the energies are much lower than the mass of the scalar propagator [39].

constructed by incorporating higher-dimensional operators to describe the effective interactions, as demonstrated in the aforementioned example. Constructing EFT operators ( $\mathcal{O}_i$ ) in a model-independent manner requires three essential ingredients: particles, symmetries, and a set of rules for counting dimensions. The construction of an effective Lagrangian that describes the underlying theory relies on three crucial components:

- **Particles:** A set of fields and derivatives that build the Operators  $\mathcal{O}_i$  in the effective field theory, encompassing all particles with masses much smaller than  $\Lambda$ .
- **Symmetries:** Operators  $\mathcal{O}_i$  must satisfy the symmetries of the SM, including Gauge and Lorentz symmetries.
- **Counting scheme (dimension counting):** This is achieved by splitting the coupling in front of each operator  $\mathcal{O}_i$  into a constant  $c_i$  and powers of the energy scale  $\Lambda$ , while organizing the operators based on their mass dimension.

### 1.4.2 Dimension-6 Operators

In this study, we will expand the SM Lagrangian into an effective Lagrangian with the following form:

$$\boxed{\mathcal{L}_{eff} = \mathcal{L}^{SM} + \mathcal{L}^{D=6}, \quad \mathcal{L}^{D=6} = \frac{1}{\Lambda^2} \sum_i c_i^{(6)} \mathcal{O}_i^{(6)}}. \quad (1.24)$$

In the above expression, operators with dimensions 5 and 7 are omitted due to their violation of either the lepton or baryon number conservations. Operators with dimension 8 are typically neglected as a factor of  $1/\Lambda^4$  suppresses them. Equation 1.24 offers further insights. Firstly, the effects of new physics are encapsulated in the  $\mathcal{O}_i^{(6)}$  operators, with Wilson coefficients  $c_i$  indicating the strength of the new interactions. Secondly, the contributions from  $\mathcal{O}_i^{(6)}$  operators scale as  $(E/\Lambda)^2$ , resulting in an increase with the energy. Lastly, it is important to note that the validity range of the EFT must always satisfy the condition  $E \ll \Lambda$ , where  $E$  represents the energy scale under consideration.

In summary, to clarify the terminology used in this thesis, we have established that EFT operators are able to capture the effects of new heavy particles that may not be observable at the energies reachable by the LHC as illustrated in Fig. 1.14. The SM framework should describe what we observe, but there may be "new" particles whose effect is not described by the SM, if the particle is very heavy with respect to LHC energies, its effect can be approximated with an EFT approach. As such, EFTs can be a powerful tool for detecting anomalies. Given that this work aims to expand the SM lagrangian using EFTs to describe the Higgs boson couplings to bosons and fermions, the main objective is to search for anomalous couplings. In the end, because everything looks in agreement with the SM prediction, we place constraints on the anomalous couplings, specifically by using the WW decay channel.

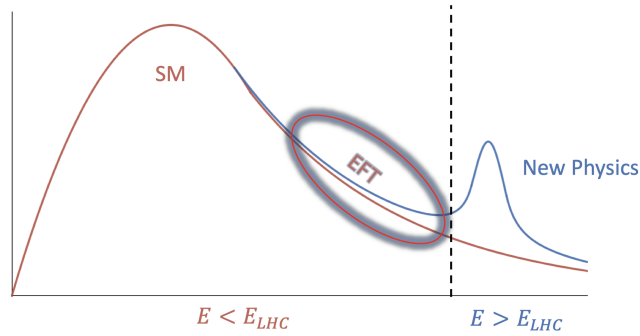


FIGURE 1.14: EFT captures the effects of new heavy particles that may not be observable at the energies reachable by the LHC. The dashed vertical line indicates the energy cut-off  $\Lambda$ .

### 1.4.3 Basis of Dimension 6 Operators

Once the effective Lagrangian is obtained, the next step is defining the operators that represent the effective interactions. As per Eq. 1.24,  $\mathcal{O}_i$  forms a complete basis of D=6 operators constructed from SM fields and derivatives. However, there are multiple ways to select a complete and non-redundant basis of operators, as discussed in Ref. [40].

One widespread basis is the *Warsaw basis* shown in Table 1.1, which originally consisted of the first complete set of (59) operators and is commonly used in the theoretical community because it corresponds to the gauge eigenstates before the SSB. Another basis is the *Higgs basis* (Eq.1.31-1.34), which is experimentally convenient as it allows for expressing the EFT in terms of mass eigenstates after the SSB. The Higgs basis establishes a more straightforward connection to measurable quantities in experiments, making it particularly suitable for studies of the Higgs boson couplings at the LHC. Another example worth mentioning is *the SILH ba-*

| Bosonic CP-even |   | Bosonic CP-odd       |   |
|-----------------|---|----------------------|---|
| $O_H$           | $[\partial_\mu(H^\dagger H)]^2$                                 |                      |   |
| $O_T$           | $(H^\dagger \overleftrightarrow{D}_\mu H)^2$                    |                      |   |
| $O_{6H}$        | $(H^\dagger H)^3$   |                      |   |
| $O_{GG}$        | $g_s^2 H^\dagger H G_{\mu\nu}^a G_{\mu\nu}^a$                   | $O_{\widetilde{GG}}$ | $g_s^2 H^\dagger H \widetilde{G}_{\mu\nu}^a G_{\mu\nu}^a$                   |
| $O_{WW}$        | $g_L^2 H^\dagger H W_{\mu\nu}^i W_{\mu\nu}^i$                   | $O_{\widetilde{WW}}$ | $g_L^2 H^\dagger H \widetilde{W}_{\mu\nu}^i W_{\mu\nu}^i$                   |
| $O_{BB}$        | $g_Y^2 H^\dagger H B_{\mu\nu} B_{\mu\nu}$                       | $O_{\widetilde{BB}}$ | $g_Y^2 H^\dagger H \widetilde{B}_{\mu\nu} B_{\mu\nu}$                       |
| $O_{WB}$        | $g_L g_Y H^\dagger \sigma^i H W_{\mu\nu}^i B_{\mu\nu}$          | $O_{\widetilde{WB}}$ | $g_L g_Y H^\dagger \sigma^i H \widetilde{W}_{\mu\nu}^i B_{\mu\nu}$          |
| $O_{3W}$        | $g_L^3 \epsilon^{ijk} W_{\mu\nu}^i W_{\nu\rho}^j W_{\rho\mu}^k$ | $O_{\widetilde{3W}}$ | $g_L^3 \epsilon^{ijk} \widetilde{W}_{\mu\nu}^i W_{\nu\rho}^j W_{\rho\mu}^k$ |
| $O_{3G}$        | $g_s^3 f^{abc} G_{\mu\nu}^a G_{\nu\rho}^b G_{\rho\mu}^c$        | $O_{\widetilde{3G}}$ | $g_s^3 f^{abc} \widetilde{G}_{\mu\nu}^a G_{\nu\rho}^b G_{\rho\mu}^c$        |

TABLE 1.1: The Warsaw basis includes 59 different kinds of bosonic D=6 operators [41]. When accounting for flavor structure and CP conjugates, the total number of distinct operators increases to 2499 [42].

*sis* [43], not used in the current analysis, which is particularly useful for describing strongly interacting Light Higgs EFTs as it establishes a more direct connection to certain BSM scenarios such as the Composite Higgs [44]. Both the Higgs basis and the SILH basis are related to the Warsaw basis through linear transformations.

## 1.5 Detector-level Higgs AC analysis

This section illustrates the phenomenology of AC of the Higgs boson. The analysis presented in this study follows the same formalism as previous investigations on anomalous couplings by using CMS datasets from Run 1 and Run 2 [45, 46]. Further in-depth discussions of the theoretical concepts of the analysis can be found in Ref. [47, 48]. The current analysis focuses on examining anomalous effects in the EWK and gluon fusion productions and the decay of the Higgs in two W bosons (HWW). The interaction between a spin-0 Higgs boson and two spin-1 gauge bosons denoted as  $V_1 V_1$  (which could be WW, ZZ,  $Z\gamma$ ,  $\gamma\gamma$ , or gg), is conveniently described by the following scattering amplitude, which provides a very generic parametrization:

$$A(HV_1 V_1) \sim \left[ a_1^{VV} + \frac{\kappa_1^{VV} q_1^2 + \kappa_2^{VV} q_2^2}{(\Lambda_1^{VV})^2} + \frac{\kappa_3^{VV} (q_1 + q_2)^2}{(\Lambda_Q^{VV})^2} \right] m_{V_1}^2 \epsilon_{V_1}^* \epsilon_{V_2}^* + a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2),\mu\nu} + a_3^{VV} \tilde{f}_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu\nu}, \quad (1.25)$$

where  $f^{(i)\mu\nu}$  is defined as  $\epsilon_{V_i}^\mu q_{V_i}^\nu - \epsilon_{V_i}^\nu q_{V_i}^\mu$ , and  $\tilde{f}^{(i)}_{\mu\nu}$  is expressed as  $\frac{1}{2} \epsilon_{\mu\nu\rho\sigma} f^{(i),\rho\sigma}$ . Here,  $\epsilon_{V_i}$  and  $m_{V_1}$  represent the polarization vector and pole mass of a gauge boson, respectively, and  $\Lambda_1$  and  $\Lambda_Q$  are the scales of BSM physics. It is worth noticing that this equation accounts for couplings to both electroweak bosons and gluons, allowing  $HV_1 V_1$  to represent either HVV or Hgg. The sign of the photon field is determined by the sign in front of the gauge fields in the covariant derivative, establishing the sign convention of the  $Z\gamma$  couplings. This analysis adopts the positive sign convention used by the JHUGEN generator [49]. At the tree-level, the dominant contributions arise from  $a_1^{ZZ}$  and  $a_1^{WW}$ . The remaining couplings for ZZ and WW are considered anomalous. Small anomalous terms arise in the SM due to loop effects. Although the SM values of these couplings are not yet accessible experimentally, new BSM contributions can make them larger. Symmetry and gauge invariance considerations require  $\kappa_1^{ZZ} = \kappa_2^{ZZ} = -\exp(i\phi_{\Lambda_1}^{ZZ})$ ,  $\kappa_1^{\gamma\gamma} = \kappa_2^{\gamma\gamma} = 0$ ,  $\kappa_1^{gg} = \kappa_2^{gg} = 0$ ,  $\kappa_1^{Z\gamma} = 0$ ,  $\kappa_2^{Z\gamma} = -\exp(i\phi_{\Lambda_1}^{Z\gamma})$ , and  $\kappa_3^{VV} = 0$ .

When considering the couplings for  $\gamma\gamma$  and  $gg$ , the first tensor structure in Eq. (1.25) is not relevant. Instead, only the  $a_2^{\gamma\gamma,gg}$  and  $a_3^{\gamma\gamma,gg}$  terms are present, with the former arising from loop effects in the SM. If there are CP-odd  $a_3^{V_1 V_1}$  couplings present in conjunction with other couplings (which are CP-even), this may result in CP violation in the respective process. To reduce the number of independent

parameters, we assume that  $a_2^{\gamma\gamma, Z\gamma}$  and  $a_3^{\gamma\gamma, Z\gamma}$  are constrained by the direct decays  $H \rightarrow \gamma\gamma$  and  $Z\gamma$ , and set them to zero. The relationship between HZZ and HWW couplings plays a crucial role in VBF production with ZZ and WW fusion in this analysis. However, it is not possible to distinguish between the two couplings based on the available data, as there are no kinematic differences.

## Analysis interpretations

The HVV results are interpreted in two ways: in terms of the AC scattering amplitude above or in the context of the SMEFT formulation. The distinction between both approaches is now discussed.

**AC interpretation (Approach 1)** It assumes custodial symmetry where HZZ and HWW couplings are equal, and thus we drop the ZZ superscript for EWK vector boson couplings, resulting in four HVV anomalous couplings to measure:  $a_2$ ,  $a_3$ ,  $\kappa_1\Lambda_1$ , and  $\kappa_2^{Z\gamma}\Lambda_1^{Z\gamma}$ . The presence of CP-odd  $a_3$  couplings together with any of the other couplings (all of them CP-even) will result in CP violation. This approach allows for a general test of the Higgs boson tensor structure and a search for CP violation in HVV interactions, similar to the Run 1 AC analysis [50].

**SMEFT interpretation (Approach 2)** By adopting the  $SU(2) \times U(1)$  symmetry into the AC amplitude, it becomes equivalent to the approach followed in the SM EFT. As a result, the number of independent anomalous couplings for HVV interactions is now reduced to three:  $a_2$ ,  $a_3$ , and  $\kappa_1\Lambda_1$ . These couplings exhibit the following relationships:

$$a_1^{WW} = a_1^{ZZ}, \quad (1.26)$$

$$a_2^{WW} = c_w^2 a_2^{ZZ}, \quad (1.27)$$

$$a_3^{WW} = c_w^2 a_3^{ZZ}, \quad (1.28)$$

$$\frac{\kappa_1^{WW}}{(\Lambda_1^{WW})^2} = \frac{1}{c_w^2 - s_w^2} \left( \frac{\kappa_1^{ZZ}}{(\Lambda_1^{ZZ})^2} - 2s_w^2 \frac{a_2^{ZZ}}{M_Z^2} \right), \quad (1.29)$$

$$\frac{\kappa_2^{Z\gamma}}{(\Lambda_1^{Z\gamma})^2} = \frac{2s_w c_w}{c_w^2 - s_w^2} \left( \frac{\kappa_1^{ZZ}}{(\Lambda_1^{ZZ})^2} - \frac{a_2^{ZZ}}{M_Z^2} \right). \quad (1.30)$$

where  $c_w = \cos \theta_W$ ,  $s_w = \sin \theta_W$ , and  $\theta_W$  represents the weak mixing angle. In the SMEFT interpretation, there exists a linear relationship between the amplitude

couplings and the EFT couplings in the *Higgs basis* (described in the previous section), as explained in [30]:

$$\delta c_z = \frac{1}{2}a_1^{ZZ} - 1, \quad (1.31)$$

$$c_{zz} = -\frac{2s_w^2 c_w^2}{e^2} a_2^{ZZ}, \quad (1.32)$$

$$\tilde{c}_{zz} = -\frac{2s_w^2 c_w^2}{e^2} a_3^{ZZ}, \quad (1.33)$$

$$c_{z\Box} = \frac{m_Z^2 s_w^2}{e^2} \frac{\kappa_1^{ZZ}}{(\Lambda_1^{ZZ})^2}. \quad (1.34)$$

$$(1.35)$$

Note: These Higgs basis couplings are 0 in the SM.

## Analysis measurements

The main emphasis of the study is put on measuring the couplings of the Higgs boson to vector bosons (HVV) and gluons (Hgg).

**HVV vertex** We perform the measurement of the effective cross-section ratios denoted as  $f_{ai}$  (or  $F_{ai}$ ) instead of the anomalous couplings themselves for convenience. Measuring ratios allows to cancel out many systematic uncertainties and constrains the physical range between -1 and 1. As such, our primary measurements are carried out in terms of cross-section ratios. For electroweak vector boson couplings, the effective fractional cross sections  $f_{ai}$  are defined as follows:

$$f_{ai} = \frac{a_i^2 \sigma_i}{\sum_j a_j^2 \sigma_j} \text{sign} \left( \frac{a_i}{a_1} \right) \quad (1.36)$$

where  $\sum_j$  encompasses all couplings considered, and  $\sigma_i$  represents the cross section for the process corresponding to  $a_i = 1$ . Table 1.2 provides the numerical values for the effective cross sections ( $\sigma_i$ ). For consistency with previous CMS measurements, the  $\sigma_i$  coefficients used to define the fractional cross sections correspond specifically to the  $gg \rightarrow H \rightarrow ZZ \rightarrow 2e2\mu$  process. These values were calculated using the JHUGEN 7.0.2 event generator [49], and sets of values are presented based on different assumptions for the coupling relationships: the AC approach with  $a_i^{ZZ} = a_i^{WW}$ , and the  $SU(2)_L \times U(1)_Y$  approach from Eqs. 1.26–1.30.

**Hgg vertex** Finally, when considering Hgg interactions, the effective fractional cross sections  $f_{a3}^{ggH}$  is defined as follows:

$$f_{a3}^{ggH} = \frac{|a_3^{gg}|^2 \sigma_3}{|a_2^{gg}|^2 \sigma_2 + |a_3^{gg}|^2 \sigma_3} \text{sign} \left( \frac{a_3^{gg}}{a_2^{gg}} \right) \quad (1.37)$$

The cross sections  $\sigma_3$  and  $\sigma_2$  correspond to  $a_3^{gg} = 1$  and  $a_2^{gg} = 1$ , respectively, and they are equal. For this case, it is impossible to distinguish the contribution to the ggH loop of heavy particles, such as the top quark or bottom quark, from the kinematic features of the event. Thus, the Hgg coupling is considered effective with the heavy degrees of freedom (quarks) integrated out. However, analyzing the Hgg coupling with  $t\bar{t}H$  and  $tH$  processes together may help identify the top quark contributions in the loop based on the relative rates of the processes.

TABLE 1.2: Cross sections for the anomalous contributions ( $\sigma_i$ ) relative to the SM value ( $\sigma_1$ ) are used to define the fractional cross sections  $f_{ai}$ . In the case of the  $\kappa_1$  and  $\kappa_2^{Z\gamma}$  couplings, the numerical values  $\Lambda_1 = \Lambda_1^{Z\gamma} = 100$  GeV are considered to keep all coefficients of a similar order of magnitude. Two sets of values corresponding to the AC ( $a_i^{ZZ} = a_i^{WW}$ ) and the SU(2) x U(1) (Eqs. 1.26–1.30) coupling relationships are shown.

| Coupling                  | AC $\sigma_i/\sigma_1$ | SMEFT $\sigma_i/\sigma_1$ |
|---------------------------|------------------------|---------------------------|
| $f_{a3}$                  | 0.153                  | 0.153                     |
| $f_{a2}$                  | 0.361                  | 6.376                     |
| $f_{\Lambda 1}$           | 0.682                  | 5.241                     |
| $f_{\Lambda 1}^{Z\gamma}$ | 1.746                  |                           |

### 1.5.1 Overview of Run 2 results

The study of anomalous Higgs couplings to bosons and fermions is currently a highly active area of research in particle physics. As proven by numerous measurements which have been conducted in various Higgs production and decay channels. Table 1.3 contains a summary of the CMS Run 2 results from AC analyses released since 2021, which establish the context for the study presented here. Of particular interest are the on-shell Higgs decays to ZZ (1) and Higgs decays to  $\tau\tau$  (3), as they share similarities with the AC HWW study conducted in this thesis. In the following sections, we provide a brief description of the HZZ and H $\tau\tau$  analyses, highlighting the main results relevant to this thesis.



| # | Channel                      | Measured vertex | Combined with  | REF  |
|---|------------------------------|-----------------|--|------|
| 1 | on-Shell $H \rightarrow ZZ$  | HVV, Hgg, Htt   | [Htt] $H \rightarrow \gamma\gamma$ [51]                  | [46] |
| 2 | off-Shell $H \rightarrow ZZ$ | HVV             | on-Shell $H \rightarrow ZZ$                              | [52] |
| 3 | $H \rightarrow \tau\tau$     | HVV, Hgg, Htt   | on-Shell $H \rightarrow ZZ + H \rightarrow \gamma\gamma$ | [53] |
| 4 | ttH and tH                   | Htt             | on-Shell $H \rightarrow ZZ + H \rightarrow \gamma\gamma$ | [54] |
| 5 | $H \rightarrow \tau\tau$     | $H\tau\tau$     | -  | [55] |

TABLE 1.3: Run 2 AC analyses with the CMS experiment released since 2021.

**on-Shell  $H \rightarrow ZZ$  (1): Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state [46]**

Studies of the on-shell Higgs boson decaying to  $ZZ$  are an essential reference for the analysis presented in this thesis, as they focus on the four-lepton channels, including  $4e$ ,  $4\mu$ , and  $2e2\mu$  final states, and consider events from ggH, VBF, VH, and ttH production mechanisms. Just like the AC Higgs boson to WW analysis, this analysis depends on dedicated measurements at the detector level. It makes use of the kinematic information of the particles involved in the production and present in the final state to construct multidimensional kinematic discriminants, which are then employed to extract the results. In particular, this analysis set constraints on anomalous couplings using the AC and SMEFT approaches. Picking the CP-odd coupling as an illustrative example, the plot on the left of Fig. 1.15 displays the likelihood scan for the  $a_3$  coupling, while that on the right shows the same results in terms of the corresponding SMEFT coupling  $\tilde{c}_{ZZ}$ . The high resolution of the four-lepton channels allows for setting stringent constraints on the Higgs boson couplings to vector bosons, gluons, and fermions.

**$H \rightarrow \tau\tau$  (3): Constraints on anomalous Higgs boson couplings to vector bosons and fermions from the production of Higgs bosons using the  $\tau\tau$  final state [53]**

The study focuses on the decay channels  $\tau_h\tau_h$ ,  $\mu\tau_h$ ,  $e\tau_h$  and  $e\mu$ , and uses several production modes depending on the vertex being measured. The VBF production analysis is used to constrain the HVV vertex, while the ggH production allows it to constrain the Hgg vertex. Discriminants sensitive to the kinematics of the associated particles in production are combined to construct 2D and 3D distributions. The results of the likelihood scan of the  $f_{a_3}$  parameter for this analysis, both

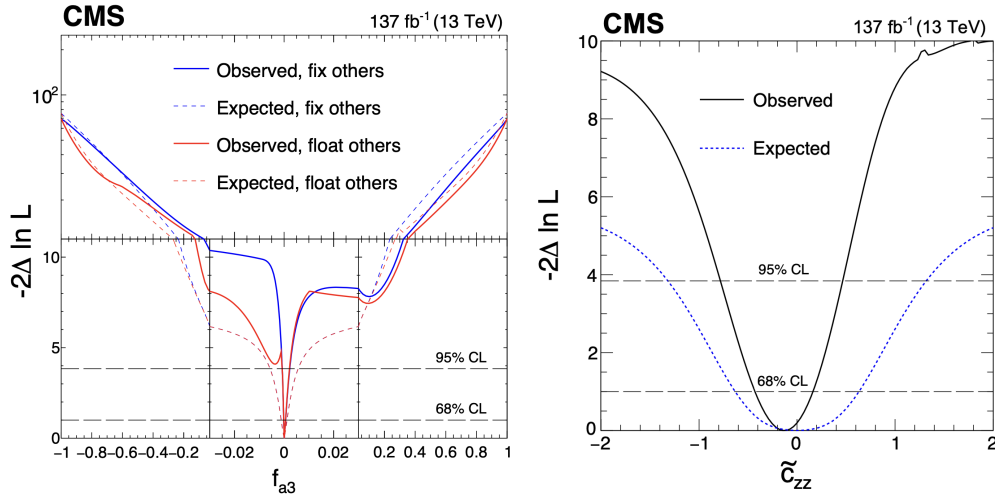


FIGURE 1.15: Observed (solid) and expected (dashed) likelihood scans from a simultaneous fit performed to data with signal and background modeling. This is shown for the parameters  $f_{a3}$  (left) and  $\tilde{c}_{ZZ}$  (right) when following the SMEFT approach.

independently and when combined with the on-shell HZZ, are shown in Fig. 1.16 on the left and right, respectively. Figure 1.17 displays the constraint on the  $f_{a3}^{ggH}$  parameter combined with the on-shell HZZ analysis mentioned above. The latest measurement excludes a pure CP-odd scenario of the Higgs boson coupling to gluons at  $2.4\sigma$  and supersedes the result presented in [46], described before.

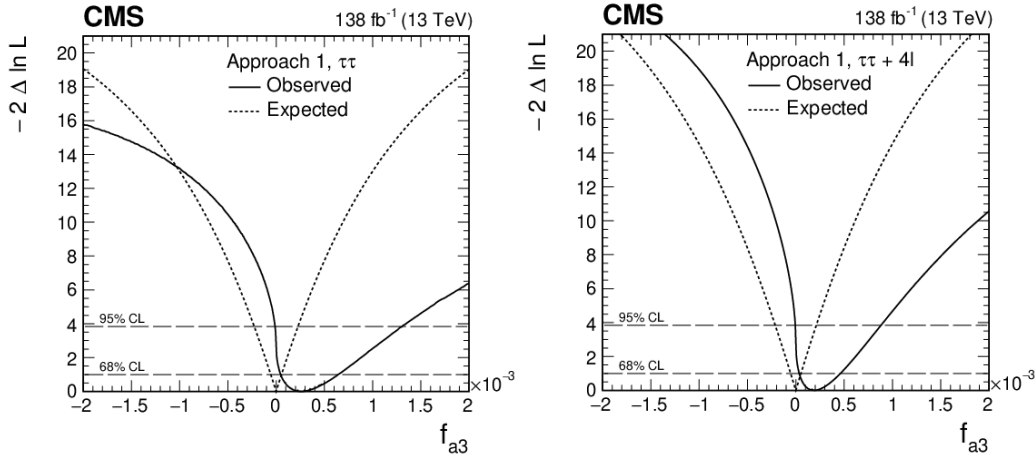


FIGURE 1.16: Observed (solid) and expected (dashed) likelihood scan of the  $f_{a3}$  parameter in the  $H \rightarrow \tau\tau$  analysis. The results are presented independently for this analysis on the left and when combined with the on-shell HZZ on the right, following the AC interpretation.

**Analyses combination** Combining multiple analyses improves our understanding of the Higgs boson couplings. This thesis represents the first use of the full

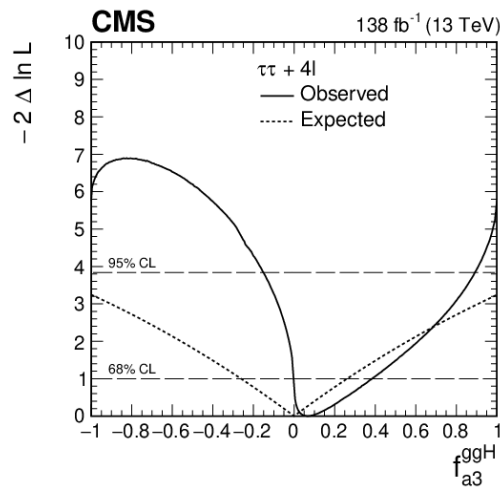


FIGURE 1.17: Observed (solid) and expected (dashed) likelihood scans of the  $f_{a3}^{ggH}$  parameter when combined with the on-shell HZZ analysis.

Run 2 dataset for the analysis of anomalous couplings of the Higgs boson in the WW decay channel. The last analysis of this type was conducted using Run 1 [50]. Additionally, the  $H \rightarrow bb$  and  $H \rightarrow \gamma\gamma$  AC analyses using the full Run 2 dataset are still forthcoming.

## ATLAS analyses

The ATLAS experiment also seeks to discover NP by conducting searches for anomalous couplings of the Higgs boson. To this purpose, the ATLAS Collaboration has performed analyses similar to those shown previously for the CMS experiment, and the references of these analyses are summarized in Table 1.4. As with the measurement of the Higgs boson mass, the LHC experiment aims to combine the results of these anomalous coupling analyses to achieve a more comprehensive understanding of the properties of the Higgs boson and any potential deviations from SM expectation. The latest result from the ATLAS experiment

| ATLAS Analysis ( $\mathcal{L} = 139 \text{ fb}^{-1}$ of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ ) | Ref  |
|--|------|
| H(125) CP $\rightarrow 4\ell$  | [56] |
| VH(125) H $\rightarrow b\bar{b}$ boosted   | [57] |
| H(125) $\rightarrow \gamma\gamma$ differential cross sections  | [58] |
| Study of top-Higgs CP properties with ttH and tH events with H decays $\rightarrow b\bar{b}$           | [59] |
| H(125) $\rightarrow \gamma\gamma$ STXS and couplings   | [60] |

TABLE 1.4: AC analyses from ATLAS using the LHC Run 2 data set.

on analyzing anomalous couplings relates to the study of the process  $H \rightarrow ZZ$ , which is analogous to the CMS analysis discussed earlier. Both experiments have employed this process to set constraints on the EFT couplings; however, there is no ATLAS AC analysis using the HWW channel as the one presented in this thesis. To illustrate one example, Fig. 1.18 presents the  $\tilde{c}_{ZZ}$  scan from Ref [56] that is consistent with the CMS result previously shown in Fig. 1.15.

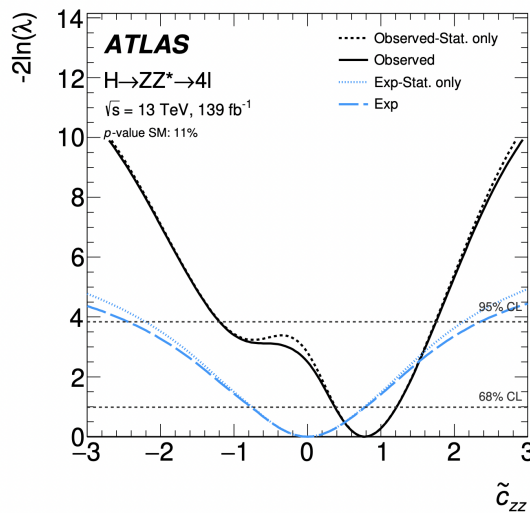


FIGURE 1.18: Likelihood scan of the EFT coupling  $\tilde{c}_{ZZ}$  [56].

## Chapter 2

# The CMS experiment and the LHC precision frontier

The *European Laboratory for Particle Physics*, which was established in 1954 near Geneva on the France-Switzerland border, is widely known as CERN standing for “*Conseil Européen pour la Recherche Nucléaire*”. CERN facility serves as the host to the most extensive and powerful particle accelerator in the world, where over 12,000 scientists from more than 110 countries work in pursuit of scientific knowledge.

Chapter 2 of the thesis focuses on the Compact Muon Solenoid (CMS) experiment and its significant role in advancing the precision of particle physics research. As one of the general-purpose experiments at the LHC, this chapter begins by providing an overview of the accelerator complex, design parameters, and operational runs. It then delves into the CMS apparatus itself, which plays a crucial role in capturing the immense volume of data generated by the proton collisions. The chapter proceeds to describe the CMS detector coordinate system, subdetectors, and trigger and data acquisition systems. Furthermore, it explores the reconstruction of various objects, including muons, electrons, photons, jets, and neutrinos, using data from the CMS detector. Lastly, it briefly touches upon the Phase 2 upgrades planned for CMS, particularly in anticipation of the High-Luminosity LHC (HL-LHC), which will significantly enhance the experimental capabilities of CMS and facilitate even more precise measurements. This chapter aims to uncover the intricate details of CMS and highlight its contributions to the forefront of particle physics research.

## 2.1 The Large Hadron Collider

The LHC is a circular particle accelerator with a perimeter of 26.7 km, buried at an average depth of approximately 100 m underground, and located in the tunnel of its predecessor, LEP [17]. The majority of the LHC tunnel is filled with superconducting magnets capable of propelling protons or heavy ion beams. While lead ion collisions (PbPb) or proton-Pb collisions are employed to analyze the dynamics of quarks and gluons in the QGP state, this thesis will not cover them. Instead, this work will focus on probing fundamental physics through pp collisions.

The peak centre-of-mass energy by design is  $\sqrt{s} = 14$  TeV and an instantaneous luminosity ( $\mathcal{L}$ , discussed in this chapter) of  $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Although the maximum energy ever achieved in operation is 13.6 TeV, the accelerator ran at 13 TeV during the 2016-2018 period when the data used in this thesis was collected.

### 2.1.1 Accelerator complex

In the LHC tunnel, two vacuum pipes accommodate the circulating proton beams, traveling in opposing directions. The beams subsequently collide at four interaction points, where the four principal experiments of the LHC are conducted. Before achieving the highest energy level, the proton beams undergo various stages delineated in the accelerator complex illustrated in Fig. 2.1. The LHC represents the final ring in a sequence of pre-existing particle accelerators.

Hydrogen gas serves as the source of protons for acceleration within the LHC. The protons are obtained by removing electrons from the hydrogen atoms via ionizing energy of 13.6 eV. Initially, the protons are driven to a Radio Frequency Quadrupole (RFQ), grouped into bunches, and accelerated to reach an energy of 750 keV. The subsequent stage involves transporting the proton bunches to the Linear Accelerator (LINAC 2), where their energy is increased to 50 MeV, enabling their transfer to the Proton Synchrotron Booster (PSB), which is the first circular accelerator in the chain that enhances the intensity of the proton bunches during its 150 m length, with the beam energy level reaching up to 1.4 GeV. Subsequently, the Proton Synchrotron (PS), with a length of 620 m, and the Super Proton Synchrotron (SPS), spanning 6,912 m, are utilized to further accelerate the proton beams up to 25 GeV and 450 GeV, respectively. Prior to injection into the

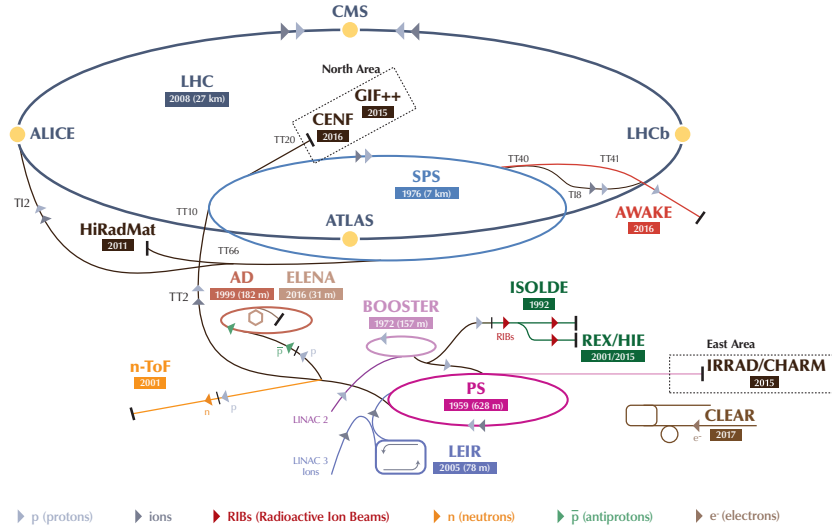


FIGURE 2.1: Schematic of the accelerator complex at CERN. Protons are accelerated and their energy levels are increased at multiple accelerators, namely LINAC 2 (in purple), BOOSTER (in pink), PS (in magenta), SPS (in blue), and ultimately at the LHC (in grey). The proton beams once circulated through the LHC, collide at the center of four particle detectors, namely CMS, ATLAS, ALICE, and LHCb.

LHC, the beam splits into two opposing parallel beamlines, employing fast kicker magnets, and subsequently accelerates to the desired energy level via high radio frequency (RF) cavities.

The design of the LHC presents a significant challenge in its endeavor to accelerate two opposing beams of positively charged protons. A total of 1232 superconducting NbTi dipoles magnets disposed into eight 2.45 km long arcs are in charge of controlling the beam optics while maintaining a circular trajectory by generating a field of 8.3 T [61]. The dipoles accelerate the protons because they are capable of inverting such a magnetic field over the span of a few cm. The radius of their curved structure is related to the momentum of the accelerated particles in a magnetic field by the expression:

$$\rho[\text{m}] = \frac{p[\text{GeV}]}{0.3 \cdot B[\text{T}]} \quad (2.1)$$

The quadrupoles are used to keep the particles focused transversely and lengthwise in the narrow size of bunches fulfilling one of the most stringent requirements of the accelerator. All the magnets are kept to a temperature of 1.9 K (-271.25°C) with superfluid helium-4 ensuring their superconducting properties.

Once the proton beams reach the nominal energy and are stable, the protons collide at four different interaction points (IP) along the LHC instrumented with particle detectors. ATLAS [62] (A Toroidal Lhc Apparatus) and CMS are the main general-purpose detectors initially designed to hunt the Higgs boson and possibly BSM scenarios at the TeV scale. They mutually cross-check physics observations at opposite points on the LHC tunnel. ALICE [63] (A Large Ion Collider Experiment) studies ion-ion collisions, and LHCb [64] (LHC beauty) consists of an asymmetric single-arm detector meant for heavy flavor physics, CP violation, rare decays by measuring decays of B-type hadrons and charm quark. Besides the four biggest-sized detectors, more than a dozen of experiments with diverse physics targets belong to the CERN-LHC complex, such as LHCf, TOTEM, FASER, SND@LHC, AMS, MoEDAL-MAPP, among others [65].

### 2.1.2 Design parameters

The aforementioned chain of accelerators yields significant energy losses due to radial acceleration, which imposes a limitation on the achievable energy in a circular collider. Therefore, protons are used because they are stable, charged, and suffer much smaller synchrotron radiation losses than electrons due to their larger mass. The nominal center-of-mass energy of the pp collisions of 14 TeV means that each proton beam has an energy of 7 TeV, while the sea of quarks and gluons that compose the protons carries a fraction of the momentum as described in section 1.2.1. Besides the beam energy, the main LHC performance indicator is the instantaneous luminosity  $\mathcal{L}$ , which characterizes the collision rate relating to the number of events per unit of time produced for a given process with a particular cross-section  $\sigma$ :

$$\frac{\partial N}{\partial t} = \mathcal{L} \times \sigma. \quad (2.2)$$

Hence, a sizeable instantaneous luminosity is essential to produce low-probability processes, but it is also a challenge for data acquisition systems, so it cannot be increased indefinitely. In turn, the integrated luminosity is obtained by integrating  $\mathcal{L}$  over time. The unit for instantaneous luminosity is commonly expressed as  $\text{cm}^{-2}\text{s}^{-1}$ , while the integrated luminosity is typically expressed in units of inverse



picobarns or femtobarns ( $\text{pb}^{-1}$  or  $\text{fb}^{-1}$ )<sup>1</sup>. The relationship between the instantaneous luminosity and the properties of the beam is given by the expression:

$$\mathcal{L} = \frac{N_p^2 n_b f \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (2.3)$$

Table 2.1 lists the values of various parameters for the LHC nominal design, including  $N_p$ , which represents the number of protons per bunch, and  $n_b$ , which is the number of bunches. The parameter  $f$  denotes the revolution frequency of the bunches, and  $\gamma_r$  is the relativistic factor. The transverse emittance  $\epsilon_n$  describes the confinement of the beam in space and momentum, while the beta function  $\beta^*$  characterizes its focus at the interaction point. Lastly,  $F$  is a geometric factor that accounts for the reduction in luminosity due to the crossing angle of the beams at the interaction point, with  $F$  being less than or equal to 1.

TABLE 2.1: LHC nominal parameters in pp collisions.

| Symbol        | Parameter                             | Nominal value                                     |
|---------------|---------------------------------------|---|
| $\sqrt{s}$    | Centre-of-mass energy                 | 14 TeV  |
| $\Delta t$    | Bunch spacing                         | 25 ns   |
| $\mathcal{L}$ | Instantaneous luminosity              | $1.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ |
| $n_b$         | Number of bunches per beam            | 2808  |
| $N_p$         | Number of protons per bunch           | $1.15 \times 10^{11}$                             |
| $f$           | Revolution frequency                  | 11245 Hz  |
| $\epsilon_n$  | Transverse emittance                  | 3.75 mm $\mu\text{m}$                             |
| $\beta^*$     | Beta function                         | 0.55 m  |
| $F$           | Geometric luminosity reduction factor | 0.95 (2011)                                       |

In the nominal design of the LHC, each beam contains 2808 proton bunches, with each bunch containing approximately  $10^{11}$  protons. The bunches are arranged in trains of 48 bunches (referred to as the “48b” scheme) spaced at intervals of 25 ns and complete about 11,000 revolutions per second, which is only 3.1 m/s slower than the speed of light. Of these, 2544 bunches collide at the CMS interaction point at a collision rate of 40 MHz, resulting in a luminosity of about  $1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  at the beginning of the fill, when proton injection is complete, and no more bunches can be accommodated.

One of the most significant drawbacks of the high instantaneous luminosity of the LHC is the large *pile-up* (PU) probability which causes multiple interactions to occur simultaneously within the same bunch-crossing. The level of pile-up is

---

<sup>1</sup>1 pb =  $10^{-36} \text{ cm}^2$

directly proportional to the instantaneous luminosity and is an average function of the beam properties. This relation is expressed by the following equation:

$$\langle PU \rangle = \mathcal{L} \frac{\sigma_{pp}^{inel}}{n_b f}, \quad (2.4)$$

At  $\sqrt{s} = 13$  TeV, the inelastic pp cross-section ( $\sigma_{pp}^{inel}$ ) is 69 mb, resulting in a nominal average PU of about 22 interactions per bunch crossing at the LHC [66]. However, the LHC has gradually increased its operational values and currently experiences a PU rate of about 60 interactions per bunch crossing. This number is expected to become even larger at the HL-LHC, planned to begin in approximately 2029, with an average PU rate of  $\langle PU \rangle = 140/200$  events. Large PU values may significantly degrade the detector efficiency and resolution of particle reconstruction. To address this issue, all LHC experiments underwent an upgrade program during the Long Shutdown 2 and plan to implement further upgrades in preparation for the HL-LHC conditions.

### 2.1.3 LHC operational runs

After more than a decade of installation and commissioning, proton beams were first circulated into the LHC in September 2008. The first high-energy p-p collision at 7 TeV took place in March 2010, marking the start of Run 1. A run refers to a collection of collision data gathered using the same detector configuration and fixed center-of-mass energy. Several runs are then combined into global LHC Runs, which typically span several years.

Although LHC Run 1 ended by 2013, the delivered luminosity of  $6 \text{ fb}^{-1}$  at 7 TeV by the end of 2011 was enough for the ATLAS and CMS Collaborations to constrain the mass range of the Higgs boson and afterward proclaim its discovery in July 2012 [1, 2] with  $23 \text{ fb}^{-1}$  collected at 8 TeV by the first half of 2012. Henceforth, the luminosity and energy of the LHC have been consistently amplified to attain its intended design value. Beginning in 2015 with Run 2, the LHC has operated at 13 TeV, ultimately generating a total luminosity of  $139 \text{ fb}^{-1}$  by the conclusion of 2018. The progress of the integrated luminosity recorded at the CMS detector and peak luminosities since the beginning of the LHC program is depicted in Fig. 2.2.

In turn, Fig. 2.3 presents an overview of the prior operational history of the LHC and the timetable for upcoming Runs. Subsequent to Run 1, the LHC was put on

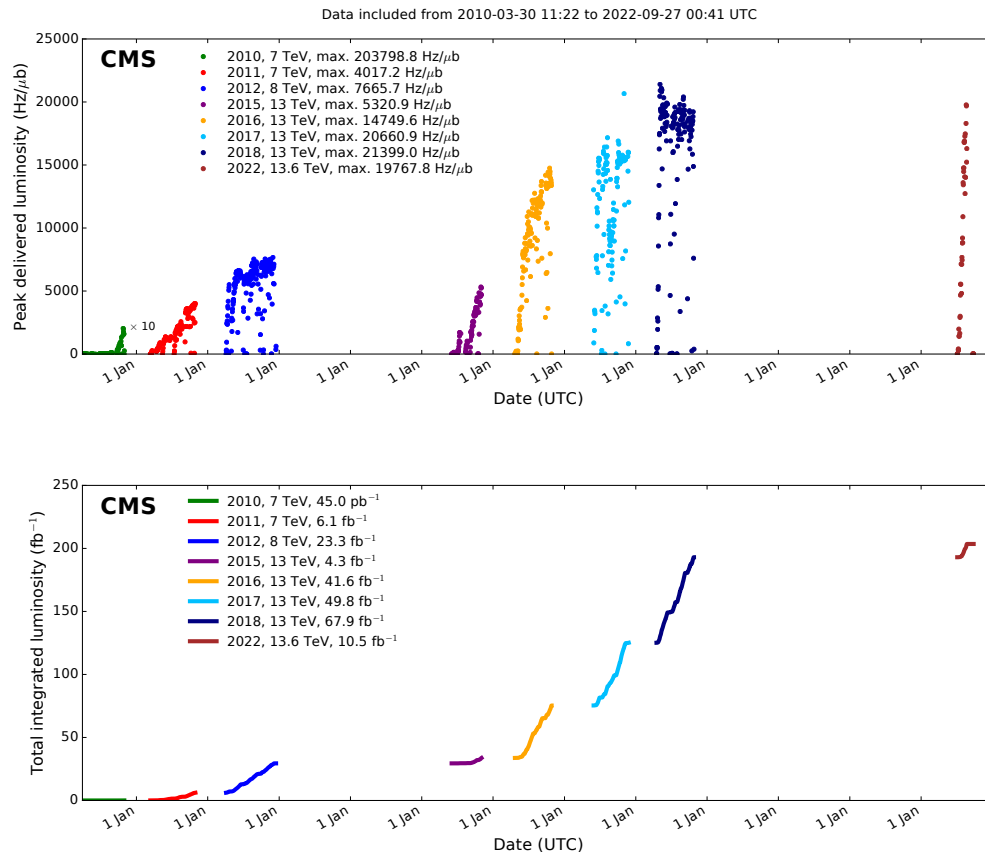


FIGURE 2.2: Peak (top) and total integrated luminosity (bottom) delivered per day to the CMS experiment during the LHC Run 1 and Run 2.

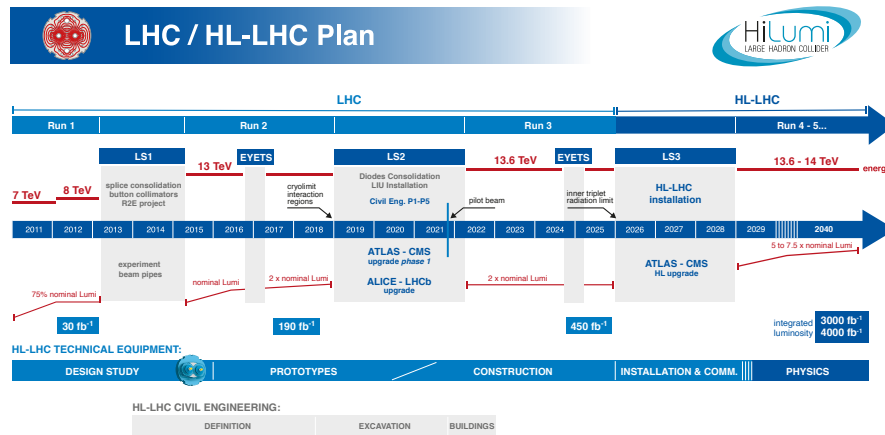


FIGURE 2.3: Baseline schedule of the LHC and HL-LHC operations.

hold for 2 years (known as Long Shutdown 1 or LS1) to upgrade the accelerator in preparation for the demands of Run 2, which included upgrades to the magnets and some of the subdetectors used in the experiments. Broadly speaking, Phase 1

refers to the time period from 2011 to 2024, while Phase 2 encompasses the period from 2029 to 2037, which is associated with the High-Luminosity era of the LHC.

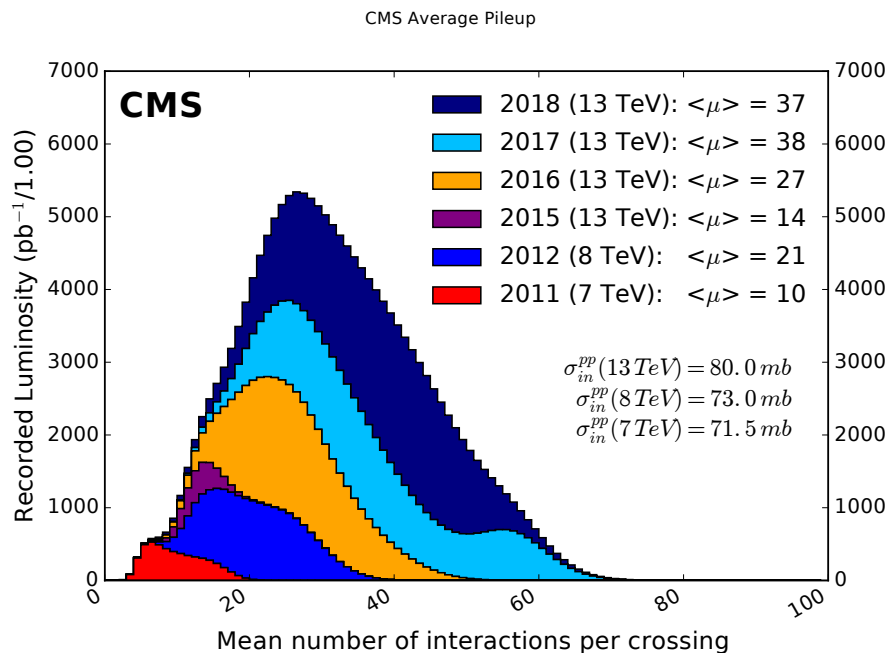


FIGURE 2.4: Distributions of the average number of interactions per bunch crossing (pileup) for pp collisions in 2015 (pink), 2016 (yellow), 2017 (blue), 2018 (purple), 2017 (light blue), and full Run 2 (gray). The overall mean values and the minimum bias cross-sections are also shown.

Figure 2.4 depicts the progressively increasing average pileup rates from 2011 to 2018 caused by higher luminosities. With the LHC Run 2 dataset spanning from 2015 to 2018, colliding at  $\sqrt{s} = 13$  TeV and delivering the designed instantaneous luminosity (as listed in Table 2.1), the possibilities for exploring new physics scenarios expanded. The CMS detector registered integrated luminosities of  $36.3 \text{ fb}^{-1}$ ,  $41.5 \text{ fb}^{-1}$ , and  $59.7 \text{ fb}^{-1}$  in 2016, 2017, and 2018, respectively, which constitute the datasets utilized for the central analysis presented in this thesis.

## 2.2 The CMS detector

In the preceding section, the CMS detector was introduced as a multi-purpose experiment situated at the CERN LHC, designed to explore various physics scenarios at the TeV energy scale and search for the Higgs boson and BSM physics.

The current section describes the structural characteristics of CMS, including the technical details of its subdetectors.

The CMS apparatus is housed in a massive experimental cavern 100 m below ground level at the LHC Interaction Point (IP) number 5, also known as P5, dedicated to CMS. The cylindrical structure of the device has a total diameter of 15 m and a length of 21.5 m. As indicated in its name, CMS is a compact apparatus in terms of its physical size, being roughly half the size of the ATLAS experiment but twice as heavy, weighing approximately 14,000 tons.

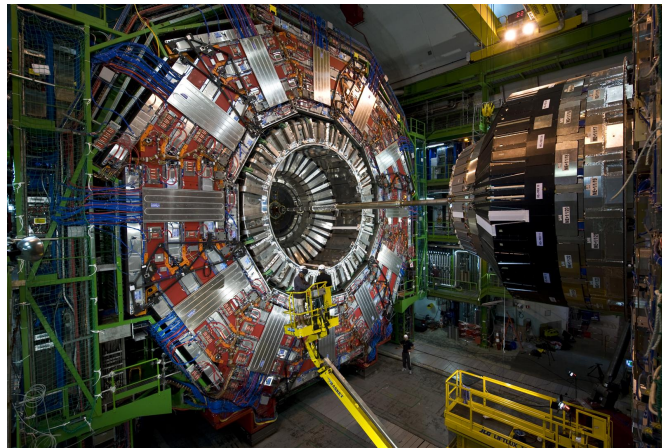


FIGURE 2.5: View of the CMS detector inside the experimental cavern at P5.

The CMS detector has a concentric subdetector design visible in Fig. 2.5. Its most important component is a 4T superconducting solenoid magnet that bends the trajectory of charged particles transversely crossing CMS, enabling the precise measurement of their momenta, paths, and the determination of the interaction vertex they emerged from. The whole inner tracker will be housed in a cylindrical support structure with a diameter of 2.4 m and a total length of 5.4 m. The tracker system includes two high-granular tracking detectors: the pixel and strip trackers. The electromagnetic and hadronic calorimeters surround the tracking system and absorb and measure the energy of electrons, photons, and hadrons. The CMS muon system is located in the outermost part of the detector, surrounding the magnet and adding an additional radius of 7 m, and taking up about 80% of the total volume of the detector.

During data-taking, pp collisions occur in the IP every 25 ns, allowing for subsequent proton bunches to collide before the decay products of the previous collision reach the active volume of the detector, which would cause an overlap known

as out-of-time pileup. Hard interactions, resulting from high momentum transfer between partons, usually split the protons, and the remnants give rise to underlying events. The high-granularity design of the CMS helps distinguish such softer processes from the signal of interest.

### 2.2.1 Coordinate system

The CMS experiment employs a right-handed coordinate system [67], illustrated in Fig. 2.6, with the origin at the IP. The x-axis points towards the center of the LHC ring, while the y-axis points upwards and is tilted 1.41% with respect to the LHC plane, fully defining the transverse plane. The longitudinal z-axis aligns with the direction of the proton beam and points anticlockwise, being the starting of the cylindrical coordinate system. The azimuthal angle  $\phi$  is defined in the XY-plane with respect to the x-axis, while the radial coordinate  $r$  is measured from the origin. The polar angle  $\theta$  is determined in the yz-plane relative to the z-axis.

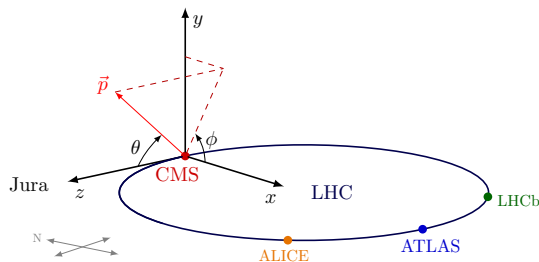


FIGURE 2.6: The coordinate system used by CMS is shown in this illustration, which also includes the LHC ring and a compass [67].

The left plot of Figure 2.7 shows the cartesian and cylindrical coordinate systems of the CMS detector. However, these coordinate systems are only applicable to describe macroscopic observables and cannot pinpoint pp collisions at the parton level. In hadronic colliders, the momentum fraction carried by each parton along the z-axis and the longitudinal boost of the collision products is unknown. Therefore, using longitudinally Lorentz boost-invariant quantities to describe the events is more convenient. The most frequently used observables to describe events in hadronic colliders are the azimuthal angle  $\phi$ , the transverse momentum, and the transverse mass. The latter two are defined as:

$$\begin{aligned} p_T^2 &= p_x^2 + p_y^2, \\ m_T^2 &= m^2 + p_x^2 + p_y^2 = E^2 - p_z^2 \end{aligned} \tag{2.5}$$

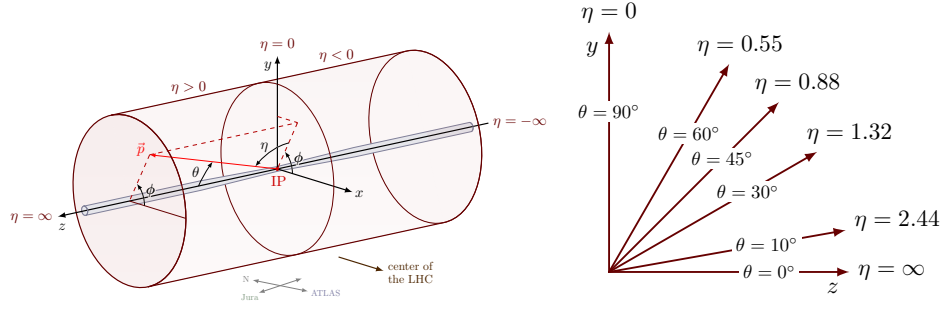


FIGURE 2.7: CMS coordinate system with the a cylindrical detector [67].

The transverse momentum is a valuable quantity in events that involve neutrinos, which are inferred through the missing transverse energy (treated later on), reflecting an imbalance in the total  $p_T$ . Additionally, the angle relative to the beam axis is expressed using the pseudo-rapidity  $\eta$ , defined as:

$$\eta = -\ln \tan(\theta/2). \quad (2.6)$$

The right side of Fig. 2.7 illustrates that the  $\eta$  value is 0 when the momentum vector of the particle is perpendicular to the beamline and approaches infinity when it becomes parallel. Detector regions with large  $\eta$  values are associated with the so-called forward direction. Notably, the pseudorapidity is an ultra-relativistic approximation ( $\beta \sim 1$ ) of the *rapidity* of a particle, which is defined as:

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right). \quad (2.7)$$

The definition of the angular separation between two particles is useful in assessing the isolation of the particle in relation to the surrounding objects and reconstructing interaction information.

$$\Delta R^2 = \Delta\phi^2 + \Delta\eta^2. \quad (2.8)$$

where  $\Delta\phi = |\phi_1 - \phi_2|$  and  $\Delta\eta = |\eta_1 - \eta_2|$ .

### 2.2.2 Subdetectors

CMS subdetectors are equipped with various materials and technologies to detect electrons, photons, hadronic particles, and muons. The information collected

from these subdetectors is crucial for offline particle reconstruction, as discussed in section 2.3.1. The following lines describe the current subdetectors in their configuration during the period of 2016-2018.

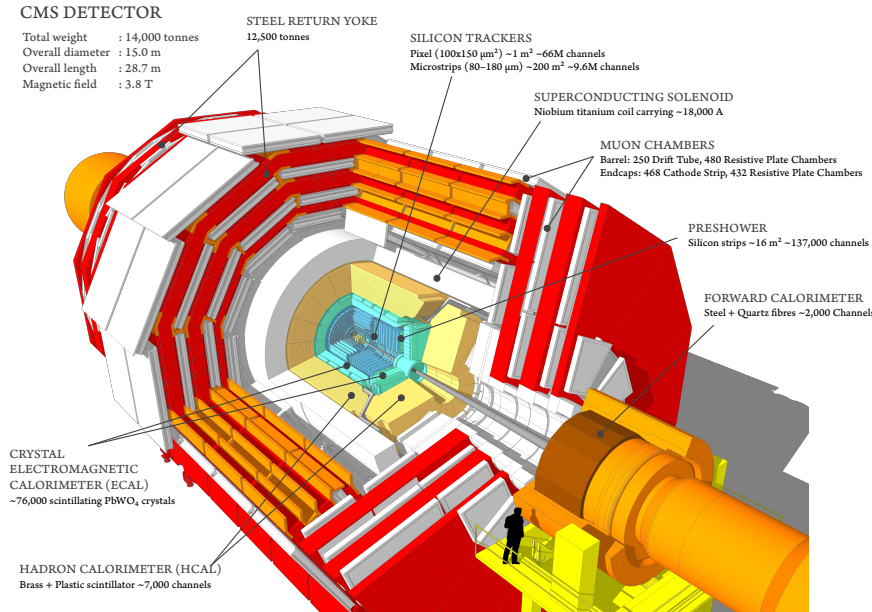


FIGURE 2.8: Schematic view of the CMS detector and its subcomponents [68].

On a broad scale, the CMS detector is divided into two main regions: the central barrel and two forward endcaps. Figure 2.8 depicts the CMS subdetectors, listed in order of proximity to the interaction point [69]:

- Silicon trackers consist of the inner pixel tracker and the outer strip tracker, which have decreasing granularity and reconstruct charged particle trajectories alongside the origin coordinate. Changes were made in 2017 and are detailed later on.
- Electromagnetic calorimeter (ECAL) measures the energy deposited by photons and electrons in the material, while the hadronic calorimeter (HCAL) measures the energy of hadrons from the resulting showers in the detector structure.
- Superconducting solenoid is an intense magnetic field that bends the trajectory of charged particles, inferring information about their charge and momentum.



- Muon chambers consist of a robust, redundant, and efficient muon spectrometer that mixes different technologies depending on the  $\eta$ -region for maximum reconstruction efficiency. This subsystem is located in the outermost part of CMS and includes drift tubes, cathode strip chambers, and resistive plate chambers.

The remainder of this section provides a more detailed view of the subsystems, starting with the solenoidal magnet due to its leading role within the subdetectors and the unique design of CMS. Any information related to the muon chambers is of special relevance to the work performed for commissioning the CMS Drift Tubes for the HL-LHC upgrades discussed in the dedicated chapter 3.

### Superconducting magnet

The superconducting solenoid magnet, which weighs 220 tons, is 12.5 m long and has a radius of 6 m, is made of NbTi alloy, and is cooled down to 4.7 K using liquid helium. Injecting currents of 18,000 A create a solenoidal magnetic field of 3.8 T that is nearly uniform and constant within its volume. This results in 2.7 GJ of energy. The magnet heavily influences the design of the whole detector and encloses the trackers and calorimeters to minimize non-active material in front of those subdetectors. The muon chambers surround the magnet and are situated in a steel return yoke that supports the structure mechanically, returns magnetic flux, and reduces the roaming field permeating the muon system in a 2 T magnetic field.

The bending of charged particles in a magnetic field with strength  $\vec{B}$  and velocity  $\vec{v}$  is associated with a Lorentz force given by  $\vec{F} = q(\vec{v} \times \vec{B})$ . Thus, the charge and momentum are obtained by analyzing the bending in conjunction with the tracker data (following Eq. 2.1). Numerical simulations are used to estimate the magnetic field strength and lines to ensure the highest accuracy and characterization of its properties and impact of the measurements. Figure 2.9 illustrates the CMS solenoid magnetic field, which is measured with a precision of 0.1% within the tracker volume.

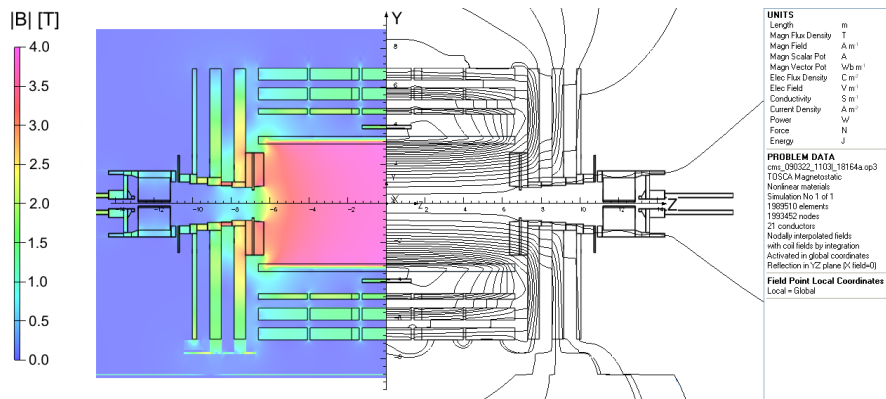


FIGURE 2.9: Value of  $|B|$  (left) and field lines (right) predicted on a longitudinal section of the CMS detector for the underground model at a central magnetic flux density of 3.8 T. Each field line represents a magnetic flux increment of 6 Wb [70].

## Tracking system

Located closest to the IP, the innermost detector of CMS is the tracker system. It covers an active area of 200 m<sup>2</sup> and has a geometrical acceptance up to  $|\eta| = 2.5$ . The acceptance of a detector refers to the range of kinematic parameters within which particles are potentially detectable, which is mainly the pseudorapidity or the transverse momentum range in case the detector is placed in the magnetic field.

Due to the decreasing detector occupancy with radial distance, the tracker is subdivided into the inner pixel tracker and the outer strip tracker. The closer to the beam pipe, the higher spatial precision is required to identify in-flight decays and discriminate primary vertexes from secondary vertexes. Hence, the tracker must satisfy stringent design requirements, detection efficiency, and radiation hardness to achieve excellent vertex reconstruction.

In addition to excellent vertex reconstruction, the tracker also aims to achieve good performance in the reconstruction of particle trajectories, energies, and charge over a wide range of particle momentum. High  $p_T$  leptons of 1 TeV are expected to be produced by high mass resonances, while the reconstruction efficiency of low  $p_T$  tracks with 100 MeV is required to measure the energy of hadronic jets.

Silicon sensors are used in the tracker due to their fine granularity and fast readout. Charged particles ionize silicon semiconductors, creating electron-hole pairs that

induce a current in the electrodes. The hits in the silicon layers determine particle trajectories, and therefore a high segmentation provides more precise spatial measurement.

Figure 2.10 shows the pixel and strip detector that compose the tracker substructure. The short black lines represent different modules that overlap each other noticeably to avoid gaps in acceptance. Starting from 3 cm from the IP, the pixel detector handles the densest particle flux. Nonetheless, the implementation of silicon pixel technology allows for outstanding spatiotemporal resolution in this harsh environment. The pixel layers are arranged in four cylindrical layers surrounding the beamline and three perpendicular disks, facilitating a 3-D particle reconstruction. Each pixel has a size of  $100 \times 150 \mu\text{m}^2$  and provides a spatial resolution ranging from  $10 \mu\text{m}$  to  $40 \mu\text{m}$  on a single hit.

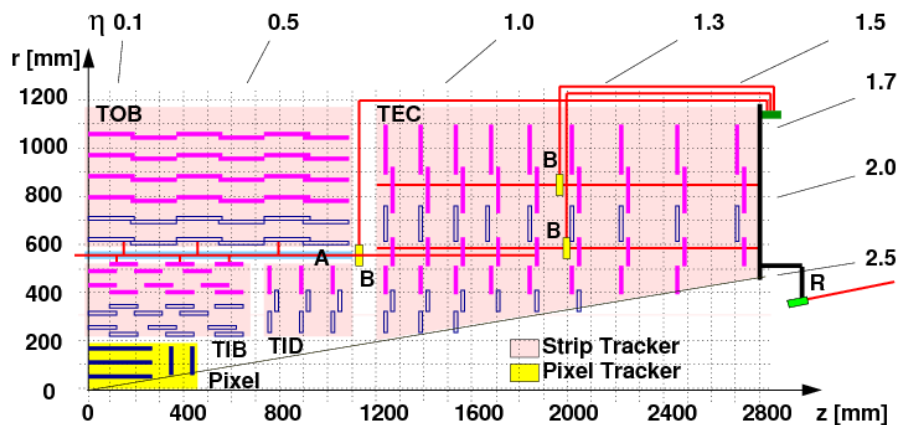


FIGURE 2.10: Shown in the  $rz$  view is one-quarter of the CMS silicon tracker, with the positions of individual modules represented by purple lines, while pairs of  $r\phi$  and stereo modules are represented by dark blue lines. Also visible are the laser rays, beam splitters (BS), and alignment tubes (AT) of the Laser Alignment System, as depicted in [71].

The pixel detector is crucial for vertex reconstruction and seed-tracking algorithms in CMS. Therefore, it is continually upgraded to accommodate higher luminosity physics programs as it suffers from radiation.

The strip tracker encases the pixel detector and comprises four parts: the tracker inner barrel (TIB), the tracker inner disks (TID), the tracker outer barrel (TOB), and the tracker endcaps (TEC). The system comprises  $\sim 9.3$  million silicon strips with 15 different geometries depending on the distance to the interaction point, ranging from  $80 \mu\text{m}$  and  $8.5 \text{ cm}$  in the inner regions to  $205 \mu\text{m}$  and  $20 \text{ cm}$  in the outer areas. The TIB covers the region  $20 < r < 55 \text{ cm}$ , and the TOB extends

up to  $r = 116$  cm, with their counterparts in the endcaps being the TID and TEC, respectively. The spatial resolution in the transverse plane is  $23\text{--}34\ \mu\text{m}$  for the TIB and  $35\text{--}52\ \mu\text{m}$  for the TOB, while it is about ten times larger in the longitudinal direction. The strip tracker has an overall hit efficiency of 99.8% for isolated muons. The double black lines in Fig. 2.10 indicate that one module is rotated relative to the other for 3D reconstructions.

The design of the tracker aims to achieve the most accurate detector performance while minimizing dead material that could lead to undesirable scatterings, *bremsstrahlung*, photon conversions, and nuclear interactions that could distort the measurements of the tracker and the calorimeters. To minimize radiation damage, the system is cooled to approximately  $-15\ ^\circ\text{C}$  using a  $\text{CO}_2$  cooling system. An upgrade in 2017 reduced the hadronic interaction length due to the passive material by about 40% in the endcaps and 10% in the barrel, improving the interaction point resolution by a factor of 1.5 in the  $z$ -direction [72, 73].

## Calorimeters

**Electromagnetic calorimeter** The tracking system is surrounded by the Electromagnetic CALorimeter (ECAL), which primarily serves to identify electrons and photons by accurately measuring their energy when an electromagnetic shower occurs, generating scintillating light. The ECAL array comprises 70,000 crystals made of lead tungstate ( $\text{PbWO}_4$ ) due to its high density, short radiation length, and small Molière radius, making it a compact and highly granular calorimeter of approximately  $25X_0$  long that guarantees the containment of full showers. This feature, coupled with its relatively quick response time of  $5\text{--}15$  ns, allows the ECAL to handle around 80% of the scintillating light emitted by particles produced in each LHC bunch crossing. Its cylindrical geometry is divided into three parts covering up to  $|\eta| = 3$ , namely the central region barrel (EB) covering  $|\eta| < 1.48$ , two endcaps (EE) between  $1.57 < |\eta| < 3$ , and a pre-shower detector in front of the endcaps with a coverage of  $1.65 < |\eta| < 2.6$ . The crystals in the barrel, of which there are 61,200, are 23 cm long and have a surface area of approximately  $22 \times 22\ \text{mm}^2$ , while the 7,324 crystals in both endcaps have a surface area of about  $28.6 \times 28.6\ \text{mm}^2$  and a length of 22 cm. To ensure there are no gaps in coverage and that they are airtight, the crystals are tilted at  $3^\circ$  relative to the interaction point in both the endcaps and the barrel. The crystals emit light, which is detected by avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs)

in the endcaps. Although the ECAL withstands moderate levels of radiation, it requires photodetectors with inner amplification due to its modest light yield of approximately 30 photons per MeV. A schematic of the ECAL layout is shown in Fig. 2.11. In front of the two endcaps, the pre-shower detector (ES) is lo-

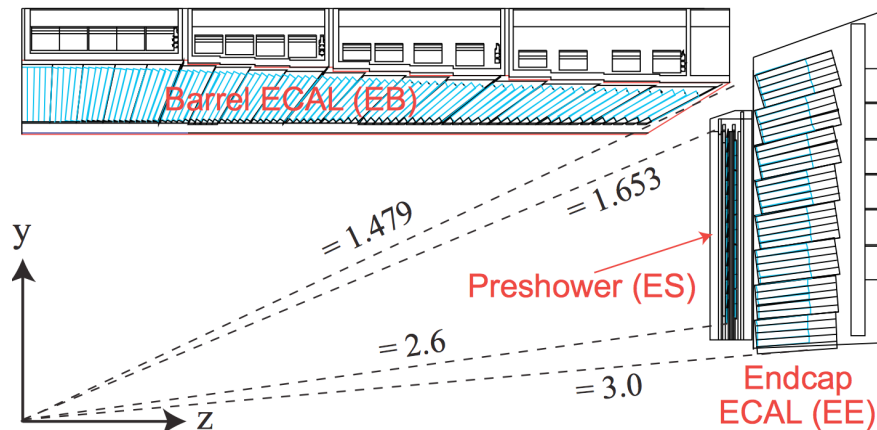


FIGURE 2.11: Schematic longitudinal layout of a quarter-section of the ECAL in the CMS detector, consisting of the barrel ECAL (EB), the endcap ECAL (EE), and the pre-shower detector (ES) [74].

cated, which is composed of a 2-layer sampling calorimeter with a lead absorber and 22 mm-long silicon strips. This detector plays a crucial role in distinguishing single high-energy photons from pairs of collimated low-energy photons produced by the in-flight decay of  $\pi^0$ . The intrinsic energy resolution of a homogeneous electromagnetic calorimeter is given by the following parameterization:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\% \text{ GeV}^{1/2}}{\sqrt{E}}\right)^2 + \left(\frac{12\% \text{ GeV}}{E}\right)^2 + (0.3\%)^2, \quad (2.9)$$

The precision of calorimeters at high energies ( $1/\sqrt{E}$ ) is highly accurate due to the first term, which represents stochastic fluctuations of the collected photoelectrons. The second term contributes to noise from the readout system (electronics), digitization chain, and pileup. The last term is responsible for non-uniformities in the detector response, miscalibrations, and energy leakage from the crystals. It is worth noting that the resolution gradually decreases over time due to the loss of crystal transparency caused by the large doses of radiation. Consequently, the ECAL requires periodic recalibration using time-dependent corrections.

**Hadronic calorimeter** The Hadronic CALorimeter (HCAL) measures approximately 70% of energy that remains in hadrons after passing through the ECAL. It is a crucial subdetector within CMS, as it measures the energy of neutral hadrons and undetectable particles.

The HCAL consists of a sampling detector with alternating layers of brass absorber (5 cm) and plastic scintillator tiles (3.7 mm thick). Hadronic-induced cascades are more challenging to measure than electromagnetic showers due to their larger fluctuations in the spatial development and energy loss and the fact that the nuclear interaction length is much larger than the electromagnetic radiation length. As a result, the HCAL is larger than the ECAL, but its size is limited by the volume constraints of the magnet.

The HCAL structure is divided into different regions, which are shown in Fig. 2.12, formed by several subdetectors located in different regions of the detector. The

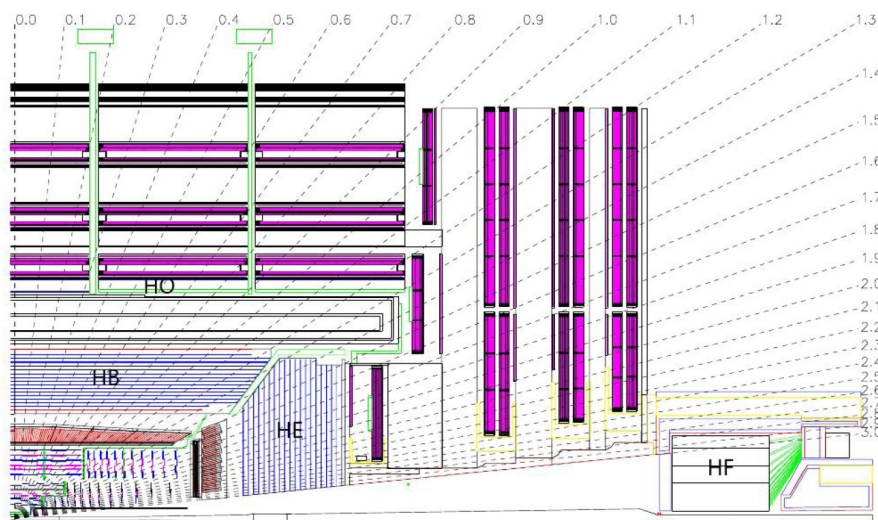


FIGURE 2.12: This figure displays a longitudinal view of the CMS quadrant, highlighting the positions and pseudorapidity ranges covered by the different hadron calorimeters: the barrel (HB), outer (HO), endcap (HE), and very forward (HF) [75].

inner region, with  $|\eta| < 1.4$ , is covered by the barrel calorimeter (HB), which has a length of approximately 7 radiation lengths ( $\lambda_i$ ). The endcap hadronic calorimeter (HE), with a thickness of around  $10\lambda_i$ , covers the region  $1.3 < |\eta| < 3.0$ . The outer hadronic calorimeter (HO) complements the barrel and endcap detectors, extending the interaction depth to approximately  $11\lambda_i$  in the central region. The HO, consisting of scintillating material, is located outside the magnet. The HF covers the very forward region, which is situated 11.2 meters away from

the interaction point, and extends the HCAL acceptance to the  $2.9 < |\eta| < 5.2$  range. The HF uses layers of steel absorber and Cherenkov-emitting quartz fibers for energy measurement.

The energy resolution is as follows:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{115\% \text{ GeV}^{1/2}}{\sqrt{E}}\right)^2 + (5.5\%)^2. \quad (2.10)$$

The HCAL performance is limited by the presence of non-Poissonian events in hadronic cascades, neutrinos, and electromagnetic components. However, this limitation is partially overcome by combining the information from HCAL with ECAL, resulting in a total energy resolution (for hadronic showers) expressed as follows:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{84.7\%}{\sqrt{E}}\right)^2 + (7.4\%)^2. \quad (2.11)$$

## Muon chambers

Muons yield a particular and clean signature after being produced in a large variety of physics processes. Due to their higher mass, muons lose less energy in their path through the detector than electrons and are therefore defined as “Minimum Ionizing particles” (MIP) in this context. This explains the location of the muon chambers in the outside part of CMS.

The CMS iron yokes encloses the muon system to immerse the chambers in a magnetic field of around 2 T. The charge of the muons may be inferred from the bending direction of their trajectory within the detector, and their transverse momentum is calculated by applying Eq. 2.1. These measurements complement those of the silicon tracker described previously.

Since it is the furthest subdetector from the interaction point, the layout of the muon system must cover a large surface area. Therefore, the choice of gas ionization chambers as the applied technology provides adequate performance at a reduced cost. There are a total of 1,400 chambers split into three gaseous detectors: Drift Tube chambers (DTs) covering the central region  $|\eta| < 1.2$ , Cathode Strip Chambers (CSCs) in the endcaps region  $0.9 < |\eta| < 2.4$ , and Resistive Plate Chambers (RPCs) up to  $|\eta| < 1.6$  for improving the trigger performance. These

pieces sum up a total of  $\sim 25,000 \text{ m}^2$  detecting area, and their arrangement is shown in Fig 2.13.

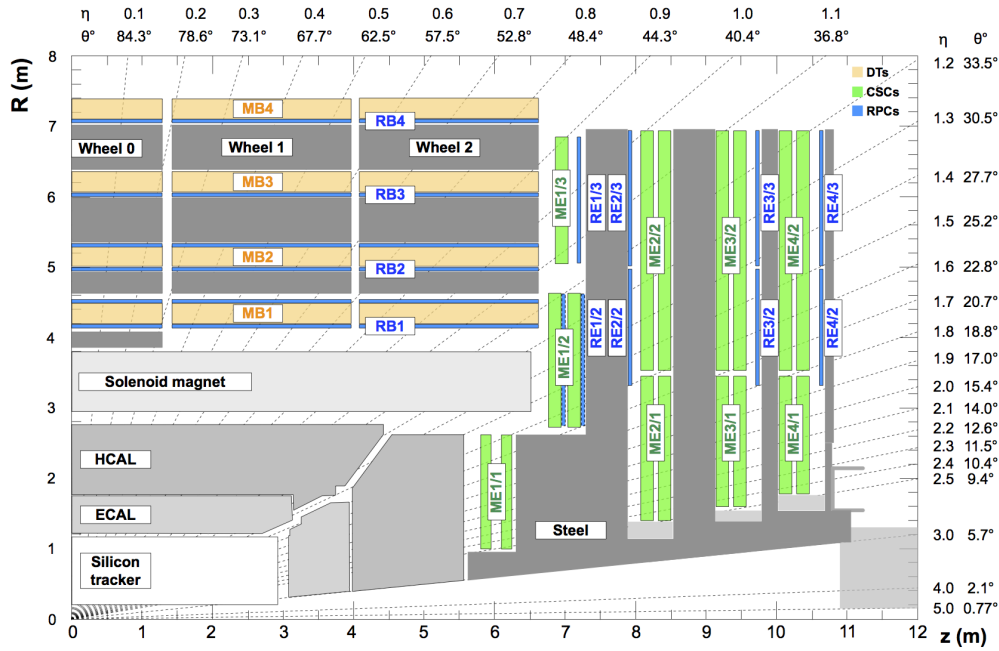


FIGURE 2.13: Shown in this  $R$ - $z$  cross-section of a quadrant of the CMS detector is the interaction point located at the lower left corner. The axis parallel to the beam ( $z$ ) runs horizontally, and the radius ( $R$ ) increases upward. The locations of the steel flux-return disks (dark areas) and various muon stations are illustrated. The drift tube stations (DTs) are labeled MB (“Muon Barrel”), and the cathode strip chambers (CSCs) are labeled ME (“Muon Endcap”). The resistive plate chambers (RPCs) are mounted in both the barrel and endcaps of CMS, where they are labeled RB and RE, respectively. [75].

The arrangement of detectors within the muon system relies on the expected background prediction rates and the magnetic field, fulfilling maximal coverage by spatial overlapping to maximize reconstruction efficiency. This is why the muon system is usually described as redundant.

In the overall functioning of the muon system, DTs are segmented into drift cells crossed by an anode wire. The position of muons is determined by measuring the drift time of an ionized electron cloud from the gas traveling to the anode wire. Arrays of positively-charged anode wires and negatively-charged copper cathode strips joined within a gas volume constitute the CSCs. These chambers determine the position of the muon when crossing the volume. RPCs use double-gap chambers to operate in avalanche mode. These chambers give information about the timing for the muon trigger complementing the DTs and CSCs.



Drift tube chambers are organized into four stations and divided into five wheels around the LHC beam pipe, providing coverage in the region  $|\eta| < 1.2$ . The cells composing the DTs are rectangular of size  $1.3 \times 4.2, \text{cm}^2$  and 2.5 m-long filled with a gas mixture of  $\text{CO}_2$  (15%) and Ar (85%). The cells are equipped with electrode strips on top and bottom, creating a constant field with a constant drift velocity of  $55 \mu\text{m/s}$ , and cathode strips on the sides. Experimentally, the drift time of a knocked-off electron is measured to reach the anode to determine muon crossing time, position, and incident angle in the chambers. The layers of the DTs are positioned to overlap each other with a half-cell offset, which effectively solves the left-right ambiguity with respect to the wire. The single-cell efficiency is 99.8%, and the spatial resolution per cell is  $\sim 180 \mu\text{m}$ , yielding an overall position resolution of  $80 - 120 \mu\text{m}$ .

The endcap regions ( $0.9 < |\eta| < 2.4$ ) that experience the highest background rates and a stronger magnetic field are equipped with CSC detectors. The trapezoidal design alternates layers of anode wires and cathode strips filled with a gas mixture of Ar (45%),  $\text{CO}_2$  (50%), and  $\text{CF}_4$  (10%). As the DTs detectors, the CSC detectors measure muon position by ionization of the gas with a spatial resolution of  $40 - 150 \mu\text{m}$ . Fast RPC detectors are located in the barrel and endcaps regions ( $|\eta| < 1.9$ ). They support DTs and CSCs with a standalone triggering system resolving tracking ambiguities in events with multiple hits within a single cell. This subsystem operates in avalanche mode, making use of two gaps of resistive 2 mm-thick layers filled with a mixture of  $\text{C}_2\text{H}_2\text{F}_4$  (96.2%),  $\text{i-C}_4\text{H}_{10}$  (3.5%) and  $\text{SF}_6$  (0.3%). When a muon crosses the RPCs, it triggers an electron shower inside the volume read out by the strips. The spatial resolution is relatively poor ( $\sim 1 \text{ cm}$ ); however, these detectors are much faster than DTs and CSCs with a time precision of 3 ns, helpful to match a bunch crossing with a muon track even in a high luminosity environment.

In 2017, Gas Electron Multiplier chambers (GEMs) were installed to extend the coverage of the muon system to the  $1.6 < |\eta| < 2.2$  region. GEMs add redundancy to the endcaps and have a slim profile, providing exceptional rate capability in high particle fluxes. Following LS2, 144 GEM detectors were incorporated into the CMS muon system.

### 2.2.3 Trigger and DAQ

Within the CMS physics program, both BSM searches and precision studies of the SM only require a small fraction of the effective pp collisions, despite the high luminosity provided by the LHC. Ideally, it would be ideal to store all information from every pp collision, but physical limitations restrict data storage capacity. As a result, it becomes necessary to strictly select the most interesting events by applying kinematic limits that exclude certain regions from consideration. These cuts are guided by the accumulated knowledge and experience in HEP. Most of the LHC experiments adopt this strategy, along with other special data-taking tactics mentioned later. Exceptionally, the LHCb experiment, due to its reduced size and because it runs at a lower luminosity than ATLAS and CMS, performs the full storage of all effective pp collision information. In experiments like ATLAS and CMS, the data corresponding to a single event occupies approximately 2 MB of disk space, whereas, for LHCb, it only requires around 30 kB. Considering the LHC p-p interaction rate of 40 MHz, permanent data storage poses one of the most significant challenges for LHC experiments. To tackle this issue, the trigger and data acquisition systems (TriDAS) in CMS reduce the rate of stored collisions from MHz to Hz on the fly, making the data more manageable to store and process for offline analysis.

SM production cross-sections of various physics processes measured by CMS are shown in Fig. 2.14. These cross-sections are at least six orders of magnitude smaller than the total pp cross-section of  $\sigma_{pp} = 10^{11}$  pb. The Trigger and DAQ systems play a crucial role in enabling online data-taking and offline analysis. The Trigger System filters events and works in two steps: the Level 1 Trigger (L1 Trigger), which is hardware-based, and the High-Level Trigger (HLT), which is software-based.

#### Level 1 trigger system

The L1 trigger system reduces the readout rate from the 40 MHz bunch-crossing frequency to a maximum of 100 kHz within a time interval (latency) of fewer than  $3.4 \mu\text{s}$ , limited by the tracker readout. It consists of hardware processors that directly access the information of the several CMS subdetectors. The L1 trigger uses FPGAs (Field Programmable Gate Arrays) and ASICs (Application Specific Integrated Circuits) to achieve a short decision time. New-generation FPGAs



## DAQ

The CMS DAQ is responsible for retrieving and assembling events that have been accepted by the initial L1 hardware trigger. Once assembled, these events are made accessible to the high-level trigger, which selects interesting events for offline storage and analysis. The system is specifically designed to handle a maximum input rate of 100 kHz and an aggregated throughput of 100 GB/s, originating from approximately 500 sources [77]. The whole process involving the dedicated triggers of each subdetector is depicted in Fig. 2.15.

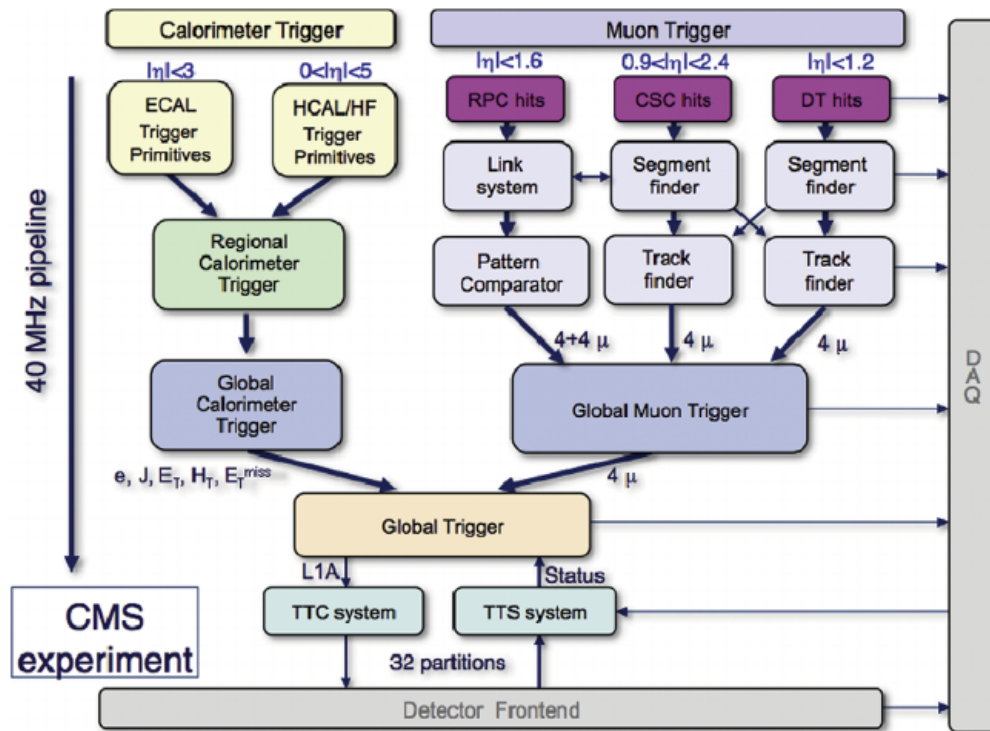


FIGURE 2.15: CMS L1 Trigger. Taken from [78]

## HLT Trigger

The HLT performs a more sophisticated selection of events that have passed the L1A event selection, reducing the information rate from 100 kHz to 2-3 kHz. HLT uses the full detector capability to perform online reconstruction of physics objects and classifies them into a Primary Dataset (PD). It comprises over 600 trigger physics signatures on subsequent steps of increasing complexity to accept the interesting events, which afterward go through a detailed reconstruction that uses information from the tracker to match objects and assemble the final trigger decision.

During Run 2, two techniques were introduced to increase the phase space accessible for physics studies. The first technique, “data scouting,” stores the HLT information of the events and is not used in precision measurements. Instead, it saves a larger number of events that would contain low-energy processes of interest for exotic physics searches such as Dark-Matter WIMPs. The second technique, called “parking”, directly streams events to be recorded on tape without any reconstruction (raw data) and is limited by the DAQ and disk storage. These stored events are kept for potential BSM physics deviations or sensitivity improvements.

## Data Storage

Finally, Tier-0 at CERN gathers the events that pass the trigger criteria for offline reconstruction and analysis. Tier-0 is responsible for data preservation, repacking, express processing, and prompt reconstruction of the interesting events acquired by the previous systems.

## 2.3 Objects reconstruction

After the trigger system selects and events are stored, those must undergo processing to reconstruct their content. To accomplish this, the CMS subdetectors gather raw information such as particles, clusters of energy deposits, and tracks. The Particle Flow (PF) algorithm utilizes this information to create a global description of the physics objects present in the event to be analyzed.

### 2.3.1 Particle Flow algorithm

The PF algorithm relies on an accurate tracking system and highly-granulated calorimeters to perform its function. Tracks are matched with a vertex, whose total associated  $p_T$  determines whether it is the main collision vertex or a pile-up vertex. Muons traverse the entire detector and are reconstructed by matching tracks with the information in the muon chambers. For electron identification, the algorithm considers *bremsstrahlung* effects and links tracks to clusters. Clusters with an associated track are identified as charged hadrons, while those without a track are classified as photons if they have electromagnetic deposits or neutral hadrons if

they have hadronic deposits. Figure 2.16 shows a schematic representation of the trajectories of the particles crossing CMS. As shown there, the charged particles are bent due to the magnetic field. Secondary vertices are used for tagging. After identifying an object, the PF reconstruction algorithm processes the related tracks and clusters in the following order.

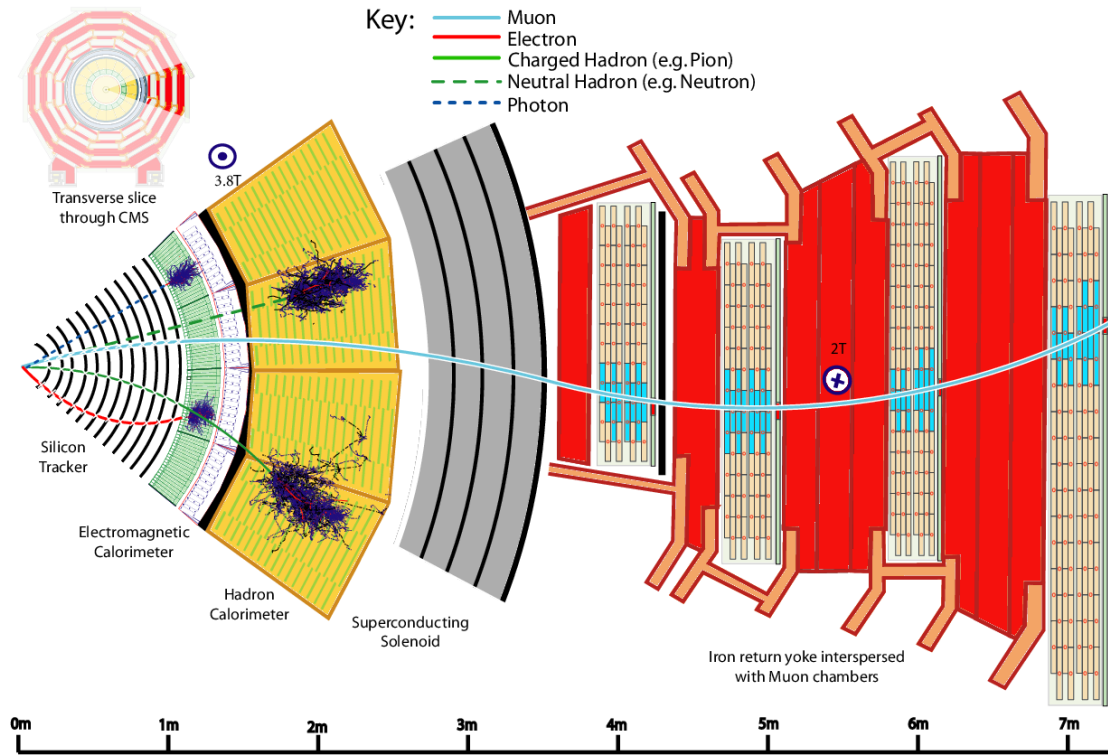


FIGURE 2.16: According to [79], particle interactions are depicted in a transverse slice of the CMS detector that extends from the beam interaction region to the muon detector. The muon and charged pion exhibit positive charges, whereas the electron displays a negative charge.

## Muons

In the CMS detector, muons have a clear and distinct signature, making them easy to identify with high efficiency. The reconstruction of muons involves combining the tracks independently reconstructed by the tracker and muon system, resulting in global muons when a tracker muon matches a standalone muon from inside-out or outside-in matching algorithms. The inside-out method involves extrapolating each tracker track with  $p_T > 0.5$  GeV and total  $p > 2.5$  GeV to the muon system and matching it with a standalone muon, while the outside-in technique extrapolates standalone muon tracks inward to match a compatible tracker

track. Both methods result in unique muon candidates, and together they achieve a reconstruction efficiency of approximately 99% for isolated muons. Normally, muons must be reconstructed as both tracker muons and global muons. The CMS experiment criteria for muon selection based on track fit quality  $\chi^2$ , the number of hits per track, and the tracker and standalone track matching degree. These requirements classify muons into different working points based on increasing purity and decreasing efficiency: loose, medium, and tight. Isolation is another property associated with muons, which refers to the activity surrounding the object. The PF relative isolation is determined by measuring the amount of energy within a cone of size  $\Delta R < 0.4$  in the  $(\eta, \phi)$  plane centered at the muon track starting from the muon production vertex. Isolation distinguishes between prompt muons and muons resulting from hadronic decays within jets. The six increasingly stringent muon isolation working points are defined by decreasing cutoffs in isolation. The Tag-and-Probe method is used to study the efficiency of muons, beginning with tracker tracks as probes. The muon identification procedure is regularly reviewed and updated to the latest version of the analysis data sets. The reconstructed hits have a spatial resolution ranging from 50 to 300  $\mu\text{m}$ , leading to a muon selection efficiency above 96% [80]. Figure 2.17 displays the efficiency plotted against  $p_T$  and  $\eta$ , while utilizing the strictest PF identification and isolation criteria.

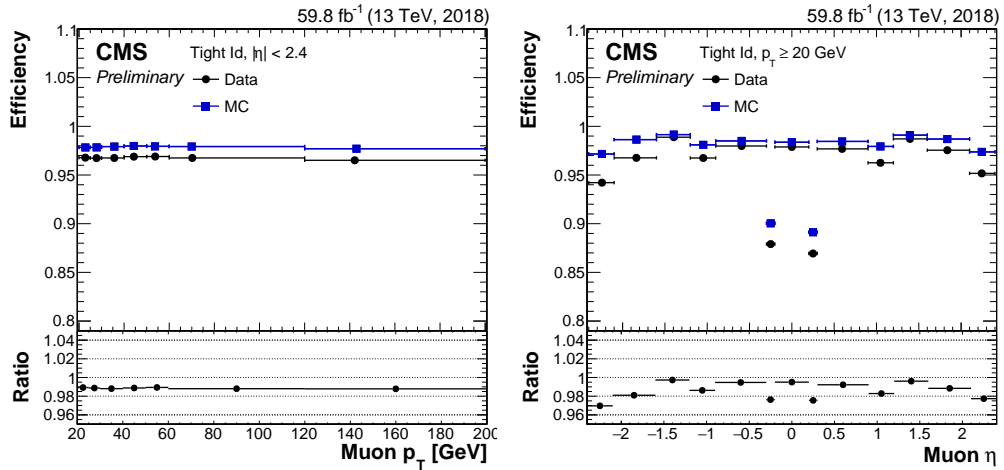


FIGURE 2.17: Tight lepton identification efficiency as a function of  $p_T$  (left) and  $\eta$  (right), using 2018 data and MC, where the denominator corresponds to tracker muons with  $p_T > 20$  GeV. No significant variation with respect to the number of primary vertices was observed, and the average scale factor was found to be 0.98. The systematic uncertainties are smaller than  $< 0.5\%$  [80].

## Electrons and photons

The electron reconstruction algorithm combines energy deposits in the ECAL with compatible charged tracks in the silicon tracker, similar to muon reconstruction. However, due to the phenomenon of *bremsstrahlung*, electrons can lose more energy and undergo more potential energy losses than muons. The amount of energy lost depends on the amount of material the electron passes through, which varies with the pseudorapidity  $\eta$ . At  $\eta \approx 0$ , the average energy loss is about 33%, while at  $\eta \approx 1.4$ , it is around 86%. Approximately half of the electrons lose 50% of their initial energy before reaching the ECAL, and about 10% lose over 95% of their energy. The amount of lost energy can be useful in distinguishing electrons from charged hadrons since the latter deposit less energy in the ECAL.

The electron reconstruction procedure involves identifying local seed clusters aggregated to energy deposits associated with *bremsstrahlung* photons. The tracker accounts for energy loss, and the clustering algorithm must collect the scattered photon energy, sometimes far from the entry point of the electron in the ECAL. Due to the magnetic field, charged particles tend to be more spread out in the  $\phi$  direction than in the  $\eta$  direction, which is toward the z-axis, as shown in Fig. 2.18. Electrons with  $p_T$  between 10-50 GeV have an energy resolution of 0.1% (0.3%) in the barrel (endcaps).

Additional identification criteria are applied using Boosted Decision Trees (BDTs) to distinguish electron candidates from pions or misidentified jets. These criteria include variables related to shower shape, track quality, track-cluster matching, isolation, and energy loss due to *bremsstrahlung*. The output of the BDT provides the probability of a candidate being a genuine electron and determines three levels of electron identification: loose, medium, and tight. These levels decrease in selection efficiency but increase in purity. The identification and isolation efficiencies for the jets faking electrons are also provided as part of the performance of the discriminator, as shown in Fig. 2.19. ECAL clusters that do not have an associated track or a matching cluster in the HCAL are considered genuine and are referred to as photons. These objects have limited interest in the analysis presented in this thesis. Ref [82] provides further information on their reconstruction and identification.



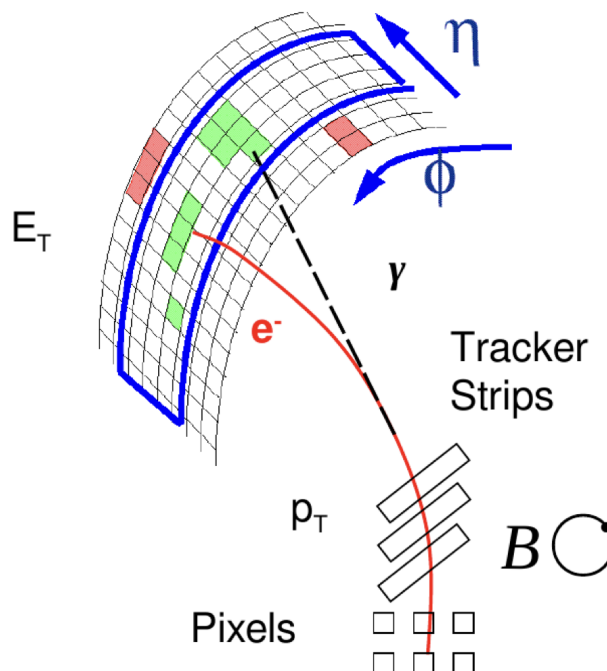


FIGURE 2.18: Display of the trajectory of an electron passing through the CMS detector. The electron interacts with the tracker material, leading to energy loss caused by *bremsstrahlung*, which emits photons in the  $\phi$  direction along the magnetic field lines. The energy deposits of both the electron and the photons in the ECAL are grouped together in superclusters and used by the GSF electron reconstruction algorithm. These superclusters are depicted by light green squares in the figure.

## Hadron reconstruction and $\tau$ leptons

After removing identified muons, electrons, and photons from the PF algorithm, the remaining tracks and energy deposits are used to identify charged and neutral hadrons. The  $\tau$  lepton, which has a mass of 1777 MeV and a relatively short lifetime of  $2.9 \times 10^{-13}$  s, decays into hadrons along with a neutrino or into a muon or an electron together with two neutrinos. Since the  $\tau$  lepton with 30 GeV may travel  $\sim 1$  mm before decaying, only its visible decay products are observed in the detector. In the cases where the  $\tau$  lepton decays into muons or electrons, the previously described methods are used for reconstruction. However, for hadronic tau decays, the hadrons-plus-strips (HPS) method is applied, where the seed is the PF jet constituents reconstructed using the anti- $k_T$  algorithm, explained in the next section. The  $\tau$  decay products are identified via the PF hadron components. The studies presented in this thesis are not focused on  $\tau$  leptons, and thus we do

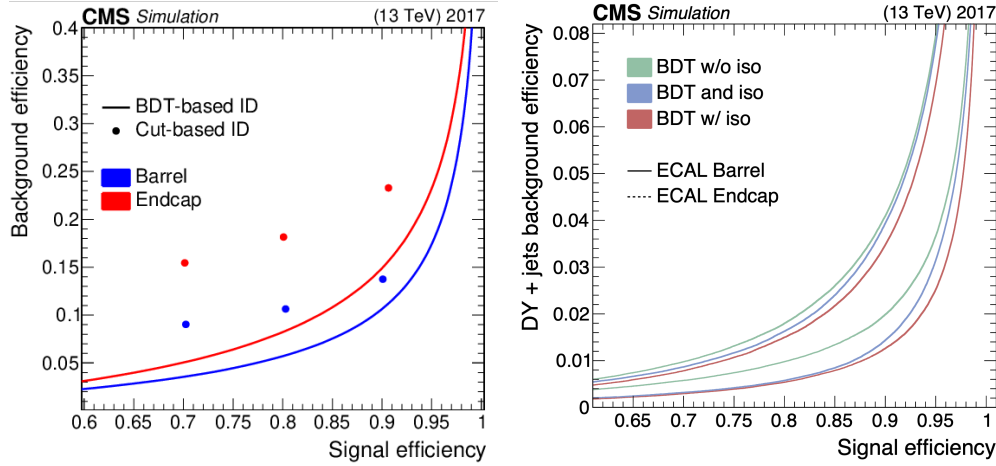


FIGURE 2.19: Performance of cut-based and BDT identification algorithms for electrons and photons in 2017 (left). The cut-based identification is presented for three different working points: tight, medium, and loose, from left to right, respectively. Performance of the BDT isolation algorithms for electrons and photons in the same year is shown (right). [81].

not describe more details on their reconstruction and identification that can be found in Ref. [82].

## Jets

In pp collisions, quarks and gluons are produced through hard interactions, which then hadronize in the detector, creating collimated groups of particles in cone-shaped jets that are reconstructed by clustering PF objects. Once electrons, muons, and photons have been identified, the remaining charged or neutral hadronic particles are used to estimate the momenta of the jets. This analysis uses the anti- $k_T$  algorithm [83] to determine jets. This algorithm recursively pairs PF candidates by defining three quantities within a set of  $N$  objects:

$$\begin{aligned}
 d_{i,B} &\equiv p_{T,i}^{-2}, \\
 d_{i,j} &\equiv \min \{p_{T,i}^{-2}, p_{T,j}^{-2}\} \left( \frac{\Delta_{ij}^2}{R^2} \right), \\
 \Delta_{ij}^2 &\equiv (y_i - y_j)^2 + (\phi_i - \phi_j)^2.
 \end{aligned} \tag{2.12}$$

The anti- $k_T$  algorithm calculates two distances for each particle and particle pair: the first distance measures the distance between the object  $i$  and the beam, while the second measures the distance between objects  $i$  and  $j$ . The distance in the

$(y, \phi)$  plane is represented by  $\Delta_{ij}^2$ , and the usual cone radius parameter,  $R$ , is used by the algorithm to cluster particles. For resolved jets, the value of  $R$  is set to 0.4 and thus referred to as AK4-jets, while for boosted-objects topologies, it is set to 0.8 and thus called AK8-jets. In both cases, the algorithm clusters higher- $p_T$  pairs first, and the cone evolves around the most energetic particle in the event, leaving the softer radiation and collinear partons to split outside the borders of the cone.

Once the distances are calculated, the algorithm merges objects  $i$  and  $j$  into a single object with 4-momenta defined as the sum of both. This process continues until the distance  $d_{i,B}$  is smaller, and the object  $i$  is tagged as a jet and removed from the input particle candidates. The algorithm stops when there are no particles left to cluster.

However, the reconstructed energy of the jet may differ from the energy of the parton that produced it due to theoretical uncertainties, non-linear detector response, and pile-up effects.  $p_T$ - and  $\eta$ -dependent jet corrections are applied to account for these discrepancies. As shown in Fig. 2.20, the typical reconstructed jet energy is about 90% of the parton energy, and its resolution is less than 15% for all  $p_T$  values when using PF candidates, which supersedes calorimeter-based reconstruction.

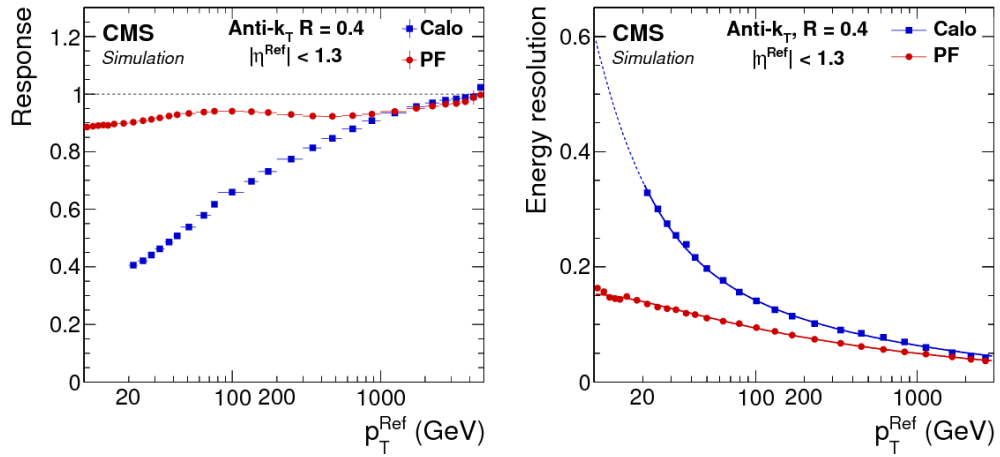


FIGURE 2.20: Jet energy response as a function of the  $p_T$  in the barrel [79].

Jets are categorized into two levels of selection: loose and tight. The criteria for selection include the fraction of charged and neutral hadrons containment, charged hadron diversity, and the amount of ECAL deposits. Jets originating from bottom quarks are of special interest. They are identified to be the signature of B hadron

decays, which are long-lived particles decaying at a secondary vertex with a relatively high particle multiplicity and an electron or muon inside the jet in 40% of cases. Machine Learning techniques are employed to analyze these features, resulting in two algorithms developed in CMS: DeepCSV and DeepJet [84, 85]. DeepJet is a more complex multi-classification DNN (Deep Neural Network) algorithm than DeepCSV, which uses several charged and low-level properties of neutral PF jet constituents, supplemented with secondary vertices properties (using the IVF [86] algorithm) associated with the jet. Typically, three working points are defined, corresponding to 10%, 1%, and 0.1% misidentification probabilities of light jets. Fig. 2.21 demonstrates the b-jet identification efficiency as a function of the  $p_T$  for both data and simulated muon jets.

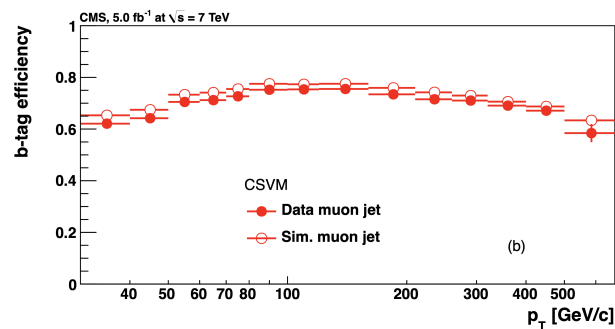


FIGURE 2.21: Measured efficiencies for the identification of b-jets in a muon jet sample using the CSV tagger. The filled circles correspond to data, while the open circles represent simulation results [87].

### 2.3.2 Missing transverse energy

Particles that cannot be detected, such as neutrinos, dark matter candidates, or BSM weakly-interacting particles, are inferred from momentum conservation and considering that the initial transverse momentum carried by the incoming partons is expected to be practically zero. The negative sum of the transverse momentum of all PF candidates in an event gives the missing transverse momentum:

$$\vec{p}_T^{miss} = - \sum_{PF} \vec{p}_T. \quad (2.13)$$

The precise measurement of this quantity requires the CMS detector to cover almost the entire solid angle to count as many particles as possible. However, inefficiencies in the tracking or clustering algorithms and non-linear responses in the calorimeters may introduce biases in  $\vec{p}_T^{miss}$ , resulting in a relative  $p_T^{miss}$  resolution of about 20% for the energies considered in this analysis (see Fig. 2.22). The relative energy resolution of the missing transverse energy is below 20-30% for all  $p_T^{miss}$  ranges.

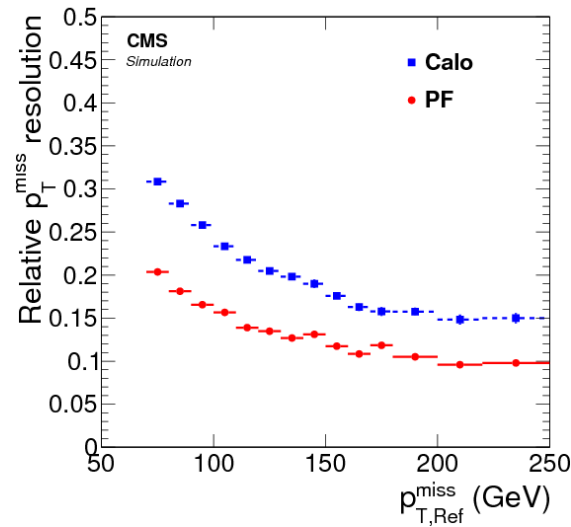


FIGURE 2.22: Dependence of the relative  $p_T^{miss}$  resolution on the  $p_{T,Ref}^{miss}$  for a simulated sample of top quark-antiquark events [79].



## Chapter 3

# CMS Muon Drift Tubes upgrade for the HL-LHC

As mentioned in the introduction of the thesis, the readout and trigger electronics of the CMS subdetectors have to be upgraded to withstand radiation levels and rates beyond their initial design parameters for future operation at the HL-LHC. In this regard, the Technology and Basic Research departments at CIEMAT collaborate to coordinate the upgrade program of the CMS Drift Tubes (DT) system. They share responsibilities in the Slice Test (ST) project developed during the CMS Phase-2 upgrade [88] aiming to test new DT electronics prototypes designed to handle the increased beam luminosity range of  $5 \cdot 10^{34}$  to  $7.5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Since LS2, rigorous testing has been conducted on the newly developed Phase-2 readout electronics. These electronics are anticipated to be fully installed during LS3, ensuring their complete operational readiness for the upcoming HL-LHC [89].

The main objective of the future upgrade is to maintain or improve the performance of the CMS in terms of muon trigger and identification. This critical goal is thoroughly investigated through the ST project, which involves precise calibration of the measured signals, primarily muons passing through the chambers. The calibration aims to achieve optimal time resolution of the DT Local Trigger, which directly impacts muon reconstruction. In this chapter, we present the studies of the performance of the Phase-2 system, specifically designed for the HL-LHC. The results are compared to the Legacy Phase-1 readout, which was operational until the LS3. Section 3.4 summarizes the results obtained for the thesis and discuss the calibration test conducted on the Phase-2 electronic signals. These tests are performed using special CMS runs in which cosmic ray data are recorded.

### 3.1 Segment reconstruction with the DT system

Based on the information regarding muon chambers described in section 2.2.2, the segment reconstruction process within the DT system involves the combination of signals from the wires within the same station that have been triggered by a muon passing through.

The DT system comprises 250 chambers that vary in size, ranging from approximately  $2\text{ m} \times 2.5\text{ m}$  to  $4\text{ m} \times 2.5\text{ m}$ . Each DT chamber is composed of 8 or 12 detection layers, with each layer consisting of up to 92 tubes. These tubes are organized into two or three “superlayers” (SL), each containing four layers. Figure 3.1 provides a visual representation of the internal structure of a DT chamber specifically designed for detecting muon tracks in the barrel region. The chamber utilizes three stacked SL, with each stack containing 4 cm wide tubes. These tubes incorporate a  $50\text{ }\mu\text{m}$  thick stretched wire at their center and contain a gas mixture of Ar and  $\text{CO}_2$ . When charged particles pass through the chamber, ionization electrons drift towards the wire, resulting in a measurable charge pulse at the point of avalanche on the wire.

In each chamber, the middle superlayer is responsible for measuring the coordinate parallel to the beam direction, while the two external superlayers measure the perpendicular coordinates. This arrangement enables a three-dimensional measurement of the muon track. Each superlayer identifies track patterns and determines the crossing time of the track by measuring the drift time of ionization electrons in each layer. This process yields four two-dimensional points, which are used to calculate the crossing position of the muon (referred to as “hits”). The ability of the system to accurately measure the timing of muon passage provides valuable inputs to the L1 Muon trigger. The recorded hits, represented by two-dimensional coordinates, undergo a preprocessing step to eliminate noise. Subsequently, a combinatorial algorithm is employed to cluster the hits that are likely to originate from the same muon. Following the clustering process, a pattern recognition algorithm is applied to identify straight-line segments that align with the anticipated trajectory of a muon traversing the detector. Linear regression is applied to these segments to determine their intercept and slope with respect to the chamber, enabling estimation of the position and direction of the muon track. Finally, the reconstructed segments from all four stations of the DT system are combined with segments from CSC and RPC detectors to form a muon candidate. This candidate is then forwarded to the global event reconstruction stage, where the



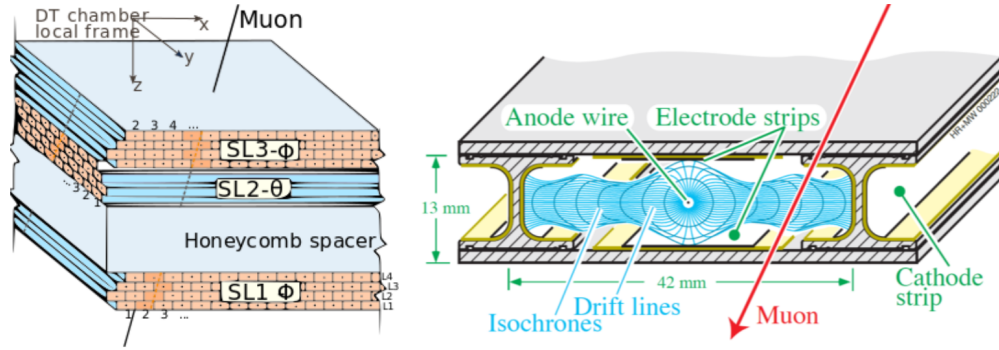


FIGURE 3.1: Left: Configuration of the DT system layers, comprising three superlayers (SL1, SL2, SL3). Right: Illustration of a muon passing through a single drift cell.

complete trajectory of the muon is identified and reconstructed. Further details can be found in section 2.2.

In the same fashion as the current offline reconstruction performed on conventional computers, the DT Phase-2 system aims to accomplish muon trajectory reconstruction tasks with a similar level of precision. However, the Phase-2 system aims to achieve this precision within the constraints of the L1 trigger, with a latency of  $[o]$  ( $1 \mu s$ ). To achieve this, dedicated programmable devices known as FPGAs are utilized in the backend electronics, running similar reconstruction algorithms. In the ST, the Analytical Method (AM) algorithm [90] was employed. The AM algorithm is specifically designed for FPGA architecture, allowing for the generation of trigger primitives from the DT detector. It focuses on reconstructing muon segments based on individual wire DT hits, which exhibit a time spread of 400 ns within a given BX (bunch crossing) due to drift time in the cell. By implementing the AM algorithm, the DT chambers can achieve the highest resolution possible, thereby enhancing the hardware system capabilities to approach the performance of offline processing.

The central objective of the hardware-dedicated analysis in this thesis is to precisely calibrate the relative timing measurements of the wires. This calibration process holds great importance in guaranteeing the reliability and accuracy of the hit reconstruction, as any inconsistencies in the wiring could potentially introduce errors in the measurements. With utmost care, we have employed dedicated techniques and procedures to calibrate all the wires meticulously. Our aim is to ensure that the calibration is both precise and consistent across all wires, thereby establishing a solid foundation for our measurements.

### 3.1.1 Hit efficiency definition

To determine the detection of a single hit within a cell, we calculate the ratio between the number of detected hits and the expected hits. The estimation of expected hits requires the use of well-reconstructed track segments that had associated hits in at least four layers in SL3 and at least one layer in SL1, with the exception of the layer being tested. The expected hit position was determined by identifying the intersection of the track segment with the layer under examination, indicating the cell where a hit was anticipated. The fraction representing the frequency of hits actually found in the cell provides crucial information about the efficiency of the specific cell under investigation. For a visual representation, please refer to the diagram in Fig. 3.2, which shows a DT chamber and a reconstructed track segment with four associated hits in SL3, extrapolated to SL1 [91].

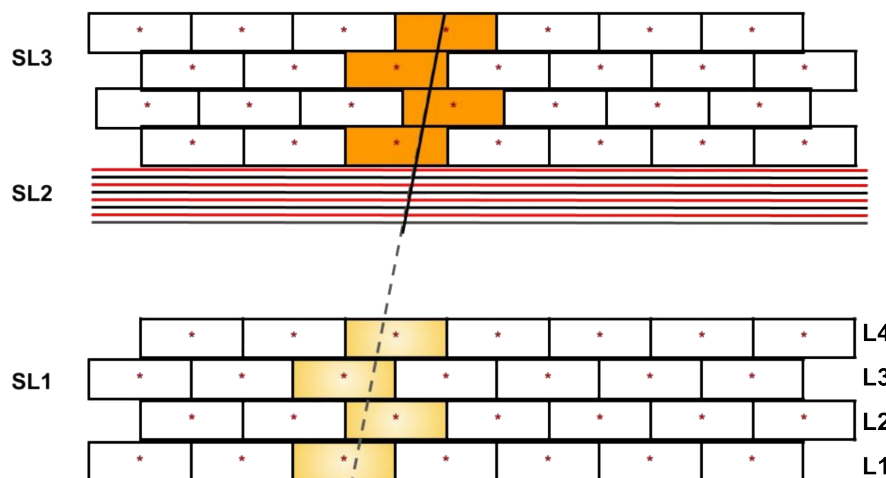


FIGURE 3.2: DT chamber with a reconstructed track segment showing four associated hits in SL3, extrapolated to SL1 (SL2 not shown to scale).

## 3.2 The Slice Test setup and goals

As described in Section 2.2, the DT detectors play a crucial role in both triggering and offline tracking within the CMS muon barrel system. For the ST project, a specific portion of the DT detector was equipped with Phase-2 on-board DT electronics (OBDT), operating in parallel with the Phase-1 (Legacy) electronics. This setup included four distinct DT chambers, namely MB1, MB2, MB3, and

MB4 Minicrate Boards, located in Sector 12 of Wheel +2 [75]. Figure 3.3 provides a diagram illustrating the two readout systems under consideration, with further details about their component elements outlined below. The Phase-1 system is visually represented by the color green, while the ST, serving as a demonstrator for the Phase-2 readout electronics, is depicted in blue. The primary objective of the

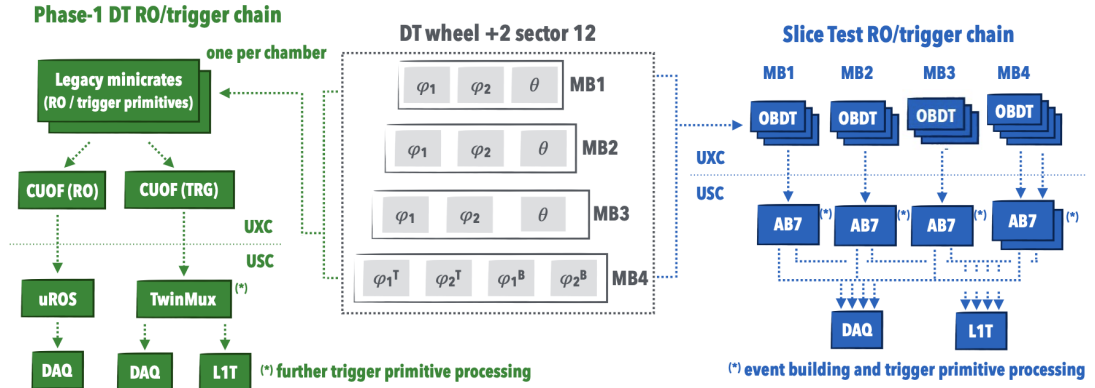


FIGURE 3.3: A schematic comparison depicting the current (Phase-1) DT readout and trigger electronics (shown in green) alongside the architecture of the DT Slice Test for Phase-2 (depicted in blue).

DT ST is to showcase improved trigger reconstruction performance comparable to that utilized in Phase-1, as well as to calibrate the Phase-2 electronics for enhanced and more accurate reconstruction of muon signals passing through the chambers. The calibration process specifically focuses on the anode wires or channels, which serve as fundamental components of the DT cells. It involves inter-channel equalization within each DT chamber, accomplished through dedicated Test-Pulse runs that inject artificial pulses simultaneously into all wires.

The analysis presented in this chapter demonstrates the stability of the Phase-2 system over time, with a maximum average difference between various data-taking runs of 0.5 ns. The developed calibration procedure, outlined below, represents a significant accomplishment of this work and serves as a valuable monitoring tool for both current and future commissioning and calibration of the DT system.

### 3.2.1 Phase-2 Slice Test On-Board DT electronics

The development of the new On-Board DT (OBDT) electronics for the CMS muon barrel system was undertaken by the CIEMAT group. The primary objective was to improve the streaming of high-resolution time-to-digital converter (TDC)

counts for the efficient identification of ionizing particle crossings. These counts are transmitted to a new back-end system located in the CMS service cavern via optical fibers. The new back-end system requires radiation-resistant electronics capable of withstanding the anticipated radiation levels in the HL-LHC.

Each OBDT samples up to 240 channels and transmits the TDC hits to the AB7 back-end electronics, which is based on the TM7 board used for triggering and readout. For each Minicrate (MB1 with 3 OBDTs, MB2 with 3 OBDTs, MB3 with 2 OBDTs, and MB4 with 4 OBDTs), up to five AB7 boards are connected. The configuration of these components is shown in Fig. 3.3.

Achieving precise synchronization among all channels is essential for the optimal performance of the trigger system. The generation of trigger primitives, which involves identifying potential tracks based on hit patterns in the chambers, relies on accurate drift time information derived from TDC measurements. Therefore, precise calibration is crucial. A trigger primitive serves as a fundamental data object that represents a potential track and contains important information such as the track position, direction, momentum, as well as the timing and quality of the associated hits. These trigger primitives are used as input to the HLT algorithms, which make decisions about event retention or discarded based on the significance of the event in terms of physics.

To achieve the optimal local trigger resolution in Phase-2, performing online calibration with a precision of approximately 1 ns is crucial, emphasizing the importance of inter-channel equalization. This calibration process is carried out within each chamber to determine individual offsets ( $t_{0i}$ ) for each channel, specifically the anode wires. The Front End Boards (FEB) located inside the gas volume are responsible for transmitting the signals produced by the electrons reaching the anode wires to the DT readout and trigger electronics. The subsequent sections provide a comprehensive overview of the results obtained from the ST, focusing on the performance evaluation of the Phase-2 electronics.

### 3.3 Phase-2 vs Phase-1 readout systems

Before discussing the calibration of the Phase-2 system, it is necessary to evaluate the hit detection performance of both the Phase-1 and Phase-2 readout systems

and compare their results. In order to do so, the recorded signal time measurements from the DT readout systems are used to reconstruct the momentum of charged particles originating from cosmic rays that pass through the chambers.

The initial step involves assessing the relative efficiency of hit detection between the Phase-1 and Phase-2 readout systems. Figure 3.4 demonstrates that hits detected by one system were also detected by the other, indicating excellent agreement between them. To conduct this comparison, only hits within a time window of 900 ns centered around the Phase-1 trigger time were taken into consideration. It is important to note that no calibration was applied during the calculation of the relative efficiencies.

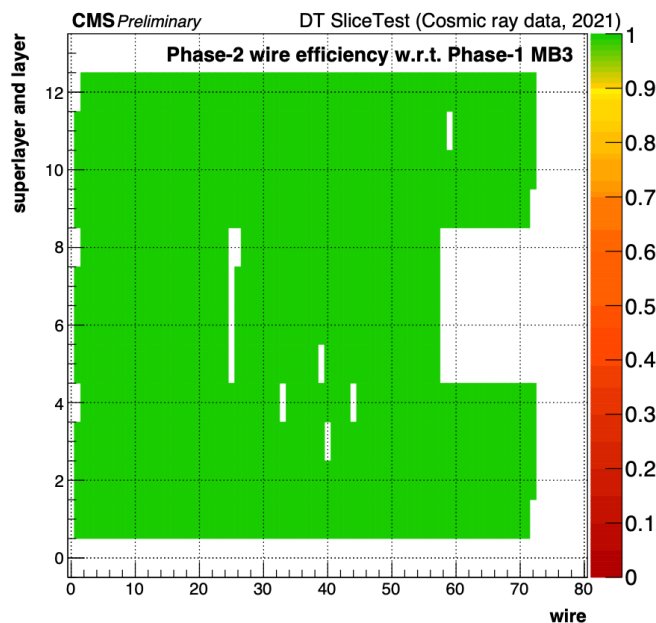


FIGURE 3.4: Comparison of hit detection efficiency between the Phase-1 and Phase-2 readout systems, demonstrating strong consistency in detecting hits, with the exception of a small number of masked wires. Adapted from [92].

The local offline segment reconstruction of cosmic rays for the MB4 station in the DT ST is shown in Fig. 3.5, using both the Phase-1 DT readout and Phase-2 prototypes with calibrated time measurements. The results exhibit a remarkable agreement between the two systems, with any slight differences attributed to the inherent precision of the calibration process or the slight variation in TDC sampling frequency. The Phase-1 TDCs have 32 counts every 25 ns, while the new system has 30 counts every 25 ns. DT segments enable the precise measurement of muon arrival times from LHC collisions with an accuracy of approximately 2-3 ns [75].

In contrast to muons produced in collisions, cosmic ray muons, as depicted in Fig. 3.5, uniformly traverse CMS over time without bunching. The distribution exhibits a square function with a width of 25 ns, smoothed by the Phase-1 local trigger.

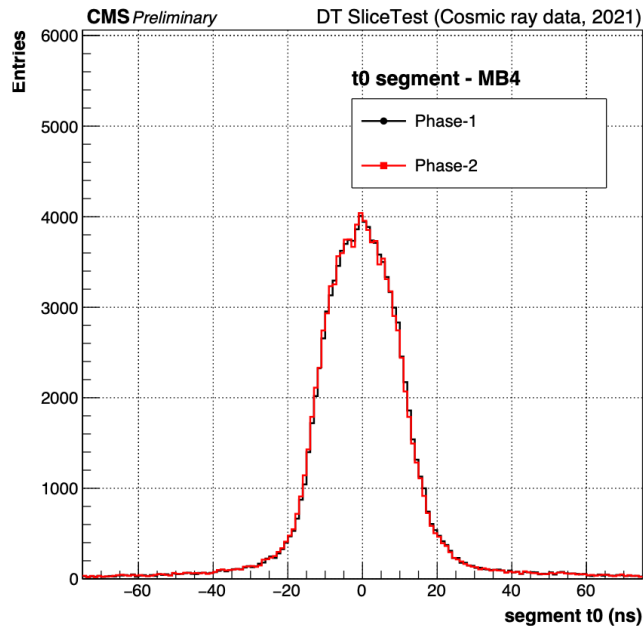


FIGURE 3.5: Segments crossing time reconstructed with Phase-1 (black) and Phase-2 (red) systems in the MB4 station of the DT Slice Test. Calibration is applied to both Phase-1 and Phase-2 time measurements, showing a strong agreement between the two systems. The results are sourced from [92].

In summary, the Phase-2 system exhibits significantly enhanced online time resolution compared to Phase-1. This is clearly illustrated in Fig. 3.6, where the solid blue line representing Phase-2 shows a lower fraction of triggers at incorrect bunch crossing (BX) compared to Phase-1 (solid red line) for a sample of unbunched cosmic muons. The comparison is based on primitives with a minimum of 4 hits (indicated as  $Q \geq 3$  in the legend) for Phase-2, to align with the Phase-1 system.

In the Phase-1 system, the trigger output is measured in BX units (25 ns intervals), depicted by the red line, which represents the convolution of a flat distribution within the BX time interval with a reconstructed time resolution for segments of 3-4 ns. In contrast, the Phase-2 system offers inherent online time resolution in the nanosecond range. This remarkable improvement in the performance of the AM reconstruction underscores the effectiveness of the Phase-2 architecture. It highlights the ability of the Phase-2 system to achieve its intended goals. The

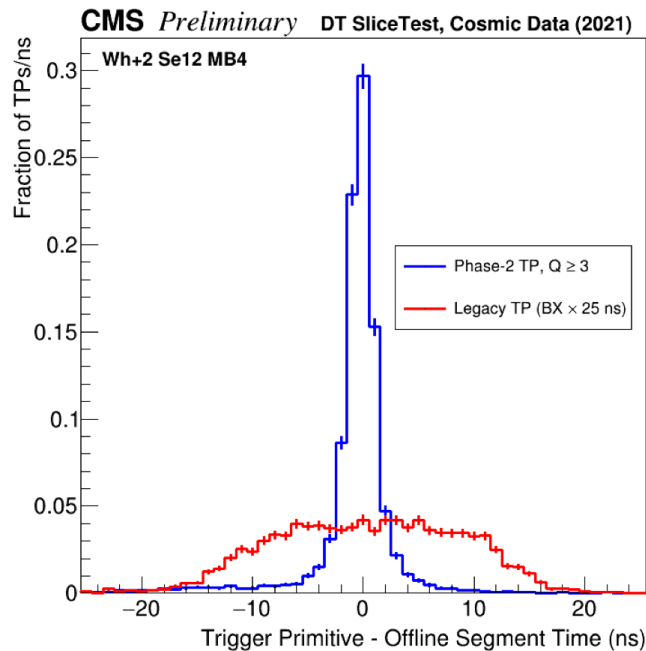


FIGURE 3.6: Discrepancy between the time of Trigger Primitives and the offline reconstructed segment time. The results are presented for Phase-2 in blue and the Legacy trigger in red, based on a cosmic muon sample obtained during the DT Slice Test. The figure is sourced from [90].

accomplishment of this result relies on the precise and reliable synchronization of the Phase-2 system, detailed in the subsequent section.

## 3.4 Phase-2 system time synchronization

The calibration methodology implemented for the Phase-2 signals allows us to eliminate any offsets caused by electronic delays or cable lengths in the signal used for muon track reconstruction.

### 3.4.1 Test Pulses run

To calibrate the electronics of the DT system, we use a Test-Pulse (TP) calibration mode which involves injecting simultaneous signals into all channels of a chamber. In the Phase-1 system, these signals are generated by activating different groups of four anode wires (channels) during each event, following specific pattern sequences that simulate vertical tracks, such as muons crossing the chamber perpendicularly at various positions. The TP signals are injected with a fixed known latency,

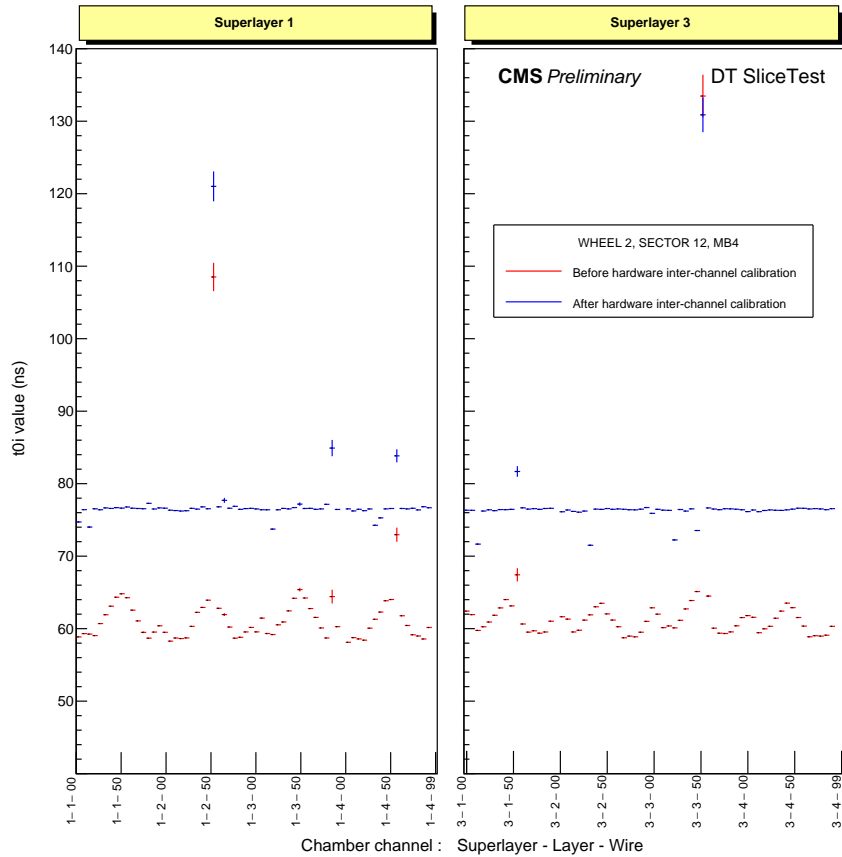


FIGURE 3.7: The pedestal value per channel ( $t_{0i}$ ) for the DT Slice Test Phase-2 is shown in the plot. The solid red line represents the pedestal values before applying the hardware inter-channel calibration, while the solid blue line represents the pedestal values after calibration using Test Pulses. Most of the values align with the expected behavior. However, a few pedestal values deviate from the expected behavior, likely due to sporadic misread channels.

and by analyzing the time readout from the TDCs, individual offsets ( $t_{0i}$ ) are determined for each wire.

Figure 3.7 visually and descriptively demonstrates the impact of hardware inter-channel calibration using TP runs, which we explain in detail in the subsequent section. The red distribution shape represents the cable lengths across the super layers. After applying the calibration, the distribution becomes uniform, although a few outliers may be observed due to sporadic misread channels that deviate from the expected behavior. To ensure the stability of synchronization, the Phase-2 readout system is subjected to testing involving "power cycles" (switching off and on) and consecutive resets of the clock source to the MOCO boards, which are involved in the configuration and monitoring of the OBDTs. Figure 3.8 presents the results of one of these tests, which involved six TP runs. The x-axis represents the reference  $t_{0i}$  value obtained from one TP run, while the y-axis shows the



differences in  $t_{0i}$  values obtained from five subsequent TP runs performed after applying an alternative operation, as described in the legend. This analysis helps to evaluate the stability of the synchronization of the Phase-2 readout system.

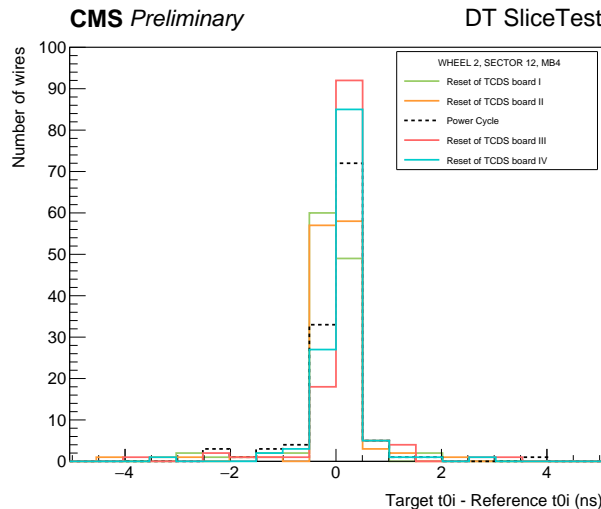


FIGURE 3.8: Stability analysis of Phase-2 readout pedestals was conducted by subjecting the system to power cycles and resets of the TCDS boards, followed by six Test-Pulse (TP) runs. The distribution of pedestal differences between the TP runs taken after applying manual operations (power cycle or TCDS board reset), and the first TP run as reference is shown. The spread of the distribution covers an eight ns window, while the central width of 1 ns aligns with the desired calibration, indicating a stable performance of the Phase-2 readout system.

### 3.4.2 Data taking in the Slice Test

Throughout the ST conducted at CERN between April and May 2021, multiple TP runs were executed to assess the stability of the Phase-2 system. The determination of the cell-by-cell  $t_{0i}$  values was repeated after each system reconfiguration and subsequent “power cycles” performed between runs. Each run was assigned a unique number, and its configuration details were recorded online to track the data-taking conditions and facilitate future verifications.

### 3.4.3 Calibration results

The stability of the  $t_{0i}$  calibration in the Phase-2 system is evaluated by comparing the average TDC times obtained from the same wire across different runs.

A reference run is selected, and the time differences between each run and the reference are calculated wire by wire. These differences are then aggregated for each chamber, and 2D histograms allow to visualize the variations.

Figure 3.9 presents an example plot showcasing the  $t_{0i}$  calibration stability for MB4 SL1, which provides position measurements on the  $r$ - $\phi$  plane. The SL consists of four layers, each containing 92 cells, represented by bins along the x-axis. Light-blue vertical bands separate cells from different layers, with their size reduced due to the larger number of cells per layer in the MB4 chamber. The y-axis represents the run numbers for the analyzed runs, with one bin per run. The z-axis displays the  $t_{0i}$  variation in nanoseconds between the target run and the reference run.

The time digitization bin has a width of 0.833 ns. Despite sporadic TP issues (affecting approximately ten wires out of nearly 3000) caused by temporary system failures, the Phase-2 electronics display good stability in time measurements. An effective monitoring tool has been developed to ensure the reproducibility of time measurements even after resets, power cycles, and reconfigurations. These delays, whose stability was extensively tested during LS2, continue to be used in 2023, highlighting the remarkable time stability of the entire electronics chain.

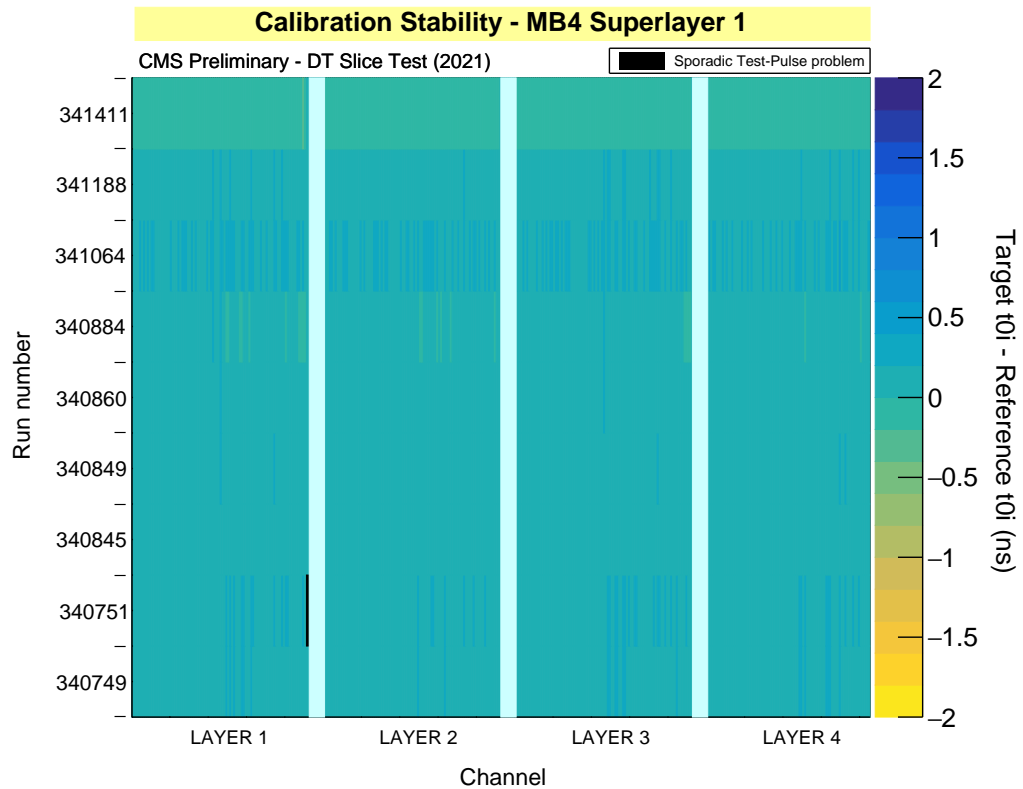


FIGURE 3.9: Stability test of  $t_{0i}$  calibration for MB4 Superlayer 1. Light-blue vertical bands separate cells of different layers, with their size adjusted to accommodate the increased number of cells in the MB4 chamber. The y-axis represents the run numbers for the analyzed runs, with one bin per run. The z-axis illustrates the  $t_{0i}$  variation in nanoseconds between the target and the reference run. Wires affected by sporadic TP injection problems are highlighted in black. Please note that the reference run is not included in the plot.



# Chapter 4

## Anomalous couplings analysis in the HWW channel

Starting from this chapter, we transition from the more technical aspects of the thesis to delve into the core of the physics analysis. In this section, we present our analysis of the anomalous couplings of the Higgs boson, specifically in the HWW channel.

For this study, we use the full Run 2 dataset obtained from the CMS experiment during the period of 2016 to 2018. The dataset comprises a total integrated luminosity of  $138 \text{ fb}^{-1}$ . As a first step, we introduce the HWW channel and offer an overview of the current understanding about it, drawing insights from the STXS (Simplified Template Cross Sections) analysis. We then proceed to delve into the signal signatures and identify the primary sources of background. Subsequently, we provide comprehensive information regarding the data and simulated samples employed in our analysis. We describe the procedures for object identification and correction, which include muons, electrons, jets, boosted V-jets, and missing transverse energy resulting from neutrinos in the final state. Moving forward, we outline our event selection criteria, which involve the categorization of events based on specific characteristics. We also discuss various observables and kinematic discriminants (KD), encompassing the kinematics in VBF and VH production for the HVV vertex, the kinematics in 2-jet ggF production for the Hgg vertex, and the kinematics of HWW decay products for the HVV vertex. Finally, we introduce our multi-dimensional kinematic discriminants and provide a study of the signal model used.

## 4.1 The HWW channel

The main focus of this thesis is the study of events on which the Higgs boson decays into a pair of W bosons. Due to the extremely brief lifetime of the W boson, which is approximately  $10^{-25}$  s, the only detectable signatures are the decay products. Similarly, the Higgs boson itself has a lifetime of  $\sim 10^{-22}$  s. The W boson may decay either hadronically into a pair of quarks or leptonically into a charged lepton and a neutrino. While the hadronic decay  $W \rightarrow q\bar{q}$  has a branching fraction of  $67.32 \pm 0.23\%$ , the leptonic decays  $W \rightarrow \ell\nu$  is  $10.89 \pm 0.09\%$  [23] per lepton (e,  $\mu$ ,  $\tau$ ). However, due to the experimental difficulty of distinguishing hadronic decays in a QCD-abundant environment like the LHC, leptonic decays, as shown in Fig. 4.1, are preferred and required for this analysis.

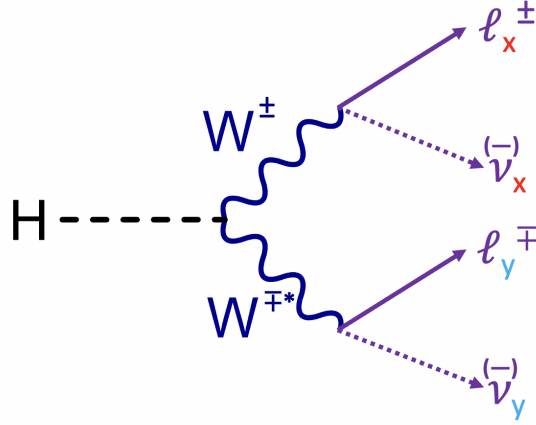


FIGURE 4.1: The decay considered in this analysis is the Higgs boson decaying into two W bosons. Each W decays into a lepton and a neutrino. This results in two oppositely charged different-flavor (DF) leptons and two neutrinos in the final state.

### 4.1.1 Measurements of HWW cross-sections

The current understanding of the HWW channel is primarily based on the SM analysis carried out in the Legacy analysis documented in Ref. [93]. This publication includes the STXS result, but it also has cross-section,  $\mu$ , and  $\kappa$  measurements. It provides not only inclusive production cross-sections but also differential cross-section measurements using the STXS framework. The STXS approach involves measuring the cross sections of various Higgs boson production mechanisms in specific regions of generator-level phase space known as STXS bins, which are

designed to enhance sensitivity to any deviations from the SM. To perform the STXS measurement, a set of analysis categories is defined to target each STXS bin, as illustrated in Fig. 4.2. The AC analysis in this thesis follows the baseline ggH, VBF, and VH selections defined for the Legacy analysis.

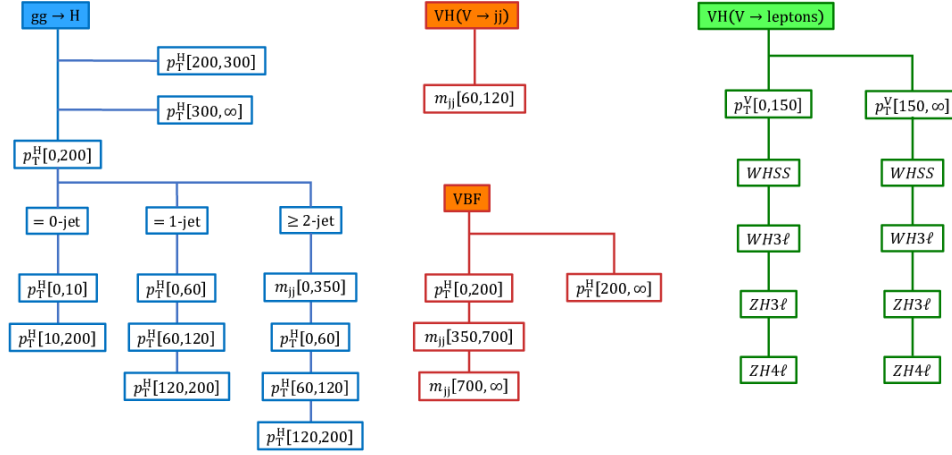


FIGURE 4.2: The categories in the analysis have been defined to target the STXS bins, with all dimensional quantities measured in units of GeV.

The figure 4.3 summarizes the main results from the Legacy analysis, including constraints on the parameters  $\kappa_V$  and  $\kappa_F$  (A), a scan of the signal strength (B), and the observed cross-sections (C) for the main Higgs production mechanisms. The combined results, considering the associated systematic and statistical uncertainties, align with the expectations of the SM.

### 4.1.2 Signal signatures

This section explores the characteristics of WW decay and the ones associated with the Higgs production mechanisms that are considered for analysis.

#### HWW decay features

The main feature of the events of interest is the presence of a pair of leptons and two neutrinos arising from the decay of the Higgs boson into two W bosons. More precisely, events containing one electron and one muon are chosen. While this selection might lead to a reduction in the expected number of signal events, opting for leptons of distinct flavors in the final state enhances the analysis sensitivity by

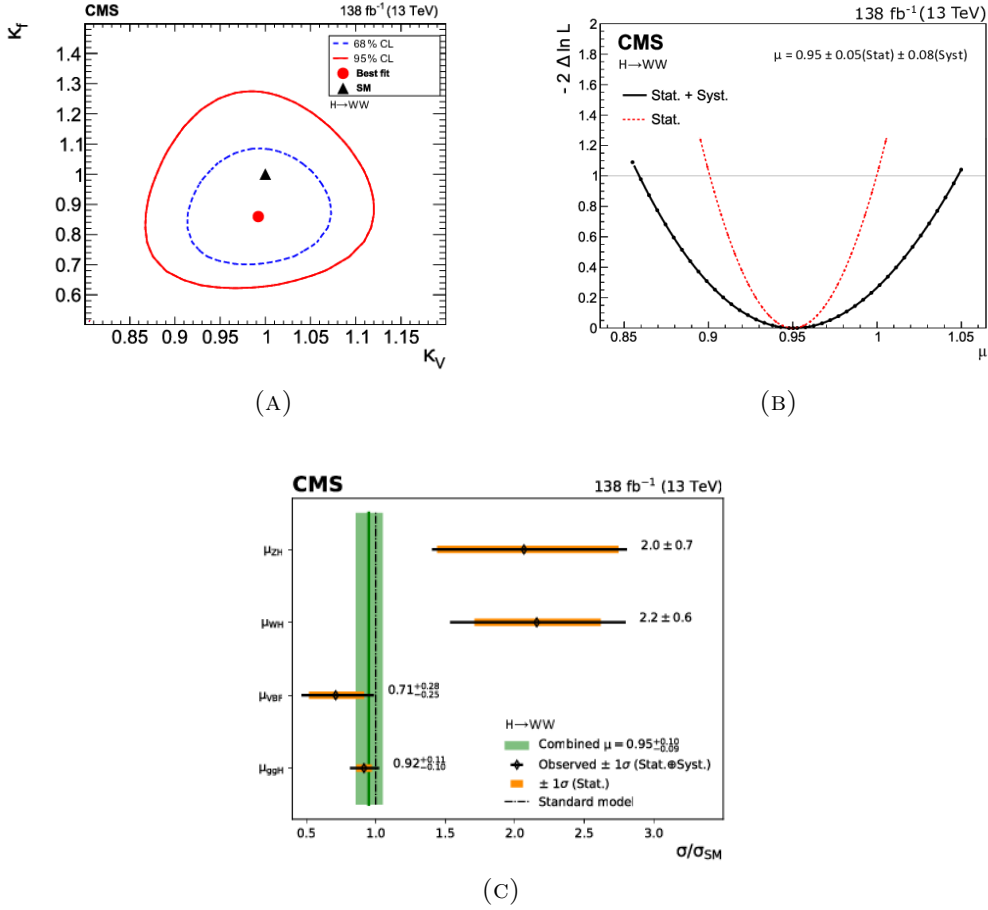


FIGURE 4.3: Results of the full Run 2 HWW Legacy analysis, including the constraints on the kappa parameters  $\kappa_V$  and  $\kappa_F$  (A), a scan of the signal strength  $\mu$  (B), and the observed cross-sections (C) for each of the main Higgs production mechanisms when all STXS bins are combined.

minimizing background processes like Drell-Yan, as elaborated upon later. Hence, the final state of interest as follows:

$$H \rightarrow WW \rightarrow e\nu_e \mu\nu_\mu. \quad (4.1)$$

The leptonic decay of the  $\tau$  leptons contributes to the signal, but it is not the main objective, as such decays present a different topology. In order to visually represent an event with those characteristics within the CMS detector volume, an illustration of a candidate decay of HWW into leptons and neutrinos is shown in Fig. 4.4. The neutrinos are not directly detected by the CMS detector and therefore manifest as missing transverse energy in the event.

An additional intriguing feature of the desired final state is the manifestation of the scalar nature of the Higgs boson through a spin correlation effect, as depicted



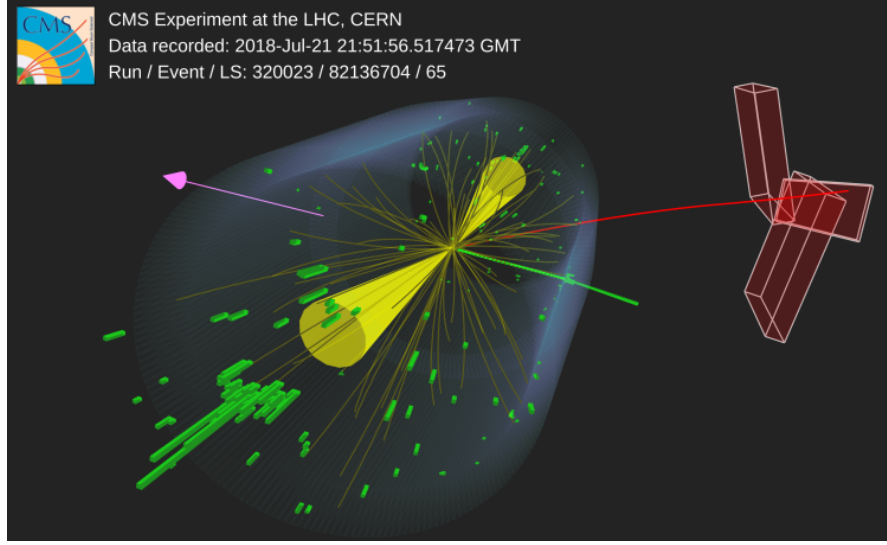


FIGURE 4.4: A candidate HWW event displaying the VBF Higgs boson production process, where the two yellow cones depict the forward jets typical of the VBF process. The green and red lines represent the electron and muon, respectively, from the W decays. The magenta arrow indicates the missing energy corresponding to the neutrinos from each W decay, which cannot be directly detected. Source [94].

in Fig. 4.5. Conservation of angular momentum necessitates that the pair of spin-1 W bosons, originating from a spin-0 resonance, exhibit opposite spin projections. When the W boson decays leptonically, it emits a neutrino (or anti-neutrino) with a momentum directed opposite to its spin, owing to its left-handed (or right-handed) nature. Consequently, a slight separation angle between the two charged leptons in opposite hemispheres arises. Another notable aspect is that one of the W bosons is emitted virtually (as  $m_H < 2m_W$ ), leading to one of the leptons possessing, on average, a lower  $p_T$  compared to cases when it is emitted from an on-shell W boson.

By highlighting these features, the intention is to mitigate or eliminate the impact of background events. In particular, two kinematic properties play a crucial role as discriminant variables in the examination of the Higgs boson in the WW channel: the transverse mass of the Higgs boson and the invariant mass of the  $e\mu$  pair in the final state. The invariant mass of the  $\ell_1\ell_2$  pair can be defined as follows:

$$m_{\ell_1\ell_2} = \sqrt{2p^{\ell_1}p^{\ell_2}(1 - \cos \alpha)}. \quad (4.2)$$

Here,  $p^{\ell_1}$  and  $p^{\ell_2}$  represent the momenta of the leading and subleading leptons, respectively, while  $\alpha$  denotes the angle between the leptons. It is important to note that the above equation assumes negligible masses for the leptons, which is

a reasonable assumption considering the energy scale of approximately 10 GeV or higher that is being considered in our analysis. This variable should be small for signal events due to the angle between the emitted leptons, which is useful to discriminate against background contributions such as the non-resonant WW production.

The presence of neutrinos in the final state cannot be directly detected in experiments; however, their contribution leads to the appearance of missing transverse momentum in the event ( $p_T^{miss}$ ). This quantity is employed to compute the transverse mass of the Higgs boson. The transverse mass is defined as the mass calculated in the transverse plane for the system consisting of the leptons and neutrinos, and it is given by the following expression:

$$m_T^H = m_T(\vec{p}_T^{\ell_1 \ell_2}, \vec{p}_T^{miss}), \quad (4.3)$$

where the transverse mass of a generic system of particles  $P_i$  is defined as:

$$m_T(\{P_i\}) = \sqrt{\left(\sum |\vec{p}_{T,i}|\right)^2 - \left|\sum \vec{p}_{T,i}\right|^2}. \quad (4.4)$$

Moreover, the variable  $m_T^H$  proves to be highly effective in distinguishing between background processes and the signal from the Higgs boson, as it possesses a higher limit at  $m_H$ , a requirement that background events are not bound to satisfy. The choice to incorporate both the  $m_{\ell\ell}$  and  $m_T^H$  variables in the analysis is justified by previous investigations, including studies such as the STXS.

## Production features

In the presence of the final state outlined in Eq. 4.1, additional jets are generated alongside the Higgs boson. The analysis takes into account three main Higgs production mechanisms: ggH, VBF, and VH.

- Regarding the ggH production mechanism depicted in Fig. 4.6 (A), the objects described in Eq. 4.1 maybe be accompanied by hadronic jets due to radiation of the gluons.
- In the VBF production mechanism illustrated in Fig. 4.6 (B), the detection of the signature relies on the presence of two jets originating from the quarks

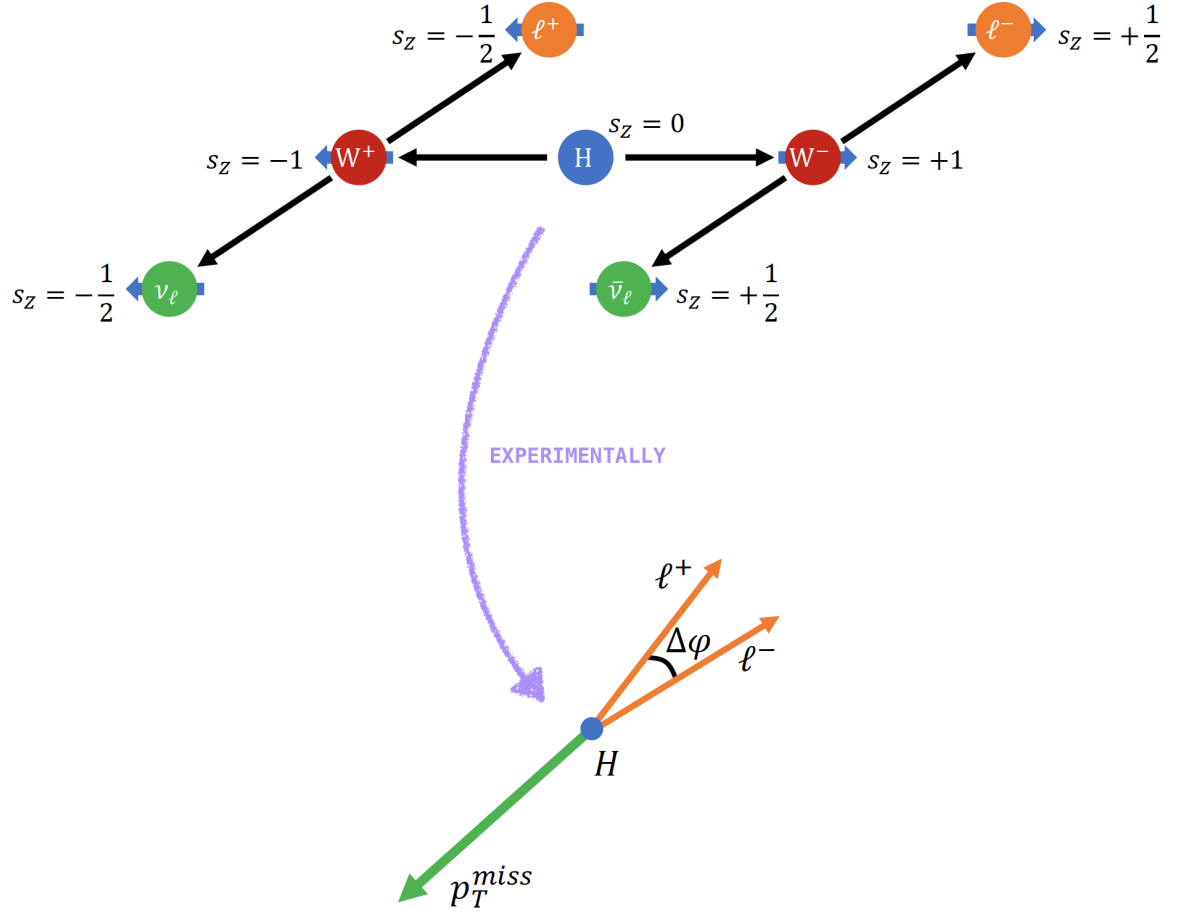


FIGURE 4.5: An illustrative depiction showing the spin correlation effect in the decay of the Higgs boson into a pair of W bosons.

radiation. The production of additional jets is significantly suppressed in this topology, as the quarks do not directly interact with each other. Consequently, there is a substantial separation within the detector barrel ( $\Delta\eta_{jj}$ ) and a high invariant mass of the dijet system ( $m_{jj}$ ). These kinematic characteristics prove valuable in distinguishing VBF jets from QCD jets in the ggH process.

- In the presented analysis, the production of the Higgs boson in association with Z and W bosons (ZH and WH, respectively) is distinguished by the presence of additional vector bosons ( $V = Z, W$ ), as depicted in Fig. 4.6 (C). VH events are considered, where the V boson decays hadronically into quarks. This decay may be reconstructed as two AK4 jets which will have a resonant invariant mass around the mass of the V boson. Additionally, we select boosted decays originating from high transverse momentum V bosons. These decays are reconstructed using a single large radius AK8 jet, which will

be discussed in section 4.3.3 due to their advantageous kinematic features for detecting BSM physics and background suppression. The reconstructed AK8 jet displays a substructure composed of two sub-jets, which has crucial importance in the analysis.

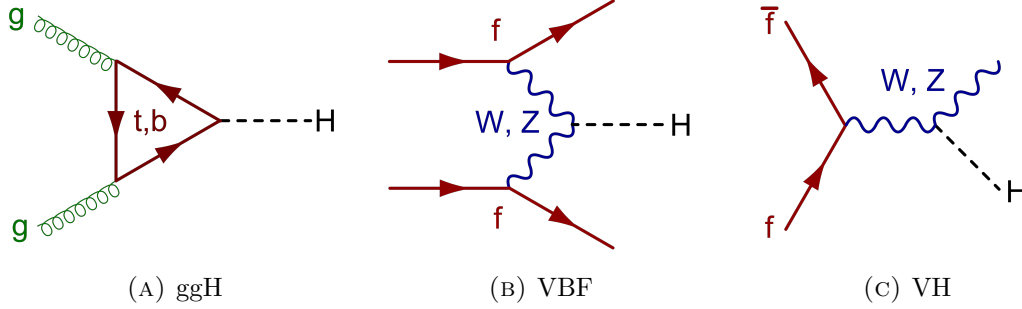


FIGURE 4.6: The three Higgs boson production mechanisms explored in this analysis are depicted as follows: ggH (A), VBF (B), and VH (C).

Therefore, the analysis initial stage involves selecting HWW leptonic events, which are subsequently categorized according to their associated jets:

- ggH production: Events with 0, 1, or 2 jets radiated.
- VBF production: Events characterized by 2 jets.
- VH production with 2 jets originating from the V boson decay.
- VH production with one AK8 jet stemming from the V boson decay.

### 4.1.3 Sources of background

The background events arise from SM processes that show the same final state as the targeted signal. However, distinctive kinematic characteristics associated with each decay process aid in identifying the origins of these events. To estimate the expected number of background events within the analysis phase space, we employ a combination of Monte Carlo (MC) simulations and data-driven techniques.

The primary backgrounds considered in this analysis include non-resonant WW production, the Drell-Yan process, top quark production, non-prompt leptons (often referred to as fakes), and multiboson production. Generally, among these backgrounds, the most significant contributors are, in order, top quark production, the Drell-Yan process, and the non-resonant WW production.

**Top quark production** The dominant background in different-flavor (DF) channels consists of hadronic jets originating from the decay of top quarks into a b quark. To distinguish these jets from those originating from light flavor quarks (u, d, s, or c) or gluons, b-tagging algorithms are employed (as described in section 2.2). These algorithms exploit distinct characteristics of b-jets, such as their long lifetime and the large mass of the bound states they form, which differentiate them from other short-lived hadrons. Figure 4.7 illustrates some example diagrams for the production of top quark pairs.

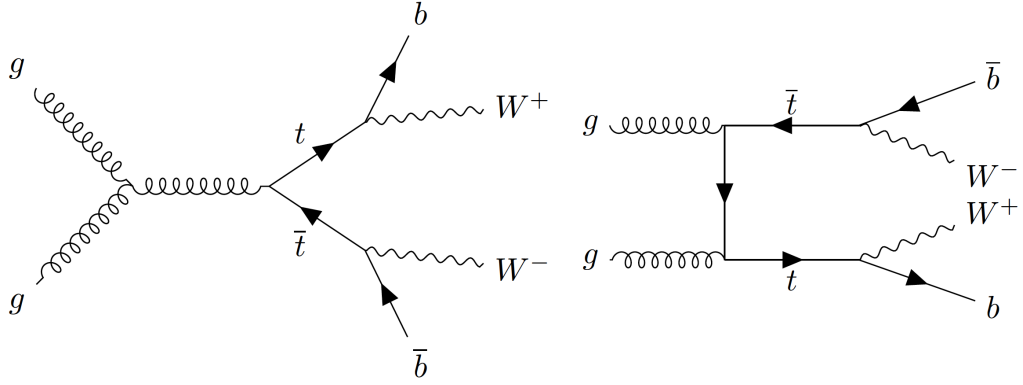


FIGURE 4.7: The diagrams serve as examples illustrating the production of pairs of top quarks.

**Drell-Yan** The Drell-Yan (DY) process involves the production of a pair of same-flavor (SF) leptons originating from either a Z boson or a virtual photon ( $\gamma^*$ ) decay. To mitigate  $DY\mu\mu$  and  $DYee$  events, the analysis selects a final state with DF leptons. However, events having  $DY \rightarrow \tau\tau \rightarrow e\mu$  are still considered as background, and thus, additional kinematic cuts such as  $E_T^{\text{miss}} > 20$  GeV and  $p_T^{\ell\ell} > 30$  GeV are applied. Estimating the DY background may involve systematic uncertainties due to the imperfect modeling of this effect. Therefore, dedicated data-driven techniques are employed, as detailed in section 5.3.

**Non-resonant WW production** To differentiate non-resonant WW pairs from Higgs decay, the analysis makes use of the  $m_{\ell\ell}$  variable. In the case of non-resonant WW, this variable exhibits higher invariant mass values compared to Higgs decay due to the angle between the emitted leptons. Non-resonant WW production is generated through both quark and gluon-initiated processes within the SM.

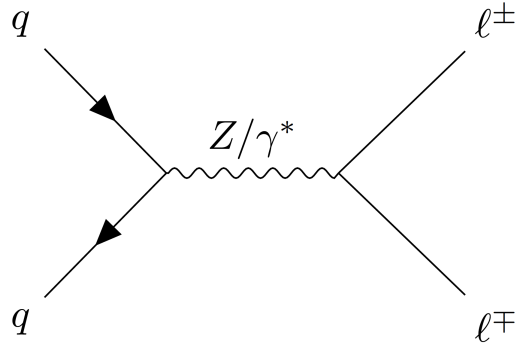


FIGURE 4.8: The diagram illustrates the LO process of generating a pair of dileptons through the Drell-Yan mechanism.

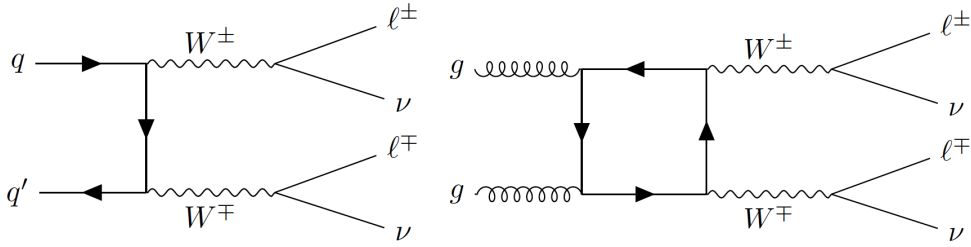


FIGURE 4.9: Non-resonant WW pair production through quark-quark (left) and gluon-gluon-initiated (right) processes.

**Non-prompt lepton background** The non-prompt leptons or the fake lepton background in this analysis arises from various sources. These sources include misidentified hadrons or jets, decays of heavy-flavor hadrons (B or D) into leptons, and photon conversions to electrons in the tracker. The main source of fakes comes from W+jets, where there is one real lepton from the W and a fake from the jet. Estimating this background is challenging due to the limited modeling of the finite resolution of the detector (as discussed in section 2.2). To reduce its contribution, additional identification and isolation criteria are applied, as explained in section 2.3.1). The analysis described in Ref [95] encountered significant discrepancies when estimating the non-prompt lepton background due to uncertainties in the fake electron modeling. As a solution, various studies were conducted, leading to the adoption of a modified electron identification approach. The AC analysis builds upon previous HWW analyses, so the derivation and discussion of the non-prompt lepton background estimation method are found in Appendix A and Ref [96]. Here, we show in Fig. 4.10 a selection of predicted and observed distributions from same-sign dilepton control regions, which are dominated by non-prompt background.

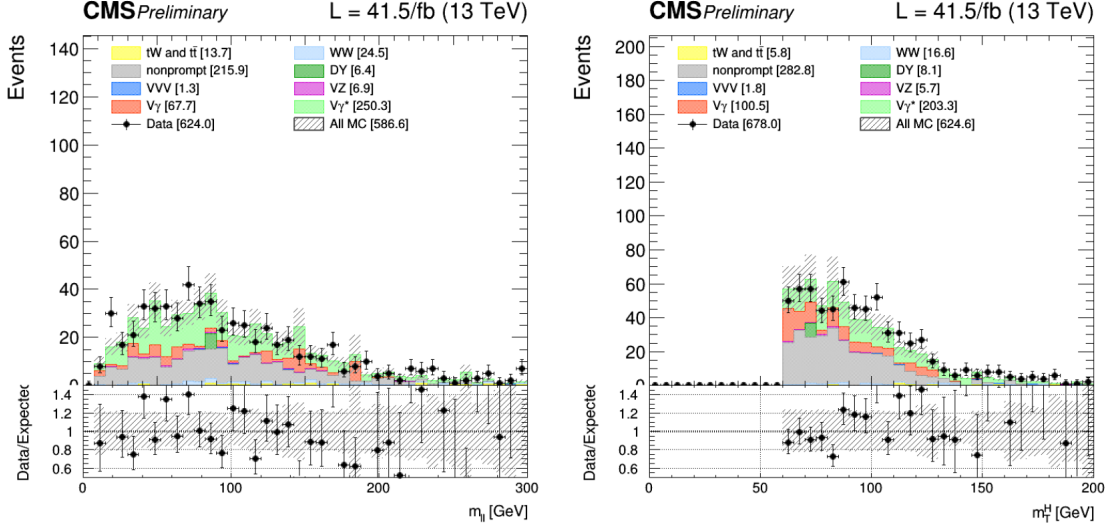


FIGURE 4.10: Same-sign control region plots 2017.  $m_{\ell\ell}$  from  $e\mu$  events (left) and  $m_T^H$  from  $\mu e$  events (right).

**Multiboson production** The production of diboson and triboson events includes  $WW$ ,  $ZZ$ ,  $ZZ$ ,  $WZ$ ,  $V\gamma^*$ ,  $V\gamma$  processes. The contribution of these backgrounds to the analysis depends on the specific final state being studied, which is discussed in subsequent sections.

## 4.2 Data and simulated samples

Moving on to the more technical aspects of the AC analysis, we will now discuss the definition of the data samples used in this study. This section provides details on the datasets employed, which encompass both the actual data obtained from the CMS experiment and the simulated MC samples. Given the evolution of the CMS and LHC configurations over time, the datasets were collected and segregated by year. Similarly, the MC samples were generated individually for each year to ensure their consistency with the corresponding datasets.

### 4.2.1 Data samples

To account for changes in the CMS and LHC configurations over time, the datasets were divided by year. The integrated luminosities for the individual years were  $36.3 \text{ fb}^{-1}$ ,  $41.5 \text{ fb}^{-1}$ , and  $59.7 \text{ fb}^{-1}$  for 2016, 2017, and 2018, respectively [97–99]. The average pileup was 23 for 2016 and 32 for the other two years. In order

to be included in the analysis, events were required to meet the reconstruction quality criteria for physics objects, particularly leptons, and to satisfy the trigger requirements implemented by the HLT system. Different final states were triggered by single and double lepton triggers with varying thresholds in  $p_T$  depending on the specific year. For the electron-muon final state, events were required to have fired at least one of the triggers listed in Table 4.1, 4.2, and 4.3 for 2016, 2017, and 2018, respectively.

In 2017 and 2018, adjustments in the triggers were made to the thresholds for electrons and muons. The thresholds for electrons were changed to 35 GeV and 32 GeV for 2017 and 2018, respectively. This was necessary because the L1 trigger relies on ECAL clusters for electron reconstruction, which is affected by noise and has difficulty distinguishing electrons from photons at low energies. On the other hand, the L1 trigger incorporates information from muon chambers for muon reconstruction, compensating for the lack of information from the silicon tracker and enabling lower trigger thresholds for muons. Using the Tag-And-Probe (T&P) method [75] described in section 2.3.1, the efficiency of the triggers for  $Z \rightarrow \ell\ell$  events was measured to be 98%.

TABLE 4.1: The HLT paths utilized in the analysis of the 2016 data.

| Dataset        | Run range     | HLT path   |
|----------------|---------------|--|
| MuonEG         | B–F (–278272) | HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL<br>HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL        |
|                | F (278273)–H  | HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ<br>HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ |
| SingleMuon     | B–H           | HLT_IsoMu24<br>HLT_IsoTkMu24   |
| SingleElectron | B–H           | HLT_Ele27_WPTight_Gsf<br>HLT_Ele25_eta2p1_WPTight_Gsf  |
| DoubleMuon     | B–G           | HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL<br>HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL                                   |
|                | H             | HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ<br>HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ                             |
| DoubleEG       | B–H           | HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ  |

## 4.2.2 Simulated samples

In this analysis, a variety of MC event generators are employed to simulate both signal and background processes, a common practice in particle physics experiments. These tools play a crucial role in optimizing event selection and efficiency,



TABLE 4.2: The HLT paths utilized in the analysis of the 2017 data.

| Dataset        | Run range | HLT path   |
|----------------|-----------|--|
| MuonEG         | B         | HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ<br>HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ |
|                | C–F       | HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL<br>HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ    |
| SingleMuon     | B–F       | HLT_IsoMu27  |
| SingleElectron | B–F       | HLT_Ele35_WPTight_Gsf  |
| DoubleMuon     | B         | HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ  |
|                | C–F       | HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8  |
| DoubleEG       | B–F       | HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL   |

TABLE 4.3: The HLT paths utilized in the analysis of the 2018 data.

| Dataset        | Run range | HLT path   |
|----------------|-----------|--|
| MuonEG         | A–D       | HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ<br>HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ |
| SingleMuon     | A–D       | HLT_IsoMu24  |
| SingleElectron | A–D       | HLT_Ele32_WPTight_Gsf  |
| DoubleMuon     | A–D       | HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8  |
| DoubleEG       | A–D       | HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL   |

estimating expected yields for different processes, and evaluating systematic uncertainties. The response of the CMS detector is modeled using the GEANT4 (GEometry AND Tracking) platform [100], which provides a detailed description of the detector components. Event generation is performed using specific MC packages tailored to the physics process of interest, and additional pile-up interactions are included using PYTHIA [101].

The same reconstruction algorithms are applied to both MC samples and real data, ensuring consistency in the analysis. MC samples from the same year share common PDFs, parton shower (PS) configurations and underlying event (UE) simulations. PDF sets are determined through fits data, as described in section 1.2.1. The PS hadronization and UE tune are provided by PYTHIA 8.2 [102]. The UE and multiple interactions (MPI) are based on the CUETP8M1 [103] and CP5 [104] tunes for the 2016 and 2017-2018 samples, respectively.

## Signal MC samples

In order to provide a comprehensive reference for the analysis, it is essential to include technical details about the event generators employed. The following summary provides pertinent information accompanied by relevant references.

**SM Higgs boson production** The focus of this analysis is the Higgs boson, and signal samples for ggH, VBF, VH, and ttH are generated using POWHEG 2.0 [105–108]. This generator reproduces the properties of these processes at next-to-leading order (NLO) accuracy in QCD. The inclusive (0-, 1-, 2-jets or more) ggH sample is generated using the POWHEG + MINLO approach and then reweighted with the NNLOPS scheme [109] to match the next-to-next-to-leading order (NNLO) accuracy in the Higgs boson  $p_T$  (transverse momentum) and the number of associated jets. The MINLO HJJ generator [110] improves the description of phase spaces with 2 jets or more, ensuring NLO accuracy in events with a number of jets greater than or equal to 2 and LO accuracy for events with 3 or more jets. The cross-sections used to normalize the samples are provided in the last column of Table 4.4.

**SM Higgs boson decay** The simulation of the decay of the Higgs boson into two W bosons and subsequent decay into leptons for ggH, VBF, and quark-initiated ZH production samples is performed using the JHUGEN generator [49]. JHUGEN version 5.2.5 is used for the 2016 samples, while version 7.1.4 is used for the 2017 and 2018 samples. For all other SM Higgs signal samples, the decay of the Higgs boson is simulated with PYTHIA 8.2 [102]. In 2016, PYTHIA 8.226 was used, and in 2017 and 2018, PYTHIA 8.230 was used. More information about the simulated samples of the SM Higgs boson are found in Table 4.4.

**HVV AC samples** Simulations of VBF and VH Higgs productions with anomalous couplings are performed using JHUGEN at LO QCD. The JHUGEN and POWHEG SM Higgs simulations are compared after parton showering, and no significant differences are found in kinematic observables as similarly reported in [46]. Therefore, the JHUGEN simulation is adopted to describe kinematics in all production modes with anomalous couplings, while the expected yields are taken from the POWHEG simulations. For ggH signals with anomalous HVV couplings, JHUGEN is used for generation. The kinematics of ggH events with

two associated jets are sensitive to the anomalous Hgg coupling at the production vertex, and therefore, events with anomalous H+2 jet coupling are also simulated.

The signal samples with anomalous couplings suitable for all three years are provided in Tables 4.5 and 4.6 for HVV and Hgg vertices, respectively. Multiple samples with different anomalous coupling hypotheses have been generated to extract various coupling components and their interference. The MELA package [111–114], which is based on JHUGEN matrix elements, simplifies the reweighting to any HVV or Hgg coupling signal hypothesis with the same production mechanism. This reweighting technique, known as *MELA reweighting*, enhances statistics by using all samples simultaneously for any signal hypothesis, allowing coverage of any model even if it has not been directly simulated.

Prior to its use, a study was conducted to validate the samples and compare the LO and higher-order SM Higgs predictions by examining the shapes of discriminant variables such as  $m_{\ell\ell}$  and  $m_T^H$  as introduced above. For the HVV signal kinematics, the JHUGEN LO simulation was used, while the expected yields were taken from the higher-order NLO simulations. The final comparison between the LO and higher-order SM Higgs predictions revealed no significant differences. As an illustrative example, Fig. 4.11 presents the distributions of the  $m_{\ell\ell}$  and  $m_T^H$  variables in the 0-jet channel for the ggH signal.

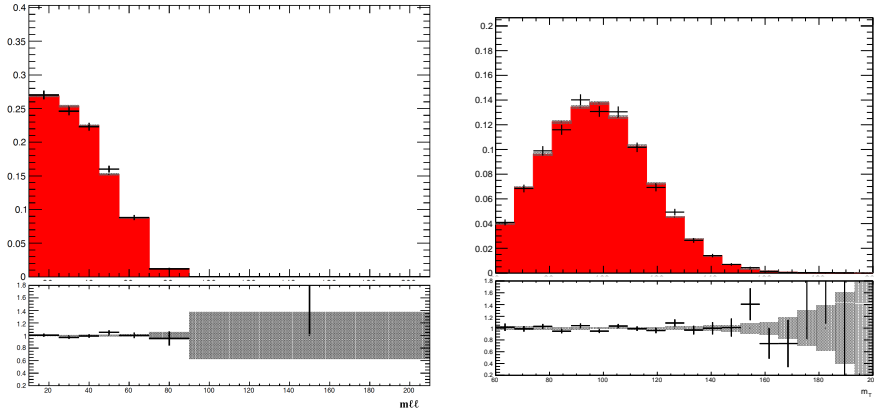


FIGURE 4.11: The  $m_{\ell\ell}$  (left) and  $m_T^H$  (right) distributions of SM Higgs at reconstruction level in the 0-jet channel are presented for both the JHUGEN (red histogram) and POWHEG + MINLO generators (black markers). All distributions are normalized to unity, and the statistical uncertainty of the MC is indicated for each distribution.

In order to account for the effects of parton showering (PS), the initial Hgg AC samples underwent a substitution mandated by the CMS internal review process.

TABLE 4.4: SM HWW signal MC samples normalized to cross-section.

| process          | year       | dataset name  | xs [pb] |
|------------------|------------|---|---------|
| ggH              | 2016       | GluGluHToWWTo2L2Nu_M125_13TeV_powheg_pythia8                      | 0.9913  |
| ggH Hjj          | 2017, 2018 | GluGluHToWWTo2L2Nu_M125_13TeV_powheg2_JHUGenV714_pythia8          | 0.8962  |
|                  |            | GluGluHToWWTo2L2Nu_M125_13TeV_powheg_minloHJJ_JHUGenV7011_pythia8 |         |
| VBF              | 2016       | VBFHToWWTo2L2Nu_M125_13TeV_powheg_JHUGen_pythia8                  | 0.0846  |
|                  | 2017, 2018 | VBFHToWWTo2L2Nu_M125_13TeV_powheg2_JHUGenV714_pythia8             |         |
| W <sup>+</sup> H | 2016       | HWplusJ_HToWW_M125_13TeV_powheg_pythia8                           | 0.135   |
|                  | 2017, 2018 | HWplusJ_HToWW_M125_13TeV_powheg_pythia8_TuneCP5                   |         |
| W <sup>-</sup> H | 2016       | HWminusJ_HToWW_M125_13TeV_powheg_pythia8                          | 0.116   |
|                  | 2017, 2018 | HWminusJ_HToWW_M125_13TeV_powheg_pythia8_TuneCP5                  |         |
| ZH               | 2016       | HZJ_HToWW_M125_13TeV_powheg_pythia8                               | 0.187   |
|                  | 2017, 2018 | HZJ_HToWWTo2L2Nu_M125_13TeV_powheg_jhugen714_pythia8_TuneCP5      |         |
| ggZH             | 2016       | GluGluZH_HToWW_M125_13TeV_powheg_pythia8                          | 0.00275 |
|                  | 2017, 2018 | GluGluZH_HToWWTo2L2Nu_M125_13TeV_powheg_pythia8_TuneCP5           |         |
| ttH              | all        | ttHToNonbb_M125_TuneCP5_13TeV_powheg_pythia8                      | 0.212   |

TABLE 4.5: AC HVV signals MC samples utilized in the analysis for all years.

|   |
|---|
| VBFHiggsOPMToWWTo2L2Nu_M-125_13TeV-JHUGenV6_pythia8       |
| VBFHiggsOPHToWWTo2L2Nu_M-125_13TeV-JHUGenV6_pythia8       |
| VBFHiggsOL1ToWWTo2L2Nu_M-125_13TeV-JHUGenV6_pythia8       |
| VBFHiggsOMToWWTo2L2Nu_M-125_13TeV-JHUGenV6_pythia8        |
| VBFHiggsOPHf05ph0ToWWTo2L2Nu_M-125_13TeV-JHUGenV6_pythia8 |
| VBFHiggsOMf05ph0ToWWTo2L2Nu_M-125_13TeV-JHUGenV6_pythia8  |
| VBFHiggsOL1f05ph0ToWWTo2L2Nu_M-125_13TeV-JHUGenV6_pythia8 |
| WHiggsOPMToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8       |
| WHiggsOPHToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8       |
| WHiggsOL1ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8       |
| WHiggsOMToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8        |
| WHiggsOPHf05ph0ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8 |
| WHiggsOMf05ph0ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8  |
| WHiggsOL1f05ph0ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8 |
| ZHiggsOPMToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8       |
| ZHiggsOPHToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8       |
| ZHiggsOL1ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8       |
| ZHiggsOMToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8        |
| ZHiggsOPHf05ph0ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8 |
| ZHiggsOMf05ph0ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8  |
| ZHiggsOL1f05ph0ToWW_2LFilter_M-125_13TeV-JHUGenV6_pythia8 |
| HiggsOPMToWWTo2L2Nu_M-125_13TeV-powheg2-JHUGenV6          |
| HiggsOPHToWWTo2L2Nu_M-125_13TeV-powheg2-JHUGenV6          |
| HiggsOMToWWTo2L2Nu_M-125_13TeV-powheg2-JHUGenV6           |
| HiggsOL1ToWWTo2L2Nu_M-125_13TeV-powheg2-JHUGenV6          |
| HiggsOPHf05ph0ToWWTo2L2Nu_M-125_13TeV-powheg2-JHUGenV6    |
| HiggsOMf05ph0ToWWTo2L2Nu_M-125_13TeV-powheg2-JHUGenV6     |
| HiggsOL1f05ph0ToWWTo2L2Nu_M-125_13TeV-powheg2-JHUGenV6    |

TABLE 4.6: AC signals MC samples in the Hgg channel for all years.

|   |
|---|
| JJHiggsOPMToWWTo2L2Nu_M-125_NNPDF31_TuneCP5_13TeV-powheg-minloHJJ-JHUGenV751-pythia8      |
| JJHiggsOMToWWTo2L2Nu_M-125_NNPDF31_TuneCP5_13TeV-powheg-minloHJJ-JHUGenV751-pythia8       |
| JJHiggsOMf05ph0ToWWTo2L2Nu_M-125_NNPDF31_TuneCP5_13TeV-powheg-minloHJJ-JHUGenV751-pythia8 |

The MINLO version of the samples, as indicated in Table 4.6, was employed to accurately quantify these effects. During the analysis review, it was noticed that the PS was not appropriate for JHUGen Hgg AC samples, which degraded the distinction between SM and AC signals. The correlation between jets, specifically the two leading jets, modeled by either JHUGen LO or MiNLO NLO simulation at the LHE level, shows good agreement. However, when Pythia hadronization is introduced, the modeling of additional jets differs. In our analysis, we need to select two jets from the multiple jets generated by Pythia+LHE. Among the jets generated by Pythia, when they are matched to JHUGen, there is a higher likelihood of selecting a jet that does not originate from LHE. In general terms, Pythia matching performs better and is more achievable with MiNLO, as up to 4 jets are already generated by the matrix element. In theory, the parton shower matching could be enhanced for either generator, and it is possible to tune the JHUGen+Pythia combination to be as effective as MiNLO+Pythia for specific purposes, but this objective has not yet been attained.

### Background MC simulation samples

Table 4.7 provides a summary of the background contributions discussed in section 4.1.3 for the three data sets conforming to full Run 2. Various event generators are utilized to model these backgrounds, including POWHEG v2 with MINLO HVJ [115], MadGraph\_aMC@NLO [116] (MadGraph5), and MCFM 7.0 [117].

| Process                   | Generator       | Accuracy |
|---------------------------|-----------------|----------|
| (qq)WW                    | POWHEG 2.0      | NLO      |
| (gg)WW                    | MCFM 7.0        | LO       |
| Top                       | POWHEG 2.0      | NLO      |
| DY                        | MadGraph5 2.4.1 | NLO      |
| $V\gamma$ and $V\gamma^*$ | MadGraph5 2.4.1 | NLO      |
| VV                        | POWHEG 2.0      | NLO      |
| VVV                       | MadGraph5 2.4.2 | NLO      |

TABLE 4.7: MC samples utilized for modeling background processes. Cross-section values listed in [118]

The backgrounds originating from non-prompt leptons, which arise from leptonic decays of heavy quarks, hadrons misidentified as leptons, and electrons from photon conversions, are suppressed by imposing identification and isolation requirements on electrons and muons. In this analysis, the primary source of the non-prompt lepton background is W+jets events, and its estimation is performed using

data-driven methods outlined in Ref [119]. The procedure involves measuring the rate of non-prompt leptons, passing loose selection criteria, and also passing a tight selection (misidentification rate). Similarly, the rate for a prompt lepton to pass the tight selection (prompt rate) is determined. The misidentification rate is measured in a data sample enriched with multijet events, while the prompt rate is measured using a T&P method in a data sample enriched with DY events. The estimation of the non-prompt background is validated using data in a CR enriched with W+jets events, where a pair of same-sign leptons is required.

The backgrounds originating from top quark processes are estimated through a combination of MC simulations and dedicated control regions (CRs) in the data. To ensure accurate modeling, a reweighting technique is employed for the top and anti-top quark  $p_T$  spectrum at the parton level in the  $t\bar{t}$  simulation. This reweighting aligns the simulation with the predictions from NNLO (next-to-next-to-leading order) and NNLL (next-to-next-to-leading logarithm) QCD calculations, incorporating the NLO (next-to-leading order) electroweak contribution [120]. The normalization of the top quark background in each considered jet multiplicity region is determined by comparing the observed data in the CRs specifically enriched with top quark events.

The estimation of nonresonant WW backgrounds involves a combination of MC simulations and dedicated regions in the data. The simulated events for quark-induced WW processes are reweighted to match the diboson  $p_T$  spectrum calculated at NNLO+NNLL QCD accuracy [121, 122]. The normalizations of the quark-induced and gluon-induced WW backgrounds are treated as free parameters in the fit, with a separate parameter assigned to each jet multiplicity region, similar to the approach used for the top quark background. In the 0- and 1-jet channels, the normalization parameters are directly constrained within the signal regions, which cover the high dilepton invariant mass phase space enriched in WW events. For the 2-jet channels, a dedicated control region is defined, which specifically selects events with high dilepton mass to enhance the presence of WW events.

The background process involving DY decays to  $\tau\tau$  is estimated using a data embedding technique [123]. In the fitting procedure to data, the normalization of this background is left unconstrained, following the same approach as for the top quark and WW backgrounds. To constrain the free normalization parameters, an orthogonal control region enriched with  $DY \rightarrow \tau\tau$  events is defined for the

0-, 1-, and 2-jet regions. The data-embedded samples utilized in this analysis cover the events that satisfy the  $e\mu$  triggers, which constitute the majority of the selected events. The remaining  $DY \rightarrow \tau\tau$  events, which contribute through the single lepton triggers (approximately 5% of the total events), are estimated using MC simulation.

The contribution from the WZ and  $W\gamma^*$  backgrounds is simulated, and a data-to-simulation scale factor is derived in a CR with three leptons, as described in Ref. [119]. The  $W\gamma$  process may also contribute as background due to photon conversions in the detector material. This process is estimated using MC simulation and validated against data in a control region that requires events with a leading  $\mu$  and a trailing  $e$  with the same sign and separation in  $\Delta R$  smaller than 0.5. The contribution of triple vector boson production is minor in all channels and is estimated using MC simulation.

### MC Corrections

To enhance the accuracy of the collision data modeling, event-by-event weights are employed for all MC samples. The efficiencies of object identification will be elaborated upon in the subsequent section. Further sample-specific corrections are outlined in section 5, where the associated systematic uncertainties are also presented.

**Pile-up** In order to align the simulated events with the data, the MC pile-up distribution is adjusted by applying weights to match the observed data. Those are determined by scaling the generated number of inelastic collisions within each event with a weight that accounts for the instantaneous luminosity. The luminosity estimation is based on an inelastic pp collision cross-section of 69.2 mb [124].

**Trigger efficiency** The HLT selection is not directly applied to MC samples. Instead, event weights are assigned based on a function that depends on the transverse momentum and pseudorapidity of the two lepton candidates. These weights are determined using the T&P method, where one lepton is designated as the tag and the other as the probe to measure the lepton efficiency. The CMS experiment ensures excellent muon reconstruction, allowing for the convenient derivation of corrections and muon identification and isolation efficiency in  $Z \rightarrow \mu^+\mu^-$  events,

both in data and MC samples. The lepton efficiency is calculated as the ratio of the number of probes satisfying the trigger conditions to the total number of probes. Separate efficiency calculations are performed for muons and electrons using appropriate equations.

The single-lepton trigger efficiencies are combined probabilistically to obtain the double-lepton trigger efficiency, which is then corrected by a factor that accounts for the separation between the leptons ( $\Delta R$ ). The lepton trigger efficiencies are considered independent as long as the isolation cones of the two leptons do not overlap.

To recover efficiency for events where the subleading lepton has a transverse momentum below the threshold of the dilepton trigger, we also use single-lepton triggers. This is possible because the thresholds for single-lepton triggers are higher than the lower thresholds in dilepton triggers, and the identification and isolation requirements are consistent between the two. The efficiency of the combined trigger as a function of the transverse momentum of the subleading lepton is compared between the simulation of the trigger and the calculation using the T&P method. This comparison is shown in Fig. 4.12 for both ggH and WW events with either a leading electron or a leading muon.

**L1 prefiring correction** During the data-taking periods of 2016 and 2017, the forward regions of the ECAL endcap experienced a loss of transparency, causing difficulties in accurately reading out jets, photons, and electrons. As a result, the L1 trigger incorrectly assigned these objects to the previous bunch crossing due to a gradual shift in the timing of the trigger primitives. This led to inefficiencies in the triggering process.

To quantify these inefficiencies, measurements were performed by the CMS Collaboration as a function of the pseudorapidity of high- $p_T$  electrons, photons, and jets. The results of these measurements, which are documented in [125], were centrally provided by the collaboration. In order to account for this effect in the MC samples, event weights were applied.



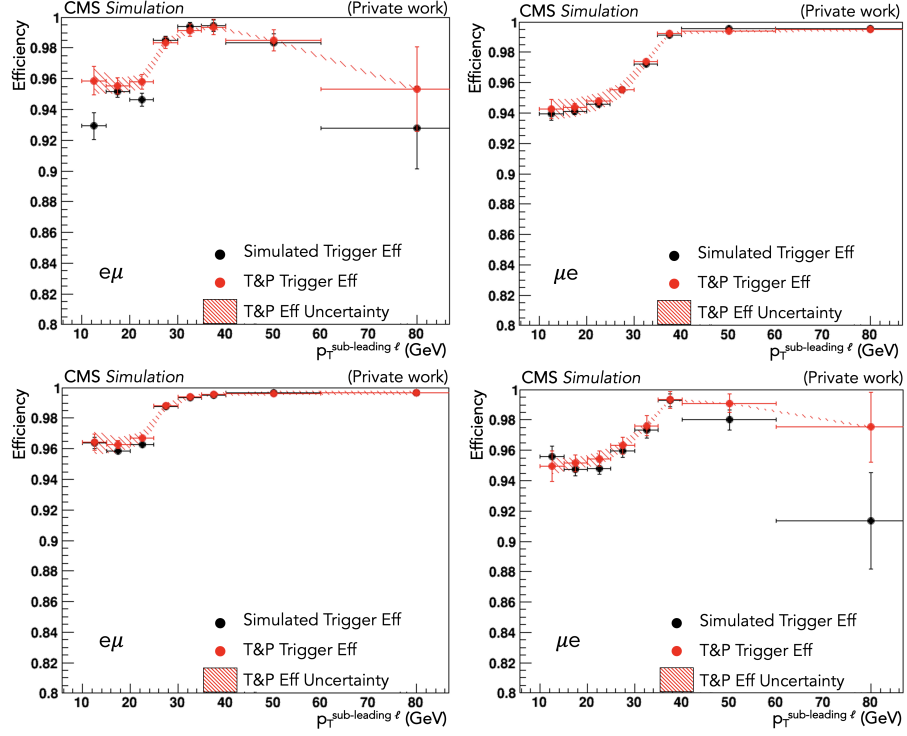


FIGURE 4.12: This figure from [93] displays a comparison of trigger efficiency between the simulated trigger (black points) and the efficiency calculated using the Tag-And-Probe method (red points) as a function of the transverse momentum of the subleading lepton. The plot shows the efficiency calculation for events with one electron-muon pair in ggH (upper row) and WW (bottom row) MC samples, with a leading electron (right column) or a leading muon (left column). Note: this plot does not include  $\Delta R$  corrections described.

### 4.3 Object identification and corrections

In this section, we present further technical details regarding the identification criteria employed to select the physics objects in the analysis. These criteria supplement the information on object reconstruction discussed in section 2.3.1. The following set of criteria is applied to ensure the quality and reliability of the selected muons, electrons, missing transverse energy (MET), AK4 jets, and boosted V-jets.

#### 4.3.1 Muons and Electrons

Dedicated selection criteria and quality requirements are applied to both muons and electrons, as they are essential objects among the final state products.

## Muons

For the selection of muons, the tight ID HWW working point, as summarized in Table 4.8 is required. Additionally, to further improve the sensitivity of the analysis and minimize the presence of fake leptons, an additional criterion is applied. Specifically, the output of the ttHMVA discriminator, which was developed for the ttH analysis, must exceed a threshold of 0.8 [126].

| Muon Tight HWW ID                                       |  |
|---|--|
| $p_T > 10 \text{ GeV}$                                  |  |
| $ \eta  < 2.4$  |  |
| $ d_{xy}  < 0.01 \text{ cm when } p_T < 20 \text{ GeV}$ |  |
| $ d_{xy}  < 0.02 \text{ cm when } p_T > 20 \text{ GeV}$ |  |
| $ d_z  < 0.1$   |  |

TABLE 4.8: Requirements for muons to pass the tight HWW working point.  $d_{xy}$  is the impact parameter of the track with respect to the primary vertex (PV), and  $d_z$  is the longitudinal distance from the track to the PV.

The total efficiency of a muon is determined by the product of several factors:

$$\epsilon_\mu = \epsilon_{\text{ttHMVA} > 0.8} \times \epsilon_{\text{Track}} \times \epsilon_{\text{Trigger}} \times \epsilon_{\text{ID}} \times \epsilon_{\text{ISO}}. \quad (4.5)$$

Although the global track reconstruction efficiency is typically 99% [127] and the trigger efficiency is 98% (as discussed in section 2.3.1), the T&P method is employed to determine the identification and isolation efficiencies. In turn, the efficiency of the Tight HWW ID is evaluated in data and a simulated DY sample, considering its dependence on  $p_T$  and  $\eta$ . The assessment is based on the following simple expression:

$$\epsilon(p_T, \eta) = \frac{N_{\text{Pass}}^{\text{Probes}}}{N_{\text{Pass}}^{\text{Probes}} + N_{\text{Fail}}^{\text{Probes}}}. \quad (4.6)$$

The results for each of the Run 2 years are presented in Fig. 4.13, where the values are extracted from the distributions shown in Fig. 4.14.

The majority of scale factors between data and MC are close to 1, with some deviations of up to 5%. The associated systematic uncertainty on the HWWID efficiency typically remains below 1%, reaching a maximum of 6%. Similarly, the isolation efficiency is determined in a comparable manner. In turn, the sum of the energy deposits of all reconstructed particles is considered within a cone of radius  $\Delta R$  around the muon. For muons with  $p_T > 20 \text{ GeV}$ , the selection criterion for ISO (isolation) is set to be  $\text{ISO} < 0.15$  within a cone of  $\Delta R = 0.4$ , resulting in

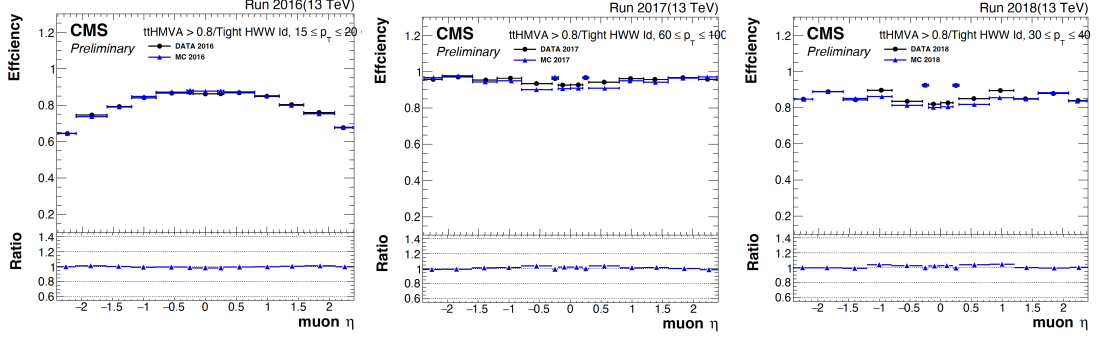


FIGURE 4.13: Efficiencies of the Tight HWWID selection for muons as a function of  $\eta$  in 2016 (left), 2017 (center), and 2018 (right) data sets are shown with different  $p_T$  ranges specified in each panel. The efficiencies are presented for both data and MC events, and the ratio of the two is used as a correction factor. The results are obtained using the TP method. The figure is taken from [93].

an efficiency of 95%. The isolation should be defined as relative to the muons  $p_T$ ; otherwise, this criterion is not meaningful. A looser  $ISO < 0.4$  selection is employed to investigate the non-prompt lepton background while maintaining the ID criteria.

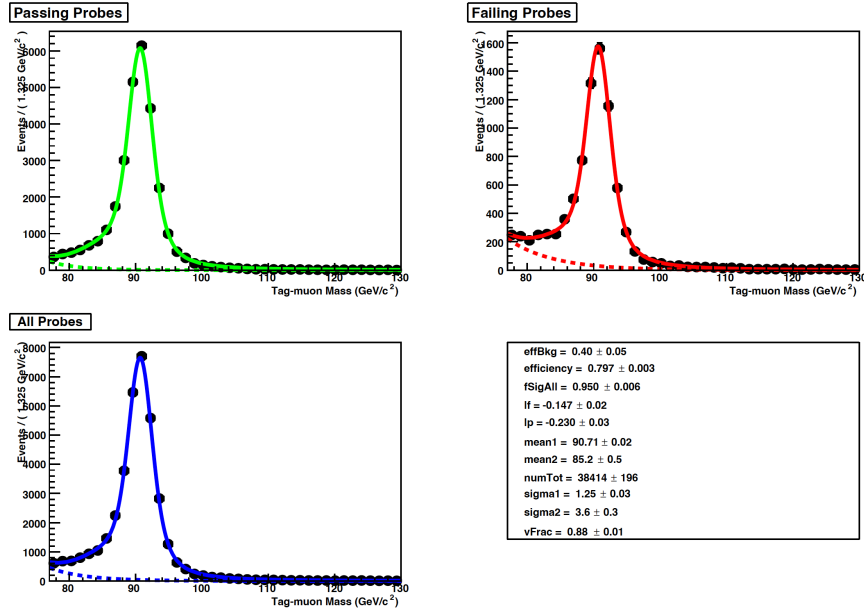


FIGURE 4.14: Performance of the Tight ISO/TightHWW ID fit in data for muons with a transverse momentum ranging from 15 GeV to 20 GeV and an absolute pseudorapidity between 1.2 and 1.6. Extracted from [93].

To reduce biases in the muon momentum assignment arising from detector misalignment, magnetic field precision, or issues in muon reconstruction, experimental scale and momentum corrections from the Rochester group [128] are employed. The net effect of these corrections is visible as a shift in the  $m_{\ell\ell}$  distribution in Fig. 4.15, which leads to a very good data/MC agreement.

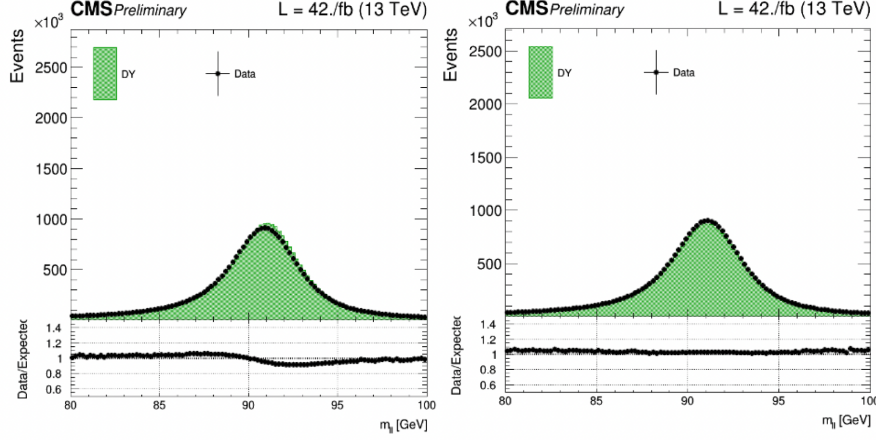


FIGURE 4.15: The left plot shows the  $\mu^+\mu^-$  invariant mass distribution in a DY control region without applying any corrections, while the right plot shows the same distribution after applying the 2017 Rochester corrections. Taken from [93] analysis.

## Electrons

In order to conduct the analysis, the electrons undergo a selection process using the stringent HWW working point, listed in Table 4.9). Same as for muons, in order to reduce the fake lepton contamination, the selection is also based on a multivariate approach (MVA) based on Boosted Decision Tree (BDT) which is anticipated to achieve a 90% efficiency rate. Additionally, the electron isolation (outlined in section 2.3.1) should be below 0.06.

| Electron Tight HWW ID   |
|---|
| $p_T > 10 \text{ GeV}$  |
| $ \eta  < 2.5$  |
| $ d_{xy}  < 0.05 \text{ cm}$ and $ d_z  < 0.1 \text{ cm}$ if $ \eta  < 1.479$ |
| $ d_{xy}  < 0.1 \text{ cm}$ and $ d_z  < 0.2 \text{ cm}$ if $ \eta  > 1.479$  |

TABLE 4.9: Requirements for electrons to pass the tight HWW working point.  $d_{xy}$  is the impact parameter of the track with respect to the primary vertex (PV), and  $d_z$  is the longitudinal distance from the track to the PV.

Like the muon selection, the efficiency of the electron selection is assessed through the T&P method using data. The selection criteria are labeled as `mva_90p_Iso2016` (for 2016) and `Fall17V1Iso_WP90` (for 2017 and 2018). The efficiency of the former working point is shown in Fig. 4.16. In turn, the electron energy scale is studied with  $Z \rightarrow ee$  [81].

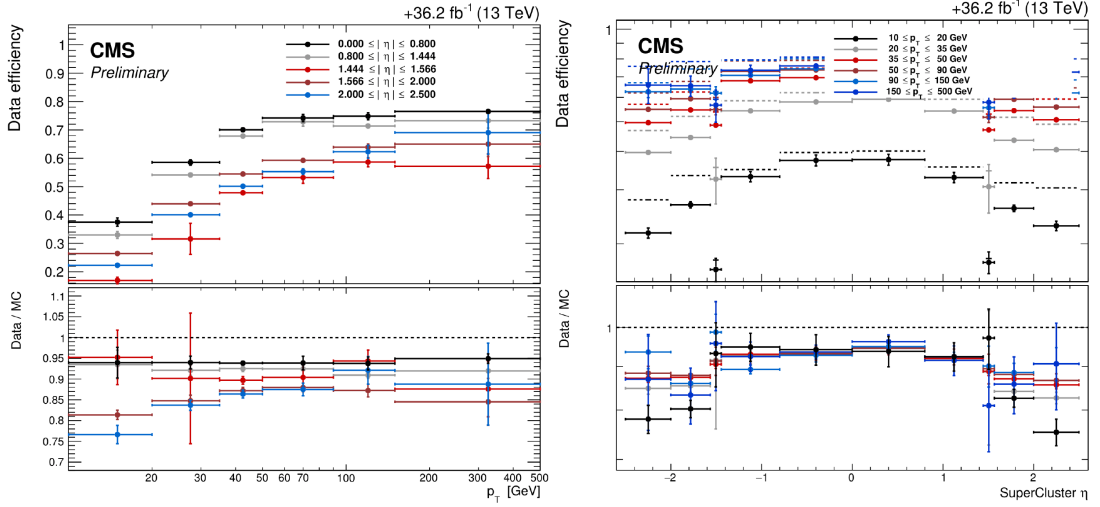


FIGURE 4.16: The figures depict the trigger efficiency results obtained from data using the `mva_90p_Iso2016` criteria. The distributions are adopted from [93].

### Lepton identification efficiency correction

The scale factors for data over MC account for the different identification efficiencies of muons and electrons, which are evaluated independently. The ultimate event weight, varying with lepton  $p_T$  and  $\eta$ , is computed by multiplying the resulting scale factors obtained using the expression:

$$SF = SF(\text{"ID"}) \times SF(\text{"ISO"}) \times SF(\text{"others"}) \quad (4.7)$$

### 4.3.2 Jets

The jets used in the analysis are AK4 PF jets identified by the anti-kT algorithm (see section 2.3.1). Charged hadron subtraction (CHS) is applied, which associates all charged hadrons composing the jet with a track originating from the primary vertex. Jets satisfying the criteria of  $p_T > 30$  GeV and  $|\eta| < 4.7$  and

passing the tight jet identification prescribed by the Jet Identification group are considered. Additionally, jets with  $p_T < 50$  GeV and passing the loose pileup jet identification [129] are also considered.

For the 2017 and 2018 campaigns, jets with  $p_T > 30$  GeV and  $|\eta| < 4.7$  that pass the tight jet ID, as defined by the Jet Identification group, undergo JER (jet energy resolution) smearing to enhance the agreement between data and MC. This correction addresses the issue of overestimated jet energy resolution during those campaigns. Furthermore, the PUPJetID is applied to filter out contaminated jets. It utilizes a BDT that separates hard-scattering jets from pileup jets based on the jet multiplicities and geometry. When tuned in bins of  $\phi$  or  $\eta$ , the BDT achieves a 99% efficiency in identifying hard scattering jets. The PUPJetID is only applied to jets with  $p_T < 50$  GeV.

### **b-tagged jets**

Moreover, in the selection of b-tagged jets, introduced in section 2.3.1, the loose working point is employed due to its superior ability to reject non-b jets. However, for the VBF channel, the DeepJet algorithm surpasses the DeepCSV algorithm in terms of signal significance and the uncertainty in signal strength [130]. For each tagger, the discriminator score is calculated for jets with  $p_T > 20$  GeV and  $|\eta| < 2.4$ , and it quantifies the probability of a jet originating from a b quark.

To enhance the agreement between the simulated and observed data distributions, we make use of a scale factor per event. This scale factor depends on variables such as the jet  $p_T$ ,  $\eta$ , flavor, and the b-tagging discriminator value. In this analysis, we apply this scale factor to the MC distributions. For each data period, a discriminator cut referred to as the “loose working point” is defined. This cut is carefully chosen to achieve a 90% efficiency rate for identifying b-jets while keeping the rate of misidentifying a light jet as a b-tagged jet at approximately 10%. The effectiveness of the veto is shown in Fig. 4.17 using green markers.

### **4.3.3 Boosted V-jets**

The production rate of heavy particles may increase with respect to that expected in the SM in scenarios involving BSM physics. Boosted topologies emerge from the high transverse momentum of these particles, leading to the merging of jets,

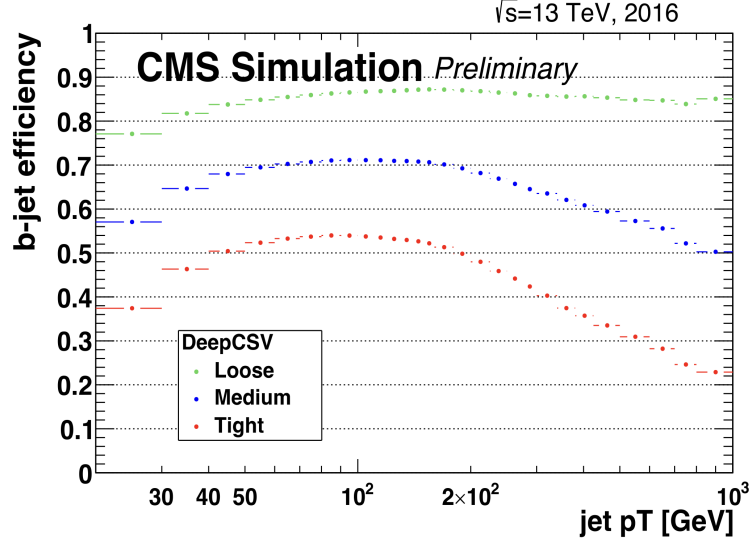


FIGURE 4.17: The figure displays the efficiency of b-jets plotted as a function of the jet transverse momentum using the DeepCSV algorithm. These efficiencies were determined based on simulated  $t\bar{t}$  events, where the jets had  $p_T > 30$  GeV and fell within the tracker acceptance region. The final bin includes any overflow events. Source: [131].

which is particularly relevant for the VH channel. In this study, an additional boosted category is introduced alongside the resolved VH category, resulting in more stringent constraints on the results, as discussed in chapter 6. Resolved decays of a given particle  $X$  are characterized by associated transverse momentum much smaller than its mass, while boosted decays exhibit the opposite behavior. As depicted in Fig. 4.18, in one case, the quark decays are resolved and are reconstructed with 2 jets ( $p_T$  less than  $m_x$ ), while in the boosted decays ( $p_T$  greater than  $m_x$ ) the quark decays are collimated such that a single large radius jet is more appropriate for reconstruction. Then substructure can be used to identify two subjects.

To establish a connection with section 2.3.1, it is important to note that boosted V-jets are a specific case within the anti- $k_T$  algorithm, characterized by a radius parameter of 0.8, which corresponds to the AK8 jet designation.

In the context of hadronic V decays, particularly in the VH production mode, the AK8 algorithm is employed to reconstruct single merged jets when the decay products are highly collimated. The presence of an HVV vertex may lead to anomalous effects that increase the fraction of boosted V bosons. By incorporating jet grooming and substructure techniques, the inclusion of the boosted VH category allows for a more aggressive suppression of background events compared to the resolved

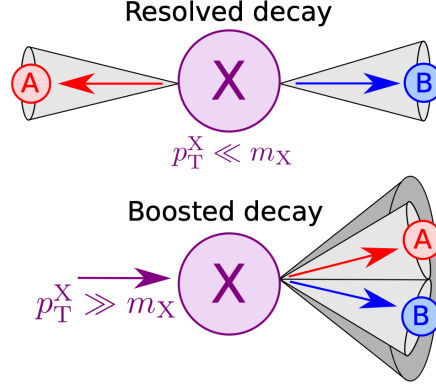


FIGURE 4.18: Boosted V jet reconstruction with large radius jets.

scenario. This addition of the boosted VH category significantly enhances the sensitivity of the analysis to potential anomalous HVV coupling effects.

For the purposes of this analysis, AK8 jets must satisfy specific kinematic requirements. They are required to have a transverse momentum  $p_T > 200$  GeV and fall within the tracker acceptance range of  $|\eta| < 2.4$ . Moreover, they need to pass the loose jet identification criteria. To mitigate the impact of pile-up and underlying event radiation, the soft drop grooming algorithm [132] is employed. This algorithm systematically breaks down the jet and iteratively removes low-momentum, wide-angle radiation. As a result, the mass values for QCD-jets decrease, while the mass of jets originating from the V-boson decay is preserved. This behavior is shown in Fig. 4.19 from Ref. [133], which examines the jet substructure at the LHC.

The compatibility between AK8 jets and a substructure hypothesis involving subjets, which is based on the energy flow aligned along N subjet axes, is assessed using the n-subjettiness variable [134]. In this analysis, the AK8 jets under consideration are associated with two subjets corresponding to the proto-jets obtained by reversing the last step of the soft drop jet declustering procedure. The ratio of 2-subjettiness to 1-subjettiness, denoted as  $\tau_2/\tau_1$ , allows for the identification of signal jets while suppressing the contribution from QCD jets.

Figure 4.20 displays the distribution of  $\tau_{21}$  for large-R jets, specifically in the context of a high-mass VH resonance search conducted at CMS using the 2016 dataset. This plot provides insights into the discriminative power of the  $\tau_{21}$  variable in the analysis. During the exploration of diboson resonances, it has been observed that the soft drop mass of the fat jet and its subjets is influenced by the jet transverse



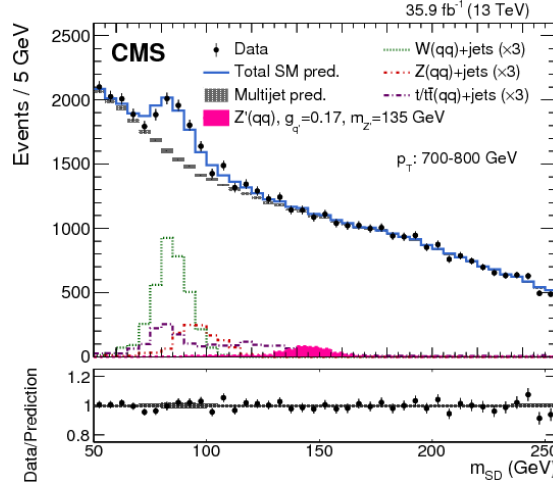


FIGURE 4.19: This figure presented in [133] displays the distribution of Soft drop jet mass for jets reconstructed using the anti- $k_t$  algorithm with a radius parameter of  $R = 0.8$ . Both data and dominant background processes are included in the plot, showcasing the comparison between their respective mass distributions.

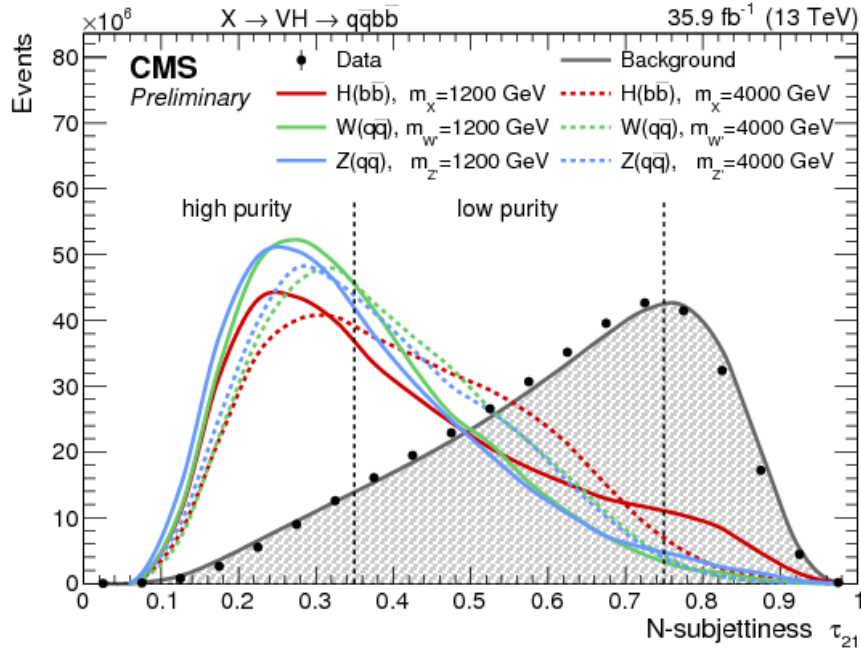


FIGURE 4.20: The figure illustrates the  $\tau_{21}$  distribution of the selected jets for data, background, and signal samples, following the application of kinematic cuts. The distributions have been scaled to match the observed number of events in the data. The dotted vertical lines in the plot represent the boundaries of the different categories as defined in the accompanying text [135].

momentum. To account for this effect, specific mass corrections, tailored to the jet  $p_T$ , are implemented. These corrections are evaluated centrally by the JMAR

(Jet MET Algorithms Reconstruction) group in CMS [136]. By applying these  $p_T$ -dependent mass corrections, the analysis aims to improve the accuracy of the reconstructed jet masses and enhance the overall sensitivity to diboson resonance signals.

#### 4.3.4 MET

To compute the MET in this analysis, the Pile-Up Per Particle Identification (PUPPI) algorithm is employed. This algorithm has demonstrated superior resolution and better agreement between data and MC simulations [137]. PUPPI assigns weights to individual particles based on the likelihood of their origin from either a primary vertex or a PU vertex, thus facilitating PU mitigation [138].

Filters are applied to reject anomalous high- $p_T^{\text{miss}}$  events. Figure 4.21, sourced from Ref. [139], presents the effect of recommended event filters on the distribution of  $p_T^{\text{miss}}$  for a dijet sample. The figure highlights that events with anomalously high  $p_T^{\text{miss}}$  are primarily attributed to electronic noise present in the calorimeters.

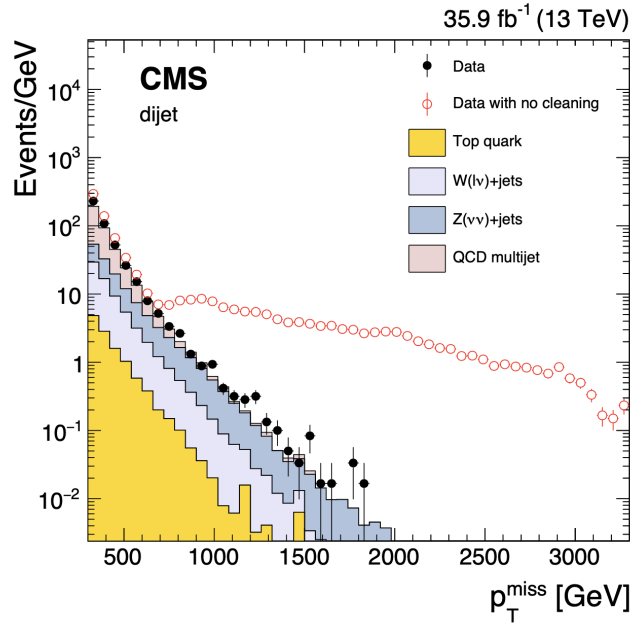


FIGURE 4.21: Comparison of the  $p_T^{\text{miss}}$  distributions for dijet events that satisfy the selection criteria, including jet identification requirements. The distribution includes events after the application of recommended event filtering algorithms (represented by filled markers), as well as events before the event filtering algorithms (represented by open markers). Simulated events are also depicted by solid histograms for reference. Sourced from [139]

## 4.4 Event selection

This section focuses on event selection and categorization, following the introduction of the analysis objects and their corresponding corrections. To be selected, an event must satisfy specific criteria. The leading lepton is required to have a  $p_T$  greater than 25 GeV, while the sub-leading lepton must have a  $p_T$  greater than 13 GeV. However, for the 2016 dataset, the  $p_T$  threshold for the sub-leading muon is lowered to 10 GeV as the lower threshold of the HLT trigger allowed that. Events with additional leptons having  $p_T > 10$  GeV are not included in the selection. Further requirements are imposed on the dilepton system. The invariant mass of the lepton pair,  $m_{\ell\ell}$ , is required to be greater than 12 GeV, and the transverse momentum of the dilepton system,  $p_T^{\ell\ell}$ , must exceed 30 GeV. In addition to the dilepton requirements, there is an additional criterion for the MET, denoted as  $p_T^{\text{miss}}$ . A transverse mass requirement of  $m_T^H > 60$  GeV is applied. Respectively,  $m_T^H$  and  $m_T^{\ell_2}$  are defined as:

$$m_T^H = \sqrt{2p_T^{\ell_1\ell_2}p_T^{\text{miss}}[1 - \cos \Delta\Phi(\vec{p}_T^{\ell_1\ell_2}, \vec{p}_T^{\text{miss}})]}, \quad (4.8)$$

$$m_T^{\ell_2} = \sqrt{2p_T^{\ell_2}p_T^{\text{miss}}[1 - \cos \Delta\Phi(\vec{p}_T^{\ell_2}, \vec{p}_T^{\text{miss}})]}. \quad (4.9)$$

Such a requirement serves two purposes: it helps suppress the DY background and ensures orthogonality with the anomalous coupling analysis in the  $H \rightarrow \tau\tau$  channel. Moreover, the cut on  $m_T^H < 125$  GeV allows for future combination with the off-shell HWW analysis while not affecting the sensitivity of the analysis. To maintain orthogonality with the non-resonant control region used in the off-shell  $H \rightarrow ZZ \rightarrow 2\ell 2\nu$  analysis [52],  $m_{\ell\ell}$  values within the range of [76.2, 106.2] GeV are excluded, enabling a future combination. It should be noted that this exclusion has minimal impact on this analysis since there are typically no signal events in this region. Finally, any event with b-tagged jets with  $p_T > 20$  GeV is vetoed to ensure the reduction of the top background. The base selection criteria are summarized in Table 4.10.

### 4.4.1 Event categorization

#### Production based categories

We categorize the analysis into different production-based categories based on the vertex under study, namely HVV or Hgg.

TABLE 4.10: Base selection criteria

---

|  |
|--|
| 2 leptons ( $e\mu$ ) of opposite charge                                    |
| $p_T^{\ell 1} > 25 \text{ GeV}$  |
| $p_T^{\ell 2} > 13 \text{ GeV}$ (10 GeV in 2016)                           |
| $\eta_\ell < 2.5$  |
| No additional lepton with $p_T > 10 \text{ GeV}$                           |
| $m_{\ell\ell} > 12 \text{ GeV}$  |
| $p_T^{\ell\ell} > 30 \text{ GeV}$  |
| $E_T^{\text{miss}} > 20 \text{ GeV}$                                       |
| $m_T^{\ell 2} > 30 \text{ GeV}$  |
| b-jet veto   |
| $m_T^H > 60 \text{ GeV}$ and $m_T^H < 125 \text{ GeV}$ (Orthogonality)     |
| $m_{\ell\ell} < 76.2 \text{ GeV}$ or $> 106.2 \text{ GeV}$ (Orthogonality) |

---

**HVV analysis** To target the ggH, VBF, and VH production processes associated with the HVV coupling, we apply exclusive selection criteria based on the level of jet activity in the event. Specifically, for the ggH categories, we require either zero or one AK4 jet in the event. In the VBF and resolved VH categories, we require two AK4 jets with a dijet mass ( $m_{jj}$ ) greater than 120 GeV and within the range of 60 GeV to 120 GeV, respectively. In the boosted VH category, the presence of a V-tagged AK8 jet is required, which has a soft drop mass ( $m_j^{\text{AK8}}$ ) between 65 GeV and 105 GeV, and satisfies the condition  $\tau_2/\tau_1 < 0.4$  [136]. To ensure orthogonality, an AK8 V jet veto is implemented in the other categories. Table 4.11 provides a summary of these production categories for the HVV coupling.

TABLE 4.11: Overview of ggH, VBF, and VH categories for the HVV analysis.

---

| Variable   | ggH | VBF                 | Resolved VH  | Boosted VH |
|------------|-----|---------------------|--------------|------------|
| AK8 V-jets | 0   | 0                   | 0            | $> 0$      |
| AK4 jets   | 0&1 | 2                   | 2            | -          |
| $m_{jj}$   | -   | $> 120 \text{ GeV}$ | 60 – 120 GeV | -          |

---

**Hgg analysis** For the Hgg coupling, which is treated independently, we adopt the same selection criteria as the VBF category described earlier, with a requirement of  $m_{jj} > 120$  GeV. This condition enhances events with a VBF-like topology, which are more sensitive to anomalous Hgg coupling effects when considering a 2-jet ggH category. The decay vertex does not exhibit sensitivity to anomalous Hgg effects, and therefore, variables based on the decay are not examined in the final Hgg analysis. This allows for a relatively strict selection of  $m_{\ell\ell} < 55$  GeV, which aids in background suppression. In addition, we include the 0-jet and 1-jet ggH categories. Although these events are not particularly sensitive to anomalous Hgg coupling effects, they provide constraints on the ggH signal strength. The categories for the Hgg coupling are summarized in Table 4.12.

TABLE 4.12: Overview of the ggH categories for the Hgg coupling.

| Variable       | ggH | ggH + 2 jets |
|----------------|-----|--------------|
| AK4 jets       | 0&1 | 2            |
| $m_{jj}$       | -   | $> 120$ GeV  |
| $m_{\ell\ell}$ | -   | $< 55$ GeV   |

### Control regions

Control regions (CRs) are established based on the base selection criteria along with additional requirements outlined in Table 4.13. These CRs serve the purpose of validating the background modeling and estimating the number of background events in the signal region.

TABLE 4.13: Control region requirements

| Variable              | DY $\tau\tau$             | top        | WW                         |
|-----------------------|---------------------------|------------|----------------------------|
| $m_{ll}$              | $> 40$ GeV and $< 80$ GeV | $> 50$ GeV | $> 106.2$ GeV              |
| $m_T^H$               | $< 60$ GeV                | -          | $> 60$ GeV and $< 125$ GeV |
| $m_T^{l2}$            | -                         | $> 30$ GeV | $> 30$ GeV                 |
| $b$ -jet veto         | yes                       | no         | yes                        |
| inverse $b$ -jet veto | no                        | yes        | no                         |

The “DY $\tau\tau$ ” CR is designed to target events originating from the Drell-Yan process  $Z \rightarrow \tau\tau$ , where both  $\tau$  decay leptonically, resulting in the  $e\mu$  final state. In turn, the top CR aims to enhance events with one or more top quarks decaying to a W boson and an associated b quark. Both of these CRs are further

divided into separate categories based on the number of jets (0-jet, 1-jet, and 2-jet) in the signal regions. In addition to these CRs, there is an additional CR specifically designed to enhance the contribution from the WW background. This CR is utilized in the 2-jet signal region. All three control regions play a crucial role in the final data fit by constraining the normalizations of the DY, top, and WW backgrounds. Furthermore, we define additional control regions, namely the boosted-DY $\tau\tau$ , boosted-top, and boosted-WW CRs, which require the presence of an AK8 V-jet candidate. These CRs serve the purpose of validating the background modeling, specifically in the boosted V jet category. However, due to limited statistical power, these CRs alone do not provide significant constraints on background normalization. Therefore, we rely on the 2-jet CRs to determine the normalizations used in the boosted VH category. We validate this approach based on the observed agreement between data and MC simulations in the boosted CRs when using background normalizations determined from the 2-jet CRs. Additionally, the  $p_T^{jj}$  distribution in the 2-jet CRs shows a good agreement between data and simulations. Figure 4.22 quantifies this agreement, where the AK8 V-jet

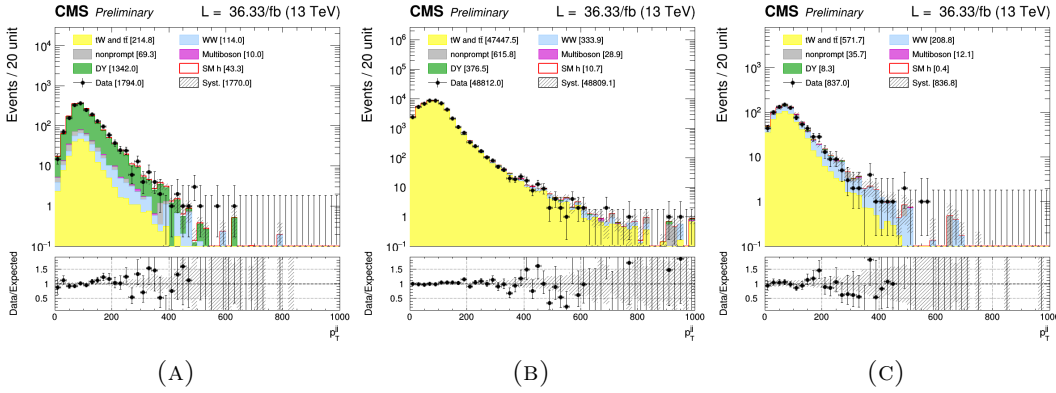


FIGURE 4.22:  $p_T^{jj}$  distribution in the DY $\tau\tau$  (A), Top (B), and WW 2-jet (C) CRs for the 2016 dataset.

veto used in the signal region is not applied. Notably, for  $p_T^{jj} > 200$  GeV, there is considerable overlap with the high  $p_T$  phase space considered in the boosted VH categories. The overall agreement between data and MC simulations across the entire  $p_T$  range further justifies the adoption of a common normalization strategy.

## 4.5 Observables and kinematic discriminants

The kinematics of the particles that are produced in the Higgs boson decay or in associating with the Higgs boson production are sensitive to AC. Specifically, this analysis aims to investigate the effects of anomalous couplings on the HVV and Hgg vertexes by studying:

1. kinematics of two quark jets from VBF and VH production (HVV);
2. kinematics of two quark jets from ggH + 2 jet production (Hgg);
3. kinematics of  $H \rightarrow WW$  decay products (HVV).

The analysis concentrates on the couplings of a spin-zero H boson, as any exotic H boson with a non-zero spin has already been excluded [140]. The considered production topologies for VBF, VH, and ggH are depicted in Fig. 4.23.

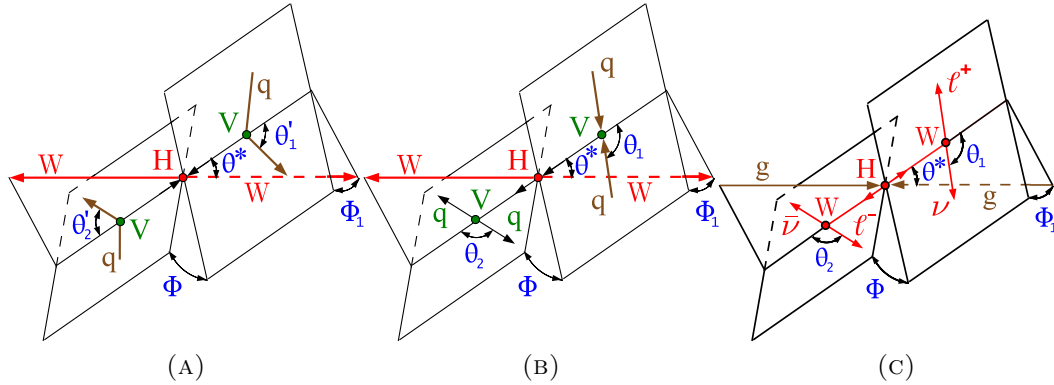


FIGURE 4.23: Topologies of the production and decay for vector boson fusion  $qq' \rightarrow qq'H$  (A),  $q\bar{q}' \rightarrow VH$  (B), and gluon fusion with decay  $gg \rightarrow H \rightarrow 2\ell 2\nu$  (C). The color-coded representation of particles in the figure is as follows: the incoming particles are depicted in brown, the intermediate vector bosons and their fermion daughters are illustrated in green, the Higgs boson and its vector boson daughters are represented in red, and the angles that describe the kinematic distributions are denoted in blue.

The five angles illustrated for the VBF and VH productions combined with the momentum transfer of the vector bosons provide a complete set of kinematic information about the production and decay of the Higgs boson.

This analysis focuses on extracting complete production kinematic information from candidate events of VBF, VH, and 2-jets ggH using discriminants derived

from matrix element calculations with the MELA package. The MELA approach, cited as a reference in [114], is employed to minimize the number of observables while retaining all essential information.

**Production-based discriminants** To construct the kinematic discriminants based on *production* in MELA, we use the kinematic information from the 2-jet system to reconstruct the 4-vector of the associated production particles, as well as the two leptons and the MET to approximate the Higgs boson 4-vector. Directly measuring this vector is challenging due to the presence of two neutrinos in the final state. Therefore, we employ a proxy method to build the Higgs boson 4-vector. To estimate the  $p_x$  and  $p_y$  components of the neutrino pair ( $\nu\nu$ ) system, we use the  $p_x$  and  $p_y$  of the missing transverse energy in the event. The  $p_z$  component of the  $\nu\nu$  system is then set to equal that of the  $\ell\ell$  system, based on the observed correlation between these variables at gen-level <sup>1</sup> for signal events. Finally, the mass of the  $\nu\nu$  system is set to equal that of the mean value observed at the gen-level for signal events. The left plot of Fig. 4.24 displays the 2D distribution of the  $\nu\nu$   $p_z$  -  $\ell\ell$   $p_z$  at gen-level for a SM VBF Higgs, while the gen-level invariant mass of the  $\nu\nu$  system for various VBF Higgs signals and the SM WW background is shown in the right plot. By combining the resulting  $\nu\nu$  4-vector with the measured 4-vector of the di-lepton system, we create a proxy Higgs boson 4-vector. It is important to note that the reconstruction of the proxy Higgs boson 4-vector has minimal impact on the final discriminants, as the MELA probability calculation for the production vertices heavily relies on the kinematics of the associated particles. We have compared the shapes of the signal discriminants with those from the HZZ analysis [46], which fully measures the Higgs decay, and found them to be consistent.

**Decay-based discriminants** In the case of ggH production without associated particles, specifically in the 0-jet or in the 1-jet category, only the HWW decay vertex is sensitive to anomalous effects in the HVV vertex. However, due to the presence of two neutrinos in the final state, it is not feasible to reconstruct the 4-vectors of all the Higgs decay products. The specific version of the MELA framework utilized in this analysis requires four inputs, corresponding to the 4 body decay, to compute matrix elements. Consequently, *decay*-based kinematic discriminants from matrix elements are not used in this analysis. Instead, we rely

---

<sup>1</sup>MC simulation before modeling the detector response on those objects.



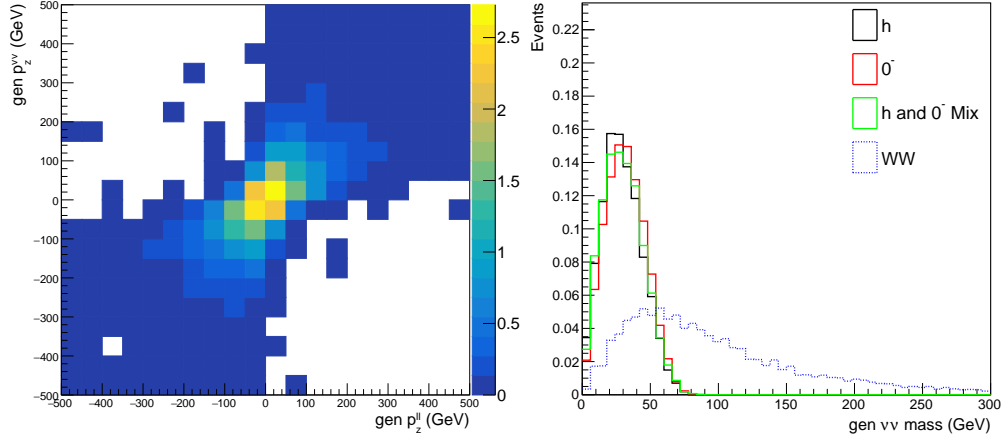


FIGURE 4.24: On the left is the gen-level 2D distribution of  $\nu\nu p_z - \ell\ell p_z$  for a SM VBF Higgs, while on the right is the gen-level invariant mass of the  $\nu\nu$  system. The right panel displays data for a SM VBF Higgs, a pure BSM pseudoscalar VBF Higgs, a combined SM-BSM mixture of these two states, as well as the background from WW processes.

on kinematic variables such as  $m_{\ell\ell}$  and  $m_T^H$  as introduced earlier in this chapter, that are related to the measured final state of the Higgs boson.

#### 4.5.1 HVV: Kinematics in VBF and VH productions

The kinematic distributions of associated particles in VBF and VH production are relevant to quantify AC effects of the Higgs boson in the HVV vertex. To investigate these couplings, a set of seven independent observables are defined in the production process, as depicted in Fig. 4.23, considering a general spin-parity hypothesis.

In the case of a spin-zero object, the variables  $\theta^*$  and  $\Phi_1$  exhibit random distributions. As a result, the electroweak production processes are described by just five variables, as stated in [113]:

- $\Omega^{\text{assoc}} = \{\theta_1^{\text{VBF}}, \theta_2^{\text{VBF}}, \Phi^{\text{VBF}}, q_1^{2,\text{VBF}}, q_2^{2,\text{VBF}}\}$  for VBF.
- $\Omega^{\text{assoc}} = \{\theta_1^{\text{VH}}, \theta_2^{\text{VH}}, \Phi^{\text{VH}}, q_1^{2,\text{VH}}, q_2^{2,\text{VH}}\}$  for VH.

We provide the final state particle 4-vectors as input to MELA where all the information from the observables are combined. Three types of discriminants are formulated for the production processes:

1. A discriminant that separates signal (sig) and background (bkg) production processes:

$$\mathcal{D}_{\text{sig}} = \frac{\mathcal{P}_{\text{sig}}(\vec{\Omega})}{\mathcal{P}_{\text{sig}}(\vec{\Omega}) + \mathcal{P}_{\text{bkg}}(\vec{\Omega})}, \quad (4.10)$$

where  $\mathcal{P}$  represents the probability.

2. A discriminant employed to separate the anomalous coupling BSM process from that of the SM:

$$\mathcal{D}_{\text{BSM}} = \frac{\mathcal{P}_{\text{BSM}}(\vec{\Omega})}{\mathcal{P}_{\text{BSM}}(\vec{\Omega}) + \mathcal{P}_{\text{SM}}(\vec{\Omega})} \quad (4.11)$$

Here, the generic label “BSM” is substituted with the specific anomalous coupling state that is being targeted. For instance, we use  $\mathcal{D}_{0-}$  and  $\mathcal{D}_{0+}$  to represent the  $a_3$  CP-odd and  $a_2$  CP-even couplings, respectively. For the  $\Lambda_1$  couplings, we utilize  $\mathcal{D}_{\Lambda_1}$  and  $\mathcal{D}_{\Lambda_1}^{Z\gamma}$ .

3. A discriminant designed to isolate the contribution of interference between different couplings:

$$\mathcal{D}_{\text{INT}} = \frac{\mathcal{P}_{\text{SM-BSM}}^{\text{int}}(\vec{\Omega})}{\mathcal{P}_{\text{SM}}(\vec{\Omega}) + \mathcal{P}_{\text{BSM}}(\vec{\Omega})} \quad (4.12)$$

The label “int” (or “INT”) denotes the interference between the two coupling contributions. For the  $a_3$  coupling, the label “CP” is typically used since the BSM hypothesis, in this case, assumes a pseudoscalar and the interference discriminant is a CP-sensitive observable. The probabilities ( $\mathcal{P}$ ) are computed from the matrix elements provided by the MELA package and are normalized to yield the same integrated cross-sections in the relevant phase space for each process. This normalization ensures a balanced distribution of events ranging between 0 and 1 for the  $\mathcal{D}_{\text{sig}}$  and  $\mathcal{D}_{\text{BSM}}$  discriminants and between -1 and +1 for the  $\mathcal{D}_{\text{INT}}$  discriminant.

The selected events are categorized into three primary production categories: VBF, resolved VH, and boosted VH. In the VBF and VH categories, the 4-vectors obtained from the two AK4 jets assigned as associated particles are employed as inputs for the MELA probability calculation. Conversely, in the boosted VH category, the 4-vectors from the 2 sub-jets of the V-tagged AK8 jet are utilized instead.

The estimate of the Higgs boson 4-vector is also required for the probability calculation and the subsequent construction of the kinematic discriminants, whose application in each production category will now be discussed.

### Discriminants in VBF channel

1. The  $\mathcal{D}_{\text{VBF}}$  discriminant is formulated according to Eq. 4.10, where  $\mathcal{P}_{\text{sig}}$  represents the probability of the VBF H production hypothesis, and  $\mathcal{P}_{\text{bkg}}$  represents the probability of 2-jet ggH production. The plot (A) in Fig. 4.25 shows the reconstructed level distributions of  $\mathcal{D}_{\text{VBF}}$  in the VBF channel for VBF Higgs, ggH, as well as the WW and top SM backgrounds. The discriminant demonstrates its capability to effectively differentiate the SM backgrounds from the VBF signal process.
2. The  $\mathcal{D}_{\text{BSM}}$  discriminant, as described by Eq. 4.11, is employed to separate AC BSM signals from SM signals. As aforementioned, the specific form of this discriminant may vary depending on the coupling being studied. It may take the label of  $\mathcal{D}_{0-}$ ,  $\mathcal{D}_{0+}$ ,  $\mathcal{D}_{\Lambda 1}$ , or  $\mathcal{D}_{\Lambda 1}^{Z\gamma}$ . As an illustrative example, the plot (B) in Fig. 4.25 presents the  $\mathcal{D}_{0-}$  discriminant for the VBF channel, considering four distinct hypotheses: the SM Higgs ( $F_{ai} = 0$ ), the BSM state ( $F_{ai} = 1$ ), and their mixture incorporating interference with different phases ( $F_{ai} = 0.5$ ,  $F_{ai} = -0.5$ ).
3. The  $\mathcal{D}_{\text{INT}}$  discriminant, as defined by Eq. 4.12, is used for event categorization. The specific form of this discriminant depends on the CP parity of the coupling being studied. It can take the label of  $\mathcal{D}_{\text{INT}}$  (CP-even) or  $\mathcal{D}_{\text{CP}}$  (CP-odd). As an example, the (C) panel in Fig. 4.25 illustrates the  $\mathcal{D}_{\text{CP}}$  discriminant for the Higgs  $0^-$  hypothesis.

**Categorization of events** The  $\mathcal{D}_{\text{BSM}}$  discriminant demonstrates notable discrimination power between a pure SM and BSM state. However, distinguishing between the two mixed SM-BSM states ( $F_{ai} = 0.5$ ,  $F_{ai} = -0.5$ ) becomes challenging due to their similar shapes. To address this issue, we leverage the distinct forward-backward asymmetry associated with  $\mathcal{D}_{\text{CP}} > 0$  and  $\mathcal{D}_{\text{CP}} < 0$ . This observable is sensitive to the interference sign between the CP-even and CP-odd contributions, resulting in an asymmetry in the distribution of events with positive and negative  $\mathcal{D}_{\text{CP}}$  values. Categorization of events based on the interference

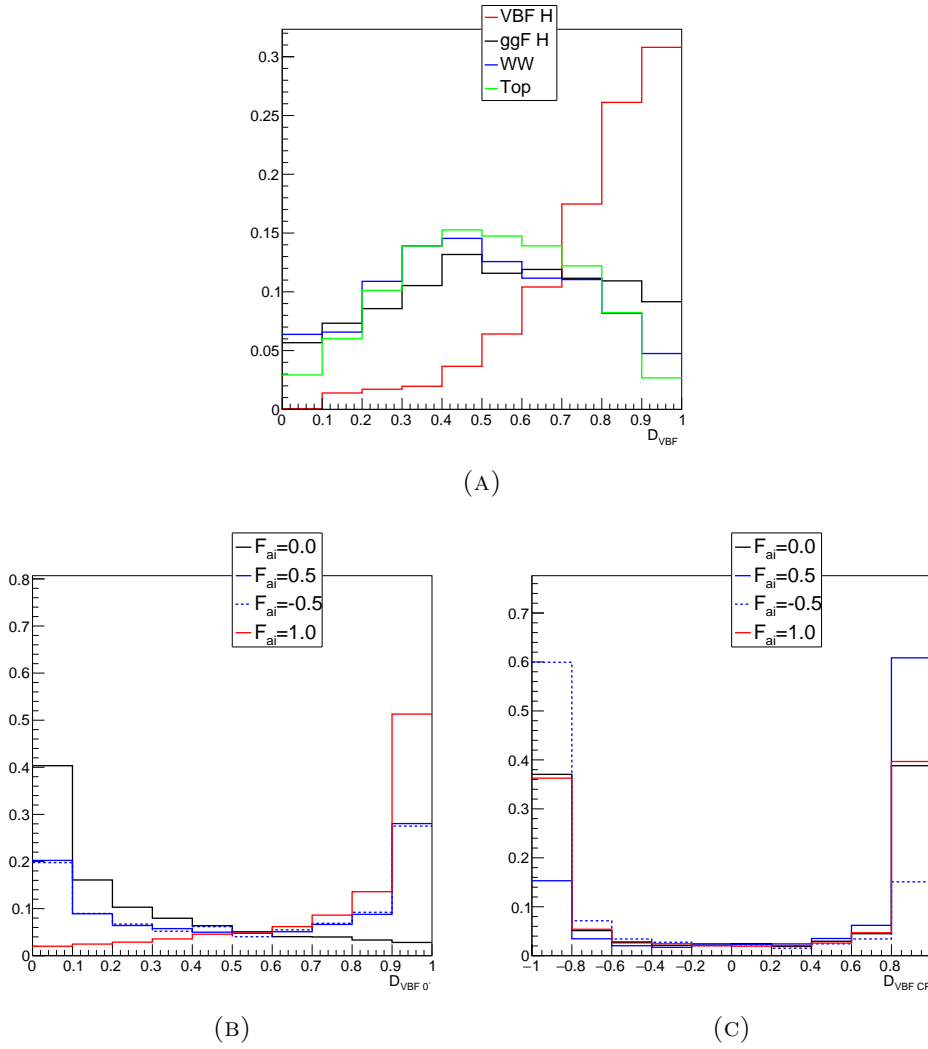


FIGURE 4.25: Discriminants in the VBF channel: Production  $\mathcal{D}_{\text{VBF}}$  (top), BSM  $\mathcal{D}_{0-}$  (bottom-left), and Interference  $\mathcal{D}_{\text{CP}}$  (bottom-right). Various hypotheses are examined, including the SM Higgs ( $F_{ai} = 0$ ), BSM state ( $F_{ai} = 1$ ), and a combination of SM and BSM states ( $F_{ai} = \pm 0.5$ ).

discriminant  $\mathcal{D}_{\text{CP}}$ , which is implemented to enhance sensitivity to any possible asymmetry being present.

Furthermore, the  $\mathcal{D}_{\text{INT}}$  observable also captures the sign of interference between the SM and BSM states for the  $a_2$  coupling. For categorizing events based on this observable, a forward-backward categorization is also employed for this case. The threshold value for  $\mathcal{D}_{\text{INT}}$  is chosen to evenly divide the SM Higgs expectation, which is 0.4. However, for the  $\Lambda_1$  and  $\Lambda_1^{Z\gamma}$  couplings, the interference discriminants exhibit a high correlation with the  $\mathcal{D}_{\text{BSM}}$  discriminants, and therefore, they are not considered in such cases.

**Validation plots in control regions** To ensure the effectiveness of the VBF channel discriminants in real data, we can check that the background estimation gives a good description of these distributions in data CRs. The CRs are used to validate the background description. This approach allows us to check the discriminants without relying on the signal region, thereby avoiding potential biases. Several examples are presented in section 4.5.4 for the  $top$ ,  $DY\tau\tau$ , and  $WW$  control regions using the 2016 dataset. Overall, a good agreement is observed between the data and MC simulations in all distributions, with the error band representing the combined statistical and systematic uncertainties. This is also true for the 2017 and 2018 datasets whose validation plots are gathered in Appendix B.

### Discriminants in the Resolved VH and boosted VH channels

In the resolved and boosted VH categories, the production discriminants specific to those categories do not exhibit a significant level of separation from  $ggH$  production or SM backgrounds. This is attributed to the relatively stringent selection criteria, which restrict the phase space to VH-like events. Consequently, the production discriminants (Eq. 4.10) are not used in the VH channels.

**Resolved VH channel** One of the  $\mathcal{D}_{BSM}$  discriminants calculated at the reconstruction level for the resolved VH channel is shown in plot (A) of Fig. 4.26. It corresponds to the  $\mathcal{D}_{0-}$  discriminant (Eq. 4.11), while the (B) panel shows the discrimination power of the  $\mathcal{D}_{CP}$  (Eq. 4.12) for the interference signals. In this channel, a V-tagged jet veto is applied to ensure orthogonality with the boosted VH category.

**Boosted VH channel** Figure 4.27 presents one of the reconstruction level  $\mathcal{D}_{BSM}$  discriminants,  $\mathcal{D}_{0-}$ , and  $\mathcal{D}_{CP}$  as calculated according to Eq. 4.11 and Eq. 4.12, respectively. The boosted VH events included in this category exhibit high- $p_T$  BSM-like V candidates. Consequently, the  $\mathcal{D}_{BSM}$  discriminants demonstrate a more pronounced peak around a value of 1. However, there remains significant discriminatory capability among various coupling states. To achieve a more uniform distribution between 0 and 1, one could consider readjusting the boosted discriminant by modifying the probability normalizations used. Due to the limited statistical data currently in the boosted category for this analysis, optimizing the discriminant further would not yield substantial benefit. Also, it was found

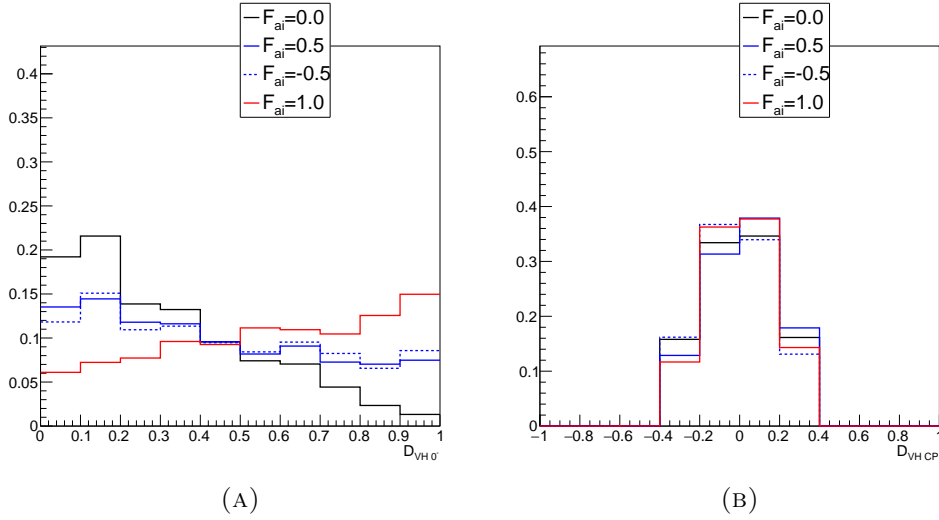


FIGURE 4.26: Discriminants in the Resolved VH channel: BSM  $\mathcal{D}_{0-}$  and Interference  $\mathcal{D}_{CP}$  for different hypotheses: SM Higgs ( $F_{ai} = 0$ ), BSM state ( $F_{ai} = 1$ ), and SM-BSM mixture ( $F_{ai} = \pm 0.5$ ).

that the SM and BSM distributions tended to shift together, which is a reflection of the high- $p_T$  phase space specific to the boosted category. Hence, optimizing the discriminant further would not yield substantial benefits in this case. However, for future analyses in Run 3 and beyond, where larger datasets become available, it would be worthwhile to explore this feature in greater detail.

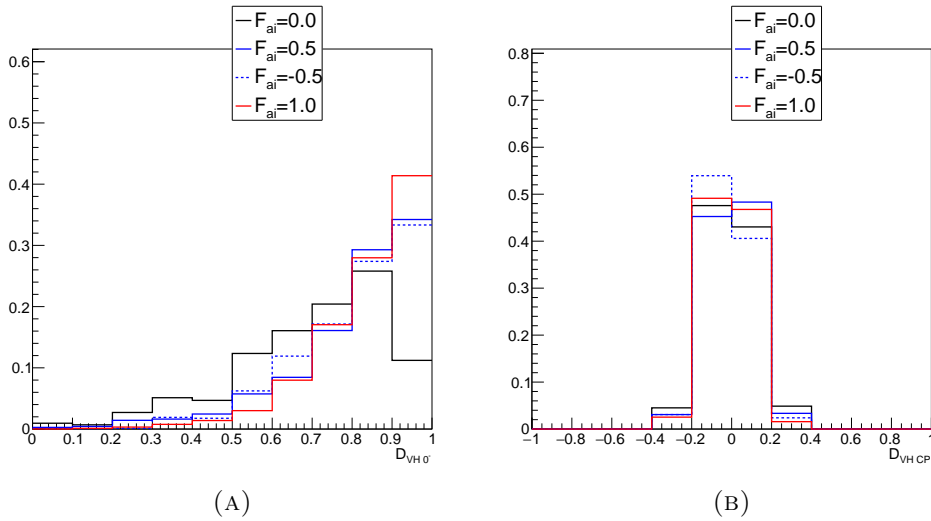


FIGURE 4.27: Discriminants in the Boosted VH channel: BSM  $\mathcal{D}_{0-}$  (A) and Interference  $\mathcal{D}_{CP}$  (B) for different hypotheses: SM Higgs ( $F_{ai} = 0$ ), BSM state ( $F_{ai} = 1$ ), and SM-BSM mixture ( $F_{ai} = \pm 0.5$ ).

**Categorization of events** Both the resolved and boosted VH channels utilize forward-backward categorizations, similar to the VBF channel, employing the  $\mathcal{D}_{CP}$  discriminant (Eq. 4.12) as described earlier.

However, in the case of the  $\mathcal{D}_{0+}$  discriminants, the two mixed states of the SM-BSM exhibit significantly distinct shapes. Additionally, there exists a correlation between  $\mathcal{D}_{INT}$  and  $\mathcal{D}_{0+}$  in these channels. Consequently, the forward-backward categorizations using  $\mathcal{D}_{INT}$  are not implemented in either the resolved or boosted VH channels.

**Validation plots in control regions:** Similar to the VBF channels, the discriminants employed in the VH categories are validated in control regions. We provide several examples of these control regions in Fig. 4.33 (top), 4.34 (DY $\tau\tau$ ), and 4.35 (WW), specifically showing the resolved VH channel (plots A and B) and the boosted VH channel (plots C and D) with the use of the 2016 dataset. These control regions provide valuable insights into the agreement between data and MC simulations. However, it is noticeable that due to the limited statistical data available for boosted VH events, the data-MC agreement is successful but also highlights the fluctuations due to the low statistics in this category.

#### 4.5.2 Hgg: Kinematics in 2-jet ggH production

The kinematic distributions of associated particles in ggH + 2 jets production exhibit sensitivity to the anomalous couplings of the Higgs boson in the Hgg vertex. The analysis follows a similar approach to the analysis of the VBF and VH categories, with a particular focus on the CP-odd  $a_3$  HVV coupling.

##### Discriminants in the 2-jet ggH channel

In this scenario, the optimized observables utilized are  $\mathcal{D}_{0-}^{ggH}$  (Eq. 4.11) and  $\mathcal{D}_{CP}^{ggH}$  (Eq. 4.12), specifically targeting the CP-odd  $a_3$  Hgg coupling. Figure 4.28 displays the distributions of these two discriminants for events in the 2-jet ggH channel, where an additional requirement of  $m_{jj} > 300$  GeV is imposed to enhance events with a VBF-like topology that are more sensitive to CP effects. For each observable, the plot shows the distributions for the SM Higgs ( $F_{ai} = 0$ ), the

BSM state ( $F_{ai} = 1$ ), and their mixture, incorporating interference with various phases ( $F_{ai} = 0.5$ ,  $F_{ai} = -0.5$ ).

**Categorization of events** Similarly to the VBF and VH channels, a forward-backward categorization is implemented in this channel utilizing the  $\mathcal{D}_{CP}^{ggH}$  observable, which is divided into two bins separated at 0.

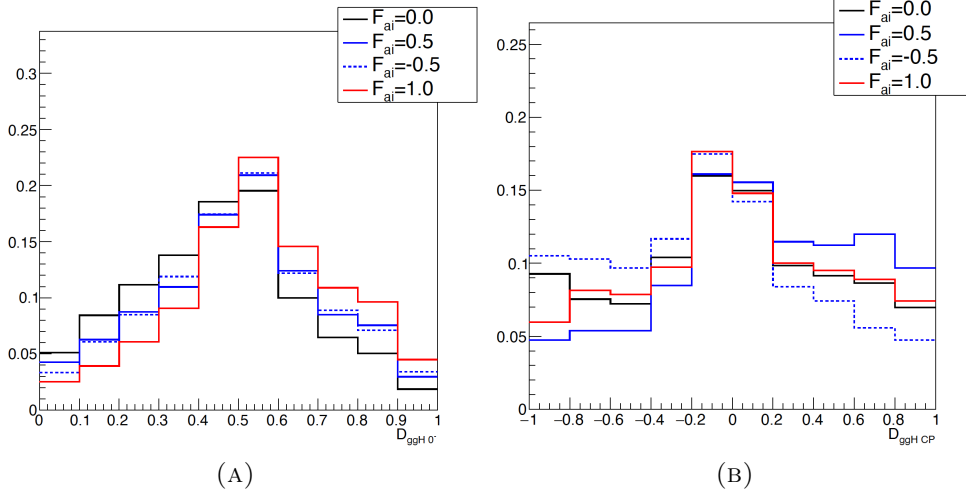


FIGURE 4.28: Discriminants in the 2-jets  $ggH$  channel: BSM  $\mathcal{D}_0^-$  (A) and Interference  $\mathcal{D}_{CP}$  (B) for different hypotheses: SM Higgs ( $F_{ai} = 0$ ), BSM state ( $F_{ai} = 1$ ), and SM-BSM mixture ( $F_{ai} = \pm 0.5$ ).

**Validation plots** The two applied discriminants are also validated in the control regions of top,  $DY\tau\tau$ , and WW, as shown in section 4.5.4. These figures exhibit the comparison between data and MC simulation, highlighting the agreement between the two. The example provided includes results from the 2016 dataset.

### 4.5.3 Kinematics of HWW decay products for HVV vertex

Kinematic distributions associated with the decay of the Higgs boson are sensitive to HVV AC. In the SM HWW analysis [93], two-dimensional templates of the dilepton mass ( $m_{\ell\ell}$ ) and the Higgs transverse mass ( $m_T^H$ ) are employed. These two variables have been identified as the most discriminatory in characterizing the SM Higgs boson.



### Discriminants in the 0-jet and 1-jet ggH channels

In this analysis, we also rely on  $m_{\ell\ell}$ - $m_T$  to gain sensitivity to anomalous HVV couplings at the HWW decay vertex. The reconstructed level  $m_{\ell\ell}$  and  $m_T^H$  variables for events in the 0 jet channel are shown in Fig. 4.29 for the SM Higgs ( $F_{ai} = 0$ ), the BSM state ( $F_{ai} = 1$ ), and their mixture, including interference with different phases ( $F_{ai} = 0.5$ ,  $F_{ai} = -0.5$ ).

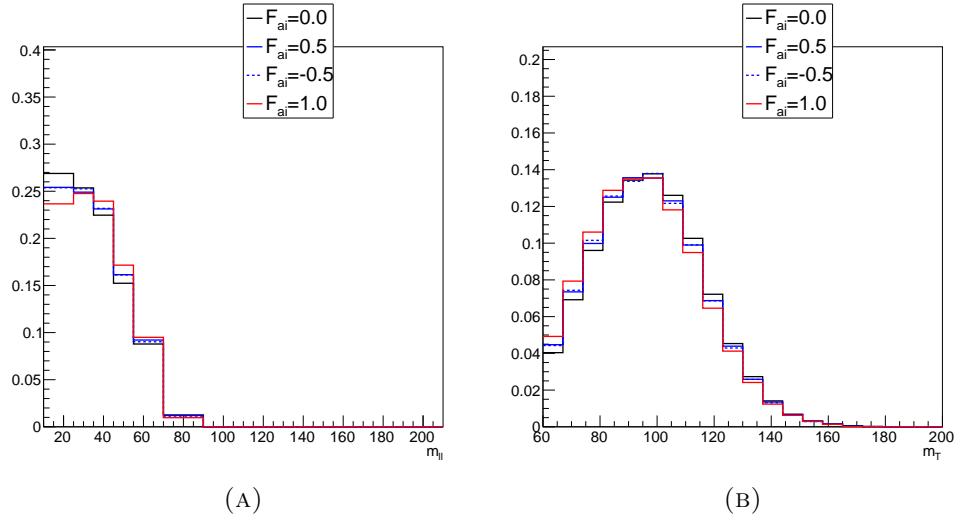


FIGURE 4.29: Discriminants in the 0-jet ggH channel:  $m_{\ell\ell}$  (A) and  $m_T$  (B) for different hypotheses: SM Higgs ( $F_{ai} = 0$ ), BSM state ( $F_{ai} = 1$ ), and SM-BSM mixture ( $F_{ai} = \pm 0.5$ ).

**Categorization of event** In this scenario, a categorization based on the number of jets is implemented.

**Validation plots** In this case, the two applied discriminants are validated in the top and  $DY\tau\tau$  CRs, as depicted in Figs. B.21 and B.22, respectively. These show the comparison between data and MC simulation, highlighting the agreement between the two. It is worth noticing that the WW CR is not utilized in this analysis. In the 0- and 1-jet channels, the normalization parameters are constrained directly in the signal regions, which span the high  $m_{\ell\ell}$  phase space enriched in WW events. The example presented shows results obtained from the 2016 dataset.

#### 4.5.4 Validation in control regions

This section aims to compile the validation plots for the applied discriminants, as mentioned and referred to throughout the preceding discussion. In the  $DY\tau\tau$ , top, and WW control regions, there is a satisfactory level of agreement observed between the background predictions and the data for the entire Run 2 dataset. However, for illustrative purposes, only the results from the 2016 dataset are presented. Figures from 4.30 to B.22 belong to this section.

## VBF channel

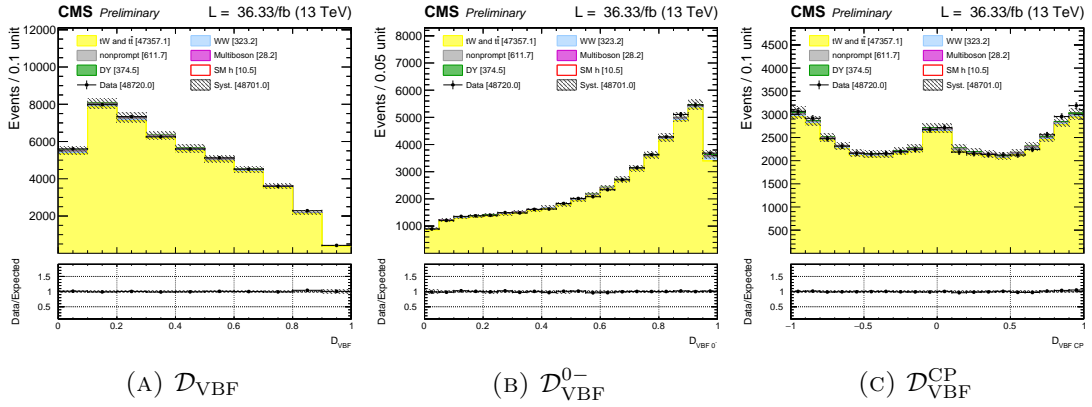


FIGURE 4.30: VBF related distributions in top 2-jet CR for the 2016 dataset.

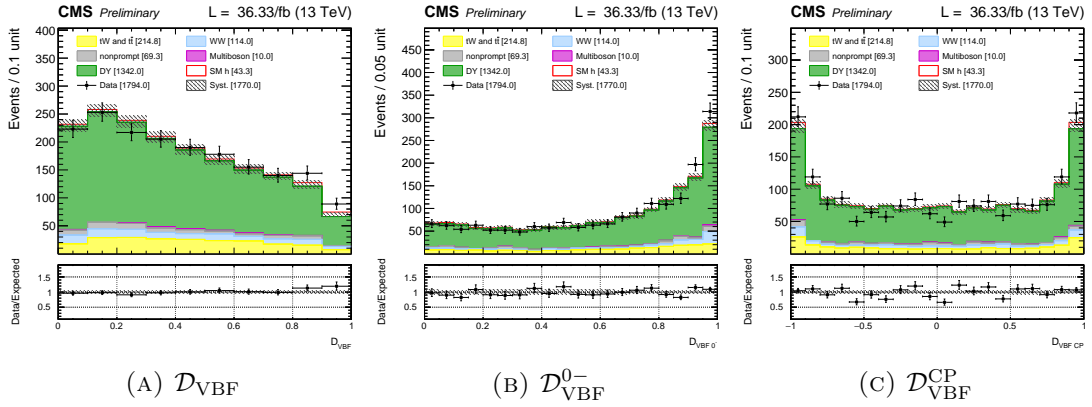
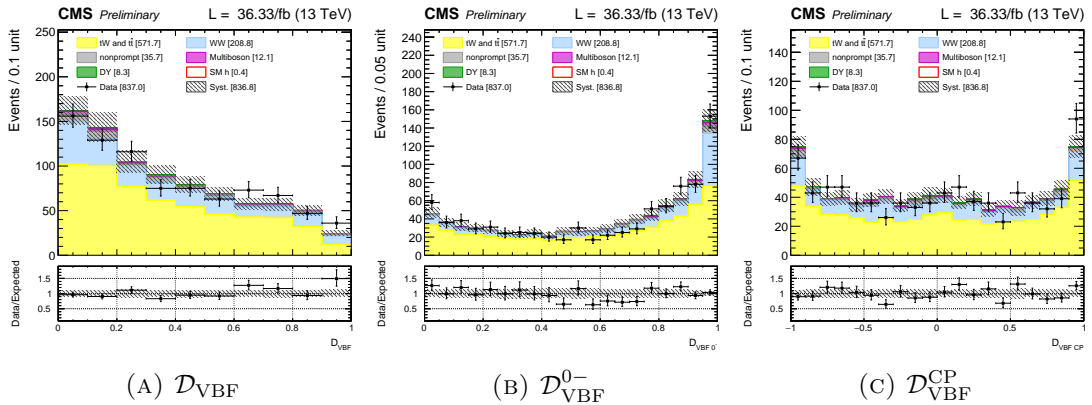
FIGURE 4.31: VBF related distributions in  $\text{DY } \tau\tau$  2-jet CR for the 2016 dataset.

FIGURE 4.32: VBF related distributions in WW 2-jet CR for the 2016 dataset.

## VH channels

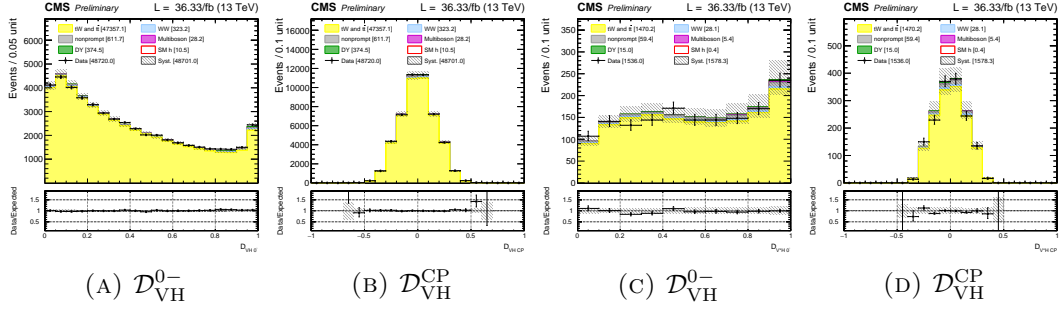


FIGURE 4.33: Resolved VH (A, B) and Boosted VH (C, D) related distributions in top 2-jet CR for the 2016 dataset.

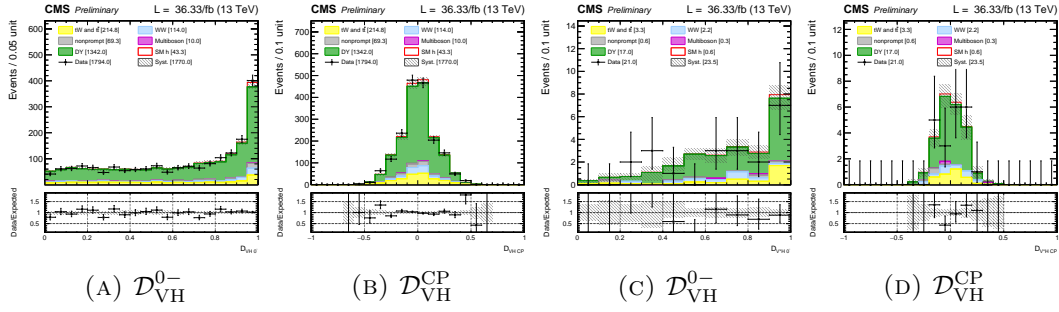


FIGURE 4.34: Resolved VH (A, B) and Boosted VH (C, D) related distributions in  $DY\tau\tau$  2-jet CR for the 2016 dataset.

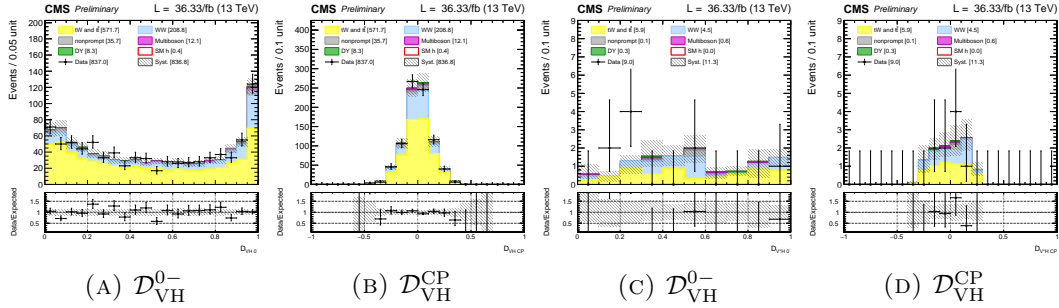


FIGURE 4.35: Resolved VH (A, B) and Boosted VH (C, D) related distributions in WW 2-jet CR for the 2016 dataset.

## 2-jet ggH channel

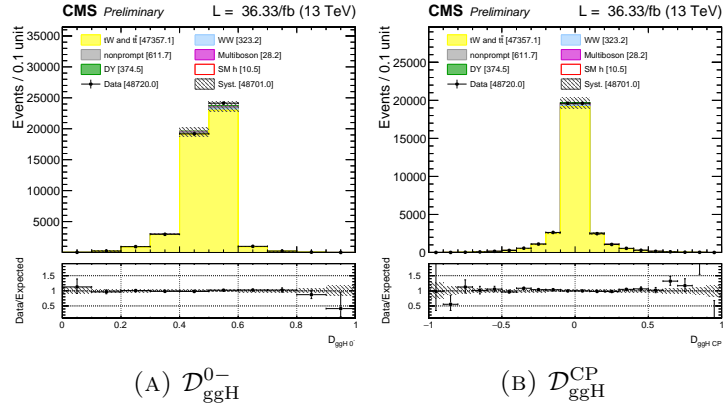


FIGURE 4.36: ggH + 2jet related distributions in the top 2-jet CR for the 2016 dataset

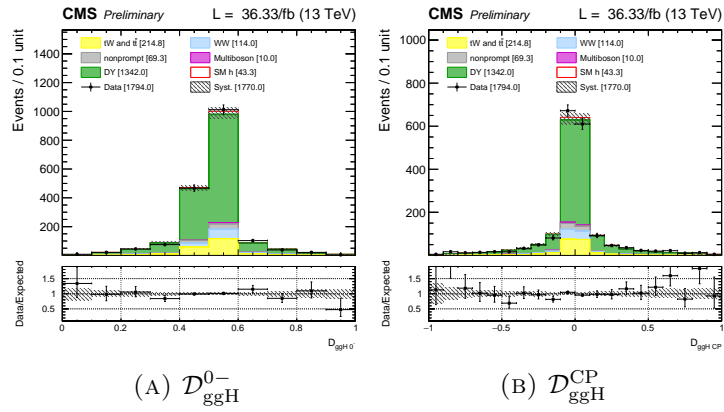
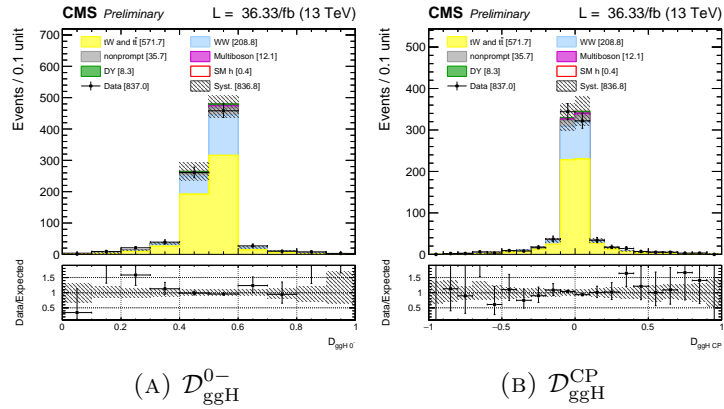
FIGURE 4.37: ggH + 2jet related distributions in the DY  $\tau\tau$  CR for the 2016 dataset

FIGURE 4.38: ggH + 2jet related distributions in the WW CR for the 2016 dataset

## 0-jet and 1-jet ggH channels

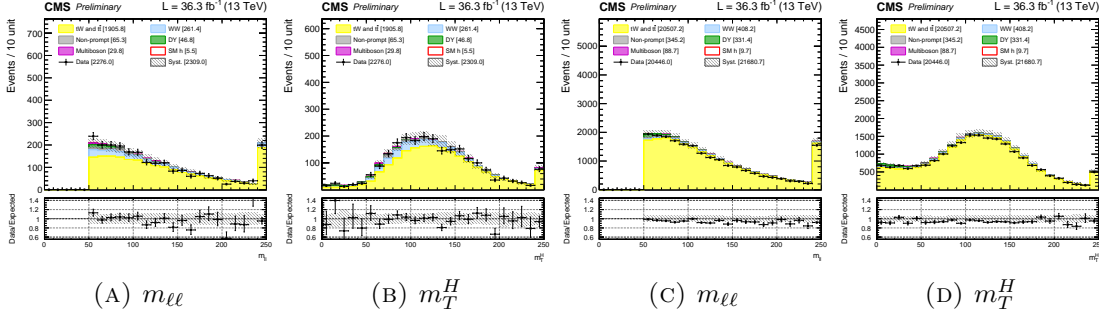
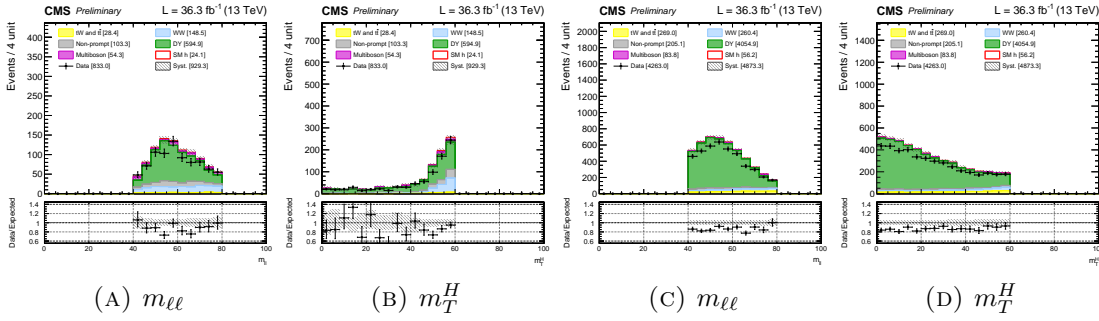


FIGURE 4.39: Top 0-jet (A, B) and 1-jet (C, D) CR distributions for 2016 datasets.

FIGURE 4.40: DY  $\tau\tau$  0-jet (A, B) and 1-jet (C, D) CR distributions for 2016 datasets.

## 4.5.5 Observed distributions in signal region

To provide an illustrative example, this section showcases the observed pre-fit 1D KD distributions in the VBF and VH signal regions using the 2016 dataset. These distributions effectively demonstrate the strong agreement between the data and the MC simulations. It is important to note that only the distributions related to the 2-jet channels are displayed, as they are the primary focus of the thesis. Additionally, it should be observed that the VBF distributions depicted in Fig. 4.41 are the same as those shown in Fig. 4.25, while the VH distributions presented in Fig. 4.42 correspond to the ones previously exhibited in Fig. 4.26.

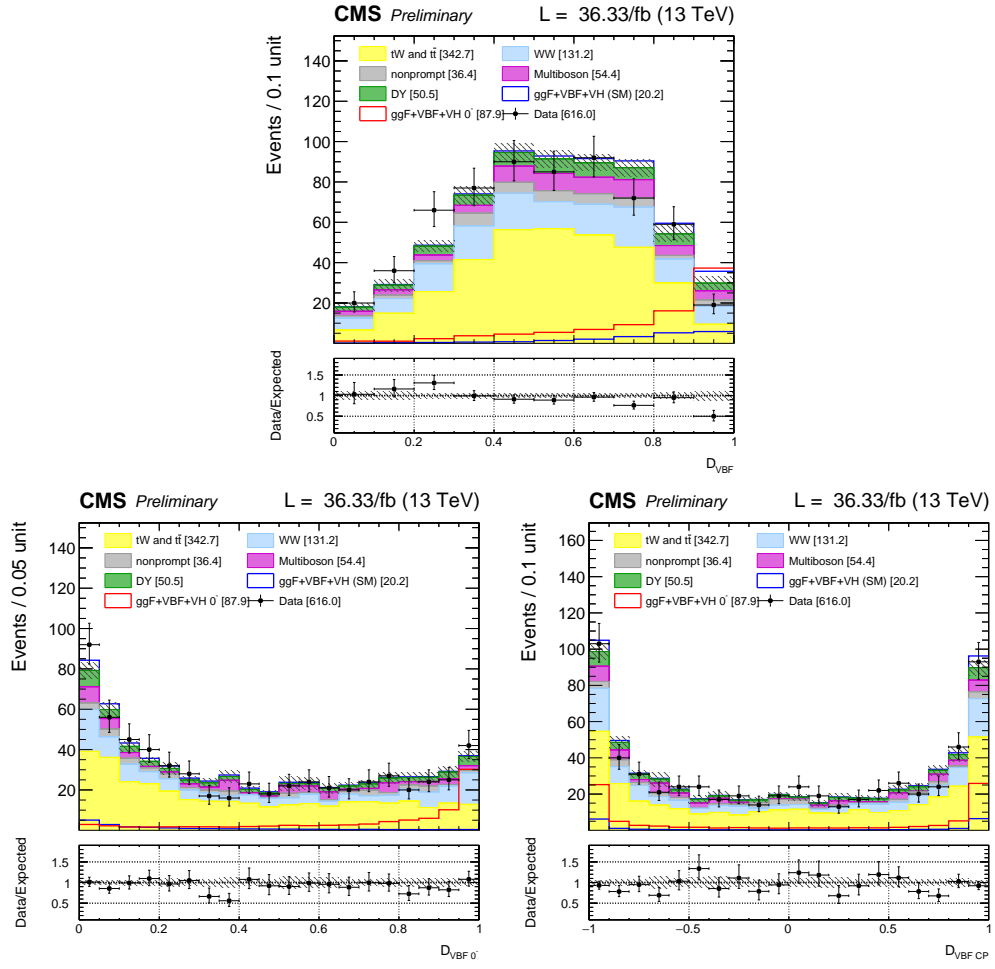


FIGURE 4.41: Observed 1D-KD corresponding to the 2-jet VBF category.

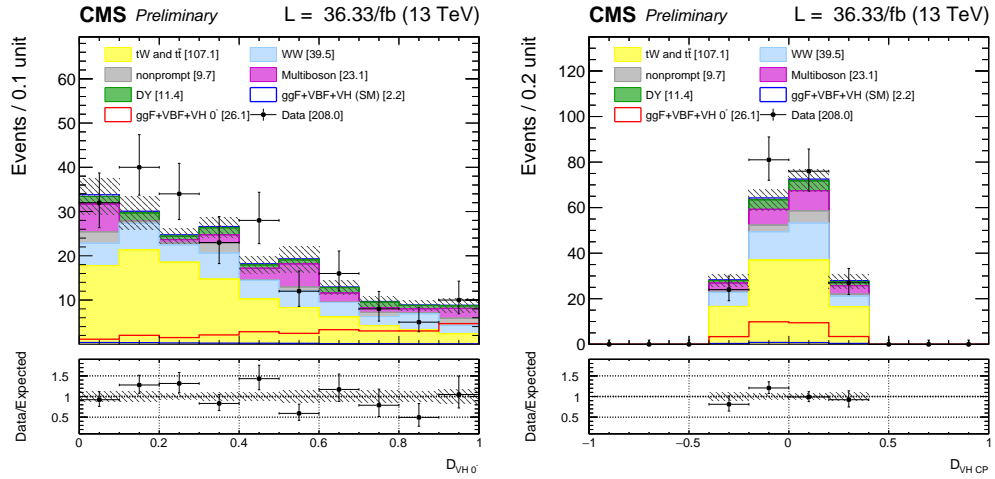


FIGURE 4.42: Observed 1D-KD corresponding to the 2-jet VH category.

## 4.6 Multi-dimensional kinematic discriminants

In both the AC and SMEFT interpretations of the analysis, the discriminants explained in the preceding sections are combined to create multidimensional kinematic discriminants. This section discusses the process of combining these discriminants, which enhances the sensitivity to AC effects at the HVV and Hgg vertices. Later in this section, Table 4.14 presents a summary of the combined discriminants, along with a reference to the corresponding final distributions used in the analysis for both AC and SMEFT interpretations.

### AC interpretation (Approach 1)

In this interpretation, the SM coupling and just one anomalous coupling at a time are studied. For the purpose of studying the HVV coupling, we construct four multi-dimensional discriminants based on MELA, specifically targeting the couplings  $a_2$ ,  $a_3$ ,  $\Lambda_1$ , and  $\Lambda_1^{Z\gamma}$  for the VBF, resolved VH, and boosted VH categories. By leveraging the discriminant power of MELA, for the  $a_3$  HVV coupling study we categorise events using the  $\mathcal{D}_{\text{CP}}$  discriminant in the VBF and VH categories. In the case of the  $a_2$  HVV coupling, an interference categorization is also implemented in the VBF category using  $\mathcal{D}_{\text{INT}}$ . Additionally, we create two 2D-unrolled discriminants for the 0-jet ggH and 1-jet ggH categories by using the  $m_{\ell\ell}$  and  $m_{\text{T}}^H$  variables.

For the Hgg vertex, we build a 2D-unrolled discriminant combining a production-based discriminant and a dedicated discriminant targeting a specific BSM coupling. Furthermore, the events are divided into two categories relying on the forward-backward asymmetry of the  $\mathcal{D}_{\text{CP}}^{ggH}$  discriminant.

The combination of discriminants for each case and the dedicated binning is discussed in the following.

**VBF category (HVV vertex)** In the VBF category, we employ a multidimensional kinematic discriminant by combining  $\mathcal{D}_{\text{VBF}}$ ,  $m_{\ell\ell}$ , and  $\mathcal{D}_{\text{BSM}}$ . This combined discriminant enhances the sensitivity to HVV couplings in both the production (through  $\mathcal{D}_{\text{VBF}}$  and  $\mathcal{D}_{\text{BSM}}$ ) and decay vertices (via  $m_{\ell\ell}$ ). For the production discriminant  $\mathcal{D}_{\text{VBF}}$ , we use one low purity and one high purity bin, separated at 0.75, as shown in Fig. 4.25 (A). The  $m_{\ell\ell}$  distribution shown in Fig. 4.29 (A) is



divided into two bins separated at 45 GeV. The selection of binning is based on variations in the discriminant power of the distributions along the x-axis of the mentioned figures. The  $\mathcal{D}_{\text{BSM}}$  discriminant, which depends on the specific coupling and can be  $\mathcal{D}_0^-$ ,  $\mathcal{D}_0^+$ ,  $\mathcal{D}_{\Lambda 1}$ , or  $\mathcal{D}_{\Lambda 1}^{Z\gamma}$ , is represented with 10 equal bins. The binning scheme for the VBF category is illustrated in Fig. 4.43, where the dashed lines define the bins for the  $m_{\ell\ell}$  variable, and the blue line indicates the separation between the two bins chosen for the production discriminant. The discriminants of the VBF category are presented as 3D unrolled distributions. To maintain sim-

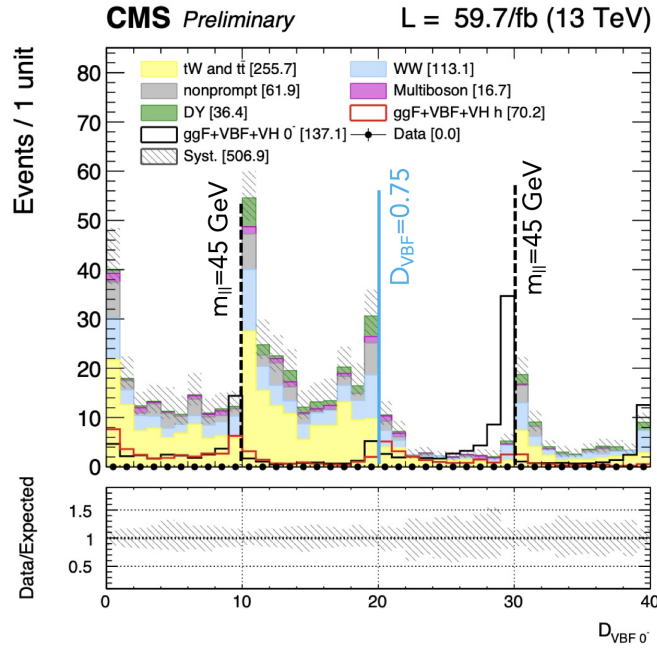


FIGURE 4.43: Example of a 3D-unrolled discriminant using  $[\mathcal{D}_{\text{VBF}}, m_{\ell\ell}, \mathcal{D}_0^-]$  in the VBF category considering the 2018 data set.

plicity, the absence of data in Fig. 4.43 is intentional. The plot including data points is shown later.

**VH categories (HV V vertex)** In the VH channels, we use a 2D-unrolled discriminant by combining  $\mathcal{D}_{\text{BSM}}$  and  $m_{\ell\ell}$ . In the resolved VH category, we divide  $\mathcal{D}_{\text{BSM}}$  into four equal-sized bins, while in the boosted VH category, we use three  $\mathcal{D}_{\text{BSM}}$  variable bins with boundaries of 0.6 and 0.8, determined based on the shape of the signal distribution shown in Fig. 4.27 (A). For  $m_{\ell\ell}$ , we consider two bins with a boundary at 45 GeV. Figure 4.44 illustrates the binning scheme for the resolved VH (A) and boosted VH (B) categories, where the dashed lines define the bins for the  $m_{\ell\ell}$  variable. It is important to note that the discriminants of the VH categories are presented as 2D unrolled distributions.

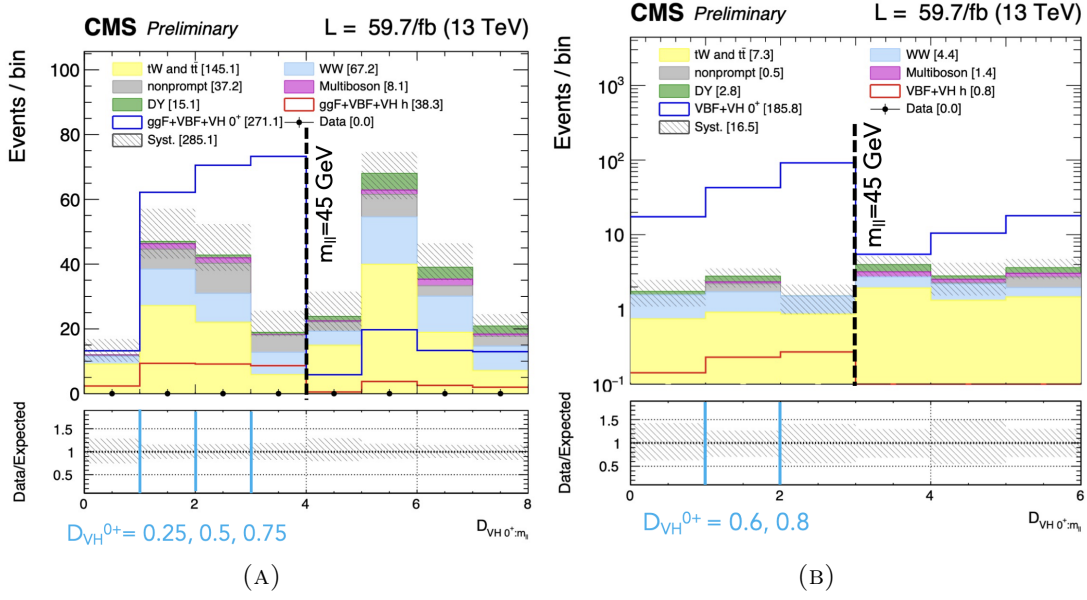


FIGURE 4.44: Example of a 2D-unrolled discriminant using  $[m_{\ell\ell}, \mathcal{D}_{0-}]$  in the Resolved VH category (A) and the Boosted VH category (B) considering the 2018 data set. The black dashed line indicates the binning in  $m_{\ell\ell}$ , and the light blue lines indicate binning chosen for the BSM discriminant.

Again, to maintain simplicity, the absence of data in Fig. 4.44 is intentional. The plot including data points is shown later.

**0-jet and 1-jet ggH categories (HVV vertex)** In this case, we utilize a 2D unrolled distribution constructed using  $m_{\ell\ell}$  and  $m_T^H$ . The chosen binning scheme for  $m_{\ell\ell}$  is [12, 17, 25, 30, 35, 40, 45, 65, 100, 200] GeV, while for  $m_T^H$ , the binning is [60, 70, 80, 90, 100, 110, 125] GeV. The corresponding distributions are shown later in Fig. 4.52 (A, B).

**2-jet ggH category (Hgg vertex)** To focus on the Hgg vertex, we construct a 2D unrolled KD using  $\mathcal{D}_{\text{VBF}}$  and  $\mathcal{D}_{0-}^{\text{ggH}}$ . It is important to note that we employ the VBF production discriminant, as we are primarily interested in events that exhibit a VBF-like topology. In this case, the variable  $m_{\ell\ell}$  is not sensitive to anomalous effects and therefore is not employed. Regarding  $\mathcal{D}_{\text{VBF}}$ , the bin boundary is relaxed to 0.5 to ensure an adequate acceptance of ggH events in the bin that resembles VBF more closely, where CP effects are most significant. In the high  $\mathcal{D}_{\text{VBF}}$  bin, eight  $\mathcal{D}_{\text{BSM}}$  bins are utilized, corresponding to [0,0.2,0.3,0.4,0.5,0.6,0.7,0.8,1]. For the low  $\mathcal{D}_{\text{VBF}}$  bin, five  $\mathcal{D}_{\text{BSM}}$  bins of [0,0.3,0.4,0.5,0.6,1] are employed. The first

and last bins are larger due to the reduction in background MC statistics in these regions.

## SMEFT interpretation (Approach 2)

Under the SMEFT interpretation, we construct a unique multi-dimensional discriminant for each channel to simultaneously constrain all HVV couplings:  $a_2$ ,  $a_3$ , and  $\Lambda_1$ . This is possible because of the signal model presented in section 4.8. Constructing and computing a signal model with 5 independent parameters, as required in the AC interpretation, poses challenges and is computationally intensive. However, by reducing the number of parameters to 4, the signal model becomes more manageable.

The discriminant distributions  $\mathcal{D}_{\text{BSM}}$  exhibit similar discrimination power between signals, as shown in Fig. 4.45. Indeed, they are correlated, and the number of bins we can have is limited by MC statistics, so we choose just 2 out of the 3  $\mathcal{D}_{\text{BSM}}$  considered. We use the CP-odd discriminant  $\mathcal{D}_{0-}$  and one of the CP-even discriminants,  $\mathcal{D}_{0+}$ , simultaneously.

For events in the VBF channel, we make 4 categories using  $\mathcal{D}_{\text{CP}}$  and  $\mathcal{D}_{\text{INT}}$ , Fig. 4.46 shows the 2D distribution of the 2 discriminants for a SM Higgs signal. The binning scheme for the interference discriminants, with divisions at 0 and 0.4, respectively, is maintained from the AC interpretation. Meanwhile, events falling into the VH channels are categorized according to the  $\mathcal{D}_{\text{CP}}$  discriminant.

We keep the  $m_{\ell\ell}$  and  $m_T^H$  discriminants for 0-jet and 1-jet ggH as for the AC approach. Additionally, it is important to note that the analysis does not study the Hgg vertex under the SMEFT interpretation.

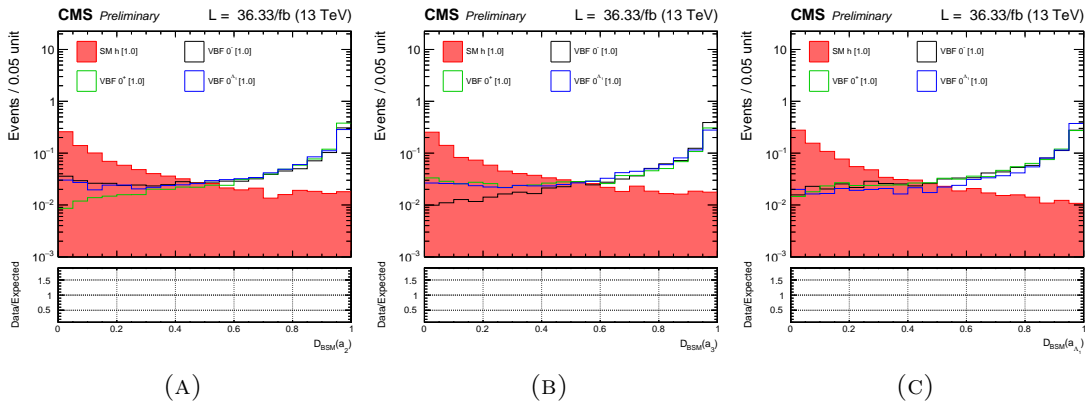


FIGURE 4.45: BSM Higgs signal discriminators  $\mathcal{D}_{0-}$  (A),  $\mathcal{D}_{0+}$  (B), and  $\mathcal{D}_{\Lambda_1}$  (C). The SM and the BSM Higgs signals are normalized to unity.

**VBF category:** The multi-KD is built by combining  $\mathcal{D}_{\text{BSM}}$ ,  $\mathcal{D}_{\text{VBF}}$  and  $m_{\ell\ell}$  discriminants. As mentioned in the introduction of this section, we use a combination of the  $\mathcal{D}_{0-}$  and  $\mathcal{D}_{0+}$  discriminants ( $[\mathcal{D}_{0-}, \mathcal{D}_{0+}]$ ) both of which are divided into 3 bins, with bin boundaries set at 0.1 and 0.9. This choice targets defining three distinct regions: one corresponding to SM-like behavior, another representing a mixture of SM and BSM effects, and a third indicating BSM-like behavior. The  $\mathcal{D}_{\text{VBF}}$  and  $m_{\ell\ell}$  variables are used in the same way as in Approach 1.

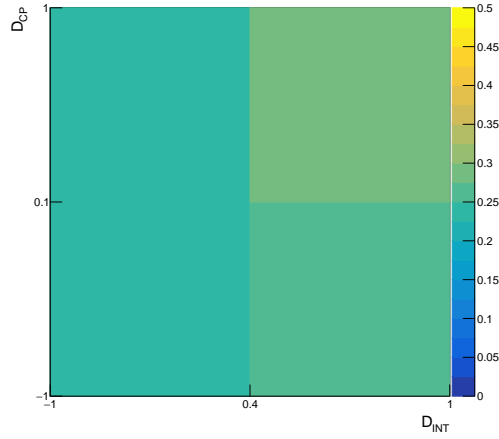


FIGURE 4.46: Expected VBF SM Higgs signal (normalized to unity) in the 4 VBF interference categories defined by the discriminants  $\mathcal{D}_{CP}$  and  $\mathcal{D}_{INT}$ .

**VH categories:** In the VH channels, both the  $m_{\ell\ell}$  distribution and the combined  $[\mathcal{D}_{0-}, \mathcal{D}_{0+}]$  discriminant are utilized. Specifically, in the resolved VH category, we employ three bins with boundaries set at 0.25 and 0.75. On the other hand, in the boosted VH category, we use two bins separated at 0.8. Additionally, the events are categorized based on the  $\mathcal{D}_{CP}$  discriminant.

**Rebinning of discriminants:** To ensure sufficient statistical precision in the predictions of all bins while retaining the kinematic information required to discriminate between the SM and anomalous coupling signal hypotheses, a rebinning strategy was implemented for both the VBF and VH channels to the unrolled  $[\mathcal{D}_{0-}, \mathcal{D}_{0+}]$  distribution. To illustrate this process, we present the expected BSM signals normalized to unity in Fig. 4.47 right before implementing the rebinning. Firstly, the bins [1, 2, 3, 4, and 7] are merged, as these bins are dominated by the SM Higgs. Secondly, bins are merged based on the background MC statistics, i.e., bins with a MC statistical fractional error greater than 0.5 or bins with no signal. The resulting distributions are shown in the section that follows.

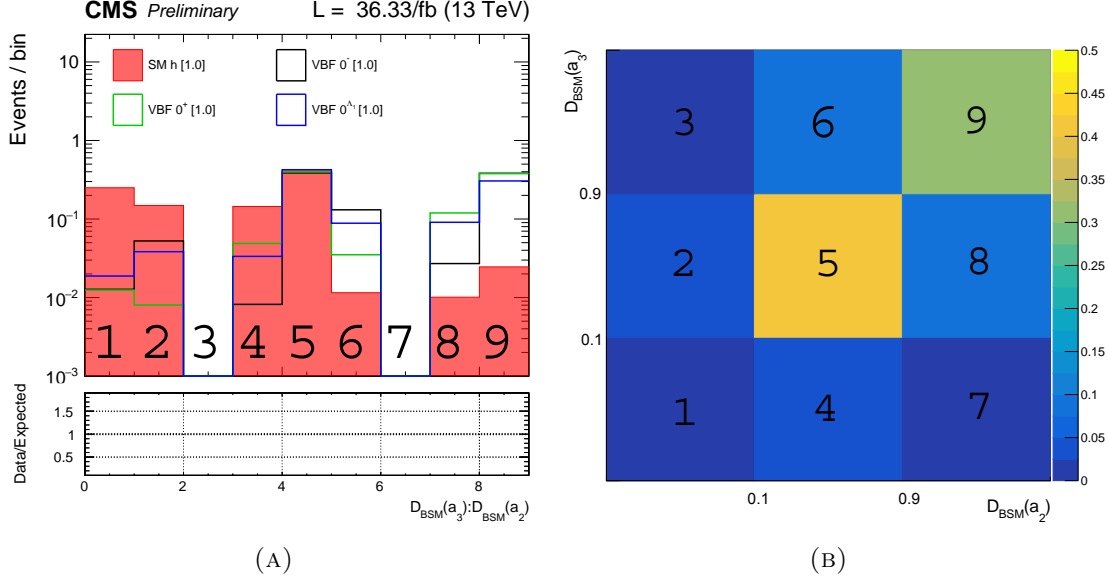


FIGURE 4.47: Considering the VBF production process only: (A) unrolled distribution of the  $[D_{0-}, D_{0+}]$  discriminant for SM and BSM Higgs signals normalized to unity, and (B) the corresponding 2D distribution for the  $a_{\Lambda 1}$  BSM signal.

TABLE 4.14: KD for both the interference-based categorization and the final discriminants employed in the data fitting process to investigate the HVV and Hgg couplings.

| Analysis | Channel  | Categorization                               | Final discriminant  | Pre-fit Fig.      |
|----------|--|--|---|-------------------|
| HVV      | VBF ( $a_3$ )                                    | $\mathcal{D}_{CP}$                           | $[\mathcal{D}_{VBF}, m_{\ell\ell}, \mathcal{D}_{0-}]$                   | 4.48 (A, D)       |
| AC       | VBF ( $a_2$ )                                    | $\mathcal{D}_{\text{int}}$                   | $[\mathcal{D}_{VBF}, m_{\ell\ell}, \mathcal{D}_{0+}]$                   | 4.49 (A, C)       |
|          | VBF ( $\kappa_1 \Lambda_1$ )                     | -  | $[\mathcal{D}_{VBF}, m_{\ell\ell}, \mathcal{D}_{\Lambda 1}]$            | 4.50 (A)          |
|          | VBF ( $\kappa_2^{Z\gamma} \Lambda_1^{Z\gamma}$ ) | -  | $[\mathcal{D}_{VBF}, m_{\ell\ell}, \mathcal{D}_{\Lambda 1}^{Z\gamma}]$  | 4.51 (A)          |
|          | VH ( $a_3$ )                                     | $\mathcal{D}_{CP}$                           | $[m_{\ell\ell}, \mathcal{D}_{0-}]$                                      | 4.48 (B, E, C, F) |
|          | VH ( $a_2$ )                                     | -  | $[m_{\ell\ell}, \mathcal{D}_{0+}]$                                      | 4.49 (B, D)       |
|          | VH ( $\kappa_1 \Lambda_1$ )                      | -  | $[m_{\ell\ell}, \mathcal{D}_{\Lambda 1}]$                               | 4.50 (B, C)       |
|          | VH ( $\kappa_2^{Z\gamma} \Lambda_1^{Z\gamma}$ )  | -  | $[m_{\ell\ell}, \mathcal{D}_{\Lambda 1}^{Z\gamma}]$                     | 4.51 (B, C)       |
|          | 0- & 1-jet ggH                                   | -  | $[m_{\ell\ell}, m_T^H]$   | 4.52 (A, B)       |
| Hgg      | 2-jet ggH  | $\mathcal{D}_{CP}$                           | $[\mathcal{D}_{VBF}, \mathcal{D}_{0-}^{ggH}]$                           | 4.53 (A, B)       |
|          | 0- & 1-jet ggH                                   | -  | $[m_{\ell\ell}, m_T^H]$   | -                 |
| HVV      | VBF  | $\mathcal{D}_{CP}, \mathcal{D}_{\text{int}}$ | $[\mathcal{D}_{VBF}, m_{\ell\ell}, \mathcal{D}_{0-}, \mathcal{D}_{0+}]$ | 4.54 (A, B, C, D) |
| SMEFT    | VH   | $\mathcal{D}_{CP}$                           | $[m_{\ell\ell}, \mathcal{D}_{0-}, \mathcal{D}_{0+}]$                    | 4.54 (E, F, G, H) |
|          | 0- & 1-jet ggH                                   | -  | $[m_{\ell\ell}, m_T^H]$   | 4.52 (A, B)       |

## 4.7 Prefit distributions

This section gathers the multidimensional KDs for all categories using the full Run 2 dataset. Each distribution shows the ratio of observed data to the combined contributions of background and SM signals before the final fit is performed (pre-fit distributions). The pure BSM contribution is also shown but is not included in the ratios. In the captions of the figures, the corresponding channel, category, and cuts utilized are specified.

### HVV discriminants in VBF and VH categories (AC)

Fig. 4.48, 4.49, 4.50, and 4.51 display the prefit distributions of the multidimensional KD for the VBF and VH channels, specifically for the  $a_3$ ,  $a_2$ ,  $\Lambda_1$ , and  $\Lambda_1^{Z\gamma}$  couplings, respectively.

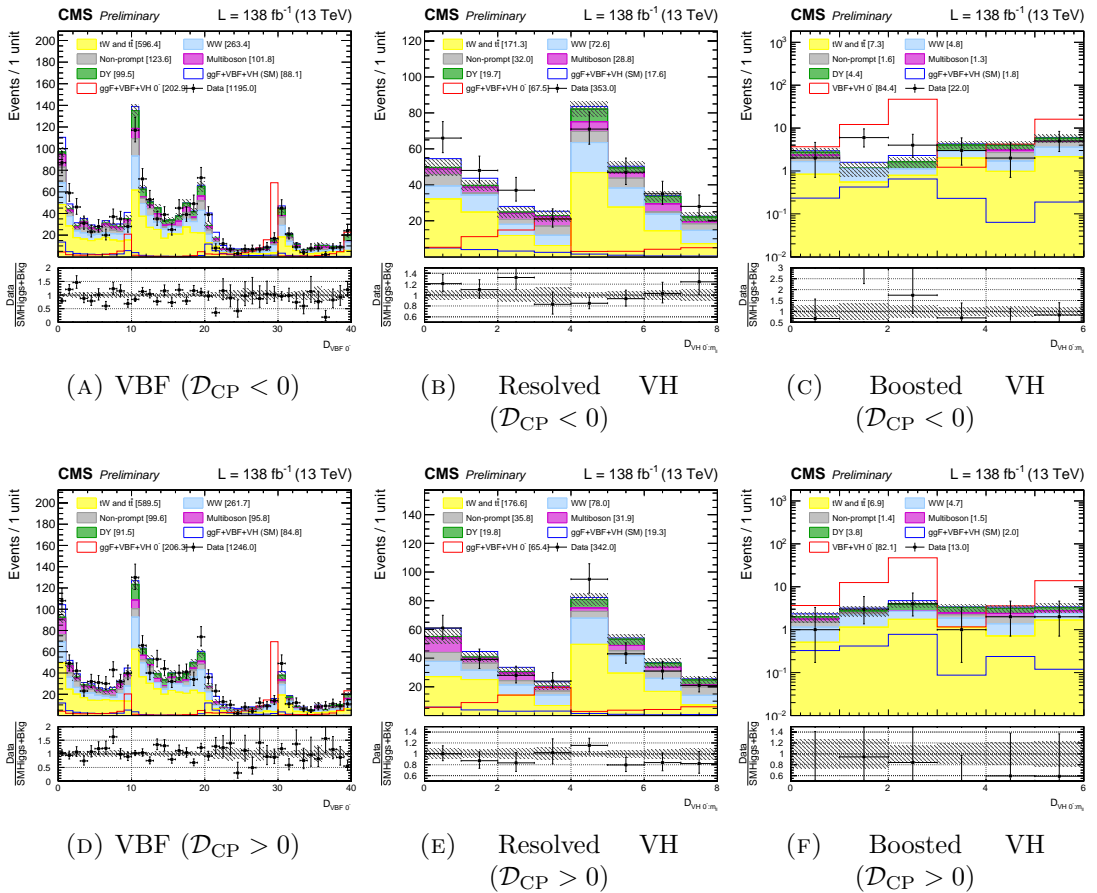


FIGURE 4.48: Multidimensional kinematic discriminants pre-fit for the  $a_3$  HVV coupling in combination of the 2016-2018 datasets.

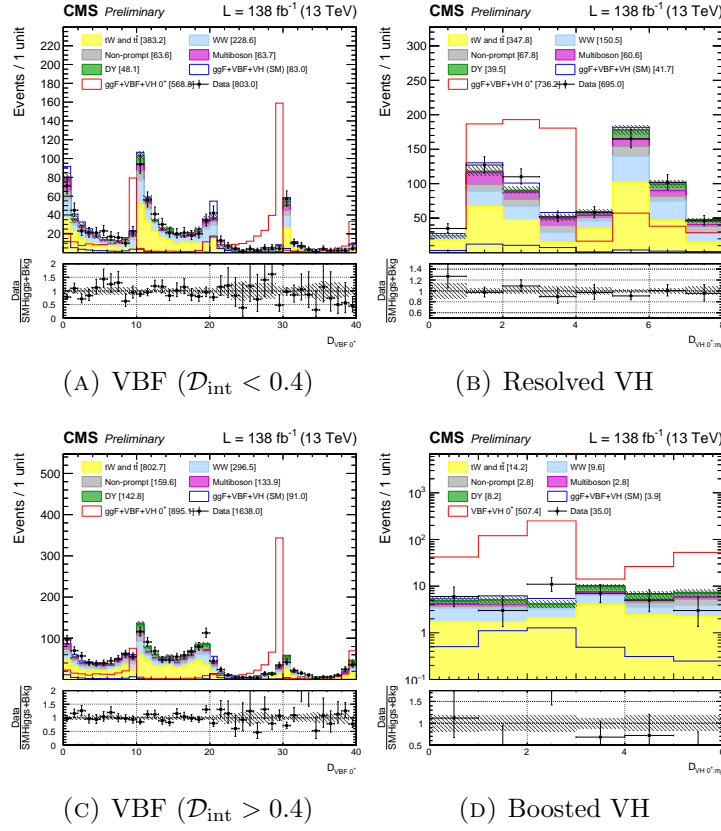


FIGURE 4.49: Multidimensional kinematic discriminants pre-fit for the  $a_2$  HVV coupling in combination of the 2016-2018 datasets. The SM and pure BSM signal expectations are shown.

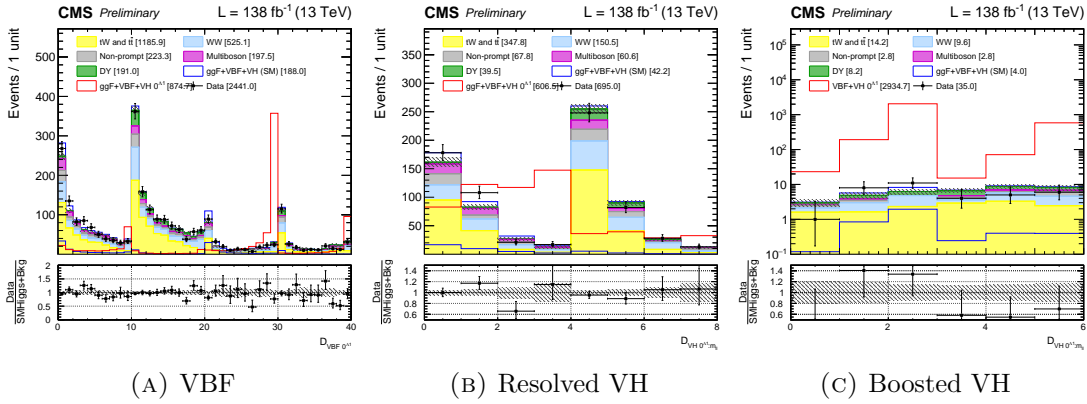


FIGURE 4.50: Multidimensional kinematic discriminants pre-fit for the  $\Lambda_1$  HVV coupling in combination of 2016-2018 datasets. The SM and pure BSM signal expectations are shown.



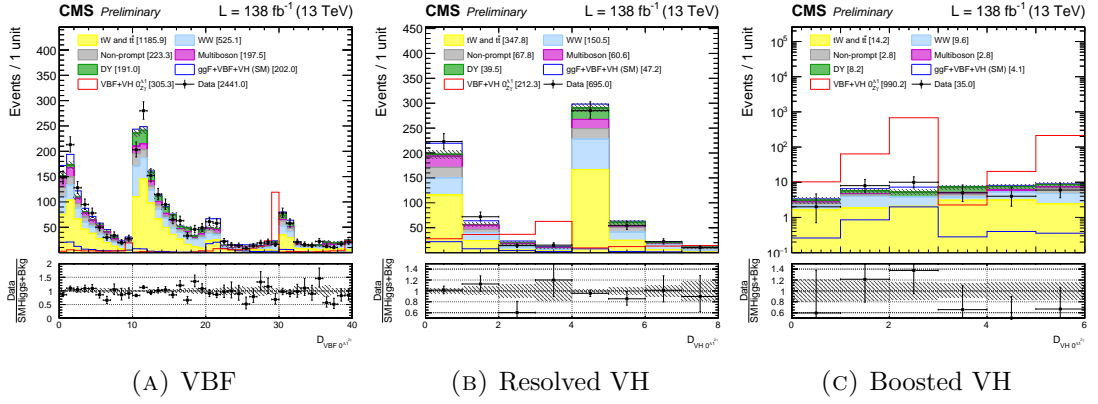


FIGURE 4.51: Multidimensional kinematic discriminants pre-fit for the  $\Lambda_1^{Z\gamma}$  HVV coupling in the combination of 2016-2018 datasets. The SM and pure BSM signal expectations are shown.

## HVV discriminants in 0-jet and 1-jet ggH categories

Figure 4.52 displays the prefit distributions of the multidimensional KD for the 0-jet and 1-jet ggH channels before performing the fit described in the following chapter. It should be noticed that the distributions are the same to constrain either  $a_3$ ,  $a_2$ ,  $\Lambda_1$ , or  $\Lambda_1^{Z\gamma}$  couplings.

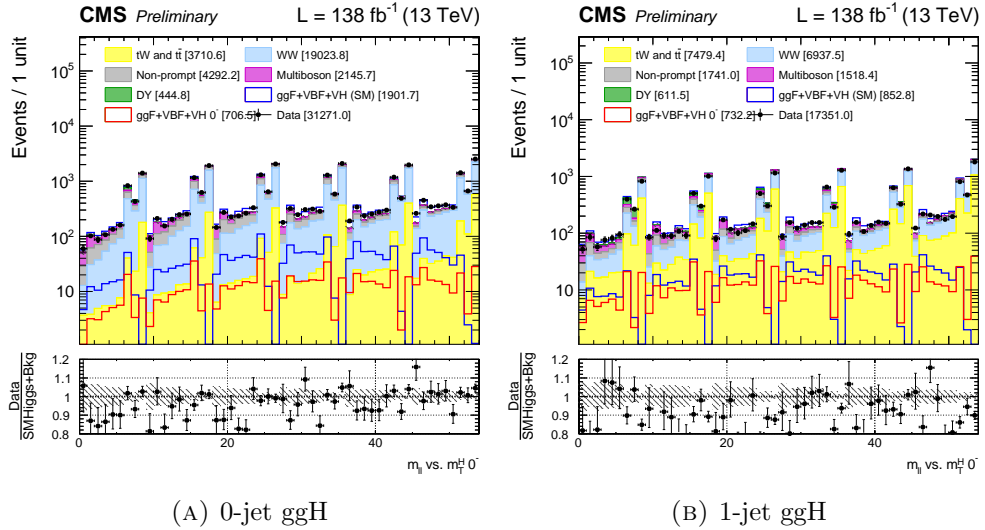


FIGURE 4.52: Two-dimensional  $m_{\ell\ell}$  vs.  $m_T$  pre-fit distributions in the 0 jet channel (A) and the 1 jet channel (B) considering the  $a_3$  HVV coupling plotted for the combination of 2016-2018 datasets.

## Hgg discriminants in the 2-jet ggH category

Figure 4.53 shows the prefit distributions of the multidimensional KD for the 2-jet ggH channel before performing the fit described in the following chapter. Simulation of the ggH + 2 jet process with Hgg anomalous couplings is done using MINLO X0JJ [141] at NLO in QCD.

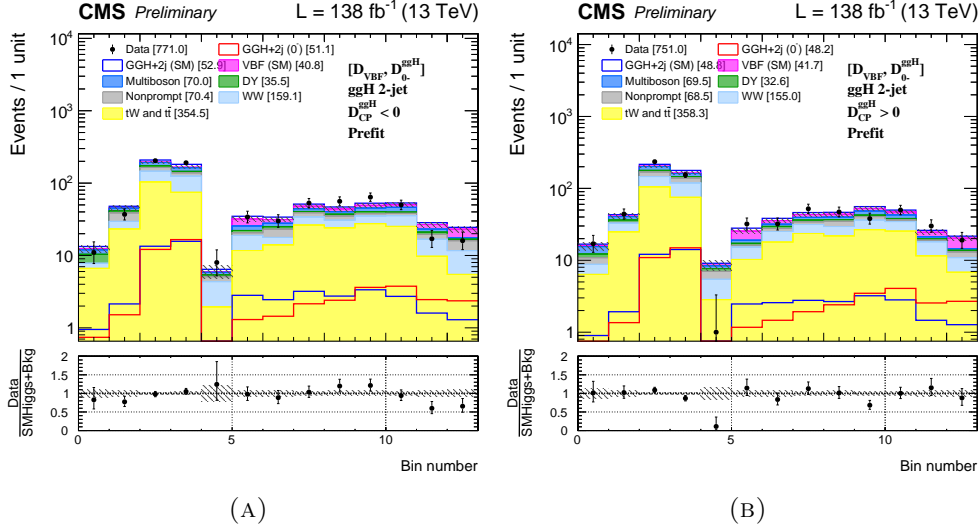


FIGURE 4.53: Prefit distribution of events in Run-2 combination of datasets for  $[\mathcal{D}_{VBF}, \mathcal{D}_{0-}^{ggH}]$  in the 2-jet ggH channel. The  $\mathcal{D}_{CP}^{ggH} < 0$  (A) and  $\mathcal{D}_{CP}^{ggH} > 0$  (B) categories are shown.

## SMEFT pre-fit full Run 2 distributions

Figure 4.54 shows the full Run 2 pre-fit distributions used for the SMEFT interpretation. Only these distributions are needed to constrain the  $a_3$ ,  $a_2$ , and  $\Lambda_1$  couplings simultaneously.

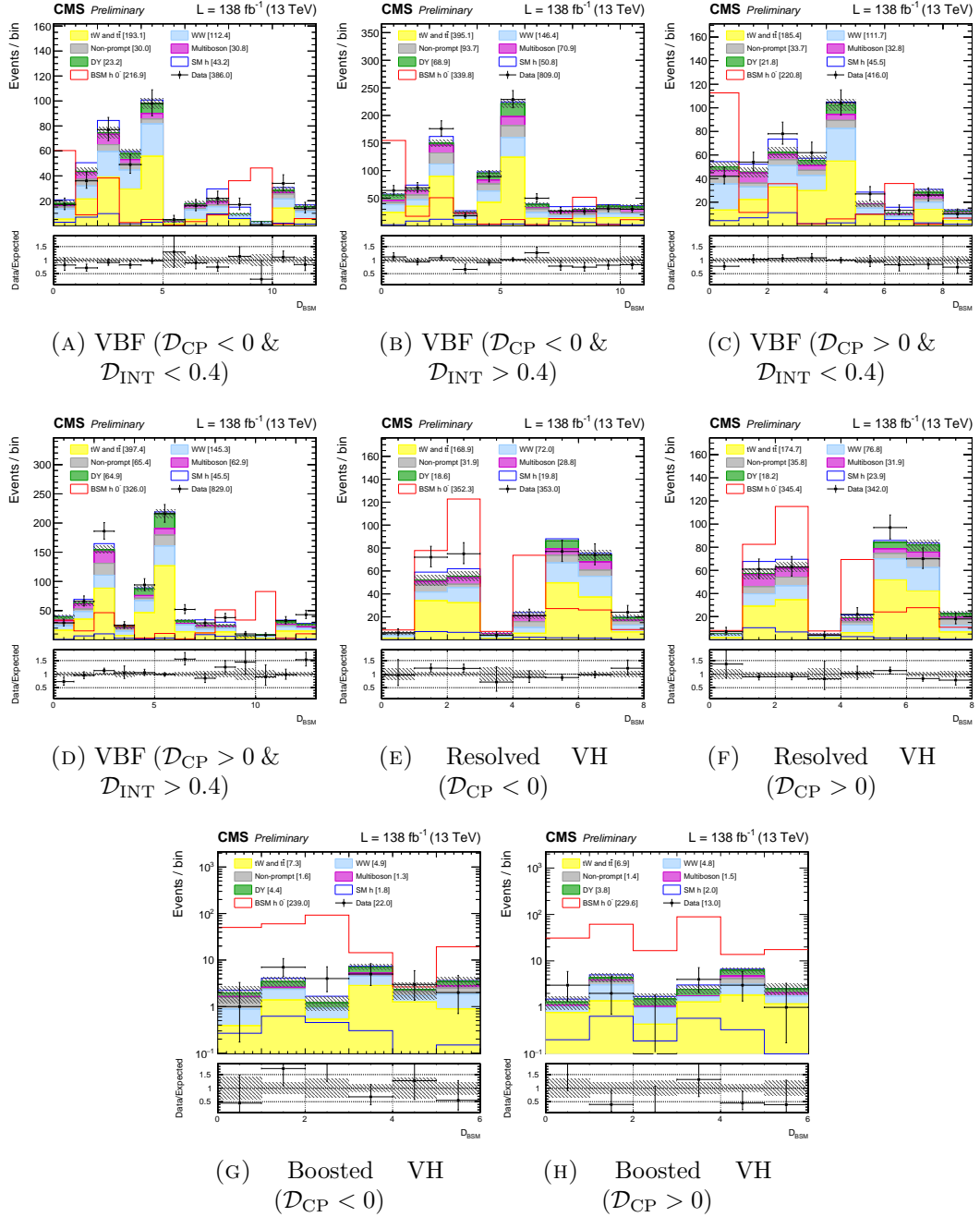


FIGURE 4.54: SMEFT Multi-KD pre-fit combining full Run 2 dataset per category. The SM and pure BSM signal expectations are shown.

## 4.8 Signal Modelling

In our study of two distinct vertices, HVV and Hgg, we require two separate signal models. Moreover, we consider two different approaches to interpret the results specifically for the HVV vertex, which impact the implementation of the signal model. In the AC approach, we examine the SM coupling along with a single BSM coupling, and in the SMEFT approach, we simultaneously examine multiple BSM couplings.

### AC model implementation

**HVV vertex** To demonstrate how the signal model is incorporated in this case, we will initially present an example for the VBF signal by examining the  $a_1$  and  $a_2$  couplings, which correspond to the SM Higgs and the CP-Even BSM Higgs, respectively. To simplify the scattering amplitude for a single *HVV vertex*, we express it as  $A(\text{HVV})$ .

$$A(\text{HVV}) = a_1 A_1 + a_2 A_2 \quad (4.13)$$

Nevertheless, the VBF production signal includes two HVV vertices: one on the production side and another on the decay side. As a result, the electroweak production amplitude is expressed as:

$$A(2\text{HVV}) = (a_1 A_1^{\text{Prod}} + a_2 A_2^{\text{Prod}}) * (a_1 A_1^{\text{Dec}} + a_2 A_2^{\text{Dec}}) \quad (4.14)$$

The square of the amplitude provides a description of a particular physical observable. Therefore, the electroweak signal model is obtained by squaring  $A(2\text{HVV})$  and organizing terms based on their couplings:

$$A(2\text{HVV})^2 = a_1^4 T_1 + a_1^3 a_2 T_2 + a_1^2 a_2^2 T_3 + a_1 a_2^3 T_4 + a_2^4 T_5 \quad (4.15)$$

where  $T_1$  denotes the purely SM term,  $(A_1^{\text{Prod}} A_1^{\text{Dec}})^2$ ,  $T_5$  represents the purely BSM term,  $(A_2^{\text{Prod}} A_2^{\text{Dec}})^2$ , and  $T_2 - T_4$  refer to SM-BSM interference terms, which are combinations of  $A_1$  and  $A_2$ . To apply this model, it is necessary to generate histograms of the pertinent discriminants for the five terms  $T_1$  through  $T_5$ , referred to as the VBF signal templates. For ease of use, we will employ the discriminant designed to distinguish the contribution of the  $a_1$  and  $a_2$  couplings in VBF production.

It is important to note that any VBF signal hypothesis,  $H(a_1, a_2)$ , may be represented using the square of  $A(2HVV)$  presented in Eq. 4.15.

To obtain all five templates, we begin by generating histograms of desirable variables for five distinct signal hypotheses, denoted as  $H_1$  through  $H_5$ . This is accomplished by applying MELA to reweigh the MC signal samples to a particular hypothesis. All MC samples are included, and the weighted average of predictions is obtained for each hypothesis. The selected hypotheses,  $H(a_1, a_2)$ , consist of pure SM, pure BSM, and three mixed-coupling hypotheses, i.e.,  $H(1,0)$ ,  $H(0,1)$ ,  $H(1,0.25)$ ,  $H(1,0.5)$ , and  $H(1,0.75)$ . The histograms for the signal hypotheses  $H_1$  to  $H_5$  are related to the required signal templates  $T_1$  to  $T_5$  by a  $5 \times 5$  matrix of the couplings used, denoted as  $G$ . This is expressed in simplified terms as  $H = G * T$ . By inverting the coupling matrix, we can express this as  $T = G^{(-1)} * H$ . Therefore, a specific template  $T_x$  is obtained through a linear combination of  $H_1$  to  $H_5$  using the inverted coupling matrix. A diagram to aid in understanding this process is presented in Fig. 4.55.

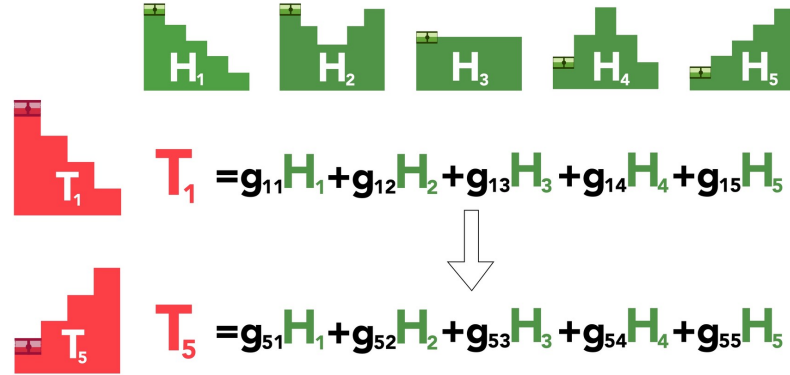


FIGURE 4.55: This figure illustrates the process for extracting the VBF/VH signal templates ( $T_1$ - $T_5$ ), which involves using 5 signal hypotheses ( $H_1$ - $H_5$ ) and the corresponding inverted coupling matrix ( $G$ ).

The WH and ZH production signals also involve 2 HVV vertices, so the same signal model and extraction procedure used for the VBF case is applied.

The ggH production signal involves only one HVV vertex, which is on the decay side. Therefore, the ggH signal model is obtained by squaring the  $A(HVV)$

amplitude and organizing the terms based on their coupling:

$$A(\text{HVV})^2 = a_1^2 T_1 + a_1 a_2 T_2 + a_2^2 T_3 \quad (4.16)$$

The pure SM term is denoted as  $T_1$ , the pure BSM term as  $T_3$ , and the SM-BSM interference term as  $T_2$ . The  $3 \times 3$  matrix is used to extract the templates  $T_1$ - $T_3$  with H1-H3 as inputs. The same methods are applied to study the  $a_3$  and  $\Lambda_1$  HVV anomalous couplings. One should notice that some templates can have a negative normalization, and thus we scale them by -1 and later introduce a factor -1 in the signal model to compensate.

**HZ $\gamma$  vertex** Regarding the  $\Lambda_1^{Z\gamma}$  coupling, there are some distinctions compared to the previous HVV couplings. In this scenario, only the VBF and ZH production vertices are affected. The amplitude that considers this anomalous coupling is expressed as:

$$A(\text{HZ}\gamma) = (a_1 A_1^{\text{Prod}} + \Lambda A_{\Lambda_1}^{\text{Prod}}) * a_1 A_1^{\text{Dec}} \quad (4.17)$$

In this case, there is no pure  $\Lambda_1^{Z\gamma}$  hypothesis, as the HWW decay channel is suppressed when the value of  $a_1$  approaches zero. To obtain the electroweak signal model, the amplitude is squared, resulting in the first three terms given in Eq. 4.14.

**Hgg vertex** The couplings of  $a_2$  and  $a_3$  of the *Hgg vertex* are the final ones considered in this analysis. These couplings correspond to the SM Higgs and the CP-Odd Higgs, respectively, and only the ggH production vertex is affected. Since we are dealing with only one vertex and two couplings, the signal model is the same as that in Eq. 4.16, but with  $a_2$  and  $a_3$  replacing  $a_1$  and  $a_2$ .

**Cross-section fractions for all vertices** As explained in Section 1.5, it is often more convenient to measure the cross-section ratio  $f_{ai}$  instead of directly measuring the couplings. To illustrate an example of how this is implemented, we will once again focus on the  $a_1$  and  $a_2$  HVV couplings. Assuming the couplings are positive, the cross-section fraction for the  $a_2$  coupling is defined as:

$$F_{a_2} = \frac{a_2^2 \sigma_2}{a_1^2 \sigma_1 + a_2^2 \sigma_2}, \quad (4.18)$$

where  $\sigma_i$  represents the signal cross-section with  $a_i = 1$  and  $a_{j \neq i} = 0$ . The cross-section fraction for the  $a_1$  coupling,  $F_{a1}$ , is obtained as  $F_{a1} = 1 - F_{a2}$ . We can also rewrite the electroweak and ggH signal models from Eqs. 4.14 and 4.16 in terms of cross-section fractions with some manipulation:

$$P_{\text{Ewk}} = \mu_{\text{Ewk}}^2 (F_{a1}^2 T_1 + g \sqrt{F_{a1}}^3 \sqrt{F_{a2}} T_2 + g^2 F_{a1} F_{a2} T_3 + g^3 \sqrt{F_{a1}} \sqrt{F_{a2}}^3 T_4 + g^4 F_{a2}^2 T_5) \quad (4.19)$$

$$P_{\text{ggH}} = \mu_{\text{ggH}} (F_{a1} T_1 + g \sqrt{F_{a1}} \sqrt{F_{a2}} T_2 + g^2 F_{a2} T_3) \quad (4.20)$$

The constant  $g$  corresponds to the ratio of cross-sections for the two couplings,  $\sqrt{\sigma_1/\sigma_2}$ , at a specific vertex. The choice of the constant  $g$  determines which  $\sigma_i$  the  $F_{ai}$  are defined with respect to. In this analysis, the ratio derived for the HZZ decay vertex, which is similar to that of the HWW decay vertex, is chosen and provided in Table 1.2. This choice allows for combination with the HZZ and  $H\tau\tau$  final states, where the same choice was made. A useful cross-check of this change in the signal model is to substitute the expressions for  $g$  and  $F_{ai}$  and use  $\mu = \frac{a_1^2 \sigma_1 + a_2^2 \sigma_2}{\sigma_1}$ . All  $\sigma_i$  cancel, and the initial signal models in terms of couplings is retrieved. The full signal model uses both  $P_{\text{Ewk}}$  and  $P_{\text{ggH}}$  to handle the VBF/VH and ggH production signals, respectively. The sign of the couplings is taken into account by scaling the  $\sqrt{F_{a2}}^x$  terms, where  $x$  has a value of 1 or 3, by  $\text{sign}(\frac{a_2}{a_1})$ . The final model has only three free parameters:  $\mu_{\text{Ewk}}$ ,  $\mu_{\text{ggH}}$ , and  $F_{a2}$ . The  $a_3$  and  $\Lambda_1$  HVV anomalous couplings can also be handled using this approach. For the  $\Lambda_1^{Z\gamma}$  coupling, only the first three terms of  $P_{\text{Ewk}}$  are needed, while for the  $a_2$  and  $a_3$  couplings of the Hgg vertex, only  $P_{\text{ggH}}$  is needed.

### SMEFT model implementation

**HVV vertex** In the AC interpretation discussed earlier, a total of five HVV couplings are considered assuming  $a_i^{ZZ} = a_i^{WW}$ . However, such a relationship is broken for the SMEFT interpretation when the full set of  $SU(2)_L \times U(1)_Y$  symmetry relationships discussed in section 1.5 is adopted. This further reduces the number of independent HVV parameters from five to four:  $a_1$ ,  $a_2$ ,  $a_3$ , and  $\Lambda_1$ . By considering these four couplings simultaneously, the scattering amplitude for one and two HVV vertices (production and decay) is expressed as:

$$A(\text{HVV}) = a_1 A_1 + a_2 A_2 + a_3 A_3 + a_4 A_4 \quad (4.21)$$

$$A(2\text{HVV}) = (a_1 A_1^{\text{Prod}} + a_2 A_2^{\text{Prod}} + a_3 A_3^{\text{Prod}} + a_4 A_4^{\text{Prod}}) * (a_1 A_1^{\text{Dec}} + a_2 A_2^{\text{Dec}} + a_3 A_3^{\text{Dec}} + a_4 A_4^{\text{Dec}}) \quad (4.22)$$

Similar to the previous section, the electroweak signal model is obtained by taking the  $A(2\text{HVV})$  amplitude and squaring it, followed by grouping the resulting terms based on their respective couplings. The resulting signal model is as follows:

$$\begin{aligned} P_{Ewk} = & a_1^4 T_1 + a_1^3 a_2 T_2 + a_1^2 a_2^2 T_3 + a_1 a_2^3 T_4 + a_2^4 T_5 \\ & + a_1^3 a_3 T_6 + a_1^2 a_3^2 T_7 + a_1 a_3^3 T_8 + a_3^4 T_9 \\ & + a_1^3 a_4 T_{10} + a_1^2 a_4^2 T_{11} + a_1 a_4^3 T_{12} + a_4^4 T_{13} \\ & + a_2^3 a_3 T_{14} + a_2^2 a_3^2 T_{15} + a_2 a_3^3 T_{16} \\ & + a_2^3 a_4 T_{17} + a_2^2 a_4^2 T_{18} + a_2 a_4^3 T_{19} \\ & + a_3^3 a_4 T_{20} + a_3^2 a_4^2 T_{21} + a_3 a_4^3 T_{22} \\ & + a_1 a_2 a_3^2 T_{23} + a_1 a_2^2 a_3 T_{24} + a_1^2 a_2 a_3 T_{25} \\ & + a_1 a_2 a_4^2 T_{26} + a_1 a_2^2 a_4 T_{27} + a_1^2 a_2 a_4 T_{28} \\ & + a_1 a_3 a_4^2 T_{29} + a_1 a_3^2 a_4 T_{30} + a_1^2 a_3 a_4 T_{31} \\ & + a_2 a_3 a_4^2 T_{32} + a_2 a_3^2 a_4 T_{33} + a_2^2 a_3 a_4 T_{34} \\ & + a_1 a_2 a_3 a_4 T_{35}, \end{aligned} \quad (4.23)$$



where the pure SM term is represented by  $T_1$ , while the pure BSM terms are represented by  $T_5$  (for  $a_2$ ),  $T_9$  (for  $a_3$ ), and  $T_{13}$  (for  $\Lambda_1$ , which is denoted as  $a_4$  for convenience). The remaining 31 terms are interference terms between pairs, triplets, and quartets of couplings. In order to extract the signal templates, 35 hypotheses are required, corresponding to pure SM, pure BSM, and various mixed-coupling hypotheses, along with the corresponding  $35 \times 35$  inverted coupling matrix.

In the case of the ggH production signal, there is only one HVV vertex on the decay side. Therefore, the ggH signal model is obtained by squaring the  $A(HVV)$  amplitude and grouping terms based on the coupling.

$$\begin{aligned}
 P_{ggH} = & a_1^2 T_1 + a_1 a_2 T_2 + a_2^2 T_3 \\
 & + a_1 a_3 T_4 + a_3^2 T_5 \\
 & + a_1 a_4 T_6 + a_4^2 T_7 \\
 & + a_2 a_3 T_8 + a_2 a_4 T_9 + a_3 a_4 T_{10}
 \end{aligned}
 \tag{4.24}$$

The pure SM term is again denoted as  $T_1$ , while the pure beyond the Standard Model terms are represented by  $T_3$  (for  $a_2$ ),  $T_5$  (for  $a_3$ ), and  $T_7$  (for  $\Lambda_1$ , which is denoted as  $a_4$  for convenience). In this case, the template extraction procedure is simpler, as only 10 hypotheses are chosen to correspond to the pure SM, pure BSM, and various mixed-coupling hypotheses. The gluon-gluon fusion signal templates are then extracted using the inverted  $10 \times 10$  coupling matrix.

**Cross-section fractions** We define the cross-section fraction for a specific anomalous contribution as:

$$F_{ai} = \frac{a_i^2 \sigma_i}{a_1^2 \sigma_1 + a_2^2 \sigma_2 + a_3^2 \sigma_3 + a_4^2 \sigma_4} \tag{4.25}$$

The cross-section fraction for the  $a_1$  coupling is calculated by subtracting the sum of the cross-section fractions for the  $a_2$ ,  $a_3$ , and  $a_4$  couplings from 1, i.e.,  $F_{a1} = 1 - F_{a2} - F_{a3} - F_{a4}$ . To interpret the signal model in terms of these cross-section fractions for the anomalous couplings, we can replace each  $a_i$  term with

$g_i \sqrt{F_{ai}}$ , where  $g_i$  represents the ratio of coupling cross-sections  $\sqrt{\sigma_1/\sigma_i}$ . We then scale each term by  $\mu_{Ewk}^2$  ( $\mu_{ggH}$ ) for the electroweak (ggH) model. The  $g_i$  coefficients used are listed in Table 1.2.

$$P_{Ewk} = \mu_{Ewk}^2 (F_{a1}^2 T_1 \cdots + g_2^4 F_{a2}^2 T_5 \cdots + g_3^3 g_4 \sqrt{F_{a3}}^3 \sqrt{F_{a4}} T_{20} \cdots + g_2 g_3 g_4 \sqrt{F_{a1}} \sqrt{F_{a2}} \sqrt{F_{a3}} \sqrt{F_{a4}} T_{35}) \quad (4.26)$$

$$P_{ggH} = \mu_{ggH} (F_{a1} T_1 \cdots + g_2^2 F_{a2} T_3 \cdots + g_4^2 F_{a4} T_7 \cdots + g_3 g_4 \sqrt{F_{a3}} \sqrt{F_{a4}} T_{10}) \quad (4.27)$$

The complete SMEFT signal model incorporates both  $P_{Ewk}$  and  $P_{ggH}$  to account for the VBF/VH and ggH production signals, respectively. To consider the sign of the couplings, the  $\sqrt{F_{ai}}^x$  terms, where  $x$  equals 1 or 3, is scaled by  $\text{sign}(\frac{a_i}{a_1})$ . The model comprises five independent parameters:  $\mu_{Ewk}$ ,  $\mu_{ggH}$ ,  $F_{a2}$ ,  $F_{a3}$ , and  $F_{a4}$ . The parameter space is constrained by the explicit requirement  $F_{a2} + F_{a3} + F_{a4} \leq 1$ , which ensures that only physical regions are probed.

In cases where the fit involves multiple floating signal parameters, a local minimum may be encountered instead of the global minimum. This issue is difficult to avoid, especially when multiple similarly deep minima are separated by significant distances. One solution is to perform the fit minimization using several starting points (varying initial guesses for  $F_{ai}$ ) and selecting the deepest minimum among all fits. In our SMEFT interpretation, we have adopted this approach to ensure the global minimum is obtained when all signal parameters are floated simultaneously.

## Signal model study

In this section, we investigate the predictions of the AC signal model in detail. In the final interpretation of the data, we perform likelihood scans of  $F_{ai}$  with the signal strengths left floating. As such, it is the signal shape that will largely determine the likelihood of a given signal hypothesis. We studied the shape of the most relevant signal in a given category for the full range of  $F_{ai}$  values for the important variables. To do this, we look at how the mean of a given variable changes as a function of  $F_{ai}$ . For any value of the model parameters, the signal expectation, corresponding to the combination of the signal templates for a given production process, can be determined. Significant features highlighted in this section are expected to be reproduced in the likelihood scans.

**0-jet and 1-jet channels** The 0 jet channel is the most sensitive channel to the ggH signal. In Fig. 4.56, we present the mean values of  $m_{\ell\ell}$  and  $m_T^H$  distributions of the ggH signal, demonstrating their shapes varying with  $F_{a2}$ ,  $F_{a3}$ , and  $F_{\Lambda1}$ . Noticeable changes in shape, deviating from the SM, occur for negative  $F_{a2}$  values, especially around -0.25, and positive  $F_{\Lambda1}$  values, particularly around 0.5. These features are also observed in the results from Run 1 [50] or HZZ [46] analyses and they are related to interference effects between SM and BSM couplings. Consequently, when performing likelihood scans for  $F_{a2}$  and  $F_{\Lambda1}$ , local maxima are expected to emerge around these non-SM-like points. Conversely, for  $F_{a3}$ , shape changes are relatively symmetric around zero.

**VBF and VH channels** In the VBF production channel, the variables  $\mathcal{D}_{\text{BSM}}$  and  $\mathcal{D}_{\text{VBF}}$  are employed. The mean values of these distributions, as functions of  $F_{a2}$ ,  $F_{a3}$ , and  $F_{\Lambda1}$ , are depicted in Fig. 4.57. Notably, for small values of  $F_{ai}$ , approximately 0.1, the  $\mathcal{D}_{\text{BSM}}$  shapes already exhibit a pronounced peak. This rapid transition from SM-like to BSM-like shapes is attributed to the electroweak production vertex sensitivity to AC effects. Consequently, likelihood scans for  $F_{ai}$  are expected to exhibit a sharp increase near the minimum at zero. For  $F_{\Lambda1}$ , the rise is relatively asymmetric, with a gentler increase on the positive side. Regarding  $\mathcal{D}_{\text{VBF}}$ , a substantial shape alteration compared to the SM is observed for  $F_{a2}$ , particularly around -0.25. The shape progression for  $F_{a3}$  and  $F_{\Lambda1}$  resembles that of  $\mathcal{D}_{\text{BSM}}$ , except for a dip around 0.6 for  $F_{\Lambda1}$ .

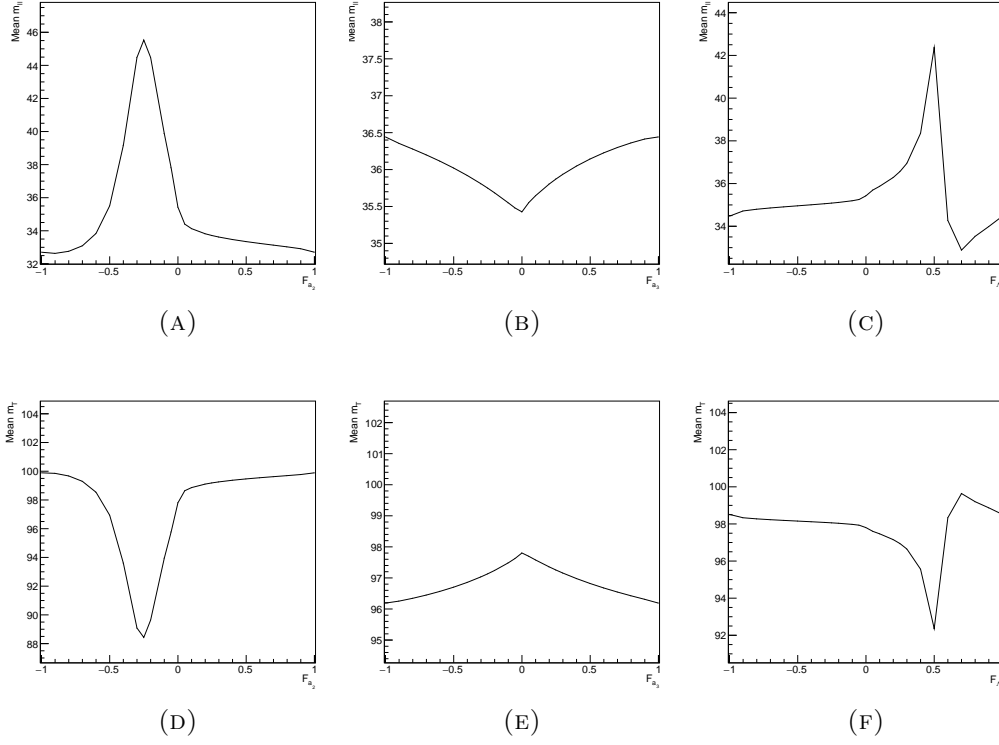


FIGURE 4.56: The mean of the ggH signal  $m_{ll}$  (top) and  $m_T$  (bottom) distributions as a function of  $f_{a_2}$ ,  $f_{a_3}$  and  $f_{\Lambda 1}$  in the 0-jet channel.

For the VH signal process, the  $\mathcal{D}_{\text{BSM}}$  variable is employed to investigate anomalous effects at the production vertex. The corresponding mean values, as functions of  $F_{a_2}$ ,  $F_{a_3}$ , and  $F_{\Lambda 1}$ , are displayed in Fig. 4.58. Similar to the VBF signal, sharp rises are observed around the minimum at zero. However, for  $F_{\Lambda 1}$ , the positive rise is sharper compared to the VBF signal, and there is also a substantial shape change with respect to the SM around 0.5. As observed for other signal variables, a similar feature is noted for  $F_{a_2}$  around -0.25.

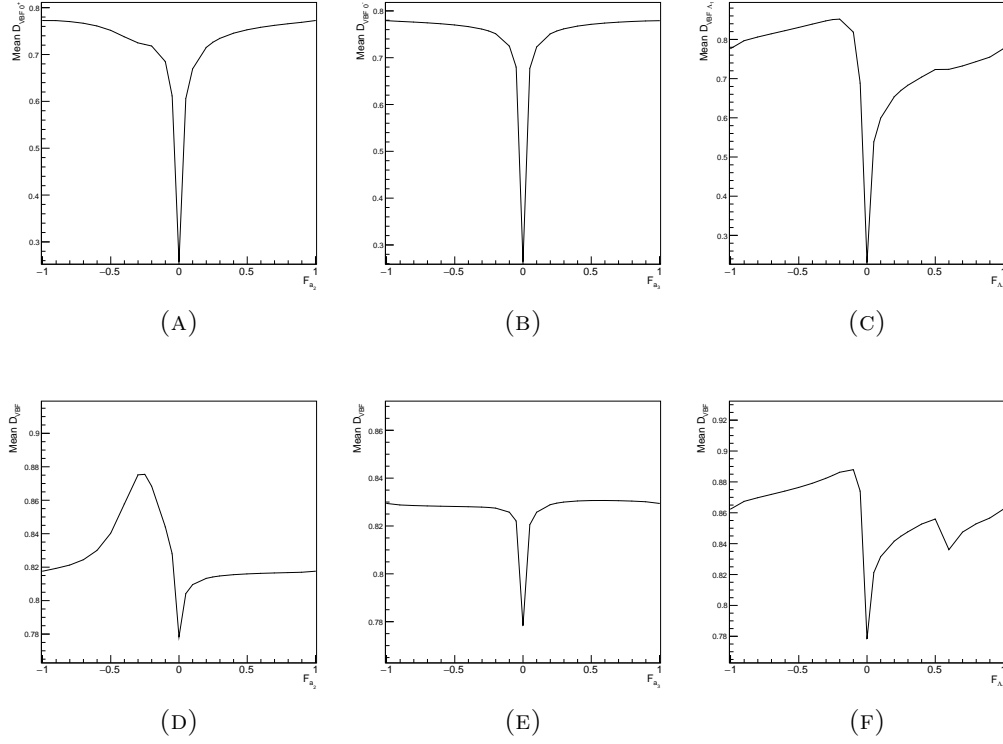


FIGURE 4.57: The mean of the VBF signal  $D_{BSM}$  (top) and  $D_{VBF}$  (bottom) distributions as a function of  $f_{a2}$ ,  $f_{a3}$  and  $f_{\Lambda1}$  in the VBF channel.

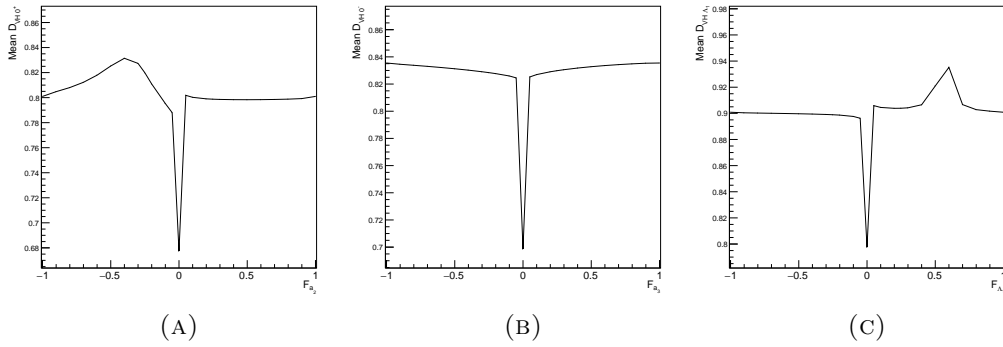


FIGURE 4.58: The mean of the VH signal  $D_{BSM}$  distribution as a function of  $f_{a2}$ ,  $f_{a3}$  and  $f_{\Lambda1}$  in the boosted VH channel.



# Chapter 5

## Signal extraction

To extract the signal of interest, we conduct a binned maximum likelihood (ML) fit on the combined data from all categories and data-taking periods. The statistical approach employed in this analysis follows the methodology developed by the ATLAS and CMS collaborations [142]. Figure 5.1 displays an overview of the categories included in the fit for the AC interpretation of the analysis. Similarly, Fig. 5.2 illustrates the categories included in the fit for the SMEFT interpretation. As mentioned earlier, the AC method exploits distinct kinematic discriminants and categorization strategies based on the investigated coupling, while the SMEFT method employs a single kinematic discriminant to constrain all couplings. Both approaches make use of the top,  $DY\tau\tau$ , and  $WW$  control regions to effectively constrain background events during the fitting process.

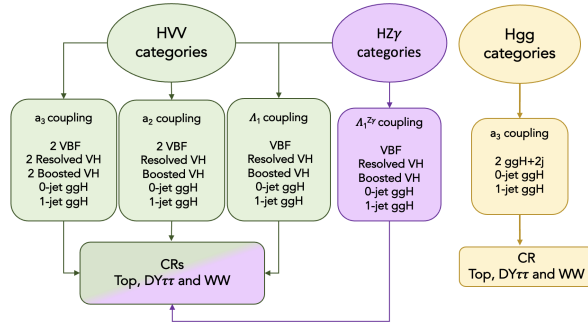


FIGURE 5.1: Diagram of categories per targetted vertex incorporated in the fit in Approach 1. For certain couplings in the VBF and VH channels, the two categories are defined based on either the  $\mathcal{D}_{CP}$  or  $\mathcal{D}_{INT}$  discriminants.

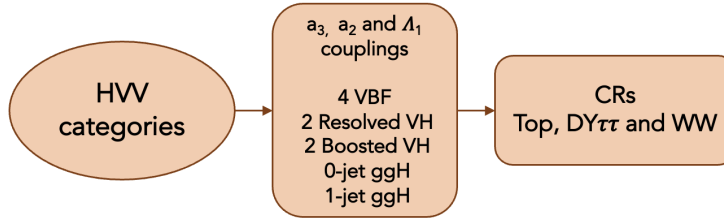


FIGURE 5.2: Diagram of categories per targetted vertex utilized in the fit in Approach 2. The  $\mathcal{D}_{\text{CP}}$  and  $\mathcal{D}_{\text{INT}}$  discriminants are used to define 4 categories in VBF. For the VH channels, two categories are defined based on the  $\mathcal{D}_{\text{CP}}$  discriminant.

## 5.1 Maximum Likelihood Fit

The methodology involves estimating a collection of unidentified parameters  $\theta$  based on a sample of observed data  $\vec{x}_{\text{obs}}$ . The likelihood of  $\theta$  being present in the observation  $\vec{x}_{\text{obs}}$  is computed by evaluating the probability distribution function  $f(\vec{x}_{\text{obs}}, \theta)$ . This function is deemed reliable in describing the data as:

$$L(\theta) = f(\vec{x}_{\text{obs}}, \theta). \quad (5.1)$$

where the set of parameters  $\theta$  is treated as the variable, and the optimal estimate of  $\theta$  in  $\vec{x}_{\text{obs}}$  is obtained by maximizing  $L(\theta)$ :

$$\hat{\theta} = \arg \max(L(\theta)), \quad (5.2)$$

being  $\hat{\theta}$  commonly known as the maximum likelihood estimator. Nonetheless, by using the logarithmic function  $\log L$  instead of  $L$  is computationally more efficient, as the former yields smaller values that are easier for a computer to handle. Additionally, Eq. 5.2 is often replaced by the negative logarithm of the likelihood, denoted as  $NLL$ :

$$\arg \max(L(\theta)) = \arg \min(-\log L(\theta)). \quad (5.3)$$

Moreover, in scenarios involving multiple independent observations  $\vec{x}_{\text{obs}}$  where the joint likelihood is the product of individual likelihoods, taking the logarithm allows the product to be expressed as a sum. The maximum likelihood estimator is beneficial as it is asymptotically unbiased, ensuring convergence in the limit of an infinite number of observations and efficiency in the large sample limit.



Incorporating systematic uncertainties into the estimation of the parameters of interest (POI) is straightforward with this method.

A crucial element of the procedure is the selection of the distribution to represent the data accurately. When the analytical form of the probability density function (pdf) is known, it is possible to calculate its value for a given combination of  $\vec{x}_{\text{obs}}$  and  $\theta$  to evaluate the likelihood, known as an unbinned ML fit. This method is suitable, for example, for the invariant mass spectrum of a resonance, which has a known analytical form. However, in the  $H \rightarrow WW$  channel, the undetected neutrinos in the final state make it impossible to experimentally reconstruct the invariant mass of the Higgs boson. Therefore, other observables, as described in the previous section, must be used. Unfortunately, an analytical form of the distributions of these observables is not available, and a binned ML fit is used instead. The significant difference is that the observable of interest is binned, i.e., a histogram of observations is constructed. This process is performed twice: once with the data sample and once with a sample from MC simulation. Therefore, the parameters  $\theta$  and the observations  $\vec{x}_{\text{obs}}$  are transformed into the numbers of expected and observed entries in each histogram bin, represented by the sum of expected signal and background events  $s_i + b_i$  and  $n_i$ , respectively. The histogram is an approximation of the unknown functional form of the pdf. Even if the analytical expression of the underlying pdf is unknown, the number of entries  $n_i$  in each bin of the histogram follows a Poisson distribution  $\mathcal{P}$ , which allows the likelihood to be explicitly written as:

$$L = \prod_{i=1}^{N_{\text{bins}}} \mathcal{P}(n_i, s_i + b_i). \quad (5.4)$$

The signal and background histograms, or templates, in the observable of choice obtained from MC simulation or other data-driven methods determine the expected signal and background entries in each bin  $s_i$  and  $b_i$ . The likelihood function is maximized in terms of the normalization of the signal histogram in order to perform a template fit. In this analysis, the background contributions that are not measured from data are set to their SM prediction within the corresponding systematic uncertainties. Additionally, the signal templates are treated similarly, except for their normalization, which is unconstrained, by introducing the signal strength modifier  $\mu$ :

$$\mu = \frac{(\sigma \times \mathcal{B})_{\text{signal}}}{(\sigma \times \mathcal{B})_{\text{SM signal}}}, \quad (5.5)$$

where  $\sigma$  is the cross-section and  $\mathcal{B}$  the Branching Fraction introduced in Section 1. The Likelihood function for signal templates results as:

$$L = \prod^{N_{\text{bins}}} \mathcal{P}(n_i, \mu \cdot s_i^{\text{SM}} + b_i^{\text{SM}}). \quad (5.6)$$

It should be acknowledged here that Eq. 5.6 ignores the signal model already defined in the previous chapter which is a function of both  $\mu$  and  $f_{ai}$ . For the given example, we deal with a simple signal model dependent on 1 parameter only.

## 5.2 Systematic uncertainties

This chapter describes how the systematic uncertainties associated with the background and signal templates used to fit the data are added to the likelihood function as *nuisance parameters*. The nuisance parameters account for how the normalization or the shape of the templates change and also for the theoretical uncertainties.

### Normalization uncertainty:

Some systematic uncertainties only affect the normalization of the templates and thus, their inclusion in the likelihood function is done by multiplying a factor (as  $\nu_{\text{norm}}$ ) to all bins.

$$L = \prod^{N_{\text{bins}}} \mathcal{P}(n_i, \mu \cdot s_i^{\text{SM}} \cdot \nu_{\text{norm}} + b_i^{\text{SM}} \cdot \nu_{\text{norm}}) \cdot \mathcal{N}(\nu_{\text{norm}}). \quad (5.7)$$

where  $\mathcal{N}(\theta)$  describes the distribution of  $\theta$  and is associated to a log-normal constraint in the likelihood function:

$$\mathcal{N}(\nu_{\text{norm}}) d\nu_{\text{norm}} = \frac{d\nu_{\text{norm}}}{\sqrt{2\pi} \nu_{\text{norm}} \log(\kappa)} \exp\left(-\frac{(\nu_{\text{norm}}/\tilde{\nu}_{\text{norm}})^2}{2(\log \kappa)^2}\right), \quad (5.8)$$

with a width driven by  $\kappa$  and centered in  $\nu_{\text{norm}} = \tilde{\nu}_{\text{norm}} = 1$  which corresponds to the nominal normalization of the template. A wider distribution of  $\nu_{\text{norm}}$  translates into an increase of the uncertainty on the measurement of  $\mu$  and vice-versa. The

choice of a log-normal constraint on the nuisance is desirable in the case of event counts in histograms because they are physically bounded to positive values.

### Shape uncertainty:

In the case of uncertainties affecting the shape of the distributions, we construct three versions of the affected template: with the nominal value (*nom*), with the nominal value plus (+1) the associated uncertainty and with the nominal value minus (-1) the associated uncertainty. An interpolation bin-by-bin is performed to express the number of events in each bin  $N_i$  as a function of the nuisance parameter:

$$N_i(\nu_{\text{shape}}) = N_i(\text{nom})(1 - \nu_{\text{shape}}^2) + \frac{\nu_{\text{shape}}(\nu_{\text{shape}} + 1)}{2} N_i(+1) + \frac{\nu_{\text{shape}}(\nu_{\text{shape}} - 1)}{2} N_i(-1). \quad (5.9)$$

In this case, a Gaussian distribution is used in the likelihood function:

$$\mathcal{N}(\nu_{\text{shape}}) d\nu_{\text{shape}} = \frac{d\nu_{\text{shape}}}{\sqrt{2\pi}} \exp\left(-\frac{\nu_{\text{shape}}^2}{2}\right). \quad (5.10)$$

The inclusion of both kinds of nuisance parameters naturally increases the dimension of the likelihood, but we reduce it by profiling the likelihood over the nuisance parameter when calculating the uncertainty on the POI. The test statistics chosen for the minimization of the profile likelihood is:

$$\ell(\mu) = \frac{L(\mu, \hat{\nu}_{\text{shape}})}{L(\hat{\mu}, \hat{\nu}_{\text{shape}})}, \quad (5.11)$$

where  $\hat{\mu}$  is the best-fit value of the POI,  $\nu_{\text{shape}}$  ( $\hat{\nu}_{\text{shape}}$ ) is the array of the (best-fit) derived nuisance parameters values, and  $\hat{\nu}_{\text{shape}}$  maximizes the likelihood for a given  $\mu$ . The 68% CL corresponds to the interval  $-\log \ell(\mu) < 0.5$ . Indeed,  $-\log(\ell(\mu))$  is always 0, otherwise the 68% CL would generally correspond to  $-\log(\ell(\mu)) + \log(\ell(\mu_{\min})) < 0.5$ .

Taking all this into consideration and gathering the normalization ( $\nu_{\text{norm}}$ ) and shape uncertainties ( $\nu_{\text{shape}}$ ) as a full set of nuisance parameters  $\theta$ , we may define a likelihood function for this analysis:

$$\mathcal{L}(\text{data}|\mu_{\text{ggH}}, \mu_{\text{EW}}, f_{ai}, \theta) = \prod_i \mathcal{P}(n_i|s_i(\mu_{\text{ggH}}, \mu_{\text{EW}}, f_{ai}, \theta) + b_i(\theta)) p(\tilde{\theta}|\theta), \quad (5.12)$$

where  $n_i$  is the observed number of data events in each bin  $i$ . The signal and background expectations ( $s_i$  and  $b_i$ , respectively) correspond to binned templates of the multidimensional discriminants described in Tab. 4.14 for both interpretations. Each signal process is parametrized as a linear combination of terms from SM and AC contributions and their interference as explained in section 4.1 where the signal model was defined. The signal expectation depends on the signal strength modifiers  $\mu_{ggH}$  and  $\mu_{EW}$  for the ggH and the VBF/VH signals, respectively, and on  $f_{ai}$  which determines the shape of the signal expectation. Those three parameters are constrained by the fit to the data. Additionally, both the signal and background expectations are functions of  $\theta$ . The  $p(\tilde{\theta}|\theta)$  term is the pdfs of the uncertainties on the nominal values of the nuisance parameters  $\tilde{\theta}$ .

### 5.3 Data-driven background normalization

The background contributions are measured in dedicated control regions, and we use *rate parameters* calculated per year. By definition, instead of giving rise to constraint terms in the likelihood function, they multiply a background yield in the signal region (SR) and the control region (CR). In the case of considering only one background, the likelihood is:

$$L = \prod_{N_{\text{bins}}^{SR}} (n_i, s_i + \alpha \cdot b_i) \prod_{N_{\text{bins}}^{CR}} (n_j, \alpha \cdot b_j), \quad (5.13)$$

$\alpha$  is the rate parameter. The rate parameter determined in the CR propagates the measured background normalization to the SR. It should be noticed that the superscript  $N_{\text{bins}}^{CR} = 1$  in Eq. 5.13, as we only consider the number of events in the CRs. Expanding the expression above for the three main backgrounds of the HWW channel (top, DY $\tau\tau$ , and WW) is trivial.

### 5.4 Sources of systematic uncertainties

The systematic uncertainties are divided into experimental or theoretical. The considered systematic uncertainties are summarized in Tables 5.1 and 5.2, where we specify the type of uncertainty, their degree of correlation, and the affected processes.

| Systematic uncertainty          | Type  | Correlation            | Process               |
|---------------------------------|-------|------------------------|-----------------------|
| Luminosity                      | norm  | Partially correlated   | All MC but WW, top    |
| Trigger efficiency              | shape | Uncorrelated           | All MC                |
| Trigger efficiency (DY)         | shape | Uncorrelated           | Data-driven DY        |
| Pre-firing weight               | shape | Uncorrelated 2016-2017 | All MC                |
| Lepton ID eff. (SR)             | shape | Uncorrelated           | All MC+Data-driven DY |
| Lepton ID eff. (CR)             | shape | Uncorrelated           | All MC+Data-driven DY |
| Lepton $p_T$ scale              | norm  | Uncorrelated           | All MC+Data-driven DY |
| Jet energy scale                | norm  | Partially correlated   | All MC                |
| Jet energy resolution           | norm  | Uncorrelated 2017-2018 | All MC                |
| AK8 JMS and JMR                 | norm  | Uncorrelated           | All MC                |
| V-tag SF                        | norm  | Uncorrelated           | Higgs, VZ and VVV     |
| MET energy scale                | norm  | Uncorrelated           | All MC                |
| Jet PU ID scale                 | shape | Uncorrelated           | All MC                |
| Fake rate (stat.+syst. origin)  | shape | Uncorrelated           | Non-prompt            |
| Fake rate (30% jet composition) | norm  | Uncorrelated           | Non-prompt            |
| B-tag SF                        | shape | Partially correlated   | All MC                |

TABLE 5.1: List of experimental systematic uncertainties considered and their correlation. Partial correlation corresponds to CMS experiment recommendations.

Some additional information:

- The total uncertainty associated with the integrated luminosity is 1.2%, 2.3%, and 2.5% for 2016 [143], 2017 [98], and 2018 [99], respectively.
- The trigger efficiency is derived from the T&P method and it affects the normalization ( $< 1\%$ ) and shape of the signal and background distributions.
- Systematic uncertainties of about 1% for electrons and 2% for muons are considered to cover any discrepancy in the lepton identification and isolation efficiencies measured between data and MC.
- The lepton momentum scale is also another source of uncertainty due to the limited data sample used for the estimation. The impact on the normalization of the signal and background templates ranges between 0.6 – 1.0% for electrons and 0.2% for muons.
- The jet energy scale uncertainty is modeled by eleven independent nuisance parameters corresponding to different correction sources. Their effects vary between 1 – 10% according mainly to the jet multiplicity in the analysis phase-space. The uncertainty associated to the jet smearing adds a normalization uncertainty between 1 – 5% depending on the jet multiplicity. Such effects are considered for AK4 and AK8 jets, additionally, for the latter, we

need to consider the boosted V-jet mass scale and resolution. V-tagging corrections with their corresponding uncertainties are considered for V-tagged AK8 jets. These variables are calibrated in a quark-initiated top sample enriched in hadronically decaying W bosons [144].

- The normalization systematic uncertainty of the missing transverse momentum takes into account the effects of the unclustered energy scale, jet energy scale, and the lepton  $p_T$  resulting in a variation of 1 – 10%.
- The jet PU identification uncertainty affects the normalization and shape of the signal and background templates and is found to be less than 1%.
- The b-tagging efficiency is modeled by seventeen nuisance parameters.
- The non-prompt background estimation is limited by the size of the data used for the misidentification rate measurements. The difference in the flavor composition of jets misidentified as leptons are treated as a nuisance parameter with an effect between a few percent to about 10% depending on the signal region. The normalization uncertainty of 30% [145] fully covers the discrepancies with respect to data.
- The limited number of simulated events is a source of statistical systematic uncertainty.

| Systematic uncertainty                          | Type                        | Correlation                               | Process                          |
|---|-----------------------------|---|----------------------------------|
| PDF (cross-section and acceptance)              | norm                        | Correlated                                | All MC but WW,top,DY             |
| QCD higher order                                | norm (Bkg)<br>shape (Higgs) | Correlated                                | All MC                           |
| Pileup modeling                                 | norm                        | Uncorrelated                              | WW,top,DY<br>ggF,VBF(SM+AC)      |
| Parton shower                                   | shape                       | Correlated                                | All MC                           |
| Underlying event                                | norm                        | Uncorrelated 2016<br>Correlated 2017-2018 | All MC but WW,top                |
| WW NNLL resummation                             | shape                       | Uncorrelated per cat.                     | WW                               |
| Single top/ $t\bar{t}$ composition              | shape                       | Correlated                                | top                              |
| $V\gamma$ , $V\gamma^*$ and $VZ$ cross-sections | norm                        | Correlated                                | $V\gamma$ , $V\gamma^*$ and $VZ$ |
| CR/SR acceptance                                | norm on CRs                 | Correlated                                | DY,top                           |

TABLE 5.2: List of considered theoretical systematic uncertainties and their correlation.

About the theoretical uncertainties:

- The uncertainties related to the choice of PDFs and  $\alpha_s$  have a minor effect on the shape of the distributions, and thus, only normalization effects related

to the event acceptance and to the cross-section are considered. These only apply to signals since the normalization for major backgrounds is constrained through data in dedicated CRs. For the Higgs boson signal processes, these uncertainties are calculated and provided by the LHC Higgs cross-section working group [48].

- The corresponding renormalization  $\mu_R$  and factorization  $\mu_F$  scales varied by factors of 0.5 or 2, and the envelopes of the varied templates were taken as one standard deviation. The background processes with their normalization constrained through data in dedicated CRs consider only the shape effect of the uncertainties. The non-resonant WW background has uncertainties derived by varying  $\mu_R$ ,  $\mu_F$ , and the resummation scale. For the ggH and VBF signal processes, this uncertainty is decoupled into multiple sources according to Ref. [48].
- The PU modeling is considered for the main backgrounds (top, DY $\tau\tau$ , and WW) and for the ggH and VBF signals. The effect is determined by varying the total inelastic pp cross-section (69.2 mb [146, 147]) within the assigned 5% uncertainty.
- The PS modeling affects the jet multiplicity, which causes the migration of events between categories resulting in template shape changes. These uncertainties are calculated by reweighting events with varied PS weights computed by PYTHIA 8.212.
- The uncertainty associated with the modeling of underlying events is evaluated by varying UE tunes used in the MC simulation. It has a minimal effect on the template shapes while affecting the normalization by about 1.5%.
- The gluon-initiated non-resonant WW production is corrected to NLO accuracy with the corresponding uncertainty of 15% applied [148]. The relative fraction with respect to the single top process and the  $t\bar{t}$  production is corrected with a systematic uncertainty of 8%. Additional uncertainties associated with a specific process, such as DY, VY,  $V\gamma$ , or  $V\gamma^*$ , are assigned respectively to the corrections applied to account for discrepancies between data and MC [93].

## 5.5 Impact of the uncertainties on the results

This section describes the impact of each systematic uncertainty on the signal cross-section fractions,  $f_{ai}$ , which is the shift in the parameter of interest when varying a systematic uncertainty up and down by  $1\sigma$ . The purpose of this study is to evaluate the correlation between the systematic source and the cross-section fraction parameter. No impactful systematic parameters were found to be noticeably pulled for the presented analysis. Asymmetric error bars assigned to each pull, defined as post-fit uncertainty divided by pre-fit uncertainty, should be smaller than  $\pm 1$  for systematic to be constrained. This was checked not to be true for all relevant high-ranked nuisance parameters.

The 30 most relevant observed impacts of the systematic uncertainties on the  $a_3$  coupling in the AC and SMEFT interpretations of the analysis can be seen in Fig. 5.3 and Fig. 5.4, respectively. The remaining impacts are gathered in Appendix C for all couplings. The right-hand side of impacts plots shows the observed impacts for small, in our case, non-zero values of  $f_{ai}$ , while the left-hand side displays the “pull” quantity  $(\theta - \theta_0)/\Delta\theta$ , where  $\theta$  and  $\theta_0$  are post-fit and pre-fit nuisance parameter values, and  $\Delta\theta$  is the pre-fit uncertainty. It should be noted that the observed impacts correspond to the  $f_{ai}$  best-fit value, whereas the expected impacts (shown in Appendix C) are displayed for small, non-zero values of  $f_{ai}$ . The signal strength parameters,  $\mu_V$ , and  $\mu_F$  have a significant impact on the result, and their precision is limited by the number of signal Higgs events. To provide further clarification on some examples, the parameters with the JES tag refer to uncertainties in jet energy scale. Parameters beginning with “prop” correspond to bin-by-bin statistical uncertainties. The names containing “THU” refer to Theoretical Uncertainties. Parameters containing “norm” refer to normalization, while those containing “btag” refer to btagging efficiency.

## 5.6 Goodness of fit

The presented analysis relies solely on MC simulations in the signal region, while the agreement between the templates and data is exclusively checked in orthogonal control regions. We do not look at the data in the signal region until the analysis structure is consistent. The CMS Collaboration implements this “blind” procedure to ensure unbiased results. During the unblinding process, one of the steps



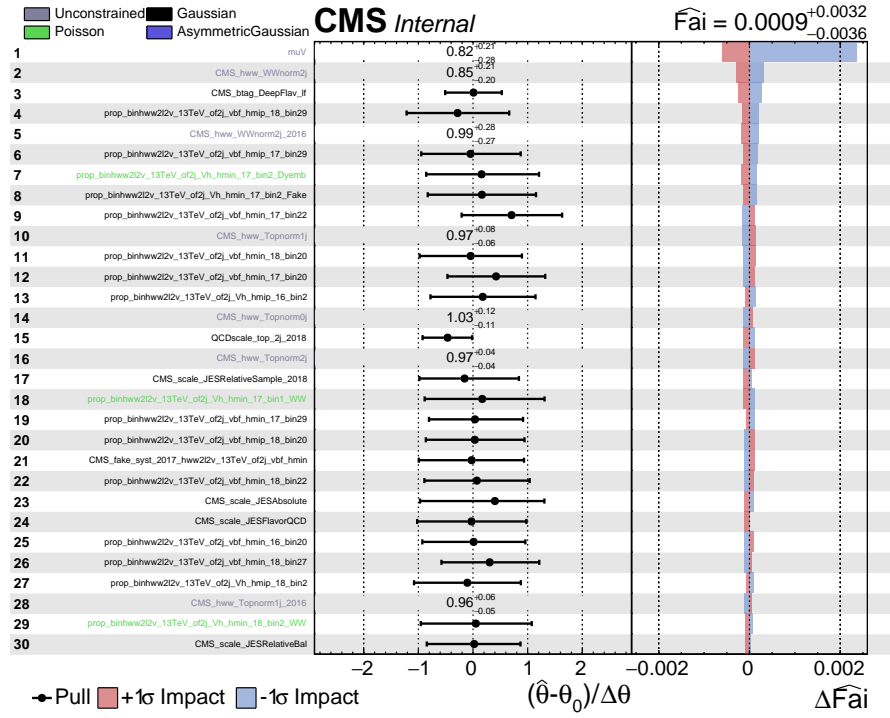


FIGURE 5.3: Observed impacts of systematic uncertainties on the cross-section fraction  $f_{a3}$  for the AC interpretation of the analysis.

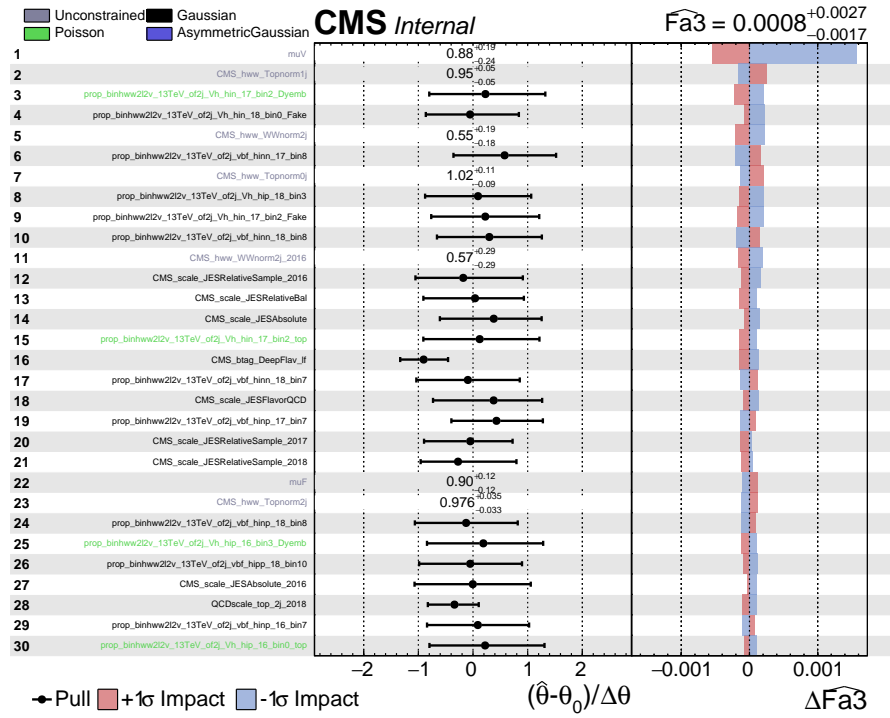


FIGURE 5.4: Observed impacts of systematic uncertainties on the cross-section fraction  $F_{a3}$  in the context of the SMEFT interpretation.

involves evaluating the adequacy of the MC templates in reproducing the data after the fit. This assessment is referred to as the Goodness of Fit (GOF), which is performed using a hypothesis test based on likelihood ratios. The Neymann-Pearson lemma [149] is utilized to maximize the discriminating power of this test. Specifically, for this analysis, the saturated [150] model of the GOF is employed to calculate a p-value, indicating how well the set of parameters extracted by the fit describes the data. The p-values obtained for both approaches are presented in Table 5.3 for reference.

TABLE 5.3: p-values corresponding to the Goodness of fit tests for the AC (Approach 1) and SMEFT (Approach 2) interpretations of the analysis.

| <b>p-values</b> | $f_{a3}$ | $f_{a2}$ | $f_{\Lambda 1}$ | $f_{\Lambda 1}^{Z\gamma}$ |
|-----------------|----------|----------|-----------------|---------------------------|
| Approach 1      | 0.006    | 0.281    | 0.009           | 0.015                     |
| Approach 2      | 0.091    | 0.093    | 0.088           | -                         |

The lower performance observed for some of the couplings was found to originate from the 1-jet ggH channel. This observation aligns with a similar conclusion made in Ref. [93], where multiple tests and studies were conducted without identifying any specific issues.

## 5.7 Post-fit distributions

Figures 5.5 - 5.11 present the KD for all channels as obtained after the final full Run 2 fit. The “Total signal” (blue histogram) stands for the sum of yields for all SM and AC signal templates, including the interference. The AC signal contribution was found to be strongly suppressed in the fit, and the total extracted signal is comparable to pure SM distribution, as can be observed in comparison with pre-fit distributions in Section 4.7

### HVV discriminants in the VBF and VH categories (AC)

Figures 5.5 - 5.8 display the post-fit distributions for the VBF and VH channels. The predicted Higgs boson signal and background distributions are shown after the fit to the data. The Higgs boson signal is shown both stacked on top of the backgrounds and superimposed. The uncertainty band corresponds to the total systematic uncertainty. The lower panel in each figure shows the ratio of the number of events observed to the total prediction.

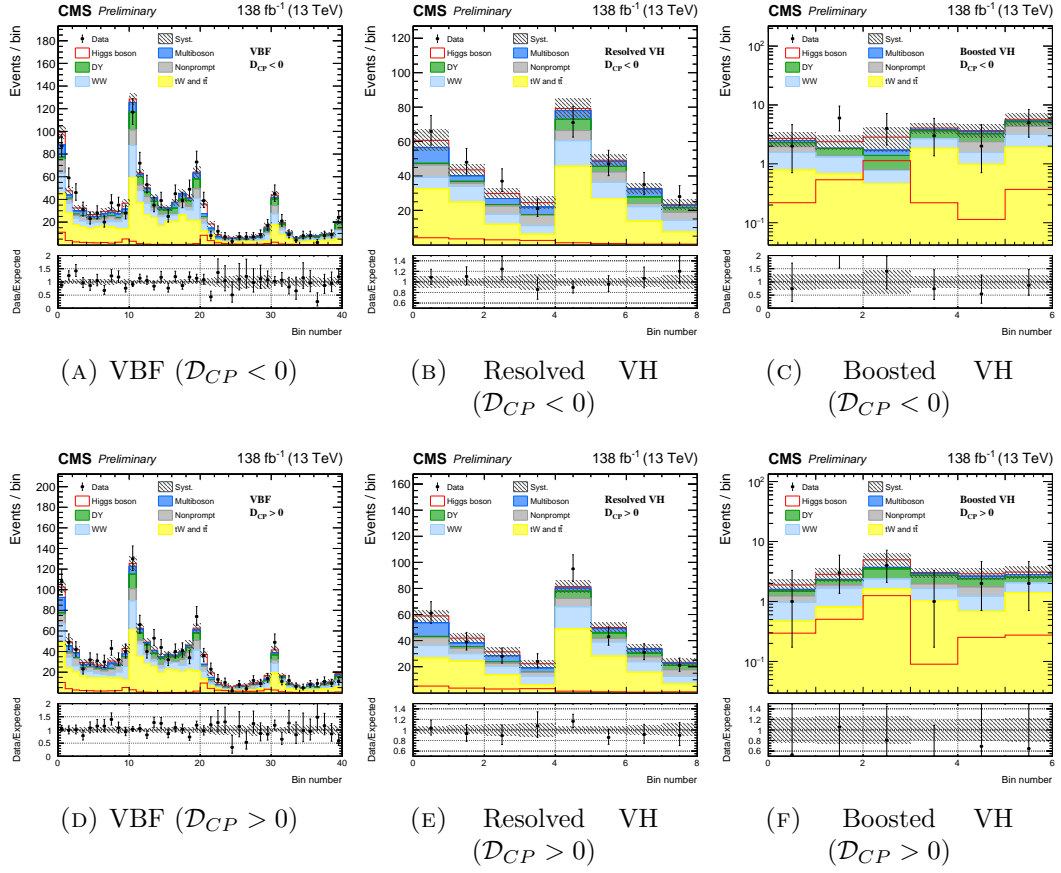


FIGURE 5.5: Multidimensional KD targeting the  $a_3$  HVV coupling for the VBF (A, D), Resolved VH (B, E), and Boosted VH (C, F).

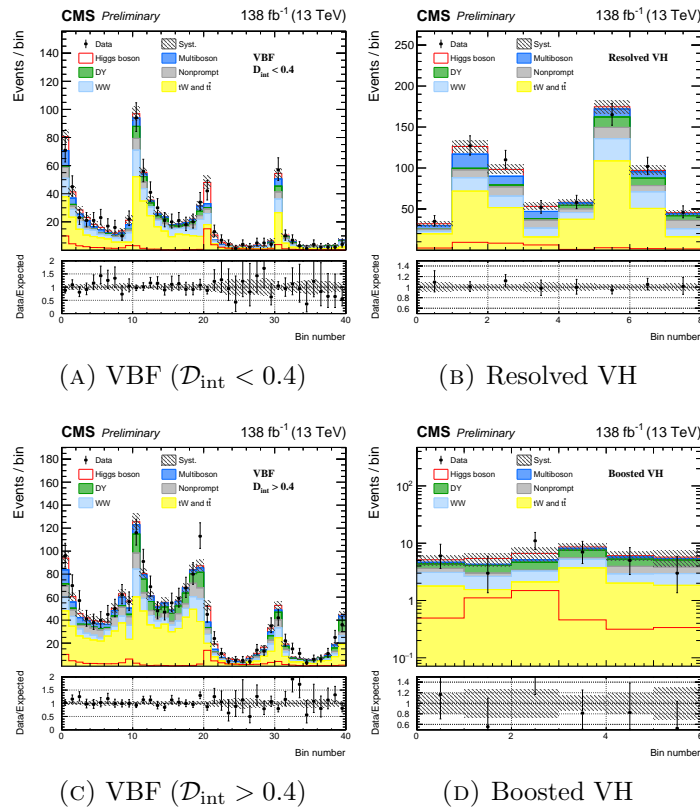


FIGURE 5.6: Multidimensional KD targeting the  $a_2$  HVV coupling for the VBF (A, C), Resolved VH (B) and Boosted VH (D) categories.

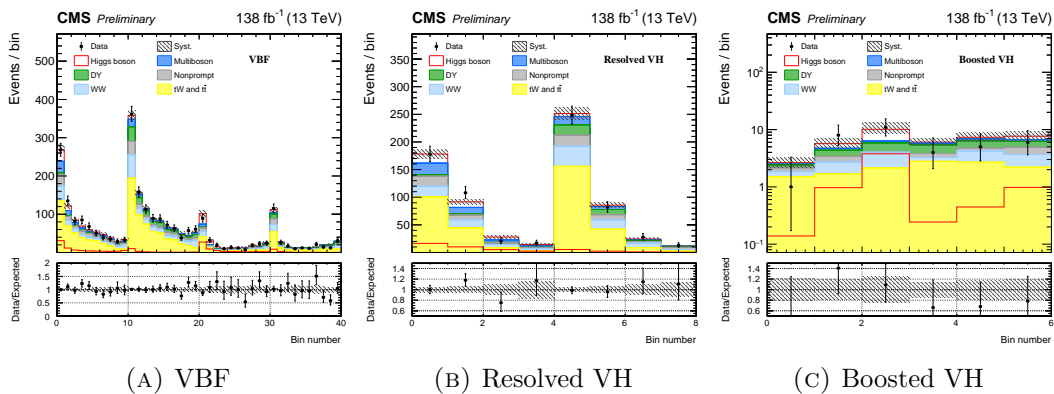


FIGURE 5.7: Multidimensional KD targeting the  $\kappa_1\Lambda_1$  HVV coupling for the VBF (A), Resolved VH (B), and Boosted VH (C) categories.

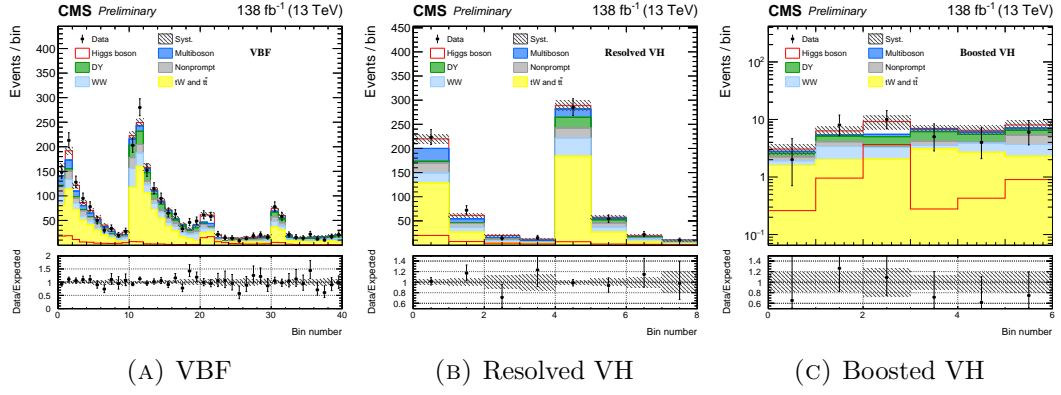


FIGURE 5.8: Multidimensional KD targeting the  $\kappa_2^{Z\gamma}\Lambda_1^{Z\gamma}$  HVV coupling for the VBF (A), Resolved VH (B) and Boosted VH (C) categories.

## HVV discriminants in 0-jet and 1-jet ggH categories

Figure 5.9 shows the post-fit Run 2 discriminants for the 0-jet and 1-jet ggH channels.

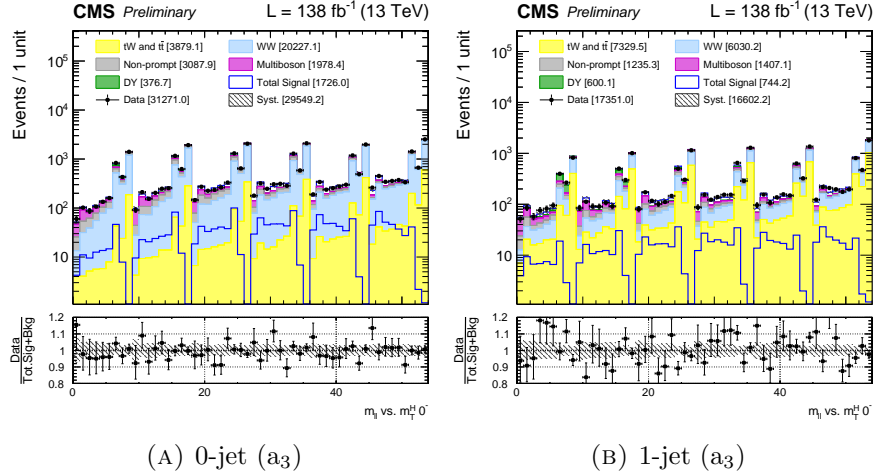


FIGURE 5.9: Two dimensional  $m_{\ell\ell}$  vs.  $m_T$  distributions in the 0-jet channel (A) and the 1-jet channel (B) after the final full Run 2 fit.

## Hgg discriminants in 2-jet ggH category

Figure 5.10 shows the post-fit Run 2 discriminants for the 0-jet and 1-jet ggH channels where the post-fit Higgs boson signal is plotted.

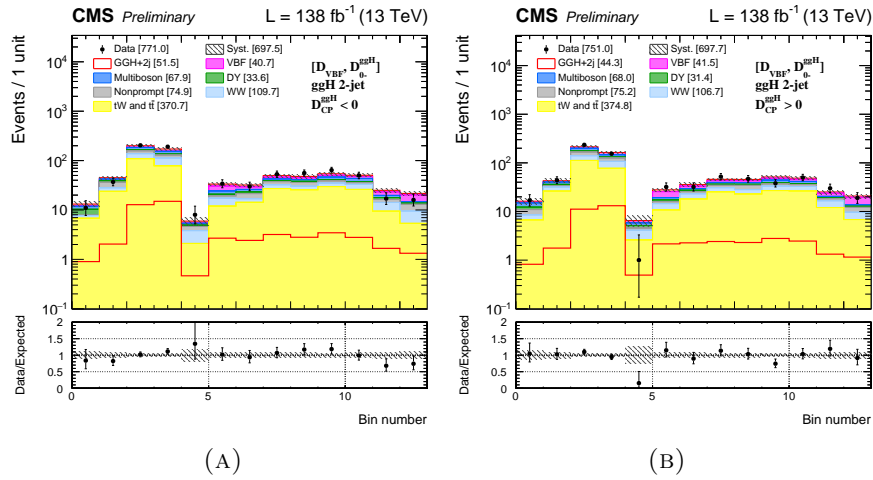


FIGURE 5.10: Distribution of events in combination of Run-2 datasets after the final full Run 2 fit for  $[\mathcal{D}_{VBF}, \mathcal{D}_0^{ggH}]$  in the ggH + 2 jets channel. The  $\mathcal{D}_{CP}^{ggH} < 0$  (left) and  $\mathcal{D}_{CP}^{ggH} > 0$  (right) categories are shown. The  $a_1$  and  $a_3$  Hgg coupling signals are considered in the fit.

## HVV discriminants in the VBF and VH categories (SMEFT)

Figure 5.11 shows the post-fit distributions from the full Run 2 dataset, which are utilized for the SMEFT interpretation of the analysis. These distributions enable the simultaneous constraint of the parameters  $a_3$ ,  $a_2$ , and  $\Lambda_1$ .

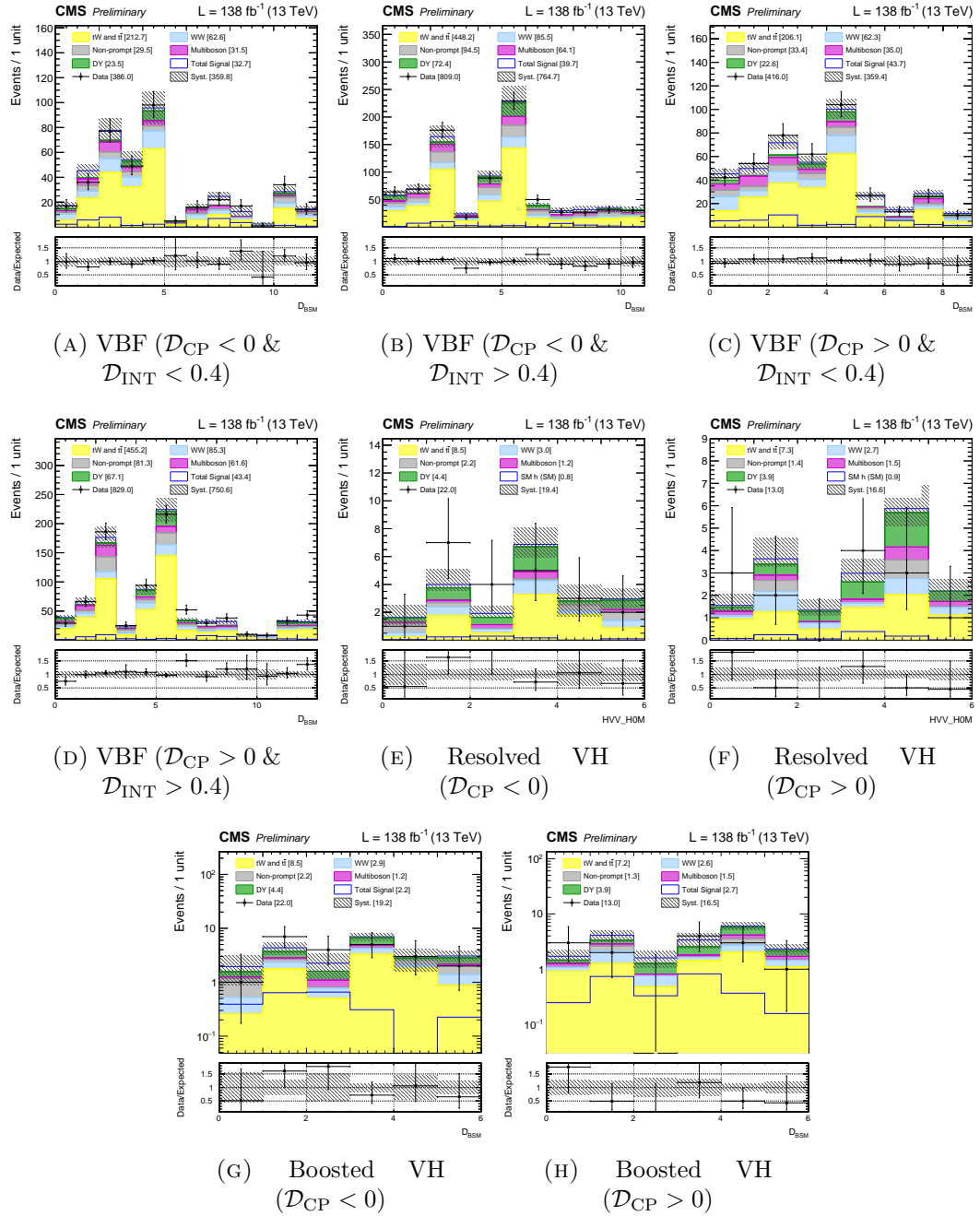


FIGURE 5.11: SMEFT Multi-KD combining the full Run 2 dataset. Four distributions correspond to the VBF channel (A, B, C and D), two to the Resolved VH (E and F) and two to the Boosted VH categories (G and H).





# Chapter 6

## Results

Signal templates and background distributions of all channels and data-taking periods are combined and evaluated by a binned maximum likelihood fit to obtain the final results. The likelihood function is defined for candidate events by Eq. 5.12 as stated in the previous chapter 5. The binned templates of multidimensional discriminants, as outlined in section 4.1, are representative of the various processes. Each signal process is expressed as a linear combination of terms arising from the SM, anomalous couplings, and their interference as explained in section 4.8. The parameters  $\mu_{\text{ggH}}$ ,  $\mu_{\text{EW}}$ , and  $f_{ai}$  determine the signal expectation and are constrained through the fit to data. The expectations of both signal and background are functions of  $\theta$ , which is the complete set of nuisance parameters accounting for the considered systematic uncertainties. The signal strength modifiers  $\mu_{\text{ggH}}$  and  $\mu_{\text{EW}}$  are associated with the ggH and VBF/VH signal yields, respectively, which are related to each other due to the same HVV couplings appear in both the production and decay of the Higgs boson. Conversely, the ggH signal is mainly initiated by the top fermion couplings and has no relationship with VBF and VH production mechanisms. Since the signal strength modifiers are free parameters in the fit, the overall signal event yield cannot differentiate between alternative signal hypotheses. The  $f_{ai}$  parameter refers to the anomalous coupling cross-section fraction, which defines the signal expectation shape. The cross-section fraction for the SM coupling is assumed to be  $1 - |f_{ai}|$ . Approach 1 investigates the SM and only one anomalous HVV coupling at a time, making it possible to study each  $f_{ai}$  independently. Depending on the anomalous coupling being constrained,  $f_{ai}$  can represent  $f_{a3}$ ,  $f_{a2}$ ,  $f_{\Lambda_1}$ , or  $f_{\Lambda_1}^{Z\gamma}$ . Approach 2, on the other hand, considers the SM and three anomalous HVV couplings simultaneously, with  $f_{ai}$  representing

either  $f_{a3}$ ,  $f_{a2}$ , and  $f_{\Lambda_1}$ . To avoid exploring an unphysical parameter space,  $|f_{a2}| + |f_{a3}| + |f_{\Lambda_1}|$  must not exceed 1. Finally, the  $p(\tilde{\theta}|\theta)$  term in Eq. 5.12 represents the probability density functions of the uncertainties associated with the nominal values of the nuisance parameters  $\tilde{\theta}$ . The systematic uncertainties that impact only the normalizations of the signal and background processes are treated as log-normal pdfs while shape-altering systematic uncertainties are treated as Gaussian pdfs. The likelihood function is optimized concerning both the signal modifier parameters and nuisance parameters. The corresponding CL intervals are determined by profile likelihood scans of the parameters. Using the profile likelihood function,  $-2\Delta \ln \mathcal{L} = 1.00$  and  $3.84$ , the 68% and 95% CL intervals, respectively, are established. The likelihood value for a given  $f_{ai}$  is determined by the signal shape of the hypothesis and the relative signal event yields across categories.

## 6.1 AC interpretation

Table 6.1 displays an example of the number of events for  $a_3$  coupling before and after performing the fit described in the previous chapter. The obtained yields are in accordance with the expected values.

TABLE 6.1: Expected (pre-fit) and observed (post-fit) number of events per category for background and signals in Approach 1 for the  $a_3$  coupling.

| Approach 1<br>$a_3$         | Post-fit (Pre-fit) |           |            |            |          | Pre-fit   |            | Data | Post-Fit     |           |
|-----------------------------|--------------------|-----------|------------|------------|----------|-----------|------------|------|--------------|-----------|
| Category/<br>Process        | Top                | WW        | Non-prompt | Multiboson | DY       | SM Signal | BSM Signal |      | Total Signal | Total Bkg |
| VBF<br>$D_{CP} < 0$         | 596 (574)          | 263 (239) | 124 (118)  | 102 (97)   | 100 (96) | 88        | 203        | 1195 | 65           | 1125      |
| VBF<br>$D_{CP} > 0$         | 590 (579)          | 262 (238) | 100 (132)  | 96 (92)    | 92 (92)  | 85        | 206        | 1246 | 79           | 1132      |
| Resolved VH<br>$D_{CP} < 0$ | 171 (169)          | 73 (66)   | 32 (39)    | 29 (34)    | 20 (19)  | 18        | 68         | 353  | 16           | 327       |
| Resolved VH<br>$D_{CP} > 0$ | 177 (171)          | 78 (70)   | 36 (34)    | 32 (32)    | 20 (19)  | 19        | 65         | 342  | 18           | 325       |
| Boosted VH<br>$D_{CP} < 0$  | 7 (7)              | 5 (4)     | 2 (3)      | 1 (1)      | 4 (4)    | 2         | 84         | 22   | 3            | 22        |
| Boosted VH<br>$D_{CP} > 0$  | 7 (6)              | 5 (4)     | 1 (1)      | 2 (1)      | 4 (4)    | 2         | 82         | 13   | 3            | 13        |

We now present the likelihood scans for the examined cross-section fractions of the HVV and Hgg vertices, considering both expected and observed scenarios. The obtained results align with the predictions of the SM.

### 6.1.1 HVV couplings in VBF and VH channels

Figure 6.1 depicts the likelihood scans for  $f_{a3}$ ,  $f_{a2}$ ,  $f_{\Lambda 1}$ , and  $f_{\Lambda 1}^{Z\gamma}$  variables in both expected and observed scenarios for the combination of the VBF, resolved VH, and boosted VH channels using the complete Run 2 dataset. The features anticipated in section 4.1 through the analysis of discriminant shapes are visible, with the likelihood sharply increasing around the expected minimum of 0. Notably, significant interference effects occur for negative values of  $f_{a2}$ , particularly around -0.25, and positive values of  $f_{\Lambda 1}$ , mainly around 0.5. These values are predicted to exhibit substantial changes in shape concerning the SM. An example worth discussing is depicted in Fig. 6.1 (A), where a notable difference between the expected and observed curves is seen. This discrepancy arises due to the inclusion of the floating parameter  $\mu_F$  in the fit, while the 0-jet and 1-jet ggH categories are excluded. Consequently, in the full combination of channels, the observed and expected curves appear more similar, as we will demonstrate later.

The 0-jet and 1-jet channels are not sensitive to  $\Lambda_1^{Z\gamma}$  coupling but are still included to constrain the ggH signal strength. On the other hand, the 0-jet and 1-jet channels are sensitive to anomalous effects in the remaining couplings, which we discuss before moving on to the full Run 2 combination of channels.

#### Effect of the Boosted VH category

Figure 6.2 exhibits the expected likelihood scans for  $f_{a3}$ ,  $f_{a2}$ ,  $f_{\Lambda 1}$ , and  $f_{\Lambda 1}^{Z\gamma}$  in black for the combination of VBF and resolved VH, which is mainly sensitive to the VBF channel. They also display the expected likelihood scans in blue for the combination of VBF, resolved VH, and boosted VH channels. By comparing these two combinations, the effect of the boosted VH channel is evaluated. The features anticipated from the discriminant shapes in section 4.8 are visible in the result, with the likelihood sharply rising around the minima at 0. The boosted VH category significantly reduces the width of the minima for  $\Lambda_1$  couplings. Additionally, significant interference effects are noticeable for negative values of  $f_{a2}$ , particularly around -0.25, and positive values of  $f_{\Lambda 1}$  around 0.5. Large changes in shape relative to the SM are predicted for these values in the signal model study.

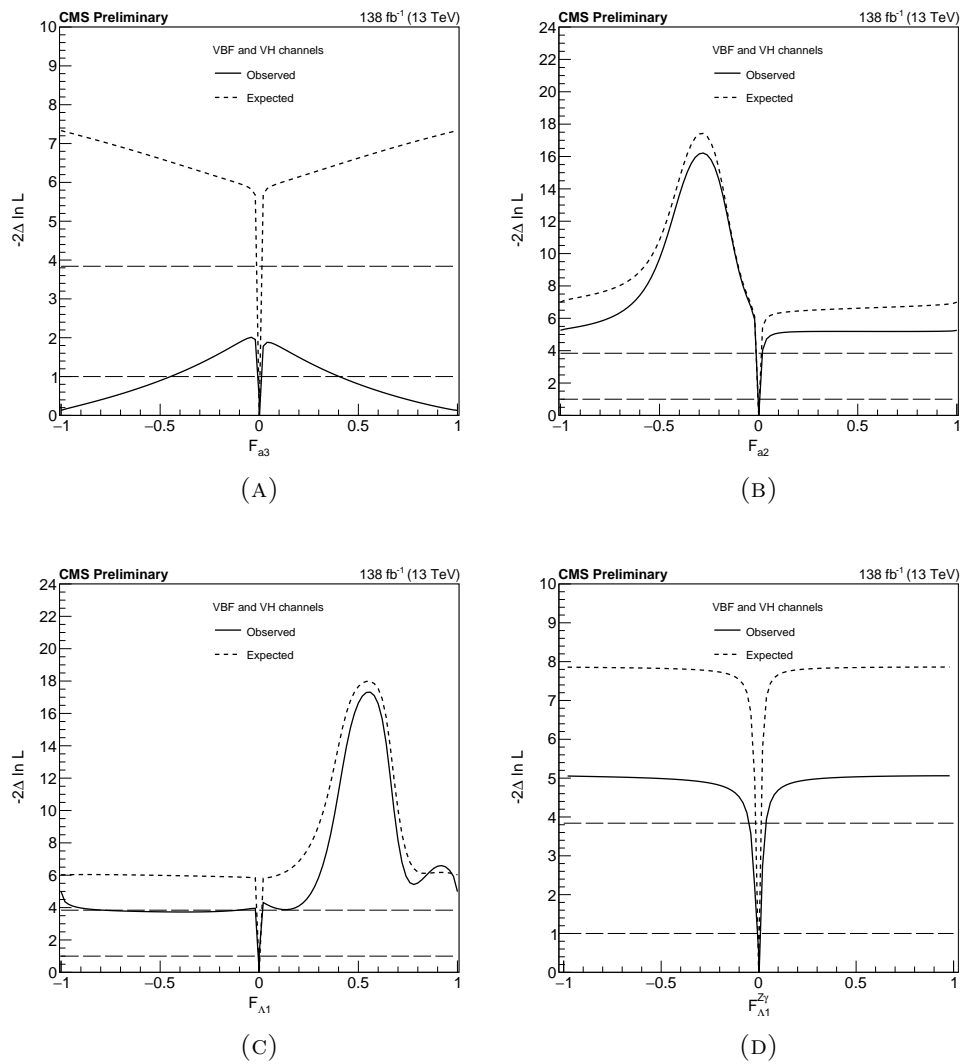


FIGURE 6.1: Expected and observed likelihood scans of  $f_{a3}$  (A),  $f_{a2}$  (B),  $f_{\Lambda 1}$  (C), and  $f_{\Lambda 1}^{Z\gamma}$  (D) combining the VBF and VH channels for the full Run 2 dataset.

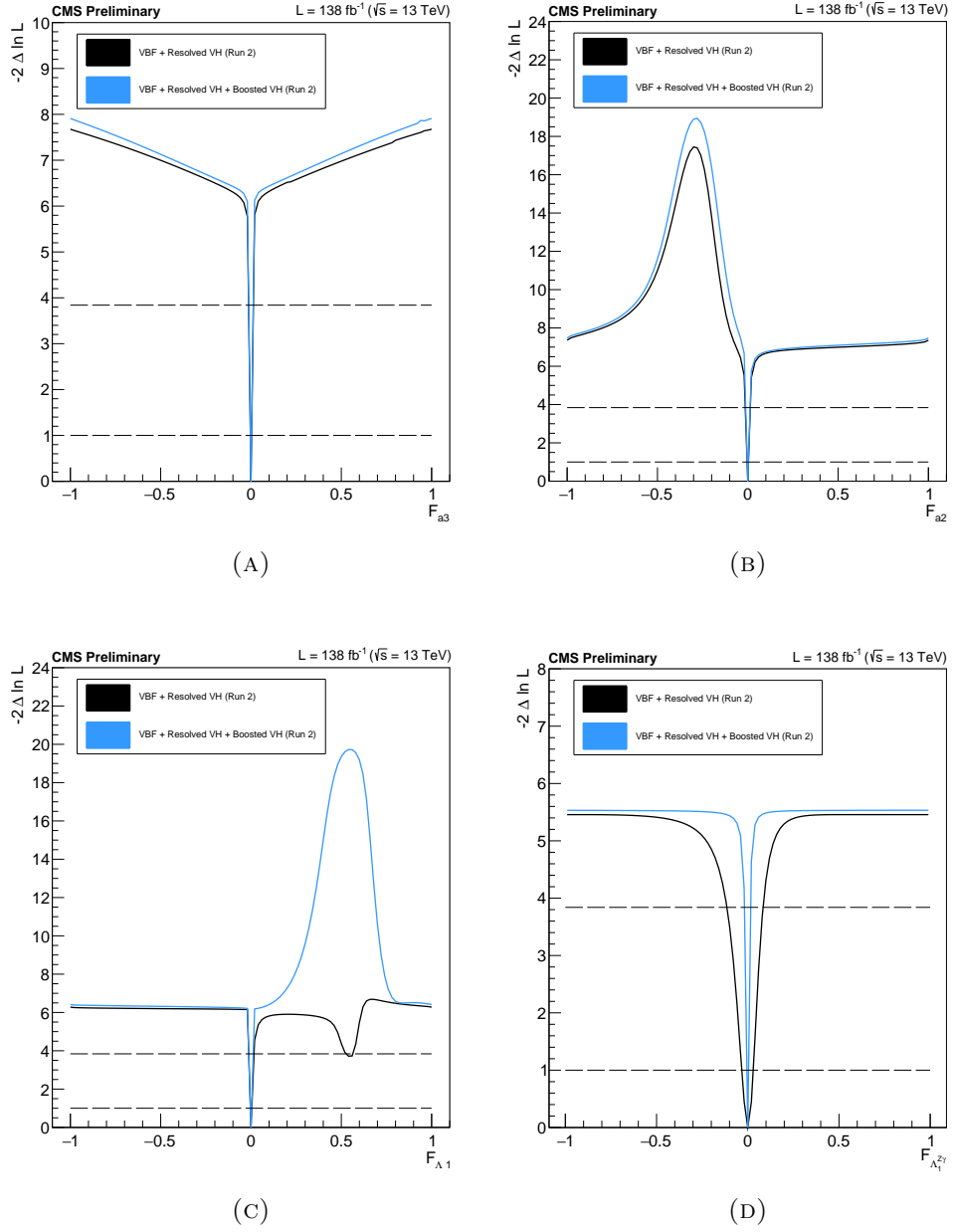


FIGURE 6.2: Expected likelihood scans of  $f_{a3}$  (A),  $f_{a2}$  (B),  $f_{\Lambda 1}$  (C), and  $f_{\Lambda 1}^{Z\gamma}$  (D) combining the VBF and resolved VH channels (black), and combining these with the boosted VH channel (blue) using the full Run 2 dataset.

### 6.1.2 HVV couplings in 0-jet and 1-jet ggH channels

Figure 6.3 displays the expected and observed likelihood scans for  $f_{a3}$ ,  $f_{a2}$ , and  $f_{A1}$  for the combination of the 0- and 1-jet channels in the full Run 2 datasets. The features predicted from the  $m_{ll}$  and  $m_T^H$  shapes in Section 4.8 are present. The most prominent features are the significant interference effects for  $f_{a2}$  around -0.25 and for  $f_{A1}$  around 0.5.

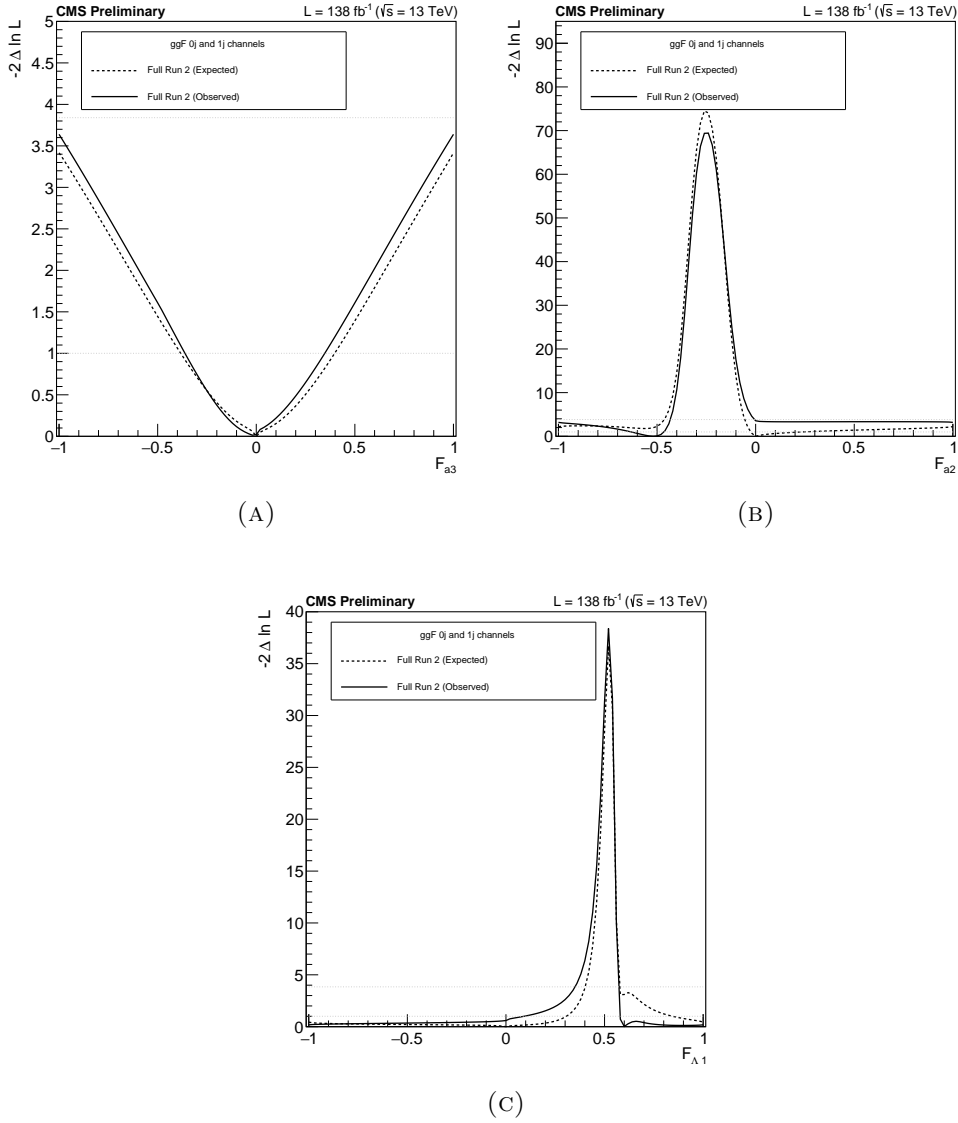


FIGURE 6.3: Full Run 2 expected and observed likelihood scans of  $f_{a3}$  (A),  $f_{a2}$  (B) and  $f_{A1}$  (C) combining the 0- and 1-jet ggH channels.

### 6.1.3 Full Run 2 combination for HVV couplings

In Fig. 6.4, we display the expected and observed likelihood scans for  $f_{a3}$ ,  $f_{a2}$ ,  $f_{\Lambda 1}$ , and  $f_{\Lambda 1}^{Z\gamma}$  for the Run 2 combination of datasets. It is important to note that the maximum  $-2\Delta \ln \mathcal{L}$  values obtained in the  $f_{a3}$  and  $f_{\Lambda 1}^{Z\gamma}$  scans are significantly lower than expected. This difference can be largely attributed to the post-fit  $\mu_{EW}$  values of 0.82 and 0.89 for the  $f_{a3}$  and  $f_{\Lambda 1}^{Z\gamma}$  fits, respectively, which can be compared to 0.95 and 0.97 for the  $f_{a2}$  and  $f_{\Lambda 1}$  fits. In the latter fits, the maximum  $-2\Delta \ln \mathcal{L}$  values are closer to the expected. It should be mentioned that the uncertainty on  $\mu_{EW}$  is approximately 20% in each case, indicating that all these values are consistent with both the SM and with each other.

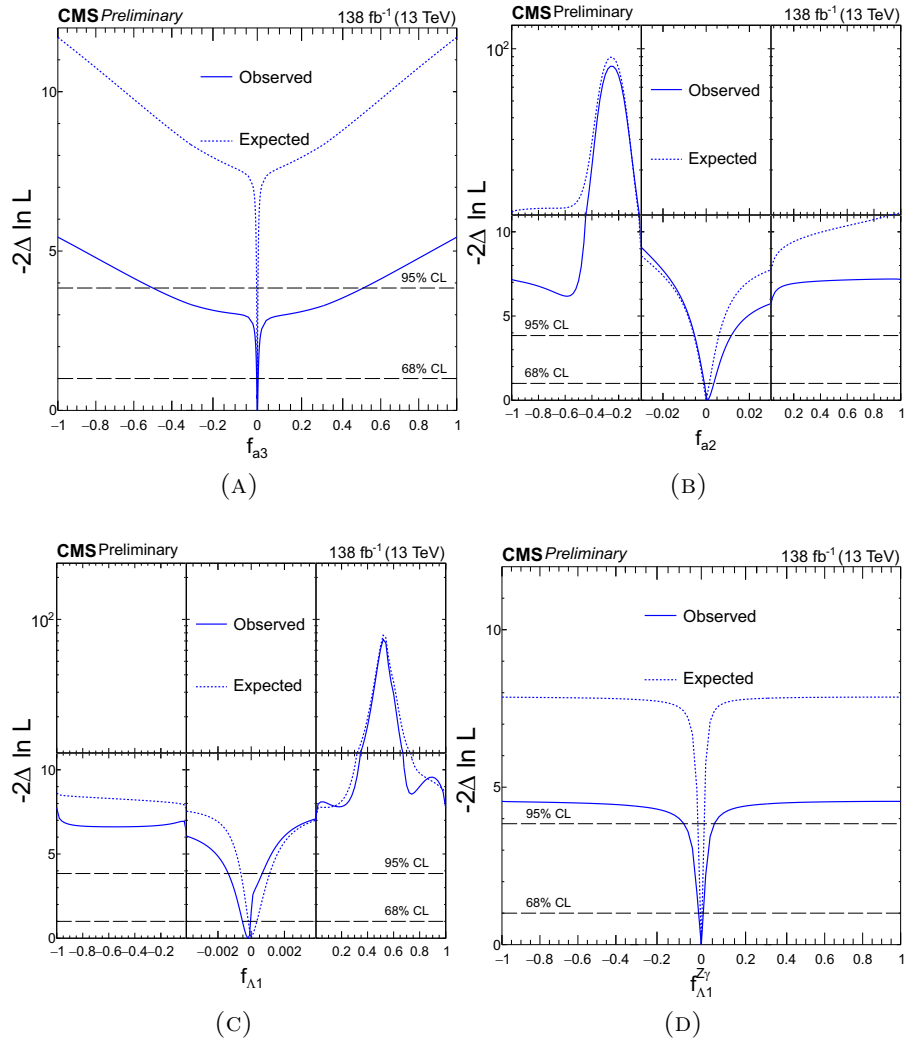


FIGURE 6.4: Expected (dashed) and observed (solid) likelihood scans of  $f_{a3}$  (A),  $f_{a2}$  (B),  $f_{\Lambda 1}$  (C), and  $f_{\Lambda 1}^{Z\gamma}$  (D) for the full combination of HVV channels using Approach 1.

## Zoom scans

Figure 6.5 provides additional information by displaying a close-up of the x-axis to obtain the 68% and 95% CL constraints, allowing for a comparison of the magnitudes of each coupling. The data gathered in Table 6.3 is also extracted from these results.

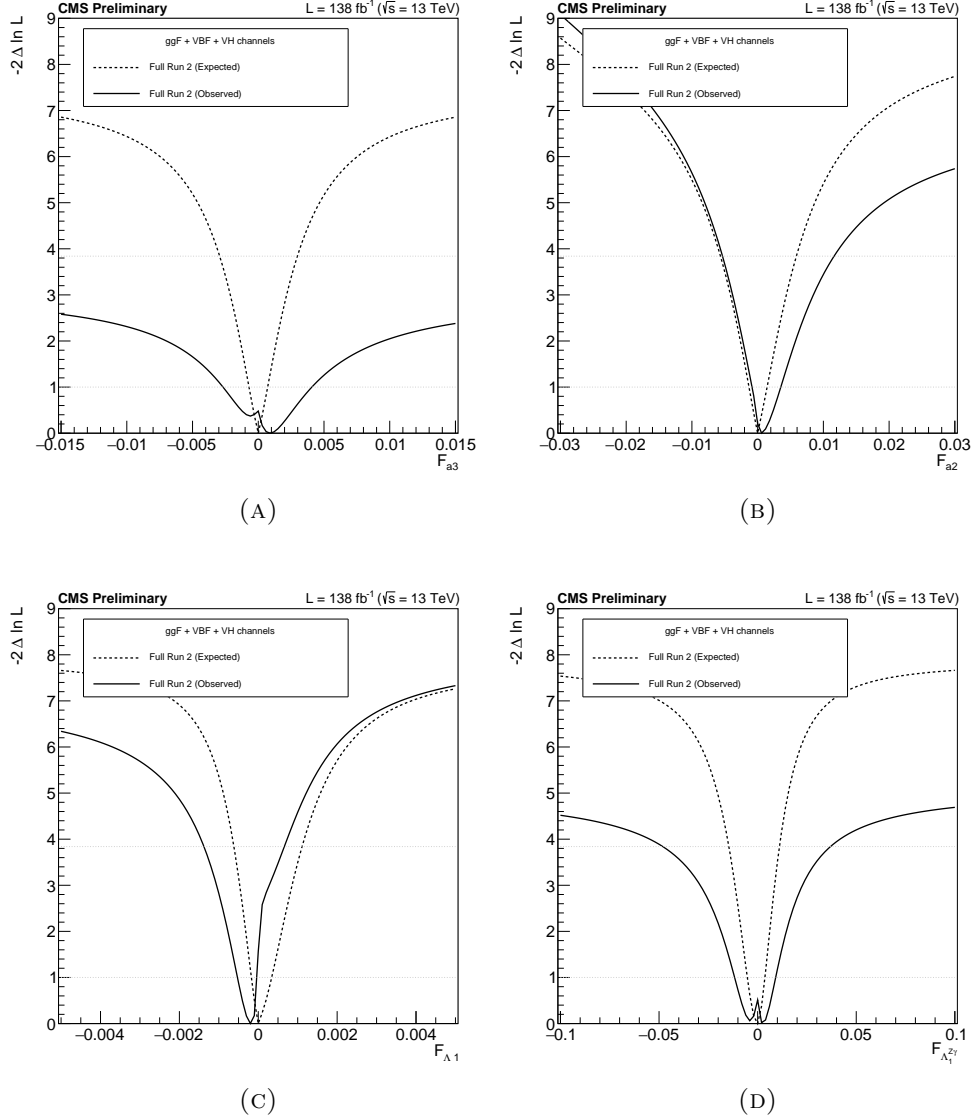


FIGURE 6.5: Observed (solid line) and expected (dashed line) likelihood scans, focusing the x-axis around the expected minima at 0, of  $f_{a3}$  (A),  $f_{a2}$  (B),  $f_{A1}$  (C) and  $f_{A1}^{Z\gamma}$  (D) using the full Run 2 dataset.



### 6.1.4 Full Run 2 Combination for Hgg couplings

In Fig. 6.6, the expected and observed  $f_{a3}^{ggH}$  likelihood scans are shown for the combination of all channels for the full Run 2 dataset. The results are obtained using the MINLO samples. The scans are shown with  $\mu_{EW}$  and  $f_{a3}$  parameters allowed to float in the fit which includes the effect of the CP-odd HVV anomalous coupling for the VBF process. It is noteworthy that the maximum  $-2\Delta \ln \mathcal{L}$  values reached is significantly bigger than expected; as such, we may place allowed 68% CL intervals on the anomalous coupling contribution. The best-fit value for  $f_{a3}^{ggH}$  is -0.034 and is constrained in the range  $[-0.721, 0.383]$  (68% CL).

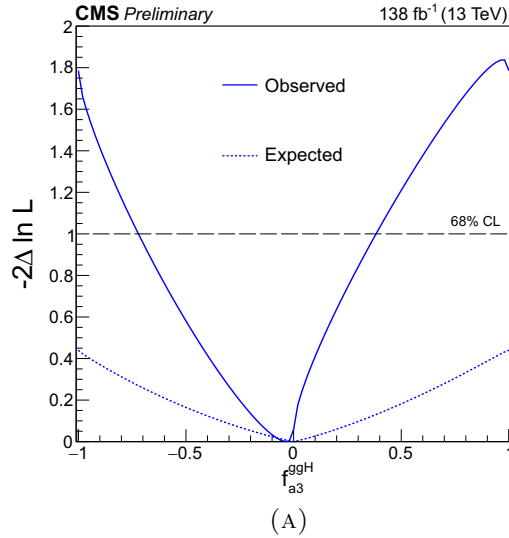


FIGURE 6.6: Expected (dashed) and observed (solid) likelihood scan of  $f_{a3}^{ggH}$ . The signal strength modifiers and the  $CP$ -odd HVV anomalous coupling cross-section fraction are treated as free parameters. The dashed horizontal line shows the 68% CL region.

## 6.2 SMEFT interpretation

Table 6.2 displays an example of the number of events before and after performing the fit described in the previous chapter. The obtained yields are in accordance with the expected values. We now present the likelihood scans for the examined cross-section fractions of the HVV vertex, considering both expected and observed scenarios. The obtained results align with the predictions of the SM.

TABLE 6.2: Expected (pre-fit) and observed (post-fit) number of events per category for background and signals in Approach 1 for the  $a_3$  coupling.

| Approach 2                             | Post-fit (Pre-fit) |          |            |            |         | Pre-fit   |            | Data | Post-Fit     |           |
|--|--------------------|----------|------------|------------|---------|-----------|------------|------|--------------|-----------|
| Category/<br>Process                   | Top                | WW       | Non-prompt | Multiboson | DY      | SM Signal | BSM Signal |      | Total Signal | Total Bkg |
| VBF<br>$D_{CP} < 0$<br>$D_{INT} < 0.4$ | 193 (213)          | 112 (63) | 30 (30)    | 31 (32)    | 23 (33) | 43        | 217        | 386  | 33           | 360       |
| VBF<br>DCP < 0<br>$D_{INT} > 0.4$      | 395 (448)          | 146 (86) | 94 (95)    | 71 (64)    | 69 (72) | 51        | 340        | 809  | 40           | 765       |
| VBF<br>$D_{CP} > 0$<br>$D_{INT} < 0.4$ | 185 (206)          | 112 (62) | 34 (33)    | 33 (35)    | 22 (23) | 46        | 221        | 416  | 44           | 359       |
| VBF<br>$D_{CP} > 0$<br>$D_{INT} > 0.4$ | 397 (455)          | 145 (85) | 65 (81)    | 63 (61)    | 65 (67) | 46        | 326        | 829  | 44           | 751       |
| Resolved VH<br>$D_{CP} < 0$            | 169 (202)          | 72 (40)  | 32 (38)    | 29 (28)    | 19 (19) | 20        | 352        | 353  | 17           | 328       |
| Resolved VH<br>$D_{CP} > 0$            | 175 (202)          | 77 (42)  | 36 (32)    | 32 (31)    | 18 (19) | 24        | 345        | 342  | 22           | 326       |
| Boosted VH<br>$D_{CP} < 0$             | 7 (9)              | 5 (3)    | 2 (1)      | 1 (1)      | 4 (4)   | 2         | 239        | 22   | 2            | 19        |
| Boosted VH<br>$D_{CP} > 0$             | 7 (7)              | 5 (3)    | 1 (1)      | 2 (2)      | 4 (4)   | 2         | 230        | 13   | 3            | 17        |

### 6.2.1 Full Run 2 Combination for HVV couplings

Figure 6.7 presents the expected and observed likelihood scans for Approach 2, adopting the  $SU(2)_L \times U(1)_Y$  coupling relationships. The results are shown for each  $f_{ai}$  separately, with the other two cross-section fractions fixed at zero or allowed to float in the fit. Interference effects are also apparent in this interpretation. However, in some cases, the maximum  $-2\Delta \ln \mathcal{L}$  values are significantly lower than expected. This is largely due to the fitted value of  $\mu_{Ewk}$ , the lowest of which is 0.82 for the  $f_{a3}$  fit in Approach 1, compared to the highest value of 0.97 for the corresponding  $f_{\Lambda 1}$  fit. The uncertainty on  $\mu_{Ewk}$  is approximately 20%, so all fitted values are consistent with the SM and with each other. Overall, all anomalous HVV coupling parameter measurements are consistent with the expectations for the SM Higgs boson. The full Approach 2 fit, where all signal parameters are simultaneously floated, is found to be 91% compatible with the SM. Table 6.3 summarizes the constraints on the anomalous HVV coupling parameters, including the best-fit values and allowed 68% and 95% CL intervals.

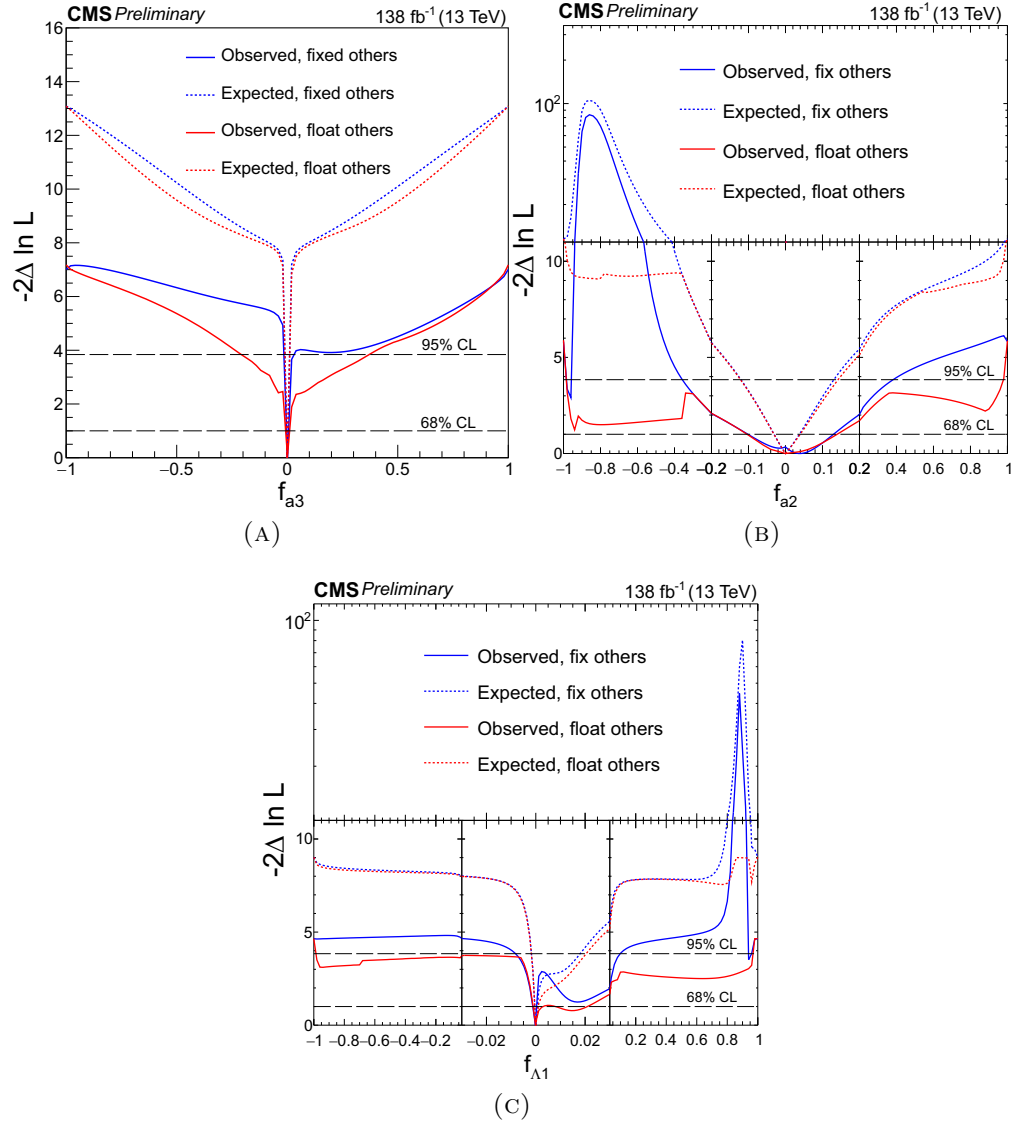


FIGURE 6.7: Expected (dashed lines) and observed (solid lines) likelihood scans of  $f_{a3}$  (A),  $f_{a2}$  (B) and  $f_{\Lambda 1}$  (C) for the full combination of HVV channels using Approach 2. The other two AC cross-section fractions are either fixed to zero (black lines) or left floating in the fit (red lines).

**Zoom Scans** A zoom-in of the x-axis is provided in Fig. 6.8 to extract the 68% and 95% CL constraints, enabling a comparison of the magnitudes of each coupling.

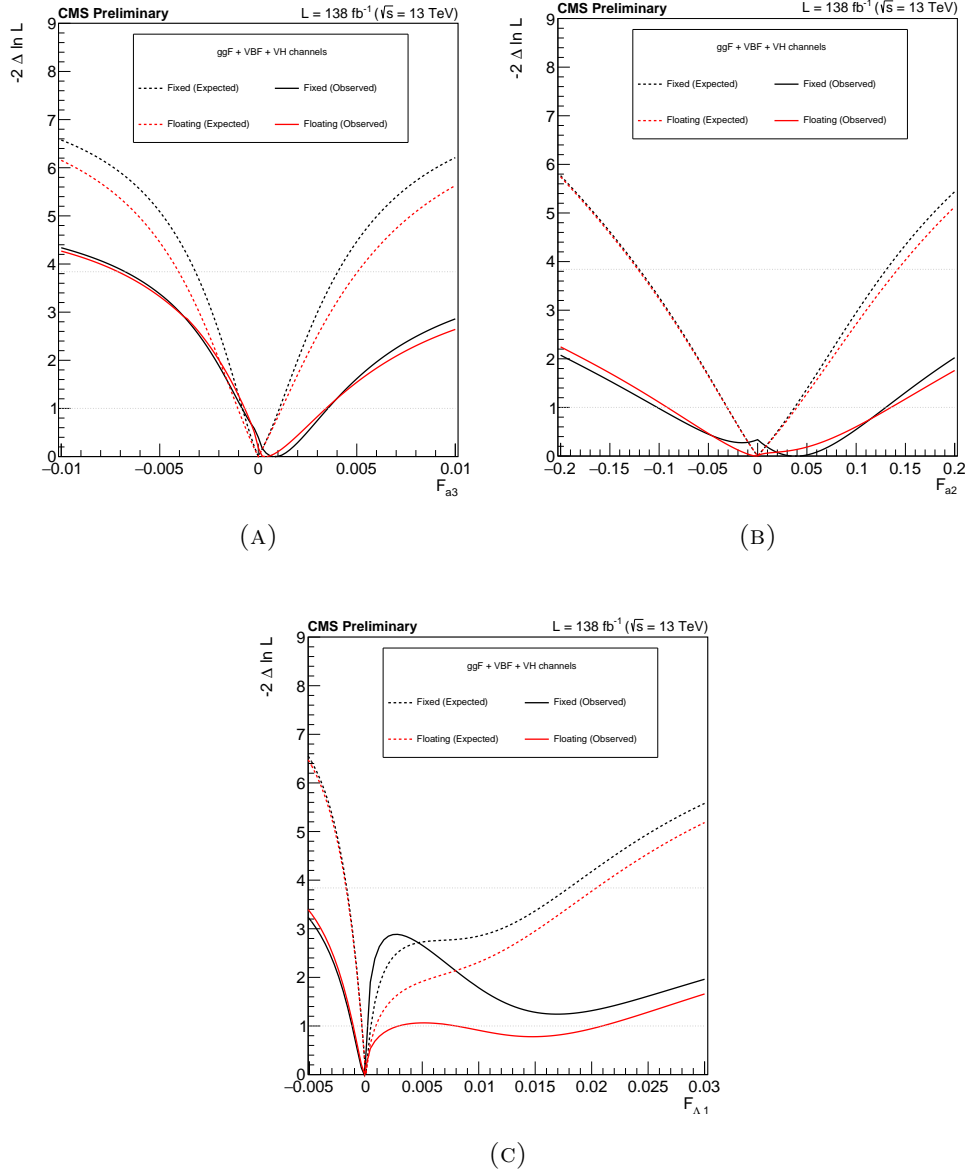


FIGURE 6.8: Full Run 2 expected (dashed lines) and observed (solid lines) likelihood scans of  $f_{a3}$  (A),  $f_{a2}$  (B),  $f_{\Lambda 1}$  (C) considering Approach 2. The other two anomalous coupling fractions are either fixed to zero (black) or left floating in the fit (red).

### 6.2.2 Constraints in terms of EFT couplings

The final results can be expressed in terms of EFT couplings  $c_Z$ ,  $c_{Z\Box}$ ,  $c_{ZZ}$ , and  $\tilde{c}_{ZZ}$  of the Higgs basis, which are related to the couplings  $a_i$  given in Eq. 1.31 - 1.34. The corresponding scans are presented in Fig. 6.9. Expressing the results in terms of EFT couplings is practical because they directly indicate the strength of the couplings and facilitate the translation of results into the Warsaw basis, which is useful for the theoretical community when interpreting these results in the theoretical context.

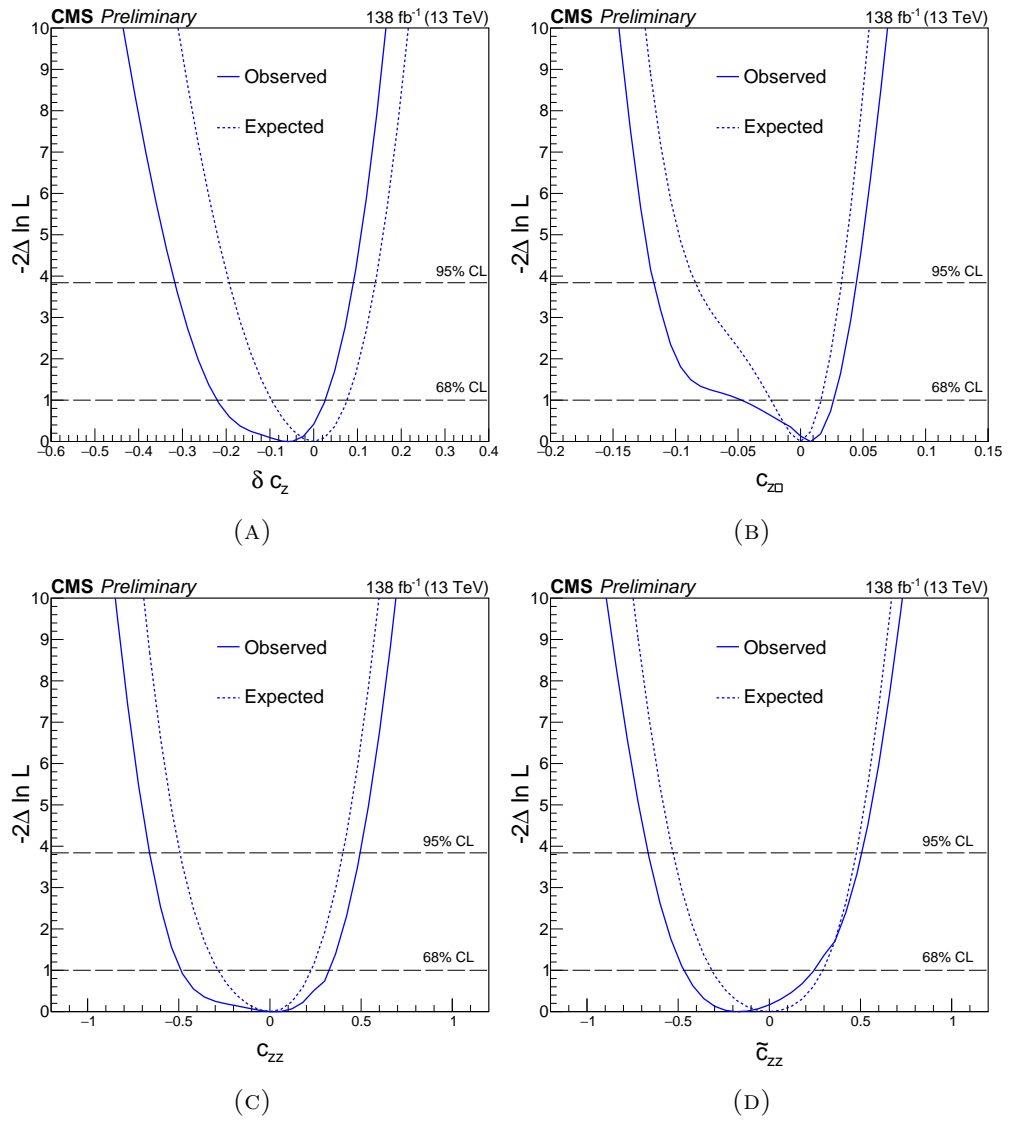


FIGURE 6.9: Full Run 2 expected (dashed line) and observed (solid line) scans of the lagrangian EFT couplings  $c_Z$  (A),  $c_{Z\Box}$  (B),  $c_{ZZ}$  (C), and  $\tilde{c}_{ZZ}$  (D) of the Higgs basis. All couplings are left floating simultaneously in the fit.

### 6.3 Summary

Table 6.3 provides a summary of the expected and observed constraints on the AC and SMEFT HVV coupling parameters for the full Run 2 combination, including the best-fit values and the allowed 68% and 95% CL intervals.

TABLE 6.3: Summary of constraints on the anomalous HVV coupling parameters with the best-fit values and allowed 68% and 95% CL (in square brackets) intervals. The symbol \* means the other parameters are left floating in the fit.

| Analysis         | $f_{ai}$                  | Observed ( $\times 10^{-3}$ )            |  | Expected ( $\times 10^{-3}$ ) |                 |
|------------------|---------------------------|--|--|-------------------------------|-----------------|
|                  |                           | 68% CL                                   | 95% CL                                     | 68% CL                        | 95% CL          |
| HV<br>Approach 1 | $f_{a3}$                  | 0.9 [-2.7, 4.1]                          | [-553, 561]                                | 0.0 [-0.7, 0.7]               | [-2.8, 2.9]     |
|                  | $f_{a2}$                  | 0.5 [-0.8, 3.5]                          | [-5.7, 12]                                 | 0.0 [-1.4, 1.3]               | [-5.2, 6.1]     |
|                  | $f_{\Lambda 1}$           | -0.2 [-0.5, 0.0]                         | [-1.4, 0.7]                                | 0.0 [-0.2, 0.5]               | [-0.6, 1.4]     |
|                  | $f_{\Lambda 1}^{Z\gamma}$ | 3.0 [-11, 9.1]                           | [-55, 42]                                  | 0.0 [-5.0, 3.8]               | [-14, 11]       |
| HV<br>Approach 2 | $f_{a3}$                  | 0.84 [-0.83, 3.5]                        | [-7.6, 58.8]                               | 0.0 [-0.8, 1.1]               | [-3.4, 4.3]     |
|                  | $f_{a3} *$                | 0.34 [-0.69, 3.4]                        | [-8.0, 361.5]                              | 0.0 [-1.0, 1.2]               | [-4.3, 5.3]     |
|                  | $f_{a2}$                  | 38 [-112.2, 129.3]                       | [-376.6, 430.0]<br>$\cup$ [-989.2, -826.3] | 0.0 [-30.9, 37.5]             | [-126.1, 136.8] |
|                  | $f_{a2} *$                | -1.0 [-104.1, 139.9]                     | [-986.4, 981.2]                            | 0.0 [-31.1, 39.8]             | [-127.5, 148.7] |
|                  | $f_{\Lambda 1}$           | -0.15 [-1.21, 0.16]                      | [-19.5, 118.5]<br>$\cup$ [909.9, 964.1]    | 0.0 [-0.4, 0.4]               | [-1.7, 18.9]    |
|                  | $f_{\Lambda 1} *$         | -0.1 [-1.08, 3.78]<br>$\cup$ [7.2, 20.7] | [-994.8, 993.9]                            | 0.0 [-0.4, 0.90]              | [-1.9, 21.4]    |
|                  |                           |  |  |                               |                 |

# Chapter 7

## Conclusions

### **CMS Muon Drift Tubes upgrade for the HL-LHC**

CMS experiment Phase-2 upgrade for the high-luminosity era of the LHC operation (HL-LHC) requires the replacement of the full electronics of the CMS Drift Tube (DT) chambers. The new design allows for a time resolution of around 1 ns for the digitization of chamber signals and will forward signals asynchronously to the service cavern at full resolution. Trigger primitives of each chamber will be built by the new backend system and used as input in the L1 CMS trigger to improve the robustness of the system against aging. The Analytical Method algorithm has been implemented for reconstructing the DT trigger primitives, with measured efficiencies of 96 to 98% and time and spatial resolutions close to the ultimate performance of the DT chambers.

As part of the hardware component of this thesis, a “Slice Test” was installed during LHC Shut Down 2 to develop and test the DT electronics designed for LHC Run 4 (Phase 2) in one sector of the CMS DT detector. In order to ensure optimal time resolution of future Local Triggers, precise calibration of the produced signals is crucial. The initial step of this calibration is inter-channel equalization, to be conducted within each DT chamber using dedicated Test-Pulse runs, which involves simultaneously injecting pulses into all wires and recording the resulting Phase 2 signals. The analysis presented in this thesis demonstrates that the new system is stable over time, with the maximum average differences between different runs amounting to only 0.5 ns. Looking ahead, this procedure will represent as an effective monitoring tool for the performance of the system.

## Anomalous Couplings of the Higgs boson

Over the course of ten years of study, the Higgs boson has become a centerpiece of particle physics research. From Run 1 data sets, it was established that the LHC data was compatible with a Standard Model Higgs of mass 125 GeV, with spin 0 and even CP [1, 2] and confirmed that the Higgs boson plays the expected role in the SM, providing mass to other particles. With the Run 2 dataset of the LHC, the precision era of Higgs physics began, enabling CP studies of the Higgs boson couplings to both bosons and fermions, which have become a major focus of research. Asymmetric matter-antimatter distribution in the universe could potentially be explained by a CP violation process, further emphasizing the importance of searching for Higgs anomalous couplings. Consequently, the study of Higgs anomalous couplings has become a hot topic in particle physics.

The  $H \rightarrow WW$  decay channel is particularly promising to probe the dynamics of the Higgs sector of the SM. This is due to its high branching ratio and relatively low-background final state, making it a prime candidate for precise measurements. The HWW SM analysis conducted in [93] is part of a larger combined CMS measurement of the Higgs boson production cross sections, branching ratios, and couplings with Run 2 data, contributing to the most comprehensive understanding of Higgs boson properties published by CMS [28].

The CMS collaboration has conducted AC analyses for various channels, including  $HZZ$ ,  $H\tau\tau$ , and  $H\gamma\gamma$ . However, a combination of these channels has not been performed since Run 1 [50]. There is a need for a concerted effort to achieve this goal, as it would expand our understanding of the properties of the Higgs boson and its role in the SM.

The conducted AC HWW analysis is as sensitive as the  $HZZ$  mostly. Furthermore, an AC analysis on this channel has not been performed since Run 1, and several improvements have been made, such as the inclusion of the VBF and VH channels, AC effects at the VBF/VH and  $ggH$  production vertexes, and naturally, the higher integrated luminosity. AC detector-level analyses complement interpretation analyses, constraining EFT couplings through an existing STXS measurement. A joint effort of analysis groups would aid in deepening our understanding of Higgs boson couplings and searching for new physics. For the first time in a AC analysis, we have shown an explicit study of the specific features of the expected shapes of the cross-section fractions scans. This is achieved by studying how the expected



signal shapes for important variables are altered by anomalous effects. The HZZ analysis reported large interference peaks were due to a big change in signal shape, for example. However, the AC HWW analysis is the first to present an explicit study in this fashion.

The inclusion of a boosted category in the AC analysis for the first time has shown an increased sensitivity to anomalous effects, resulting in improved constraints that approach the SM values ( $f_{ai} = 0$ ). Nevertheless, due to statistical limitations in the Run 2 data set, this category has limited effectiveness. Therefore, conducting the same analysis using future LHC datasets could significantly enhance the constraints in the HWW channel.

The analysis has followed two interpretations: Approach 1 (AC) considers one SM coupling and one BSM coupling at a time, while Approach 2 (SMEFT) constrains all anomalous couplings simultaneously. To provide a general overview of the strength of our findings compared to similar analyses, we can state that we have effectively constrained the available phase space for anomalous couplings to the per mil level. The best constraints are set for the  $f_{\Lambda 1}$  and  $f_{a3}$  for the AC and the SMEFT interpretations, respectively. Table 7.1 compares our results with the constraints set by the HZZ and  $H\tau\tau$  analyses.

TABLE 7.1: Comparison of the best HVV constraints set by the HWW analysis to the HZZ and  $H\tau\tau$  analyses.

| Analysis                     | $f_{ai}$        | Observed ( $\times 10^{-3}$ ) |               | Expected ( $\times 10^{-3}$ ) |               |
|------------------------------|-----------------|-------------------------------|---------------|-------------------------------|---------------|
|                              |                 | 68% CL                        | 95% CL        | 68% CL                        | 95% CL        |
| HVV Approach 1               | $f_{\Lambda 1}$ |                               |               |                               |               |
| HWW                          |                 | -0.2 [-0.5, 0.0]              | [-1.4, 0.7]   | 0.0 [-0.2, 0.5]               | [-0.6, 1.4]   |
| HZZ                          |                 | 0.04 [0.02, 0.22]             | [0.14, 0.60]  | 0.0 [0.16, 0.26]              | [0.69, 1.10]  |
| $H\tau\tau$                  |                 | -0.12 [0.08, -0.10]           | [-0.34, 0.01] | 0.0 [0.19, -0.05]             | [-0.15, 0.55] |
| Combination HZZ+ $H\tau\tau$ |                 | -0.04 [0.04, -0.08]           | [-0.22, 0.16] | 0.0 [0.11, -0.04]             | [-0.11, 0.38] |
| HVV Approach 2               | $f_{a3}$        |                               |               |                               |               |
| HWW                          |                 | 0.34 [-0.69, 3.4]             | [-8.0, 361.5] | 0.0 [-1.0, 1.2]               | [-4.3, 5.3]   |
| HZZ                          |                 | 0.05 [0.10, 0.61]             | [0.72, 2.18]  | 0.0 [1.2, 1.2]                | [5.7, 5.7]    |
| $H\tau\tau$                  |                 | 0.4 [-0.33, 0.53]             | [-0.01, 1.90] | 0.0 [-0.08, 0.08]             | [-0.33, 0.33] |
| Combination HZZ+ $H\tau\tau$ |                 | 0.28 [0.39, -0.23]            | [-0.01, 1.28] | 0.0 [-0.08, 0.08]             | [-0.30, 0.30] |

## Personal contributions

In the CMS publication, CMS Performance Note CMS DP-2022-001 [151], I shared the findings of my research contributions related to the calibration of new electronics and the stability of time measurements. Specifically, I conducted a study on stability by utilizing Test Pulse signals generated by the Phase-1 system. This involved injecting simultaneous pulses into all wires and analyzing the resulting TDC times and Phase-2 electronics readouts. The study demonstrated that the Phase-2 system exhibits remarkable stability and performs even better than the required specification of 1 ns. As part of the Slice Test project for CIEMAT, I participated in the development of a valuable monitoring tool for the HL-LHC DT chambers. Lastly, I had the opportunity to present these significant findings at the LHCC Conference in November 2021.

Regarding the Higgs boson AC studies, I served as the contact person for the analysis, which implies coordinating activities for the analysis team, keeping the results up to date and visible to the HEP community. Within the technical part of the analysis, my role involved analyzing the complete set of Run 2 pp collision data collected by the CMS detector at 13 TeV, with a focus on ggH, VBF, and VH channels. This analysis was performed for the first time utilizing this set of data. Specifically, I conducted a full analysis of the VBF and VH channels and combined all categories to provide an interpretation in terms of a SM EFT. We also included for the first time in a Higgs AC analysis a boosted VH channel to exploit the features of AK8 jets and their substructure. To extract meaningful results within the context of the SM EFT interpretation, I developed multidimensional kinematic discriminants that allowed for the simultaneous constraint of multiple anomalous couplings. Additionally, I continuously led the efforts of the analysis team to move the AC analysis through the CMS internal review and regularly present our findings to the CMS community at conferences and meetings. As part of this engagement, I participated in the Higgs 2022 conference held in Pisa, where I gave a parallel talk about EFT-related analysis of the CMS Collaboration.

At the beginning of my Ph.D., I made significant contributions to the Lepton Flavour Violation analysis, focusing on two channels that are forbidden by the Standard Model:  $H \rightarrow \mu\tau$  and  $H \rightarrow e\tau$ . Specifically, this analysis explored scenarios where the  $\tau$  lepton decays either leptonically or hadronically, aiming to search for off-diagonal couplings in the Yukawa matrix that deviate from zero. To conduct this study, we utilized the miniAOD samples corresponding to the CMS

Run 2 dataset, with particular emphasis on the ggH and VBF production modes. The results were obtained through a fitting procedure using a Boosted Decision Tree (BDT). Additionally, I performed a cross-check analysis using the nanoAOD version of the samples, focusing on the  $\mu\tau_e$  channel within the 2018 dataset. I presented these findings at the ICHEP 2020 conference [152].



# Appendix A

## Non-prompt background estimation method

Prompt leptons are leptons that directly emerge from the decay of particles resulting from a collision or from the collision itself. In the case of a fully leptonic decay of  $H \rightarrow WW$ , two prompt leptons are expected to appear in the final state, arising from the decay of the two W bosons. However, there are alternative sources of leptons in the final state, including genuine leptons originating from the decay of heavy-flavored hadrons or misidentified jets that are falsely classified as isolated leptons. These instances contribute to the fake lepton background. In the  $H \rightarrow WW$  analysis, the main origin of background due to misidentified leptons in the 0-jet category stems from events where W boson plus jets are mistakenly identified. These events consist of a prompt lepton and neutrinos resulting from the decay of the W boson. If one of the jets is incorrectly labeled as a prompt lepton, the final state exhibits two leptons and missing energy, resembling the desired state of interest. This misidentification is more prone to occur with lower transverse momentum leptons, but it remains relevant across the entire range of lepton phase space. In contrast, the 1-jet and 2-jet categories primarily encounter the lepton misidentification background from semileptonic decays of top quark pairs. Other background processes like QCD multijet events and fully hadronic top quark pair decays could also generate non-prompt leptons. However, since these processes involve two non-prompt leptons instead of just one, they are less likely to contribute significantly to the background in the signal region. The contribution of non-prompt leptons is not calculated using MC simulations due to the inherent limitations in accurately reproducing instrumental behavior. Instead, a

data-driven approach is employed to estimate the impact of this background. This appendix outlines the methodology and the tests to validate its effectiveness.

## The method

The objective is to count the number of events containing either a prompt lepton ( $N_p$ ) or a non-prompt lepton ( $N_f$ ) within a sample of one-lepton events. These events are selected based on specific identification criteria similar to those employed in the  $H \rightarrow WW$  analysis. To determine these counts, we utilize information regarding the number of events with one loosely identified lepton ( $N_l$ ). These events can either contain one tightly identified lepton ( $N_{t1}$ ) or no tightly identified leptons ( $N_{t0}$ ). The following relationships apply:

$$\begin{cases} N_l = N_p + N_f = N_{t0} + N_{t1} \\ N_{t0} = (1 - p)N_p + (1 - f)N_f \\ N_{t1} = pN_p + fN_f \end{cases} \quad (\text{A.1})$$

where  $f$  and  $p$  are defined as follows:

- The fake rate ( $f$ ) represents the likelihood of a nonprompt lepton meeting the loose criteria and further satisfying the corresponding stringent selection as well.
- The prompt rate ( $p$ ) is the probability of a real prompt lepton that passes the loose criteria to also pass the corresponding tight selection. It is therefore very close to one and much higher than the fake rate.

By inverting these equations, it is possible to derive the contribution from prompt and nonprompt leptons as a function of the probabilities of lepton identification as tight and the observed count of tight leptons.

$$\begin{cases} N_p = \frac{1}{p-f} \cdot [(1-f)N_{t1} - fN_{t0}] \\ N_f = \frac{1}{p-f} \cdot [pN_{t0} - (1-p)N_{t1}] \end{cases} \quad (\text{A.2})$$

Using the provided expression, the nonprompt contribution ( $N_f$ ) can be obtained from a data sample. This involves multiplying the number of events by a weight of  $p$  if there is a loose but not tight lepton or  $1-p$  if there is a tight lepton. In

section 2.3.1, the definitions for electrons and muons were outlined, specifying both “tight” and “loose” criteria for each particle. Now, we utilize those definitions to compute the fake and prompt rates.

Eq. A.1 can be extended to the two-lepton final state as:

$$\begin{cases} N_{\ell\ell} = N_{pp} + N_{fp} + N_{ff} = N_{t2} + N_{t1} + N_0 \\ N_{t0} = (1-p)^2 N_{pp} + (1-p)(1-f)N_{fp} + (1-f)^2 N_{ff} \\ N_{t1} = 2p(1-p)N_{pp} + (f(1-p) + p(1-f))N_{fp} + 2f(1-f)N_{ff} \\ N_{t2} = p^2 N_{pp} + pfN_{fp} + f^2 N_{ff} \end{cases} \quad (\text{A.3})$$

In the case of the two-lepton final state, where  $N_{\ell\ell}$  represents the number of events with two leptons satisfying the loose criteria,  $N_{pp}$ ,  $N_{fp}$ , and  $N_{ff}$  correspond to the number of events with zero, one, or two non-prompt leptons, and  $N_{t0}$ ,  $N_{t1}$ , and  $N_{t2}$  indicate the number of events with zero, one, or two tight leptons. These equations can once again be inverted to deduce the contribution of fake leptons, establishing a relationship between the number of prompt and fake leptons and the number of tight and loose leptons.

$$\begin{pmatrix} N_{pp} \\ N_{fp} \\ N_{ff} \end{pmatrix} = \frac{1}{(p-f)^2} \begin{pmatrix} f^2 & -f(1-f) & (1-f)^2 \\ -2fp & p(1-f) + f(1-p) & -2(1-p)(1-f) \\ p^2 & -p(1-p) & (1-p)^2 \end{pmatrix} \begin{pmatrix} N_{t0} \\ N_{t1} \\ N_{t2} \end{pmatrix} \quad (\text{A.4})$$

By utilizing the lepton fake and prompt rates, it becomes possible to establish a connection between events featuring zero, one, or two tight leptons and events with zero, one, or two non-prompt leptons. This information is leveraged to estimate the non-prompt background by analyzing the data events that involve two loose leptons fulfilling the remaining selection criteria while applying the appropriate weights computed using Eq. A.4.

The computation of the fake rate involves utilizing a control region that contains a significant number of non-prompt leptons. It is important to note that the method assumes the fake rate in the signal region is equivalent to that observed in the control regions where it is determined, although this may not always hold true. One of the primary factors that can lead to variations is the detector response, which can differ for leptons with distinct kinematic properties. To address this, the rates are computed as a function of lepton  $\eta$  and  $p_T$ , thereby taking into account the potential disparities between processes.





# Appendix B

## Control region plots

2017

VBF channel

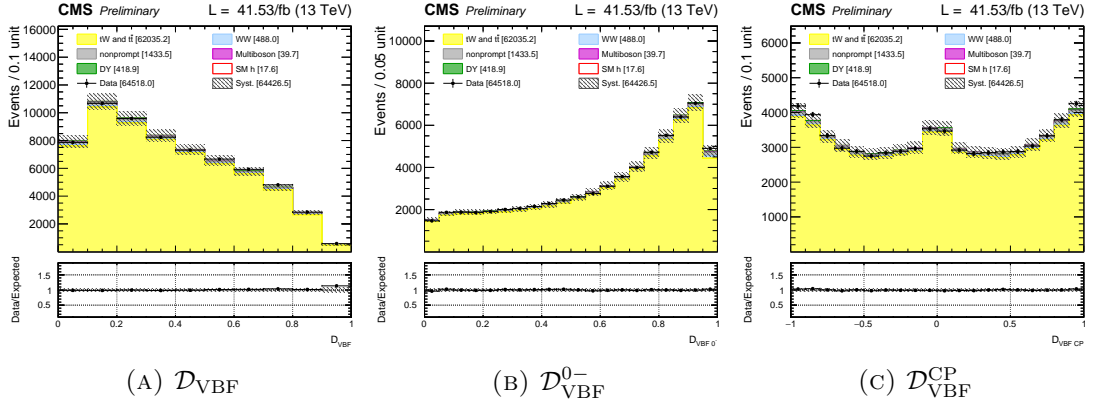


FIGURE B.1: VBF related distributions in top 2-jet CR for the 2017 dataset.

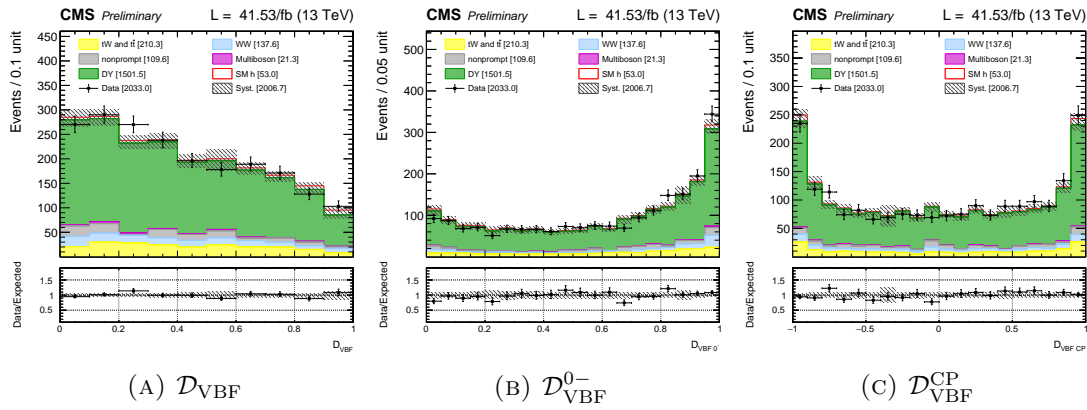
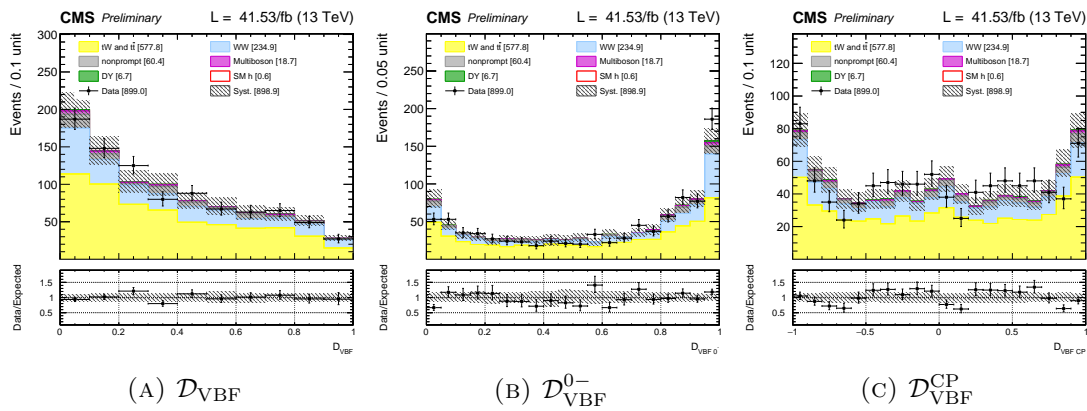

 FIGURE B.2: VBF related distributions in  $DY\tau\tau$  2-jet CR for the 2017 dataset.


FIGURE B.3: VBF related distributions in WW 2-jet CR for the 2017 dataset.

## VH channels

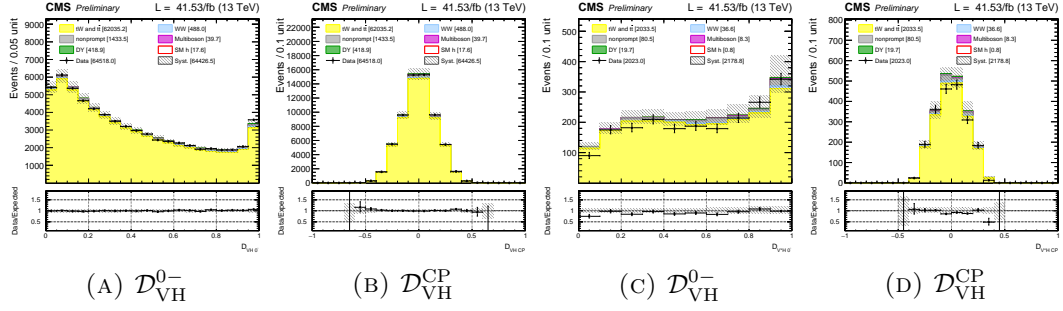


FIGURE B.4: Resolved VH (A, B) and Boosted VH (C, D) related distributions in top 2-jet CR for the 2017 dataset.

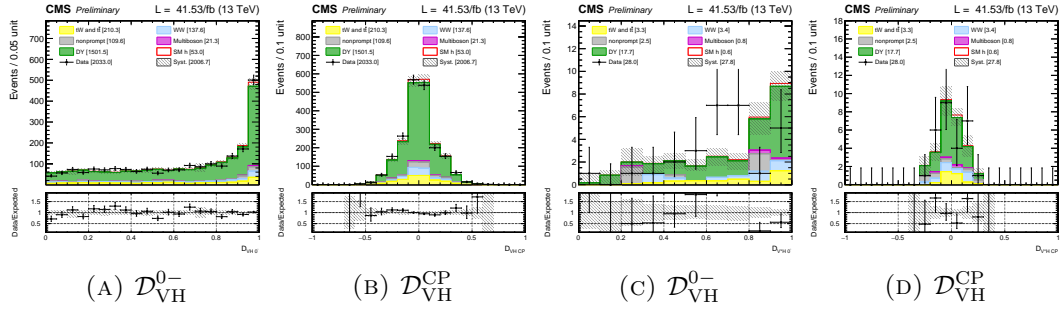


FIGURE B.5: Resolved VH (A, B) and Boosted VH (C, D) related distributions in DY  $\tau\tau$  2-jet CR for the 2017 dataset.

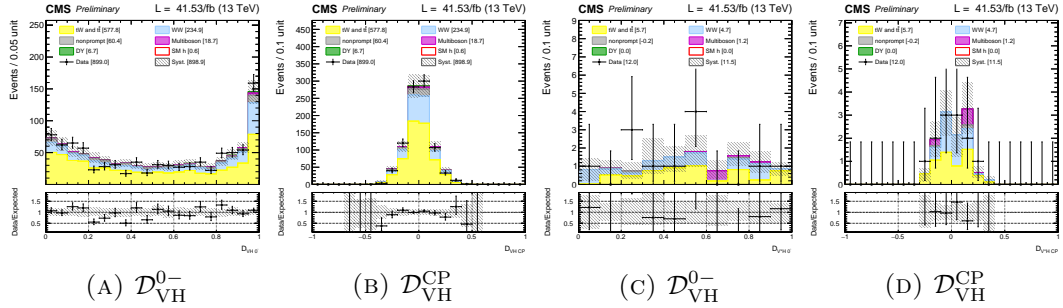


FIGURE B.6: Resolved VH (A, B) and Boosted VH (C, D) related distributions in WW 2-jet CR for the 2017 dataset.

## 2-jet ggH channel

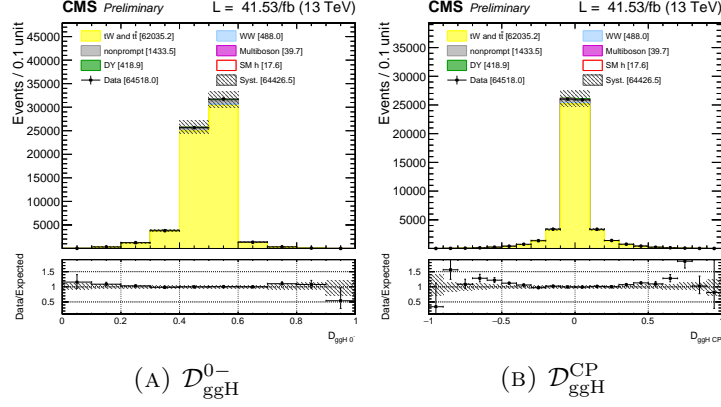


FIGURE B.7: 2-jet ggH related distributions in the top 2-jet CR for the 2017 dataset

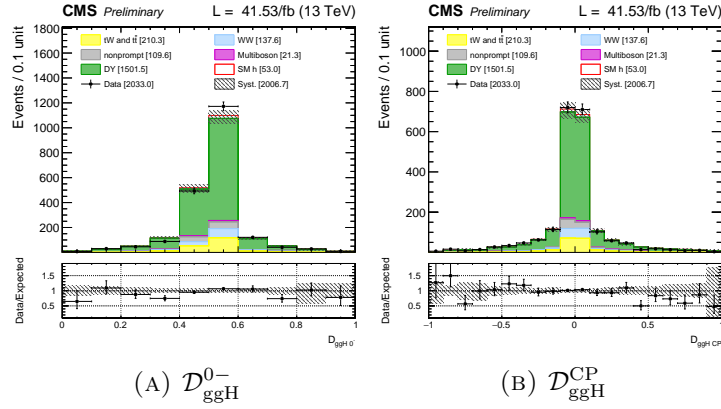
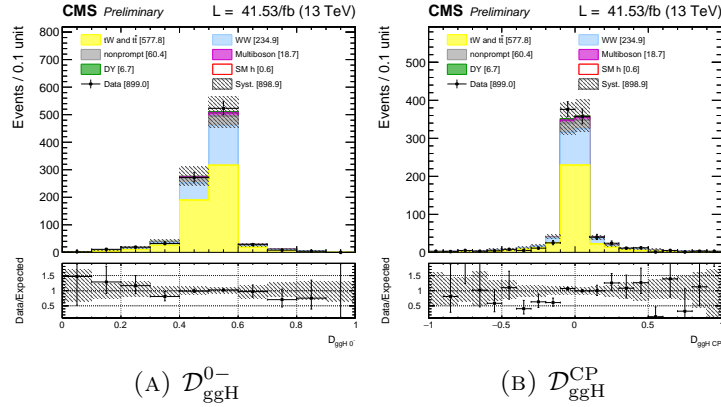
FIGURE B.8: 2-jet ggH related distributions in the DY  $\tau\tau$  CR for the 2017 dataset

FIGURE B.9: 2-jet ggH related distributions in the WW CR for the 2017 dataset

## 0-jet and 1-jet ggH channels

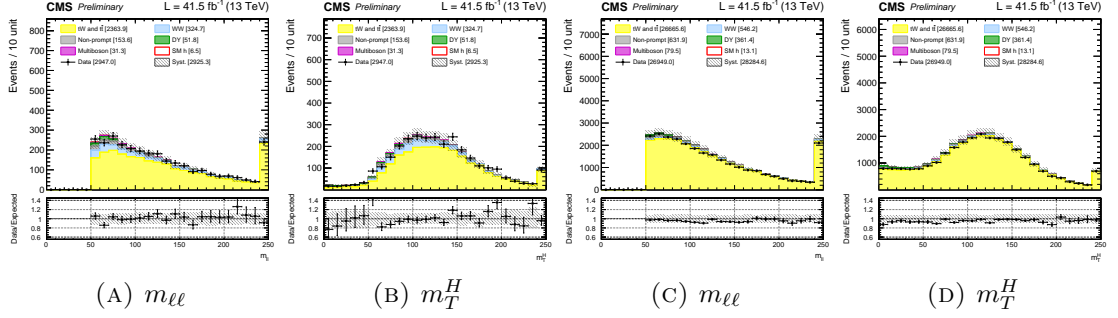
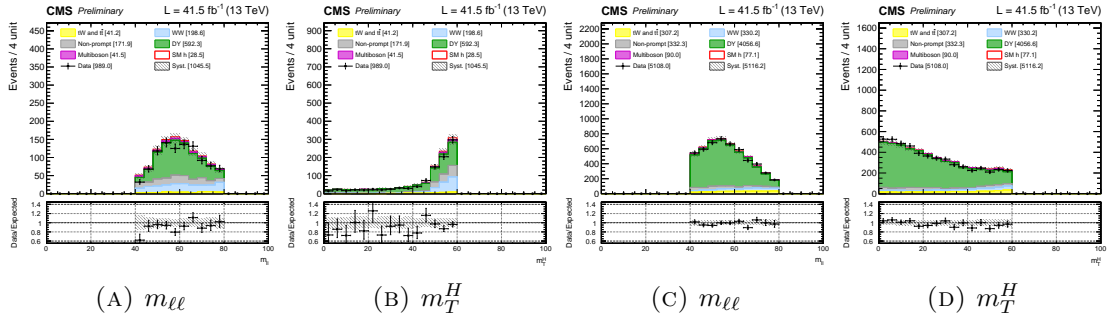


FIGURE B.10: Top 0-jet (A, B) and 1-jet (C, D) CR distributions for 2017 datasets.

FIGURE B.11: DY  $\tau\tau$  0-jet (A, B) and 1-jet (C, D) CR distributions for 2017 datasets.

## 2018

## VBF channel

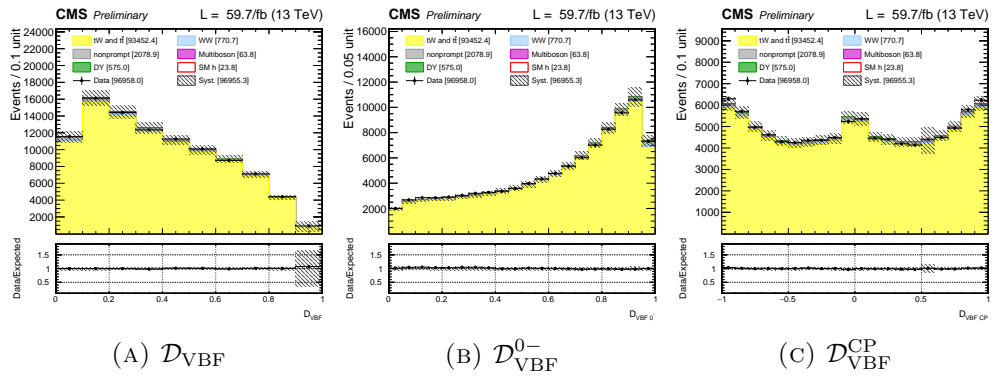
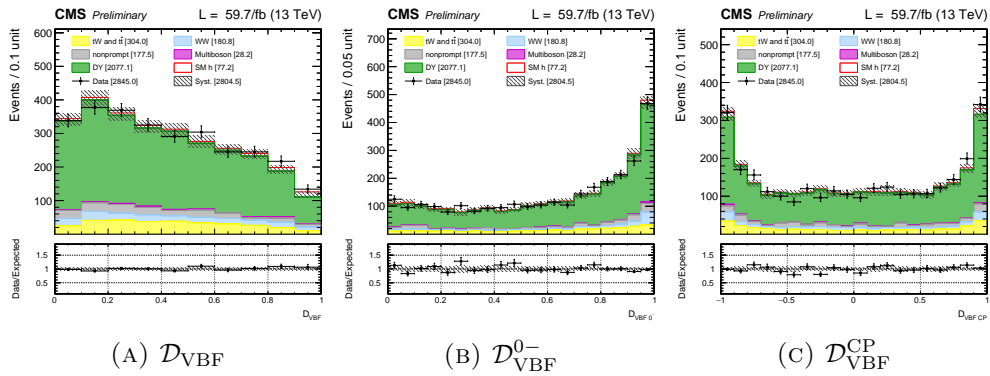
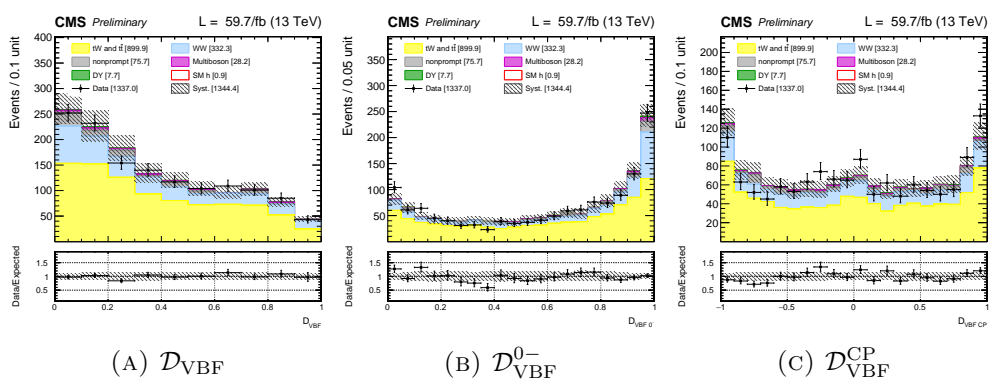


FIGURE B.12: VBF related distributions in top 2-jet CR for the 2018 dataset.


 FIGURE B.13: VBF related distributions in  $DY \tau\tau$  2-jet CR for the 2018 dataset.

 FIGURE B.14: VBF related distributions in  $WW$  2-jet CR for the 2018 dataset.

## VH channels

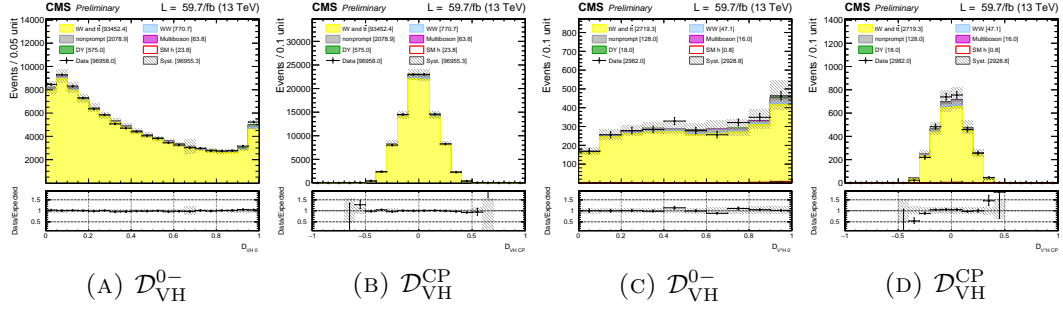


FIGURE B.15: Resolved VH (A, B) and Boosted VH (C, D) related distributions in top 2-jet CR for the 2018 dataset.

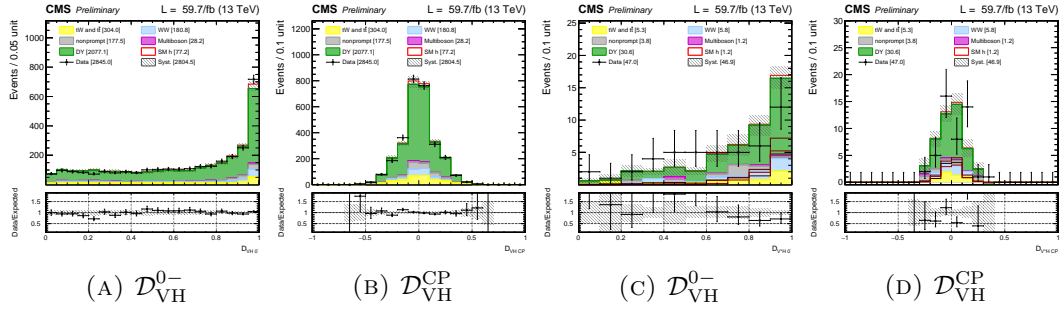


FIGURE B.16: Resolved VH (A, B) and Boosted VH (C, D) related distributions in DY  $\tau\tau$  2-jet CR for the 2018 dataset.

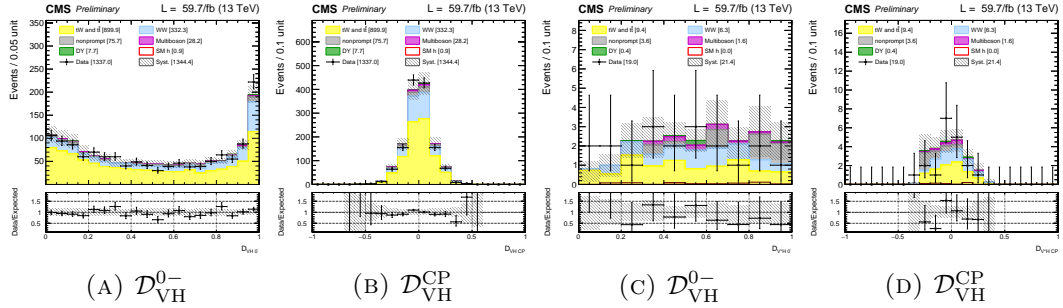


FIGURE B.17: Resolved VH (A, B) and Boosted VH (C, D) related distributions in WW 2-jet CR for the 2018 dataset.

## 2-jet ggH channel

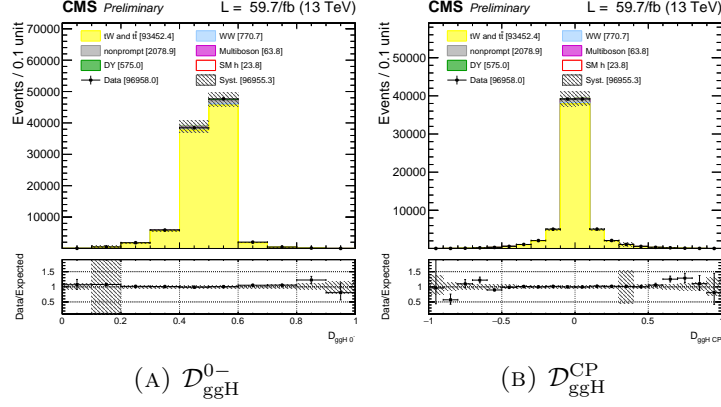


FIGURE B.18: 2-jet ggH related distributions in the top 2-jet CR for the 2018 dataset

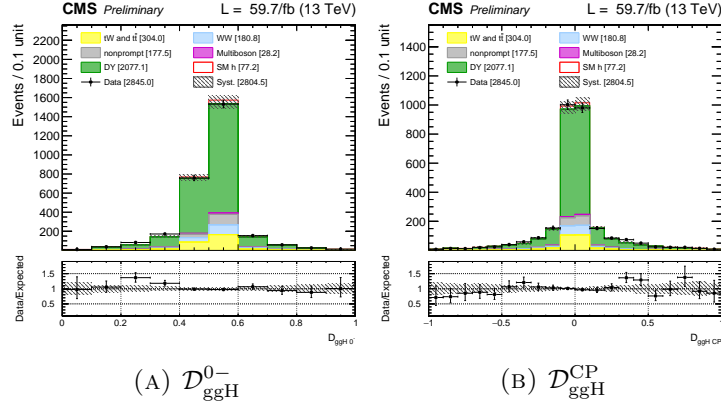
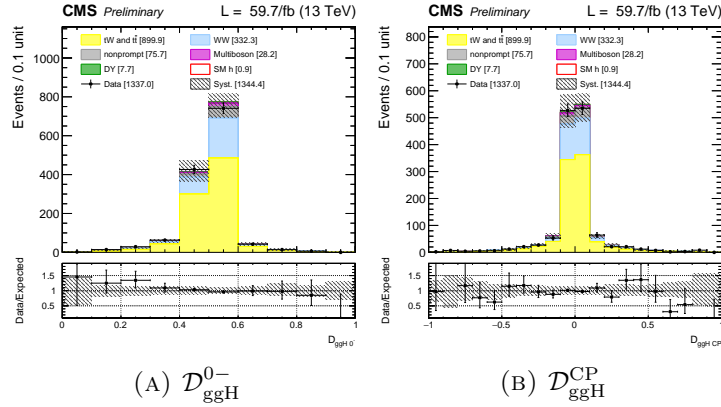
FIGURE B.19: 2-jet ggH related distributions in the  $\text{DY } \tau\tau$  CR for the 2018 dataset

FIGURE B.20: 2-jet ggH related distributions in the WW CR for the 2018 dataset



## 0-jet and 1-jet ggH channels

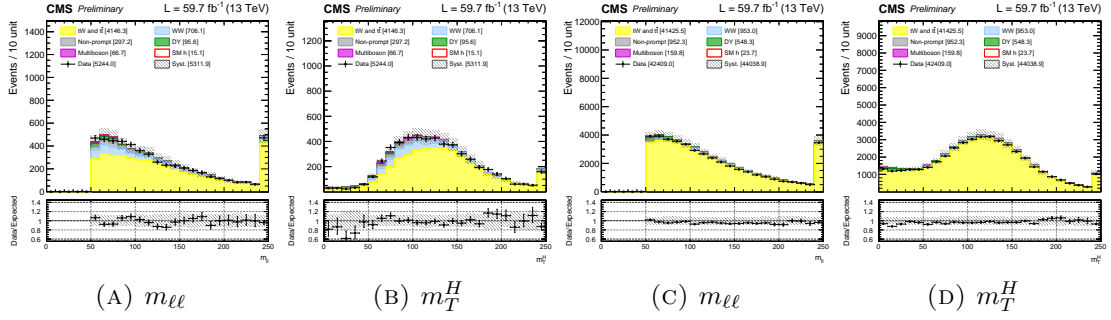
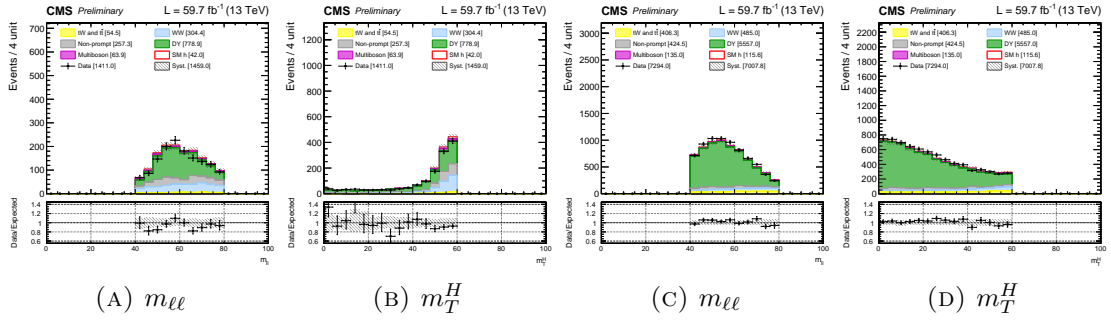


FIGURE B.21: Top 0-jet (A, B) and 1-jet (C, D) CR distributions for 2018 datasets.

FIGURE B.22: DY  $\tau\tau$  0-jet (A, B) and 1-jet (C, D) CR distributions for 2018 datasets.



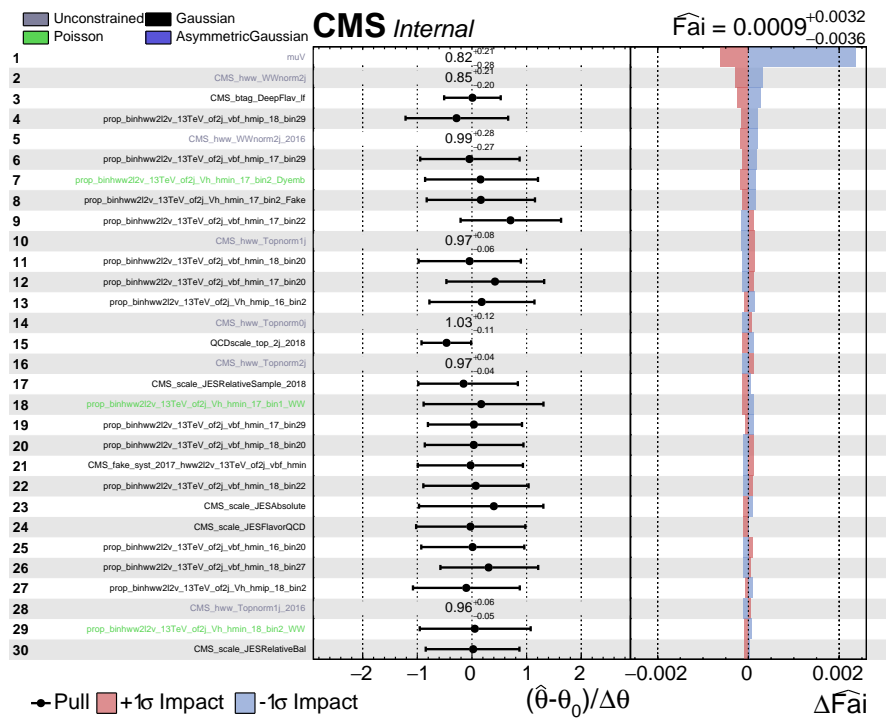
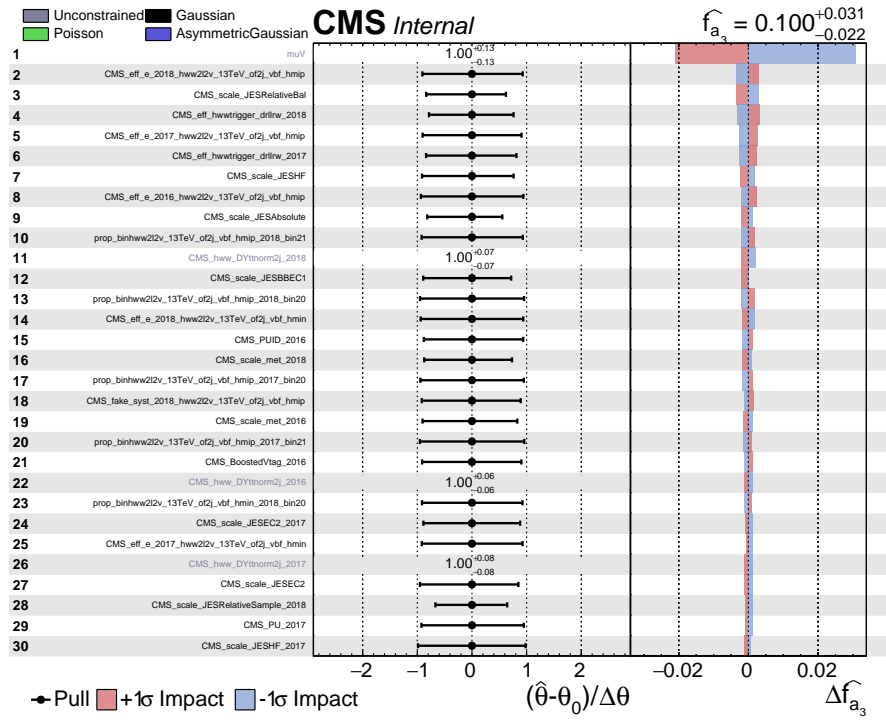
# Appendix C

## Impacts

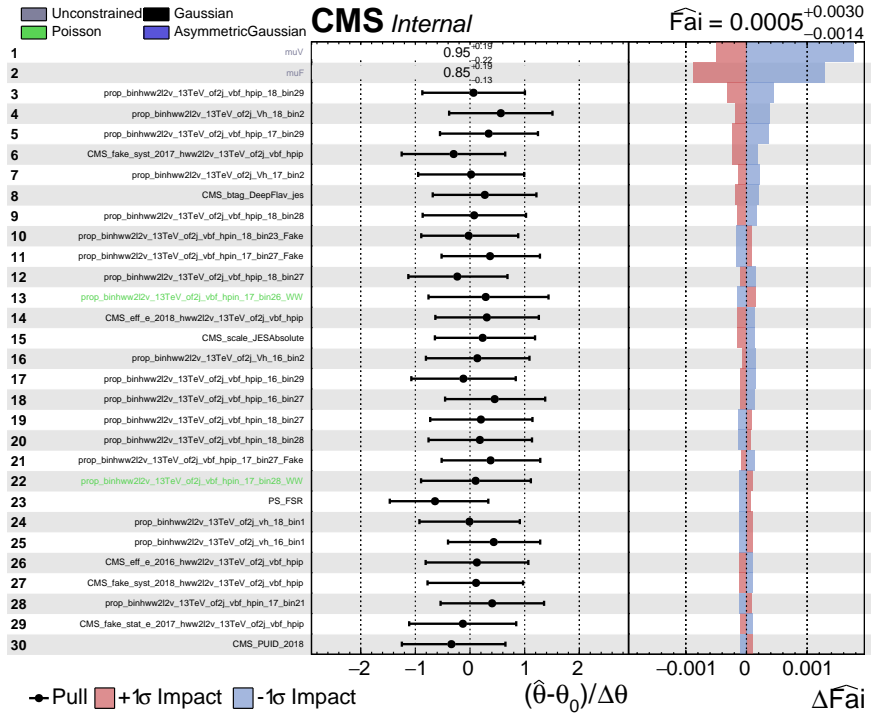
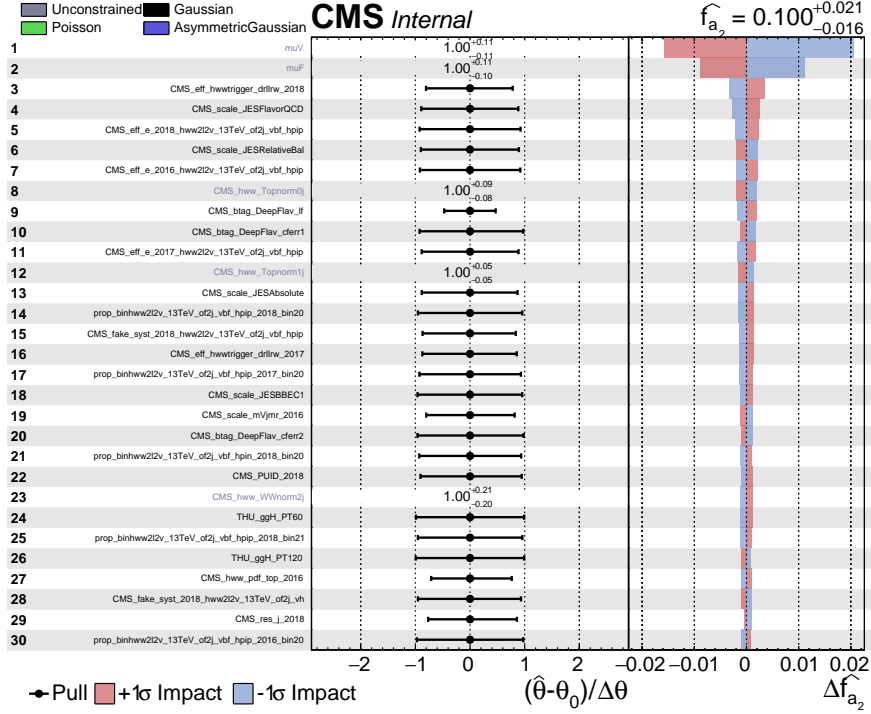
Both expected (fit on MC toys) impacts before unblinding and observed (fit on data) impacts are presented. Expected impacts are shown for small but non-zero values of the  $f_{ai}$  while in the case of observed impacts, correlations correspond to the  $f_{ai}$  best-fit value. For both cases, the signal strength parameters,  $\mu_V$  and  $\mu_F$ , have a significant impact on the result. The precision on these parameters, which are always profiled, is limited by the number of signal Higgs events. The following pages present:

1. HVV  $a_3$  coupling - expected and observed impacts.
2. HVV  $a_2$  coupling - expected and observed impacts.
3. HVV  $\Lambda_1$  coupling - expected and observed impacts.
4. HVV  $\Lambda_1^{Z\gamma}$  coupling - expected and observed impacts.
5. Hgg  $a_3$  coupling - expected and observed impacts.
6. SMEFT HVV  $a_3$  coupling - expected and observed impacts.
7. SMEFT HVV  $a_2$  coupling - expected and observed impacts.
8. SMEFT HVV  $\Lambda_1$  coupling - expected and observed impacts.

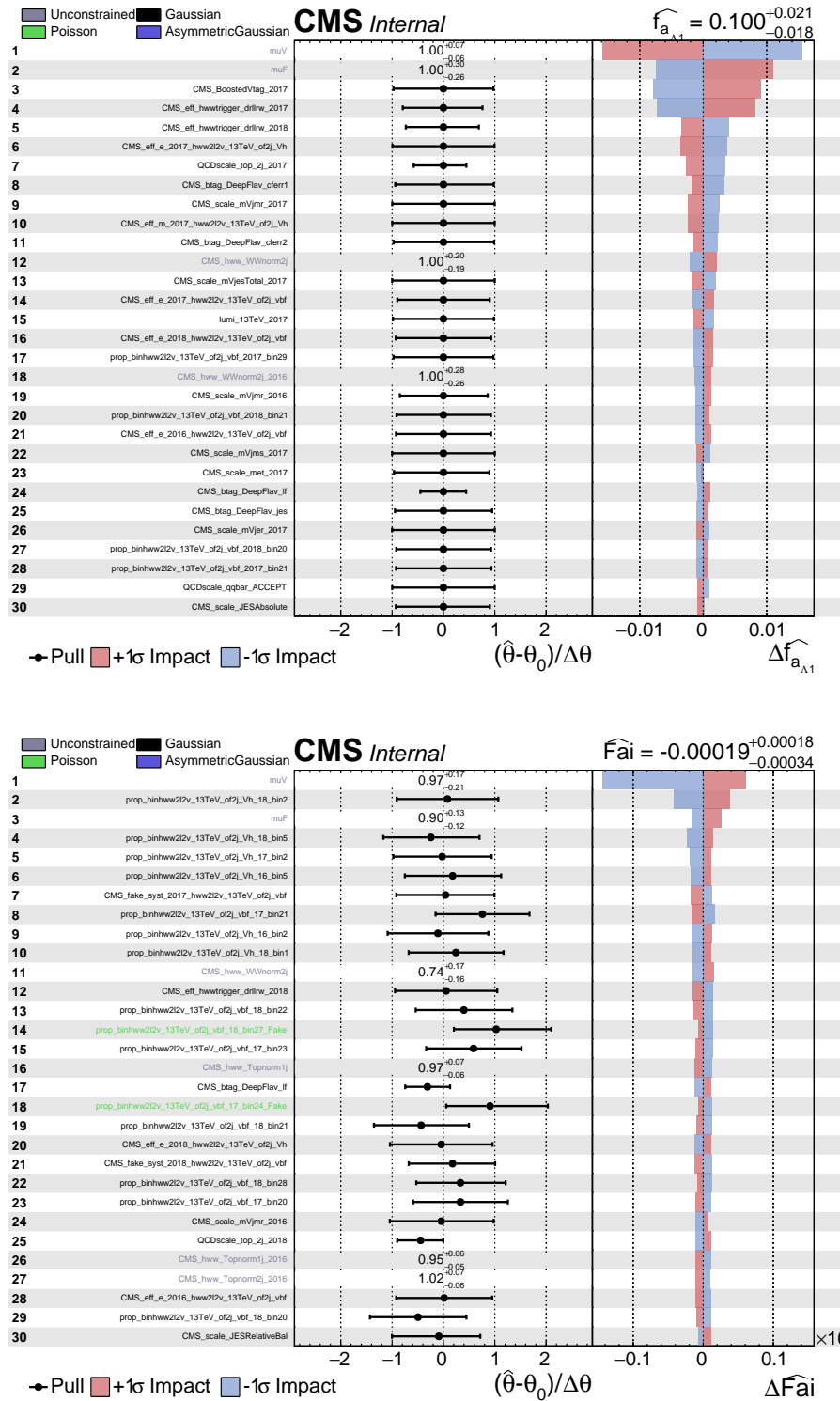
# Expected (top) and observed (bottom) impacts for the HVV $a_3$ coupling



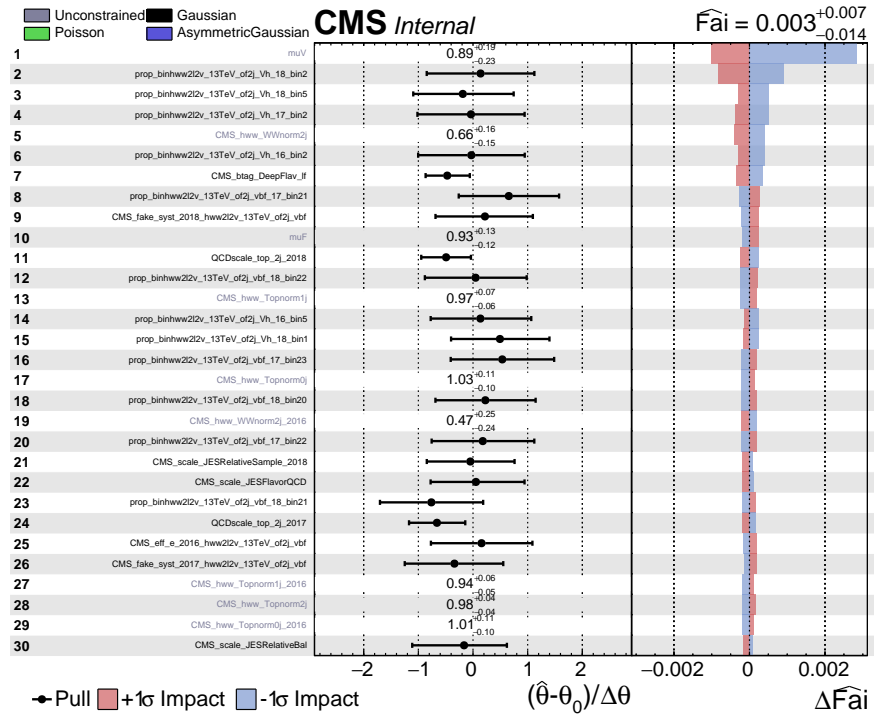
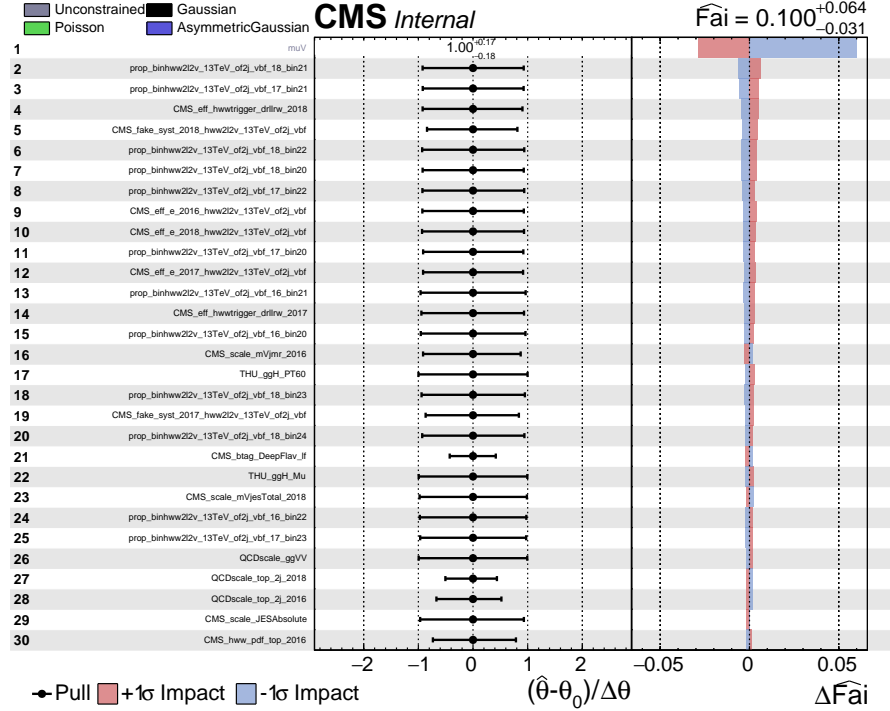
## Expected (top) and observed (bottom) impacts for the HVV $a_2$ coupling

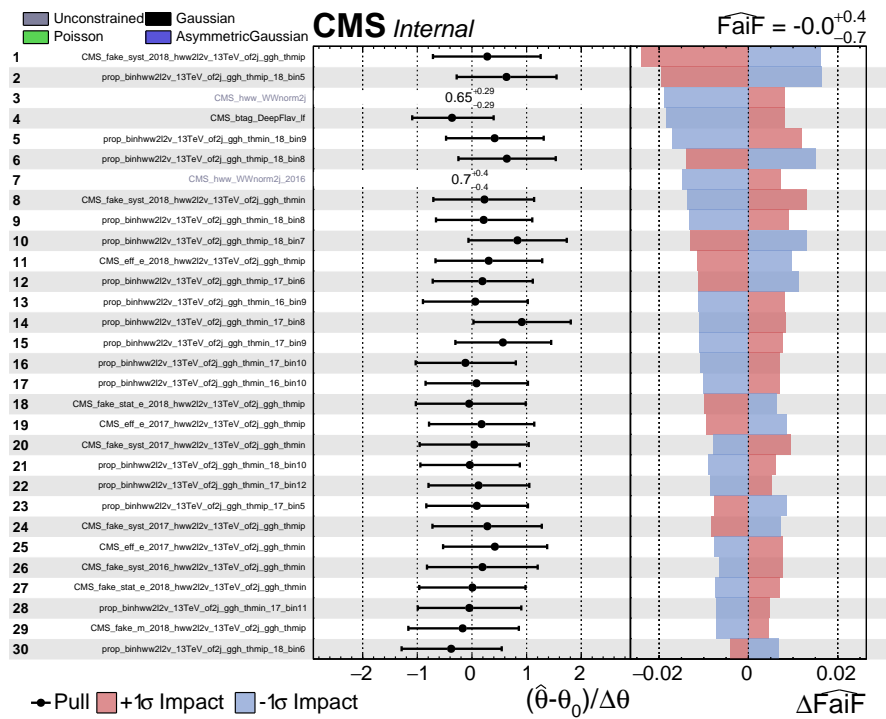
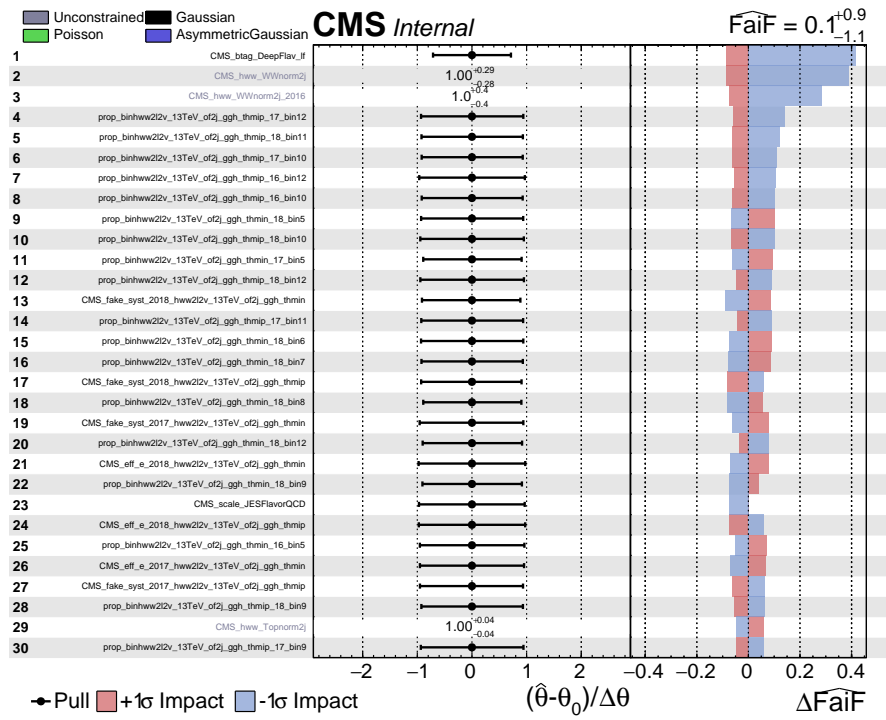


# Expected (top) and observed (bottom) impacts for the HVV $\Lambda_1$ coupling



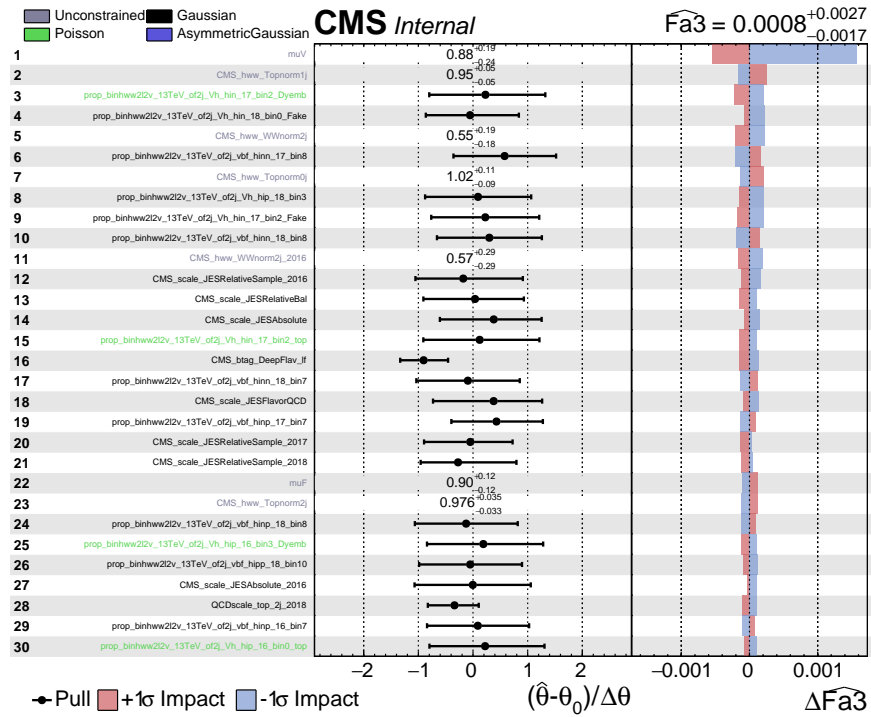
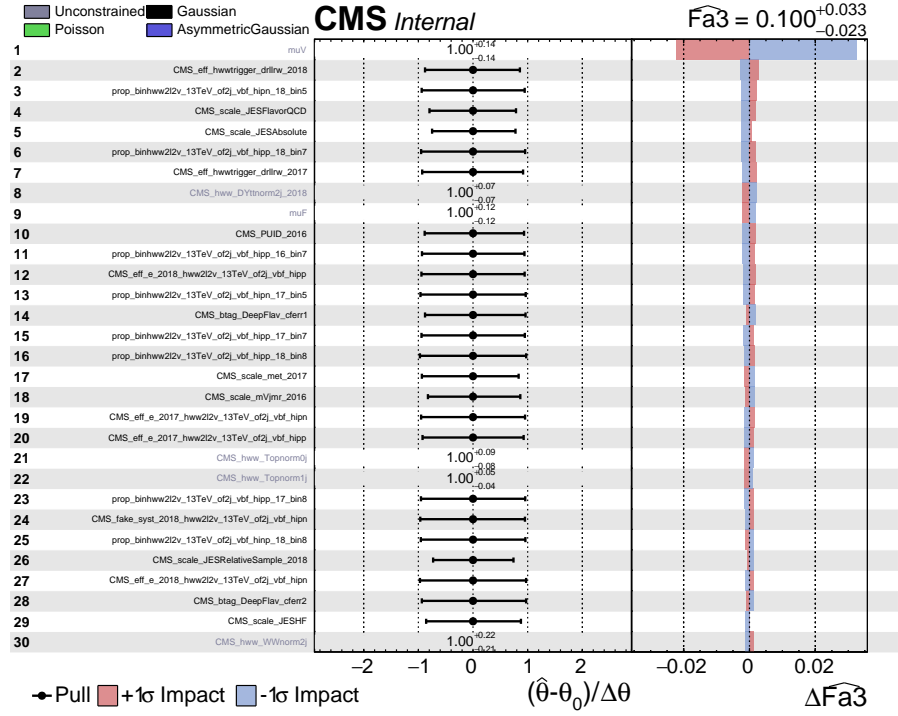
# Expected (top) and observed (bottom) impacts for the HVV $\Lambda_1^{Z\gamma}$ coupling



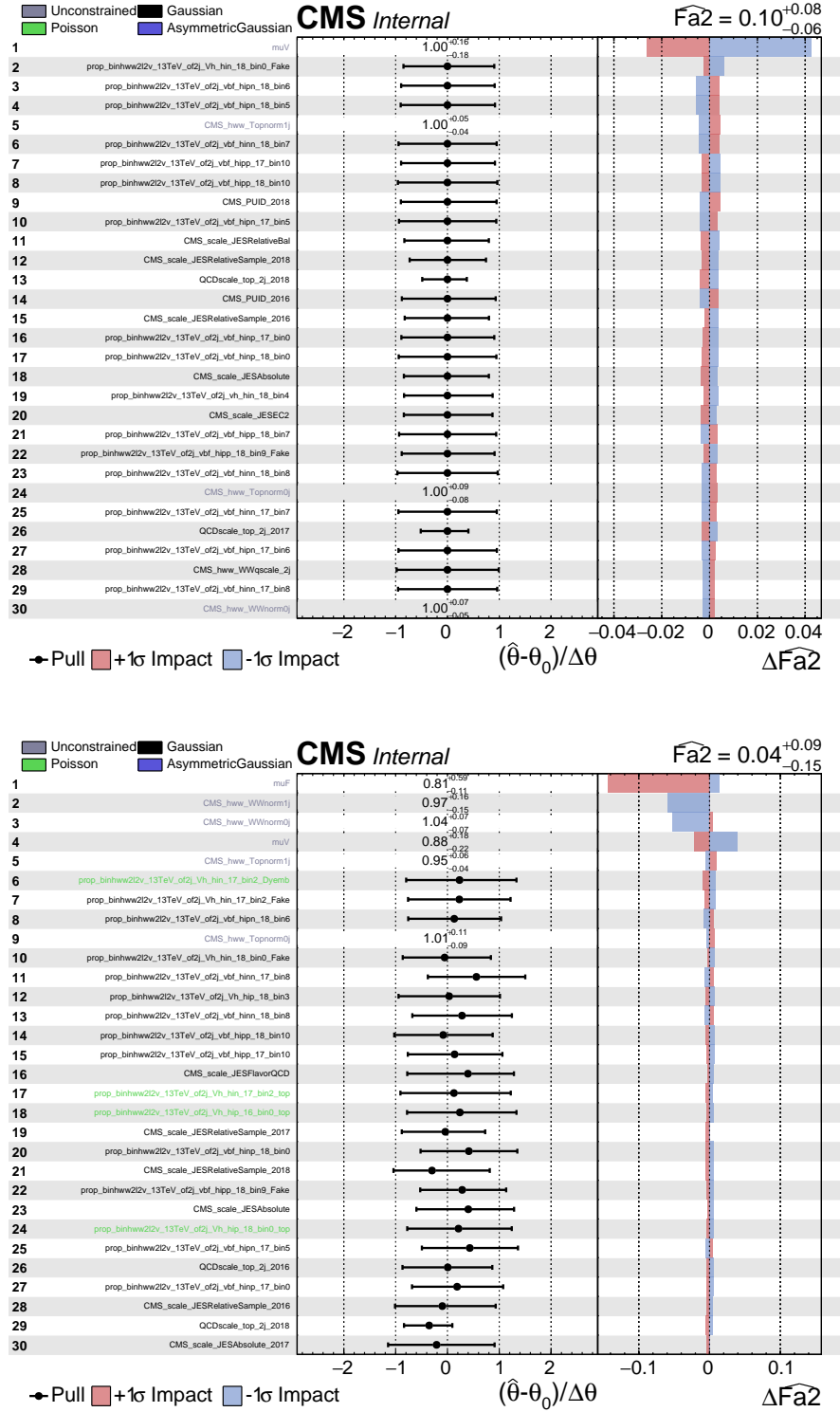
Hgg  $a_3$  coupling - expected impacts



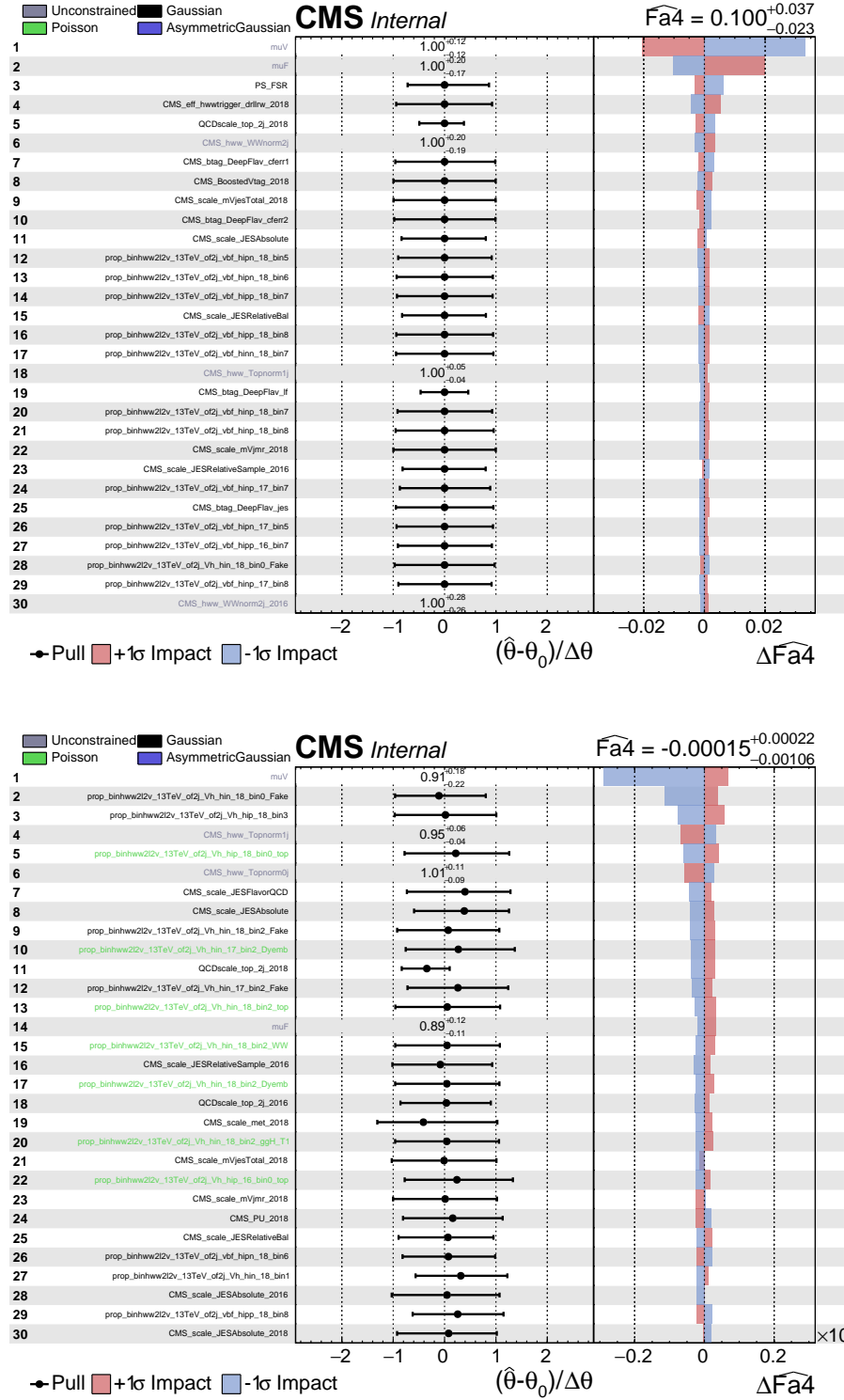
## Expected (top) and observed (bottom) impacts for the SMEFT HVV $a_3$ coupling



# Expected (top) and observed (bottom) impacts for the SMEFT HVV $a_2$ coupling



## Expected (top) and observed (bottom) impacts for the SMEFT HVV $\Lambda_1$ coupling





# Bibliography

- [1] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, sep 2012. doi: 10.1016/j.physletb.2012.08.021. URL <https://doi.org/10.1016%2Fj.physletb.2012.08.021>.
- [2] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, sep 2012. doi: 10.1016/j.physletb.2012.08.020. URL <https://doi.org/10.1016%2Fj.physletb.2012.08.020>.
- [3] J. I. Illana and A. J. Cano. Quantum field theory and the structure of the Standard Model. 2022. doi: 10.48550/arXiv.2211.14636. URL <https://doi.org/10.48550/arXiv.2211.14636>.
- [4] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. doi: [10.1103/PhysRevLett.81.1562](https://doi.org/10.1103/PhysRevLett.81.1562). URL <https://doi.org/10.1103/PhysRevLett.81.1562>.
- [5] T2K Collaboration. The T2K experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 659(1):106–135, dec 2011. doi: 10.1016/j.nima.2011.06.067. URL <https://doi.org/10.1016%2Fj.nima.2011.06.067>.
- [6] NO $\nu$ A Collaboration. Improved measurement of neutrino oscillation parameters by the NO $\nu$ A experiment. *Physical Review D*, 106(3), aug 2022. doi: 10.1103/physrevd.106.032004. URL <https://doi.org/10.1103%2Fphysrevd.106.032004>.
- [7] T. W. B. Kibble. History of electroweak symmetry breaking. 2015. URL <https://doi.org/10.48550/arXiv.1502.06276>.

- [8] A. Kirillov, Jr. *An Introduction to Lie Groups and Lie Algebras*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2008. doi: 10.1017/CBO9780511755156.
- [9] A. Beekman et al. An introduction to spontaneous symmetry breaking. *Sci-Post Physics Lecture Notes*, dec 2019. doi: 10.21468/scipostphyslectnotes.11. URL <https://doi.org/10.21468%2Fscipostphyslectnotes.11>.
- [10] J. F. Donoghue et al. Likely values of the Higgs vacuum expectation value. *Physical Review D*, 81(7), apr 2010. doi: 10.1103/physrevd.81.073003. URL <https://doi.org/10.1103%2Fphysrevd.81.073003>.
- [11] CMS Collaboration. A measurement of the Higgs boson mass in the diphoton decay channel. *Physics Letters B*, 805:135425, jun 2020. doi: 10.1016/j.physletb.2020.135425. URL <https://doi.org/10.1016%2Fj.physletb.2020.135425>.
- [12] Simon J. D. King et al. Discovering the origin of yukawa couplings at the LHC with a singlet Higgs and vector-like quarks. *Journal of High Energy Physics*, 2021(5), may 2021. doi: 10.1007/jhep05(2021)144. URL <https://doi.org/10.1007%2Fjhep05%282021%29144>.
- [13] I. de Medeiros Varzielas and S. F. King. Origin of yukawa couplings for Higgs bosons and leptiquarks. *Physical Review D*, 99(9), may 2019. doi: 10.1103/physrevd.99.095029. URL <https://doi.org/10.1103%2Fphysrevd.99.095029>.
- [14] M. Bargiotti et al. Present knowledge of the cabibbo-kobayashi-maskawa matrix. *La Rivista del Nuovo Cimento*, 23(3):1–71, mar 2000. doi: 10.1007/bf03548883. URL <https://doi.org/10.1007%2Fbf03548883>.
- [15] P. Hansen. Results from the UA1 and UA2 Experiments. In *International School of Physics Enrico Fermi: Elementary Particles*, 6 1984. URL <https://cds.cern.ch/record/182774/files/cer-000094196.pdf>.
- [16] Elizabeth Locci. Status of W and Z Physics in the UA1 and UA2 Experiments. In *7th European Symposium on Nucleon Antinucleon Interactions: ANTIPROTON 86*, 10 1986. URL <https://cds.cern.ch/record/172153>.

- [17] T. Taylor and D. Treille. The Large Electron Positron Collider (LEP): Probing the Standard Model. *Adv. Ser. Direct. High Energy Phys.*, 27:217–261, 2017. doi: 10.1142/9789814749145\_0007. URL <https://cds.cern.ch/record/2312570>.
- [18] CDF and D0 Collaborations. Higgs boson studies at the Tevatron. *Physical Review D*, 88(5), sep 2013. doi: 10.1103/physrevd.88.052014. URL <https://doi.org/10.1103%2Fphysrevd.88.052014>.
- [19] A. Deur et al. The QCD running coupling. *Progress in Particle and Nuclear Physics*, 90:1–74, sep 2016. doi: 10.1016/j.pnpnp.2016.04.003. URL <https://doi.org/10.1016%2Fj.pnpnp.2016.04.003>.
- [20] A. Di Giacomo. Understanding color confinement. *EPJ Web of Conferences*, 70:00019, 2014. doi: 10.1051/epjconf/20147000019. URL <https://doi.org/10.1051%2Fepjconf%2F20147000019>.
- [21] Frank Wilczek. Nobel lecture: Asymptotic freedom: From paradox to paradigm. *Reviews of Modern Physics*, 77(3):857–870, sep 2005. doi: 10.1103/revmodphys.77.857. URL <https://doi.org/10.1103%2Frevmodphys.77.857>.
- [22] Laszlo L. Jenkovszky, V. K. Magas, and F. Paccanoni. Duality in the Bjorken variable between small and large x. In *Crimean Summer School-Seminar on New Trends in High-Energy Physics (Crimean 2000)*, pages 121–127, 5 2000.
- [23] R L Workman et al. (Particle Data Group). Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2022(8), 08 2022. ISSN 2050-3911. doi: 10.1093/ptep/ptac097. URL <https://doi.org/10.1093/ptep/ptac097>. 083C01.
- [24] M. Klein and R. Yoshida. Collider physics at HERA. *Progress in Particle and Nuclear Physics*, 61(2):343–393, oct 2008. doi: 10.1016/j.pnpnp.2008.05.002. URL <https://doi.org/10.1016%2Fj.pnpnp.2008.05.002>.
- [25] DF Collaboration. Combined search for the Standard Model Higgs boson decaying to a  $b\bar{b}$  pair using the full cdf data set. *Phys. Rev. Lett.*, 109:111802, Sep 2012. doi: 10.1103/PhysRevLett.109.111802. URL <https://link.aps.org/doi/10.1103/PhysRevLett.109.111802>.

- [26] D0 Collaboration. Combined search for the Standard Model Higgs boson decaying to  $b\bar{b}$  using the d0 run ii data set. *Phys. Rev. Lett.*, 109:121802, Sep 2012. doi: 10.1103/PhysRevLett.109.121802. URL <https://link.aps.org/doi/10.1103/PhysRevLett.109.121802>.
- [27] CDF Collaboration and D0 Collaboration. Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron. *Phys. Rev. Lett.*, 109:071804, Aug 2012. doi: 10.1103/PhysRevLett.109.071804. URL <https://link.aps.org/doi/10.1103/PhysRevLett.109.071804>.
- [28] A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature Physics*, 607(7917):60–68, jul 2022. doi: 10.1038/s41586-022-04892-x. URL <https://doi.org/10.1038/2Fs41586-022-04892-x>.
- [29] LHC Higgs Cross Section Working Group:. Handbook of lhc Higgs cross sections: 3. Higgs properties: Report of the lhc Higgs cross section working group, 2013. URL <http://cds.cern.ch/record/1559921>.
- [30] CERN. Cern yellow reports: Monographs, vol 2 (2017): Handbook of lhc Higgs cross sections: 4. deciphering the nature of the Higgs sector, 2017. URL <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32>.
- [31] CMS Collaboration. Evidence for the direct decay of the 125 GeV Higgs boson to fermions. *Nature Physics*, 10, 01 2014. doi: 10.1038/nphys3005.
- [32] J. de Blas et al. Higgs boson studies at future particle colliders. *Journal of High Energy Physics*, 2020(1), jan 2020. doi: 10.1007/jhep01(2020)139. URL <https://doi.org/10.1007/2Fjhep01%282020%29139>.
- [33] CMS Collaboration. Search for non-resonant Higgs boson pair production in the final state with two bottom quarks and two tau leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV, 2022. URL <https://doi.org/10.48550/arXiv.2206.09401>.
- [34] M. et al. Tanabashi. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018. doi: 10.1103/PhysRevD.98.030001. URL <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.



- [35] CMS Collaboration. Search for invisible decays of the Higgs boson produced via vector boson fusion in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Physical Review D*, 105(9), may 2022. doi: 10.1103/physrevd.105.092007. URL <https://doi.org/10.1103%2Fphysrevd.105.092007>.
- [36] CMS Collaboration. Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Journal of High Energy Physics*, 2021(11), nov 2021. doi: 10.1007/jhep11(2021)153. URL <https://doi.org/10.1007%2Fjhep11%282021%29153>.
- [37] Meenakshi Narain et al. The Future of US Particle Physics – The Snowmass 2021 Energy Frontier Report, 2023. URL <https://doi.org/10.48550/arXiv.2211.11084>.
- [38] A. Falkowski. Effective field theory approach to LHC Higgs data. *Pramana*, 87(3), aug 2016. doi: 10.1007/s12043-016-1251-5. URL <https://doi.org/10.1007%2Fs12043-016-1251-5>.
- [39] Michele Maggiore. *A Modern introduction to quantum field theory*. Oxford Master Series in Physics. 2005. ISBN 978-0-19-852074-0.
- [40] A. Falkowski. Higgs Basis: Proposal for an EFT basis choice for LHC HXSWG. 2015. URL <https://cds.cern.ch/record/2001958>.
- [41] B. Grzadkowski et al. Dimension-six terms in the Standard Model lagrangian. *Journal of High Energy Physics*, 2010(10), oct 2010. doi: 10.1007/jhep10(2010)085. URL <https://doi.org/10.1007%2Fjhep10%282010%29085>.
- [42] R. Alonso et al. Renormalization group evolution of the Standard Model dimension six operators III: gauge coupling dependence and phenomenology. *Journal of High Energy Physics*, 2014(4), apr 2014. doi: 10.1007/jhep04(2014)159. URL <https://doi.org/10.1007%2Fjhep04%282014%29159>.
- [43] Gerhard Buchalla et al. A systematic approach to the SILH lagrangian. *Nuclear Physics B*, 894:602–620, may 2015. doi: 10.1016/j.nuclphysb.2015.03.024. URL <https://doi.org/10.1016%2Fj.nuclphysb.2015.03.024>.
- [44] Oliver Witzel. Review on composite Higgs models. 2019. doi: <https://doi.org/10.48550/arXiv.1901.08216>.

- [45] CMS Collaboration. Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state. *Physics Letters B*, 775:1–24, dec 2017. doi: 10.1016/j.physletb.2017.10.021. URL <https://doi.org/10.1016%2Fj.physletb.2017.10.021>.
- [46] CMS Collaboration. Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state. *Physical Review D*, 104(5), sep 2021. doi: 10.1103/physrevd.104.052004. URL <https://doi.org/10.1103%2Fphysrevd.104.052004>.
- [47] Andrei V. Gritsan, Jeffrey Roskes, Ulascan Sarica, Markus Schulze, Meng Xiao, and Yaofu Zhou. New features in the jhu generator framework: Constraining Higgs boson properties from on-shell and off-shell production. *Phys. Rev. D*, 102:056022, Sep 2020. doi: 10.1103/PhysRevD.102.056022. URL <https://link.aps.org/doi/10.1103/PhysRevD.102.056022>.
- [48] CERN. Cern yellow reports: Monographs, vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, 2017. URL <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32>.
- [49] Sara Bolognesi et al. Spin and parity of a single-produced resonance at the LHC. *Physical Review D*, 86(9), nov 2012. doi: 10.1103/physrevd.86.095031. URL <https://doi.org/10.1103%2Fphysrevd.86.095031>.
- [50] CMS Collaboration. Constraints on the spin-parity and anomalous couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Physical Review D*, 92(1), jul 2015. doi: 10.1103/physrevd.92.012004. URL <https://doi.org/10.1103%2Fphysrevd.92.012004>.
- [51] CMS Collaboration. Measurements of  $t\bar{t}H$  production and the cp structure of the yukawa interaction between the Higgs boson and top quark in the diphoton decay channel. *Phys. Rev. Lett.*, 125(6):061801, 2020. doi: 10.1103/PhysRevLett.125.061801. URL <https://cds.cern.ch/record/2713705>.
- [52] CMS Collaboration. Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production, 2022. URL <https://doi.org/10.1038/s41567-022-01682-0>.
- [53] CMS Collaboration. Constraints on anomalous Higgs boson couplings to vector bosons and fermions from the production of Higgs bosons using the  $\tau\tau$  final state, 2022. URL <https://doi.org/10.48550/arXiv.2205.05120>.

- [54] CMS Collaboration. Search for CP violation in  $t\bar{t}h$  and  $th$  production in multilepton channels in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2022. doi: doi.org/10.48550/arXiv.2208.02686. URL <https://doi.org/10.48550/arXiv.2208.02686>.
- [55] CMS Collaboration. Analysis of the CP structure of the Yukawa coupling between the Higgs boson and leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Journal of High Energy Physics*, 2022(6), jun 2022. doi: 10.1007/jhep06(2022)012. URL <https://doi.org/10.1007/JHEP06%282022%29012>.
- [56] Georges Aad et al. Test of CP-invariance of the Higgs boson in vector-boson fusion production and its decay into four leptons. 4 2023. URL <https://doi.org/10.48550/arXiv.2304.09612>.
- [57] Georges Aad et al. Measurement of the associated production of a Higgs boson decaying into  $b$ -quarks with a vector boson at high transverse momentum in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Lett. B*, 816:136204, 2021. doi: 10.1016/j.physletb.2021.136204.
- [58] Georges Aad et al. Measurements of the Higgs boson inclusive and differential fiducial cross-sections in the diphoton decay channel with  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *JHEP*, 08:027, 2022. doi: 10.1007/JHEP08(2022)027.
- [59] ATLAS Collaboration. Probing the CP nature of the top-Higgs Yukawa coupling in  $t\bar{t}H$  and  $tH$  events with  $H \rightarrow b\bar{b}$  decays using the ATLAS detector at the LHC. 3 2023. URL <https://doi.org/10.48550/arXiv.2303.05974>.
- [60] ATLAS Collaboration. Measurement of the properties of Higgs boson production at  $\sqrt{s} = 13$  TeV in the  $H \rightarrow \gamma\gamma$  channel using  $139 \text{ fb}^{-1}$  of  $pp$  collision data with the ATLAS experiment. 7 2022. doi: doi.org/10.48550/arXiv.2207.00348. URL <https://doi.org/10.48550/arXiv.2207.00348>.
- [61] Lucio Rossi. The LHC Main Dipoles and Quadrupoles Toward Series Production. *IEEE Transactions on Applied Superconductivity*, June 2003.
- [62] ATLAS Collaboration. ATLAS data quality operations and performance for 2015–2018 data-taking. *Journal of Instrumentation*, 15(04):P04003–P04003, apr 2020. doi: 10.1088/1748-0221/15/04/p04003. URL <https://doi.org/10.1088/1748-0221/15/04/p04003>.

- [63] ALICE Collaboration. Performance of the ALICE experiment at the CERN LHC. *International Journal of Modern Physics A*, 29(24):1430044, sep 2014. doi: 10.1142/s0217751x14300440. URL <https://doi.org/10.1142%2Fs0217751x14300440>.
- [64] I. Belyaev et al. The history of LHCb. *The European Physical Journal H*, 46(1), mar 2021. doi: 10.1140/epjh/s13129-021-00002-z. URL <https://doi.org/10.1140%2Fepjh%2Fs13129-021-00002-z>.
- [65] CERN Home. Experiments. <https://home.cern/science/experiments>. [Online; accessed 22-March-2023].
- [66] CMS Collaboration. Cms luminosity - public results. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>. [Online; accessed 22-March-2023].
- [67] I. Neutelings. TikZ.net CMS coordinate system. URL [https://tikz.net/axis3d\\_cms/](https://tikz.net/axis3d_cms/).
- [68] Tai Sakuma. Cutaway diagrams of CMS detector. 2019. URL <https://cds.cern.ch/record/2665537>.
- [69] G. L. Bayatian et al. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. 2006.
- [70] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 5(03):T03021–T03021, mar 2010. doi: 10.1088/1748-0221/5/03/t03021. URL <https://doi.org/10.1088%2F1748-0221%2F5%2F03%2Ft03021>.
- [71] CMS Collaboration. Alignment of the CMS silicon tracker during commissioning with cosmic rays. *JINST*, 5:T03009. 39 p, 2010. doi: 10.1088/1748-0221/5/03/T03009. URL <https://cds.cern.ch/record/1211825>.
- [72] A. Dominguez et al. CMS Technical Design Report for the Pixel Detector Upgrade. Technical report, 2012. URL <https://cds.cern.ch/record/1481838>.
- [73] Martin Lipinski. The Phase-1 Upgrade of the CMS Pixel Detector. Technical Report 06, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2265423>.

- [74] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. CERN, Geneva, 2006. URL <https://cds.cern.ch/record/922757>. There is an error on cover due to a technical problem for some items.
- [75] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Journal of Instrumentation*, 13(10):1748–0221, June 2018. URL <https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.1088%2F1748-0221%2F13%2F06%2FP06015&v=7eff5fa6>.
- [76] CMS Collaboration. Summaries of CMS cross Section Measurements. . URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>.
- [77] G Bauer et al. The data-acquisition system of the cms experiment at the lhc. *Journal of Physics: Conference Series*, 331(2):022021, dec 2011. doi: 10.1088/1742-6596/331/2/022021. URL <https://dx.doi.org/10.1088/1742-6596/331/2/022021>.
- [78] Manfred Jeitler. Upgrade of the trigger system of cms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 718:11–15, 08 2013. doi: 10.1016/j.nima.2012.08.091.
- [79] CMS Collaboration. *JINST*, 12:P10003. 82 p, Oct 2017. doi: 10.1088/1748-0221/12/10/P10003. URL <https://cds.cern.ch/record/2270046>.
- [80] CMS Collaboration. Muon reconstruction performance during Run II. 2019. URL <https://cds.cern.ch/record/2682902>.
- [81] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16:P05014. 81 p, May 2021. doi: 10.1088/1748-0221/16/05/P05014. URL <https://cds.cern.ch/record/2747266>.
- [82] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, May . doi: 10.1088/1748-0221/16/05/p05014.

- [83] M. Cacciari et al. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, apr 2008. doi: 10.1088/1126-6708/2008/04/063. URL <https://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [84] CMS Collaboration. Performance of the Deepjet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector. *Technical Report CMS-DP-2018-058*, November 2018.
- [85] CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(05):P05011–P05011, may 2018. doi: 10.1088/1748-0221/13/05/p05011.
- [86] CMS Collaboration. B-tagging performance of the CMS Legacy dataset 2018. 2021. URL <https://cds.cern.ch/record/2759970>.
- [87] CMS collaboration. Identification of b-quark jets with the CMS experiment. *Journal of Instrumentation*, 8(04):P04013–P04013, apr 2013. doi: 10.1088/1748-0221/8/04/p04013. URL <https://doi.org/10.1088%2F1748-0221%2F8%2F04%2Fp04013>.
- [88] CMS Collaboration. The Phase-2 Upgrade of the CMS Beam Radiation Instrumentation and Luminosity Detectors. Technical report, CERN, Geneva, 2021. URL <https://cds.cern.ch/record/2759074>. This is the final version, approved by the LHCC.
- [89] G Apollinari, I Béjar Alonso, O Brüning, M Lamont, and L Rossi. *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2015. doi: 10.5170/CERN-2015-005. URL <https://cds.cern.ch/record/2116337>.
- [90] CMS Collaboration. Analytical Method (AM) for DT Trigger Primitive Generation in Phase 2. Performance results on simulation. 2021. URL <https://cds.cern.ch/record/2792319>.
- [91] CERN. Geneva. The LHC Experiments Committee. The Phase-2 Upgrade of the CMS Muon Detectors. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2283189>. This is the final version, approved by the LHCC.
- [92] CMS Collaboration. Performance of the DT Phase-2 readout and local reconstruction with cosmic ray data from the DT slice-test. 2021. URL <https://cds.cern.ch/record/2799473>.

- [93] CMS Collaboration. Measurements of the Higgs boson production cross section and couplings in the W boson pair decay channel in proton-proton collisions at  $\sqrt{s} = 13$  TeV, 2022.
- [94] CMS Collaboration. Measuring the Higgs boson decay to WW is 90% physics. The other half is teamwork!, . URL <https://cms.cern/news/measuring-higgs-boson-decay-ww-90-physics-other-half-teamwork>.
- [95] CMS Collaboration. Measurement of the inclusive and differential Higgs boson production cross sections in the leptonic WW decay mode at  $\sqrt{s} = 13$  TeV. *Journal of High Energy Physics*, 2021(3), mar 2021. doi: 10.1007/jhep03(2021)003. URL <https://doi.org/10.1007%2Fjhep03%282021%29003>.
- [96] Adrian Alvarez Fernandez. Measurement of Higgs boson production via gluon fusion in the HWW decay channel with the CMS experiment. Medida de la producción del bosón de Higgs mediante fusión de gluones en el canal de desintegración HWW con el experimento CMS, 2020. URL <https://cds.cern.ch/record/2750126>. Presented 26 Jan 2021.
- [97] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2257069>.
- [98] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at  $\sqrt{s} = 13$  TeV. Technical report, CERN, Geneva, 2018. URL <https://cds.cern.ch/record/2621960>.
- [99] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at  $\sqrt{s} = 13$  TeV. Technical report, CERN, Geneva, 2019. URL <https://cds.cern.ch/record/2676164>.
- [100] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. ISSN 0168-9002. doi: doi.org/10.1016/S0168-9002(03)01368-8. URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [101] T. Sjöstrand et al. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026–026, may 2006. doi: 10.1088/1126-6708/2006/05/026. URL <https://doi.org/10.1088%2F1126-6708%2F2006%2F05%2F026>.

- [102] T. Sjöstrand et al. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, jun 2015. doi: 10.1016/j.cpc.2015.01.024. URL <https://doi.org/10.1016%2Fj.cpc.2015.01.024>.
- [103] V. Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *The European Physical Journal C*, 76(3), mar 2016. doi: 10.1140/epjc/s10052-016-3988-x. URL <https://doi.org/10.1140%2Fepjc%2Fs10052-016-3988-x>.
- [104] CMS Collaboration. Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1), jan 2020. doi: 10.1140/epjc/s10052-019-7499-4. URL <https://doi.org/10.1140%2Fepjc%2Fs10052-019-7499-4>.
- [105] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO vector-boson production matched with shower in POWHEG. *JHEP*, 07:060, 2008. doi: 10.1088/1126-6708/2008/07/060.
- [106] Paolo Nason. A new method for combining NLO QCD with shower monte carlo algorithms. *Journal of High Energy Physics*, 2004(11):040–040, nov 2004. doi: 10.1088/1126-6708/2004/11/040. URL <https://doi.org/10.1088%2F1126-6708%2F2004%2F11%2F040>.
- [107] S. Frixione et al. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007(11):070–070, nov 2007. doi: 10.1088/1126-6708/2007/11/070. URL <https://doi.org/10.1088%2F1126-6708%2F2007%2F11%2F070>.
- [108] S. Alioli et al. A general framework for implementing NLO calculations in shower monte carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6), jun 2010. doi: 10.1007/jhep06(2010)043. URL <https://doi.org/10.1007%2Fjhep06%282010%29043>.
- [109] K. Hamilton et al. Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator, 2015. URL <https://doi.org/10.48550/arXiv.1501.04637>.
- [110] R. Frederix and K. Hamilton. Extending the minlo method. *Journal of High Energy Physics*, 2016, 12 2015. doi: 10.1007/JHEP05(2016)042.



- [111] Sara Bolognesi et al. Spin and parity of a single-produced resonance at the LHC. *Physical Review D*, 86(9), nov 2012. doi: 10.1103/physrevd.86.095031. URL <https://doi.org/10.1103%2Fphysrevd.86.095031>.
- [112] Yanyan Gao et al. Spin determination of single-produced resonances at hadron colliders. *Physical Review D*, 81(7), apr 2010. doi: 10.1103/physrevd.81.075022. URL <https://doi.org/10.1103%2Fphysrevd.81.075022>.
- [113] Ian Anderson et al. Constraining anomalous HVV interactions at proton and lepton colliders. *Physical Review D*, 89(3), feb 2014. doi: 10.1103/physrevd.89.035007. URL <https://doi.org/10.1103%2Fphysrevd.89.035007>.
- [114] A. V. Gritsan et al. Constraining anomalous Higgs boson couplings to the heavy-flavor fermions using matrix element techniques. *Physical Review D*, 94(5), sep 2016. doi: 10.1103/physrevd.94.055023. URL <https://doi.org/10.1103%2Fphysrevd.94.055023>.
- [115] G. Luisoni et al. HW  $\pm$ /HZ 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *Journal of High Energy Physics*, 2013(10), oct 2013. doi: 10.1007/jhep10(2013)083. URL <https://doi.org/10.1007%2Fjhep10%282013%29083>.
- [116] 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. *Journal of High Energy Physics*, 2014(7), jul 2014. doi: 10.1007/jhep07(2014)079. URL <https://doi.org/10.1007%2Fjhep07%282014%29079>.
- [117] John M. Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. *Nuclear Physics B Proceedings Supplements*, 205:10–15, August 2010. doi: 10.1016/j.nuclphysbps.2010.08.011.
- [118] HWW Latinos Framework. Samples cross-sections. URL <https://github.com/latinos/LatinoAnalysis/blob/master/NanoGardener/python/framework/samples/samplesCrossSections2016.py>.
- [119] CMS Collaboration. Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at  $\sqrt{s}=13$  Tev. *Physics Letters B*, 791: 96–129, 2019. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2018.12.073>. URL <https://www.sciencedirect.com/science/article/pii/S0370269319301169>.

- [120] M. Czakon et al. Top-pair production at the LHC through NNLO QCD and NLO EW. *Journal of High Energy Physics*, 2017(10), oct 2017. doi: 10.1007/jhep10(2017)186. URL <https://doi.org/10.1007%2Fjhep10%282017%29186>.
- [121] P. Meade et al. Transverse momentum resummation effects in  $w^+w^-$  measurements. *Physical Review D*, 90(11), dec 2014. doi: 10.1103/physrevd.90.114006. URL <https://doi.org/10.1103%2Fphysrevd.90.114006>.
- [122] P. Jaiswal and T. Okui. Explanation of the  $ww$  excess at the LHC by jet-veto resummation. *Physical Review D*, 90(7), oct 2014. doi: 10.1103/physrevd.90.073009. URL <https://doi.org/10.1103%2Fphysrevd.90.073009>.
- [123] CMS Collaboration. An embedding technique to determine backgrounds in proton-proton collision data. *Journal of Instrumentation*, 14(06):P06032–P06032, jun 2019. doi: 10.1088/1748-0221/14/06/p06032. URL <https://doi.org/10.1088%2F1748-0221%2F14%2F06%2Fp06032>.
- [124] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018–P09018, sep 2020. doi: 10.1088/1748-0221/15/09/p09018. URL <https://doi.org/10.1088%2F1748-0221%2F15%2F09%2Fp09018>.
- [125] CMS Collaboration. Performance of the CMS level-1 trigger in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Journal of Instrumentation*, 15(10):P10017–P10017, oct 2020. doi: 10.1088/1748-0221/15/10/p10017. URL <https://doi.org/10.1088%2F1748-0221%2F15%2F10%2Fp10017>.
- [126] CMS Collaboration. Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at  $\sqrt{s} = 13$  tev, 2021.
- [127] CMS Collaboration. Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-probe technique. 2020. URL <https://cds.cern.ch/record/2724492>.
- [128] A. Bodek et al. Extracting muon momentum scale corrections for hadron collider experiments. *The European Physical Journal C*, 72(10), oct 2012. doi: 10.1140/epjc/s10052-012-2194-8. URL <https://doi.org/10.1140%2Fepjc%2Fs10052-012-2194-8>.

- [129] CMS Collaboration. Jet PU ID: efficiency and scale factors with Z+jets, . URL <https://indico.cern.ch/event/1132840819/contributions/3527893/attachments/1892944/3122168/puid.pdf>.
- [130] CMS Collaboration. CMS Phase 1 heavy flavour identification performance and developments. 2017. URL <http://cds.cern.ch/record/2263802>.
- [131] CMS Collaboration. Heavy flavor identification at CMS with deep neural networks. 2017. URL <https://cds.cern.ch/record/2255736>.
- [132] J. et al. Larkoski, A. Soft Drop. *JHEP*, 05:146, 2014. doi: 10.1007/JHEP05(2014)146.
- [133] R. et al. Kogler. Jet Substructure at the Large Hadron Collider. Jet Substructure at the Large Hadron Collider : Experimental Review. *Rev. Mod. Phys.*, 91:045003. 44 p, Dec 2019. doi: 10.1103/RevModPhys.91.045003. URL <https://cds.cern.ch/record/2641634>.
- [134] J Thaler and K. Van Tilburg. Maximizing Boosted Top Identification by Minimizing N-subjettiness. *JHEP*, 02:093, 2012. doi: 10.1007/JHEP02(2012)093.
- [135] Huang Huang. Searches for diboson resonances at CMS. Technical report, 2017. URL <https://cds.cern.ch/record/2641627>.
- [136] CMS Collaboration. W/z-tagging of jets, working points and scale factors. Twiki, CERN, 2017. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JetWtagging>.
- [137] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018–P09018, sep 2020. doi: 10.1088/1748-0221/15/09/p09018. URL <https://doi.org/10.1088%2F1748-0221%2F15%2F09%2Fp09018>.
- [138] CMS Collaboration. Pileup-per-particle identification: optimisation for Run 2 Legacy and beyond. 2021. URL <http://cds.cern.ch/record/2751563>.
- [139] CMS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s}=13$  TeV using the CMS detector. *Journal of Instrumentation*, 14(07):P07004–P07004, jul 2019. doi: 10.1088/1748-0221/14/07/p07004. URL <https://doi.org/10.1088%2F1748-0221%2F14%2F07%2Fp07004>.

- [140] ATLAS Collaboration. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Physics Letters B*, 726(1-3):120–144, oct 2013. doi: 10.1016/j.physletb.2013.08.026. URL <https://doi.org/10.1016%2Fj.physletb.2013.08.026>.
- [141] Paolo Nason, Carlo Oleari, Marco Rocco, and Marco Zaro. An interface between the powheg box and MadGraph5\_aMC@NLO. *The European Physical Journal C*, 80(10), oct 2020. doi: 10.1140/epjc/s10052-020-08559-7. URL <https://doi.org/10.1140%2Fepjc%2Fs10052-020-08559-7>.
- [142] ATLAS, CMS Collaborations, and LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, 2011. URL <http://cds.cern.ch/record/1379837>.
- [143] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at  $\sqrt{s} = 13$  TeV in 2015 and 2016 at cms, 2021.
- [144] CMS Collaboration. Identification techniques for highly boosted w bosons that decay into hadrons. *Journal of High Energy Physics*, 2014(12), dec 2014. doi: 10.1007/jhep12(2014)017. URL <https://doi.org/10.1007%2Fjhep12%282014%29017>.
- [145] CMS Collaboration. Measurements of properties of the Higgs boson decaying to a w boson pair in pp collisions at  $\sqrt{s} = 13$  TeV. *Physics Letters B*, 791:96–129, apr 2019. doi: 10.1016/j.physletb.2018.12.073. URL <https://doi.org/10.1016%2Fj.physletb.2018.12.073>.
- [146] ATLAS Collaboration. Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13$  TeV with the ATLAS detector at the LHC. *Physical Review Letters*, 117(18), oct 2016. doi: 10.1103/physrevlett.117.182002. URL <https://doi.org/10.1103%2Fphysrevlett.117.182002>.
- [147] CMS Collaboration. Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 13$  TeV. *Journal of High Energy Physics*, 2018(7), jul 2018. doi: 10.1007/jhep07(2018)161. URL <https://doi.org/10.1007%2Fjhep07%282018%29161>.
- [148] Giampiero Passarino. Higgs CAT. *The European Physical Journal C*, 74(5), may 2014. doi: 10.1140/epjc/s10052-014-2866-7. URL <https://doi.org/10.1140%2Fepjc%2Fs10052-014-2866-7>.

- 
- [149] Thomas Rahlf. *Statistical Inference*, pages 1–31. Springer Berlin Heidelberg, Berlin, Heidelberg, 2021. ISBN 978-3-642-40458-0. doi: 10.1007/978-3-642-40458-0\_16-2. URL [https://doi.org/10.1007/978-3-642-40458-0\\_16-2](https://doi.org/10.1007/978-3-642-40458-0_16-2).
- [150] Robert D. Cousins. Generalization of chisquare goodness-of fit test for binned data using saturated models , with application to histograms. 2013. URL [https://www.physics.ucla.edu/~cousins/stats/cousins\\_saturated.pdf](https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf).
- [151] CMS Collaboration. DT Slice Test 2021 calibration stability plots. (2022-001), January 2022. URL [https://cds.cern.ch/record/2801640/files/DP2022\\_001.pdf](https://cds.cern.ch/record/2801640/files/DP2022_001.pdf).
- [152] Lourdes Urda. Search for lepton flavour violating decays of the Higgs boson with Run II data. *PoS*, ICHEP2020:108, 2021. doi: 10.22323/1.390.0108. URL <https://cds.cern.ch/record/2790752>.