Constraining anomalous Higgs boson couplings using the CMS detector at the Large Hadron Collider

Thesis submitted for the degree of

Doctorat en Sciences

at the

Université Libre de Bruxelles

and the degree of

Doctor of Science: Physics

at the

University of Antwerp

to be defended by

Tomáš Kello

Promotors:
Prof. dr. Pierre Van Mechelen (UA)
Prof. dr. Laurent Favart (ULB)
Supervisor:
dr. Tahys Janssen (UA)

Antwerp, 2023

Jury members

University of Antwerp	Prof. dr. Johan Verbeeck	(chair)
	Prof. dr. Pierre Van Mechelen	(promotor)
	dr. Tahys Janssen	(supervisor)
Université Libre de Bruxelles	Prof. dr. Pascal Vanlaer	(secretary)
	Prof. dr. Laurent Favart	(promotor)
CIEMAT Madrid	dr. Maria Cepeda	

Abstract

This thesis presents constraints on the anomalous couplings (AC) of the Higgs boson (H) with vector bosons as obtained in analysis studying Higgs boson production and decay into the pair of W bosons. Various alternative Higgs spin-parity scenarios were considered, including CP violating effects. The study is performed on proton-proton collisions at a center-of-mass energy of 13 TeV, corresponding to 138 fb^{-1} of integrated luminosity, collected by the CMS detector at the LHC during the 2016-2018 period of data-taking. The analysis targets the dilepton different flavor ($e\mu$) final state with up to two associated jets, providing additional kinematic information used in combination with the matrix element technique to increase sensitivity to anomalous contribution at the production vertex. In cases where the analysis phase space has limited kinematic information, we rely on decay kinematics, also affected by anomalous effects. Up to four anomalous couplings are studied both independently and simultaneously, within the standard model effective field theory framework assumed in the latter case. All measurements and the corresponding confidence intervals for their values we obtained were found to be consistent with the expectations for the Standard Model Higgs boson. These results significantly surpass previous AC constraints obtained in the $H \to WW^*$ channel and are competitive with results obtained in other Higgs boson decay channels.

Abstract

Deze thesis presenteert limieten op de anomale koppeling (AK) van het Higgs-boson (H) met vectorbosonen, verkregen in de analyse van de productie van het Higgs-boson en het verval naar een paar W-bosonen. Meerdere alternatieve Higgs-spin-pariteit scenario's werden overwogen, inclusief CP-schendende effecten. De studie werd uitgevoerd bij proton-protonbotsingen met een massamiddelpuntsenergie van 13 TeV, en met een geïntegreerde luminositeit van 138 fb^{-1}, verzameld door de CMS-detector bij de LHC in 2016-2018. De analyse richt zich op de dilepton ($e\mu$) eindtoestand met maximaal twee geassocieerde jets, die aanvullende kinematische informatie verschaffen voor gebruik in combinatie met de matrixelementtechniek om de gevoeligheid voor anomale bijdragen aan de productievertex te vergroten. In gevallen waar de analyse faseruimte beperkte kinematische informatie heeft, gerbuiken we de verval-kinematica, die eveneens wordt beïnvloed door anomale effecten. Tot vier anomale koppelingen worden zowel onafhankelijk bestudeerd als samen in het kader van de effectieve veldentheorie van het standaardmodel. Alle metingen en de bijbehorende betrouwbaarheidsintervallen die we hebben verkregen bleken consistent te zijn met de verwachtingen voor het Higgs-boson van het Standaardmodel. Deze resultaten overtreffen aanzienlijk eerdere limietbepalingen op AK die werden verkregen in het $H \to WW^*$-kanaal en zijn competitief in vergelijking met resultaten verkregen in andere verval-kanalen van het Higgs-boson.

Abstrait

Cette thèse présente les contraintes sur les couplages anormaux (CA) du boson de Higgs (H) avec les bosons vecteurs obtenues en étudiant la production et la désintégration du boson de Higgs en une paire de bosons W. Divers scénarios alternatifs de parité et de spin du boson Higgs ont été envisagés, y compris les effets de violation de CP. Cette étude est réalisée sur des collisions proton-proton à une énergie dans le centre de masse de 13 TeV, correspondant à 138 fb^{-1} de luminosité intégrée, collectée par le détecteur CMS au LHC pendant la période d'acquisition de données 2016-2018. L'analyse cible l'état final de saveurs différentes de la paire de leptons ($e\mu$) en laquelle se désintègre les 2 bosons W avec jusqu'à deux jets associés, fournissant des informations cinématiques supplémentaires utilisées en combinaison avec la technique des éléments matriciels pour augmenter la sensibilité à la contribution anormale au vertex de production. Dans les cas où l'espace des phases d'analyse a des informations cinématiques limitées, nous nous appuyons sur la cinématique de désintégration qui est également affectée par des effets anormaux. Jusqu'à quatre couplages anormaux sont étudiés à la fois indépendamment et simultanément, dans le cadre de la théorie des champs effectifs du modèle standard. Toutes les mesures se sont avérées conformes aux attentes pour le boson de Higgs du modèle standard et de limites sur le valeurs possibles ces couplages ont été établies. Ces résultats dépassent largement les contraintes précédentes obtenues dans le canal $H \to WW^*$ et sont compétitifs avec les résultats obtenus dans d'autres canaux de désintégration du boson de Higgs.

Acknowledgements

With this thesis, I draw a close on over four years of dedicated data analysis and the development of cutting-edge software and hardware technology, both essential for the advancement of our research and with broader applications in future experiments. Foremost, I would like to express my sincere gratitude to the Excellence of Science (EOS) program funded through FNRS (the National Fund for Scientific Research) and FWO (the Research Foundation - Flanders) for their financial support, which made this research possible. I reserve my deepest appreciation for my promotors, Prof. dr. Pierre Van Mechelen (University of Antwerp) and Prof. dr. Laurent Favart (Université Libre de Bruxelles) for providing the base ground, support and ultimately their time to make this goal possible. I also wish to thank dr. Tahys Janssen for her supervision, guidance, and support throughout my doctoral studies. Special thanks go to Pierre, Tahys and Laurent for a thorough review of this manuscript.

I consider myself very lucky that I have had the privilege of collaborating with exceptional minds in the field of (not only) particle physics at CMS, which has greatly contributed to my professional and personal growth. Let me explicitly mention at least a small fraction of them: a huge thanks go to Dermot Moran and Lourdes Urda for their dedicated work on this research; to Lorenzo Viliani, Piergiulio Lenzi, Roberto Seidita and Ulascan Sarica for their helpful advice, and availability for countless discussions concerning this analysis; to numerous colleagues involved in the CMS HWW group that created a healthy and motivating environment throughout the years; to Hans Van Haevermaet for sharing our office as well as for daily discussions on every imaginable topic; to my colleagues involved in the CMS Tracker Phase-2 Upgrade project, particularly Pascal Vanlaer, Inna Makarenko, Eric Roose and Wim Beaumont for our fruitful collaboration; to Aleksandra Lelek, Patrick Connor and Senne Van Putte for their immense support in work related to the CMS Tracker Alignment; to Mees Van Kampen and Kamiel Janssens for being more than colleagues, especially during the difficult times of the Covid-19 pandemic; to Sarah Van Mierlo for having a big heart and also control of all bureaucratic procedures; to all of my UAntwerpen and ULB colleagues for creating a welcoming workspace, and finally, to all of my CMS colleagues worldwide for sharing their comments, questions, and ideas that have fueled my professional growth. I would also like to express my gratitude to those who contributed to my education before my doctoral studies, without whom I would not have embarked on this remarkable journey.

Another, equally important group of people that made my journey possible are my dearest friends. Both present in Antwerp and scattered around the world, they have always been there for me, providing emotional support, hugs, and the much-needed escape from the everyday duties. The pure fact that I am unable to create a list of them which would not require becoming a separate chapter is a real blessing to me.

Most importantly, let me express my greatest gratitude to my family, grand-parents, parents, and brothers, for their support and belief in me.

In addition, I extend my heartfelt thanks to Michal Souček and Juliana Slabeyová for their exceptional work on the beautiful cover design of this thesis.

Last but not least, I extend my thanks to the esteemed jury members, Prof. dr. Johan Verbeeck (UAntwerpen), Prof. dr. Pascal Vanlaer (ULB), and dr. Maria Cepeda (CIEMAT Madrid) for accepting their role and dedicating time to the fair review of my thesis.

CONTENTS

INTRODUCTION

Throughout human history, only sizeable and valuable advancements have been driven by our innate desire to explore the unknown. While there have been some dark moments, our motivation to uncover new knowledge has always been guided by noble causes, limited only by imagination and available technology. From the discovery of fire, we have been continuously asking questions and revealing answers – two aspects of human nature expanding the limited understanding of what is called the Universe. Despite the length of known human history, it is roughly the last 100 years of discoveries that have profoundly shaped our perception of natural laws, encompassing everything from the description of the vast space of macroscopic universe to the mechanics governing the smallest particles of matter.

In this thesis and related studies, we will be addressing the latter end of this broad spectrum, hopefully uncovering new details of the elementary universe. Specifically, the focus is on the particle that plays a special role in the hierarchy of all elementary particles – the Higgs boson (H). Notably, it is due to the interaction of elementary particles with the Higgs boson that these particles acquire their masses. Despite the Higgs boson's discovery in 2012, it is far premature to claim we fully understand its true purpose and properties. Moreover, its uniqueness has repeatedly drawn interest of scientists using this phenomenon as a potential gateway to new physics, e.g. within the scope of the Effective Field Theory (EFT).

One possible approach to probing Higgs boson properties, such as spin, spatial or charge conjugation (CP) symmetries, is studying how it interacts with other elementary particles, such as vector bosons. The measure of this interaction is expressed by means of the so called coupling constants. Thanks to the technological marvels of the late 20^{th} and 21^{st} centuries, such as the Large Hadron Collider (LHC) at CERN, we can now produce the Higgs boson in high-energy proton-proton (pp) collisions at an increasingly higher production rate than was possible since its discovery. This allows for excellent sensitivity to the measurements of the Higgs boson couplings and opens up the opportunity to explore potential anomalous discrepancies hinting for signs of new physics.

This thesis analyzes datasets collected by the Compact Muon Solenoid (CMS) detector during 2016-2018 period of data-taking. The targeted pp-collision signal events involve the Higgs boson decaying into a pair of electroweak W bosons. As such, both the Higgs boson production and decay vertices with vector bosons V (HVV) are studied for the presence of anomalous contribution. In particular, final state kinematic information and dedicated matrix element methods are employed to define discriminants sensitive to anomalous effects. Special attention is given to the anomalous CP-odd coupling, which, if nonzero, would indicate CP-violation. Nevertheless, all anomalous couplings accessible

in this phase space are studied and constrained in the fit on data.

The ultimate goal of this study is to improve previously obtained coupling constraints (where applicable) and provide competitive results as compared to analyses in Higgs boson channels with different final state topology. As a result of the dedicated group effort, the relevant analysis summary is currently in the approval process for publication. The work presented in this thesis also serves as a short overview of the author's parallel research studies concerning the reconstruction of physical objects common for a wider range of published analyses, work on the alignment of the CMS Tracker subdetector, and work on the Phase-2 CMS Tracker detector upgrade.

The organisation of this thesis is as follows. Chapter 1 provides an overview of the state-of-the-art knowledge of particle physics. In this chapter, the reader is reminded of the standard model of elementary particles and interactions (baseline theory), followed by sections dedicated to the application of the EFT, specifically in the scope of HVV interaction vertices. The CMS detector and its integral components and functionalities are discussed in Chapter 2. The first part of Chapter 3 is dedicated to the general description of Monte Carlo (MC) simulation of pp-collisions at LHC, including detector response. The other part describes both general and analysis-specific procedures of reconstructing physical objects, such as leptons or jets. Chapter 4 provides information about the CMS triggers used in the collection of data. Further in this chapter, dedicated sections describe procedures employed in estimating major background processes and provide details on the construction of signal templates used in the final data fit. The analysis strategy, requirements defining the signal region phase space, and the construction of multi-dimensional discriminants are explored in Chapter 5. The statistical approach of performing the profile likelihood scans in the process of obtaining constraints on anomalous Higgs boson couplings is described in Chapter 6. It also includes an exhaustive list of all systematic uncertainties assumed in the final fit and their impact on this study. The analysis results are summarised and discussed in Chapters 7 and 8, respectively. Constraints on anomalous HVV couplings obtained in this study are compared to the constraints acquired in the past as well as to results obtained in parallel analyses probing other Higgs boson decay channels. Appendix A of this thesis consists of a short overview of the CMS Tracker alignment efforts with a focus on evaluating detector's performance during the 2016-2018 period of data-taking. Finally, Appendix B summarises the author's contribution to the planned CMS Tracker Phase-2 upgrade, with a focus on the development of a testing procedure for the new generation of silicon-based sensors and modules.

CHAPTER

ONE

STANDARD MODEL OF ELEMENTARY PARTICLES AND BEYOND

The discovery of the Higgs Boson as announced by the CMS [1] and A Toroidal LHC ApparatuS (ATLAS) [2] collaborations on the 4th of July 2012 marked a crucial moment in the history of particle physics, enticing both the scientific community and the general public to seek more about the nature and structure of our universe. By the time of the discovery, the existence of a scalar spin-0 elementary particle, such as the case for the Higgs boson, had been long-predicted by the Brout-Englert-Higgs (BEH), i.e. a concept explaining the origin of the elementary particle masses. About that time, the Higgs boson had been considered a missing piece for the understanding of the state-of-the-art theory of the fundamental forces and the building blocks forming our universe, commonly referred to as the Standard Model (SM). The excitement from reaching this extraordinary milestone roughly 50 years after its theoretical prediction thrives till this very moment, multiplied even more by the time of writing this thesis when celebrations of the 10th anniversary of the Higgs boson discovery are ongoing.

To fully comprehend the importance of the Higgs boson discovery in the context of the SM, we will start with a brief introduction into the history of the discoveries of the elementary particles and their properties in Section 1.1. A particular focus is put on Higgs boson production processes and selected decay modes as this will play an important role in choosing the phase space suitable to study new phenomena elaborated in this thesis. Further understanding of interactions between the SM particles would not be possible without an appropriate mathematical framework to be briefly introduced in Section 1.2 and to be benefited from in Section 1.3 explaining the basic ideas behind the realization of fundamental forces. Section 1.4 is dedicated to the BEH mechanism and the Higgs boson's unique position in the SM hierarchy.

Since the discovery of the Higgs boson, the SM has been continuously challenged by the experimental particle physicists employing the most up-to-date analytic techniques and technologies as well as analyzing increasingly larger quantities of data collected by numerous experiments around the world. Many predictions of the SM features were proven, confirming the well known theories and occasionally hinting for possible effects beyond the scope of the SM (BSM). It is because of the effort of a countless number of people involved in particle physics across the globe that the SM is a continuously evolving theory supported by experimental results obtained with a state-of-the-art precision. It is also known to be an incomplete theory, namely not involving an extensive theory of

gravity neither considering the presence of dark matter particles (if the concept is right), explaining neutrino oscillations and others. There is plenty of possible extensions of the SM while explaining all of them would be beyond the scope of this thesis. In Section 1.5 we will repeat some of the ideas behind the EFT, in particular we focus on the minimal expansion of the SM in the frame of the so called Anomalous Coupling (AC) and Standard Model Effective Field Theory (SMEFT) which are particularly useful in understanding the Higgs boson production and decay modes studied in this analysis.

1.1 Journey through discoveries

The beginnings of the human search for the structure and mechanics of the known universe can be traced back to ancient Greek (around the 5th century BC) when the first atomic theory was formulated by Democritus (and possibly his tutor Leucippus). While the idea of the universe being formed by physically indivisible atoms surrounded by a "void" (an ancient predecessor of a vacuum) could serve as an inspiration for generations of philosophers to come, its formulation was hardly understood as a rigorous mathematical concept as we know it from the modern physics era. Instead, many consider the electron discovery by J. J. Thomson in 1897 to mark the beginnings of particle physics. In Thomson's experiment, cathode rays composed of electrons have their trajectories curved as they pass through an electric field which, based on the curvature of the observed trajectory, was a manifestation of the electron's negative electric charge $-e$. Here $e \approx 1.6 \times 10^{-19}$ C stands for an elementary electric charge measured a few years later in the Millikan oil-drop experiment [3]. Combining Millikan's and Thomsons' experiments, one could easily calculate an invariant mass of the electron. Its presently known value is $m_{\mathrm{e}} \approx 0.511$ MeV$/c^2$, where c is the speed of light in a vacuum. The mass of the particle is conventionally expressed in multiples of an electron volt (eV) unit, which is an equivalent of energy acquired by the electron passing through a potential difference of 1 volt. It is also convenient not to write "$1/c^2$" throughout this text, however it is important to keep this factor in mind whenever it is needed to express the particle mass in standard units, e.g. grams. In addition, we will be using electron volts (usually billions of them or GeV) as units for the particle's momentum while we will omit the factor "$1/c$" needed to recover the usual units. The discovery of the electron and its properties was just a beginning of the more-than-century-long journey full of remarkable findings forming SM as we know it today – see Figure 1.1 for the full table of currently known elementary particles and their properties. These and many other historical remarks or milestones mentioned in this text should only serve as a baseline for a much more complex journey through particle physics. An exhaustive list of events leading to important discoveries can be also reviewed in textbooks written by D. Griffiths [4] or T. Davídek, R. Leitner [5] among many others.

Shortly after the discovery of the electron, Thomson correctly predicted that electrons are constituents of atoms. Furthermore, Rutherford carried out his famous scattering experiment disclosing that otherwise electrically negative atoms have a small and heavy, positively charged center called a nucleus. In example, the lightest element of the periodic table, hydrogen H^1, is a bound state of the positively charged $(+e)$ nucleus, historically named the proton p, and one electron with an energy state determined by the electron shell principle. According to this principle, bound particles, such as electrons in atoms or protons in nuclei, can only occupy quantized (discrete) energy levels. The number of the particles in the same energy level is then restricted by the Pauli exclusion principle which states that two or more identical half-integer spin particles cannot coexist in the same

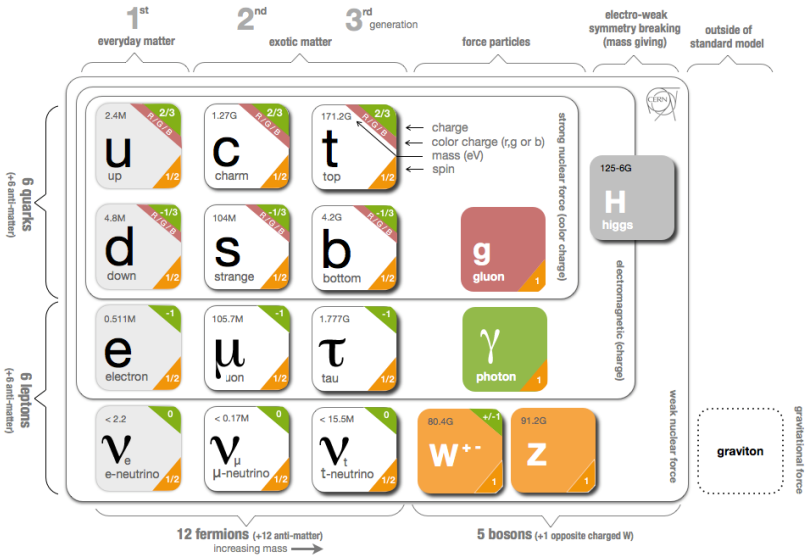

Figure 1.1: Elementary particles of matter and fundamental interactions within the Standard Model. Source: [6]

energy state. E.g., two electrons (both with spin $s = 1/2$) can occupy the same ground state of a helium atom only thanks to their different spin projection $m_s = +1/2$ and $m_s = -1/2$. Thus, spin is an important intrinsic property for all particles (such as their mass or electric charge, etc.) which existence is related to the principles discussed further in the text. As the spin can only be of a half-integer or integer value, we will be referring to the corresponding particles as fermions and bosons, respectively. In the following text we will be discussing (almost exclusively) particles of matter and their properties while implicitly keeping on mind their antimatter counterparts. While antiparticles bear exactly the same mass as the corresponding particle they differ by opposite charge (and/or other quantum numbers in general). Whenever needed, we will adopt a usual convention to depict antimatter particles with an emphasis on their opposite charge (positron – antiparticle of the electron, will be displayed as e^+) or by using symbols with a bar ($\bar{p}$ is used for an antiproton).

1.1.1 Elementary fermions

It was soon clear that atoms must consist of other particles as their mass is not corresponding to the number of protons in their nuclei. Dating back to 1932, Chadwick discovered an electrically neutral version of the proton, consequently renamed to the neutron [7]. Neutrons are only ≈ 1.3 MeV heavier than protons, with a proton mass equal to ≈ 938.3 MeV. Together with protons, they form all the nuclei of atoms and are complemented by electrons in their respective energy states. While electrons have no known internal structure, it was not until 1961 that Gell-Mann and Zweig predicted that protons and neutrons consist of even more elementary particles called "quarks". In particular, the proton was proposed to be a compound state of two so called "up"

quarks (or u) with electric charge $+2/3e$ and one "down" quark (or d) with electric charge $-1/3e$. Similarly to the proton, the neutron would be a compound state of two d quarks and one u quark. Figure 1.1 discloses there are four more "flavors" of quarks, namely "strange" (s), "charming" (c), "top" (t) and "bottom" (b), hinting for many more possible combinations of quark bound states – summarily called "hadrons".

Before we proceed to the discussion of general concepts behind hadron formation, let us focus on the first column of the SM table above, known as the first generation of elementary particles. It consists of all "stable matter" elementary building blocks of the universe – u and d quarks as they form atom nuclei and electrons as they are present in atom shells. The first column consists yet of another electrically neutral particle with spin 1/2 called "electron neutrino" ν_e. Not less important, it is its antimatter counter particle, the electron antineutrino, that soon played a crucial role in explaining experimentally confirmed continuous electron energy spectrum for the known β decay for some of the unstable atoms. Particularly, in β^- decay, one of the neutrons in the atom nucleus is converted to a proton while emitting an electron and antineutrino

$$n \to p + e^- + \bar{\nu}_e. \tag{1.1}$$

The presence of antineutrino assures the continuous energy spectrum of detected electrons (as a straightforward consequence of $1 \to 3$ decay kinematics), as well as the angular momentum is preserved in this process. Needed to say, the electron (anti)neutrino is extremely elusive particle, given its poor ability to interact with matter, and was only discovered in 1956 [8].

Considering only the low-energy scale processes, it would be possible to describe most of the known universe only by using the first generation of particles. Increasing this energy scale, for example by looking at high-energy collisions at particle accelerators, one can reveal further complexity. In fact, each of the four first-generation particles has two heavier siblings belonging to the second and third generation. For instance, muon μ is ≈ 200 times heavier version of electron while mass of tau lepton, τ, is significantly larger ($m_\tau \approx 1.7$ GeV). All three "charged leptons" (that is how electron, muon and tau lepton are commonly addressed) have the same spin 1/2 as well as an electric charge while they are only deviating in the aforementioned masses. Generally, the heavier the particle is, the more unstable it appears, i.e. is profound to decay. With a relatively long decay length of $c\tau \approx 659$ m (here τ stands for the lifetime of the particle), it was possible to detect the muon directly in a cloud chamber as demonstrated for the first time by Anderson and Neddermayer in 1937 [9]. The much more unstable τ lepton ($c\tau \approx 87$ μm) could only be discovered through its decay products as it was first time confirmed in several consequent experiments performed at SLAC laboratory in years after 1974 [10].

The first-generation electron neutrino ν_e was already mentioned with regards to the continuous electron spectrum observed in β decay. A few years later after its discovery, muon-flavored neutrino, ν_μ, was confirmed in an experiment carried out at the AGS accelerator [11]. Proton beams were interacting with a beryllium target, and unstable charged hadrons were created in this interaction, consequently decaying into muons and neutrinos. While muons from this decay were shielded, neutrinos were passing through the shielding material into a spark chamber, occasionally interacting with chamber electrodes which in majority of cases resulted in another muon in a final state and almost never an electron. This phenomenon could only be explained by the existence of a muon-flavored neutrino. Lastly, the tau neutrino, ν_τ, was discovered at the beginning of 21st century in a similar fixed-target experiment carried out at the Tevatron accelerator [12]. On purpose, we had not discussed the mass of neutrinos until all three neutrino flavors

were introduced. As a matter of fact, all three neutrino flavor states: ν_e, ν_μ and ν_τ are propagated as a quantum-mechanical linear superposition of their fundamental mass states (simply denoted as ν_1, ν_2 and ν_3) until their wave function is reduced into a single observable eigenstate thanks to the interaction. Moreover, an evidence was found that neutrino flavor can be transformed over long distances as demonstrated in experiments by the Super-Kamiokande Collaboration [13], the Sudbury Neutrino Observatory [14] and others. These observations can be explained by the phenomenon denoted as neutrino oscillation which directly implies that neutrinos are not massless particles as originally assumed in SM. Up-to-date, the exact values of neutrino masses are still unknown, however, based on recent measurements they can be limited to less than 1 eV.

Our list of elementary fermions would not be complete without the already mentioned three generations of quarks, ranging in mass from 2 MeV for the u quark up to as much as 173 GeV for the t quark making it the heaviest known elementary particle. Contrary to any of the previously listed elementary particles, quarks were not yet observed independently (thanks to the properties of the bounding strong force to be discussed later). In nature, quarks can only exist in a bound state of two or more quarks and/or antiquarks in accordance with the general Pauli principle applied for fermions. This is only possible if quarks posses another quantum number denoted as "color" (unrelated to the visible light) with three possible values, "red", "green" and "blue", while hadrons appear to be "colorless". The two main quark constellations are allowed:

- "mesons" (bound states of a quark and antiquark $q\bar{q}$) and
- "baryons" (bound states of three quarks qqq or three antiquarks $\bar{q}\bar{q}\bar{q}$).

In case of mesons, color and anticolor value is canceled out while for baryons each color is presented in the same amount, hence the association with otherwise unrelated color of the visible light. More exotic hadrons can exist as a combination of multiple mesons or baryons, e.g. "tetraquarks" $q\bar{q}q\bar{q}$ [15] and "pentaquarks" $qqqq\bar{q}$ [16], as confirmed by several discoveries announced by the LHCb detector (LHCb) experiment. For the brevity of this text, hadrons are formed by a limited number of "valence" quarks (see rules above) which are the only constituents contributing to the hadron quantum numbers, while they can also consist of an unrestricted number of "sea" quarks (virtual $q\bar{q}$ pairs) possibly undergoing hadronization processes which a short discussion is delayed for Chapter 3 in this thesis.

History of quark discoveries, though indirect due to their nature, roots back to 1968. In years that followed, several electron-proton scattering experiments performed by MIT-SLAC Collaboration [17, 18] disclosed that nucleons seem to possess an inner structure. Discovered scattering patterns were in favor of well known Gell-Mann and Zweig symmetry model [19, 20] which already assumed presence of a strange quark (together with up and down) and dictated rules for forming baryons and mesons. Discovery of the fourth quark, the charming c, is considered a milestone in particles physics, often referred to as "November revolution", as it helped to confirm the so called GIM mechanism [21]. Physicists Glashow, Iliopoulos and Maiani figured out that including the c quark into the theory explains the absence of certain processes in which quark flavor is exchanged, e.g. $s \to d$. Existence of the c quark is proven through discovery of $J/\psi = (c\bar{c})$ meson in two different experiments performed at Brookhaven National Laboratory [22] and SLAC [23] in 1974. Soon after in 1977, history was repeated and physicists at Fermilab observed new resonance $\Upsilon = (b\bar{b})$ [24] marking discovery of the fifth quark, the bottom b. Finally, discovery of the heaviest of quarks needed to wait until the top production became accessible via proton-antiproton collisions at the Tevatron accelerator (with a beam energy

of 900 GeV). The top quark t was consequently discovered by two independent Tevatron experiments, CDF [25] and D0 [26] in 1994.

1.1.2 Gauge bosons

In conclusion, twelve fundamental particles (plus an equal number of antiparticles) in three families (quarks, charged leptons and neutrinos) divided into three generations are forming the fermionic part of the SM table. All fermions can interact with each other while the particular form of interaction is a manifestation of four fundamental forces: electromagnetic force govern by the principles of Quantum Electrodynamics (QED), strong interactions driven by the Quantum Chromodynamics (QCD), weak force responsible for the nuclear fusion or β decay and gravity which effects on elementary particles can be efficiently neglected compared to other forces given the energy scale usually considered and the long distances gravity is acting on. Gravity interaction is believed to be mediated by the hypothetical (possibly massless) spin-2 tensor boson called "graviton". Confirmation of this theory is far beyond the experimental capabilities of the modern particle physics and will not be discussed in this text. The rest of the relevant interactions between fundamental particles could be imagined as an exchange of spin-1 gauge bosons. We recognize

- the massless photon γ acting as a messenger between electrically charged particles,
- the massless octet of gluons intermediating strong interaction "felt" only by quarks in hadrons (as gluons are changing color of quarks they exist in 8 color combinations),
- the heavy vector bosons $W^{\pm}$ ($m_W \approx 80$ GeV) and Z^0 ($m_Z \approx 91$ GeV) responsible for the weak charged and neutral currents felt by all fermions.

The true nature of these fundamental forces can be described in terms of Quantum Field Theory (QFT) as will be discussed in more detail in Section 1.3, however, a short induction to the history of gauge boson discoveries is in place.

The story of the photon can be traced back to 1905 when it was first time hypothesized by Albert Einstein and later used as an explanation to many experiments, e.g. to clarify the photoelectric effect. Nevertheless, it was not until 1924 when the photon got fully accepted after the Einstein's reaction on the Compton effect discovery [27], and that was through his comment on photons existing as single energy quanta. The history of W and Z boson discoveries on particle accelerators is much younger, dating to 1983. Eventhough the W was considered a necessary ingredient to explain β emission long time before that, it was UA1 [28] and UA2 [29] experiments at SPS proton-antiproton collider at CERN (with a beam energy of 270 GeV) that marked the W boson discovery in interactions of type

$$u + \bar{d} \to W^+ \to e^+/\mu^+ + \nu_e/\nu_\mu, \tag{1.2}$$

while the Z boson was found in processes such as

$$u + \bar{u} \to Z^0 \to e^+/\mu^+ + e^-/\mu^-. \tag{1.3}$$

Finally, similar in age, yet preceding the discovery of W and Z by four years, is the confirmation of gluon as an intermediating particle for strong interactions in the PETRA experiment in DESY (Germany). In particular, gluon was discovered directly through the "gluon Bremsstrahlung" process [30]

$$e^+ + e^- \to q + \bar{q} + g, \tag{1.4}$$

where the gluon is considered to be emitted from one of the final state quarks and consequently undergos the hadronisation process (turning into a measurable object called "jet" as further described in the text).

1.1.3 The Higgs boson

Last but not least, the Higgs boson H is the remaining and only fundamental scalar (spin-0) particle known to the SM. Existence of the scalar boson had been debated long time before its discovery as it would solve discrepancies arising during the construction of the SM theory, namely the problem of tree-unitarity preservation in the electroweak processes such as $e^+e^- \rightarrow W^+W^-$. Moreover, it is through the interaction with the Higgs field (see Section 1.4) that all other SM particles gain their masses (including massive intermediate vector bosons $W^\pm$ and Z). However, the mass of the Higgs boson itself is a free parameter of the SM and needs to be determined experimentally.

In 2012, CMS and ATLAS collaborations announced the discovery of the Higgs-like particle h (let us denote the possibility of the particle not being the SM Higgs boson by using a lower case letter whenever it is not clear by context). The new particle with a mass of 125.25 GeV (currently the most precise value [31]) and spin-0 was discovered by combining results from several decay channels. Among those, the most significant excess of signal events over background was observed in a phase space with two photons in the final state and with the candidate particle decaying into two Z bosons. Evidence for the Higgs-like boson was consequently fortified by independent analyses carried out by CMS and ATLAS collaborations in most of the accessible decay channels.

To fully understand the nature of the newly discovered boson and its properties, it is particularly important to study its interactions with the rest of the SM particles. Any deviation from the SM prediction both in the Higgs production mechanism and at the level of its decay rates could hint at the presence of new physics. In the following subsections we will introduce the most important SM decay channels as well as the processes driving Higgs boson production in proton-proton colliders which are relevant in the scope of this analysis. More general concepts of the mathematical framework used to describe these processes beyond the scope of the SM are summarised in Section 1.5.

Higgs boson decay modes

The SM Higgs boson branching ratio $BR(H \rightarrow XY)$ is plotted in Figure 1.2 (left) around the currently known value of the Higgs boson mass. It is defined as the fraction of events when the unstable particle, in this case the Higgs boson, decays within a particular decay mode XY and is related to the partial Higgs decay rate (decay witdth) $\Gamma(H \rightarrow XY)$ as follows:

$$BR(H \rightarrow XY) = \frac{\Gamma(H \rightarrow XY)}{\Gamma(H)}, \tag{1.5}$$

where $\Gamma(H)$ is a sum over all possible partial Higgs boson decay widths. In general, the decay width depends on the strength of the interaction between the mother particle and decay products as well on the general properties of the particles in question, such as the number of possible spin states and their mass. One can observe from Figure 1.2 (left) that the coupling (a measure of interaction strength) of the Higgs boson with fermions grows with their mass while other effects will play a role for interactions with bosons. An obvious exception applies for the massless particles such as gluons or photons where the higher order fermion-loop contributions need to be taken into account (the general

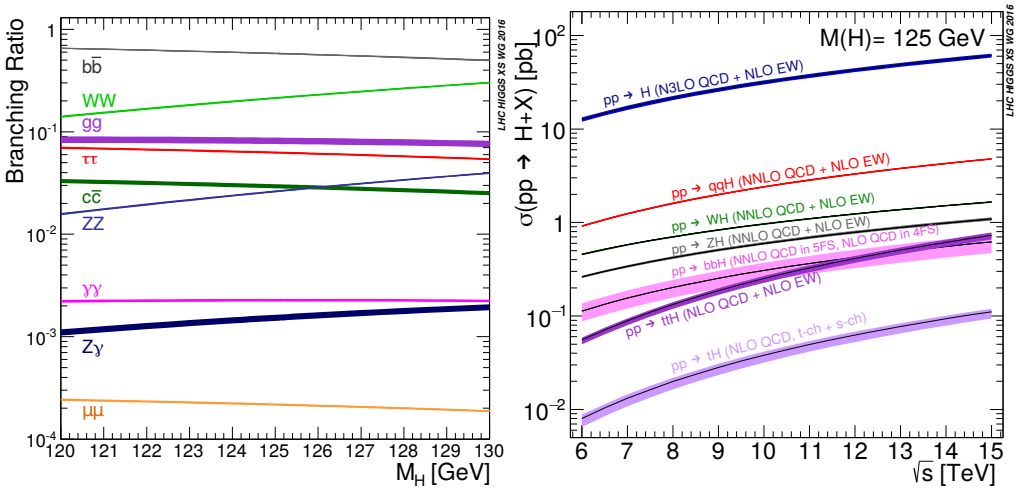

Figure 1.2: The SM Higgs boson decay branching ratios near the Higgs boson mass $m_H \approx 125$ GeV (left) and the SM Higgs boson production cross-sections as a function of the center-of-mass energy $\sqrt{s}$ (right) with VBF production denoted as qqH. The theoretical uncertainties on displayed quantities are expressed by the width of bands. Source [32].

concept of direct tree level and loop interactions will be further discussed in Section 1.2). As a matter of fact, the Higgs boson partial decay width in Yukawa type interaction with fermions $\Gamma(H \to ff)$ is proportional to the square of their mass. Therefore, the Higgs boson decays into the pair of bottom quarks is dominant (top quarks would be kinematically suppressed due to their large mass) compared to the possible decays into the pair of lighter fermions.

Second to the first in terms of branching ratio and the main focus of this thesis is the Higgs boson decay into the pair of W gauge bosons. This process would not be possible, assuming the rest mass of $m_W \approx 80$ GeV and applying conservation laws, unless at least one particle in this process is "off-shell". We use this term to describe the so called virtual particles that do not need to obey the relativistic energy-momentum equation

$$P^2 = E^2 - p^2 = m^2 \tag{1.6}$$

where $P \equiv (E, \vec{p})$ is a particle's four-vector momentum (four-momentum), E stands for a particle energy and p is the magnitude of its momentum vector $\vec{p}$. Such virtual particles, having the same mass m as their real counterparts, can only exist for a limited time needed to propagate into the observable final state particles. Whenever it is needed, the virtual nature of the particle is emphasized by using the superscript $*$ next to the usual symbol. In this analysis we only consider the case when the Higgs boson is defined "on-shell" (obeying Equation 1.6) leaving one of the W bosons necessarily off-shell.

As it was previously mentioned, the W bosons are heavy and therefore unstable. In about 2/3 of cases they decay hadronically, predominantly into charged pions $\pi^{\pm}$ ($u\bar{d}, d\bar{u}$) and kaons $K^{\pm}$ ($u\bar{s}, s\bar{u}$). Unfortunately, the hadronic final state is not significantly distinguishable from those observed in decays of other unstable particles. In the remaining $\approx 1/3$ of cases the $W^{\pm}$ boson decays leptonically into the pair of one charged lepton and the corresponding neutrino $l^{\pm}\nu_l$, which allows for a kinematically cleaner phase space.

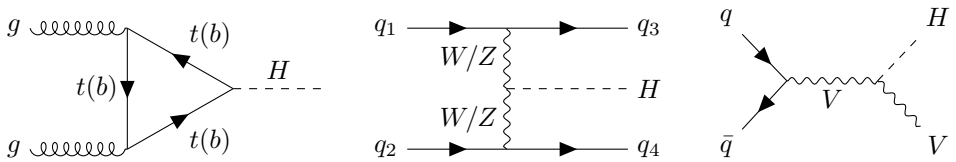

Figure 1.3: The SM Higgs boson production Feynman diagrams as they occur at the lowest order in the perturbation series. Displayed are: gluon-gluon fusion (left), VBF production (middle), quark-initiated VH production (right). Diagrams are shown in an ascending order from the left to right based on the production rate at the LHC.

Selecting events with both the W bosons decaying leptonically:

$$H \to WW^* \to 2l2\nu_l \tag{1.7}$$

we create a sufficiently stringent set of requirements to study this decay mode while reasonably suppressing possible background processes. Finally, we request that visible leptons in the final state form an electron-muon pair ($e\mu$) to create even cleaner phase space suitable for studying the delicate nature of the new physics phenomena.

Remaining decay modes mentioned in Figure 1.2 (left) are targeted by analyses beyond the scope of this thesis. It is, however, important to take into account possible final state topologies, such as the fully leptonic $H \to \tau\tau \to 2l2\nu_l 2\nu_\tau$ decay, that might be falsely identified as those occurring in our $H \to WW$ signal region. More about the particular choice of the studied phase space, selected decay mode and strategies employed to estimate possible background processes is discussed in Chapter 4 and 5.

Higgs boson production modes

Besides the dedicated $H \to WW^*$ decay mode, it is possible to study the strength of the Higgs boson interaction with vector bosons through some of the Higgs boson production mechanisms dominating at proton-proton colliders. It is also convenient to divide the studied phase space into the categories by targeting events originating in the particular Higgs production processes. This is achieved by using relevant kinematic information and will be discussed in Chapter 5. Following the SM Higgs production processes are ranked from the most common to the least abundant and will be considered in this analysis:

1. Gluon-gluon fusion (ggF),
2. Vector boson fusion (VBF),
3. Vector-boson-associated production (VH).

Gluon-gluon fusion process dominates thanks to the aforementioned tendency of the Higgs boson to couple with heavy SM particles. In this case, the Higss boson coupling to the pair of gluons (Hgg) is mediated through a loop of virtual heavy quarks, top quarks having the leading contribution, as depicted by the diagram in Figure 1.3 (left). In other words, while ggF process on its own (without considering further Higgs decay) does not allow to study Higgs interaction with vector bosons, it opens an opportunity to indirectly probe the Hgg coupling through Yukawa-type interaction with quarks.

VBF is the second most frequent production process followed by VH production as it is also expressed by means of the SM production cross-section (a measure of the probability

for the process to happen) in Figure 1.2 (right). The VBF process is often referred to as "weak boson Bremsstrahlung" given the obvious similarities with the original term. In this case, however, it is the virtual weak bosons (W or Z) that are radiated by quarks undergoing hard scattering processes in proton-proton collisions as depicted in Figure 1.3 (middle). On the other hand, the "Higgs Strahlung" term is used to describe the topology of the VH production as it would be the Higgs boson that is radiated from W/Z boson (see diagram in Figure 1.3 (right)). In both cases, the Higgs boson is directly coupled to the pair of vector bosons making these channels predominantly sensitive to the measurement of HWW/HZZ coupling.

Other processes such as $t\bar{t}$-associated Higgs production ($t\bar{t}$H) or gluon-initiated ZH production are significantly suppressed and will only be considered at the level of SM theory without accounting for any possible Beyond the Standard Model (BSM) contribution.

Since the discovery of the Higgs-like particle, several analyses were carried out aiming to confirm its SM properties while keeping an eye on any deviations from the SM theory. The particular focus was put on the so called spin (J) and parity (CP) intrinsic properties that will be explained in the context of quantum-mechanical wave-like representation of elementary particles in the following section. Several "exotic" scenarios with spin-1 or spin-2 representations were tested and rejected by CMS [33] and ATLAS [34], both collaboration presenting results in favor of the SM prediction of $J^{CP} = 0^{++}$ (defining properties of the scalar boson particle). While these results were compelling, they did not fully reject the possibility for a small ("anomalous") contribution to the Higgs boson coupling with gauge bosons HVV (V standing for W, Z, γ and gluons in general). By anomalous contribution we refer to both the entirely new BSM processes that are significantly suppressed by the scale of new physics and higher order loop corrections to the SM processes which are very small by definition (but could be enhanced by BSM effects). For the brevity of this text, we will commonly refer to both types of anomalous contributions as BSM unless we want to explicitly point out the origin of the contribution or its mechanism. Moreover, we will only assume processes corresponding to the 0-spin representation while allowing terms changing parity properties. More details about this minimal expansion of the SM Higgs boson can be reviewed in Section 1.5.3 while in the following text we still owe to the general concepts of the SM theory.

1.2 Brief introduction to the Quantum Field Theory

So far we have been discussing the elementary particles without paying too much attention to the suitable framework that would properly describe the nature of their existence and yield results compatible with experimental findings. The best theory used to successfully describe the SM is known as the QFT and some of the founding ideas will be discussed as follows.

QFT is based on both Quantum Mechanics (QM) in order to respect the wave-like properties of the particles and the Special Relativity (SPR) principles adopted to cope with the relativistic energies typical for particle physics experiments. In QM, free particles are inherently delocalised objects with a probability to occur at a particular space ($\vec{x}$) and time (t) described by a wave function $\psi(\vec{x}, t)$. In classical mechanics, any observable such as momentum, energy or coordinates, is a single function of the phase space. On the contrary, QM postulates that any relevant information needed to describe the par-

ticle state is contained by the corresponding wave function and can be obtained by the formalism of quantum-mechanical operators $\hat{A}$ associated with the measured observables A. Without going into further details (see any QM textbook for more information), any result of such a measurement will be one of the eigenvalues a satisfying the operator equation

$$\hat{A}\psi = a\psi. \tag{1.8}$$

The time evolution of the particle wave function is then governed by the Schrodinger equation which allows us to study the system dynamics. The following example displays the one-dimensional time-dependent Schrodinger equation

$$i\frac{\partial\psi(\vec{x},t)}{\partial t} = \hat{H}\psi(\vec{x},t) = -\frac{1}{2m}\frac{\partial^2\psi(\vec{x},t)}{\partial x^2} + \hat{V}\psi(\vec{x},t), \tag{1.9}$$

where the non-relativistic Hamiltonian operator $\hat{H} = \hat{T} + \hat{V}$ corresponding to the sum of the total kinetic (T) and potential (V) energy was assumed. Apparently, QM is a useful tool to describe the states of the non-relativistic particles that exist in a particular phase space with a certain probability. On the other hand, it is not comprised in QM how exactly the particles are born in this phase space and how they annihilate. Moreover, space coordinates and time enter the Schrodinger equation (Equation 1.9) under a different order of partial derivatives. This is a direct violation of Lorentz invariance, i.e. the requirement for physical observables to be invariant under Lorentz transformation between any two inertial reference frames (see any SPR textbook for more information). The SPR also tells us that energy can be turned into mass following the Einstein relation

$$E^2 = |\vec{p}|^2 + m^2, \tag{1.10}$$

while this is not possible in QM. In this regard, QFT represents a compromise between both theories by respecting their basic principles which will become more evident once this theory is fully constructed.

Owing to the fact that the best classical theory we posses is a field theory, e.g. Maxwell's equations of electrodynamics, it is reasonable to start building the QFT based on the classical field approach. The dynamics of a classical system with N degrees of freedom is summarised by the Lagrangian $L(q_i, \dot{q}_i)$ as a function of the generalised coordinated q_i ($i \in \{1, 2, ..., N\}$) and their time derivatives $\dot{q}_i$. The lagrangian in classical mechanics is by convention defined as

$$L = T - V, \tag{1.11}$$

with a minus sign not to be confused with a plus sign in the Hamiltonian. The time evolution of such a system between two end-points in time, t_1 and t_2, corresponding to the coordinates $q_i^{(1)} = q_i(t_1)$ and $q_i^{(2)} = q_i(t_2)$, is then described by the action $\mathcal{S}[q_i]$. The action is a numerical quantity (so called functional) defined as an integral of the Lagrangian over time:

$$\mathcal{S}[q_i] = \int_{t_1}^{t_2} L(q_i, \dot{q}_i, t)\, dt. \tag{1.12}$$

In principle, the action sums up information about all possible paths that the system can follow while choosing only one which is eventually implemented in nature and for which equations of motion exist. The natural choice is done under the assumption of the so called least action principle which advocates that a path q between endpoints $q_i^{(1)}$ and $q_i^{(2)}$ is chosen by the system in a way which leaves the action unchanged under a small first

order perturbation δq. This can be imagined in the famous analogy [35] with light (as a wave) always choosing the shortest trajectory between two endpoints. If light arrived at a particular point in a different time, its phase would also be shifted. The total light amplitude at this spot is then a sum over amplitudes corresponding to all possible paths that light could take. The sum serves as a smoothening factor for which all essentially different phases cancel out while paths with reasonably similar phases remain. Finally, the path preferred by the system is the one with many adjacent paths with similar phases. Applying this principle one can derive the Euler-Lagrange equations of motion (see the rigorous recipe in [36]) in a differential form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0. \tag{1.13}$$

So far we have been assuming a system with a discrete degree of freedom. That is in contrary to the field theory where any point defined by the space-time coordinates x^μ is assigned a value represented by a continuous field quantity $\phi_i(x^\mu)$ (index i is here reserved for a possible additional dependence on a discrete observable). A typical example of a continuous field quantity is the temperature of the system $T(x^\mu)$, however, vector values such as the magnetic field $\vec{B}(x^\mu)$, or more complicated tensor structures can be appointed. In this context, the Lagrangian function adopted to represent states of the discrete system can be replaced by the continuous Lagrangian density $\mathcal{L}(\phi_i, \partial_\mu \phi_i)$ which is now a function of the field itself and its space-time derivatives

$$\partial_\mu \phi_i \equiv \frac{\partial \phi_i(x^\mu)}{\partial x^\mu}. \tag{1.14}$$

Applying the same least action principle it can be shown (see [4] or any QFT textbook) that the Euler-Lagrange equations for the fields will acquire an equivalent form

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0, \tag{1.15}$$

which, similarly to the classical version, yields the field equations of motion once the particular form of the Lagrangian density is substituted.

In summary, the quantum-mechanical wave function representation of particles is replaced by the quantum fields satisfying their corresponding field equations. In other words, particles in QFT are understood as (energy) excitations of the fields that allow for particle creation and annihilation at any point in space and time (though this process still needs to be addressed in more details). For the sake of completeness, note that the actual form of the aforementioned field relations derived to target the QFT framework would also fit into any classical field theory textbook as we did not discuss any of their quantum-mechanical aspects. Formally, the field quantities ϕ_i introduced above should undergo a quantization procedure in which they become operators themselves in a similar manner as momentum or position observables will get their corresponding operators in QM. However, it is not necessary to discuss the operator-like properties of the fields as any important aspects of QFT required for the construction of the SM can be derived (or at least introduced) while keeping this as an underlying concept. Instead, we will focus on possible Lagrangian density forms that have proven to be useful in the description of the SM. As it is not beneficial to further differentiate between these two terms, henceforth we will refer to the Lagrangian density by using only the term Lagrangian.

1.2.1 Free particle Lagrangians

It takes considerable ingenuity to construct a Lagrangian as there is no straightforward analogy with kinetic and potential terms in the classical Equation 1.11. In QFT, the Lagrangian is understood as an axiomatic manifestation of the system which it is intended to describe and needs to be "found" in a form that accurately represents the nature around us. It would be unpractical (and we arguably do not observe it) if the universe behaved differently here and at the opposite side of the globe or 10 years ago compared to now. There is a natural set of global requirements emerging such as:

- invariance under translation and rotation in space,
- time translation invariance,
- and already mentioned Lorentz boost invariance,

that certainly need to be demanded for any real-world Lagrangian. This is summarised in the requirement for any field entering the Lagrangian to be transformable under the Poincare group. As a matter of fact, symmetries generated by the Poincare group are not satisfactory in the description of fundamental interactions, however this will be left for discussion in Section 1.3.

In practice, free particle Lagrangians (no interactions yet) were constructed to determine the dynamics for different field quantities such as scalars, vectors, etc., which are directly related to the particles known in the SM (also efforts were made to describe some more exotic particles). In this regard, spin-0 particles are defined as an excitation of the scalar field $\phi(x)$ and the corresponding Lagrangian acquires the form

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}m^2\phi^2. \tag{1.16}$$

where m stands for the particle mass and we employ the usual Einstein summation notation

$$\partial_0 \phi \equiv \partial^0 \phi \text{ and } \partial_k \phi \equiv -\partial^k \phi \tag{1.17}$$

with $0 \to t$ and $k = 1, 2, 3 \to x, y, z$. It can be shown that substituting the Lagrangian from Equation 1.16 into Equation 1.15 yields the (multiparticle) Klein-Gordon equation of motion

$$\partial_\mu \partial^\mu \phi + m^2 \phi = 0. \tag{1.18}$$

Contrary to the Schrodinger equation, the Klein-Gordon equation is conveniently found in its Lorentz invariant form (derivatives appear within the same order) confirming the relativistic nature of QFT. This will be true also for the rest of the equations of motion shown in this section.

Most of the elementary particles in the SM table are fermions. Analogous for spin-half (spin-$\frac{1}{2}$) particles, (Dirac) Lagrangian can be found as follows

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi, \tag{1.19}$$

where

$$\psi(x) = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix} \tag{1.20}$$

is a four-component complex spinor ($\psi_j(x) = a_j(x) + ib_j(x)$, $j = 1, 2, 3, 4$). The spinor is an element of a complex vector space, however, contrary to real vectors it has an interesting feature - it needs to be rotated by $2 \times 360°$ in order to get to its initial state. The fact it consists of four independent components means that γ^μ objects acting on spinors in Lagrangian 1.19 are represented by four so called Dirac matrices (one for each covariant index μ) with a dimension of 4×4. Dirac matrices are required to have specific properties under Hermitean conjugation (symbol †) so that the following is true

$$\gamma_0^\dagger = \gamma_0 \equiv \gamma^0, \gamma_k^\dagger = -\gamma_k \equiv \gamma^k \tag{1.21}$$

where we have assumed an arbitrary choice of the metric tensor $g_{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$ when lowering index for Dirac matrices using relation $\gamma_\mu = g_{\mu\nu}\gamma^\nu$. Using this notation it is also useful to introduce Dirac adjoint spinors

$$\bar{\psi} = \psi^\dagger \gamma_0. \tag{1.22}$$

It can be shown that applying the Lagrangian 1.19 into Equation 1.15 (assuming $\phi \to \bar{\psi}$ and $\partial_\mu \phi \to \partial_\mu \bar{\psi}$) results in the famous free-particle Dirac equation for fermions

$$i\partial\!\!\!/\psi - m\psi = 0 \tag{1.23}$$

where we have employed the Feynman "slash" notation $a\!\!\!/ \equiv a_\mu \gamma^\mu$. At this point we should mention that Dirac's motivation to derive this equation was originally driven by the problem of negative probability densities appearing in the non-relativistic Klein-Gordon equation (not to be confused with its relativistic version in Equation 1.18) and it was only by chance that it also provided a feasible description of spin-half particles.

Before we proceed, there is one particular aspect of this equation that helps us in understanding of the SM and deserves our attention. For that we need to have a look at explicit form of the solution for Dirac's equation. In particular, we consider two independent prescriptions in the form of definite momentum and energy plane waves

$$\psi_u(x) = u(p)e^{-ipx} \text{ and } \psi_v(x) = v(p)e^{ipx}, \tag{1.24}$$

where $u(p)$ and $v(p)$ are spinors defined in momentum space (their explicit form will not be important in the scope of this text). Substituting these solutions, one can get two forms of the "momentum" Dirac's equation:

$$(p\!\!\!/ - m)u(p) = 0 \tag{1.25}$$

and

$$(p\!\!\!/ + m)v(p) = 0. \tag{1.26}$$

It can be shown that the former corresponds to a positive energy solution $E = \sqrt{|\vec{p}|^2 + m^2}$ while the latter yields a solution with negative energy $E = -\sqrt{|\vec{p}|^2 + m^2}$. It is strangely beautiful that nature actually implements this negative energy solution and it is rather our point of view that needs to be changed. The usual interpretation employed by Feynman says that a nonphysical negative particle solution (propagating backwards in time) corresponds to a physical positive energy antiparticle solution with opposite charge (and propagating forward in time). Indeed, Dirac's equation served as a useful tool to explain both fermion and antifermion particles observed in experiments.

Finally, assuming the free particle Lagrangian for a vector field A_μ

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m^2 A^\nu A_\nu \tag{1.27}$$

leads to the derivation of the Proca equation of motion for massive spin-1 particles:

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0 \tag{1.28}$$

where $F^{\mu\nu}$ stands for a field-strength tensor

$$F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{1.29}$$

In the case of a massless vector boson, such is a case for photons $m_\gamma = 0$, the Proca equation becomes identical to the Maxwell equation for empty space. This observation can serve as a starting base for building the theory of QED.

The Equations 1.18, 1.23 and 1.28 were discussed in this text as they appropriately determine dynamics for the free particle spin-0, spin-half and spin-1 fields. On the other hand, the full SM picture would not be complete without understanding how particles interact between themselves. In this regard, the Lagrangians introduced in this section need to be upgraded for interaction terms. Before we fully divert our attention into this matter, let us first summarise how individual fields behave under certain discrete transformations such as parity and charge conjugation.

1.2.2 Spin-parity properties

As indicated in the previous section, the particular character of free-particle fields decides on spin value assigned to associated particles, e.g.

- scalars $\leftrightarrow$ spin-0,
- spinors $\leftrightarrow$ spin-half,
- vectors $\leftrightarrow$ spin-1,
- tensors $\leftrightarrow$ spin-2.

The pioneering idea of spin as an additional quantum property has arisen from the study of Dirac equation. It can be shown (see rigorous proof in [36]) that the total angular momentum $\vec{J}$ of the system described by the Dirac Lagrangian can only become a constant of the motion (conserved quantity) if its corresponding operator can be written as

$$\hat{\vec{J}} = \hat{\vec{L}} + \hat{\vec{S}} \tag{1.30}$$

where $\hat{L}^a, a = 1, 2, 3$ are components of the system's angular momentum, and $\hat{S}^a, a = 1, 2, 3$ stands for components of newly defined spin-angular momentum

$$\hat{S}^a = \frac{1}{2}\begin{pmatrix} \hat{\tau}^a & 0 \\ 0 & \hat{\tau}^a \end{pmatrix} \tag{1.31}$$

written in terms of 4×4 matrices with 2×2 block elements, including $\tau_a, a = 1, 2, 3$ representing Pauli matrices. With some effort it can be shown that the following is true for any Dirac spinor $\psi(x)$

$$\hat{\vec{S}}^2\psi = s\,(s+1)\,\psi = \frac{3}{4}\psi \tag{1.32}$$

where the corresponding spin for elementary particles associated with Dirac spinors is determined as $s = \frac{1}{2}$. This is an inspiring result and a great success of the Dirac equation. In general, matrices $\hat{S}^a$ are representing rotational generators of the Lorentz group, which differ for each irreducible representation of the group, residing in a particular space where field observables are constructed. Integer-spin values are then naturally obtained only for those field observables defined under the relevant group representation. In other words, spin should be viewed as a defining abstract quantity arising from the construction of field observables and reflecting their rotational symmetry.

Besides spin, field observables generally depend on the space-time coordinates. Parity $\mathcal{P}$ is another defining property which describes how field observables behave upon inversion of space coordinates, such as

$$x = (t, \vec{x}) \xrightarrow{\mathcal{P}} \tilde{x} = (t, -\vec{x}). \tag{1.33}$$

In the example with Dirac spinors, it can be shown that the corresponding parity operator is equal to γ_0 matrix, and Dirac spinors transform as

$$\psi(x) \xrightarrow{\mathcal{P}} \psi(\tilde{x}) = \hat{\mathcal{P}}\psi(x) = \gamma_0\psi(x). \tag{1.34}$$

Intrinsic parity for individual fermions is then defined as an eigenvalue (conserved quantity) of the parity operator acting on particles at rest, i.e. $x = \tilde{x} = (t, \vec{0})$. In momentum space, this translates to

$$\begin{aligned} \gamma_0 u(m, 0) &= +u(m, 0) \\ \gamma_0 v(m, 0) &= -v(m, 0) \end{aligned} \tag{1.35}$$

with intrinsic parity assigned to fermions (antifermions) equal to $+1(-1)$. Contrary to fermions, bosons and antibosons possess identical intrinsic parity. Needed to add, parity conservation in particle decay (or generally during interaction) is a subject often studied in high energy physics experiments. The nature of electromagnetic and strong currents, to be discussed in the text below, preserves parity, while parity is not conserved in weak interactions. This fact was first time experimentally confirmed in weak β decay [37] by observing that-time-unprecedented angular distribution of positrons coming from decay.

In addition to parity, let us define the charge conjugation operator $\hat{\mathcal{C}}$ that exchanges each particle for antiparticle and vice versa. In terms of the Dirac-Pauli representation, we can write

$$\psi(x) \xrightarrow{\mathcal{C}} \bar{\psi}(x) = \hat{\mathcal{C}}\psi(x) = i\gamma^2\psi^* \tag{1.36}$$

where ψ^* stands for the complex conjugate of the Dirac spinor. $\hat{\mathcal{C}}$ eigenvalues ("C-parity") are only defined for electrically neutral particles (or fully neutral systems of particles) as only those are truly identical to their antiparticle counterparts. An example of such a particle is the photon with intrinsic C-parity equal -1 (see rigorous proof in [36]).

Spin, space parity, and C-parity are often summarily denoted as spin-parity properties J^{CP}. Figuring out correct J^{CP} values for the Higgs boson is an important step in understanding the SM and a possible probe of BSM physics. Up to this date, "exotic" scenarios for the Higgs boson with non-zero integer-spin were successfully excluded in several Higgs decay channels. Nevertheless, a small deviation from SM $J^{CP} = 0^{++}$ prediction at the level of CP-parity properties was not yet excluded and served as a motivation for studies performed in this thesis and related publication [38]. In the following text, any observable (or operator) is considered CP-odd if it changes its sign under the combined $\mathcal{CP}$ transformation and is CP-even otherwise.

1.3 Fundamental interactions

In a universe without fundamental interactions quarks would not be confined in protons (no strong force), stars would not be capable of nuclear fusion (no weak force), planets would never be formed (no gravity), and this thesis would slip out of your hands as there would be no concept of friction (no electromagnetic force). Elementary particles undoubtedly experience fundamental forces (though, as previously stated, gravity will not be discussed), and this information should be inherently included in the Lagrangian of the system. It is also reasonable to assume that such a Lagrangian should exhibit certain symmetries which would account for the universality of physics they describe. Let us take the Dirac Lagrangian 1.19 as an example and suppose a global phase transformation of a complex Dirac spinor

$$\psi(x) \to \psi'(x) = e^{i\theta}\psi(x) \tag{1.37}$$

which represents a unitary transformation of the Abelian (commutative) group U(1). In other words, it is expected that a rotation of a Dirac spinor in the complex plane by a constant angle θ should not change the physics of the Dirac Lagrangian. It is important to realize that θ does not depend on space-time coordinates x_μ (therefore a "global" transformation), so derivatives in the Dirac Lagrangian will effectively ignore the factor of $e^{i\theta}$ leaving the original form of the Dirac Lagrangian unchanged

$$\mathcal{L}(\psi(x)) \to \mathcal{L}'(\psi'(x)) \equiv \mathcal{L}(\psi(x)). \tag{1.38}$$

It can be shown that the more general requirement of global invariance under a non-Abelian special unitary group SU(n) is fulfilled for any complex field of rank n. In this case, a unitary transformation is represented by $n \times n$ matrix with determinant equal to 1 (therefore a "special" group), and simple multiplication by a real value of θ is replaced for non-commutative matrix multiplication. In other words, the universality of physical law under constant rotation of fields arises naturally in the same way as there is a freedom to change the temperature scale from degrees of Celsius to Fahrenheit without altering the fact it is freezing outside. Henceforth, we will refer to this requirement with the term: global gauge invariance. Let us expand this general concept in the following section using an example which will naturally lead us into an introduction of QED.

1.3.1 Local gauge invariance and QED

Suppose even more stringent requirement of local gauge invariance for which the rotation angle $\theta(x)$ depends on space-time coordinates

$$\psi(x) \to \psi'(x) = e^{i\theta(x)}\psi(x). \tag{1.39}$$

Considering the same example, partial space-time derivatives in the Dirac Lagrangian cannot ignore the rotation angle anymore. As such, the free particle Dirac Lagrangian is not invariant under a local U(1) transformation, i.e.

$$\mathcal{L}(\psi(x)) \to \mathcal{L}'(\psi'(x)) = \mathcal{L}(\psi(x)) - \bar{\psi}\gamma^\mu(\partial_\mu\theta)\psi. \tag{1.40}$$

Fortunately, local gauge invariance can be restored by assuming an additional vector field $A_\mu(x)$ with certain gauge transformation properties,

$$A_\mu \to A'_\mu = A_\mu + \frac{1}{g}\,\partial_\mu\theta, \tag{1.41}$$

and by replacing partial derivatives for their covariant form $\mathcal{D}_\mu$, i.e.

$$\partial_\mu \to \mathcal{D}_\mu = \partial_\mu - igA_\mu. \tag{1.42}$$

Substituting this into the free particle Dirac Lagrangian results in a form

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^\mu A_\mu \psi \tag{1.43}$$

which is now invariant under a local gauge U(1) transformation. An extra term that is now present in the Lagrangian describes the interaction between Dirac fields associated with fermions and the newly assumed gauge vector field A_μ representing the photon exchanged through the neutral electromagnetic vector current $j^\mu = g\bar{\psi}\gamma^\mu\psi$. Constant $g = -q \equiv e$ can be recognised as an electromagnetic coupling equal to electron's charge.

Contrary to global gauge invariance, expecting that physics universality would hold also for any local phase transformation can be somehow less intuitive. In analogy with different temperature scales, this would be an equivalent of asking every country in the world to use its own innovative scale and regularly change it every month while expecting that meteorologists will still be coming up with the same conclusions about weather. It might be unconventional to ask for this, however, not impossible. As a consequence, this requirement comes with a price – in a real world example, it is an introduction of electromagnetic interaction between fermions and the photon that restores the balance. For the sake of completeness, the full version of QED Lagrangian is finalized by adding the familiar kinetic term for massless photon fields (which is already invariant by definition):

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^\mu A_\mu \psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{1.44}$$

1.3.2 Weak interactions

Building upon acquired experience with the Abelian case, let us have a closer look at the requirement of non-Abelian (non-commutative) gauge invariance under the SU(2) group. Suppose the same form of the free particle Dirac Lagrangian

$$\mathcal{L}_{e\nu_e} = i\bar{\Psi}\gamma^\mu \partial_\mu \Psi - m\bar{\Psi}\Psi, \tag{1.45}$$

however, Ψ is now representing a "doublet" of Dirac spinors, e.g. a doublet of electron and electron neutrino fields, which transforms as

$$\Psi(x) \equiv \begin{pmatrix} \nu_e(x) \\ e(x) \end{pmatrix} \to \Psi'(x) = \hat{U}\Psi(x). \tag{1.46}$$

employing a unitary matrix (with determinant equal to 1):

$$\hat{U} = \exp(i\theta^a T^a) \equiv \exp(i\theta^1 T^1 + i\theta^2 T^2 + i\theta^3 T^3) \tag{1.47}$$

where the "weak isospin" group generators $T^a = \frac{1}{2}\sigma^a$ can be identified as Pauli matrices $\sigma^a, a = 1, 2, 3$ and θ^a correspond to three rotation angles. In analogy with the U(1) example, suppose that $\theta^a(x)$ becomes dependent on space-time coordinates under a local SU(2) transformation. Similar to the previous case, the Lagrangian 1.45 requires the extra interaction terms in order to become invariant. Generalizing the concepts employed in the Abelian scenario, derivatives in the Dirac Lagrangian should be replaced for their covariant form

$$\partial_\mu \to \mathcal{D}_\mu = \partial_\mu - ig_W W_\mu \equiv \partial_\mu - ig_W W_\mu^a(x) T^a, \tag{1.48}$$

where the matrix W_μ is constructed from three new Yang-Mills vector fields $W_\mu^a, a = 1, 2, 3$. Required gauge transformation properties of these fields can be determined by explicitly writing out the covariant transformation for derivatives while applying it on an arbitrary function $f(x)$:

$$\begin{aligned}\mathcal{D}'_\mu f \equiv (\partial_\mu - ig_W W'_\mu)f &= \hat{U}\mathcal{D}_\mu \hat{U}^{-1} f \\ &= \hat{U}(\partial_\mu - ig_W W_\mu)\hat{U}^{-1} f \\ &= \hat{U}\,\partial_\mu(\hat{U}^{-1}f) - ig_W \hat{U} W_\mu \hat{U}^{-1} f \\ &= \hat{U}\,\partial_\mu(\hat{U}^{-1})f + \hat{U}\hat{U}^{-1}\,\partial_\mu f - ig_W \hat{U} W_\mu \hat{U}^{-1} f \\ &= (\partial_\mu - ig_W \hat{U} W_\mu \hat{U}^{-1} + \hat{U}\,\partial_\mu \hat{U}^{-1})f.\end{aligned} \tag{1.49}$$

From there, comparing the first and the last line yields the desired transformation properties:

$$W'_\mu = \hat{U} W_\mu \hat{U}^{-1} + \frac{i}{g_W}\hat{U}\,\partial_\mu \hat{U}^{-1}. \tag{1.50}$$

Recasting this formula into terms of matrix components in their finite form, however, is technically involved and can only be achieved under an assumption of gauge transformation within an infinitesimally small rotation angle $\epsilon^a(x)$. In this approach, the gauge transformation unitary matrix $\hat{U}$ is expanded in a Taylor series while neglecting terms of order $\mathcal{O}(\epsilon^a(x)^2)$ and higher:

$$\hat{U} = 1 + i\epsilon^a(x) T^a \tag{1.51}$$

Substituting $\hat{U}$ into formula 1.50 and after some manipulation (see [36] or [39] for details) we get transformation properties for individual Yang-Mills fields:

$$W_\mu^a \to W_\mu^{a'} = W_\mu^a - f^{abc}\epsilon^b W_\mu^c + \frac{1}{g_W}\,\partial_\mu \epsilon^a \tag{1.52}$$

where f^{abc} is the fully antisymmetric Levi-Civita symbol. Keeping this in mind, the Lagrangian 1.45 can now be written in its invariant form with the extra weak interaction term

$$\mathcal{L}_{e\nu_e} = i\bar{\Psi}\gamma^\mu\,\partial_\mu \Psi - m\bar{\Psi}\Psi + g_W \bar{\Psi}\gamma^\mu W_\mu \Psi. \tag{1.53}$$

It appears that the requirement for non-Abelian local gauge invariance initially proposed by C. N. Yang and R. Mills [40] in the early 1950s leads to the introduction of three weak currents $j_a^\mu = \frac{g_W}{2}\bar{\Psi}\gamma^\mu\sigma_a\Psi$ that couple to Yang-Mills fields with a coupling constant g_W. Contrary to QED, Yang-Mills fields cannot be directly associated with physical particle states as one would suppose. Nevertheless, there is experimental evidence for a weak charged current interaction in which the space parity is not preserved, e.g. in atomic β-decay. Fortunately, this can be achieved when assuming the proper chirality structure of fermion spinors.

Indeed, Dirac spinors ψ can be naturally decomposed into left- and right-handed chiral states such as $\psi = \psi_L + \psi_R$. Similar to spin, chirality is an intrinsic quantum property of particles. It is useful to imagine this concept as an example with Dirac spinor rotated by 360° in the complex plane. In this example, the Dirac spinor gets a complex phase of -1 (remember that spinors must be rotated by two times the full angle to get the initial phase of $+1$). If visualized on a regular analog clock, the small hand travels from 12 am (phase $+1$) to 6 pm (phase -1) in the conventional clockwise (right-handed) direction, however, nothing prevents us to design clocks with an anti-clockwise (left-handed) mechanism. In the same analogy, a Dirac spinor can have distinguishable

left- and right-handed components. On the other hand, while spin projection is assumed as a varying property of the same particle, particles with different chirality states should be understood as really distinct particles.

Considering the chirality of fermions, it can be shown that charged weak currents can only couple to left-handed chiral particle states and right-handed chiral antiparticle states (other possible combinations are equal to zero). Therefore, a doublet of Dirac spinors under SU(2) symmetry can only be formed by left-handed particles (right-handed antiparticles), while right-handed particles (left-handed antiparticles) are treated as singlets under SU(2). Weak charged current $j_{\pm}^{\mu}$ with properties as observed in the experiment, i.e. associated with real world physical $W^{\pm}$ bosons, can be found as a linear combination of weak currents j_1^{μ} and j_2^{μ}

$$j_{\pm}^{\mu} = \frac{1}{\sqrt{2}}(j_1^{\mu} \pm i j_2^{\mu}) = \frac{g_W}{2\sqrt{2}} \bar{\Psi}_L \gamma^{\mu} (\sigma_1 \pm i\sigma_2) \Psi_L. \tag{1.54}$$

Weak charged bosons $W^{\pm}$ are then defined as a linear combination of Yang-Mills fields

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}}(W_{\mu}^1 \mp i W_{\mu}^2). \tag{1.55}$$

It would be convenient for the remaining j_3^{μ} expression in Lagrangian 1.53 to form a weak neutral current which directly couples with the W_{μ}^3 field. However, this would be in contradiction with experimental findings for the neutral weak boson Z interacting with particles of both left-handed and right-handed chirality. This dispute finally led to the recognition of the late 1960s idea of Glashow [41], Weinberg [42] and Salam [43] (GWS) introducing the unified model of "electroweak" (EWK) interaction.

1.3.3 Electroweak unification

At low energies (below 246 GeV), both the electromagnetic and parity-violating weak interactions are acting as fully distinct forces. On the other hand, in a hot and dense environment of our early Universe (or in the experimental conditions achieved in high-energy particle collisions), both interactions seem to be manifested by a single unified force. This sudden change of "phase" at a certain energy threshold is a direct consequence of spontaneous symmetry breakdown which is left for further discussion in Section 1.4. Instead, we will focus on principles proposed by GWS model in which the fermion state chirality decomposition plays a crucial role.

GWS model of electroweak unification allows us to incorporate two neutral and two charged gauge fields which will be indirectly associated to the experimentally observed photon γ, Z boson and charged $W^{\pm}$ bosons. A natural choice for such a construction leads to the requirement of local gauge invariance under

$$\mathrm{SU}(2) \times \mathrm{U}(1) \tag{1.56}$$

symmetry with three Yang-Mills fields W_{μ}^a corresponding to the non-Abelian SU(2) group and a new neutral vector field B_{μ} corresponding to the Abelian group U(1), suggesting that B_{μ} transforms identically to Equation 1.41. The particular choice of U(1) group representation, however, is considered as a degree of freedom (as there is an infinite number of non-equivalent options). In fact, we can express the general form of the transformation properties for a multiplet of Dirac spinors

$$\Psi \to \Psi' = e^{iY\theta(x)}\Psi, \tag{1.57}$$

where Y stands for the "weak hypercharge" defined as an eigenvalue (conserved quantity) of U(1) group symmetry generators. Taking an example from Yang-Mills theory, employing fermion chirality decomposition and assigning different transformation properties to the left-handed ($\Psi_L = (\nu_L, e_L)^T$) and right-handed (e_R, ν_R) fermion components affects the way how covariant derivatives are acting on these states which eventually leads to the simultaneous formation of desired QED and weak currents. For the sake of simplicity, following example of (incomplete) GWS Lagrangian is representing the lepton sector of the electroweak interactions, while keeping in mind that the same could be applied also for quarks:

$$\begin{aligned}\mathcal{L}_{lepton} &= i\bar{\Psi}_L\gamma^\mu(\partial_\mu - ig_W W_\mu^a T^a - ig'_W Y_L B_\mu)\Psi_L \\ &+ i\bar{e}_R\gamma^\mu(\partial_\mu - ig'_W Y_R^e B_\mu)e_R \\ &+ i\bar{\nu}_R\gamma^\mu(\partial_\mu - ig'_W Y_R^\nu B_\mu)\nu_R.\end{aligned} \tag{1.58}$$

The weak hypercharge quantities Y_L, Y_R^e and Y_R^ν appearing in this formula are, in general, different real numbers with their true value essentially bonded by the principles of underlying symmetry and electric charge conservation (see [36] or [39] for details). In this regard, it can be shown that the relationship between the electromagnetic coupling e and two of the weak couplings g_W, g'_W is given by the expression

$$e = g_W \sin(\varphi_W) = g'_W \cos(\varphi_W), \tag{1.59}$$

where φ_W is a free parameter of the GWS model, often described as the "weak mixing angle". Eventually, the physical photon and Z boson states can be identified as a linear combination of fields B_μ and W_μ^3:

$$\begin{aligned}A_\mu &= W_\mu^3 \sin(\varphi_W) + B_\mu \cos(\varphi_W) \\ Z_\mu &= W_\mu^3 \cos(\varphi_W) - B_\mu \sin(\varphi_W).\end{aligned} \tag{1.60}$$

For the sake of completeness, the appropriate kinetic terms need to be added to the electroweak lagrangian to successfully reveal the Euler-Lagrangian equations for the gauge fields W_μ^a and B_μ. The kinetic term for the Abelian field is constructed from the field-strength tensor of a familiar form

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.61}$$

while Yang and Mills have proven that the same pattern cannot be applied in the non-Abelian case. Instead, they came with the idea of employing the covariant field-strength tensor

$$F_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g_W f^{abc} W_\mu^b W_\nu^c \tag{1.62}$$

which naturally leads to the desired gauge invariant kinetic term (as well as extra terms corresponding to the self-interactions of gauge bosons). Finally, the electroweak lagrangian of the SM lepton sector reads

$$\mathcal{L}_{lepton}^{EWK} = \mathcal{L}_{lepton} - \frac{1}{4}F_{\mu\nu}^a F^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \tag{1.63}$$

1.3.4 Strong interactions

The requirement of local gauge invariance under a particular symmetry group has proven its usefulness in the introduction of the gauge bosons governing both electromagnetic and

weak fundamental forces. The same analogy can be applied also for strong interactions with underlying symmetry principles driven by the SU(3) group. Required transformation properties for Dirac spinors are given by

$$\psi_c(x) \to \psi_c'(x) = \exp(i\theta^a T^a)\psi_c(x), \tag{1.64}$$

where $T^a = \lambda^a/2$ stands for eight independent generators of the SU(3) group represented by 3×3 Gell-Mann matrices λ^a, $a = 1, ..., 8$. Because of increased rank of group generators to $n = 3$, Dirac spinors must hold an additional information about quarks, denoted by index c and corresponding to already mentioned "color" charge. Color charge can assume three possible values which can be expressed by the three-component vectors

$$r = \begin{pmatrix}1\\0\\0\end{pmatrix}, g = \begin{pmatrix}0\\1\\0\end{pmatrix}, b = \begin{pmatrix}0\\0\\1\end{pmatrix} \tag{1.65}$$

conveniently resembling the red, green and blue color encoding in an (unrelated) optical sense. This implies that all quarks regardless of flavor (u, d, ...) can exist in three orthogonal color states (r, g, b). Furthermore, eight independent group generators will give a rise to eight new vector gauge fields G_μ^a associated with QCD gluons. In this regard, the covariant derivative takes the form

$$\mathcal{D}_\mu = \partial_\mu - ig_S G_\mu^a T^a \tag{1.66}$$

where g_S is the strong interaction coupling constant. Suppose that QCD gauge fields transform under the SU(3) as

$$G_\mu^a \to G_\mu^{a'} = G_\mu^a - f^{abc}\epsilon^b G_\mu^c + \frac{1}{g_S}\,\partial_\mu \epsilon^a \tag{1.67}$$

where we have assumed transformation with infinitesimally small rotation angles $\theta^a \to \epsilon^a$. Employing the covariant field-strength tensor $G_{\mu\nu}^a$,

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_S f^{abc} G_\mu^b G_\nu^c, \tag{1.68}$$

as it is forming the additional kinetic term, we can write the QCD Lagrangian in its invariant form

$$\mathcal{L}_{QCD} = i\bar{\psi}_c\gamma^\mu\,\partial_\mu\psi_c + g_S\bar{\psi}_c\gamma^\mu G_\mu^a T^a\psi_c - \frac{1}{4}G_{\mu\nu}^a G^{a\mu\nu}. \tag{1.69}$$

Note that the requirement of local gauge invariance has again led to the introduction of an extra interaction term between fermions and gluon fields. However, in analogy with the electric charge and QED, it is only quarks with a non-zero color charge (represented by the color part of a Dirac spinor) that can participate in strong interactions with gluons acting as mediators. To preserve color charge in strong processes, gluons also provide a necessary exchange mechanism between interacting particles by "taking" some of the color charge with them. This implies that gluons necessary possess a color, which could possibly explain why free-standing gluons and quarks are lacking in the Nature (phenomenon known as "color confinement"). Natural preference for color particle singlets, e.g. colorless hadrons, has been experimentally supported, however, there is no conceptual proof for this hypothesis up to this date.

In conclusion, fundamental interactions briefly explained in this and preceding chapters are inherently contained in the Lagrangian of the system after enforcing suitable

gauge transformation properties in accordance with Nature. On the other hand, we have purposely skipped any discussion of particle masses and omitted corresponding mass terms. Besides partially pragmatic reasons (mass terms were not needed for explanation of fundamental interactions), including terms of type

$$m\bar{\Psi}\Psi \xrightarrow{\text{ch}} m(\bar{\Psi}_L\Psi_R + \bar{\Psi}_R\Psi_L) \tag{1.70}$$

apparently leads to problems with gauge invariance when accounting for the chirality decomposition of Dirac spinors. Given that both quarks and leptons are affected by electroweak interactions makes this a major problem, which only got resolved through introduction of spontaneous symmetry breakdown further discussed in Section 1.4. Before we proceed, let us yet make use of the known interaction Lagrangian forms in order to exploit their potential in a practical sense of calculating experimentally interesting quantities.

1.3.5 Feynman diagrams

In Section 1.2 we have introduced the action $\mathcal{S}$ as a quantity describing the time evolution for any process in the system. As an example, imagine a particle a propagating from initial state i to final state f. In QM, such a process can be assigned a probability amplitude $\mathcal{A}_{fi}$ proportional to $e^{i\mathcal{S}/\hbar}$, where $\hbar$ is a Planck's constant with the same dimension as action. In order to find the probability

$$\mathcal{P}_{fi} = |\mathcal{A}_{fi}|^2 \tag{1.71}$$

for this process to occur, one needs to calculate the amplitude for each feasible path between states i and f. This happens to be a tedious task that is usually addressed by employing perturbation theory where any interaction between particles is viewed as a small "perturbation" to the free particle state. In the first approximation, particle a propagates between states i and f as it would be scattering in a static potential invoked by other particles. A problem occurs when particle a is allowed to interact with other (virtual) particles on the path from state i to f. Perturbation series are expanded for new terms when particle a assumes intermediate state j. The more interactions are allowed the more terms in the perturbation series (higher orders) need to be evaluated. This approach might serve as a sufficient approximation in the classical picture of interactions, however it assumes that particles exchange their momenta instantly regardless of distance and without any mediator which eventually leads to causality violation. As shown in the previous section, this issue should be rather approached from the QFT point of view by requesting local gauge invariance which naturally gives rise to gauge fields acting as messengers between interacting particles. The remaining question is how to fully exploit this information once we are left with an invariant form of the Lagrangian.

Fortunately, the QFT offers an elegant framework dealing with the perturbation series developed by Richard Feynman. The set of "Feynman rules" involves a practical procedure how to read any Lagrangian term and transform it into the corresponding factor entering the formula for the invariant transition matrix element $\mathcal{M}_{fi}$ (equivalent to probability amplitude A_{fi}). This is achieved without any need to explicitly perform space-time integration of the Lagrangian each time there is a new process to consider. Moreover, Feynman found a way to assign each physical process a diagram that is both functional in prescribing how $\mathcal{M}_{fi}$ should look like as well as efficient in visualizing individual processes as they occur in time.

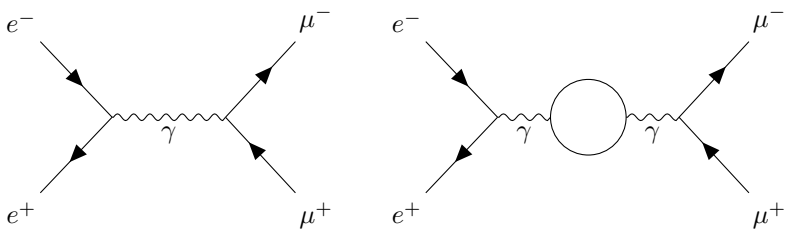

Figure 1.4: Feynman diagram for the LO QED annihilation process $e^+e^- \to \mu^+\mu^-$ (left) and an example of the NLO (1-loop) diagram contributing to the same process (right).

Some examples of Feynman diagrams were already displayed in Figure 1.3. Any line in these drawings corresponds to a certain particle (type of line also differentiates whether it is scalar, spinor or vector), while vertices are interpreted as interaction points. Each vertex is assigned a unique coupling factor interpreted as the strength of interaction between particles (lines) entering and leaving this vertex. The more interaction points, the higher the order of that diagram, and more difficult it is to calculate the full matrix element. Contrary to the Leading Order (LO) diagrams (sometime denoted as tree-level), higher order diagrams can be recognised by the presence of one or more loops – see example (1.4 (right) with extra interaction vertices forming a loop made of fermions. Henceforth, higher order diagrams in the SM will be commonly denoted as Next-to-Leading Order (NLO) in case they are limited to one-loop contributions, Next-to-NLO (NNLO) in case of two-loops contributions, etc. depending on the level of added complexity (3NLO and more). Let us stress out that coupling factors are multiplicative. In some cases, their value is very small, making processes which only arise with multiple loops very rare to occur.

It can be shown that the only ingredient needed to obtain the concrete form of SM coupling factor is contained by the corresponding Lagrangian interaction term. For simplicity, imagine the tree-level QED process $e^+e^- \to \mu^+\mu^-$ 1.4 (left) with two vertices of type $V_{\gamma ee} = V_{\gamma\mu\mu} = V_{\gamma \mathrm{ll}}$. Assuming the interaction term from the Lagrangian 1.44

$$i\mathcal{L}^{int}_{QED} = ig\bar{\psi}\gamma^\mu A_\mu \psi, \tag{1.72}$$

the corresponding QED vertex can be immediately recognized as

$$V_{\gamma \mathrm{ll}} = ig\gamma^\mu, \tag{1.73}$$

with the QED coupling equal to the unit of electric charge $g = e$. Another rule assigns the momentum-space spinors $u(\bar{v})$ to the fermions (anti-fermions) coming into the vertex and their corresponding adjoint version $\bar{u}(v)$ to the outcoming fermions (anti-fermions). The remaining internal line will get the specific factor called "propagator". Propagators can be conveniently identified by looking at the free-particle (momentum-space) field equations. In the case of a massless photon with momentum p and after imposing the Lorentz gauge $\partial_\mu A^\mu = 0$ as a freedom of choice, Proca equation 1.28 becomes

$$-p^2 g_{\mu\nu} A^\nu = 0 \tag{1.74}$$

and the propagator (inversed term standing next to A^ν in the formula) reads

$$P_\gamma = \frac{-ig_{\mu\nu}}{p^2}. \tag{1.75}$$

Finally, the formula for the $\mathcal{M}_{fi}$ element can be written as a product of all individual factors.

In general, there can be multiple Feynman diagrams contributing to the considered process and different spin polarisations for the involved particles should also be considered. This thesis, however, does not aim to provide an exhaustive list of (B)SM coupling factors, propagators and other necessary ingredients for the calculation of matrix elements. Nevertheless, it is useful to understand basic concepts as we will often refer to the theoretical values for the cross-section or decay width quantities per particular processes under study. As it is shortly summarised in the following section, the proper form of the invariant matrix element is the necessary ingredient for their calculation.

1.3.6 Useful applications of QFT

Suppose that an unstable particle of weight M decays into n daughter particles via a process defined by matrix element $\mathcal{M}_{fi}$. Corresponding element of the partial decay width is given by formula (originating in Fermi's Golden rule)

$$\mathrm{d}\Gamma = \frac{1}{2M} \left|\mathcal{M}_{fi}\right|^2 \mathrm{dLIPS}_n, \tag{1.76}$$

where dLIPS_n is an element of the Lorentz Invariant Phase Space volume [36]. In the example of a common two-body decay calculated in the center-of-mass frame of the mother particle, dLIPS_2 is reduced to the form

$$\mathrm{dLIPS}_2 = \frac{1}{16\pi^2} \frac{p_{\mathrm{cms}}}{M} \,\mathrm{d}\cos\theta \,\mathrm{d}\phi. \tag{1.77}$$

Trivial dependence on the polar θ and azimuth ϕ angles implies there is a priori no preferred flight direction for the decaying particles. Nevertheless, the possible angular dependence of the transition matrix element $\mathcal{M}_{fi}$, induced by particles with non-zero spin, can eventually yield a non-trivial result upon integration. In that regard, kinematic and angular properties of the particles involved are playing a significant role in the decay width calculation.

Cross-section term σ originally refers to the mind experiment in which a single particle of type a flies into a fixed target made of particles of type b, placed in the plane perpendicular to the direction of flight. The interaction cross-section between particle a and the fixed target is then given by the area S hit by this particle. Cross-section is measured in units of (femto)barn where

$$1\mathrm{fb} = 10^{-15}\mathrm{b} = 10^{-43}\mathrm{m}^2. \tag{1.78}$$

In experiments on particle accelerators, a bunch of particles a with velocity $\beta_a = \left|\vec{p}_a\right| / E_a$ is in the collision course with another bunch of particles b traveling in the opposite direction with velocity $\beta_b = \left|\vec{p}_b\right| / E_b$. In this case, the number of interactions per time (or rate) r in a unit of volume V is proportional to the interaction cross-section by relation

$$r = \frac{(\beta_a + \beta_b)}{V} \sigma. \tag{1.79}$$

Suppose the scattering experiment $a + b \to a' + b'$ in the center-of-mass frame $\vec{p}_a = -\vec{p}_b = \vec{p}_{\mathrm{cms}}$. Assuming relation 1.79 while recasting the transition interaction rate by employing Fermi's Golden rule yields a formula for the differential cross-section

$$\mathrm{d}\sigma = \frac{1}{4 p_{\mathrm{cms}} \sqrt{s}} \left|\mathcal{M}_{fi}\right|^2 \mathrm{dLIPS}_2, \tag{1.80}$$

where we have denoted a center-of-mass energy $\sqrt{s} = E_a^{\text{cms}} + E_b^{\text{cms}}$. The total cross-section σ is then obtained via integration over possible angular/kinematic dependencies, similar to the decay width calculation.

In general, both the decay width and cross-section can be determined with increasingly better precision by including higher order diagrams entering the formula for the transition matrix element. On the other hand, in experiments on proton-proton colliders, it is eventually the partons (quarks and gluons) inside of protons with a priori unknown kinematic properties that undergo hard scattering processes. Because of this missing information, scientists needed to come up with other approximative techniques that ultimately lead to a satisfactory solution. More about issues related to the modeling of proton-proton interactions is discussed in Section 3.1.

1.4 Brout-Englert-Higgs mechanism

The invariant mass (also "rest mass") is an intrinsic defining property for all SM particles. In some special cases, the invariant mass is zero, as it is for photons which allows them to travel at the speed of light. Gluons are also massless, nevertheless, they are always (strictly speaking) found confined in their compound states with quarks. For most of the known particles, the invariant mass assumes non-zero values in accordance with special relativity. In the previous section, we have shown how the nature of the fundamental interactions is encoded in the Lagrangian of the system upon requesting local gauge invariance. In a similar manner, information about particle masses is included in the corresponding terms entering the Lagrangian. On the other hand, if introduced on an ad-hoc basis, mass terms can undesirably violate the Lagrangian's symmetry. A typical example is the non-invariant mass term corresponding to a massive vector boson appearing in Lagrangian 1.27 or a similar issue with fermion masses depicted by formula 1.70.

In the following chapter, we will briefly describe the Brout-Englert-Higgs mechanism (BEH) which ultimately led into the generation of particle masses via the ingenious introduction of a new scalar field and its interactions with the SM particles. The BEH mechanism is based on the Goldstone's model [44] of a Lagrangian in the context of spontaneous symmetry breakdown [45] – an idea later independently studied by several groups of scientists, such as F. Englert and R. Brout [46], P. Higgs [47] or G. S. Guralnik, C. R. Hagen and T. W. Kibble [48] and is further discussed in Section 1.4.1. An explicit application of this theory in the scope of an Abelian symmetry group was originally elaborated by P. Higgs [49, 50] and further expanded to the non-Abelian case by T. W. Kibble [51]. A brief summary of their work is presented in Sections 1.4.2 – 1.4.3 with a focus on the generation of vector boson masses and their interaction with the Higgs boson. For the sake of completeness, the Yukawa-type of coupling [42] between the Higgs boson and leptons/quarks is mentioned in Section 1.4.4.

1.4.1 Goldestone's model

The Goldstone's type of Lagrangian defined for a complex scalar field ϕ (and its complex conjugate ϕ^*) can be written as

$$\mathcal{L}_{\text{Goldstone}} = \partial_\mu \phi \, \partial^\mu \phi^* + \mu^2 \phi \phi^* - \lambda (\phi \phi^*)^2 \tag{1.81}$$

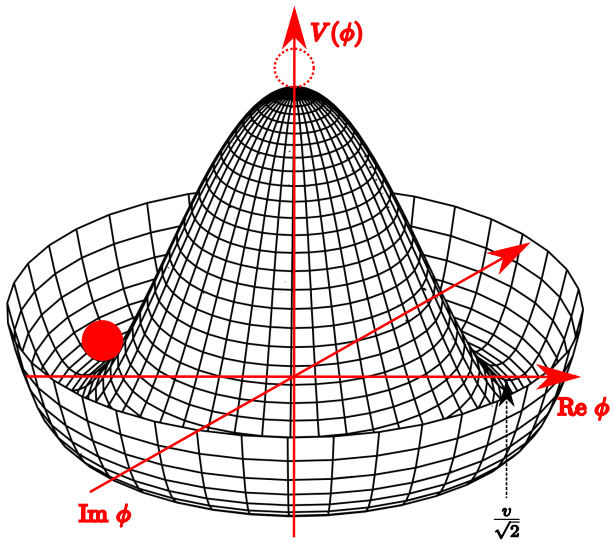

Figure 1.5: Goldstone potential visualized in the complex plane based on the description given by Equation 1.82.

where both μ and λ are real positive parameters, the former with a dimension of mass and the latter representing a dimensionless coupling constant. The last two terms

$$V(\phi) = -\mu^2\phi\phi^* + \lambda(\phi\phi^*)^2 \tag{1.82}$$

are often referred to as the Goldstone potential. A peculiar thing about this formula is the opposite sign appearing in front of the mass term (compared to Lagrangian 1.16), preventing us from the usual free-particle Lagrangian interpretation. For a better understanding of underlying concepts, the Goldstone potential (z-axis) is visualized in the complex plane with the real (imaginary) part of the complex scalar field profiled to the x-axis (y-axis) in Figure 1.5. After some investigation, it becomes apparent that the energy density described by $V(\phi)$ acquires a local maximum for $\phi_1 = 0$. However, this extremum is unstable, i.e. any infinitesimally small deviation $\delta\phi$ from this point leads to values $V(\phi_1 + \delta\phi)$ smaller than $V(\phi_1)$. Another extremum is found for the constant set of values $\phi_0 = \frac{v}{\sqrt{2}}\exp(i\theta)$ parametrised by an arbitrary real number θ. It represents a continuously degenerate spectrum of ground states (global minima) located on a complex plane circle with radius $v/\sqrt{2}$, where

$$v = \frac{\mu}{\sqrt{\lambda}} \tag{1.83}$$

was metaphorically named "vacuum" (from the state with the lowest energy). The observation that the ground state of the system, described by the Goldstone potential, is given by non-zero continuous value of ϕ_0 offers an idea to probe an alternative point of view: The complex scalar field can be "shifted" by a magnitude of ϕ_0. To further elaborate on this statement, let us recast the original complex field $\phi(x)$ in terms of real "polar" field variables $\sigma(x)$ and $\pi(x)$ by writing

$$\phi(x) = \frac{(\sigma(x) + v)}{\sqrt{2}}\exp\left(i\frac{\pi(x)}{v}\right) \tag{1.84}$$

where we have explicitly shifted the radial part $\sigma(x)$ by the non-zero vacuum value v. It can be shown (see [39] for details) that substituting relation 1.84 into Lagrangian 1.81 yields

$$\mathcal{L}_{\text{Goldstone}} = \frac{1}{2}\,\partial_\mu\sigma\,\partial^\mu\sigma + \frac{1}{2}\,\partial_\mu\pi\,\partial^\mu\pi - \lambda v^2\sigma^2 + \mathcal{L}_{\text{int}} \tag{1.85}$$

with $\mathcal{L}_{\text{int}}$ covering for all interaction terms cubic and quartic in σ and π that are not important for this discussion. Goldstone's Lagrangian now describes a one real scalar field σ with a mass term showing proper sign and which yields $m_\sigma = \sqrt{2\lambda v^2} = \sqrt{2}\mu$. Furthermore, it describes a one massless $m_\pi = 0$ real scalar field excitation π denoted as a Goldstone boson.

None of this was obvious from the initial form of Goldstone's Lagrangian 1.81 before we performed the shift from the vacuum value. In other words, the existence of a massless Goldstone boson is a manifestation of the broken symmetry of Goldstone's Lagrangian which originates in the presence of a non-invariant degenerate ground state in otherwise globally invariant Lagrangian. The whole term usually used for such a symmetry scheme – "spontaneous symmetry breakdown" – refers to the fact that system spontaneously chooses the lowest energy state under arbitrarily small perturbation. In general, the Goldstone model is a necessary ingredient to study new massive scalar field while the true meaning of the new-appearing Goldstone boson will be revealed through the familiar requirement of local gauge invariance under a particular symmetry group.

1.4.2 Abelian BEH mechanism

Ingenuity of the BEH mechanism in its simplistic form can already be exposed by requesting local gauge invariance under Abelian group U(1) within the scope of Goldstone's Lagrangian. Similar to the formal procedure enlisted in Section 1.3.1, ordinary partial derivatives are exchanged for their covariant form giving rise to the new Abelian gauge field A_μ. Assuming the usual transformation properties expected for an Abelian symmetry case while employing parametrisation 1.84 we can summarily write

$$\begin{aligned} \sigma(x) \to \sigma'(x) &= \sigma(x), \\ \pi(x) \to \pi'(x) &= \pi(x) + v\theta(x), \\ A_\mu(x) \to A'_\mu(x) &= A_\mu(x) + \frac{1}{g}\,\partial_\mu\theta(x). \end{aligned} \tag{1.86}$$

Aforementioned set of transformation properties shows a certain degree of freedom such as it is possible to choose $\theta(x) = -\pi(x)/v$ without changing the physical nature of the original Lagrangian. Transformation properties will now explicitly read

$$\begin{aligned} \sigma(x) \to \sigma^{(U)}(x) &= \sigma(x), \\ \pi(x) \to \pi^{(U)}(x) &= 0, \\ A_\mu(x) \to A^{(U)}_\mu(x) &= A_\mu(x) - \frac{1}{gv}\,\partial_\mu\pi(x). \end{aligned} \tag{1.87}$$

where transformed $\pi^{(U)}(x)$ boson field has fully vanished. In other words, the Goldstone boson has become nonphysical in the scope of proper gauge choice, commonly denoted as U-gauge. In that regard, the Higgs-like adaptation of the Goldstone's Lagrangian, introducing interaction with an Abelian field and employing U-gauge transformation, can be written as

$$\mathcal{L}^{(U)}_{\text{Higgs}} = \frac{1}{2}\,\partial_\mu\sigma\,\partial^\mu\sigma - \lambda v^2\sigma^2 - \frac{1}{4}F^{(U)}_{\mu\nu}F^{(U)\mu\nu} + \frac{1}{2}g^2v^2A^{(U)}_\mu A^{(U)\mu} + \mathcal{L}^{(U)}_{\text{int}} \tag{1.88}$$

where we have employed the usual relation for the field-strength tensor $F_{\mu\nu}^{(U)} = \partial_\mu A_\nu^{(U)} - \partial_\nu A_\mu^{(U)}$ and denoted $\mathcal{L}_{\text{int}}^{(U)}$ for all remaining interaction terms between the massive scalar field σ and the Abelian vector field $A_\mu^{(U)}$ (including self-interactions). Notice, that thanks to the BEH mechanism described in this section, the vector field $A_\mu^{(U)}$ now possesses a new mass term such as one can readily recognise

$$m_A^{(U)} = gv. \tag{1.89}$$

In conclusion, two real scalar fields were introduced in the scope of Goldstone's model, one of which is massless in the physical spectrum. BEH mechanism describes a procedure in which originally massless Goldstone boson disappears for the price of gauge field gaining mass. Originally massive scalar boson remains in the physical spectrum and is non-trivially coupled to the gauge vector boson.

1.4.3 Non-Abelian BEH mechanism

Idea of the BEH mechanism may be expanded to the non-Abelian case, originally studied by T. W. Kibble and subsequently applied by S. Weinberg and A. Salam in a context of EWK unification. In analogy with Section 1.3.3, Goldstone's Lagrangian can be recast as

$$\mathcal{L}_{\text{Goldstone}} = \partial_\mu \Phi^\dagger \, \partial^\mu \Phi + \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 \tag{1.90}$$

where

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.91}$$

is a doublet of complex scalar fields ϕ^+ and ϕ^0. Note that the shape of Goldstone potential remains the same, leaving the previous discussion about energy density extrema essentially unchanged. Similar to the Abelian case, $\Phi(x)$ can be written employing parametrisation by "polar" real scalar field variables

$$\Phi(x) = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(\sigma(x) + v) \end{pmatrix} \exp\left(i \frac{\pi^a(x) \tau^a}{v} \right) \tag{1.92}$$

with the massive field $\sigma(x)$ shifted by the vacuum value v; $\pi^a, a = 1, 2, 3$ associated with three massless Goldstone bosons and τ^a corresponding to Pauli matrices. Considering the U-gauge choice, the scalar doublet form is simplified to

$$\Phi^{(U)}(x) = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(H(x) + v) \end{pmatrix} \tag{1.93}$$

where we have explicitly denoted a more suitable symbol for the massive scalar field $\sigma(x) \to H(x)$ and will refer to its excitations by using the name "Higgs boson". Following the usual procedure of obtaining a gauge invariant Lagrangian, ordinary derivatives are exchanged for their covariant form and gauge invariance is restored through interaction of the Higgs field with familiar Yang-Mills gauge fields corresponding to the SU(2) and U(1) symmetry. Without further elaboration on technical details (see [39] for more), the preceding discussion leads to the formation of the Higgs Lagrangian in the scope of $SU(2) \times U(1)$ symmetry:

$$\begin{aligned} \mathcal{L}_{\text{Higgs}}^{(U)} = & \frac{1}{2} \partial_\mu H \, \partial^\mu H - \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4} \lambda H^4 \\ & + \frac{1}{8} (v + H)^2 \left[2 g^2 W_\mu^- W^{+\mu} + (g^2 + g'^2) Z_\mu Z^\mu \right] + \mathcal{L}_{\text{gauge}}^{(U)} \end{aligned} \tag{1.94}$$

where formulae 1.55 and 1.60 were used to express Higgs boson interaction with physical fields $W_\mu^\pm$, Z_μ associated with EWK gauge bosons and g, g' are corresponding coupling constants related via Equation 1.59 using the weak mixing angle φ_W. $\mathcal{L}_{\text{gauge}}^{(U)}$ term covers for all the remaining gauge field self-interactions and kinetic terms already known from the previous chapters.

The second line in the formula for the Higgs Lagrangian 1.94 is especially illustrative of the fact that only thanks to the idea to shift the scalar field from its vacuum value, there can be non-zero mass terms associated to the gauge bosons (quadratic in field variable and proportional to m_W^2 for the two W bosons and to $\frac{1}{2}m_Z^2$ for the Z boson). It can be immediately recognized that

$$\begin{aligned} m_{W^\pm} &= \frac{1}{2}vg \\ m_Z &= \frac{1}{2}v\sqrt{g^2 + g'^2}. \end{aligned} \tag{1.95}$$

In other words, three massless Goldstone bosons vanished, and as a consequence, three gauge fields have acquired mass. Also note that the Higgs boson mass term remains the same as in the Abelian case. In general, EWK gauge boson and Higgs boson masses depend on particular coupling constants, which are not a priori known and need to be obtained experimentally or vice versa. Additionally, it can be assumed that due to the Higgs boson having no electric and color charge, there are no mass terms associated with photons and gluons, explaining their massless nature.

Another important aspect of the SM Higgs Lagrangian are the newly arising interaction terms between the Higgs boson and vector gauge bosons. For the brevity of this text, we will now focus on the SM interaction vertices of type HVV, where the only allowed tree-level contributions arising from this Lagrangian are respective to the choice of $VV := \{WW, ZZ\}$. In particular, Lagrangian terms

$$\begin{aligned} \mathcal{L}_{HWW} &= gm_W H W_\mu^- W^{+\mu} \\ \mathcal{L}_{HZZ} &= \frac{1}{2\cos\varphi_W} gm_Z H Z_\mu Z^\mu \end{aligned} \tag{1.96}$$

depict SM-allowed coupling factors assigned to HWW and HZZ vertices. Any other possible combinations respecting Higgs boson scalar nature, such as $VV := \{\gamma\gamma, Z\gamma, gg\}$, only become possible by adding higher order loop contributions which will be further discussed in Section 1.5. For the sake of completeness, the Higgs boson also appears to couple to itself through the cubic ($\propto H^3$) and quartic ($\propto H^4$) self-interactions.

1.4.4 Yukawa-type interactions

The idea of the Higgs boson interaction with gauge boson fields led to the non-zero EWK boson masses emerging via the BEH mechanism. A similar discussion can be done for the Yukawa-type interaction between the Higgs boson and fermions, eventually leading to the generation of fermion masses. The corresponding Yukawa-type Lagrangian for all types and generally different fermion flavors f, f' can be summarily written as

$$\mathcal{L}_{\text{Yukawa}} = -\sum_{f,f'} \bar{\Psi}_L^f \left[g_{f,f'}\Phi + \tilde{g}_{f,f'}\tilde{\Phi} \right] \Psi_R^{f'} + \text{h.c.} \tag{1.97}$$

where $\tilde{\Phi} = i\tau_2\Phi^*$ is the charge conjugate of the original scalar doublet Φ; $g_{f,f'}, \tilde{g}_{f,f'}$ are corresponding coupling constants, and h.c. was added to symbolize all terms obtained

as the hermitian conjugate to the terms explicitly written in the formula. The new $\tilde{\Phi}$ scalar field is a necessary addition to obtain physical mass terms for all types of fermions including neutrinos (eventhough other mechanisms of generating neutrino masses are being studied) and quarks, in accordance with the chirality decomposition of Dirac spinors. In the simplified example with the lefthanded SU(2) doublet $\Psi_L = (\nu_L, e_R)^T$ and singlet e_R, the Yukawa-type Lagrangian yields

$$\mathcal{L}^{(U)}_{\text{Yukawa}} = -\frac{1}{\sqrt{2}} g_e (H + v)(\bar{e}_L e_R + \bar{e}_R e_L) \tag{1.98}$$

where we have assumed U-gauge transformation properties for the scalar field $\Phi^{(U)}$ 1.93. It is now straightforward to recognize the electron mass $m_e = \frac{1}{\sqrt{2}} g_e v$ and the Yukawa interaction vertex

$$\mathcal{L}_{Hee} = -\frac{1}{\sqrt{2}} g_e H \bar{e} e. \tag{1.99}$$

Including the Yukawa-type Lagrangian to the overall picture concludes on the currently established and experimentally tested SM of elementary particles and fundamental interactions. As it was advertised in the beginning of this chapter, the SM theory with a handful of free parameters is yet incomplete. Particular attention is received by studies of the Higgs boson spin-parity properties as any deviation from the SM values could hint for new BSM physics.

1.5 Beyond the Standard Model

The Physics Beyond the Standard Model (BSM) term is continuously evolving in its definition as the new "standards" are re-defined by the currently accepted understanding of the full Standard Model (SM) picture. A prime example of dynamics in the High Energy Physics (HEP) field is the fact that not too long ago, e.g. see the summary of theoretical and experimental challenges of BSM physics by J. Ellis [52] from 2009, the BEH mechanism was considered yet a possible extension of the "old" SM theory. About ten years after the discovery of the massive $m_h \approx 125$ GeV scalar resonance, we consider the Higgs boson to be a canonical part of the SM, while keeping an eye on possible anomalous discrepancies from the predicted Higgs boson quantum properties.

To contradict the famous quotation of the former US patent officer Charles H. Duell saying: "[...] everything that can be invented has been invented.", there is clearly a handful of questions puzzling physicists up to this date, not all of them necessarily related to the Higgs boson. Among those we can name a few potentially accessible by the HEP programme:

1. The matter-antimatter asymmetry – It is not yet resolved why there is a dominance of matter over antimatter in our Universe. Could it be the CP-violation in anomalous h scalar field interactions responsible for this asymmetry? Could the existence of a heavy "Majorana" neutrino decays that explains the abundance of matter?

2. Dark matter – About 26% of our Universe is made of dark matter while its origin is fully unknown. Astrophysical observations (such as gravitational lensing) are in favor of the existence of exotic matter which causes the abnormal angular velocities of objects rotating around the centres of the galaxies. Up to this date, no known particle candidates were discovered that would explain this phenomenon.

3. Flavor puzzle – There are three generations/flavors of fermions, however it is not known why. It is also observed there is a certain hierarchy in fermion masses which is yet not understood.

Other theories adding up to the coverage of the BSM physics, such as grand unification of fundamental forces or quantum theory of gravity are currently beyond our technological capabilities as they are expected to occur at scales no smaller than $10^{16} - 10^{19}$ TeV.

Walking on the edge of philosophical discussion, any scientific theory should eventually become fundamental, thus describing the Nature everywhere and at any scale. This is often not possible given the limitations in experimental capabilities or simply because of a lack of knowledge. When searching for the BSM theories to build on top of the SM, it is often useful to approach problematics within a particular energy (or distance) scale. This approximation reflects on the fact that certain processes cannot be visible for a certain scale, nevertheless it does not affect any underlying (fundamental) theory that would explain such a behavior, though it is yet unknown. The family of the theories which operate below given energy scale Λ, while being (almost fully) decoupled from what is happening at high energies is described as "effective".

The idea behind the effective theories is not alien also in other research fields. Historically, physicists tend to describe natural phenomena with limited information or margin of the physical scale such as the provided explanation is, indeed, valid within a desired accuracy. Typical example is calculating the precession rate of Mercury's perihelion [53] using Newton's gravity law compared to results obtained with corrections from Einstein's general relativity which matches observations with superior precision. Another pioneering moment is the application of effective theories in studying critical phenomena in condensed matter systems [54] that served as a base for modern EFT employed in HEP analyses. While other theoretical models might also bring some light into the BSM physics, it seems natural to exploit the capabilities of the EFT models, especially that we now possess sufficient sensitivity with modern particle accelerators.

In the following sections we will address the basic concepts of the EFT expansion of the SM Lagrangian. We will focus on terms affecting the SM HVV vertex, where VV stands for any of the following combinations of gauge bosons: $WW, ZZ, \gamma\gamma, Z\gamma, gg$. We will further translate relevant EFT Lagrangian terms into the language of the scattering amplitude as it offers a more intuitive overview of possible BSM contributions to the Higgs boson production and decay (HVV) vertices. Finally we will introduce the Anomalous Couplings (AC) and Standard Model EFT (SMEFT) schemes providing mutually alternative interpretation of the observed results that are naturally compatible with the EFT models.

1.5.1 Effective Field Theories

The driving idea behind the EFT is to introduce a scheme in which the system's dynamics at a low energy scale (typically below 1 TeV) is isolated from the processes occurring at high energies. This is a reasonable assumption as any new physics occurring at low energies would likely be noticed in previous experiments while underlying physics might only be accessible at energies beyond our current reach. Quantitatively, the energy scale Λ (possibly multiple independent scales) serves as a factor suppressing any terms in the new BSM Lagrangian that are accounting for the EFT contribution. In particular we can write

$$\mathcal{L}_{\text{BSM}} = \mathcal{L}_{\text{SM}} + \mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^{d_i-4}} \mathcal{O}_i, \tag{1.100}$$

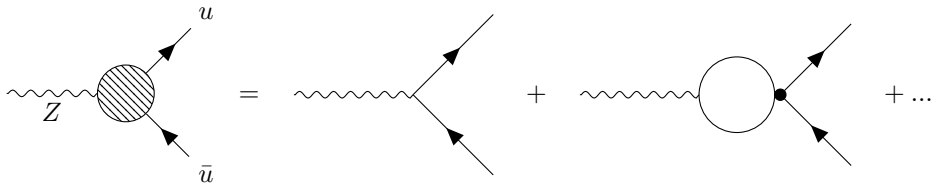

Figure 1.6: Tree-level contribution and 1-loop corrections (RHS) to the effective $Z\bar{u}u$ interaction vertex (LHS). Solid blob in 1-loop diagram depicts the fact that either dimension 6 or 8 operators are contributing to the four-quark vertex.

where EFT operators with a (mass) dimension of $[\mathcal{O}_i] = d_i$ contribute proportionally to the dimensionless Wilson coefficients c_i suppressed by a corresponding power of the Λ scale. In this regard, the EFT is constructed without violating well-established SM as it is decoupled from high energy scale processes and which only manifests indirectly by placing constraints on a priori unknown Wilson coefficients. Moreover, new operators entering the BSM Lagrangian take into account any of the underlying symmetries, such as $SU(3) \times SU(2) \times U(1)$ and Lorentz invariance required by the SM. As such, the BSM Lagrangian of this type is in literature often interchangeably denoted as the Standard Model EFT Lagrangian (SMEFT). On the other hand, Equation 1.100 might imply problems with the predictive power of the action arising from an infinite sum of operators entering calculations. Before we address this issue, let us take one step back to make a general comment on the dimensionality of operators.

The SM Lagrangian operators can be found with a dimension of ≤ 4, e.g.:

- boson mass terms with $d_i = 2$,
- fermion mass terms or ϕ^3 scalar self-interactions with $d_i = 3$,
- various kinetic terms, QED or QCD interactions with $d_i = 4$.

Note that the previous statements might be confirmed from a dimensional analysis of quantum fields (and their derivatives) forming the SM operators, e.g. using $[\phi] = 1, [\psi] = 3/2, [A_\mu] = 1, [F_{\mu\nu}] = 2$. These operators are denoted as renormalizable, as they lead to a finite cross-section valid up to energies E/Λ where Λ can stand for the mass M of a heavy resonance or other natural energy threshold. On the other hand, EFT operators start with a dimension of $d_i > 4$, often inducing EFT-loop corrections to the terms that would normally appear in the SM Lagrangian (i.e. corrections to the tree-level but also SM-loop contributions in principle).

Figure 1.6 shows an example of the 1-loop (EFT) corrections to the $Z\bar{u}u$ interaction vertex on top of the SM tree-level contribution. Four-quark vertex that needs to be added to the 1-loop diagram is depicted as a solid blob to emphasize that either $d_i = 6$ or $d_i = 8$ EFT operators were included (see the proof that loop-order is not a priori equal to the order of terms in the Lagrangian expansion [55]). Unfortunately, EFT loop corrections (similar as in case of higher order perturbation in the SM theory) are causing undesired divergences. As can be derived from the Feynman rules, particular contributions to $Z\bar{u}u$

vertex are proportional to values

$$\begin{aligned} I_{d_i=6} &\sim \frac{1}{M^2}\int \frac{1}{p^2}\ \mathrm{d}^4p \approx \frac{1}{M^2}\Lambda^2 \approx \mathcal{O}(1), \\ I_{d_i=8} &\sim \frac{1}{M^4}\int \frac{1}{p^2}p^2\ \mathrm{d}^4p \approx \frac{1}{M^4}\Lambda^4 \approx \mathcal{O}(1), \end{aligned} \tag{1.101}$$

where in the integral computation we have employed a momentum cut-off at typical energy scale $\Lambda = M$ to show that both higher-dimensional contributions are of the order of unity. Moreover, a possibly infinite sum of higher-dimensional operators that could contribute to this vertex (2-loop and higher order diagrams) also adds up to the divergent behavior. In other words, seemingly an infinite sum of counterterms would need to be added by hand to recover physical (finite) results, breaking up any predictive power that EFT could offer.

In practice, this problem is usually approached by implementing suitable (mass-independent) renormalisation scheme such as dimensional regularisation and minimal subtraction schemes, which detailed description is left for expert literature [56, 57, 58]. Nevertheless, let us give an example of dimensional regularisation in treating the aforementioned divergent contributions, i.e. introducing an arbitrary renormalisation scale μ such as $\mathrm{d}^4p \to \mu^{4-d_i}\,\mathrm{d}^{d_i}p$. Such a regularisation is typically leading to integrals proportional to values

$$\begin{aligned} I_{d_i=6} &\approx \frac{m^2}{M^2}\log\mu, \\ I_{d_i=8} &\approx \frac{m^4}{M^4}\log\mu, \end{aligned} \tag{1.102}$$

where the renormalisation scale μ only appears in a logarithm (and not in the explicit powers such as μ^2). Both integrals are now small, assuming that newly appearing mass scale m, e.g. mass of the quark entering loop, is sufficiently smaller than scale of the new physics M. Generally, if the renormalisation procedure is performed up to given E/Λ power in Lagrangian expansion, i.e. up to desired precision $\epsilon < (E/\Lambda)^{d_i-4}$, the EFT Lagrangian yields definite cross-section values and restores its predictive powers. For the sake of clarity, physical observables naturally do not depend on the choice of a renormalisations scheme, however any prediction is done up to the given order, thus results will be affected by the choice and should be understood as approximative.

Another famous and possibly more intuitive application of EFT theory dates back to the middle of 20$^{\text{th}}$ century when particular focus was put on experiments demonstrating weakly interacting particles, such as measurement of muon decay to electron,

$$\mu^- \to e^- + \bar{\nu}_e + \nu_\mu. \tag{1.103}$$

In this process scientists observed a continuous electron energy spectrum, thus assumed 3-body particle decay hinting for a presence of a bounding force between muon, electron and both neutrinos. At that time, explanation via Fermi's four-fermion interaction vertex, see Figure 1.7 (left) where the presence of an effective $d_i = 6$ operator is denoted by a black solid blob, appeared as a natural solution as it remarkably well described observed muon decay rates. According to Fermi's theory [39], muon decay width is driven by factor of $\approx G_F^2 m_\mu^5$, where m_μ is the mass of the muon and G_F stands for the Fermi constant. Low-energy weak scattering processes were also described using the same theory, however this could only be achieved with limited precision. Furthermore, discrepancies between

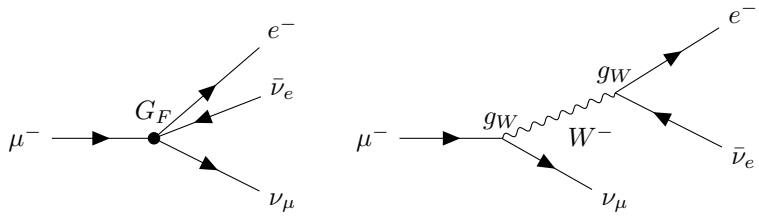

Figure 1.7: Feynman diagrams for muon decaying into electron, electron anti-neutrino and muon neutrino under an assumption of effective Fermi's four-lepton interaction theory (left) and underlying theory with intermediate W boson (right).

expectation and experiment started to appear for processes with momentum exchange typically above the 1 GeV threshold, questioning limitations of Fermi's theory. Indeed, it can be shown that the coupling factor appearing at Fermi's type of vertex assumes a form of $G_F \approx g_W^2/m_W^2$, i.e. the validity region for effective Fermi's theory is bound by suppression factor of $1/m_W^2$ while it fails to describe processes at higher energies. Eventually, the discovery of the charged W boson by the end of 20^{th} century has shed a new light on the underlying theory for muon decays (and similar processes) by introducing weak charged currents as depicted by the diagram in Figure 1.7 (right). As such, the prediction of a model with an exchange of a charged W boson is indistinguishable from an effective Fermi's theory only if the momentum transfer is smaller than m_W. On the other hand, the weak charged current has become a canonical part of the SM and can retrospectively be assumed as an underlying theory for Fermi's four-fermion interaction.

1.5.2 Effective Lagrangian in HVV sector

Let us have a closer look at the possible BSM contributions into the HVV interaction vertex. It appears there is only one possible dimension $d_i = 5$ SMEFT operator, assumed to affect the generation of neutrino masses [59], which is not relevant in our case. Moreover it belongs to the family of operators that do not preserve the lepton number $\mathbb{L}$ – a requirement fulfilled in the SM. Similarly, dimension $d_i = 7$ operators will not be considered as they do not preserve the variable defined as the $\mathbb{B} - \mathbb{L}$ [60], where $\mathbb{B}$ is a baryon number. Dimension $d_i \geq 8$ operators are suppressed by at least a Λ^4 scale factor and higher, thus can be neglected, given our current experimental capabilities. Remaining considerable SMEFT operators have a dimension of $d_i = 6$, they preserve both lepton and baryon numbers and can be constructed from the SM field operators, i.e. $X_\mu \in \{G_\mu^a, W_\mu^a, B_\mu\}$ and an SU(2) doublet Φ, as well as the even number of covariant derivatives D_μ.

The smallest non-redundant independent operator basis corresponding to the dimension $d_i = 6$ consists of 59 operators, out of which only 9 might contribute to the effective HVV vertex. In practice, multiple equivalent operator bases are considered across analyses, some of them more suitable to access observables sensitive to the studied phenomena than others. Explicitly, we assume 9 operators originally defined in the so called Warsaw

basis [61]

$$
\begin{aligned}
&\mathcal{O}_{\Phi G} = \Phi^\dagger \Phi G^a_{\mu\nu} G^{a\mu\nu},\ \mathcal{O}_{\Phi \tilde{G}} = \Phi^\dagger \Phi \tilde{G}^a_{\mu\nu} G^{a\mu\nu},\\
&\mathcal{O}_{\Phi W} = \Phi^\dagger \Phi W^a_{\mu\nu} W^{a\mu\nu},\ \mathcal{O}_{\Phi \tilde{W}} = \Phi^\dagger \Phi \tilde{W}^a_{\mu\nu} W^{a\mu\nu},\\
&\mathcal{O}_{\Phi B} = \Phi^\dagger \Phi B_{\mu\nu} B^{\mu\nu},\ \mathcal{O}_{\Phi \tilde{B}} = \Phi^\dagger \Phi \tilde{B}_{\mu\nu} B^{\mu\nu},\\
&\mathcal{O}_{\Phi WB} = \Phi^\dagger \sigma^a \Phi W^a_{\mu\nu} B^{\mu\nu},\ \mathcal{O}_{\Phi \tilde{W} B} = \Phi^\dagger \sigma^a \Phi \tilde{W}^a_{\mu\nu} B^{\mu\nu},\\
&\mathcal{O}_{\Phi D} = (\Phi^\dagger D^\mu \Phi)^* (\Phi^\dagger D_\mu \Phi),
\end{aligned}
\tag{1.104}
$$

where σ^a stand for Pauli matrices, operators consisting only of conventional gauge field-strength tensors $X_{\mu\nu} \in \{G^a_{\mu\nu}, W^a_{\mu\nu}, B_{\mu\nu}\}$ (see definition in Equations 1.68, 1.62 and 1.61) are CP-even and operators containing at least one dual field-strength tensor $\tilde{X}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma} X^{\rho\sigma}$ are CP-odd. Note that observing non-zero values of Wilson coefficients (rather their Λ^2-suppressed magnitude) corresponding to the CP-odd operators would indicate CP-parity violation.

Furthermore, it is convenient to recast the effective part of the SMEFT Lagrangian in terms of the physical fields appearing after spontaneous symmetry breakdown, e.g. gluon, $\gamma, W^\pm, Z$, Higgs boson and other mass eigenstates. In particular, one could divide resulting Lagrangian terms based on the number of physical fields participating in (self)interactions (or kinetic terms), e.g.

$$
\mathcal{L}_{\text{EFT}} = \mathcal{L}_{\text{kinetic}} + \mathcal{L}_{\text{3V}} + \mathcal{L}_{\text{4V}} + \mathcal{L}_{\text{HVV}} + \mathcal{L}_{\text{H}^2} + \mathcal{L}_{\text{H,self}} + \mathcal{L}_{\text{other}},
\tag{1.105}
$$

where individual parts of the Lagrangian stand for (in this order) kinetic terms, terms with 3 and 4 gauge vectors, terms contributing to the HVV vertex, quartic terms with 2 Higgs fields and 2 other fields, multi-Higgs self-interaction terms and other terms accounting for combinations including fermion fields. For the convenience of the reader and relevance in this analysis we write down terms corresponding to the HVV vertex:

$$
\begin{aligned}
\mathcal{L}_{\text{HVV}} = \frac{H}{v}\Bigg[&(1+\delta c_w)\frac{g^2 v^2}{2} W^+_\mu W^{-,\mu} + (1+\delta c_z)\frac{(g^2+g'^2)v^2}{4} Z_\mu Z^\mu\\
&+ c_{ww}\frac{g^2}{2} W^+_{\mu\nu} W^{-,\mu\nu} + \tilde{c}_{ww}\frac{g^2}{2} W^+_{\mu\nu} \tilde{W}^{-,\mu\nu} + c_{w\Box} g^2 (W^-_\mu\, \partial_\nu W^{+,\mu\nu} + \text{h.c.})\\
&+ c_{gg}\frac{g_s^2}{4} G^a_{\mu\nu} G^{a,\mu\nu} + c_{\gamma\gamma}\frac{e^2}{4} A_{\mu\nu} A^{\mu\nu} + c_{Z\gamma}\frac{e\sqrt{g^2+g'^2}}{2} Z_{\mu\nu} A^{\mu\nu}\\
&+ c_{ZZ}\frac{g^2+g'^2}{4} Z_{\mu\nu} Z^{\mu\nu} + c_{Z\Box} g^2 Z_\mu\, \partial_\nu Z^{\mu\nu}\\
&+ \tilde{c}_{gg}\frac{g_s^2}{4} G^a_{\mu\nu} \tilde{G}^{a,\mu\nu} + \tilde{c}_{\gamma\gamma}\frac{e^2}{4} A_{\mu\nu} \tilde{A}^{\mu\nu} + \tilde{c}_{Z\gamma}\frac{e\sqrt{g^2+g'^2}}{2} Z_{\mu\nu} \tilde{A}^{\mu\nu}\\
&+ \tilde{c}_{ZZ}\frac{g^2+g'^2}{4} Z_{\mu\nu} \tilde{Z}^{\mu\nu} + c_{\gamma\Box} g g' Z_\mu\, \partial_\nu A^{\mu\nu}\Bigg]
\end{aligned}
\tag{1.106}
$$

Note that each term is proportional to one of the Wilson coefficients expressed in the so called Higgs basis [62], e.g. coefficients marked with the d'Alembert symbol $\Box$ are related to terms which are quadratic in partial derivatives. The particular form of the operator parametrisation and linear relations between Wilson coefficients in the Warsaw and Higgs basis is to be found explicitly determined in [63]. Also note that the first line in 1.106 has the same form as the SM tree-level contributions to the HWW and HZZ vertices 1.96 (omitting coefficients δc_w and δc_z which presence would only manifest by

the changes in the normalisation of the SM process) while other lines consist of CP-even and CP-odd operators with a tensor structure not being present in the SM Lagrangian. The missing Λ^2 suppression factor is implicitly included and the original notation can be restored by performing the $c_i \to \frac{v^2}{\Lambda^2} c_i$.

1.5.3 Spin-0 scattering amplitude

General discussion in previous sections comprised of possible higher-dimensional EFT extension to the SM HVV vertex. We have so far omitted any discussion of the pure SM corrections emerging from higher orders in the perturbation series as they are currently experimentally inaccessible. For example, the magnitude of the expected SM loop contribution to the HVV vertex corresponding to the a_2 coupling is of the $\mathcal{O}(10^{-2})$ order, while an even more tiny $\mathcal{O}(10^{-11})$ correction is expected from the a_3 coupling three-loops contribution. Nevertheless, coupling constants a_i related to the SM corrections might get enhanced by the EFT contributions, i.e. $a_i \approx a_{\text{SM}} + a_{\text{EFT}}/\Lambda_i^2$ (see details on normalisation in [64]). As such, we will treat them together with pure EFT higher-dimensional terms which normally do not appear in the SM on their own. Combining both types of possible contributions to the SM (summarily denoted as BSM/AC effects) is a strategic move to define the most general parametrisation of new undiscovered physics phenomena.

In this thesis (and related analysis [38]) we aim to measure the magnitude of the aforementioned BSM/AC effects to the HVV interaction vertex (both in Higgs boson production and decay) relative to the known SM values. In theory, this could be achieved directly by determining values of the Wilson coefficients (in arbitrary basis) as retrieved from a fit on data (plus measuring loop-order SM corrections individually). From the experimental point of view, however, it is more practical to turn from the Lagrangian "language" and employ the probabilistic terminology by rewriting formula 1.106 into generalised tensor form of the HVV scattering amplitude (e.g. by using Feynman rules). We also restrain ourselves to consider only amplitude terms related to the spin-0 Higgs boson hypothesis, given that more exotic scenarios with a spin-1 or spin-2 Higgs boson structures were already constrained in analyses with the Run-1 dataset at CMS [33] and ATLAS [34] experiments. In particular we assume an amplitude

$$\begin{aligned}\mathcal{A}(\text{HV}_1\text{V}_2) \approx & \left[a_1^{\text{VV}} + \frac{\kappa_1^{\text{VV}} q_{\text{V1}}^2 + \kappa_2^{\text{VV}} q_{\text{V2}}^2}{(\Lambda_1^{\text{VV}})^2} + \frac{\kappa_3^{\text{VV}} (q_{\text{V1}} + q_{\text{V2}})^2}{(\Lambda_Q^{\text{VV}})^2} \right] m_{\text{V1}}^2 \epsilon_{\text{V1}}^* \epsilon_{\text{V2}}^* \\ & + \frac{1}{v} a_2^{\text{VV}} f_{\mu\nu}^{*(1)} f^{*(2),\mu\nu} + \frac{1}{v} a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu\nu},\end{aligned} \tag{1.107}$$

where q_{Vi}, m_{Vi} and ϵ_{Vi} are the gauge boson momentum, pole mass and the polarisation vector; the field-strength tensor and its dual version are in terms of the scattering amplitude defined as $f^{(i),\mu\nu} = \epsilon_{\text{Vi}}^{\mu} q_{\text{Vi}}^{\nu} - \epsilon_{\text{Vi}}^{\nu} q_{\text{Vi}}^{\mu}$ and $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} f^{(i),\rho\sigma}$; Λ_1^{VV} and Λ_Q^{VV} are the EFT scales explicitly suppressing terms that do not occur in the SM; a_1^{VV} is a coupling constant corresponding to the SM amplitude and, finally, couplings $\kappa_1^{\text{VV}}, \kappa_2^{\text{VV}}, \kappa_3^{\text{VV}}$ and $a_2^{\text{VV}}, a_3^{\text{VV}}$ represent AC and will be further discussed per each combination of gauge vectors VV. In the following list, we summarise applicable symmetry requirements as well as our empirical knowledge and remarks about all present couplings that further simplify the formula for the scattering amplitude:

- only the SM tree-level contribution to the scattering amplitude emerges from the relations $a_1^{WW} \neq 0$, $a_1^{ZZ} \neq 0$ and $a_{i\neq 1}^{\text{VV}} = 0$; there is no SM tree-level contribution to the H$\gamma\gamma$, Hgg and HZγ vertices (i.e. $a_1^{\gamma\gamma} = a_1^{gg} = a_1^{Z\gamma} = 0$),

- requirements of symmetry and gauge invariance imply $\kappa_1^{gg} = \kappa_2^{gg} = 0$, $\kappa_1^{\gamma\gamma} = \kappa_2^{\gamma\gamma} = 0$, $\kappa_1^{Z\gamma} = 0$, $\kappa_3^{\mathrm{VV}} = 0$, $\kappa_1^{ZZ} = \kappa_2^{ZZ} = -\exp(i\phi_{\Lambda_1}^{ZZ})$ and $\kappa_2^{Z\gamma} = -\exp(i\phi_{\Lambda_1}^{Z\gamma})$, where $\phi_{\Lambda_1}^{ZZ}$ and $\phi_{\Lambda_1}^{Z\gamma}$ are anomalous phases,
- in case of H$\gamma\gamma$ and Hgg vertices $m_{\mathrm{V1}} = 0$, therefore only $a_2^{\gamma\gamma,gg}$ and $a_3^{\gamma\gamma,gg}$ AC appear in the formula,
- $a_2^{\gamma\gamma,Z\gamma}$ and $a_3^{\gamma\gamma,Z\gamma}$ couplings are assumed to get fully constrained in analyses with direct $H \to \gamma\gamma$ and $H \to Z\gamma$ decay channels and we are effectively setting them to zero,
- it is possible to study a_3^{gg} AC (while assuming a_2^{gg} receives only contribution from the SM loop corrections), however it is not considered for discussion in this thesis,
- terms proportional to a_3^{VV} couplings are CP-odd, their presence (i.e. $a_3^{\mathrm{VV}} \neq 0$) would indicate CP-parity violation.

More assumptions can be done depending on the choice of our interpretation as discussed below.

1.5.4 AC vs. SMEFT interpretation

Relative correspondence of the a_i^{ZZ} and a_i^{WW} couplings from the scattering amplitude 1.107 is mostly relevant for the Higgs boson VBF production process. Practically, there are hardly any measurable kinematic differences between the two fusion modes which allows to further reduce the number of independent AC. This degree of freedom might be addressed from two different perspectives denoted as:

1. AC scheme,
2. SMEFT scheme.

In case of the AC scheme interpretation, it can simply be assumed that $a_i^{WW} = a_i^{ZZ} = a_i$, where we have intentionally removed the superscript ZZ/WW. List of the relevant (independent) AC is then reduced to the CP-even a_2, $a_{\Lambda_1} \equiv \kappa_1^{WW}/(\Lambda_1^{WW})^2 = \kappa_1^{ZZ}/(\Lambda_1^{ZZ})^2$, $a_{\Lambda_1}^{Z\gamma} \equiv \kappa_2^{Z\gamma}/(\Lambda_1^{Z\gamma})^2$ and the CP-odd a_3. Dedicated strategy to study all four couplings independently (by only keeping one non-zero coupling at the time) is employed and discussed in following chapters.

In case of the SMEFT scheme interpretation, the $SU(2) \times U(1)$ symmetry is required (as already discussed in the general Section 1.5.1), implying validity of the following set of relations:

$$\begin{aligned} a_1^{WW} &= a_1^{ZZ}, \\ a_2^{WW} &= c_W^2 a_2^{ZZ}, \\ a_3^{WW} &= c_W^2 a_3^{ZZ}, \\ \frac{\kappa_1^{WW}}{(\Lambda_1^{WW})^2} &= \frac{1}{c_W^2 - s_W^2}\left(\frac{\kappa_1^{ZZ}}{(\Lambda_1^{ZZ})^2} - 2s_W^2 \frac{a_2^{ZZ}}{m_Z^2}\right) \\ \frac{\kappa_2^{Z\gamma}}{(\Lambda_1^{Z\gamma})^2} &= \frac{2 s_W c_W}{c_W^2 - s_W^2}\left(\frac{\kappa_1^{ZZ}}{(\Lambda_1^{ZZ})^2} - \frac{a_2^{ZZ}}{m_Z^2}\right) \end{aligned} \tag{1.108}$$

where $s_{\mathrm{W}} = \sin(\varphi_W)$ and $c_{\mathrm{W}} = \cos(\varphi_W)$. As a consequence, number of independent anomalous couplings in the SMEFT scheme (or SMEFT couplings) is reduced to three:

a_2, $a_{\Lambda_1} \equiv \kappa_1^{WW}/(\Lambda_1^{WW})^2$ and a_3. The SMEFT couplings are studied both independently with other couplings fixed to zero and simultaneously with other couplings left floating in the fit. For the sake of completeness, relevant terms in the scattering amplitude 1.107 can be explicitly matched to the corresponding terms in the effective Lagrangian 1.106 by assuming following relations between the SMEFT couplings and Wilson coefficients in Higgs basis:

$$\begin{aligned} \delta c_Z &= \frac{1}{2} a_1^{ZZ} - 1 \\ c_{ZZ} &= -2\frac{s_W^2 c_W^2}{e^2} a_2^{ZZ} \\ \tilde{c}_{ZZ} &= -2\frac{s_W^2 c_W^2}{e^2} a_3^{ZZ} \\ c_{Z\Box} &= \frac{m_Z^2 s_W^2}{e^2} \frac{\kappa_1^{ZZ}}{(\Lambda_1^{ZZ})^2} \end{aligned} \tag{1.109}$$

As a direct consequence, any constraints set upon the AC values in measurement can be expressed in the arbitrary SMEFT basis. Note that other Wilson coefficient that are not explicitly mentioned in 1.109 are irrelevant for the HVV vertex, become reducible due to the symmetry requirement applied in SMEFT scheme or were set to zero for various reasons discussed in the previous section.

Finally, better than directly constraining anomalous couplings (in the AC or SMEFT schemes), it is convenient to set constraints on an effective fractional cross-section defined as

$$f_{a_i} = \frac{|a_i|^2 \sigma_i}{\sum_j |a_j|^2 \sigma_j} \text{sign}\left(\frac{a_i}{a_1}\right), \tag{1.110}$$

where the sum runs over all considered couplings and σ_i is the cross-section for the process corresponding to $a_i = 1$. Possible f_{a_i} values lie between -1 and 1 where $f_{a_i} = 0$ can be conveniently interpreted as a pure SM scenario. In other words, constraining f_{a_i} by rejecting all regions besides a small interval around zero would indicate compatibility with the SM, while allowing for any $f_{a_i} \neq 0$ value could hint for signs of new physics. As a bonus, many of the systematic uncertainties cancel out in the ratio as compared to constraining the a_i values directly. For the compatibility reasons with measurements previously carried out by the CMS experiment, σ_i values are chosen to correspond to the $gg \to H \to ZZ \to 2e2\mu$ process and are enlisted (relative to the SM cross-section) in Table 1.1 separately for AC and SMEFT schemes. This choice does not affect the interpretability of the final results while it is allowing for future combination with similar analyses in different final state decay channels where the same approach was chosen.

Analytical form of the HVV scattering amplitude 1.107 re-phrased in terms of the effective fractional cross-section introduced in this chapter serves as a base for constructing the physical model used in the final fit. Detailed description of the signal model is provided in Section 4.4.3 while more about the general fitting method employed to constrain anomalous couplings from data is available in Section 6.2.1.

f_{a_i}	AC scheme σ_i/σ_1	SMEFT scheme σ_i/σ_1
f_{a_3}	0.153	0.153
f_{a_2}	0.361	6.376
$f_{a_{\Lambda 1}}$	0.682	5.241
$f_{a_{\Lambda 1}^{Z\gamma}}$	1.746	–

Table 1.1: Ratio of the cross-section defined for the process $a_i = 1$ over cross-section corresponding to the SM process as calculated by event generator JHUGen 7.0.2. In case of the $\kappa_1^{WW,ZZ}$ and $\kappa_2^{Z\gamma}$ couplings, the numerical values of $\Lambda_1^{WW,ZZ,Z\gamma} = 100$ GeV were chosen to keep all coefficients of similar order of magnitude. Ratios are displayed for the AC and SMEFT schemes separately.

CHAPTER

TWO

COMPACT MUON SOLENOID AT LHC

The history of experimental particle physics has shown that the nature of elementary particles and fundamental interactions is likely to be revealed through both non-accelerator and accelerator particle experiments. The former typically includes, but is not exclusive to, experiments with high-energy cosmic rays, neutrino experiments, or dark matter searches. The latter employs accelerator technology in fixed target experiments, or more importantly, in experiments with particle colliders operating in a beam-to-beam mode. In any case, individual particles incoming from natural sources or arising from primary collisions are identified, and their kinematic properties are measured using dedicated detection techniques. Ideally, the higher the detector's resolution, precision, angular coverage and efficiency, the better the understanding of physical processes under study.

This chapter aims to introduce the experimental setup used to collect data relevant to the studies reported in this thesis. Section 2.1 provides an introduction to the basic accelerator principles, followed by a short overview of LHC technologies. A more detailed description of the CMS detector and its delicate subsystems is presented in Section 2.2. It is worth noting that the information provided about the LHC and CMS components is relevant for the LHC operation during the years 2015-2018, also known as the Run-2 period of data taking, unless stated otherwise.

2.1 Particle accelerators

With a bit of imagination, particle accelerators can be viewed as large "microscopes", operating with particle beams that act as "probes" into the blocks of fundamental matter. The larger the beam energy, the better the spacial resolution and the smaller the objects that can be studied. From another point of view, colliding particles undergo inelastic scattering processes leading to the production of new particles, some of which yet need to be discovered. The higher the energy put into the beam collision, quantified by means of the center-of-mass energy $\sqrt{s}$, the heavier the particles that might be produced. Assuming particle beams with energy E and containing particles of mass m, experiments with beam-to-beam particle colliders profit from a higher energy outcome $\sqrt{s} \propto 2E$, as compared to fixed target experiments with $\sqrt{s} \propto \sqrt{2mE}$.

On the other hand, the probability of a particle beam interacting with dense target material is higher than that of interaction in the head-on collision of two anti-parallel beams. To improve these odds, both beams are segmented into multiple concentrated

particle bunches, increasing the interaction chance per each bunch crossing. An important quantity describing collider performance in terms of the interaction rate r is called the instantaneous luminosity $\mathscr{L}$, defined as

$$r = \mathscr{L}\sigma \tag{2.1}$$

where σ stands for the interaction cross-section defined in the previous chapter. Assuming both colliding bunches contain n identical particles and that the beam density in the transverse profile is given by a Gaussian distribution described by width parameters s_x and s_y, the instantaneous luminosity is given by the formula

$$\mathscr{L} = f\frac{n^2}{4\pi s_x s_y} \tag{2.2}$$

where f is the collision frequency. To estimate the amount of collected data, an integral of the instantaneous luminosity over a selected time period is determined and will be denoted as the integrated luminosity L measured in units of inversed femtobarns (fb^{-1}).

Another limitation comes with the particular accelerator design. Linear accelerators (linacs) have their capability to reach higher energies directly restricted by the accelerator length, while circular colliders (cyclotrones, synchrotrones, etc.) allow for multiple acceleration rounds. In most cases, acceleration technology is based on Radio Frequency (RF) cavities providing an electrical impulse to all charged particles passing through the cavity volume. Depending on the phase difference between incoming particles and the oscillation frequency of the electric field inside the RF cavity, the particles will be boosted, slowed down or left untouched at desired energy peak values. By definition, circular colliders need to keep the particle beams on stable radius trajectories, placing stringent requirements on sophisticated magnet systems.

The concrete choice of the accelerated particles also plays a significant role, i.e. the usage of electrons at circular colliders is limited by severe radiative losses as compared to the much milder effect in the case of heavier protons. These and many other technical aspects need to be taken into account, arguably leading to the trend of choosing the circular design over the linear alternative. The following example of the world-class circular collider puts most of the state-of-the-art accelerator technology into practice.

2.1.1 Large Hadron Collider

The LHC is the world's largest-ever built circular particle accelerator. Housed by the "Conseil Européen pour la Recherche Nucléaire", known as the CERN laboratory, it is located near Geneva, overlapping the Switzerland-France borders. The LHC machine is placed in an underground tunnel with an internal diameter of 3.7 m and a circumference of almost 27 km. It is situated at depths varying from 50 to 175 m. Its position provides an advantage both to limit the area occupied above the surface as well as to protect against external radiation sources and other disruptive factors.

The LHC was originally designed to study high-energy pp-collisions primarily aiming for clarification of the mechanism behind electroweak symmetry breakdown – a theory now widely believed to be confirmed. In that regard, the LHC consists of two parallel rings equipped with fully separated vacuum tubes capable to accommodate two proton beams oriented in mutually opposite direction. The proton beams reach the main LHC rings pre-boosted as they pass through a chain of several pre-acceleration machines. Everything starts with an electric field stripping electrons from hydrogen gas atoms. The separated protons are sent to LINAC 2 for an initial 50 GeV boost, followed by several

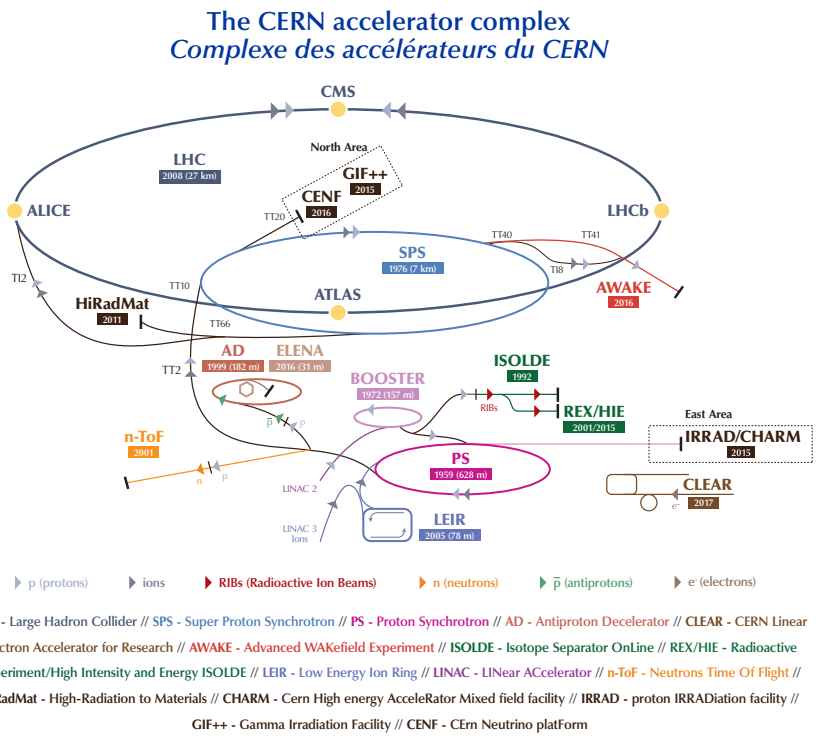

Figure 2.1: The CERN accelerator complex (August 2018). Injection chain in case of pp-collisions follows the path: LINAC 2 → BOOSTER → PS → SPS → LHC. Source: [65].

acceleration rounds in the Circular Proton Synchrotron Booster (BOOSTER) increasing the beam energy to 1.4 GeV. As a next step, the proton beams are segmented into multiple bunches separated by a 25 ns time gap employing RF cavities installed at the Parton Showers (PS). Additionally, the PS increases the beam energy to 25 GeV. Prior to injection into the LHC, the proton beams are sent to the Super Proton Synchrotron (SPS) for a final pre-acceleration cycle. The SPS has a circumference of 7 km allowing to reach a beam energy of 450 GeV. The entire acceleration complex (including the location of the experiments) is depicted in Figure 2.1.

Finally, the proton beams are passed to the LHC acceleration rings. Two separate superconducting RF systems (one for each ring), operating at $\approx$ 400 MHz frequency, provide an energy boost of $\approx$ 0.5 MeV per acceleration round up to maximum designed beam energy of 7 TeV (corresponding to a maximum center-of-mass energy of $\sqrt{s} =$ 14 TeV). The proton beams are kept at a stable circular trajectory thanks to the more than 1200 dipole superconducting magnets situated alongside the acceleration rings. To provide a magnetic field of 8 T, the dipoles are cooled down to < 2 K by using superfluid helium, making the LHC one of the coldest places in the known universe. More than 400 quadrupoles and other multipole magnets are installed to further improve beam stability and increase integrated luminosity by actively controlling beam focusing. In case of poor beam quality (or emergency), the proton beams are redirected to a dedicated dump area where the contained energy can be safely dissipated in an $\approx$ 8 m long Graphite tube. On the oher hand, beams reaching target energy and passing quality checks are sent to

a collision course with bunch-crossing happening at the nominal Interaction Point (IP) located in experimental caverns alongside the LHC circumference.

The LHC design [66] was commissioned in Winter 2009 and fully started its operations by 2010, marking the early beginning of the so called Run-1 period of data taking lasting till Spring 2013. Since then, the LHC is continuously operating with an exception of regular Year-End Technical Stops (YETS) dedicated to small or critical maintenance and Long Shutdown (LS) breaks after each Run period of data-taking intended for time consuming technical upgrades. During the Run-2 period, the proton beams reached a maximum energy of 6.5 TeV and the LHC delivered 160 fb^{-1} of integrated luminosity with a record luminosity peak measured in 2018 [67]. The LHC also operates in a lead-ion beam regime targeting studies of the QCD matter under extreme conditions (the so called quark-gluon plasma), however this subject will not be further discussed in this thesis.

2.2 The Compact Muon Solenoid

Four major experiments are placed at the LHC interaction points (IP):

- The Compact Muon Solenoid (CMS),
- A Toroidal LHC ApparatuS (ATLAS),
- The LHCb detector (LHCb), and
- A Large Ion Collider Experiment (ALICE),

among other smaller experiments. The study case to be discussed in this thesis was carried out over a dataset corresponding to pp-collisions with a center-of-mass energy of $\sqrt{s} = 13$ TeV collected during the Run-2 period using the CMS detector. During this time, CMS recorded 138 fb^{-1} of integrated luminosity relevant for physics studies. An average number of 60 interactions per bunch crossing, or Pile-up (PU), was measured during Run-2 data-taking, each interaction resulting in multiples of particles flying in all directions. Such a high interaction rate was predicted and reflected in the high granularity design of the CMS components, radiation hardness of the detector sub-systems placed in close proximity to the IP and full spacial coverage.

The CMS detector with its cylindrical geometry (more than 21 m long and about 15 m in diameter) is situated around the nominal IP. The CMS uses a standard right-handed coordinate system depicted in Figure 2.2 with its center assigned to the IP. The z-axis is parallel to the beamline direction pointing out of the CMS center (geographically towards the Jura mountain). The y-axis is pointing vertically towards the surface direction while the x-axis points radially inwards the LHC center. Taking advantage of the detector geometry, it is a usual practice to adopt a cylindrical coordinate system (r, ϕ, θ) with r being the radial coordinate, the azimuth angle ϕ measured in xy-plane and the polar angle θ determined w.r.t the z-axis. Particles accelerated alongside the beam axis have their momentum determined mostly by the longitudinal component (p_z) while particles moving in the direction perpendicular to the beam (e.g created in the collision) will have their transverse momenta p_{T} calculated from their x and y momentum components. It is useful to define the rapidity of a particle:

$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_{\text{z}}}{E - p_{\text{z}}} \right), \tag{2.3}$$

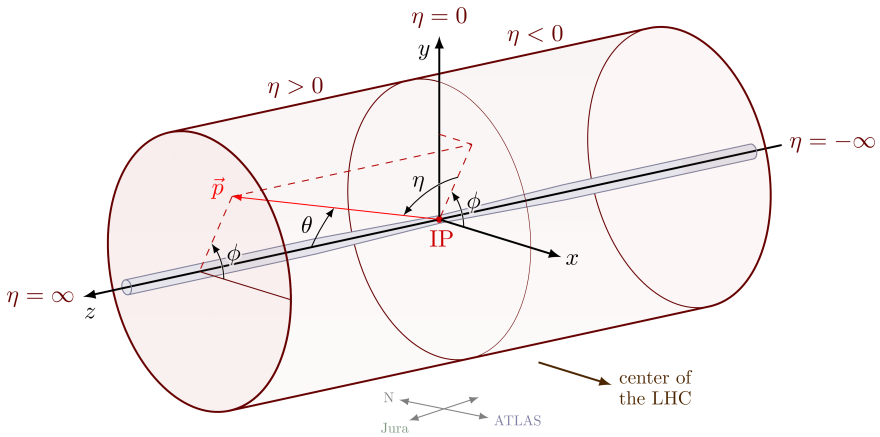

Figure 2.2: CMS cylindrical geometry and coordinate system. Arbitrary particle coming from interaction point (IP) is described by three-momentum vector $\vec{p}$. Source: [68].

as it is equal to zero for particles with $p_z = 0$ and maximal in absolute values for particles moving parallel to the beam direction. It can also be shown that a difference between rapidities of two particles is a Lorentz invariant quantity w.r.t to a boost alongside the z-axis. Given the relativistic behavior of the particles measured in the detector it is even more convenient to define the pseudorapidity

$$\eta = -\ln[\tan(\theta/2)], \tag{2.4}$$

which in terms of high energy limits is equivalent to rapidity y and depends only on the polar angle.

The CMS detector is designed to identify and reconstruct physical objects passing through its detection subsystems and measure their kinematic properties with a state-of-the-art precision. The CMS layout is depicted in Figure 2.3. It consists of several components situated in axially symmetric shell-like layers forming the CMS barrel section while the full detector coverage is assured by enclosing the barrel in end-cap sections on each side. The dominant feature of the CMS detector is a superconducting solenoid generating a homogeneous field of 3.8 T. The presence of a powerful magnetic field is essential to measure the momenta of electrically charged particles as their trajectories are bent under the influence of a magnetic force. Generously large dimensions of the solenoid (12.5 m long with an internal diameter of 6 m) allow to place detection systems inside of its core. Enclosed by the solenoid are an inner tracker designed to measure trajectories of charged particles, a lead tungstate crystal electromagnetic calorimeter and a brass and scintillator hadron calorimeter, both capable of a destructive measurement of the energy deposited by particles primarily interacting electromagnetically and by the strong interaction, respectively. Outside of the solenoid volume is a sophisticated system of muon stations allowing to track muons and measure their momentum with an outstanding resolution. The muon stations are embedded in the outermost layered structure together with $\approx$ 10000 tonnes heavy flux-return yoke helping to contain and form the magnetic field. While these detector systems are sufficient to identify and measure most of the particles passing through, dedicated algorithms (discussed in Chapter 3) need to be applied to account for the elusive nature of neutrinos (or possibly other undiscovered feebly interacting particles) escaping the detector volume.

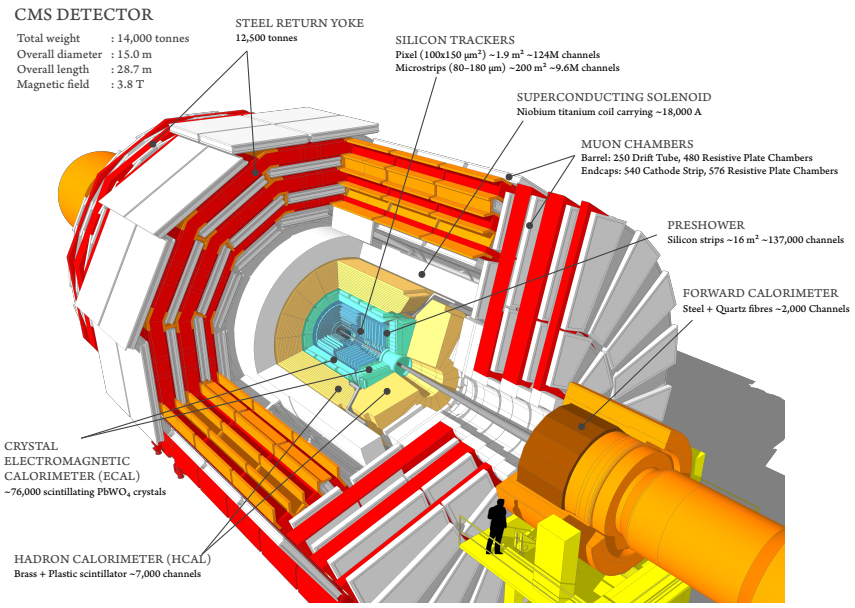

Figure 2.3: Cutaway view of the CMS detector operational during Run-2 period of data-taking. Phase-1 upgrade of Pixel Tracker detector was installed in March 2017. Source: [69].

In the following text we will be discussing individual CMS subsystems in more details via Sections 2.2.1 - 2.2.4. The final Section 2.2.5 is dedicated to the multilevel trigger system and storage solutions deployed in order to handle the vast amount of data acquired by CMS during its operation. The original design of the CMS detector in its completeness can be reviewed in [70].

2.2.1 The Inner Tracker

The tracker is the innermost CMS component situated in the immediate vicinity of the IP. It was constructed to efficiently map trajectories (also called tracks) of electrically charged particles as well as to reconstruct secondary vertices which originate in decays of short-lived particles emerging from the collisions. The tracker system is fully based on silicon semi-conductor technology providing high resolution measurements and fast detector response while considering the extreme operational conditions given the large radiation flux. The CMS tracker, 5.8 m long and 2.5 m in diameter, consists of two main dedicated sub-systems: the silicon pixel detector and the silicon strip detector.

The silicon pixel detector resides at the closest distance from the IP providing excellent spacial resolution, especially for the reconstruction of secondary vertices. The pixel detector initial design (denoted as Phase-0) consisted of 3 concentric barrel layers (BPIX) matching 2 endcap disks (FPIX) on each side. To adapt for the growing values of PU and luminosity, an upgraded Phase-1 version was installed during the LHC YETS 2016/2017 [73]. One extra layer was added in the barrel region (allowing to deliver four spacial points per trajectory) while three disks were presented at each of the endcaps as depicted in Figure 2.5. The BPIX is assembled from two half-barrel structures each consisting of four layers called half-shells covering the range of $[-270, 270]$ mm w.r.t

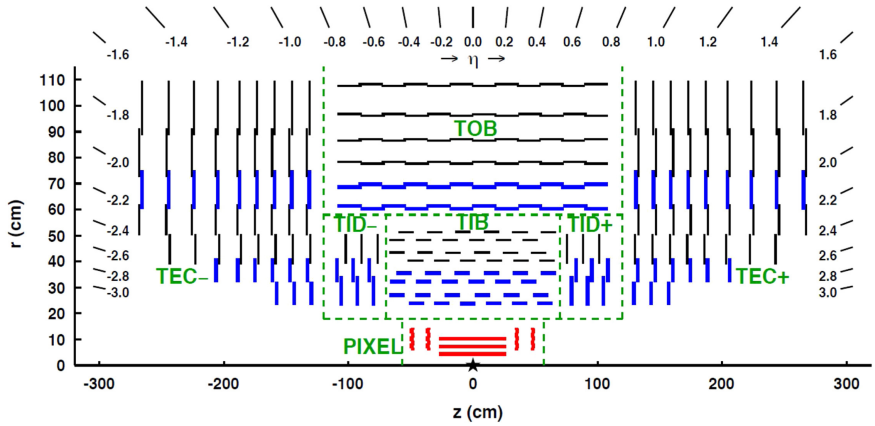

Figure 2.4: One half ($r > 0$) of the CMS Tracker detector layout in the $r - z$ plane prior to Pixel Phase-1 upgrade. Single-sided strip modules are shown in thin black lines, stereo strip modules in thick blue lines and pixel modules are depicted in red lines. Source: [71].

the IP alongside the z-axis. Comparably the FPIX disks, providing radial coverage in $[45, 161]$ mm range, are divided into two half-disks each consisting of multiple inner and outer half-rings. At the end of 2018, the CMS BPIX and FPIX detectors were holding as much as 1856 silicon pixel modules providing the overall pseudorapidity coverage of $|\eta| < 2.5$. The pixel modules are built from pixel sensors bump-bonded to the embedded read-out electronics. The pixel sensors consist of 160×416 cells with dimensions of 100×150 μm^2 and pixel pitch of 100 µm allowing for a resolution between 15 µm to 20 µm depending on the radial position of the module as well as the angle of the measured track.

The silicon strip detector extends the tracker radial coverage from 20 cm to 116 cm, providing at least 9 single point measurements (out of which at least ≈ 4 having an extra second dimension) within the $|\eta| < 2.4$ range. It consists of several independent sub-systems. The silicon pixel detector is surrounded by the Inner Barrel detector (TIB) formed by 4 barrel layers and by the composition of 3 matching Inner Disks (TID) situated from each side. The radial tracker coverage is then finalized by six layers of the Tracker Outer Barrel (TOB) while 9 disks of the Tracker EndCaps (TEC) at each side of the tracker fill up the range between 124 cm and 282 cm in the direction of the z-axis. Altogether, the silicon strip detector consists of 15148 silicon strip modules varying in shape and thickness of the sensors as well as in the definition of the strip pitch leading to the single point resolution as good as 23 µm in the first two layers of the TIB and 35 µm in the last two layers of the TOB. Some of the silicon strip modules contain an extra strip sensor attached back-to-back to the first sensor with a small stereo angle of 100 mrad (stereo modules) allowing for a measurement of the second dimension (z in the barrel and r in the endcaps). One half ($r > 0$) of the CMS Tracker detector with all its sub-systems prior to the Pixel Phase-1 upgrade is schematically displayed in Figure 2.4.

The general tracking principles applicable for both silicon pixel and strip sensors are based on the ability of high energy charged particles passing through layers of semiconductor material to ionize surrounding atoms only loosing a small fraction of their energy. As a result, a number of electrons and electron holes (a concept of missing electron in ionized atoms) is formed in the detector material alongside the trajectory of the energetic particles. Electrons and holes are acting upon the applied electric field

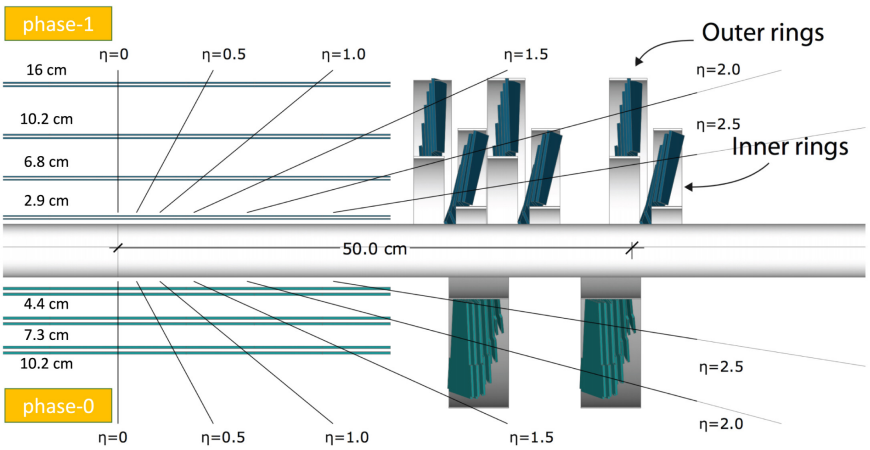

Figure 2.5: Comparison of the original Phase-0 CMS Pixel Tracker detector layout in the $r - z$ plane w.r.t to Phase-1 upgrade. Source: [72].

and travel to the corresponding finely distributed electrodes where the resulting pulse is collected, amplified and read-out by the electronics for further processing by trigger system.

With regard to both the number of sensors and the total active silicon area, the CMS tracker is the largest silicon tracker ever built. Unfortunately, the designed spacial resolution of $\mathcal{O}(10\ \mu\text{m})$ for position measurements is restricted by the maximal mechanical alignment precision of $\mathcal{O}(100\ \mu\text{m})$ achieved during the installation of the tracker. It is therefore necessary to deploy suitable corrections to account for any discrepancy in position or orientation of the modules or possibly larger segments of the detector. A dedicated alignment procedure and its further validation is performed at multiple campaigns during and before the actual data-taking. My personal contribution to this procedure and evaluation of the tracker performance during Run 2 is presented in Appendix A and has been summarised in publication [74].

The upcoming high luminosity upgrade of the LHC (HL-LHC), intended to start its operation around 2029, is aiming for a challenging 3000 fb^{-1} of integrated luminosity to be collected by the CMS detector which is more than ≈ 18 times the total integrated luminosity acquired during Run 2 and ≈ 10 times more than what is planned for the Run 3 period starting in 2022. Several CMS components including the CMS Tracker have to be adapted accordingly. My personal contribution to the Phase-2 CMS tracker upgrade efforts is summarised in Appendix B.

2.2.2 The electromagnetic calorimeter

The CMS Electromagnetic Calorimeter (ECAL) is a high resolution detector component capable to measure energy deposits left by electrons and photons emerging from the collision or as a product of decaying secondary particles. The ECAL is constructed from $75,848$ lead tungstate crystal ($PbWO_4$) scintillators mounted in modules homogeneously situated around the inner tracker system where it is constantly affected by the magnetic field of 3.8 T. The particular material choice is motivated by the ability of the incoming energetic electrons or photons to loose most of their energy by the radiative effects. An electron with its trajectory bent by the magnetic field will be decelerated, in most of the

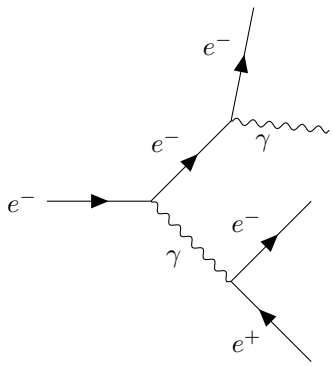

Figure 2.6: Example of electromagnetic shower propagating through scintillator material.

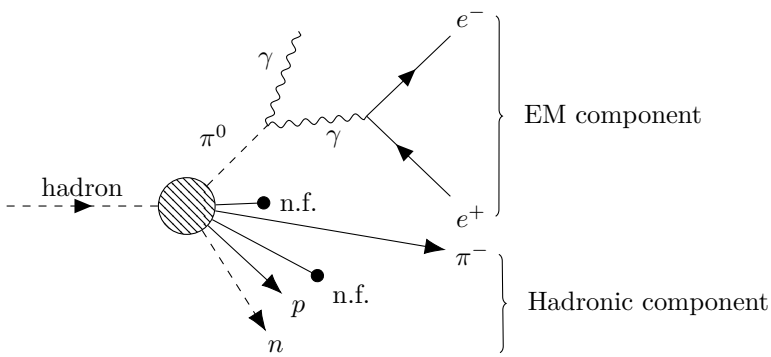

Figure 2.7: Scheme of hadronic shower propagating through brass absorber. Neutral hadrons (dashed arrow), charged hadrons (solid arrow) and nuclear fragments (circled arrow) are forming pure hadronic component. Neutral π^0 meson triggers propagation of electromagnetic component of hadronic shower.

cases, by the emission of a "breaking photon" – process known as the Bremmstrahlung. On the other hand, a photon with energy $E_\gamma > 2m_e$ ($E_\gamma > 4m_e$) will most likely interact with the electric field of an atomic nucleus A (shell electrons) by the production of electron-positron pairs:

$$\gamma + A \to A + e^- + e^+. \tag{2.5}$$

While a Bremmstrahlung photon can cause additional e^-e^+ pair production, the electron from this pair is again affected by the radiative losses and the positron will inevitably annihilate with another electron resulting in two photons

$$e^- + e^+ \to \gamma + \gamma, \tag{2.6}$$

which again will cause another e^-e^+ pair production and so on. Multiple consequent iterations of the aforementioned processes will form a so called electromagnetic shower as pictured by the diagram in Figure 2.6. The electromagnetic shower will eventually stop when the energies of the photons are below the threshold for e^-e^+ pair production and the energies of the electrons drop below the threshold of ionisation effects prevailing over radiative effects.

The ability of the material to propagate electromagnetic showers is given by two characteristic constants: the radiation length X_0 and the Moliere radius R_{M}. The radiation length is defined as the mean distance over which the electron energy drops to the $1/e$ value of its initial energy by Bremmstrahlung (e being the Euler number), while the Moliere radius is defined as the radius of a cylinder that contains 90% of the whole electromagnetic shower energy deposit. Short radiation length $X_0 = 0.89$ cm and a small value of the Moliere radius $R_{\mathrm{M}} = 2.2$ cm make $PbWO_4$ the appropriate choice for the ECAL. Another reason to employ $PbWO_4$ is its ability as an inorganic crystal to provide a rapid scintillation response.

Scintillation is a phenomenon where particles, such as those contained in an electromagnetic shower, excite electrons inside of the atoms forming the crystal grid of the material. The excited atoms will eventually relax back to their ground state followed by the emission of a photon with a characteristic wave length in the visible light spectrum (blue-green range of colors in the case of $PbWO_4$). In this regard, a rapid scintillation response means that 80% of the emitted light will be collected at the end of crystal in about 25 ns, which is the time-span between two bunch crossings at the LHC.

The lead tungstate crystals in the ECAL barrel region (EB) are situated in modular structures providing a pseudorapidity coverage of $|\eta| < 1.48$. Two endcaps (EC) are enclosing the ECAL volume on each side, expanding its coverage to $|\eta| < 3.0$. The particular geometrical configuration in the transverse plane section of the ECAL is depicted in Figure 2.8 showing one ECAL Barrel (EB) "supermodule" (with 4 modules, each consisting of several sub-modules, each containing exactly 5 crystals of the same shape) and a section of one half of the EC (with several "supercrystal" structures made of crystals in a 5x5 group configuration). The EB crystals are made 230 mm long, corresponding to $\approx 26X_0$, and their front face cross-section is equal to 22×22 mm^2. The EC crystals are slightly shorter with 220 mm in length ($\approx 26X_0$) possessing a bit larger front face area of 28.6×28.6 mm^2. All crystals are oriented in a space such that their length is describing approximately the same pseudorapidity values. The relatively small scintillation light yield of the $PbWO_4$ crystals requires an amplification of the signal from the incident photons before the signal can be converted into an electrical pulse for further processing. This is achieved by photodetector mounted on the back side of each crystal (further from the CMS center) which is then connected to the CMS trigger system.

An additional detection element, the CMS Preshower detector (ES), was installed in front of the ECAL endcaps targeting identification of neutral pions π^0 predominantly decaying into a pair of photons. Secondary photons coming from the π^0 decay can be potentially mistaken for energetic photons emerging from the collision. This effect is reduced by adding two extra layers of lead radiator ($2X_0$ and $1X_0$ thick) invoking the production of electromagnetic showers, each followed by a plane of silicon strip detectors allowing to measure shower energy deposits. Moreover, two planes of silicon strip detectors have their strips oriented mutually orthogonal, allowing to study transverse profiles of the electromagnetic showers. The CMS Preshower is covering the forward pseudorapidity region of $1.65 < |\eta| < 2.6$, improving the otherwise relatively low spacial resolution given by the small granularity of EC crystals.

2.2.3 The hadron calorimeter

The CMS Hadron Calorimeter (HCAL) provides a complementary energy measurement for all charged and neutral hadrons which undergo inelastic scattering processes with atomic nuclei in the absorber material. Depending on the cross-section of such an interaction, the incident hadron triggers a chain production of secondary particles, e.g. ad-

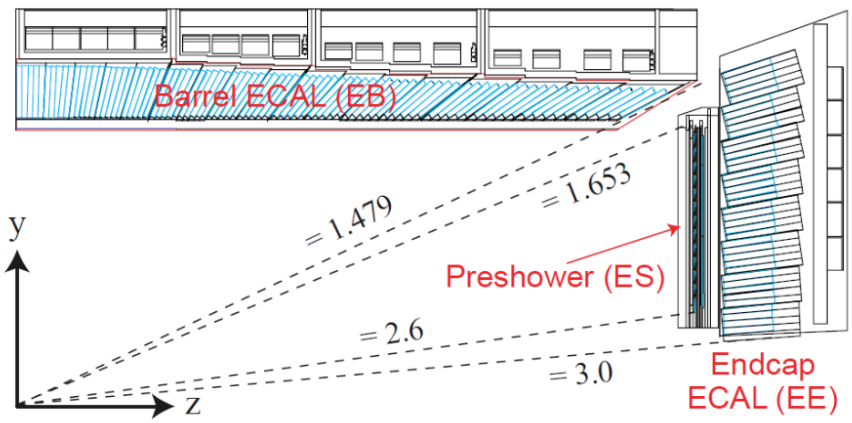

Figure 2.8: Transverse plane section of the CMS electromagnetic calorimeter. Source: [75].

ditional energetic hadrons, low-energy nucleons or excited atomic nuclei. Many of these particles cause tertiary interactions, decay and possibly undergo nuclear de-excitation or fission, contributing to the propagation of the hadronic shower. Given the underlying principles of the strong interaction, hadronic showers tend to expand to much larger dimensions compared to electromagnetic showers. In fact, the hadronic interaction length λ_h, the quantity typically used to describe the profile of the hadronic shower (as it was in case with radiation length X_0), rises with the nucleon number A for a given material as

$$\lambda_h \approx A^{\frac{4}{3}} X_0. \tag{2.7}$$

This argument places a crucial requirement on the material selection and depth of the active HCAL volume. The complex shape of hadronic showers often contains a short-profiled electromagnetic sub-shower component triggered by the $\pi^0 \rightarrow \gamma\gamma$ decay as schemed in Figure 2.7. A more compact energy distribution, typical for the electromagnetic sub-shower, requires a calorimetry system with a higher granularity of detection layers as compared to the pure hadronic component. Most of the energy deposits by secondary particles is then directly or indirectly detectable by means of ionisation losses. On the other hand, a significant amount of energy is "consumed" by the nuclear recoil, excitation or escapes in form of the neutrinos, inevitably affecting the measurement resolution.

The main HCAL subsystems are placed within the CMS solenoid volume, therefore restricting construction options to non-magnetic materials. Both, the HCAL Barrel (HB) and the HCAL Endcap (HE) are formed from interchanging layers of brass absorber, acting as inactive showering medium, and plastic scintillator layers, serving as active measurement components. The overall cross-section of the HCAL layout is depicted in Figure 2.9. The HB consists of two half-barrel sub-structures situated alongside the z-axis, each constructed from 18 wedges corresponding to azimuth $\phi = 20°$ and providing a total pseudorapidity coverage of $|\eta| < 1.39$. Following the same ϕ geometry, HE consists of 18 sectors hermetically enclosing the inner solenoid volume from both sides and extending the HCAL coverage to the overlapping range of $1.30 < |\eta| < 3.00$. The HB wedges and HE sectors are designed to accommodate brass layers, ranging approximately $5-8$ cm in thickness, and about 0.4 cm thick layers of plastic scintillator. Exceptionally,

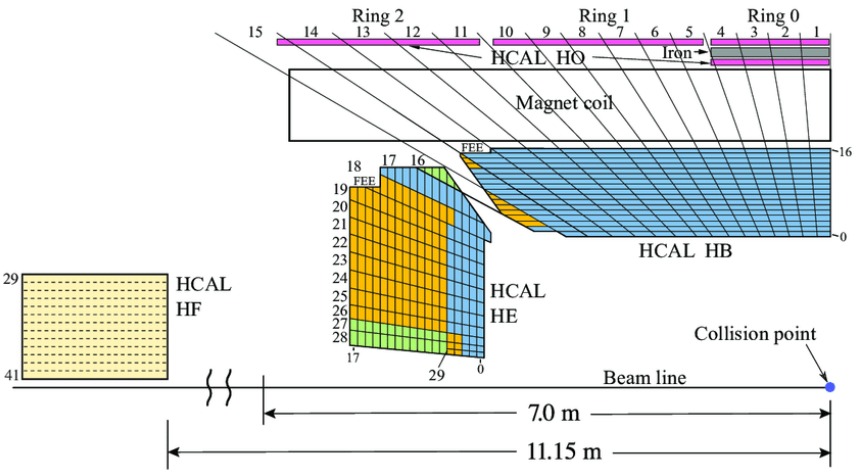

Figure 2.9: One quarter section of the CMS Hadronic Calorimeter layout. Source: [76].

the innermost and outermost HB layers are made of ≈ 0.9 cm stainless steel to enhance structural integrity of the barrel section.

The HCAL Outer Barrel (HO) is an additional detector component installed outside of the CMS solenoid to improve the efficiency of the energy measurements for incident hadrons escaping the inner calorimetry volume. It is divided to 5 rings with $\eta - \phi$ segmentation consistent with the HB. Each ring is formed of a 1 cm thick scintillator layer with an exception for the central ring which possesses two scintillator layers on both sides of an extra stainless steel absorber block. The HO component covers for $|\eta| < 2.6$ region.

The remaining pseudorapidity range (up to $|\eta| \approx 5.2$) is covered by the HCAL Forward Calorimeter (HF) placed at ≈ 11 m from the IP alongside the z-axis in both directions. Its presence significantly reduces the amount of undetected particles that would otherwise be adding up to the reconstructed transverse missing energy. The HF is designed to sustain extreme impact rates typical for a very forward region close to the beam. It consists of grooved absorber plates made of steel with inserted radiation hard quartz fibers serving as active component. Instead of collecting light emitted by scintillator, the HF detection principles are based on Čerenkov radiation induced by charged particles traversing quartz material with a speed higher than speed of light in this material. Apart from speed, the intensity of Čerenkov radiation depends on the angle w.r.t to the trajectory of incident particles, allowing to reconstruct their kinematic properties.

2.2.4 The muon systems

Contrary to the approximately 200× lighter electron, the muon is much less susceptible to radiative losses and unlikely to trigger Electromagnetic (EM) cascades in the ECAL absorber material. The leptonic nature of muons also reduces the possibility of the strong inelastic interactions with the HCAL layers. Besides the occasional hits in the CMS Tracker, marking its initial trajectory, the muon tends to leave the solenoid volume. A system of muon stations (MS) [77], [78], embedded in the magnetic field return-yoke was designed to identify escaping muons and measure their kinematic properties using three types of gaseous detectors. In general, this technology is based on the ionisation of gas

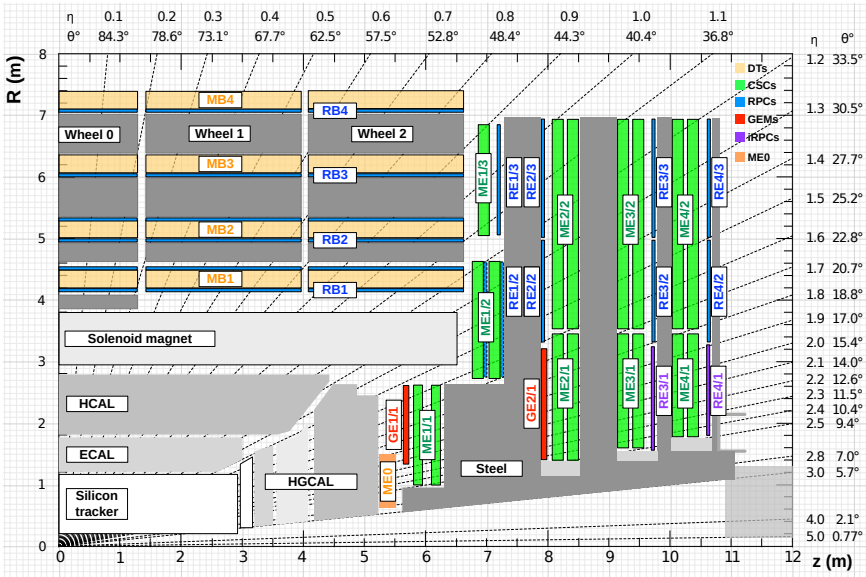

Figure 2.10: Layout of the CMS Muon detector subsystems in preparation for the Run 3 period of data-taking. Drift Tubes (DT) in the barrel section is displayed in yellow, Cathode Strip Chamber (CSC) in the endcap section is shown in green, blue is used for Resistive Plate Chambers (RPC). Components installed during LS 2 (with a small exception for Gas Electron Multiplier (GEM) detectors in GE1/1 section installed in 2017) are displayed in red, orange and purple. Source: [78].

atoms by energetic muons passing through gas chambers stationed with high granularity. Electrons knocked out of the gas atoms are placed in an electric field and drift to the positively charged cathodes. The accumulated electric charge is collected and amplified to form a detectable signal. The measurement of the drift time needed for electrons to reach the cathode is then translated into precise information about the muon's position and direction of flight. Multiple self-operating chambers placed at constant distance from the beamline (barrel) or interaction point (endcap) are commonly denoted as muon stations.

One quarter of the CMS muon system layout is displayed in Figure 2.10. The MS barrel section (MSB) consists of 4 concentric cylinder stations situated at $4-7$ m radius R from the beamline and covering the $|\eta| < 1.2$ range. Each station is equipped with up to 12 layers of rectangular DT filled with a mixture of $Ar(85\%)/CO_2(15\%)$ gas, $1.9-4.1$m long (depending on R) and with a transverse cross-section of 42 mm^2. Located in the center of DT is a negatively charged anode wire made of stainless steel creating a potential difference with the cathode plates at the DT walls and between layers. Maximum drift time values for this setup reach $\approx$ 400 ns. The DT are placed both horizontally and vertically (w.r.t beam axis) to provide measurements of the $R-\phi$ and $R-z$ coordinates, respectively. The spacial resolution for the DT subsystem depends on the particular gas chamber position and is of order of ≈ 250 µm.

The Muon System (MS) endcap section (MSE), ranging $0.9 < |\eta| < 2.4$ in pseudorapidity, is affected by a much higher muon rates and heterogeneous magnetic field (compared to the MSB), hence it required a muon subsystem with a minimal response time and fine segmentation. In that regard, four disk-shaped sections consisting of CSC

detection technology are installed on both sides of the CMS detector perpendicular to the beam axis. The gas mixture employed in this case consists of 40% Ar, 50% CO_2 and 10% CF_4. Each CSC is constructed from 6 layers of multi-wire proportional chambers providing readout from cathode strips placed radially outwards of the detector at constant ϕ values and anode wires placed perpendicular to the cathode orientation with very fine ≈ 3 mm spacing. By construction, muon hits in the CSC subsystem are measured as two-dimensional space points compared to the one dimension obtained by the DT subsystem. The resolution achieved by the CSC varies with the non-uniformity of the magnetic field within $40 - 140$ µm range.

The output from DT and CSC stations contributes to the offline reconstruction of muons trajectories as well as serves for the purpose of triggering in order to select events with muons above a certain p_{T} threshold. To improve triggering capabilities, the MSB and MSE sections are both equipped with extra RPC focused on precise bunch crossing time measurements. The RPC stations are placed in close vicinity with corresponding DT and CSC stations following the same barrel and endcap geometry and providing total pseudorapidity coverage of $|\eta| < 1.9$. Each RPC station consists of two pairs of parallel Bakelit plates, each pair forming a 2 mm wide leak-proof gap filled by water vapors and a mixture of Freon (95.2%), isobutan (4.5%) and sulfur hexafluorid (0.3%) gases. The RPC station is operated at the avalanche mode with high-voltage applied on the electrodes attached to the outer surface of Bakelit plates coated with a thin layer of conductive graphite. The signal from avalanches caused by a muon entering the gas gaps is collected by a readout strip placed in between of two pairs of Bakelit plates. The overall time resolution (including the propagation of the signal in readout strips and processing by electronics) achieved by a standalone RPC setup was estimated to < 3 ns.

For the sake of completeness, the first segment of the new detection subsystem denoted as GEM [79] was installed in the MSE section during the LHC YETS 2017/2018 and was operating since then (for testing purposes only). Other components, including the rest of GEM stations, improving capabilities of the muon and other CMS subsystems are planned towards the Phase-2 upgrade intended for future Run-3 and High Luminosity LHC (HL-LHC) operations.

2.2.5 Trigger system and data storage

The LHC interaction rate of 40 MHz by far exceeds the CMS allowed bandwidth levels. Moreover, the majority of the LHC collisions is coming from processes unlikely to disclose new physics, therefore disregarded due to the small importance to the CMS physics program. In that regard, CMS possesses a dedicated trigger system serving as an online filter for the "events" (all activity presumably originating from a single interaction) showing physics-relevant features. The CMS trigger system is two-tiered. It consists of a firmware-based Level-1 trigger (L1) initially reducing the event rate to 100 kHz and a High Level Trigger (HLT) operating on a software basis and further decreasing the event rate to a bearable 1 kHz.

The L1 trigger uses limited information from both the ECAL and HCAL calorimeters as well as the CMS Muon system. It takes about 3.2 μs for the signal to travel from the detector front end electronics to the L1 trigger processors and back. Out of this period, no more than 1 μs is dedicated to the calculations needed for a prompt decision to catch up with the high bunch-crossing rate. An event is accepted or rejected based on the characteristic energy deposits and detector hits correlated with a detector trace typically left by physical objects such as electrons, muons, tau leptons, photons or "jets" (sprays of the hadronic particles coming from the IP). Physics-relevant information is

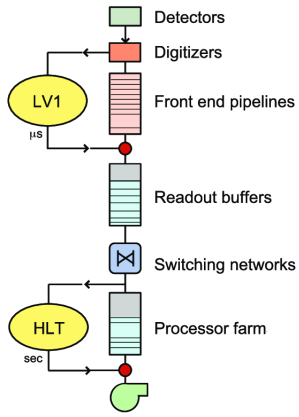

Figure 2.11: Scheme of the CMS trigger system. Source: [81].

then selected based on a certain number of these objects passing required energy or momentum thresholds, requirements on a global sum of transverse energy E_{T} or a total transverse energy missing in the detector $E_{\mathrm{T}}^{\mathrm{miss}}$. Dedicated algorithms to identify physical objects with limited L1 trigger information are summarised in [80].

Until the decision is made by the L1 trigger, the complete event information is stored in a buffer, awaiting to get dumped or allowed for further processing by HLT trigger. More than 400 different HLT paths (sequences of algorithms and selection requirements) were in use during the Run-2 period of data-taking. The algorithms employed by HLT software resemble those used in an offline reconstruction, given the completeness of information received from detector. On the other hand, reconstruction of physical objects and tracks is restricted to the region in close vicinity of the L1 objects to achieve a compromise with CPU time limitations of the HLT processor farm. Relevant HLT paths required for this analysis will be enlisted in Chapter 4.

The chain of the CMS trigger system is schematically depicted in Figure 2.11. After the HLT decision is made, raw data containing full and detailed information about the events (including detector hits, accumulated charge, etc.) is stored at the CMS Tier-0 (T0) computing center located at CERN. At the T0, available detector information is used for data calibration and full reconstruction of physical objects. Reconstructed datasets are additionally processed to limit their size and stored at Tier-1 and Tier-2 computing centers situated at various places around the world.

CHAPTER

THREE

DATA SIMULATION AND RECONSTRUCTION

The nature of the High Energy Physics (HEP) experiments at the LHC requires a deep understanding of the processes involved in proton-proton collisions. To ensure that our observations align with the SM expectations, or conversely, show signs of new physics, it is common practice to compare measured data with simulated events. Prerequisites for such simulation involve a good knowledge of scattering processes occurring at the parton level as well as the full information about the CMS response to particles passing through the detector volume. This is achieved by means of the Monte Carlo (MC) method originally developed by J. Neumann and currently employed by numerous MC generators as an essential part of the HEP data analysis [82].

The CMS Collaboration has a hierarchical structure allowing the sharing of the analysis workflow among several levels of management and working groups. The CMS MC group centrally produces ensembles of the LHC simulated events (also called the "MC samples") for most of the known SM and BSM processes categorized by defining physics assumptions and/or final state event topology. To reflect on the CMS-specific analysis requirements, the MC generators are configured and sometimes interfaced with the CMS SoftWare (CMSSW) [83] developed in-house. In Section 3.1, we discuss most of the general aspects and ingredients needed in the MC simulation. For the particular choice of the MC samples considered in this analysis, we refer to Chapter 4.

Low-level information, e.g. tracker signal ("hits") and energy deposits in calorimeter cells, retrieved from the CMS detector readout, is not always suitable for standard analysis procedures. Therefore, this information is translated into high-level physical objects, such as electrons or jets, using dedicated reconstruction algorithms approved by the CMS Physics Object Groups (POGs) and discussed in Section 3.2. The selection and strictness of proper reconstruction settings are often subject to the final tuning depending on particular analysis needs. The decisions made in the context of this thesis and their justification will also be discussed, with a special focus on the reconstruction of jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ objects as I had significantly contributed to this effort.

To complete the entire analysis chain, the CMS Physics Analysis Groups (PAGs) select physical topics suitable for studies and oversee dedicated working groups performing statistical analyses and interpreting results w.r.t theory and phenomenological models.

3.1 Monte Carlo simulation

The MC generators build upon the idea that a physical system can be described by probability density functions. The non-analytic nature of the MC method implies that event simulation does not suffer from the large dimensionality of the parametric phase space typical for LHC collisions with a high bunch-crossing rate and hundreds of secondary particles at stake. For the convenience of the reader, the general concept of the MC integration method is briefly summarised in Section 3.1.1.

Most of the signal and background processes relevant to the MC simulation, such as the production of heavy resonances (Higgs boson, W, etc.), originate from the hard scattering of proton constituents (quarks and gluons). The high energy scale (or alternatively, the short distance scale) is a property defining hard scattering processes and is manifested by high momentum transfer between colliding partons. Thanks to the nature of strong interactions, the magnitude of the coupling strength g_S (in the literature also re-parametrized as $\alpha_S = g_S^2/4\pi$) is steadily reduced towards high momentum exchange which allows to treat hard scattering in the Perturbative QCD (pQCD) regime.

Hard processes might provide a foundation for successful event generation, however, partons inherently carry color charge, which significantly complicates simulation in a two-fold way. First, analogous to the QED, where an accelerated electron loses part of its energy by radiating photons, scattered partons are prone to radiate energetic gluons driven by the QCD principles. Gluons, contrary to electrically neutral photons, also carry color charge and can themselves radiate other gluons, contributing to the evolution of the PS. Secondly, as the energy scale of the PS evolves towards low energies, partons become affected by color confinement and start to form (often unstable) colorless hadrons. At energies below 1 GeV it is no longer possible to employ the perturbative approach, therefore processes like "hadronisation" need to be treated by non-perturbative models. In fact, the majority of processes in proton-proton collisions happens at a low energy scale, classified as soft scattering. Another example of scattering in the soft regime is the mutual interaction of incoming partons not originally involved in the hard scattering. When such a process occurs during the same bunch-crossing, it is significantly affecting the outcome of the studied hard process and contributes to the effect commonly denoted as the Underlying Event (UE).

In the following sections, we will address each of the aforementioned processes in more detail. An example of a typical event structure is schematically visualized in Figure 3.1. Finally, the MC simulation of the detector response and the propagation of secondary particles through the detector material is discussed.

3.1.1 General principles

The MC method can be essentially viewed as a numerical integration technique employed by means of generating pseudo-random numbers. Schematically, the problem of the MC generation can be narrowed down to the integration of the probability density $f(x)$, where x represents an n-dimensional vector corresponding to n particles (and their kinematic properties) distributed randomly in the phase space volume V. In other words, when randomly generated x values fall within the $f(x)$ domain (bounded by V), the event is accepted, and vice versa. The particular form of the function $f(x)$ depends on multiple factors, such as matrix elements or parton distributions, and will be discussed in the following sections. Meanwhile, we generally suppose the integration result

$$I_n = \int_V f(x)\ \mathrm{d}^n x \tag{3.1}$$

and repeat the integration for N events (each with a different set of values $x_i, i = 1, ..., N$). Out of the principles of the central limit theorem of statistics, an unbiased estimate of the I_n value can be found as the mean value of the function $f(x)$ (multiplied by the volume V), such as

$$I_n \hat{=} V \langle f(x) \rangle = \frac{V}{N} \sum_{i=1}^{N} f(x_i). \tag{3.2}$$

An important aspect of this result is that the estimated error for this evaluation does not depend on the dimensionality of the phase space n, but rather on the number of generated events N. In particular

$$\mathrm{Err}_n \hat{=} V \sqrt{\frac{\mathrm{Var}[f(x)]}{N-1}}, \tag{3.3}$$

where $\mathrm{Var}[f(x)] = \left\langle f^2(x) \right\rangle - \langle f(x) \rangle^2$ stands for the variance of $f(x)$. Notably, the estimated error for the estimated value of the $f(x)$ integration decreases as $\approx \frac{1}{\sqrt{N}}$. This is true when generating "unweighted" events. In practice, events might be generated with weights w_i different from 1 to account for the MC efficiency in event selection and/or additional weights might be calculated a posteriori to reflect on effects not considered initially and causing discrepancies between data and the MC simulation.

3.1.2 Hard scattering

The important realisation in simulating hard processes is that high momentum transfer Q between scattering partons allows us to treat those partons as asymptotically free QCD energy quanta (as α_S becomes small), thus making the application of perturbative approach possible. Let us further elaborate on this statement.

As previously stated, pQCD is hitting its limitations when accounting for soft gluon radiation. Moreover, it can be shown that collinear gluon radiation (defined by a small angle between parton and radiated gluon) causes another sort of logarithmic divergences in (parton scattering) cross-section calculation. Fortunately, the latter issue gets resolved in the so called KLN theorem [85, 86], proving that collinear divergences disappear in the calculation of an inclusive proton-proton cross-section when accounting for virtual corrections, individually per each order in the perturbative approach. The full independence of hard scattering subprocesses from soft gluon radiation is then assured by applying the "factorisation theorem" [87], such as an inclusive cross-section $\sigma_{pp\to X}$ for proton-proton scattering can be factorised as

$$\begin{aligned} \sigma_{pp\to X} &= \sum_{i,j} \int_0^1 \mathrm{d}x_i \, \mathrm{d}x_j \int f_i^{p_{(1)}}(x_i, \mu_F) f_j^{p_{(2)}}(x_j, \mu_F) \, \mathrm{d}\hat{\sigma}_{ij\to X}(\mu_F, \mu_R) \\ &= \sum_{i,j} \int_0^1 \mathrm{d}x_i \, \mathrm{d}x_i \int f_i^{p_{(1)}}(x_i, \mu_F) f_j^{p_{(2)}}(x_j, \mu_F) \frac{1}{2\hat{s}} \left|\mathcal{M}_{ij\to X}\right|^2 (\Phi_X, \mu_F, \mu_R) \, \mathrm{d}\Phi_X \end{aligned} \tag{3.4}$$

where

- $x_{i,j} \in (0, 1)$ is a fraction of parton momentum inside of the scattering proton $p_{(1,2)}$ such as one can write $\vec{p}_{i,j} = x_{i,j}\vec{p}_{(1,2)}$, with indices i, j running over all parton types in the initial state protons.

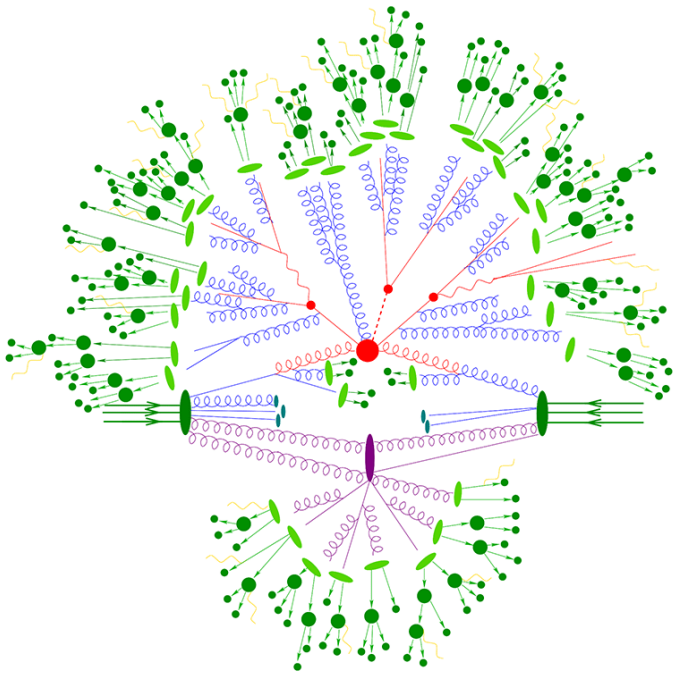

Figure 3.1: Schematic view of the hard scattering process involving a neutral scalar particle, two quarks and their decay products in the final state (red). This process takes place during a proton-proton collision at the LHC. The initial state, as well as the final state partons undergo parton showering (blue), which eventually leads to the hadronisation (light green), followed by hadronic decays (dark green). Any charged particle can sustain EM radiative losses (yellow). The original partons not involved in the hard scattering undergo secondary interactions known as the underlying event (purple), while some of the partons might cause additional pile-up activity outside of the event's time frame (cyan). Source: [84].

- μ_F and μ_R stand for the factorisation and renormalisation scales, respectively. Their purpose is to draw a cut-off line in parton momentum transfer spectra to prevent it from divergences coming from soft radiation (μ_F affects all terms in the formula) and possible divergences from virtual particles in higher order Feynman loops (μ_R enters the calculation of matrix elements). Nevertheless, the final result should not a priori depend on μ_F and μ_R values, as those are chosen arbitrarily to account for fixed order corrections in the perturbation series and would fully disappear if we could perform an infinite sum over all orders. Reasonable choice for a hard scale in question would be variables such as momentum transfer Q, center-of-mass energy squared s or other Lorentz invariant variables that are capable to properly describe the scattering event. In case of hard processes with up to two particles in the final state X, it is a usual practice to assign $\mu_F^2 = \mu_R^2 = Q^2$ or $\mu_F^2 = \mu_R^2 = \hat{s} = x_i x_j s$. To estimate systematic uncertainties driven by an arbitrary scale choice, μ_F and μ_R are varied by a factor of 2 up and down.

- The Parton Distribution Functions (PDF) $f_{i,j}^{p_{(1,2)}}(x_{i,j}, \mu_F)$ parameterise the prob-

ability of finding a parton of type i, j and with momentum fraction $x_{i,j}$ in the final state hadron(s). Unfortunately, it is not possible to find their explicit form analytically. Instead, the PDF are extracted from a global fit on data accumulated from numerous scattering experiments involving protons. In fact, the PDF highly depend on the internal proton dynamics (partons can exchange virtual gluons or quark-antiquark pairs) and with currently available computational tools can be evolved up to the NNLO in the QCD. Furthermore, the PDF are varied w.r.t factorisation scale to provide the best MC-to-data agreement – an example of $\mu_F^2 = Q^2$ dependence is depicted in Figure 3.2 for $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right). The PDF sets employed in the MC simulation on samples used in this analysis are provided by the NNPDF Collaboration. In particular, the MC samples corresponding to 2016 dataset are using NNPDF3.0 [88] version of distribution functions while NNPDF3.1 [89] version was used for simulation in 2017 and 2018.

- The Parton-level cross-section $\hat{\sigma}_{ij\to X}$ depends on both μ_F and μ_R scales. It is obtained by integration over an element of phase space $\mathrm{d}\Phi_X = \mathrm{dLIPS}_n$ (see Section 1.3.6) corresponding to the n particles in the final state X. Thanks to the factorisation theorem, $\hat{\sigma}_{ij\to X}$ does not hold information about the parton distribution, which narrows problem down to finding corresponding parton-level Matrix Elements (ME) squared $|\mathcal{M}_{ij\to X}|^2$ (and averaged over spin and color initial states in accordance with general discussion in Section 1.3.6). Whenever possible, the ME squared are evaluated for the desired order in perturbation series, e.g. by applying the Feynman rules. On the other hand, this task becomes increasingly cumbersome with more than 2-3 particles in the final state, especially at the level of summation over possible spin states and quantum numbers. Thus, an algebraic approach is often limited for the simplest low order diagrams while numerical summation and MC integration is preferred for processes with higher final state multiplicities.

Some of the available MC generators are designed to tackle individual points in the above list – from evaluation of the ME through integrating over the correct PDF for each point in the phase space towards successful simulation of hard scattering processes. Others are specialized for the incorporation of parton shower effects, hadronisation or other soft regime processes. The final MC event is a result of multiple MC generators that are mutually compatible and can be interfaced on top of each other in a logical order from those simulating hard processes up to the soft edge of the energy scale and final state particle decays. As an example, most of the SM samples in this analysis have their hard scattering part simulated by POWHEG [90, 91, 92] or MADGRAPH5_aMC@NLO [93] software with a precision up to NLO in QCD. In case of the BSM signal processes, we take advantage of JHUGEN [94] MC generator sensitive to spin and parity properties of heavy resonances.

3.1.3 Parton showering and hadronisation

Any calculation performed to simulate hard scattering processes is of the fixed order in pQCD. To simulate an effect of the remaining higher orders, MC generators employ dedicated PS algorithms capable to describe event evolution from the high energy scale down to the low-Q threshold typical for soft scattering. Both incoming partons as well as partons created in the hard scattering can trigger QCD radiation contributing to the evolution of the PS. Depending on the place of parton splitting, PS are recognised as of the Initial State Radiation (ISR) and Final State Radiation (FSR) origin, respectively.

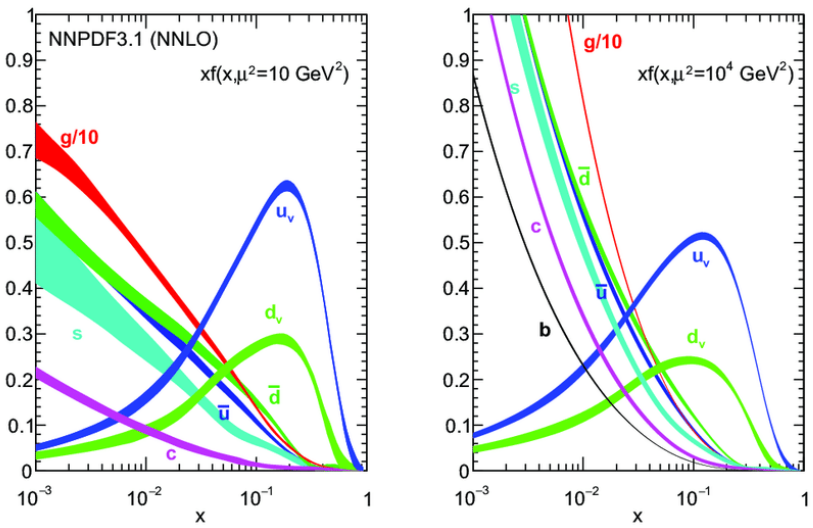

Figure 3.2: NNPDF3.1 version of the parton distribution function in dependence on the parton momentum fraction x and for two fixed values of the factorisation scale $Q^2 = 10$ GeV^2 (left) and $Q^2 = 10^4$ GeV^2 (right). The PDF are multiplied by the value of x for the visualisation reasons and displayed for each parton types with uncertainties depicted by the width of the line. Source: [89].

Similar to the hard scattering case, matrix elements describing parton radiation suffer from collinear and soft divergences. These are again resolved by means of a factorisation scheme with the PS cut-off scale introduced to assure finite cross-section values.

A very instructive is an iterative factorisation scheme implemented for the PS in the collinear limit $\theta \to 0$, where θ is the splitting angle between radiated and parent partons. Consider that hard scattering process $pp \to X$ producing final state partons of general type f has a cross-section $\sigma_{pp\to X}$ and is now accompanied by an extra parton a, e.g. radiated gluon. In good approximation, the cross-section for such a process can be factorised as a product of the hard scattering and parton radiation terms, such as

$$\sigma_{pp\to aX} \approx \sigma_{pp\to X} \sum_f \int \frac{\alpha_S}{2\pi} \frac{P_{af}(z,\phi)}{\theta^2} \, \mathrm{d}\theta^2 \, \mathrm{d}z \, \mathrm{d}\phi, \tag{3.5}$$

where z is a momentum fraction taken by parton a and $P_{af}(z,\phi)$ is a splitting function for a process $f \to f + a$ (see [82] for the explicit form) that generally depends on the parton flavor and spin (through azimuth ϕ). Formula 3.5 is typically used to simulate one collinear splitting for the FSR parton which can later undergo another hard scattering and possibly another collinear splitting, etc. Iterative pattern is repeated until the PS cut-off factorisation bound is reached.

Approaching the end of the PS evolution, momentum transfer eventually drops to $Q <$ 1 GeV, making perturbative approach inapplicable. At this stage, the soft regime QCD becomes strongly interactive, forcing partons to form colorless hadrons in the process of hadronisation. Several non-perturbative models are employed to describe this process. One of the most illustrative ones [95] depicts the interaction between two relatively close

partons $a - b$ as an elastic string with one parton at each side. If the string is pulled significantly apart (in analogy to a large distance scale in the soft regime), the string breaks into two new strings $a - a', b' - b$, each with a new parton attached. Production of new primary hadrons ends when any of the two original partons from the PS is too far apart (> 1.5 fm) to get influenced. PS evolution and hadronisation modeling is typically handled by PYTHIA [96] MC generator.

3.1.4 Underlying event and pile-up

It is experimentally observed that all the hard scattering processes are accompanied by additional activity that does not originate in ISR, FSR, or hadronisation and is summarily denoted as the UE. Most of the UE activity is coming from remaining partons that do not participate in hard processes, however they interact in a rather soft regime. In terms of the MC simulation, UE is treated by the Multiple-Parton Interaction (MPI) model, which in its first approximation allows for additional parton scattering of $2 \to 2$ type. In fact, at this hard tail of the UE spectrum, a perturbation approach is still possible and typically yields two back-to-back oriented reconstructible jets with a small total transverse momentum of the di-jet system. Much more often, secondary partons as well as beam remnants (partons leaving the hard scattering event almost intact) will interact by exchanging color charge and only a small amount of momentum, with no visible jets reconstructed, thus significantly affecting the soft regime production of final state hadrons. Similar to hadronisation, the soft part of the UE requires dedicated non-perturbative models and is also addressed by PYTHIA.

While the UE describes additional interaction per the same proton-proton collision, further simultaneous activity in the CMS detector causes a non-trivial confusion around the original hard process under study. Inclusively, these activities are denoted as Pile-Up (PU) and can originate in other pp-collisions happening during the same, or alternatively, previous/next bunch crossing. Simulation of PU events is obtained using a dedicated CMSSW module in parallel with standard MC simulation as follows:

1. the mean number $\langle N \rangle$ of interactions per bunch-crossing is retrieved from data,
2. a simulated number of random interactions is sampled by a Poisson distribution assuming that the distribution mean is equal to $\langle N \rangle$,
3. an actual event is simulated using this sampled number of random interactions.

Alternatively to the third point, PU interactions are simulated as independent samples which are later interfaced with hard scattering processes using PYTHIA. The PU simulation is further studied at the level of reconstructed physical objects and various methods [97] are employed to mitigate its effect in data.

3.1.5 The CMS detector response

General purpose MC generators described in the previous sections define the physical event including (generator-level) kinematic information for all of the particles in the final state (and sometimes also intermediate states). Finally, this information is propagated through the CMS detector volume. For the datasets employed in this analysis, this is achieved using the GEANT4 [98] simulation toolkit. Various aspects of the detector subsystems are taken into account: detector geometry and positioning, material composition including active, passive or dead (cables, cooling pipes) components, propagation

of the magnetic field through the material or environmental conditions. GEANT4 provides complex tracking simulation for all charged particles in the magnetic field while storing information about detector hits and interaction vertices. Both, electromagnetic and hadronic showers are simulated while accounting for underlying physical principles for electrons, photons or hadrons interacting with absorber material. Conversion of analog to digital signal is also included in the simulation, as well as readout and simulation of electronics response. In conclusion, a large set of detector parameters is retrieved by means of both data-driven (e.g. test beam calibration) and theory models in order to simulate event propagation in accordance with real-time detector performance. The actual output of GEANT4 simulation is later interfaced with various algorithms performing physical object reconstruction (same as for real data) as described in the following section.

3.2 Reconstruction of Physical objects

The CMS detector provides a comprehensive list of raw detector readout information as a response to particles passing through and interacting with the detector subsystems. In practice, combined information from ideally all detector components is used to identify and describe individual particles using the Particle Flow (PF) [99] algorithm. More detailed description is provided in Section 3.2.1, though its simplified version can be summarised into several steps. First, CMS Tracker and Muon System (MS) chambers are used to reconstruct trajectories ("tracks") belonging to muons. Remaining tracks are extrapolated to the HCAL volume and matched to the reconstructed clusters (a set of calorimetry cells with energy deposits sharing the same hit topology) to identify charged hadrons. A special algorithm is employed to match electron tracks with corresponding ECAL clusters and clusters associated with the Bremsstrahlung photons radiated by electrons on their way to ECAL. The rest of ECAL clusters is assigned to stable isolated photons not originating in Bremsstrahlung. Lastly, remaining HCAL clusters are associated to neutral hadrons.

The same level of information from the CMS detector needed to reconstruct individual particles is simulated at the generator-level, thus both data and MC samples are further processed to obtain reconstruction-level physical objects. Namely, analysers are interested in

- jets comprised (mostly) of collimated hadrons and photons with most of their energy deposited in the ECAL and HCAL volume,
- isolated electrons and photons with their energy reconstructed by the ECAL,
- the visible trace left by isolated muons reaching the MS chambers,
- dedicated tags marking hadronic tau decays and b-quark initiated jets (b-jets),
- and finally missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ manifesting the presence of neutrinos and possibly other feebly interacting particles.

Reconstruction procedures are performed according to the official Physics Object Group (POG) guidelines. Some of the procedures (when explicitly stated) are analysis-specific and were tuned to meet corresponding requirements. My personal contribution to the workload related to object reconstruction is comprised of assuring proper reconstruction of jets and $E_{\mathrm{T}}^{\mathrm{miss}}$ objects entering the $H \to WW$ analysis. Explicitly, I have contributed by implementing and studying:

- effects of preliminary versions of jet energy corrections,
- effects of applying jet momentum resolution smearing,
- effects of applying various jet PU identification criteria and data-to-MC scale factors,
- uncertainties related to the jet corrections/smearing/PU and reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$,
- translation of relevant jet corrections onto $E_{\mathrm{T}}^{\mathrm{miss}}$ objects.

Most significant contribution is summarised and more details on the aforementioned topics are presented in Sections 3.2.4-3.2.5.

3.2.1 The Particle Flow algorithm

Three main aspects of the PF algorithm form a skeleton of event reconstruction and deserve more detailed explanation: the tracking algorithm (or vertex and track reconstruction), the clustering algorithm (or cluster reconstruction in calorimeters) and the linking algorithm to perform the matching of the PF elements (tracks and clusters).

Tracking algorithm

Knowing the trajectory of all charged particles passing through the CMS detector is a crucial element for evaluating the momenta and flight directions of these particles. The tracking performance is typically quantified by calculating tracking efficiency (capability to reconstruct all tracks) and tracking "fake" rate (the ratio of incorrectly recognised tracks over all reconstructed tracks). An inefficient tracker system could lead to the false interpretation of energy deposits in the CMS HCAL and ECAL as standalone energy clusters, thereby resulting in an increased count of reconstructed neutral hadrons or photons. On the other hand, fake tracks can naturally affect the final momentum and energy spectra, especially when tracks are accounted in the reconstruction of jets. The CMS tracking algorithm described below is developed to target both, maximal efficiency and a minimal tracking fake rate.

In most cases, individual tracks are reconstructed by employing the Combinatorial Kalman Filter (CKF) [100] . Initially, detector hits are searched layer by layer for suitable building blocks, also called "seeds" [101]. Seeds are usually selected as a pair of consecutive hits in the CMS Tracker Pixel detector and assuming that the trajectory passes through the center of the beam (and/or known vertex in case of an iterative approach). Initial track parameters are then estimated using the seed configuration and detector (both CMS Tracker and MS) is searched for the next hit compatible with dynamics typical for charged particles in the magnetic field. This step is repeated and track parameters are updated (in the fit) for each added hit candidate until the last detector layer is reached or another stopping condition is fulfilled (e.g. maximum number of hits). Final fit is performed when all hits are included and corresponding "fit quality flag" is assigned to each track.

To avoid possible overlap in exploiting detector information, tracks are constructed in parallel and selected hits are removed from the pool of possible candidates. Furthermore, to increase tracking efficiency and minimize the fake rate, the CMS uses the so called iterative tracking algorithm [102]. During the initial round, the CKF is applied for

the seeds selected with tighter criteria (e.g. high track p_{T}), targeting the raise in the track reconstruction purity. Seed selection criteria are steadily loosen during following rounds, increasing tracking efficiency, but keeping the fake rate reasonably low given that combinatorics are now significantly reduced.

Finally, combined information of multiple reconstructed tracks originating in the area compatible with the beamline is fitted to reconstruct the position of the Primary interaction Vertex (PV) [103, 104]. In case of multiple PV candidates, the one with outgoing tracks corresponding to the maximal sum of track p_{T} is selected. The pool of available reconstructed tracks (not originating in the PV) is further scanned for the secondary vertices which might define the origin of unstable secondary particles or mark the beginning of jet evolution (e.g. triggered by b-quark).

Clustering algorithm

To enhance kinematic information on both charged and neutral particles, it is important to recover the energy deposits for each cell of the CMS calorimetry system in order to fully estimate single-particle contributions. While in the case of charged particles, reconstructed trajectories are assumed to get matched with calorimetry signal, in the case of neutral particles, calorimetry readout is the only available information. In that regard, calorimeters serve not only to measure energy of incoming particles but also to distinguish charged hadrons from neutral ones as well as to identify electrons and photons effectively stopped by the ECAL absorber material (not leaving a signal in the HCAL).

Initially, the readout signal from individual cells is searched for energy deposits exclusively larger than 1 GeV energy threshold within any subgroup of neighboring cells (including those connected only by a corner). These cells are denoted as cluster seeds as they mark the starting point for construction of the so called topological clusters. These objects yet do not constrain single-particle energy deposits, on the contrary, multiple particles can contribute to the same topological area. When the energy deposit measured in the cell adjacent to the cell in the topological cluster (starting from the seed) is higher than twice the noise level set for this cell, then this cell is attached to the topological cluster in an iterative manner. Eventually, more than one seed can be present in the same topological cluster.

Single-particle clusters within the topological cluster are reconstructed using a maximum likelihood fit on the Gaussian-mixture model [105]. In this model, all seeds included in the topological cluster are assumed to give a rise to the Gaussian energy distributions (each seed corresponds to one Gaussian distribution), thus populating energy deposits in all cells included in the topological cluster. Fitted Gaussian parameter values, such as position and total energy are then considered as defining properties for the clusters associated with individual particles.

CMS detector components deploying the clustering algorithm are the EB, ECAL Endcaps (EC), layers of ES as well as the HB and HE. Each of the components performs this algorithm independently. In case of the CMS HCAL forward calorimeter, the corresponding readout from individual cells is, by design, separated into hadronic and electromagnetic shower contributions which are directly assigned to the hadronic and electromagnetic clusters.

Linking algorithm

Once the individual PF elements are reconstructed, the PF linking algorithm [106]

searches in their nearest proximity for the suitable candidates to form PF pairs. Besides proximity, additional requirements are placed upon particular types of pair combinations as further discussed.

Tracks reconstructed in the CMS Inner Tracker are extrapolated to the ECAL and HCAL layers up to the depth expected for the typical longitudinal profile of electromagnetic and hadronic showers. A link is established between the extrapolated track and the position of the reconstructed cluster in the overlapping area. In the case of multiple clusters fulfilling this condition, the one located within the shortest distance in the η-ϕ plane is selected. Similar linking (cluster-to-cluster) is performed between clusters originating in the ECAL and HCAL layers. Furthermore, ECAL clusters originating from Bremsstrahlung photons are linked to the parent electron track if any of the tangent lines extrapolated from the intersection point of the electron track and the CMS Inner Tracker layers is overlapping with the cluster position in the ECAL. Moreover, tracks reconstructed in the central tracker are extrapolated to the MS where they are checked to match with possible muon candidates. Remaining PF elements are searched for the combination of three tracks sharing the same misplaced intersection (secondary vertex), while at most one of these three tracks has its origin in the primary interaction vertex (e.g. an unstable hadron was produced and consequently decays).

Once matched, PF pairs are associated within one PF block given there is at least one direct or indirect link between them. PF blocks then serve as input for the reconstruction of physical objects in the logical order that prevents from double-usage of PF elements and is respected by the ordering of the following sections.

3.2.2 Muon reconstruction

In practice, three types of muon tracks are recognised in the scope of the PF algorithm depending on a level of information and an approach used in their reconstruction:

1. Standalone muon tracks – using only the track elements reconstructed from the MS chambers,

2. Global muon tracks – reconstructed by matching the standalone tracks with the tracker tracks ("outside-in" approach),

3. Tracker muon tracks – obtained from the tracker tracks extrapolated to the MS and matched to the standalone tracks ("inside-out" approach).

Standalone muon tracks are reconstructed employing the CKF method while using only the most limited subset of information restricted to the signal from the DT, CSC and RPC chambers. Because of these limitations, standalone tracks suffer from small spatial resolution and are rather used as a backup method in case that options 2. and 3. fail. Fortunately, about 99% of all muon tracks in the CMS detector acceptance are reconstructed as global or tracker muon tracks.

The starting point for the reconstruction of the global muon tracks are the robust-enough standalone muon tracks assembled by requiring a sufficient number of hits in the MS chambers. This way it is assured that the muon reconstruction fake rate gets significantly reduced as compared to the tracker-only tracks not matched to the standalone muon tracks. As a next step, both the standalone muon tracks and the tracker tracks are extrapolated to the common surface area and matched based on the mutual distance. The global muon track is then a result of a global fit combining information from both matched tracks.

Parameter	Selection criteria
Number of hits in the Tracker Strip	≥ 5
Number of hits in the Tracker Pixel	≥ 1
Hits in the MS (Tracker muon track)	≥ 2
Hits in the MS (Global muon track)	≥ 1
χ^2/dof (Global muon track)	< 10
$\lvert d_{xy} \rvert$	< 0.2 cm
$\lvert d_z \rvert$	< 0.5 cm

Table 3.1: The Muon POG definition of tight ID selection criteria in addition to the muon being reconstructed as global and tracker.

In the case of muons with $p_{\mathrm{T}} > 200$ GeV, the measurement of the global muon track p_{T} is superior to p_{T} measurements with any other track type. On the other hand, an advantage of the tracker muon track is high reconstruction efficiency for low-p_{T} muons, especially in detector regions with an excess of dead material. The reconstruction procedure in this case goes as follows. All the tracker tracks with $p_{\mathrm{T}} > 0.5$ GeV and total momentum $p > 2.5$ GeV are extrapolated to the MS and searched for a possible match with hits in at least one of the DT or CSC segments. Unfortunately, such a loose matching algorithm is susceptible to the presence of charged hadron remnants occasionally reaching innermost MS stations and being misidentified as muons.

Depending on the probed analysis phase space, analysers can opt for different track collections. Very often, however, the muon track is reconstructed as both a global and tracker muon track. In that case, two options are merged into a single muon track candidate.

Muon identification and isolation

Assuming different reconstruction procedures, the final collection of muon objects might differ in the quality of their reconstruction and isolation from other physical objects inside of the detector. For that purpose, CMS Muon POG provides analysers with a set of selection and isolation criteria, varying by the level of stringency [107]. Analysers might choose from the available selection sets, also denoted as working points.

As an example, three levels of muon Identification (ID) working points are widely used among analysers: loose, medium and tight. Loose ID targets identification of prompt muons (with tracks originating near to the PV, e.g. from W decay) as well as muons coming from the semileptonic quark decays. Medium ID places the requirement for loose muons, such as they need to result from the fit using more than 80% of the hits in the Tracker, therefore allowing to target muons coming from the subset of semileptonic heavy flavor quark decays. Finally, tight ID is defined to strictly suppress nonpromt muons coming from semileptonic quark decays as well as to reduce charged hadron fake rate. Besides muon requested to be reconstructed as both global and tracker muon track, several parameters were tested to have significant impact on "tight muon" definition. Among these are: number of hits in the Strip and Pixel Tracker, number of hits in the

MS chambers, quality of the fit expressed by χ^2/dof test and compatibility of muon track with the PV position. The last point is calculated by means of the impact parameters d_{xy} and d_z defined as a distance of the PV and muon track at the closest approach, profiled to the x-y plane and z-axis, respectively. All muons selected in this analysis need to pass Muon POG tight ID selection criteria summarised in Table 3.1.

Additionally, muons can be categorised based on their spatial Isolation (ISO). Typically, Muon POG provides two basic recipes to estimate muon ISO. In both cases, volume around reconstructed muons, restricted by the geometrical cone of

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} < R_{\text{cone}} \tag{3.6}$$

with an origin in the PV and cone radius R_{cone} chosen depending on the method, is searched for the presence of adjacent particles. Besides that, methods vary as follows:

1. $\Delta R < 0.3$ cone is searched for all reconstructed tracks (ISO_{track}),
2. $\Delta R < 0.4$ cone is searched for all charged and neutral hadrons from the PF algorithm ($ISO_{0.4}$),

where both isolation classifiers are obtained as a relative ratio of corresponding p_{T} and E_{T} contributions from considered PF elements over transverse momentum $p_{\text{T}}(\mu)$ of the reconstructed muon. In this analysis, the $ISO_{0.4}$ was selected as a base to assure proper muon isolation as it also accounts for the PU corrections to the neural hadron energy deposits. It was estimated based on simulation that such a correction factor can be obtained as half the sum of energy deposits corresponding to the charged hadrons originating in the PU vertices. The entire formula reads:

$$ISO_{0.4} = \frac{1}{p_{\text{T}}(\mu)} \left(\sum_{\text{PV}\to\text{ch.h.}} p_{\text{T}} + \max\left(0, \sum_{\text{n.h.}} E_{\text{T}} + \sum_{\gamma} E_{\text{T}} - 0.5 \times \sum_{\text{PU}\to\text{ch.h.}} p_{\text{T}}\right)\right). \tag{3.7}$$

Muon definition in HWW analysis

Targeting leptonic W decays in $H \to WW^*$ channel, it is necessary to further study (and improve) quality of prompt muon selection criteria as well as our capability to

Parameter	Selection criteria
Muon POG tight ID	yes
Muon pseudorapidity $\lvert\eta\rvert$	< 2.4
$\lvert d_z\rvert$	< 0.1 cm
$\lvert d_{xy}\rvert$ for muons with $p_{\text{T}} \leq 20$ GeV	< 0.01 cm
$\lvert d_{xy}\rvert$ for muons with $p_{\text{T}} > 20$ GeV	< 0.02 cm

Table 3.2: Muon selection criteria common for both TightHWW and LooseHWW working points.

Parameter	TightHWW selection	LooseHWW selection
$ISO_{0.4}$	≤ 0.15	≤ 0.4
ttHMVA	> 0.8	no

Table 3.3: Additional muon selection criteria optimized for TightHWW and LooseHWW working points separately.

estimate nonpromt background. Former study subject requires updated muon definition denoted as "TightHWW" muon. Latter subject is closely related to the method (see more in Section 4.3.4) employed to estimate presence of nonprompt muons (leptons in general) and requires introduction of "LooseHWW" muon with relaxed isolation and identification criteria. Extra requirements that are common for both HWW working points are summarised in Table 3.2 while selection criteria specific for TightHWW and LooseHWW muons separately are pointed out in Table 3.3.

Particular attention should be paid to the selection criteria based on a ttHMVA discriminant obtained via Multivariate Analysis (MVA) of muon kinematic and reconstruction properties (as well as information about nearest reconstructed jet). Originally, this method was developed in scope of $t\bar{t}H$ production study [108] which aimed to maximize capability of selecting prompt/nonprompt lepton candidates by employing Boosted Decision Tree (BDT) technique. In the scope of this analysis, optimal selection value of ttHMVA > 0.8 was found by maximizing efficiency of selecting prompt muons.

3.2.3 Electron and photon reconstruction

As it is known from previous discussion, electrons passing through magnetic field are prone to radiate Bremsstrahlung photons. Moreover, electrons and photons are hardly ever traversing the detector material without triggering EM showers. Because of much higher multiplicity of PF elements associated to the original electron or photon, it is convenient to attempt for the reconstruction of both object types simultaneously, only deciding on their particular type by applying final selection criteria.

Firstly, ECAL energy clusters defined by the PF clustering algorithm are merged into so called Superclusters (SCs) if they appear in the same η-ϕ window around the original cluster seed. This way, the Supercluster (SC) can account also for the accompanying energy deposits, e.g. from Bremsstrahlung photons. Secondly, the CKF method typically used for the track reconstruction was updated for the Gaussian Sum Filter (GSF) method [109] which performs necessary track corrections for substantial radiative losses. To reduce computational complexity, the GSF electron tracks are built starting from the predefined selection of two seed types:

1. ECAL-driven – selected from the available SCs,

2. Tracker-driven – selected from the generic CKF tracks in the Tracker.

In case of tracker-driven GSF electron tracks, if any available CKF track (i.e. not used in muon reconstruction) that is found in the Tracker is compatible with any ECAL cluster, it is used as a seed to the GSF algorithm. On the contrary, ECAL-driven electron tracks are built based on the SCs compatible with the track seeds (doublets and triplets of hits

Parameter	EB	EE
$\sigma_{\eta\eta}$	< 0.011	< 0.031
$\lvert\eta_{\mathrm{SC}} - \eta_{\mathrm{track}}\rvert$	< 0.004	no
$\lvert\phi_{\mathrm{SC}} - \phi_{\mathrm{track}}\rvert$	< 0.02	no
H/E	< 0.06	< 0.065
$\lvert 1/E_{\mathrm{SC}} - 1/p_{\mathrm{track}}\rvert$	< 0.013	
Cluster PF isolation (ECAL)	< 0.16	< 0.12
Cluster PF isolation (HCAL)	< 0.12	
Tracker PF isolation	< 0.08	
GSF electron track χ^2/ndof	< 3	

Table 3.4: The EGM POG electron cut-based selection criteria for the 2016 dataset. Selection might differ for the ECAL barrel ($|\eta| \leq 1.479$) and endcap ($|\eta| > 1.479$) regions.

in the Tracker). More details about energy thresholds required for the SC to be selected as ECAL-driven seed are to be found in [110].

As a third step, the GSF electron tracks and SCs are associated using dedicated BDT classifier [105] trained on the available track and supercluster kinematic variables as well as the variables defining an overall quality of the fit. In case of the ECAL-driven GSF tracks, an alternative matching procedure based on the closest approach of the GSF track and the associated SC is allowed within $|\Delta\eta| < 0.02$ and $|\Delta\phi| < 0.15$ windows. In parallel, dedicated algorithm [111] is employed to identify electron tracks that originate in photon conversion to e^+e^- pairs.

Eventually, all available information: the SCs, ECAL clusters, GSF electron tracks, generic CKF tracks associated to electrons and conversion electron tracks are fed into PF linking algorithm to identify PF blocks with a mixture of electron-photon (e/γ) object candidates. Both electrons and isolated photons are then assembled starting from the SCs. If the e/γ object has a link to the GSF electron track (i.e. this link survives BDT classification discussed above), it is tagged as electron. On the other hand e/γ objects with no association to the GSF tracks are tagged as isolated photons. Note that remaining ECAL clusters that were not included to the considered SC or were not recovered as energy deposits from Bremsstrahlung radiation, are still available for the reconstruction of non-isolated photons contained in jets.

Electron identification and isolation

No isolated photons are requested in the final state considered in this analysis, therefore, in the following text we will focus exclusively on electron identification and isolation properties. As it is in case of muons, Electron-Gamma POG (EGM POG) studied variables related to electrons and adjacent objects in order to differentiate between prompt electrons (with tracks originating near to the PV, e.g. from W decay) and those from

Parameter	EB	EE
$\sigma_{\eta\eta}$	< 0.0105	< 0.0309
$\lvert\eta_{\mathrm{SC}} - \eta_{\mathrm{track}}\rvert$	< 0.00365	<0.00625
$\lvert\phi_{\mathrm{SC}} - \phi_{\mathrm{track}}\rvert$	< 0.0588	<0.0355
H/E	$< 0.026 + \frac{1.12}{E_{\mathrm{SC}}} + 0.0368\frac{\rho}{E_{\mathrm{SC}}}$	$< 0.026 + \frac{0.5}{E_{\mathrm{SC}}} + 0.201\frac{\rho}{E_{\mathrm{SC}}}$
$ISO_{0.3}$	< 0.0718	< 0.143
$\lvert 1/E_{\mathrm{SC}} - 1/p_{\mathrm{track}}\rvert$	< 0.0327	< 0.0335
Missing inner hits	≤ 1	
Conversion veto	yes	

Table 3.5: The EGM POG electron cut-based selection criteria for the 2017/18 datasets. Selection might differ for the ECAL barrel ($|\eta| \leq 1.479$) and endcap ($|\eta| > 1.479$) regions.

the background sources. In particular, electrons from semileptonic quark decays, conversion electrons and jets misidentified as electrons will be all tagged as nonprompt. Two complementary methods were developed by EGM POG for electron identification and isolation: one is based on BDT technique while second directly applies selection criteria on the sensitive quantities, henceforth denoted as "cut-based" method. Some of the most relevant variables addressed by both methods are summarised below:

- Ratio of hadronic over electromagnetic energy H/E: the HCAL deposits in a cone of $\Delta R < 0.15$ around the SC direction are summed and compared to the SC deposits in the ECAL.
- The second moment $\sigma_{\eta\eta}$: the Root Mean Square (RMS) weighted distribution of the ECAL cluster crystal energy deposits calculated along η in 5×5 matrix inside of the SC.
- SC-track pseudorapidity distance $|\eta_{\mathrm{SC}} - \eta_{\mathrm{track}}|$: distance between the SC energy-weighted position and the extrapolated inner track in η.
- SC-track azimuth distance $|\phi_{\mathrm{SC}} - \phi_{\mathrm{track}}|$: distance between the SC energy-weighted position and the extrapolated inner track in ϕ.
- $|1/E_{\mathrm{SC}} - 1/p_{\mathrm{track}}|$ variable: here E_{SC} stands for the SC energy and p_{tracks} is the momentum of the track at the point of closest approach to the PV.
- The median of the transverse energy density per unit area ρ calculated in the center of detector.
- Cluster/Tracker PF isolation: quantity defined by the ratio of summed energy / momenta deposits around the electron direction in $\Delta R < 0.3$ cone relative to the electron p_{T}.
- Combined PF isolation $ISO_{0.3}$: robust definition of electron isolation using all available PF information defined in $\Delta R < 0.3$ cone and relative to the electron p_{T}. See formula 3.7 for more details.

Dataset	Parameter	EB	EE
2016	Missing inner hits	< 1	
2017-2018	$\sigma_{\eta\eta}$	no	< 0.03
	$\lvert 1/E_{\mathrm{SC}} - 1/p_{\mathrm{track}} \rvert$	no	< 0.014
2016-2018	Electron pseudorapidity $\lvert\eta\rvert$	< 2.5	
	$\lvert d_{xy} \rvert$	< 0.05 cm	< 0.1 cm
	$\lvert d_z \rvert$	< 0.1 cm	< 0.2 cm

Table 3.6: Additional electron selection criteria common for both TightHWW and LooseHWW working points.

- Number of missing hits in the innermost Tracker layers: given that prompt electrons are expected to originate near to the beam line, this number is expected small or zero.
- Conversion veto: assures that electron is not matched to the conversion photon.

Several cut-based working points are defined by EGM POG to address analyses with different level of electron background contamination. Because of high number of nonprompt leptons present in our analysis phase space, all considered electrons need to pass relatively tight selection criteria derived from official working points. Particular selection slightly differ for 2016 and 2017/18 datasets given the CMS Tracker Pixel Phase-1 upgrade as well as the advancement in reconstruction procedures and are summarized in Table 3.4 and Table 3.5, respectively. Eventually, tighter selection criteria might be applied for electrons selected in this analysis (both cut- and BDT-based) as further described in the following section.

Electron definition in HWW analysis

On top of the cut-based ID working points, we define two electron working points denoted as TightHWW and LooseHWW. Identically to the muon case, LooseHWW electrons are subject to the procedure of estimating nonpromt lepton background while TightHWW electrons are designed to optimize selection of prompt final state electrons. Extra selection requirements common for both analysis-specific working points are summarised in Table 3.6 separately for 2016 and 2017/18 datasets.

In addition to that, TightHWW electrons are required to pass the BDT-based ID provided by the EGM POG. Two multivariate classifiers are trained individually in the 2016 and 2017/18 campaigns using a superset of kinematic and fit-quality variables already introduced in the scope of the cut-based selection. BDT response is varied to reflect different levels of efficiency required by analysers. In our case, BDT-based ID working point was picked to correspond to the efficiency of 90% (for both classifiers). Finally, TightHWW electrons are required to pass more stringent isolation criteria summarised in Table 3.7.

Dataset	Parameter	EB	EE
2016	$ISO_{0.3}$	< 0.0588	< 0.0571
2017-2018	$ISO_{0.03}$	< 0.06	

Table 3.7: Additional isolation criteria applied for electrons passing TightHWW BDT-based selection.

3.2.4 Jet reconstruction

Once the all isolated electrons, muons and photons are reconstructed in the PF algorithm described in previous sections, all PF elements used in this process are removed from the pool of available information. Remaining tracks, ECAL and HCAL clusters become a subject to the further PF linking as they will serve for an identification of charged hadrons (e.g. $\pi^{\pm}$ or protons), neutral hadrons (e.g. kaons K_L^0 or neutrons), non-isolated photons (e.g. from π^0 decays) and very occasionally additional muons.

All the left out ECAL and HCAL clusters that are not linked to any of the remaining tracks will be associated to the photons and hadrons. Inside of the Tracker acceptance ($|\eta| < 2.5$) it is safe to assume that each of such ECAL clusters will a priori give rise to photons and not neutral hadrons. This assumption is based on observation [105] that neutral hadrons contained in jet leave only about 3% of the jet's energy in the ECAL volume. Outside of the Tracker acceptance, ECAL clusters that are also linked to the HCAL clusters (but not to the track) will be considered as deposits from the same neutral hadron. Only those ECAL clusters not linked to the HCAL clusters will still be associated to photons.

The ECAL and HCAL clusters that are mutually linked to at least one of the remaining tracks will be associated to the charged hadrons with their energy and momentum reconstructed using combined information from both calorimetry systems and the sum of the track momenta. Rarely, reconstructed energy of charge hadron candidate is significantly smaller than reconstructed momentum. In this case, an alternative muon PF reconstruction is launched with relaxed selection criteria to recover muon not identified in the previous steps.

Anti-kT algorithm and jet assembly

Once the reconstruction of all PF particles is completed, dedicated Anti-kT (AK) algorithm [112] is used to reconstruct jets, i.e. conic-shaped sprays of collimated particles with closely localized energy deposits, ideally originating from the single parton. The AK algorithm can be simplified into following steps:

1. loop over available PF particles and pseudojets (recombined multiples of PF particles) starting from index i,
2. for given index i, loop over other particles/pseudojets with index $j \neq i$ and determine its mutual distance measures d_{ij},
3. also determine its distance measure from the beam d_{iB},
4. find the smallest of the distances $\min\{d_{ij}\}$,

5. if $\min\{d_{ij}\} < d_{iB}$, recombine particles i and j into the new pseudojet and repeat steps 1. to 5.,

6. else, mark pseudojet i as a hard jet, remove all PF particles contained in this jet from the available choices and start another iteration until no more PF particles/pseudojets are left in the pool.

Distance measures used in this algorithm are defined as

$$\begin{aligned} d_{ij} &= \min\left\{\frac{1}{p_{\mathrm{T}i}^2}, \frac{1}{p_{\mathrm{T}j}^2}\right\}\frac{\Delta_{ij}^2}{R_{\mathrm{cone}}^2}, \\ d_{iB} &= \frac{1}{p_{\mathrm{T}i}^2}, \end{aligned} \tag{3.8}$$

where $\Delta_{ij}^2 = \Delta y_{ij}^2 + \Delta\phi_{ij}^2$ with y, ϕ and p_{T} standing for rapidity (see definition 2.3), azimuth and transverse momenta for considered particles/pseudojets. Resolution parameter R_{cone} sets a radius of a cone defining geometry of the reconstructed jets. Jet kinematic properties are determined from the vectorial sum of the four-momenta of all PF particles contained in jet.

An advantage of the anti-kT algorithm is that the overall jet shape is not dictated by the presence of soft particles (with low p_{T}) coming from showering processes. Instead, pseudojets are formed around hard particles which are only merged into the single jet if their mutual distance is smaller than parameter R_{cone} (otherwise they form individual hard jets). Henceforth, jets reconstructed via anti-kT algorithm with $R_{\mathrm{cone}} = 0.4$ ($R_{\mathrm{cone}} = 0.8$) are denoted as AK4 (AK8) jets.

Pile-up mitigation

Additional PU interactions might lead to extra particles present in the event and being falsely reconstructed as jets. Several techniques were developed to mitigate PU effects, while the most commonly used among the CMS analysers are the Charged Hadron Subtraction (CHS) and Pile-up Per Particle Identification (PUPPI) algorithms [113]. Both methods start with exploiting information about charged hadron tracks and their associated primary vertices. Single hard-scatter vertex is selected out of the all available primary vertices, if this vertex corresponds to the maximal quadratic sum of associated tracks transverse momenta, $\sum p_{\mathrm{T}}^2$. Other primary vertices with sub-leading density of high-p_{T} tracks are denoted as PU vertices. CHS method then removes all charged hadrons with their tracks originating in PU vertices from the pool of available PF particles, so they cannot be used in jet reconstruction via anti-kT algorithm. Note that CHS technique does not account for neutral hadrons and photons coming from PU vertices.

On the other hand, PUPPI method operates purely at the particle level while employing all available information about tracking, PU properties and distribution of particles in the detector volume. In case of charged hadrons, algorithm resembles that of the CHS method. For neutral particles, the PUPPI algorithm reflects on QCD structure of collimated jets which assumes that all particles in the same shower will be located close to each other as compared to PU particles which are scattered more homogeneously. Based on this information, each particle candidate is assigned a weight in range between 0 and 1, parametrizing probability of particle coming from PU vertex or hard-scatter vertex, respectively. The PUPPI-weighted PF particles are then passed to the anti-kT algorithm for further clustering into jets.

Additional corrections on reconstructed jet kinematic properties might be required depending on chosen PU mitigation method, detector response, data-to-MC agreement, jet (quark/gluon) flavor composition, resolution of momentum measurement and other aspects discussed in the following sections.

Jet energy corrections

Proper determination of jet four-momentum is of the key importance for analyses selecting events with jets in the final state. The Jet Energy Corrections (JEC) summarily refer to the calibration of jet four-momenta by applying the appropriate Scale Factors (SF) depending on jet kinematic properties and parton composition. Each source of possible discrepancy is addressed by a separate correction which is applied sequentially in order defined by the CMS Jet and Missing E_{T} POG (JetMET POG) [114]. Overall scaling procedure is depicted by infographics 3.3 and is further discussed below.

In the first stage, jets reconstructed with CHS method need additional treatment (that is not needed with PUPPI method) to account for neutral hadrons and photons originating in PU vertices. It is assumed that energy contribution coming from PU particles is proportional to the jet area A. As such, corrections are parameterised as a function of A as well as jet η, p_{T} and median energy density in the event to account for additional dependencies. The SF are derived by comparing jet distributions reconstructed in simulated QCD samples enriched by di-jet events (events with at least 2 jets in the final state) with and w/o additional layer of PU interactions. Corrections are then applied to both data and MC samples employed in this analysis. Note that the SF related to PU effect used in data contain an extra layer of corrections to address residual discrepancies using the Random Cone (RC) method (see Chapter 4.3 in [114]).

In the second stage, the CMS detector response, simulated by GEANT4 and corrected for initial parton showering by PYTHIA generator, is studied in simulated di-jet QCD samples. The response SF are determined by comparing jet p_{T} distributions for the generated particle-level jets (i.e. jets are clustered using stable generator-level candidates with an exception of neutrinos) and the reconstruction-level PF jets that are geometrically matched to their particle-level counterparts. Both data and MC samples are corrected by the response SF parameterised as a function of jet η and p_{T}.

During the third stage, small (order of %) corrections on residual differences of the jet response in data (observed in comparison to the MC) are applied. The SF are derived independently (contrary to the stage two corrections) as a function of jet η and p_{T}, respectively. In both cases, corrections target possible imbalance between jets momenta (or overall event energy recoil) for jets selected ("probed") in various datasets respective

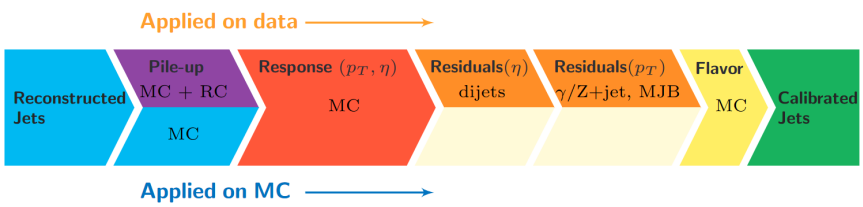

Figure 3.3: Consecutive list of jet energy corrections as defined by the JetMET POG. Corrections applied to data (upper row) are visually separated from those applied to simulated samples (bottom row). Source: [114].

to the reference ("tag") object selected from the same event, however measured in the specific η or p_{T} regions. In case of the η-dependent SF, probed are jets selected in di-jet events irrespective to their η and p_{T}, while tagged is the jet selected in detector barrel region $|\eta| < 1.3$. In case of the p_{T}-dependent SF, probed are jets from $Z/\gamma + \mathrm{jet}$ (multi-jet) samples, located in detector barrel region with momenta in range between 30 GeV and 800 GeV (> 800 GeV for multi-jet samples), while tagged is the precisely measured Z/γ object (overall energy recoil for multi-jet samples).

In the last correction stage, different flavor composition of jets is studied in $Z/\gamma + \mathrm{jet}$ simulation as it is directly affecting jet particle composition and causes variations in parton fragmentation energy. Similar to the second stage approach, the SF are determined by comparing jet p_{T} distribution between generator particle-level jets with known flavor and the matching reconstruction-level jets.

Jet energy resolution

The particle-level Jet Energy Resolution (JER) in the MC simulation is defined as a width of $p_{\mathrm{T}}^{\mathrm{reco}}/p_{\mathrm{T}}^{\mathrm{gen}}$ distribution, where $p_{\mathrm{T}}^{\mathrm{reco}}$ and $p_{\mathrm{T}}^{\mathrm{gen}}$ stand for the reconstruction-level and generator particle-level jet transverse momenta, respectively. In practice, value of JER in simulation is estimated as σ_{JER} obtained from the Gaussian-core fit within $\pm 2\sigma_{\mathrm{JER}}$ range around the mean of this distribution. Unfortunately, resolution achieved in data is suboptimal compared to one achieved in the MC. This is mostly due to the inactive ECAL material and/or energetic particles punching through HCAL volume, i.e. effects that can be reasonably addressed in simulation by more advanced fitting models (e.g. two-sided Crystal Ball function). Moreover, resolution in data suffers from the presence of elusive neutrinos that are otherwise not included in reconstruction of generator particle-level jets.

On the other hand, resolution in data is determined by means of the fitting the width of the distribution of the di-jet system momentum assymetry $\mathcal{A}$ between jet leading (index 1) and sub-leading (index 2) in momentum

$$\mathcal{A} = \frac{p_{\mathrm{T}}^{\mathrm{jet1}} - p_{\mathrm{T}}^{\mathrm{jet2}}}{p_{\mathrm{T}}^{\mathrm{jet1}} + p_{\mathrm{T}}^{\mathrm{jet2}}} \tag{3.9}$$

as measured in QCD di-jet events selected by dedicated HLT paths [115].

To match jet resolution in the MC simulation with resolution obtained in data, reconstructed jet p_{T} distribution is smeared by applying multiplicative factor c_{JER}. Following the JetMET POG suggestion, particular implementation of JER smearing varies depending on the jet origin (hard-scatter or PU). In that regard, reconstructed jets are marked as genuine if they pass two requirements: geometric matching with generator-level jets inside of cone defined by $\Delta R < R_{\mathrm{cone}}/2$ and resolution matching. Latter requirement is expressed by means of relative resolution σ_{JER} mentioned above, such as following is true:

$$|p_{\mathrm{T}}^{\mathrm{reco}} - p_{\mathrm{T}}^{\mathrm{gen}}| < 3\sigma_{\mathrm{JER}} p_{\mathrm{T}}^{\mathrm{reco}} \tag{3.10}$$

According to the hybrid JER procedure, genuine jets are scaled by multiplicative factor defined as

$$c_{\mathrm{JER}} = 1 + (s_{\mathrm{JER}} - 1)\frac{p_{\mathrm{T}}^{\mathrm{reco}} - p_{\mathrm{T}}^{\mathrm{gen}}}{p_{\mathrm{T}}^{\mathrm{reco}}} \tag{3.11}$$

where s_{JER} is data-to-MC scaling factor obtained by comparing the balance between resolution obtained by fitting distribution 3.9 in data and simulation. On the other

Parameter	$\lvert\eta\rvert \leq 2.4$	$2.4 < \lvert\eta\rvert \leq 2.7$	$2.7 < \lvert\eta\rvert \leq 3$	$\lvert\eta\rvert > 3$
Neutral hadron fraction (AK4)	< 0.9	< 0.9	< 0.98	–
Neutral hadron fraction (AK8)	< 0.9	< 0.9	–	–
Neutral EM fraction (AK4)	< 0.9	< 0.9	> 0.01	< 0.9
Neutral EM fraction (AK8)	< 0.9	< 0.9	–	–
Charged hadron fraction	> 0	–	–	–
Charged EM fraction	< 0.99	–	–	–
N. of neutral particles (AK4)	–	–	> 2	> 10
N. of neutral particles (AK8)	–	–	–	–
N. of charged particles	> 0	–	–	–
N. of constituents	> 1	> 1	–	–

Table 3.8: Jet tight ID selection criteria defined by the JetMET POG for 2016 datasets.

hand, non-genuine jets are smeared stochastically using formula

$$c_{\mathrm{JER}} = 1 + N(0, \sigma_{\mathrm{JER}})\sqrt{\max(s^2_{\mathrm{JER}} - 1, 0)} \tag{3.12}$$

where N denotes a random number sampled from a normal distribution with a zero mean and variance σ^2_{JER}. The JER smearing is applied on top of the calibrated jet four-momenta for both AK4 and AK8 jets.

Jet identification

Following an example of electrons and muons, different level of efficiency and purity in jet reconstruction might be requested depending on analysis needs. The JetMET POG defines a set of variables sensitive to selecting hard-scatter jets over badly reconstructed jets or jets resulting from the ECAL/HCAL noise. Most of them defines a fractional composition of the reconstructed jet energy as well as a number of PF particle constituents. Example of recommended jet tight ID selection criteria, corresponding to $> 99\%$ efficiency and $> 98\%$ background rejection for AK4 (AK8) jets in $|\eta| < 3$ ($|\eta| < 2.7$) region, are summarised in Tables 3.8-3.10 for all three datasets. Note that selection slightly differ depending on PU mitigation algorithm in use, therefore AK8 jets will not receive the same selection as AK4 jets by default.

In some cases, it is useful to identify jet's origin in more detail. In particular, we tackle the presence of top quark background production by identifying jets originating from bottom quarks, also denoted as "b-tagged" jets, as top quark decays around 1/3 of time into the $Wb \to l\nu b$ final state. Corresponding "b-tagging" procedure [116] is based on the multivariate classification and is discussed in Section 4.3.2.

Another example is selecting jets likely to originate from "boosted" V hadronic decay from associated VH production. Typically, hadronic V decay at rest (or with V not significantly boosted) is resolved by the presence of two AK4 jets with a large subtended

Parameter	$\lvert\eta\rvert \leq 2.4$	$2.4 < \lvert\eta\rvert \leq 2.7$	$2.7 < \lvert\eta\rvert \leq 3$	$\lvert\eta\rvert > 3$
Neutral hadron fraction (AK4)	< 0.9	< 0.9	–	> 0.02
Neutral hadron fraction (AK8)	< 0.9	< 0.9	< 0.99	> 0.02
Neutral EM fraction (AK4)	< 0.9	< 0.9	$[0.02, 0.99]$	< 0.9
Neutral EM fraction (AK8)	< 0.9	< 0.9	–	< 0.9
Charged hadron fraction	> 0	–	–	–
Charged EM fraction	–	–	–	–
N. of neutral particles (AK4)	–	–	> 2	> 10
N. of neutral particles (AK8)	–	–	–	$[2, 15]$
N. of charged particles	> 0	–	–	–
N. of constituents	> 1	> 1	–	–

Table 3.9: Jet Tight ID selection criteria defined by the JetMET POG for 2017 datasets.

angle. Such a topology, however, is also common for more ordinary parton-initiated jets from QCD production. To suppress QCD background, it is effective to select cases with high-p_{T} (boosted) V boson decaying into two highly collimated jets, which are consequently reconstructed as one AK8 "V-tagged" jet. Some of the variables used in the "V-tagging" process [117] are:

- N-subjetiness τ_N – this variable describes a generalised jet shape under an assumption it contains N subjets. It is defined as p_{T}-weighted distance between all jet constituents (index k) and the nearest subjet axis:

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T}}^{(k)} \min(\Delta R_1^{(k)}, ..., \Delta R_N^{(k)}), \tag{3.13}$$

  where the normalisation factor $d_0 = \sum_k p_{\mathrm{T}}^{(k)} R_{\mathrm{cone}}$ is defined using original jet resolution parameter R_{cone} used in jet reconstruction.

- softdrop groomed jet mass $m^{\mathrm{AK8}}_{\mathrm{softdrop}}$ – here " softdrop groomed" refers to the algoritm [118] removing low-p_{T} components of jets which are likely to originate from ISR, UE and PU processes. In case of QCD jets, groomed mass will appear to get smaller as compared to groomed mass for V-tagged jets measured relatively close to the value of V boson mass.

Jet definition in HWW analysis

Jets defined in the scope of the $H \to WW^*$ analysis are of two types:

1. AK4 PF jets with CHS algorithm, henceforth denoted simply as jets (also as "resolved jets" in context of associated VH production topology),
2. AK8 PF jets with PUPPI algorithm, henceforth denoted as "boosted V-jets" (or simply V-jets) as these are specifically used as V-tagged jets.

Parameter	$\lvert\eta\rvert \leq 2.6$	$2.6 < \lvert\eta\rvert \leq 2.7$	$2.7 < \lvert\eta\rvert \leq 3$	$\lvert\eta\rvert > 3$
Neutral hadron fraction (AK4)	> 0.9	> 0.9	–	> 0.2
Neutral hadron fraction (AK8)	> 0.9	> 0.9	> 0.99	> 0.02
Neutral EM fraction (AK4)	< 0.9	< 0.99	$[0.02, 0.99]$	< 0.9
Neutral EM fraction (AK8)	< 0.9	< 0.99	–	< 0.9
Charged hadron fraction	> 0	–	–	–
Charged EM fraction	–	–	–	–
N. of neutral particles (AK4)	–	–	> 2	> 10
N. of neutral particles (AK8)	–	–	–	$[2, 15]$
N. of charged particles (AK4)	> 0	> 0	–	–
N. of charged particles (AK8)	> 0	–	–	–
N. of constituents	> 1	–	–	–

Table 3.10: Jet Tight ID selection criteria defined by the JetMET POG for 2018 datasets.

Both jets and V-jets have their four-momenta calibrated by the JEC described above and are requested to pass JetMET POG tight ID criteria. On top of that, analysis-specific "cleaning" procedure was defined for both types independently.

In particular:

- jets associated to LooseHWW leptons (see definition in Sections 3.2.2-3.2.3) via geometric matching inside of $\Delta R < 0.3$ cone radius (only leptons with $p_{\mathrm{T}} > 10$ GeV are affected) are removed,

- jets with $p_{\mathrm{T}} < 50$ GeV are requested to pass loose PU ID selection criteria (described below) to address additional contamination by PU jets,

- finally only jets with $p_{\mathrm{T}} > 30$ GeV within detector acceptance $|\eta| < 4.7$ are counted.

PU ID working points (not to get confused with Jet ID working point) are defined as cuts on dedicated BDT response, trained to identify PU jets using variables defining event and jet topology. Full list of variables and details on BDT training can be accessed in reference [119]. Particular choice of PU ID working point in this analysis, however, is a result of balancing the capability to reject jets from PU interactions and an agreement between data and MC achieved in a fraction of the jet veto efficiency.

Former aspect is studied from the measurement of efficiency (Figure 3.4 left) and purity (Figure 3.4 rifgt) in selecting jets as a function of jet pseudorapidity at the generator-level η^{gen} and reconstruction-level η^{reco} , respectively. Efficiency Eff and purity Pur distributions are measured in simulation of Drell-Yan events (in particular the $Z \to \mu^{+}\mu^{-}$

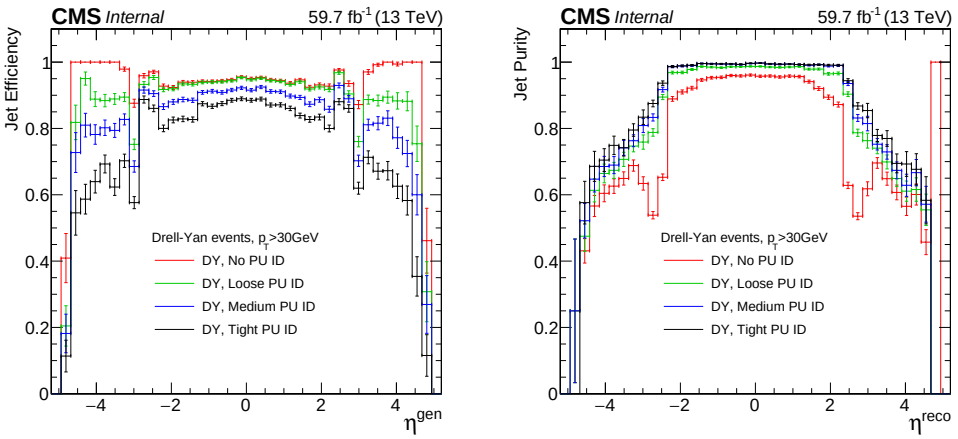

Figure 3.4: Jet selection efficiency (left) and purity (right) as a function of jet pseudorapidity for generator-level η^{gen} and reconstruction-level η^{reco} jets, respectively, as measured in simulation of 2018 Drell-Yan events. Distributions are displayed per jets passing loose PU ID (green), medium PU ID (blue) and tight PU ID (black) in comparison to the case of not applying any PU ID selection (red).

process) within narrow Z boson mass window of $[80, 100]$ GeV and are defined as:

$$\begin{aligned} Eff &= \frac{N_{\text{jet}}^{\text{reco/gen}}}{N_{\text{jet}}^{\text{gen}}}(\eta^{\text{gen}}), \\ Pur &= \frac{N_{\text{jet}}^{\text{reco/gen}}}{N_{\text{jet}}^{\text{reco}}}(\eta^{\text{reco}}), \end{aligned} \tag{3.14}$$

where $N_{\text{jet}}^{\text{reco/gen}}$ is a number of the reconstruction-level jets geometrically matched to the generator-level jets within $\Delta R = 0.4$ cone radius and $N_{\text{jet}}^{\text{reco}}/N_{\text{jet}}^{\text{gen}}$ are total numbers of reconstruction-level and generator-level jets in the event, respectively. Latter aspect is studied by measuring the fraction of 0- and 1-jet events as a function of the number of primary vertices, both in Drell-Yan simulation and in data collected with HLT trigger requiring two muons in the final state and targeting the same Drell-Yan event topology. As depicted in Figure 3.5, reasonable agreement between the MC simulation and data is achieved within statistical uncertainty for all considered PU ID working points. As expected, with tighter PU ID selection criteria, more PU jets is removed from the jet collection, thus distributions appear more flat. Unavoidably, such a reduction comes with a degrading effect for jet efficiency and purity defined by formulae 3.14. Eventually, loose PU ID was selected as a good compromise between two studied aspects.

Different cleaning procedure is employed for V-jets in this analysis. In particular:

- V-jets associated to any leptons via geometric matching inside of $\Delta R < 0.8$ cone radius are removed from the V-jet collection.
- Targeting boosted V hadronic decay in VH production mode, V-jets are selected in the softdrop groomed mass window of $[65, 105]$ GeV. Additionally, ratio of the 2-subjetiness and 1-subjetiness $\tau_2/\tau_1 < 0.45$ was chosen to further differentiate between V-tagged jets consisting of 2 subjets and QCD jets with a single subjet.

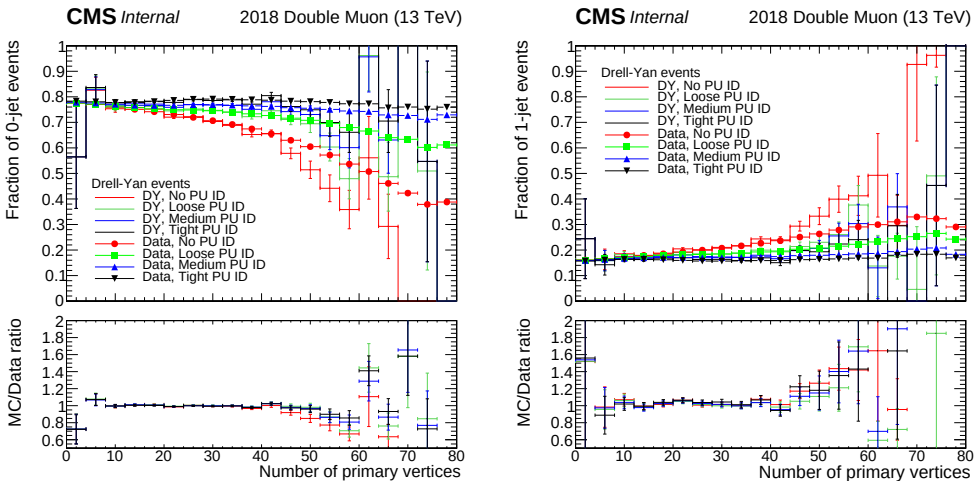

Figure 3.5: Fraction of 0-jet (left) and 1-jet (right) events as a function of a number of primary vertices as measured in 2018 Drell-Yan simulation and Data. Distributions are displayed per jets passing loose PU ID (green), medium PU ID (blue) and tight PU ID (black) in comparison to the case of not applying any PU ID selection (red). Bottom pad is showing MC-to-data agreement respective to each considered PU ID scenario.

- Finally only V-jets with $p_{\mathrm{T}} > 200$ GeV within $|\eta| < 2.4$ detector acceptance are counted.

Eventually, both jet types have their transverse momenta smeared by corresponding c_{JER} factors. Smearing procedure appeared to be particularly useful in correcting the data-to-MC discrepancies observed in 2017 Drell-Yan control region as shown in preliminary jet studies displayed in Figure 3.6. On the other hand, it was decided not to apply the JER smearing for jets in 2016 MC samples. That is because of the presence of residual PU noise in so called "jet horn" region ($2.8\,|\eta| < 3.0$) that undesirably affects calculation of c_{JER} factors (as a result of an excess of non-genuine jets smeared stochastically). Moreover, data-to-MC agreement in 2016 simulation is reasonable also before applying the JER smearing, thus we only keep the JER smearing in 2017 and 2018 MC samples. Furthermore, the JER smearing on V-jet transverse momenta is properly translated onto groomed V-jet mass to account for any residual data-to-MC discrepancy in mass distribution.

3.2.5 Missing transverse energy

Feebly interacting particles, such as neutrinos or possible not-yet-discovered BSM particles are elusive for the CMS detector subsystems. Their presence, however, is manifested by the transverse momentum missing in the overall event energy balance, thus can be reconstructed. Assuming initial conditions with near-to-zero transverse component of the net momentum for the head-on beam collisions, the total transverse momentum reconstructed in the event final state should be compensated by the negative (vector) sum of all transverse momenta associated to the reconstructed PF particles. In other words,

raw missing transverse momentum vector can be defined as

$$\vec{p}_{\mathrm{T,raw}}^{\,\mathrm{miss}} = -\sum_{i=1}^{\text{PF particles}} \vec{p}_{\mathrm{T,i}}. \tag{3.15}$$

Additional corrections are needed to address any inefficiency of reconstructing tracks, non-linearity of the calorimetry response for hadrons and necessity to set minimal p_{T} and energy thresholds in the calorimetry system. This is achieved by propagating any jet related momentum corrections to the $\vec{p}_{\mathrm{T,raw}}^{\,\mathrm{miss}}$:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = \vec{p}_{\mathrm{T,raw}}^{\,\mathrm{miss}} - \sum_{j}^{\text{PF jets}} (\vec{p}_{\mathrm{T,j}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,j}}), \tag{3.16}$$

where the sum is over all uncorrected jets with $p_{\mathrm{T}} > 15$ GeV (to reduce contamination from PU jets) and corresponding $\vec{p}_{\mathrm{T,j}}^{\,\mathrm{corr}}$ stands for the corrected jet p_{T}. Alternatively, expression 3.16, also denoted as PF Missing E_{T} (PF MET), can be recast employing the PUPPI-weighted PF particle candidates (and PUPPI jets) to suppress $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ dependence on PU interactions. In such a case we will use the PUPPI Missing E_{T} (PUPPI MET) notation. Also note that term "missing transverse momentum" used together with notation $p_{\mathrm{T}}^{\mathrm{miss}}$ is often used interchangeably with term "missing transverse energy" and notation $E_{\mathrm{T}}^{\mathrm{miss}}$. To avoid any confusion, both terms refer to the same quantity defined simply by the size of the missing transverse momentum vector $\left|\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\right|$.

MET definition in HWW analysis

Missing transverse momentum reconstruction is sensitive to detector noise (e.g. PU particles, cosmic radiation), beam-halo effect [120], inactive detector material (e.g. non-functional ECAL cells) or any failure in reconstruction of the PF particles. Any anomalous or fake contribution to the $p_{\mathrm{T}}^{\mathrm{miss}}$ object might be affecting analysis phase space selection, especially in case with neutrinos in the final state. To avoid any $p_{\mathrm{T}}^{\mathrm{miss}}$ misreconstruction, we apply dedicated Missing Transverse Momentum (MET) quality filters provided by the JetMET POG to remove both MC and data events affected by this issue.

As already mentioned, PU mitigation in PF reconstruction is directly affecting quality of $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. An example of PF MET and PUPPI MET distributions for 2017 simulation and data is depicted in Figure 3.7. A significant improvement in data-to-MC agreement is achieved in favor of PUPPI MET object. Similar improvement is observed for 2018 dataset, while there is little effect observed in 2016. Henceforth, when applying selection criteria on $p_{\mathrm{T}}^{\mathrm{miss}}$ object in this analysis, we refer to PUPPI-based version of reconstruction, which for the sake of consistency, is employed in all three datasets. The remaining data-to-MC discrepancies in the PUPPI MET reconstruction, as seen in Figure 3.7, are addressed by the corresponding systematic uncertainties on the MET object discussed in Section 6.2.2. Unfortunately, as we use AK4 PF CHS version of jets, it is not possible to propagate the JER smearing to the reconstruction of PUPPI MET (caused by the incompatibility of calculating c_{JER} factors). Propagation of the JEC is not affected and is properly accounted in $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction.

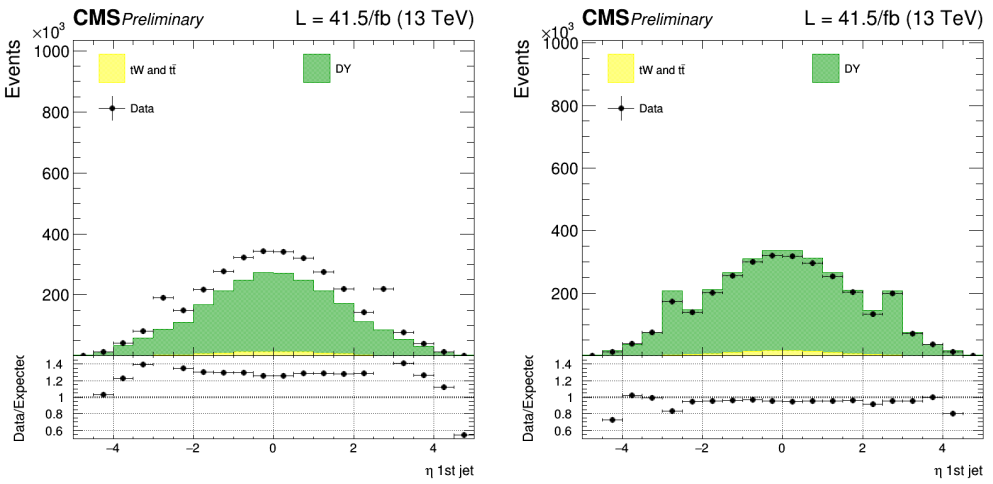

Figure 3.6: Distribution of the leading jet pseudorapidity in 2017 Drell-Yan control region in simulation (Drell-Yan and top production events) and data (double muon HLT trigger) before the jet transverse momentum is smeared (left) and after this procedure (right). Bottom pad displays ratio of events in data over expected MC prediction. This preliminary test served for a validation of the JER factors applied in $H \to WW^*$ analysis for 2017 datasets.

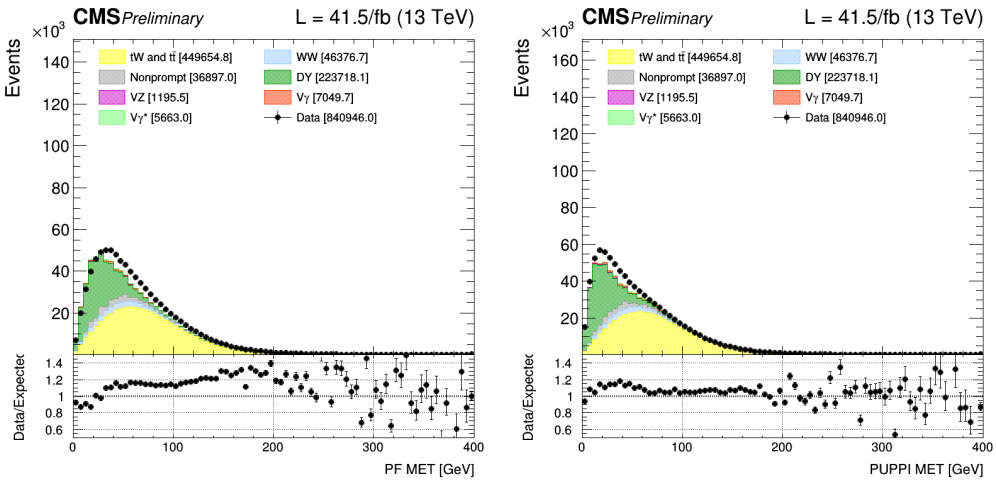

Figure 3.7: Distributions of PF MET (left) and PUPPI MET (right) in jet-inclusive 2017 Drell-Yan control region. Bottom pad displays ratio of events in data over expected MC prediction. Source: [121].

CHAPTER

FOUR

ANALYSIS SETUP

The main focus of this thesis and its related publication [38] is to select and examine pp-collision events that can be identified as candidates for the $H \to WW^* \to e\mu+\text{MET}$ production and decay chain, while searching for possible deviation from the SM expectation. The visible part of the studied final state particle spectrum consists of an electron-muon pair with opposite electric charge (see more details on the analysis selection strategy in Chapter 5). Due to this phase space choice, it is adequate to perform the analysis on datasets collected using single and double lepton triggers, which will be further discussed in Section 4.1. To be able to interpret the observed data and validate each step of this analysis, an appropriate selection of possible background and signal (both SM and BSM) samples was simulated and provided by the CMS MC group. The general configuration used in MC generation is discussed in Section 4.2, while Section 4.3 elaborates further on analysis-specific details regarding the estimation of background processes. Finally, Section 4.4 describes an analytic procedure for enhancing MC statistics (i.e. increasing the number of events) when modeling the BSM signals based on the Matrix Element Likelihood Approach (MELA) [94, 122, 123, 124] – an analytic tool provided by the JHUGen group.

4.1 Datasets and triggers

The datasets employed in this analysis correspond to 138 fb^{-1} of integrated luminosity recorded by the CMS detector during the Run-2 period (2016-2018) of data-taking at the LHC. The datasets are further divided into three parts based on the year when the data was collected,

- 2016: $L = 36.3\ \text{fb}^{-1}$,
- 2017: $L = 41.5\ \text{fb}^{-1}$,
- 2018: $L = 59.7\ \text{fb}^{-1}$.

These parts will be treated as individual analyses to account for different experimental conditions during data-taking. Furthermore, the datasets are subdivided into smaller segments called eras, which are denoted with capital letters in alphabetical order.

The general description of the CMS trigger system was introduced in Section 2.2.5. HLT triggers that were activated during the collection of datasets used in this analysis

Dataset	Era	HLT path
SingleMuon	B–H	HLT_IsoMu24
		HLT_IsoTkMu24
SingleElectr.	B–H	HLT_Ele27_WPTight_Gsf
		HLT_Ele25_eta2p1_WPTight_Gsf
DoubleMuon	B–G	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL
	H	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ
DoubleEG	B–H	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
MuonEG	B–F*	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL
		HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
	F*–H	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ
		HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ

Table 4.1: HLT trigger paths used to collect the 2016 single and double lepton datasets. *Era 2016-F is specific to using a different set of MuonEG triggers up to LHC run number 278272 than above this number.

are summarised independently per each year in Tables 4.1 – 4.3. If at least one of the enlisted triggers was fired, the incident event was saved to the corresponding dataset with a name denoting the minimum number (and flavor) of leptons requested by the trigger logic. In that regard, the primary datasets can be divided into single lepton and double lepton samples, specifically:

- SingleMuon (μ), SingleElectron (e),
- DoubleMuon ($\mu + \mu$), DoubleEG ($e/\gamma + e/\gamma$) and MuonEG ($\mu + e/\gamma$) datasets.

Event exclusivity is assured by prioritizing events firing the more stringent single lepton triggers (thereby falling into the single lepton dataset) over double lepton triggers with relaxed selection criteria. For example, an event that would normally fire both SingleMuon and DoubleMuon triggers is assigned to the SingleMuon dataset.

Apart from the obvious period of trigger validity (i.e. not all triggers were available for all eras), Tables 4.1 – 4.3 display so called HLT path, which encodes the series of consequent requirements placed upon HLT lepton candidates to determine whether the relevant trigger can be fired. These requirements resemble those used for offline lepton reconstruction, as described in Section 3.2. However, the related algorithms are simplified

Dataset	Era	HLT path
SingleMuon	B–F	HLT_IsoMu27
SingleElectr.	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
MuonEG	B	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
		HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL DZ
	C–F	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
		HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ

Table 4.2: HLT trigger paths used to collect the 2017 single and double lepton datasets.

to meet demanding time constraints in online data processing. Tags used in the HLT path are referring to:

- electron/muon HLT candidate track isolation definition ("IsoVL", "TrkIsoVVL" and "Iso" tags),
- identification requirements for electron to be matched with HLT track ("TrackIdL") and/or ECAL cluster ("CaloIdL"),
- tighter requirements for electron identification in case of SingleElectron dataset ("WPTight"),
- the application of the Gaussian Sum Filter method in electron tracking ("Gsf"),
- requirements for muon candidates to be recognised as Tracker muons ("TkMu"),
- various kinematic requirements, e.g. "Ele25_eta2p1" selects HLT electron candidates with $p_{\mathrm{T}} > 25$ GeV and pseudorapidity $|\eta| < 2.1$; the "DZ" tag places additional requirements on the longitudinal displacement (along the beamline) between two leptons required by double lepton triggers.

Once collected, all datasets used in this analysis are subject to the same offline object reconstruction (and later, the phase space selection) as in the simulated samples.

4.2 MC configuration

Details on various aspects of the MC simulation were discussed in Section 3.1. Both the Higgs boson signal samples (SM or BSM) and most of the background processes were

Dataset	Era	HLT path
SingleMuon	A–D	HLT_IsoMu24
SingleElectr.	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
MuonEG	A–D	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ

Table 4.3: HLT trigger paths used to collect the 2018 single and double lepton datasets.

modeled exclusively using the MC generators as summarised in the following sections as well as in Table 4.4 and Table 4.5. All the samples are simulated using the same Parton Distribution Functions (PDF) provided by the NNPDF Collaboration. In particular, the 2016 samples are using PDF version NNPDF3.0 [88], while version NNPDF3.1 [89] was used for the simulation in 2017 and 2018. The corresponding PDF order in QCD (LO or NLO) matches the order of the MC generator used for the particular sample. For all the considered MC samples, processes such as parton showering, hadronisation and the underlying event (UE) are addressed by Pythia 8.2. The full set of parameters used to simulate the UE using the Multiple-Parton Interaction (MPI) method is denoted as a "tune". In this analysis, we follow the recommendation of the CMS MC generator group and use the CUETP8M1 tune [125] for the 2016 samples and the CP5 tune [126] for the 2017 and 2018 samples. Some of the background processes, e.g. the nonprompt lepton background or Drell-Yan production, were estimated using data-driven methods, and we leave this discussion for Section 4.3.

4.2.1 Signal samples

The SM Higgs boson signal samples for gluon-gluon fusion (ggF), vector boson fusion (VBF) and vector-associated Higgs boson production (WH or ZH) are simulated using Powheg (2016) and Powheg2 (2017/18) MC generators at NLO accuracy in QCD. Jet-inclusive ggF production samples are generated using the MiNLO approach and re-weighted to NNLO accuracy in both Higgs boson transverse momentum and the number of hadronic jets employing the NNLOPS scheme [127]. The MiNLO HVj [128] extension of the Powheg generator is used to provide NLO accuracy for the WH and quark-initiated ZH production (see the diagram in Figure 1.3 right) with 0 or 1 hadronic jets in the final state. No additional treatment is used in generating SM VBF processes. In case of ggF, VBF and quark-initiated ZH production, the Higgs boson decay into a pair of W bosons (decaying leptonically) is simulated using JHUGen 5.2.5 (2016) and JHUGen 7.1.4 (2017/18) MC generators. All other SM signal samples (including gluon-induced ggZH production) have their decay chain simulated using Pythia 8.226 (2016) and Pythia 8.230 (2017/18). WH production samples were simulated with an inclusive

Process	Sub-process	σ[fb]	Production/Decay generator
ggF		991.3	2016: POWHEG/JHUGEN 5.2.5, 2017/18: POWHEG2/JHUGEN 7.1.4
VBF		84.6	2016: POWHEG/JHUGEN 5.2.5, 2017/18: POWHEG2/JHUGEN 7.1.4
WH	W^+H	135.0*	2016: POWHEG/PYTHIA 8.226, 2017/18: POWHEG2/PYTHIA 8.230
	W^-H	116.0*	2016: POWHEG/PYTHIA 8.226, 2017/18: POWHEG2/PYTHIA 8.230
ZH	quark-initiated	17.1	2016: POWHEG/JHUGEN 5.2.5, 2017/18: POWHEG2/ JHUGEN 7.1.4
	ggZH	2.75	2016: POWHEG/PYTHIA 8.226, 2017/18: POWHEG2/PYTHIA 8.230

Table 4.4: List of the SM signal samples considered in this analysis. The displayed cross-section σ values correspond to the Higgs boson production and the decay chain $H \to WW^* \to 2l2\nu$ with the exception (*) of the HW production samples simulated with an inclusive final state decay, i.e. $H \to WW^*$. The version of the employed MC generators differs per dataset.

final state decay, i.e. $H \to WW^*$. Nevertheless, proper analysis phase space is assured by employing selection cuts discussed in the next chapter. Other SM signal processes such as $t\bar{t}$H and $b\bar{b}$H were simulated, however, they were found to have a negligible impact on this analysis. The Higgs boson mass in the event generation is assumed to be 125 GeV, while value of 125.38 GeV [129] is used for the cross-section calculation.

For each Higgs boson production topology that is also considered in the SM case (i.e. ggF, VBF, WH and ZH) we assume a possible AC contribution to the HVV vertex. Alternative signal samples were generated for the a_{Λ_1}, a_2 and a_3 coupling scenarios independently, considering the full BSM effect (i.e. $f_{a_i} = 1$). Additionally, mixed SM-BSM scenario samples (i.e. $f_{a_i} = 0.5$) were also produced per each of the production processes. Both types of the alternative samples were simulated using JHUGEN 6 (2016) and JHUGEN 7.1.0 (2017/18) MC generators at LO in QCD. The consequent $H \to WW^* \to 2l2\nu$ decay is simulated employing the same MC generators and assuming identical coupling values as in the production vertex. Alternative signal samples for the $a_{\Lambda_1}^{Z\gamma}$ scenario were obtained by the MELA re-weighting procedure (see Section 4.4).

As a matter of fact, SM signal samples were generated with the NLO MC generator (POWHEG) as compared to the LO (JHUGEN) used in generation of alternative signal samples. To address possible issues, we have studied the shape of the kinematic discriminants (variables constructed to be sensitive to the AC signal – see more in Chapter 5) for

Process	Sub-process	σ[pb]	Generator
Top production	Single $t/\bar{t}$ (t-channel)	70.69	POWHEG2
	Single t (s-channel)	3.36	POWHEG2
	Single $tW/\bar{t}W$	71.20	POWHEG2
	$t\bar{t}$	87.31	POWHEG2
Nonresonant WW	quark-initiated WW	12.18	POWHEG2
	gluon-induced WW	0.57	MCFM v7.0
	EWK WW	0.43	MADGRAPH
Drell-Yan	$Z \to ee/\mu\mu/\tau\tau$	24635.20	MADGRAPH
	$Z \to \tau\tau \to e\mu + \text{MET}$	255.63	MADGRAPH
Multiboson	Vγ^*($>$ 100 MeV)	58.83	POWHEG2
	Wγ^*($<$ 100 MeV)	1758.00	PYTHIA 8.212
	Wγ	1816.83	MADGRAPH
	VZ	10.59	POWHEG2
	VVV	0.42	MADGRAPH

Table 4.5: List of the simulated background samples considered in this analysis. The displayed cross-section σ values correspond to the inclusive two-lepton final state. The "V" symbol stands for both the W and Z bosons.

the same SM signal process generated by different order generators. An example of such a test with no significant differences observed can be viewed in Figure 4.1. For the sake of consistency with similar analyses, the alternative signal samples were re-weighted to match the expected yields of the SM signal samples. Note that the nature of the studied phenomena is measured by means of a quantity relative to the SM values (e.g. effective fractional cross-section) implying that re-weighting of this kind have no measurable effect.

4.2.2 Background samples

Any process occurring in the pp-collision at the LHC that results in a similar final state topology as our studied signal phase space is considered background. Even though we study signal processes with anomalous contributions to the Higgs boson production and decay, we rely on background composition that is known from SM measurements, e.g. [130], as the measurable kinematic differences between SM and BSM signal contributions are negligibly small relative to background contamination.

One of the most relevant background contributions comes from the $t\bar{t}$ pair (Figure 4.2 left), single top tW, single top s-channel and t-channel production processes generated

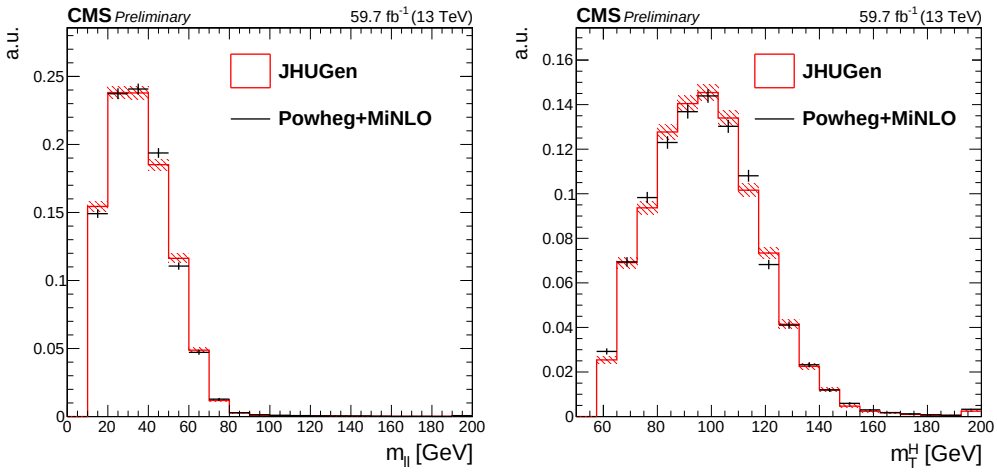

Figure 4.1: Reconstruction-level $m_{\ell\ell}$ (left) and $m_{\mathrm{T}}^{\mathrm{H}}$ (right) distributions (see definitions in Equations 5.1 and 5.2) for the SM Higgs signal samples in the ggF + 0-jet signal region, as defined in Table 5.2. The compared distributions are reconstructed from samples generated by JHUGen (red histogram) and Powheg+MiNLO (black points) approach. All distributions are normalised to unity and display MC statistical uncertainty (shaded bands on the histogram and vertical bars on the points).

with Powheg2. Top quarks decay in $\approx 1/3$ of the cases to the Wb pair (Figure 4.2 right) with W eventually decaying leptonically, thus mimicking a topology with the W boson from the Higgs boson decay.

Another significant source of contamination is caused by nonresonant WW production. We consider the quark-initiated WW process (Figure 4.3 left) generated with Powheg2 at NLO accuracy, gluon-induced WW process simulated with the MCFM v7.0 generator [131, 132, 133] at LO accuracy that is renormalized to match NLO cross-section [134], and nonresonant Electroweak (EWK) WW process with two jets in the final state, modeled at LO accuracy with MadGraph5_aMC@NLO v2.4.2.

The Drell-Yan (DY) production (Figure 4.3 right) of a charged lepton pair is simulated with MadGraph5_aMC@NLO v2.4.2 at NLO accuracy to complement the data-driven estimation of DY background discussed in Section 4.3.1. This background is largely reduced thanks to the requirement of a different flavor ($e\mu$) lepton pair in the final state, however, it is not negligible, as the $Z/\gamma^* \to \tau\tau$ decay can lead to observed $e\mu$ pairs.

Diboson processes with at least one Z boson or one virtual photon γ^* with a mass above 100 MeV are simulated at NLO accuracy with Powheg2, while Wγ^* production for a mass below 100 MeV was generated using Pythia 8.212 in the parton showering of Wγ events. Simulated is also the process with a W boson produced in association with a photon from initial state radiation (Wγ) and triboson (VVV) production process, both generated at NLO accuracy with MadGraph5_aMC@NLO v2.4.2. All processes mentioned in this paragraph are summarily denoted as "Multiboson" production.

Other background processes, such as production of the Higgs boson in the $H \to \tau\tau$ decay channel, were simulated, however, they were found to have a negligible effect on this analysis.

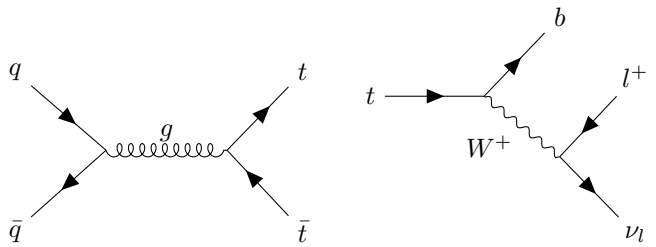

Figure 4.2: Example of LO $t\bar{t}$ production at the LHC (left). Top quarks decay in $\approx 1/3$ of the cases into a pair of W boson decaying leptonically and b quark inducing the evolution of the b-jet (right).

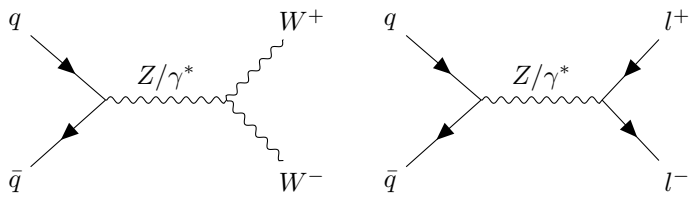

Figure 4.3: Example diagrams of LO non-resonant WW production (left) and the Drell-Yan process (right) with $l = e, \mu, \tau$.

4.2.3 Data-to-MC agreement

One of the most important tasks in every HEP analysis is to assure consistency between collected data and simulation at every possible predictable level of understanding of the studied processes. A couple of effects deteriorating the observed data-to-MC agreement, which are directly related to the dataset triggers, PU effects, heavy-flavor tagging and object reconstruction, are discussed as follows. Wherever applicable, remaining discrepancies are addressed by assigning appropriate systematic uncertainties as discussed in Chapter 6.

Trigger efficiency corrections

Trigger logic is not applied to MC samples, which gives a rise to differences in event selection efficiency between collected datasets and simulation. Any data-to-MC discrepancy caused by inefficiency of triggers needs to be addressed. This is performed by calculating the event-by-event weights using a method known as "Tag&Probe" and applying them on top of the generated MC samples. As this method is commonly used to calculate efficiency for arbitrary selection, identification or isolation criteria, we will introduce it on a more general basis and later focus on its specific application in calculating trigger efficiency.

The Tag&Probe method uses well-known mass resonances, such as the Z boson, with both the "tag" and "probe" potentially identified as leptons coming from the Z decay. The tag lepton is usually requested to pass very tight selection criteria to ensure it genuinely originates from the Z boson decay. Probes are leptons selected with less strict criteria, nevertheless, they are matched to the tag lepton only if they form a lepton pair with a dilepton invariant mass close to the m_Z value. Probes are further requested to pass a particular selection working point – the one for which efficiency is measured. In

particular, efficiency can be found as a function of lepton pseudorapidity η and transverse momentum p_{T}, given by the equation:

$$\epsilon(\eta, p_{\mathrm{T}}) = \frac{N_{\mathrm{pass}}^{\mathrm{probes}}}{N_{\mathrm{pass}}^{\mathrm{probes}} + N_{\mathrm{fail}}^{\mathrm{probes}}}, \tag{4.1}$$

where $N_{\mathrm{pass/fail}}^{\mathrm{probes}}$ are the numbers of probes passing/failing the desired selection working point in a given (η, p_{T}) region. Efficiency is further determined for both the data region enriched in targeted resonance decays and MC simulation. The desired MC weight is given as a fraction of these values and applied event-by-event. Note that leptons used in this example could, in principle, be generalised for any known physical object as long as they are originating from a known resonance decay.

In the case of calculating lepton trigger efficiency, a particularly clean region was targeted by selecting events with two leptons of opposite charge and a dilepton invariant mass $m_{ll} \in [60, 120]$ GeV. Single lepton efficiencies were calculated by requesting probes to pass the corresponding HLT trigger path filters. The probability of an event firing a single lepton trigger, denoted as P(S), is then directly given by the single lepton efficiency ϵ_{S} (with S standing for SingleElectron or SingleMuon). Calculating double muon efficiencies requires a more involved combinatorial approach. The probability of an event firing a double lepton trigger P(D), for example D = MuonEG, is given by the formula:

$$\begin{aligned} \mathrm{P}(\mathrm{MuonEG}) &= \mathrm{P}(e\mu) + \mathrm{P}(\mu e) - \mathrm{P}(e\mu \wedge \mu e) \\ &= \epsilon_{e\mu}^{\mathrm{leg1}} \epsilon_{e\mu}^{\mathrm{leg2}} + \epsilon_{\mu e}^{\mathrm{leg1}} \epsilon_{\mu e}^{\mathrm{leg2}} - \epsilon_{e\mu}^{\mathrm{leg1}} \epsilon_{\mu e}^{\mathrm{leg1}}. \end{aligned} \tag{4.2}$$

Here, $\mathrm{P}(e\mu)$ and $\mathrm{P}(\mu e)$ stand for the "per-leg" probabilities of an event passing the MuonEG trigger with the higher p_{T} cut ("leg1") on the electron and muon candidate, respectively. Term $\mathrm{P}(e\mu \wedge \mu e)$ is then correcting for the combinatorial overlap of possible cases. The second row of Equation 4.2 expresses the same idea using the per-leg double lepton efficiencies calculated by the Tag&Probe method. The single electron and single muon per-leg efficiencies exceed 80% for the entire pseudorapidity range (where the electrons/muons are reconstructed), as measured for the 2017 and 2018 primary datasets. For the 2016 datasets, the single electron (muon) efficiency remains in the $50 - 80\%$ ($> 80\%$) range. The double lepton efficiencies exceed 90% in 2017 and 2018. In the 2016 case, the double electron (double muon) efficiency stays above 90% (80%).

Eventually, per-event lepton trigger efficiencies $\mathrm{P}(\mathrm{S} \vee \mathrm{D})$ are extracted by a similar combinatorial approach, combining single and double lepton probabilities:

$$\mathrm{P}(\mathrm{S} \vee \mathrm{D}) = \mathrm{P}(\mathrm{S}) + \mathrm{P}(\mathrm{D}) - \mathrm{P}(\mathrm{S} \wedge \mathrm{D}). \tag{4.3}$$

As indicated by some of the double lepton HLT paths with a "DZ" tag, the overall trigger efficiency also depends on the requirement for the longitudinal displacement between two leptons. The corresponding DZ-efficiency is determined as a ratio between the number of events passing the double lepton trigger with DZ requierements (where applicable) and those passing it without this selection criteria. DZ-efficiency was derived as a function of the number of PU interactions, ranging between the $91 - 99\%$.

Given that triggers are solely applied to the data, the per-event efficiencies (both the lepton trigger and DZ trigger) are calculated in the data and then directly applied as per-event weights in MC.

Prefiring corrections

In 2016 and 2017, a gradual timing shift of the ECAL endcap was incorrectly propagated to the L1 trigger, causing highly energetic readout from the e/γ and/or jet candidate deposits to be incorrectly assigned to the previous bunch-crossing. Moreover, the L1 trigger is set up to rule out two consecutive bunch-crossings from firing, thus allowing events to veto themselves in cases with a significant amount of ECAL energy in the $2 < |\eta| < 3$ forward region. Due to the nature of this issue, this effect is commonly known as "prefiring". As implied above, the magnitude of prefiring effect depends on the pseudorapidity and transverse momentum of the firing object candidate, therefore it is partially process-specific, i.e. it ranges between 3% yield loss for Drell-Yan events to 5% loss for VBF Higgs boson production with associated energetic jets expected in the forward region.

As this is a systematic effect introduced for a limited period of time, it was not simulated in MC samples. To account for possible discrepancies, corresponding corrections in the form of event-by-event weights were provided by the JetMET POG group assuming the following formula:

$$w = \prod_{i=\{e/\gamma,\text{jet}\}} (1 - \mathrm{P}_i(\eta, p_\mathrm{T}))\,, \tag{4.4}$$

with P_i representing the probability of e/γ or jet candidate firing the L1 trigger as calculated on datasets selected specifically to contain only non-prefirable events.

Pile-up re-weighting

MC samples are simulated taking into account the expected PU effects for each targeted period of data-taking, as discussed in Section 3.1.4. To correct the remaining discrepancy, the real PU conditions are retrieved from actual data-taking (per each year) by measuring the ratio of the normalised distributions of inelastic collisions over the simulated MC pile-up profile. Weights are determined per event as a function of the generated number of inelastic collisions. The positive impact of the PU re-weighting on the data-to-MC agreement can be directly observed in Figure 4.4, which displays a distribution of a number of primary vertices both with and without PU weights.

Jet tagging scale factors

In Section 3.2.4, we mentioned the V-tagging and b-tagging procedures used to identify jets coming from boosted hadronic V boson decay and those induced by bottom quarks, respectively. Such procedures, whether based on multivariate techniques or simple selection cuts, are prone to have different efficiencies when performed in MC simulation or data. For both cases, tagging efficiencies (corrected for the mistagging rate) are measured by the Tag&Probe method and are used to construct correction scale factors applicable to each jet. In particular, the b-tagging (V-tagging) efficiency for the working point selected in this analysis is typically around 84% (65%). Eventually, the event-by-event weights are produced as a product of jet-by-jet scale factors for all jets passing or failing the given tagging procedure and are derived as a function of jet η, p_T, flavor, and the tagging discriminator itself. A detailed procedure, including the efficiency and mistagging rate studies, can be reviewed in [116] (b-tagging) and [117] (V-tagging) as provided by the b-tag & vertexing POG (BTV POG) and JetMET POG, respectively. We also

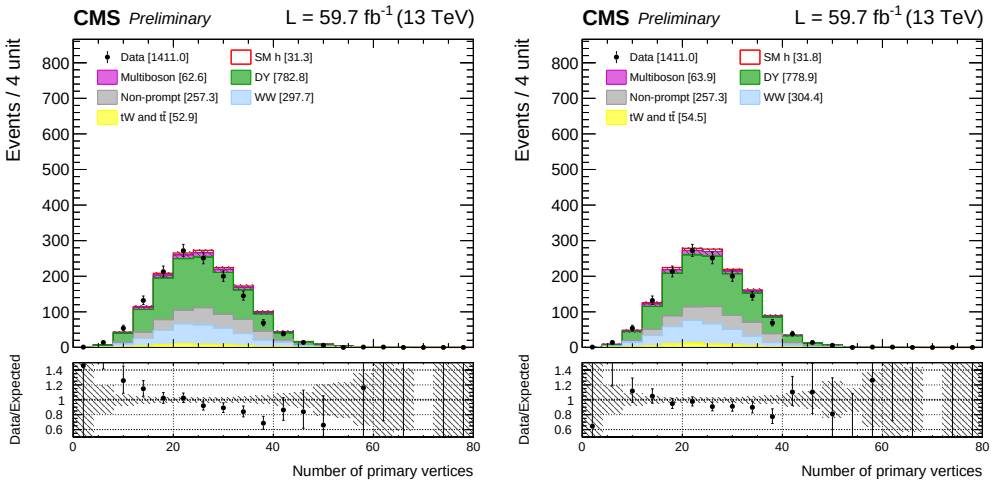

Figure 4.4: Distributions of the number of primary vertices with (right) and without (left) applying PU weights. These distributions are shown for the control region rich on Drell-Yan events with zero associated jets in the final state.

pay extra attention to the b-tagging procedure in Section 4.3.2, as it is an integral part of the top background estimation.

Corrections on object identification

In general, the efficiency of reconstructing leptons and/or jets in MC simulation differs from the efficiency achieved in the data. To account for the corresponding differences in MC, dedicated scale factors are derived and applied in the form of event-by-event weights.

The lepton identification scale factors are derived as the ratio of the efficiency in selecting leptons with the ID working point between MC and the data. Efficiencies are measured employing the Tag&Probe method as a function of lepton p_{T} and η. In particular, the efficiency to reconstruct and identify muons is greater than 96% for most of the kinematic regions, while for electrons, it ranges between $60 - 80\%$ depending on the electron's p_{T}.

In the case of jets, the reconstruction efficiency/purity was shown to heavily depend on the selected PU ID working point. As shown in Figure 3.4 for the Loose PU ID, the efficiency/purity exceeds 90% in the central pseudorapidity region and drops for higher $|\eta|$ values. Similar study was performed centrally by the CMS JetMET POG, which came to the same conclusion. Jet PU ID efficiency/purity was measured in both MC and data samples targeting $Z+$jets events. The employed measurement method relied on fitting azimuth angle distribution ϕ between reconstructed Z boson and jets. In this topology, genuine jets are expected to fly back-to-back with reconstructed Z boson (forming a peak around $\phi = \pi$), while PU jets appear randomly distributed. Performing this fit simultaneously for both jets passing/failing the PU ID working point is used to constrain the efficiency/purity for all jets entering the analysis phase space.

4.3 Background estimation

This section is dedicated to details on the background estimation procedure for cases when plain MC simulation discussed in previous sections could be improved. Alternative data-driven methods are introduced in the case of nonprompt lepton background (also called "fake") and Drell-Yan (DY) $Z/\gamma^* \to \tau\tau$ process. Empirical knowledge of dedicated phase space Control Region (CR) in data, i.e. kinematic regions known to be enriched in a particular process, is often used to determine further corrections on background shape and expected yields (normalisation) in top, WW, DY as well as WZ/Wγ^*/Wγ background contributions. Several improvements in simulating kinematic distributions are also discussed.

Note that the methodology for treating background processes, presented below, was originally developed for various analyses outside of the scope of this thesis and has since been adopted as an integral part of this and similar studies probing $H \to WW^*$ decay channels.

4.3.1 Drell-Yan events

Background from the DY process $Z/\gamma^* \to \tau\tau$ is estimated using the data embedding technique [135]. As a first step, embedded samples are constructed from data by selecting events with the $Z/\gamma^* \to \mu\mu$ topology. Consequently, the building blocks (tracks, calorimeter cells, etc.) used in the reconstruction of two muons are removed from this event and replaced with the simulated pair of tau leptons, with kinematic properties copied from the properties of removed muons. This method allows us to estimate the contribution of events passing the double lepton triggers with significantly mitigated uncertainties on the description of object kinematics as compared to the generated MC samples. The remaining DY contribution from events passing the single lepton triggers (only around 5%) is estimated using pure MC simulation.

While embedded DY samples were provided by the $H\tau\tau$ Physics Analysis Group (PAG), it is necessary to account for possible discrepancies between data and embedded samples. In particular, the efficiencies for the identification, isolation and trigger requirements applied in the selection and lepton (electron or muon) reconstruction need to be considered and are determined by means of the Tag&Probe method. Similar to the procedure described in Section 4.2.3, pairs of probe leptons with looser selection criteria and tagged leptons with tighter selection are tested to pass/fail particular identification/isolation/trigger working points under study. In this case, however, we do not possess MC samples serving as a reference. Instead, $H\tau\tau$ PAG released special data samples with both muons in the event replaced by the simulated pair of muons or electrons (similarly as in the case of producing tau embedded samples). Efficiencies are obtained as a function of lepton pseudorapidity and momentum, with additional dependence on azimuth angle in the case of efficiency to identify electrons, and are further used in determining the event-by-event weights.

It appears that about 2% of events with two muons selected at the beginning of the embedding procedure do not originate from the DY process. To prevent double counting of events contaminating embedded samples, which are also simulated in the MC samples, we scale DY simulation events with the so called DY veto. In that regard, simulated events with two final state leptons originating in tau decay are removed.

Eventually, to address any residual discrepancy in the overall population of DY events in the studied signal region, we define a dedicated DY Control Region (DY CR) enriched in DY events and leave the parameter controlling this normalisation freely floating in

the signal extraction fit. More about the DY CR and its definition is to be found in Section 5.1.3.

4.3.2 Top production

As mentioned in Section 4.2.2, single top and $t\bar{t}$ production processes are among the most relevant sources of contamination in the dilepton final state signal region. Fortunately, the top production topology has a recognisable signature, as one can search for the presence of b-jets in the event, indicating the $t \to Wb$ decay. Requiring exactly zero b-jets in the signal region (also referred to as b-veto) significantly reduces the top background. Under these circumstances, the development of an efficient b-tagging procedure becomes of utmost importance. Several algorithms for computing b-tagging discriminators were provided by the BTV POG, with the one selected for this analysis discussed below.

B-tagging is a probabilistic method where each jet is assigned a "likelihood" of containing hadrons constituted of bottom quarks (b-hadrons). This is made possible by the characteristic properties of b-hadrons, as observed in the studies of track and secondary vertex information. For example, b-hadrons typically have a lifetime on the order of 1.5 ps, while hadrons containing c quarks have a lifetime of 1 ps or less, resulting in a specific pattern for b-hadrons. Notably, the tracks used to reconstruct the secondary vertex marking the b-hadron decay exhibit distinctive displacement characteristics compared to tracks originating from c-hadron decays. Kinematic variables describing this kind of information, such as impact parameters between the displaced tracks and the primary vertex, were used as an input to the Deep Combined Secondary Vertex (DeepCSV) algorithm [116] – a deep neural network trained to distinguish between heavy-flavor and light-flavor jets. The particular choice of the loose working point (i.e. the cut on the b-tagging discriminator) was motivated by the best performance in selecting signal events after applying b-veto in simulated top and WW background samples. It corresponds to a b-tagging efficiency of $\approx 84\%$ and mistagging rate of $\approx 10\%$ ($\approx 40\%$) for the light-flavor jets (c-jets). The difference between b-tagging efficiency in simulation and data is taken into account by applying dedicated MC scale factors obtained using the Tag&Probe method in various control samples.

In addition to efficiency corrections, it was observed that the top p_{T} spectrum is softer in data as compared to the simulated $t\bar{t}$ samples. Event-by-event weights,

$$w_{t\bar{t}} = \sqrt{f(p_{\mathrm{T}}^{\mathrm{top}}) f(p_{\mathrm{T}}^{\text{anti-top}})}, \tag{4.5}$$

were derived to correct on this issue and applied exclusively to $t\bar{t}$ events. In this formula, the function f depends on the transverse momenta of the top and anti-top quarks, respectively. Implicitly, it also depends on the choice of the tune version employed in simulating UE effects. The particular form used in this analysis for the 2017/18 samples (CP5 tune) reads

$$f_{\mathrm{CP5}}(p_{\mathrm{T}}) = 0.103 \exp(-0.0118 p_{\mathrm{T}}) - 0.000134 p_{\mathrm{T}} + 0.973. \tag{4.6}$$

The parameters entering this function were determined from the fit on the $t\bar{t}$ samples rescaled to match NNLO QCD and NLO EWK prediction based on the theoretical cross-section obtained from reference [136] and according to prescription provided by the Top PAG. An additional re-weighting factor for the CUET tune employed in the 2016 samples is derived in the fit by comparing p_{T} distributions in 2017/18 and 2016:

$$f_{\mathrm{CP5}\to\mathrm{CUET}}(p_{\mathrm{T}}) = \exp(1.615 \times 10^{-3} + 3.467 \times 10^{-6} p_{\mathrm{T}} - 8.906 \times 10^{-8} p_{\mathrm{T}}^2). \tag{4.7}$$

Similar to dealing with process normalisation with DY background, we define a dedicated Top Control Region (Top CR) enriched in top production events and leave parameter controlling this normalisation freely floating in the signal extraction fit. More about the Top CR and its definition is to be found in Section 5.1.3.

4.3.3 Nonresonant WW background

Nonresonant WW background was estimated using a combination of MC simulation and dedicated control regions in the data used to constrain WW normalisation.

Both quark-initiated and gluon-induced nonresonant WW production processes are considered in simulation. Events produced in the former case are re-weighted to match the diboson p_{T} spectrum computed at NNLO QCD accuracy [137, 138]. Gluon-initiated production accounts for only $\approx 5\%$ of the total cross-section, but remains considered. Additionally, we simulate the EWK production of WW pairs with 2 associated jets, which is a non-negligible process, especially in regions targeting VBF and VH signals.

Similar to dealing with process normalisation for top and DY backgrounds, we define a dedicated WW Control Region (WW CR) enriched in WW production – this time by selecting events with a high dilepton invariant mass m_{ll} (see more in Section 5.1.3). The parameter controlling WW normalisation is left freely floating in the signal extraction fit, as in the previous cases. This approach, however, is only applied for the phase space targeting VBF/VH topology with 2 associated jets. In the case of ggF topology with 0 or 1 associated jets, the normalization parameters are constrained directly in the signal regions (i.e. no WW CR is defined for 0/1-jet channels), which span the high m_{ll} phase space rich in WW events.

4.3.4 Nonprompt lepton background

Nonprompt leptons originating from leptonic heavy quark decays, hadrons misidentified as leptons, and electrons from photon conversion are suppressed by applying lepton identification and isolation criteria, as discussed in Sections 3.2.2 and 3.2.3. The remaining nonprompt leptons originate mostly from W+jets production, where one of the leptons is genuine (coming from W decay) and the second corresponds to the jet misreconstructed as a lepton. The final state topology with two leptons, one of which is fake, and with missing transverse momentum, is kinematically identical to our targeted signal phase space, thus it cannot be reduced by applying the usual selection cuts.

Instead, the residual nonprompt lepton background is estimated by selecting a control sample of events enriched in fake leptons and extrapolating their abundance in the designated signal region. This method is entirely data-driven as both the selected control sample and the extrapolation factor are derived from data rather than MC simulation. In particular, a control sample is selected from data by requesting one lepton that passes the TightHWW working point (WP), i.e. it is a likely-to-be-genuine lepton, and another that fails this working point but passes the LooseHWW requirements, i.e. it is a likely-to-be-fake lepton. To properly account for the presence of nonprompt background in the signal region, the extrapolation factor is derived by determining the so called "fake" (f) and "prompt" (p) rates and is applied to the control sample in the form of event-by-event weights. In this step, it is assumed that once the lepton kinematics are taken into account, the same fake and prompt rates can be employed in different analysis signal regions.

In this analysis, the fake rate measures the fraction of jets satisfying (lepton) LooseHWW WP while also passing (lepton) TightHWW criteria, over jets passing only (lep-

ton) LooseHWW criteria. It can be seen as a probability of a nonprompt lepton to be identified as a genuine lepton. The fake rate is measured in the dataset collected with HLT triggers requesting an abundance of QCD multijet events. On the other hand, the prompt rate defines the ratio of genuine leptons passing the same selection criteria, i.e. it defines a probability of LooseHWW genuine lepton to be identified also as TightHWW genuine lepton. As such, it can be used to remove prompt lepton contamination from the control sample. The prompt rate is measured by means of the Tag&Probe method using the dataset enriched in DY events.

Once the f and p rates are determined, they are incorporated into the weights w applied to the data events passing the usual analysis selection with only the LooseHWW criteria required for leptons. In particular, an $\approx fp/(p-f)$ multiplicative factor enters the w formula when the lepton satisfies only the LooseHWW requirement, while an $\approx f(p-1)/(p-f)$ factor is used for the case when the lepton is also passing the TightHWW criteria. More details about constructing the extrapolation factor and derivation of the prompt and fake rates can be reviewed in [130].

4.3.5 Multiboson production

The WZ background contribution is estimated in MC simulation together with the Wγ^* process, while these two modes are distinguished by a 4 GeV cut on the Z/γ^* candidate mass. In the two charged lepton final state targeted in this analysis, WZ and Wγ^* processes can only contribute if one out of the three leptons (from W and Z/γ^* decay) is not identified. In that regard, data-to-MC scale factors are derived in a three lepton CR, as defined in [130]. As mentioned in Section 4.2.2, the Wγ process was also estimated in MC, as incoming photons might experience conversion in detector material. MC simulation is further validated using data in CR, requesting events with a leading μ and a sub-leading e with the same electric charge sign and angular separation within a $\Delta R < 0.5$ cone. The VVV triple-boson production is a minor background and is estimated in MC simulation.

4.4 Signal modeling

As discussed in Section 4.2.1, three types of the signal MC samples were simulated: the pure SM hypothesis, the pure AC hypothesis and a mixture of both types of events. The presence of an anomalous contribution to the HVV scattering amplitude, however, introduces all sorts of the interference effects and needs to be treated at the quantum-mechanical (i.e. probabilistic) level. Note that the SM-BSM mixed sample was produced assuming a single point corresponding to the $f_{a_i} = 0.5$ scenario, thus cannot be considered as a full parameterisation of the interference.

This section aims to introduce an analytic procedure of constructing a template-based signal model that can eventually be used for the extraction of the AC signal in the fit. Given the big number of assumed hypotheses and corresponding interference terms, this procedure largely relies on the ability of the MELA analytic tool to re-weight existing MC samples to any other desired hypothesis (assuming the same production process). In addition, employing the analytic procedure of enhancing MC statistics significantly reduces the computational complexity of MC generation.

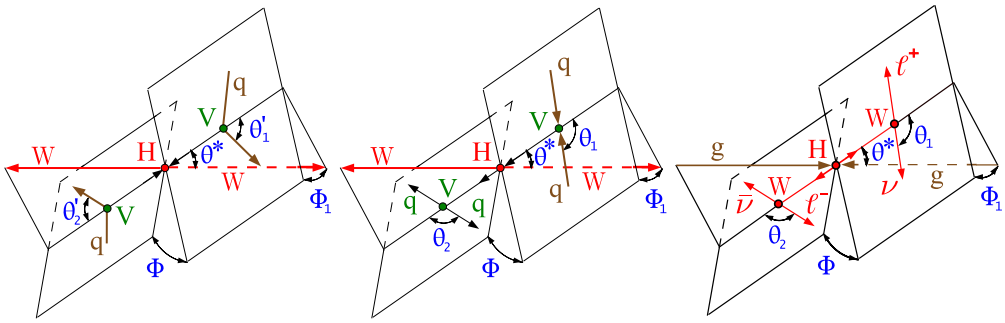

Figure 4.5: Topologies of Higgs boson production and decay for the VBF process $qq' \to qq'H$ (left), VH process $qq' \to VH$ (middle), and the ggF process with the $gg \to H \to WW \to 2l2\nu$ decay chain (right). The incoming partons are shown in brown, the intermediate vector bosons and their decay products are shown in green, the Higgs boson and its vector boson daughters as well as their leptonic decay products are shown in red. The angles characterizing kinematic distributions are shown in blue. Figure is based on the original from [64].

4.4.1 Topology of the signal events

The Higgs boson production and decay topology are explicitly depicted in Figure 4.5 for the VBF, VH and ggF processes. This topology depends on the assumed AC hypothesis, as illustrated in Figure 4.6, which displays several angular distributions (see definitions below) for the generator-level particles involved in or emerging from the studied processes. This is a profound observation which is fully employed by the MELA procedure of calculating the production and decay Matrix Elements (ME) as further discussed in the following subsection. For a more detailed description, refer to [123], [94] or [139], which is also discussed below.

The MELA analytic tool was designed to identify and reduce the number of independent parameters needed to keep full kinematic information about Higgs production and decay. In the most general case, a set of 8 parameters can be used to describe these events:

- the square of the vector boson momentum exchange four-vectors $q^2_{\mathrm{V}1}$ and $q^2_{\mathrm{V}2}$,
- the azimuth ϕ^* between the final state W bosons and initial state partons as measured in the Higgs boson rest frame,
- the polar θ^* angle between the final state W bosons (in the Higgs boson rest frame) and the z-axis direction,
- the polar angles θ_1 and θ_2 between the final state leptons (i.e. $l_1\nu_1$ and $l_2\nu_2$ pairs) decaying back-to-back in the W rest frames and the direction of the parent W bosons (measured in the Higgs boson rest frame),
- the Φ angle between the two W decay planes defined in terms of lepton momenta in the Higgs boson rest frame,
- the angle Φ_1 between one of the W decay and production planes defined in terms of lepton momenta in the Higgs boson rest frame. The angle Φ_2 corresponding to the another W boson is redundant thanks to the relation $\Phi = \Phi_1 + \Phi_2$.

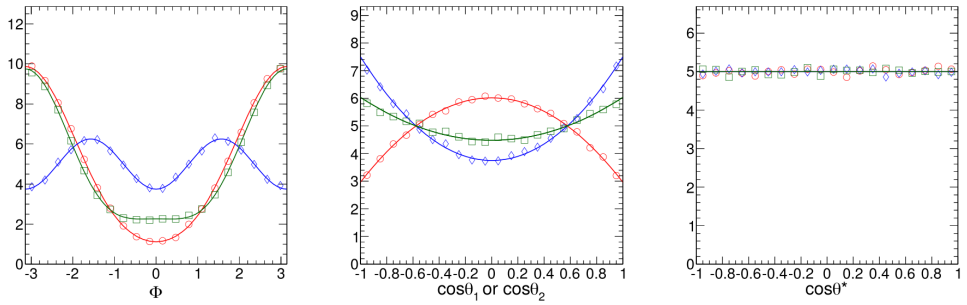

Figure 4.6: Distributions of typical generator-level observables describing Higgs boson production and the $H \to WW^*$ decay mode. Displayed are the Φ angle between two W decay planes (left), the cosine of the polar θ_1/θ_2 angles represented, due to the similarity, by a single figure (middle), and the cosine of the polar θ^* angle (right), which appears randomly distributed in case of the spin-0 scenario. Compared are signal events assuming a pure BSM contribution of the CP-odd a_3 coupling (blue diamond), a CP-even a_2 coupling (green square), and assuming a pure SM contribution (red circle). Source [94].

Note that the azimuth angle ϕ^* (not shown in Figure 4.6) is always trivial and can be easily integrated out in the scattering amplitude calculation. Another redundancy appears by assuming only spin-0 amplitude terms, in which case, the angles θ^* and Φ_1 are randomly distributed, making their contribution also trivial. The remaining parameters are summarily dubbed as $\vec{\Omega} = (q^2_{\mathrm{V}1}, q^2_{\mathrm{V}2}, \theta_1, \theta_2, \Phi)$ where we have explicitly denoted the q^2-dependence of the scattering amplitude rather than plain four-vector information. Let us stress that one of the q^2 distributions should correspond to the on-shell W boson (peaked around ≈ 80 GeV value) while the other to the off-shell W boson.

The discussion in the following subsection is focused on the application of generator-level ME computed by MELA while accounting for $\vec{\Omega}$ parameter values. Ideally, the ME technique could also be employed at the reconstruction-level to define discriminants capable of distinguishing between different topologies and using them in the process of signal extraction. Nevertheless, the presence of neutrinos in the final state prevents us from the full kinematic reconstruction of the event, calling for a more refined strategy as further discussed in Chapter 5.

4.4.2 Signal templates

The generator-level ME are used to determine transition weights between samples of different signal hypotheses, possibly including those not simulated by MC generators (such as $a^{Z\gamma}_{\Lambda 1}$ model). As an example, the transition weight defined by the formula

$$w_{\mathrm{H}_1(a_2)\to\mathrm{H}_1(a_3)} = w^{\mathrm{xs}}_{\mathrm{H}_1(a_2)\to\mathrm{H}_1(a_3)} \frac{\mathrm{ME}_{\mathrm{H}_1(a_3)}(\vec{\Omega})}{\mathrm{ME}_{\mathrm{H}_1(a_2)}(\vec{\Omega})} \tag{4.8}$$

is used to re-weight generated CP-even signal samples, i.e. the hypothesis $\mathrm{H}(a_1, a_2) = \mathrm{H}^{(a_2)}(0,1) \equiv \mathrm{H}_1(a_2)$, into the signal samples assuming a pure CP-odd coupling contribution, i.e. the hypothesis $\mathrm{H}(a_1, a_3) = \mathrm{H}^{(a_3)}(0,1) \equiv \mathrm{H}_1(a_3)$. The ratio between ME corresponding to the initial and target hypotheses is corrected by the factor $w^{\mathrm{xs}}_{\mathrm{H}_1(a_2)\to\mathrm{H}_1(a_3)}$ to account for proper cross-section normalisation in the target sample, as would be expected if the sample was generated in MC simulation. In practice, up to 7 originally

simulated MC samples can be re-weighted to any of the considered hypotheses (1× SM sample, 3× pure BSM sample and 3× SM-BSM mixed sample). Note that weights in formula 4.8 might only be used to re-weight samples corresponding to the same type of production topology (i.e. VBF signals to another VBF hypothesis, etc.). Eventually, the target signal sample corresponding to the particular hypothesis, e.g. $\mathrm{H}_1(a_3)$ from the previous simplified example, is obtained as a weighted average of all (seven) re-weighted MC samples.

The re-weighting procedure described in the previous paragraph might not only be used to increase the statistical precision of the MC prediction for existing simulated samples, but also to target sample hypotheses assuming any desired mixture of coupling values. In practice, a finite set of the signal hypotheses H is needed to fully parametrise the kinematics of the signal events, including the SM-BSM interference effect. Henceforth, we will refer to any kinematic distribution reconstructed using re-weighted signal samples as a "template" T. Particular observables used to represent kinematics for different hypotheses are picked based on their capability to discriminate signal, which will be further discussed in Chapter 5. As noted in Section 1.5.4, AC and SMEFT interpretation schemes of the underlying BSM physics also received different approach in treating the simultaneous contribution of multiple anomalous couplings at the same time. In that regard, the construction of the templates will be discussed separately for both approaches.

Templates in the AC scheme using Approach 1

In the AC interpretation scheme, we choose to consider one anomalous coupling a_i with a non-zero value at a time, also denoted as Approach 1. That way, we target the following individual sets of signal hypotheses: $\mathrm{H}(a_1, a_2)$, $\mathrm{H}(a_1, a_3)$, $\mathrm{H}(a_1, a_{\Lambda 1})$ and $\mathrm{H}(a_1, a_{\Lambda 1}^{Z\gamma})$, where we implicitly assume different values of a_1 and a_i in each of the sets. To illustrate how signal templates are created, we will assume VBF production topology as it offers the most general case with two HVV vertices affected by anomalous contribution – production ("prod") and decay ("dec") vertex. For the brevity of this text, we will demonstrate the construction of templates for the $\mathrm{H}(a_1, a_2)$ case and comment on any irregularities later on.

Suppose the simplified version of the scattering amplitude, $\mathcal{A}(\mathrm{HVV})$, in the form

$$\mathcal{A}(\mathrm{HVV}) = a_1 A_1 + a_2 A_2. \tag{4.9}$$

In the case of VBF production topology and decay, also denoted as electroweak (EWK) signal model, two vertices enter the amplitude formula as independent multiplicative factors, i.e.

$$\mathcal{A}(2\mathrm{HVV}) = (a_1 A_1^{\mathrm{prod}} + a_2 A_2^{\mathrm{prod}}) * (a_1 A_1^{\mathrm{dec}} + a_2 A_2^{\mathrm{dec}}). \tag{4.10}$$

By squaring this amplitude, we obtain the probability distribution for the EWK signal model that reads:

$$\mathcal{P}_{\mathrm{EWK}} = \mathcal{A}^2(2\mathrm{HVV}) = a_1^4 \mathrm{T}_1 + a_1^3 a_2 \mathrm{T}_2 + a_1^2 a_2^2 \mathrm{T}_3 + a_1 a_2^3 \mathrm{T}_4 + a_2^4 \mathrm{T}_5, \tag{4.11}$$

where $\mathrm{T}_1 = (A_1^{\mathrm{prod}} * A_1^{\mathrm{dec}})^2$ is the pure SM term, $\mathrm{T}_5 = (A_2^{\mathrm{prod}} * A_2^{\mathrm{dec}})^2$ is the pure BSM contribution for the a_2 coupling and templates $\mathrm{T}_2 - \mathrm{T}_4$ correspond to the SM-BSM interference (terms mixing A_1 and A_2). To implement this EWK model (such as it can be used in signal fit extraction), VBF templates $\mathrm{T}_1 - \mathrm{T}_5$ need to be constructed.

Note that any signal hypothesis can be parametrised by a function of the form prescribed by the probability relation 4.11. In the VBF production case (and the example

with the a_2 coupling) we can write

$$\mathrm{H}(a_1, a_2) = \mathcal{P}_{\mathrm{EWK}}(a_1, a_2). \tag{4.12}$$

Following the previous discussion, for $N = 5$ templates T needed in the EWK model (right-hand side in formula 4.12), exactly the same number N of alternative hypotheses (left-hand side of formula 4.12) is needed. In this analysis, we have chosen arbitrary set of hypotheses, $\mathrm{H}_1 - \mathrm{H}_5$, assuming the prescription

$$(a_1, a_2) = \{(1, 0); (0, 1); (1, 0.25); (1, 0.5); (1, 0.75)\}, \tag{4.13}$$

where each of them was obtained by means of MELA re-weighting. In summary, one can recast the relation between an arbitrary hypothesis H and signal templates T using a 5×5 matrix G, explicitly reading:

$$\begin{pmatrix} \mathrm{H}_1 \\ \mathrm{H}_2 \\ \mathrm{H}_3 \\ \mathrm{H}_4 \\ \mathrm{H}_5 \end{pmatrix} = \begin{pmatrix} 1^4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1^4 \\ 1^4 & 0.25 & 0.25^2 & 0.25^3 & 0.25^4 \\ 1^4 & 0.5 & 0.5^2 & 0.5^3 & 0.5^4 \\ 1^4 & 0.75 & 0.75^2 & 0.75^3 & 0.75^4 \end{pmatrix} \begin{pmatrix} \mathrm{T}_1 \\ \mathrm{T}_2 \\ \mathrm{T}_3 \\ \mathrm{T}_4 \\ \mathrm{T}_5 \end{pmatrix}. \tag{4.14}$$

Eventually, finding the particular template distribution is narrowed down to calculating an inverse of matrix G, as can be deduced from the inversion of Equation 4.14, i.e. $\vec{T} = G^{-1}\vec{H}$. The VH production topology is similar to that of the VBF process, therefore we employ the same EWK model.

As an example, kinematic distributions corresponding to 5 different WH hypotheses displayed in Figure 4.7 are transformed into 5 templates shown in Figure 4.9. The analysis phase space used in this example is known as the "ggF + 1-jet" category, i.e. selection cuts applied to define this signal region (see full overview in Chapter 5) are chosen to mimic ggF the production topology with exactly one associated hadronic jet in the final state. It is assumed that SM templates constructed from WH, ZH and VBF production MC samples will not contribute significantly in this region, however, we cannot fully neglect it. Moreover, the predicted contribution from pure BSM templates and/or SM-BSM interference might also play a crucial role. For some of the cases where original MC simulation was available, e.g. pure SM/BSM hypotheses, we were able to compare validity of analytically-constructed templates. Reasonable agreement between analytical and simulated distributions was observed, especially for production samples statistically dominating in selected signal regions. More studies concerning the validity of signal templates were performed and will be discussed in Section 4.4.4.

In the case of ggF production topology, only the decay vertex contributes to the HVV amplitude, leading to the simplified probability model:

$$\mathcal{P}_{\mathrm{ggF}} = \mathcal{A}^2(\mathrm{HVV}) = a_1^2\mathrm{T}_1 + a_1 a_2 \mathrm{T}_2 + a_2^2 \mathrm{T}_3, \tag{4.15}$$

where $\mathrm{T}_1 = (A_1^{\mathrm{dec}})^2$ is the pure SM term, $\mathrm{T}_3 = (A_2^{\mathrm{dec}})^2$ is the pure BSM term and $\mathrm{T}_2 \approx A_1^{\mathrm{dec}} A_2^{\mathrm{dec}}$ is representing the interference of both. Hypotheses $\mathrm{H}_1 - \mathrm{H}_3$ were chosen to describe the full SM case, full BSM case, and an equal mixture of both:

$$(a_1, a_2) = \{(1, 0); (0, 1); (1, 0.27)\}. \tag{4.16}$$

Note that the $a_2 = 0.27$ value in the mixture scenario can be derived from Equation 1.110 by assuming $f_{a_2} = 0.5$. Similar to the EWK model, kinematic ggF templates $\mathrm{T}_1 - \mathrm{T}_3$ are

derived by inverting corresponding 3×3 matrix G. An example of kinematic distributions for 3 different ggF hypotheses and 3 corresponding ggF signal templates is displayed in Figure 4.8 and Figure 4.10, respectively.

The template construction procedure, as described in the lines above, can also be employed for the a_3 and $a_{\Lambda1}$ couplings – see an additional example of ggF production templates in the ggF + 1-jet signal region in Figure 4.11 and Figure 4.12. On the other hand, the HZγ vertex and related $a_{\Lambda1}^{Z\gamma}$ coupling appear purely as new BSM physics by introducing higher-order loop corrections. As such, it is only VBF and ZH production vertices that might be affected by this anomalous effect. Unlike any other AC considered, there is no pure BSM hypothesis for the $a_{\Lambda1}^{Z\gamma}$ coupling that is allowed in the $H \to WW^*$ decay channel, as with $a_1 \to 0$, this decay is higly supressed. In other words, the scattering amplitude corresponding to the HZγ vertex can be written as:

$$\mathcal{A}(\mathrm{HZ}\gamma) = (a_1 A_1^{\mathrm{prod}} + a_{\Lambda1}^{Z\gamma} A_{Z\gamma}^{\mathrm{prod}}) * a_1 A_1^{\mathrm{dec}}. \tag{4.17}$$

Squaring this amplitude leads to the first three terms appearing in the EWK signal model shown in Equation 4.11: the pure SM term and two interference terms. Following the same template construction procedure, three alternative hypotheses, $\mathrm{H}_1 - \mathrm{H}_3$, were provided by the MELA re-weighting method, corresponding to the pure SM and two mixed coupling hypotheses:

$$(a_1, a_{\Lambda1}^{Z\gamma}) = \{(1,0); (1,0.25); (1,0.75)\}. \tag{4.18}$$

Finally, inverting the 3×3 G matrix in its relevant form gives three kinematic templates $\mathrm{T}_1 - \mathrm{T}_3$.

Templates in the SMEFT scheme using Approach 2

In the SMEFT interpretation scheme, we construct the signal model by assuming all SMEFT couplings at the same time, also denoted as Approach 2. In that regard, we can study SMEFT couplings both independently, with other couplings fixed to zero, and simultaneously with the rest of couplings profiled (floating) in the fit. An arbitrary SMEFT signal hypothesis, $\mathrm{H}(a_1, a_2, a_3, a_{\Lambda1})$, can be expressed by assuming a particular set of $(a_1, a_2, a_3, a_{\Lambda1})$ values. Note that thanks to the set of underlying relations in Equation 1.108, the coupling $a_{\Lambda1}^{Z\gamma}$ is not independent and can be omitted from the signal parameterisation. It is important to emphasize, that for the remaining three SMEFT couplings, a_2, a_3 and $a_{\Lambda1}$, we use the same notation as in the AC scheme and Approach 1. However, because of the same set of the relations in Equation 1.108, these couplings should be considered fundamentally different between the two interpretation schemes.

The template construction procedure in Approach 2 is similar to the one employed in Approach 1 with an obvious increase in complexity due to the presence of multiple couplings at the same time. In particular, suppose the simplified version of the scattering amplitude, $\mathcal{A}(\mathrm{HVV})$, in the following form:

$$\mathcal{A}(\mathrm{HVV}) = a_1 A_1 + a_2 A_2 + a_3 A_3 + a_{\Lambda1} A_{\Lambda1}. \tag{4.19}$$

In the case of EWK signal model representing either VBF or VH production processes with both production and decay vertices, we can express it as follows:

$$\begin{aligned}\mathcal{A}(\mathrm{2HVV}) = &(a_1 A_1^{\mathrm{prod}} + a_2 A_2^{\mathrm{prod}} + a_3 A_3^{\mathrm{prod}} + a_{\Lambda1} A_{\Lambda1}^{\mathrm{prod}}) \\ &* (a_1 A_1^{\mathrm{dec}} + a_2 A_2^{\mathrm{dec}} + a_3 A_3^{\mathrm{dec}} + a_{\Lambda1} A_{\Lambda1}^{\mathrm{dec}}).\end{aligned} \tag{4.20}$$

Squaring this amplitude leads to 35 terms:

$$\begin{aligned}\mathcal{P}_{\text{EWK}} = \mathcal{A}^2(\text{2HVV}) = {} & a_1^4\text{T}_1 + a_1^3a_2\text{T}_2 + a_1^2a_2^2\text{T}_3 + a_1a_2^3\text{T}_4 + a_2^4\text{T}_5 \\ & + a_1^3a_3\text{T}_6 + a_1^2a_3^2\text{T}_7 + a_1a_3^3\text{T}_8 + a_3^4\text{T}_9 \\ & + a_1^3a_{\Lambda1}\text{T}_{10} + a_1^2a_{\Lambda1}^2\text{T}_{11} + a_1a_{\Lambda1}^3\text{T}_{12} + a_{\Lambda1}^4\text{T}_{13} \\ & + a_2^3a_3\text{T}_{14} + a_2^2a_3^2\text{T}_{15} + a_2a_3^3\text{T}_{16} \\ & + a_2^3a_{\Lambda1}\text{T}_{17} + a_2^2a_{\Lambda1}^2\text{T}_{18} + a_2a_{\Lambda1}^3\text{T}_{19} \\ & + a_3^3a_{\Lambda1}\text{T}_{20} + a_3^2a_{\Lambda1}^2\text{T}_{21} + a_3a_{\Lambda1}^3\text{T}_{22} \\ & + a_1a_2a_3^2\text{T}_{23} + a_1a_2^2a_3\text{T}_{24} + a_1^2a_2a_3\text{T}_{25} \\ & + a_1a_2a_{\Lambda1}^2\text{T}_{26} + a_1a_2^2a_{\Lambda1}\text{T}_{27} + a_1^2a_2a_{\Lambda1}\text{T}_{28} \\ & + a_1a_3a_{\Lambda1}^2\text{T}_{29} + a_1a_3^2a_{\Lambda1}\text{T}_{30} + a_1^2a_3a_{\Lambda1}\text{T}_{31} \\ & + a_2a_3a_{\Lambda1}^2\text{T}_{32} + a_2a_3^2a_{\Lambda1}\text{T}_{33} + a_2^2a_3a_{\Lambda1}\text{T}_{34} \\ & + a_1a_2a_3a_{\Lambda1}\text{T}_{35},\end{aligned} \tag{4.21}$$

where T_1 is the pure SM term, while T_5, T_9 and T_{13} are the pure BSM terms corresponding to a_2, a_3 and $a_{\Lambda1}$ couplings, respectively. The rest of the terms represents various interference contributions.

Summarily, an equivalent of Equation 4.14 is constructed with the matrix G now formed of 35×35 elements, describing relations between SMEFT signal templates, $\text{T}_1 - \text{T}_{35}$, and arbitrary SMEFT hypotheses. Following the familiar procedure, 35 different alternative hypotheses, $\text{H}_1 - \text{H}_{35}$, were constructed using the MELA re-weighting method, assuming coupling values corresponding to the pure SM, pure BSM and multiple mixed coupling scenarios. The EWK signal templates are then extracted using the inverted G matrix. A selection of several SMEFT templates, assuming WH production and decay model, is displayed in Figure 4.14.

In the case of ggF production model, we perform similar calculations with a single (decay) vertex. Therefore we arrive at the probability relation consisting of 10 terms, i.e.

$$\begin{aligned}\mathcal{P}_{\text{ggF}} = \mathcal{A}^2(\text{HVV}) = {} & a_1^2\text{T}_1 + a_1a_2\text{T}_2 + a_2^2\text{T}_3 \\ & + a_1a_3\text{T}_4 + a_3^2\text{T}_5 \\ & + a_1a_{\Lambda1}\text{T}_6 + a_{\Lambda1}^2\text{T}_7 \\ & + a_2a_3\text{T}_8 + a_2a_{\Lambda1}\text{T}_9 + a_3a_{\Lambda1}\text{T}_{10},\end{aligned} \tag{4.22}$$

where T_1 is the pure SM term, while T_3, T_5 and T_7 are the pure BSM terms corresponding to a_2, a_3 and $a_{\Lambda1}$ couplings, respectively. As usual, the rest of the terms describe various interference contributions. The ggF signal templates, $\text{T}_1 - \text{T}_{10}$, are extracted using the inverted 10×10 G matrix and a vector formed of 10 alternative hypotheses, $\text{H}_1 - \text{H}_{10}$, corresponding to the pure SM, pure BSM and multiple mixed coupling scenarios. An example of several SMEFT templates assuming the HVV decay vertex in the ggF production model is shown in Figure 4.13.

Finally, note that we could choose to apply Approach 2 also in the case of the AC scheme. Nevertheless, simultaneous treatment of 5 independent couplings (1× SM and 4× BSM) leads to an unprecedented number of templates, which would extremely complicate the convergence of the final fit. Therefore this option is not considered, and Approach1 (Approach 2) remains inherently related to the AC scheme (SMEFT scheme) throughout this text.

4.4.3 Signal extraction model

As noted at the end of Section 1.5.4, it is convenient to parameterize probability distributions $\mathcal{P}_{\text{ggF}}$ and $\mathcal{P}_{\text{EWK}}$ by an effective fractional cross-section f_{a_i} (see Equation 1.110). In practice, this can be performed by replacing each coupling a_i in the formulae 4.11 and 4.15 (AC scheme) as well as formulae 4.21 and 4.22 (SMEFT scheme) by the corresponding $g_i * f_{a_i}$ term, where $g_i = \sqrt{\sigma_1/\sigma_i}$ is derived from values in Table 1.1, and by introducing ggF and EWK signal strength modifiers μ_{ggF} and μ_{EWK}.

The ggF and EWK signal models in the case of the AC scheme with Approach 1 can now be written as

$$\begin{aligned}\mathcal{P}_{\text{ggF}} &= \mu_{\text{ggF}}(f_{a_1}\text{T}_1 + g_i\sqrt{f_{a_1}}\sqrt{f_{a_i}}\text{T}_2 + g_i^2 f_{a_i}\text{T}_3),\\ \mathcal{P}_{\text{EWK}} &= \mu^2_{\text{EWK}}(f^2_{a_1}\text{T}_1 + g_i\sqrt{f_{a_1}}^3\sqrt{f_{a_i}}\text{T}_2 + g_i^2 f_{a_1}f_{a_i}\text{T}_3 + g_i^3\sqrt{f_{a_1}}\sqrt{f_{a_i}}^3\text{T}_4 + g_i^4 f^2_{a_i}\text{T}_5),\end{aligned} \tag{4.23}$$

where the SM fractional cross-section f_{a_1} is found as $f_{a_1} = 1 - f_{a_i}$, with a_i standing for one of four AC considered: a_2, a_3, $a_{\Lambda 1}$ or $a^{Z\gamma}_{\Lambda 1}$. Note that by setting $\mu_{\text{ggF/EWK}} = (a_1^2\sigma_1 + a_i^2\sigma_i)/\sigma_1$ and writing out g and f_{a_i} terms, it is possible to recover the original probability relations featuring direct parameterisation by anomalous couplings.

In a similar manner, the ggF and EWK signal models can be derived while accounting for the SMEFT scheme with Approach 2. In particular,

$$\begin{aligned}\mathcal{P}_{\text{ggF}} &= \mu_{\text{ggF}}(f_{a_1}\text{T}_1 ... + g_2^2 f_{a_2}\text{T}_3 ... + g^2_{\Lambda 1} f_{a_{\Lambda 1}}\text{T}_7 ... + g_3 g_{\Lambda 1}\sqrt{f_{a_3}}\sqrt{f_{a_{\Lambda 1}}}\text{T}_{10}),\\ \mathcal{P}_{\text{EWK}} &= \mu^2_{\text{EWK}}(f^2_{a_1}\text{T}_1 ... + g_2^4 f^2_{a_2}\text{T}_5 ... + g_3^3 g_{\Lambda 1}\sqrt{f_{a_3}}^3\sqrt{f_{a_{\Lambda 1}}}\text{T}_{20} ...\\ &\quad + g_2 g_3 g_{\Lambda 1}\sqrt{f_{a_1}}\sqrt{f_{a_2}}\sqrt{f_{a_3}}\sqrt{f_{a_{\Lambda 1}}}\text{T}_{35}),\end{aligned} \tag{4.24}$$

where the SM fractional cross-section f_{a_1} is restricted by the physical threshold expressed by the formula $f_{a_1} = 1 - f_{a_2} - f_{a_3} - f_{a_{\Lambda 1}}$. In both approaches, the sign of the couplings is implicitly considered, thus terms proportional to $\sqrt{f_{a_i}}^x$ for $x = 1$ or $x = 3$ are scaled by $\text{sign}(a_i/a_1)$.

Let us stress that both ggF and EWK signal models are considered simultaneously in the final fit to treat ggF and VBF/VH production signals, respectively. An exception occurs in the case of the AC scheme with the $a^{Z\gamma}_{\Lambda 1}$ coupling, where only the first three terms in $\mathcal{P}_{\text{EWK}}$ are required. Also, note that only one signal strength modifier μ_{EWK} is needed to describe the normalisation of both VBF and VH production signals. This could be done because the signal yields for the VBF and VH processes are related, assuming that the same couplings entering the HVV production vertex also enter the HVV decay vertex. This is not true in the case of ggF production, which is predominantly initiated via coupling with the top quark loop, thus, the corresponding μ_{ggF} signal strength is treated individually. Both μ_{ggF} and μ_{EWK} are considered free parameters of the fit (i.e. they are always profiled). As a consequence, the overall signal event yield is not used to discriminate between alternative signal hypotheses. Constraints on f_{a_i} values in both AC and SMEFT schemes are obtained via Maximum Likelihood method, the discussion of which is left for the dedicated Section 6.2.1.

4.4.4 Signal validation

Some of the signal templates entering the relations for $\mathcal{P}_{\text{ggF}}$ and $\mathcal{P}_{\text{VBF}}$ appear to have negative normalisation (T_2 template in Figure 4.9) or fluctuate around zero value (T_2 template in Figure 4.11). These effects reflect the quasi-probabilistic nature of the quantum-mechanical interference, which can a priori yield negative values (former case) or manifest

as fully destructive (latter case). In fact, vanishing interference is typical, but is not exclusive to cases where the interplay between CP-odd and CP-even operators is dominant. Templates showing this behavior can be safely set to zero as it was tested they do not have an effect on constraining anomalous couplings.

Templates that appear with negative normalisation are flipped to positive values, while a minus sign is introduced at the corresponding place in formulae 4.23 and 4.24. Since individual templates can be negative, the overall signal model distributions $\mathcal{P}_{\text{ggF}}$ and $\mathcal{P}_{\text{EWK}}$ represent physical probabilities, and their normalisations must be positive. This fact was checked separately for all signal regions entering the analysis phase space and all production processes with a non-negligible contribution. The performed test could be described in the following steps:

1. Signal templates are constructed in MC simulation according to the usual procedure described above per each considered production process.

2. For a chosen value $f_{a_i} = y$, factor g_i and template normalisation sign ± 1, we define a scaling factor $N_x(y, g_i, \pm 1)$ standing in front of template T_x according to formulae $\mathcal{P}_{\text{ggF}}$ or $\mathcal{P}_{\text{VBF}}$.

3. Signal strength modifiers are set to $\mu_{\text{ggF}} = \mu_{\text{EWK}} = 1$.

4. Expected number of events is defined as an integral (yield) Y of the final distribution constructed as a weighted sum of templates, $\sum_x N_x T_x$.

5. Steps $2 - 4$ are repeated for several values $y \in [-1, 1]$ and the corresponding set of integral yields is plotted.

An example of the normalisation test in the ggF phase space with 0/1 associated jets, performed for a spectrum of possible f_{a_2}, f_{a_3} and $f_{a_{\Lambda 1}}$ values, is shown separately for ggF, WH and ZH production processes in Figure 4.15. The VBF process is not shown as in this region it is significantly suppressed across the f_{a_i} spectra (to the extent it could essentially be neglected). Similar tests were performed in all remaining signal regions concluding on a positivity of the probability normalisation.

In summary, the final set of the signal templates for both the AC and SMEFT schemes, as well as the analytical formulae for $\mathcal{P}_{\text{ggF}}$ and $\mathcal{P}_{\text{VBF}}$ describing our signal model, are the necessary ingredients for building the statistical method of extracting constraints on anomalous couplings. As it will be shown in the next chapter, the particular choice of observables used to define individual template distributions is a strategic step in stipulating our sensitivity to alternative signal. As such, we will return to the description of template shapes, as these can help us in understanding the final fit results.

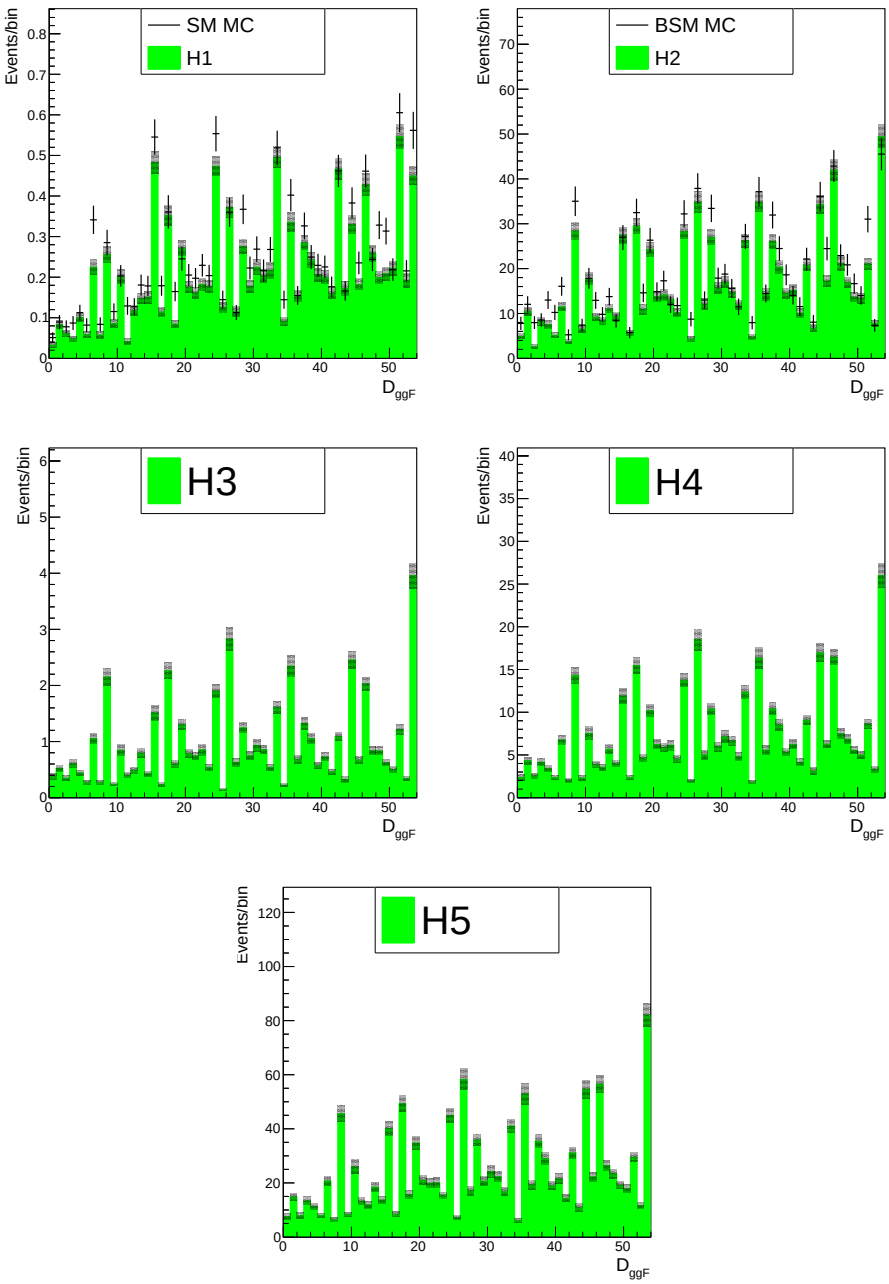

Figure 4.7: The WH signal hypotheses $H_1 - H_5$ (green histogram) considered for the a_1 and a_2 HVV couplings in the ggF + 1-jet signal region represented by distributions of D_{ggF} defined for signal extraction in the ggF categories. Comparison with the original SM and BSM MC distributions (black points) is displayed in the case of the pure SM hypothesis (H_1) and the pure BSM hypothesis (H_2). The shaded area and vertical bars correspond to the MC statistical error on the number of events.

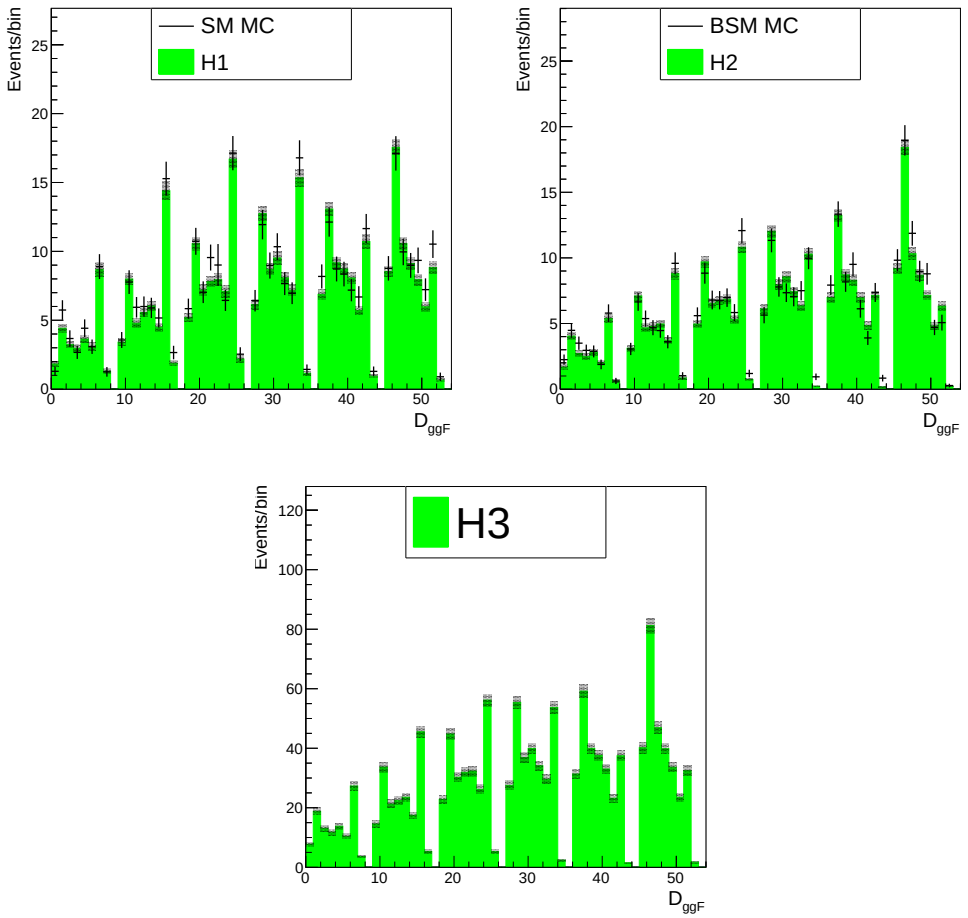

Figure 4.8: The ggF signal hypotheses $H_1 - H_3$ (green histogram) considered for the a_1 and a_2 HVV couplings in the ggF + 1-jet signal region represented by distributions of D_{ggF} defined for signal extraction in the ggF categories. Comparison with the original SM and BSM MC distributions (black points) is displayed in the case of the pure SM hypothesis (H_1) and the pure BSM hypothesis (H_2). The shaded area and vertical bars correspond to the MC statistical error on the number of events.

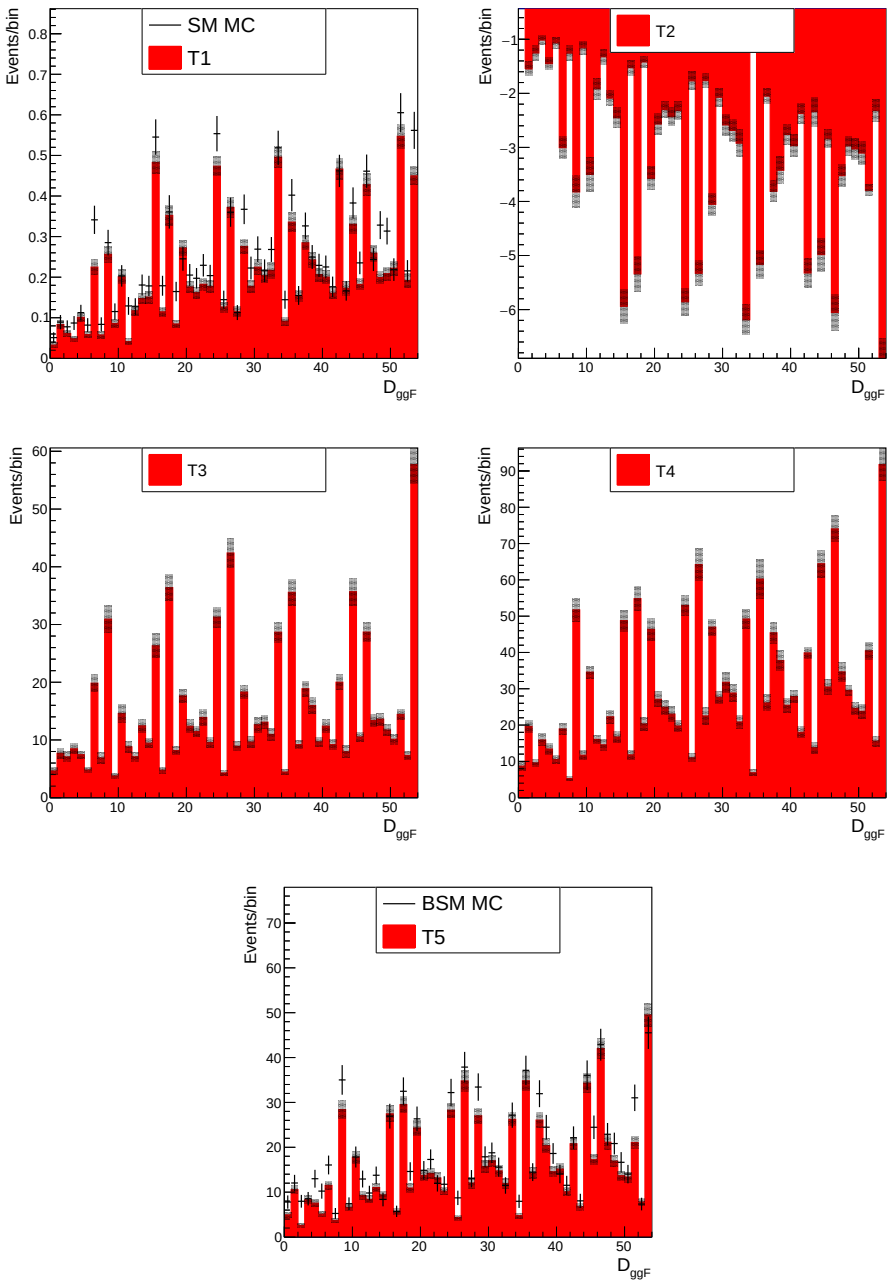

Figure 4.9: The WH signal templates $T_1 - T_5$ (red histogram) considered for the a_1 and a_2 HVV couplings in the ggF + 1-jet signal region represented by distributions of D_{ggF} defined for signal extraction in the ggF categories. Comparison with the original SM and BSM MC distributions (black points) is displayed in the case of the pure SM template (T_1) and the pure BSM template (T_5). The shaded area and vertical bars correspond to the MC statistical error on the number of events.

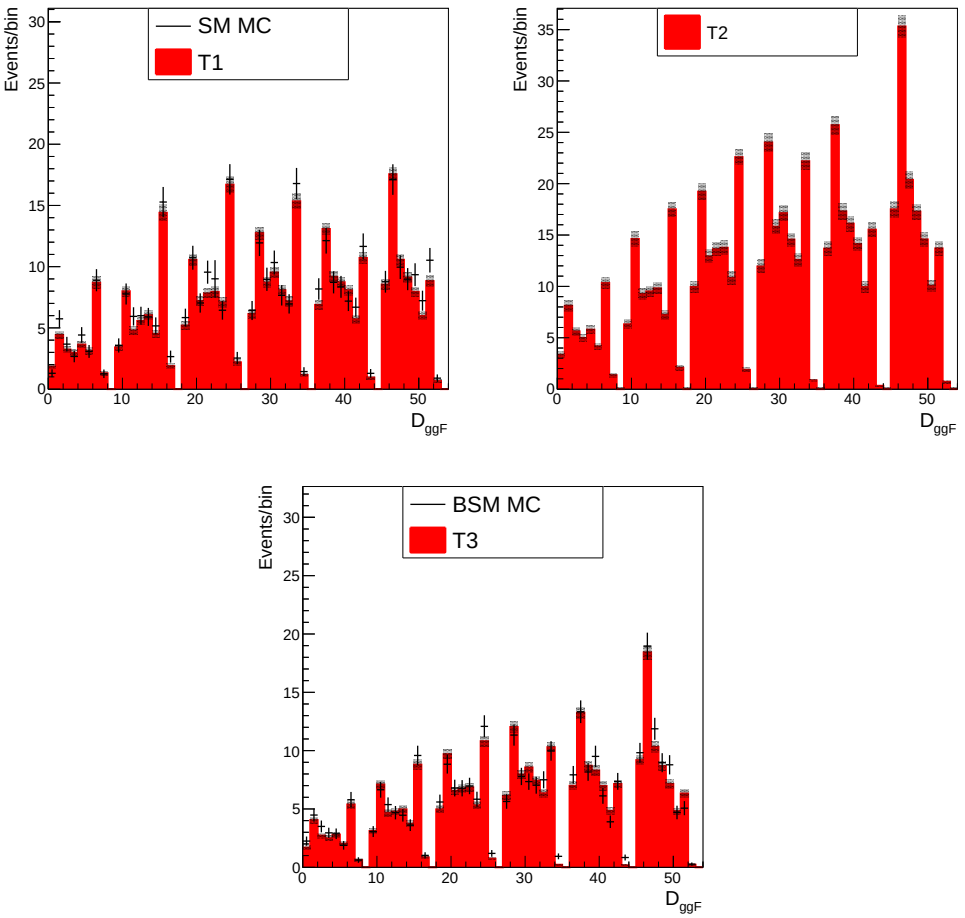

Figure 4.10: The ggF signal templates $T_1 - T_3$ (red histogram) considered for the a_1 and a_2 HVV couplings in theggF + 1-jet signal region represented by distributions of D_{ggF} defined for signal extraction in ggF categories. Comparison with the original SM and BSM MC distributions (black points) is displayed in the case of the pure SM template (T_1) and the pure BSM template (T_3). The shaded area and vertical bars correspond to the MC statistical error on the number of events.

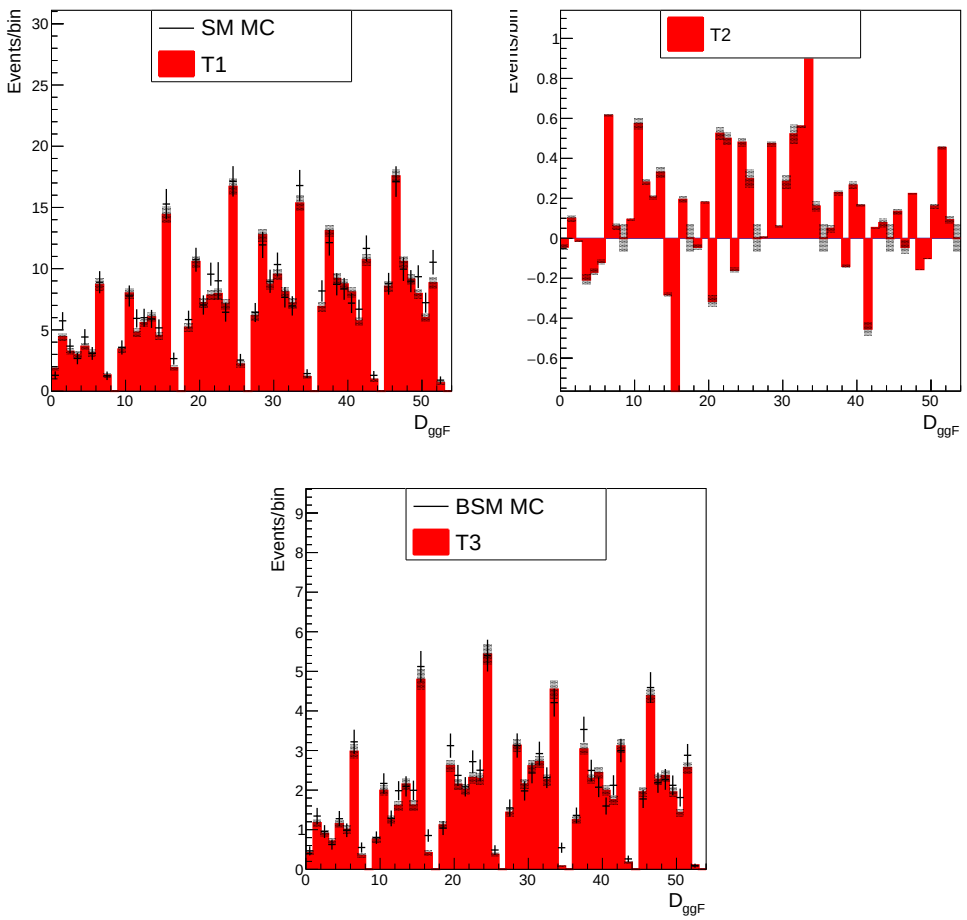

Figure 4.11: The ggF signal templates $T_1 - T_3$ (red histogram) considered for the a_1 and a_3 HVV couplings in the ggF + 1-jet signal region represented by distributions of D_{ggF} defined for signal extraction in ggF categories. Comparison with the original SM and BSM MC distributions (black points) is displayed in the case of the pure SM template (T_1) and the pure BSM template (T_3). The shaded area and vertical bars correspond to the MC statistical error on the number of events.

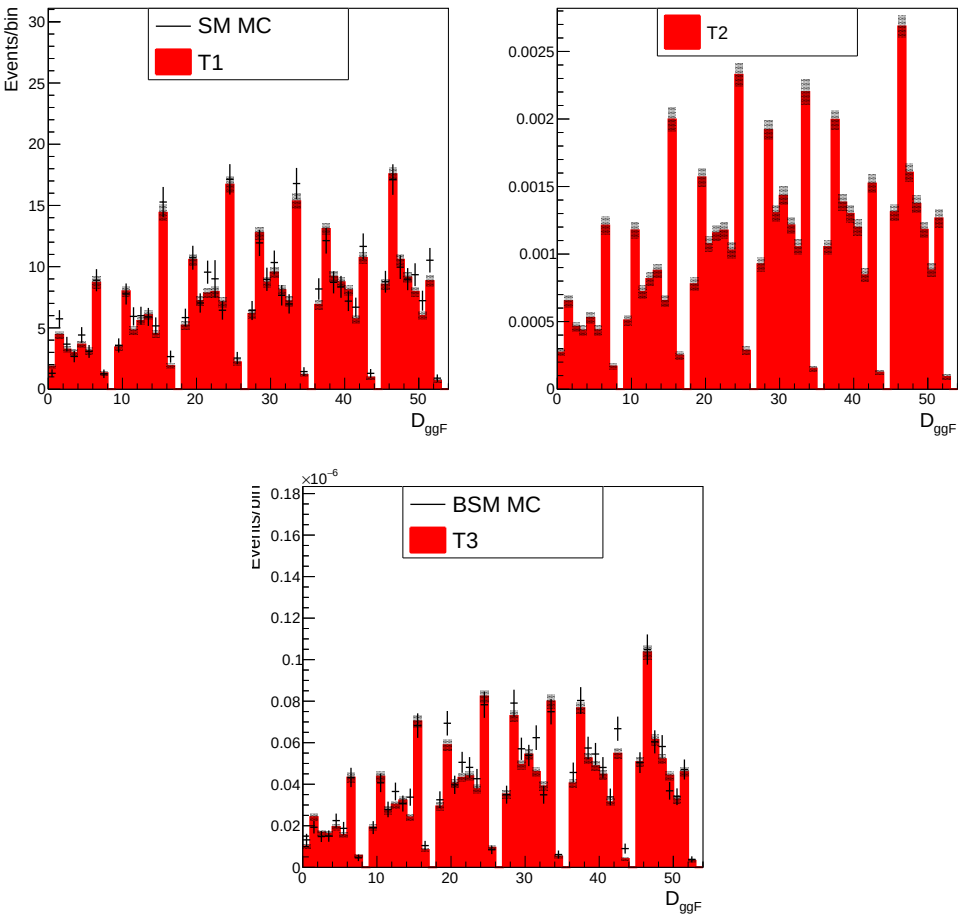

Figure 4.12: The ggF signal templates $\mathrm{T}_1 - \mathrm{T}_3$ (red histogram) considered for the a_1 and $a_{\Lambda 1}$ HVV couplings in the ggF + 1-jet signal region represented by distributions of D_{ggF} defined for signal extraction in ggF categories. Comparison with the original SM and BSM MC distributions (black points) is displayed in the case of the pure SM template (T_1) and the pure BSM template (T_3). The shaded area and vertical bars correspond to the MC statistical error on the number of events.

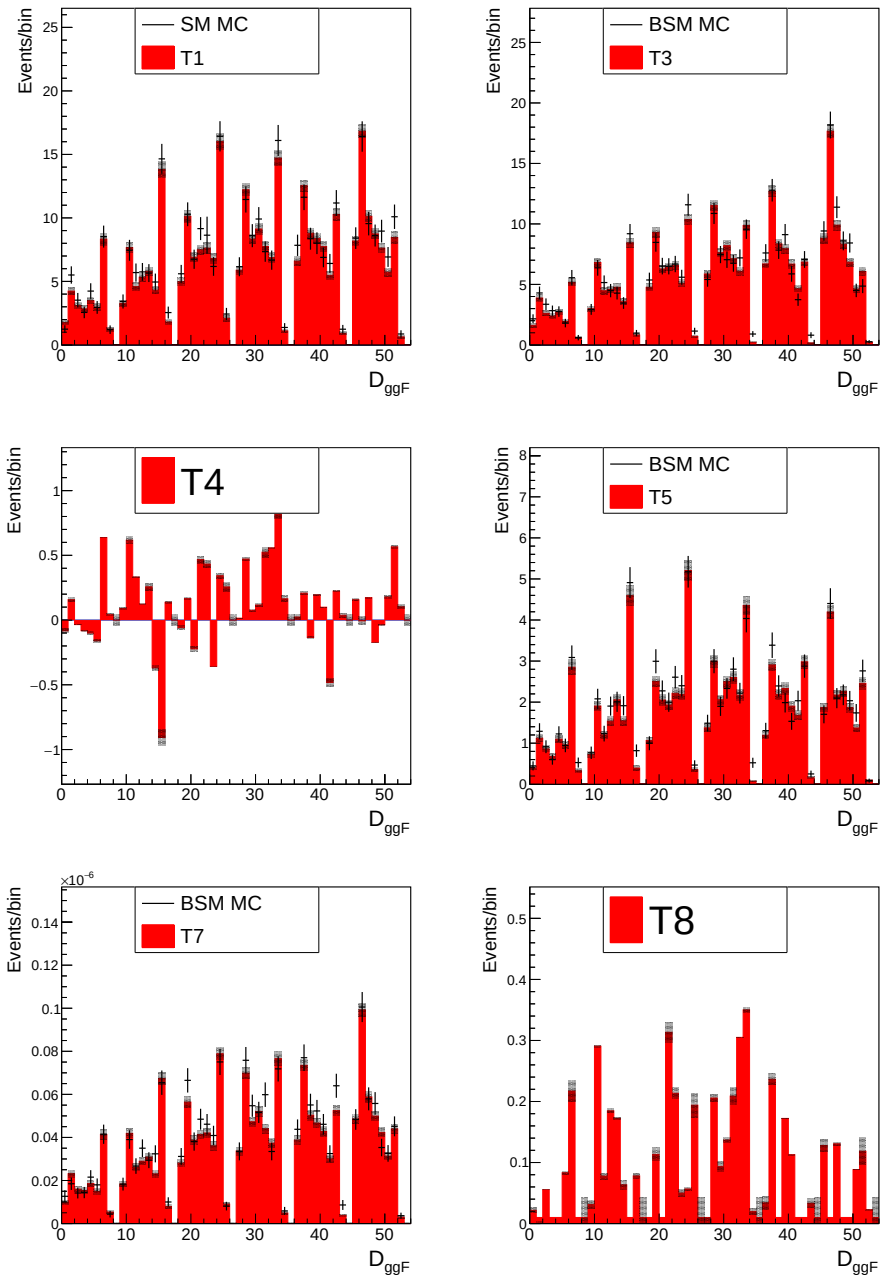

Figure 4.13: Selection of the SMEFT signal templates assuming the HVV decay vertex in ggF production is shown for the pure SM case (T_1), the pure BSM case corresponding to the a_2 coupling (T_3), the a_3 coupling (T_5), the $a_{\Lambda 1}$ coupling (T_7), and interference between the a_3 coupling and SM (T_4) as well as between the a_2 and a_3 couplings (T_8). All distributions are shown in the ggF + 1-jet signal region and correspond to the D_{ggF} discriminant defined for optimal signal extraction in ggF categories. Comparison with the original SM and BSM MC distributions (black points) is also displayed. The shaded area and vertical bars correspond to the MC statistical error on the number of events.

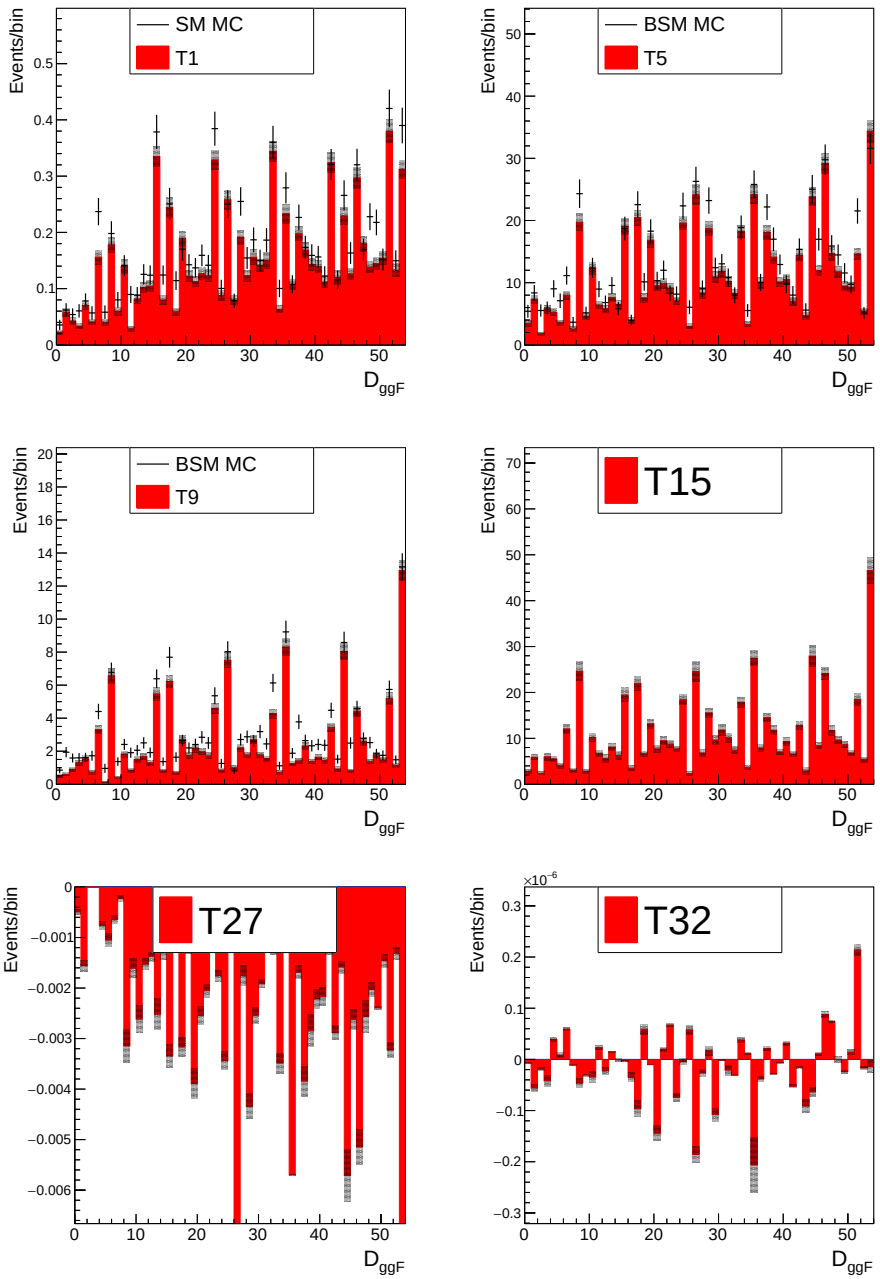

Figure 4.14: Selection of the SMEFT signal templates assuming WH production and decay HVV vertex is shown for the pure SM case (T_1), the pure BSM case corresponding to the a_2 coupling (T_5) or the a_3 coupling (T_9) and interference between several couplings (T_{15}, T_{27}, T_{32}). All distributions are shown in the ggF + 1-jet signal region and correspond to the D_{ggF} discriminant defined for optimal signal extraction in ggF categories. Comparison with the original SM and BSM MC distributions (black points) is also displayed. The shaded area and vertical bars correspond to the MC statistical error on the number of events.

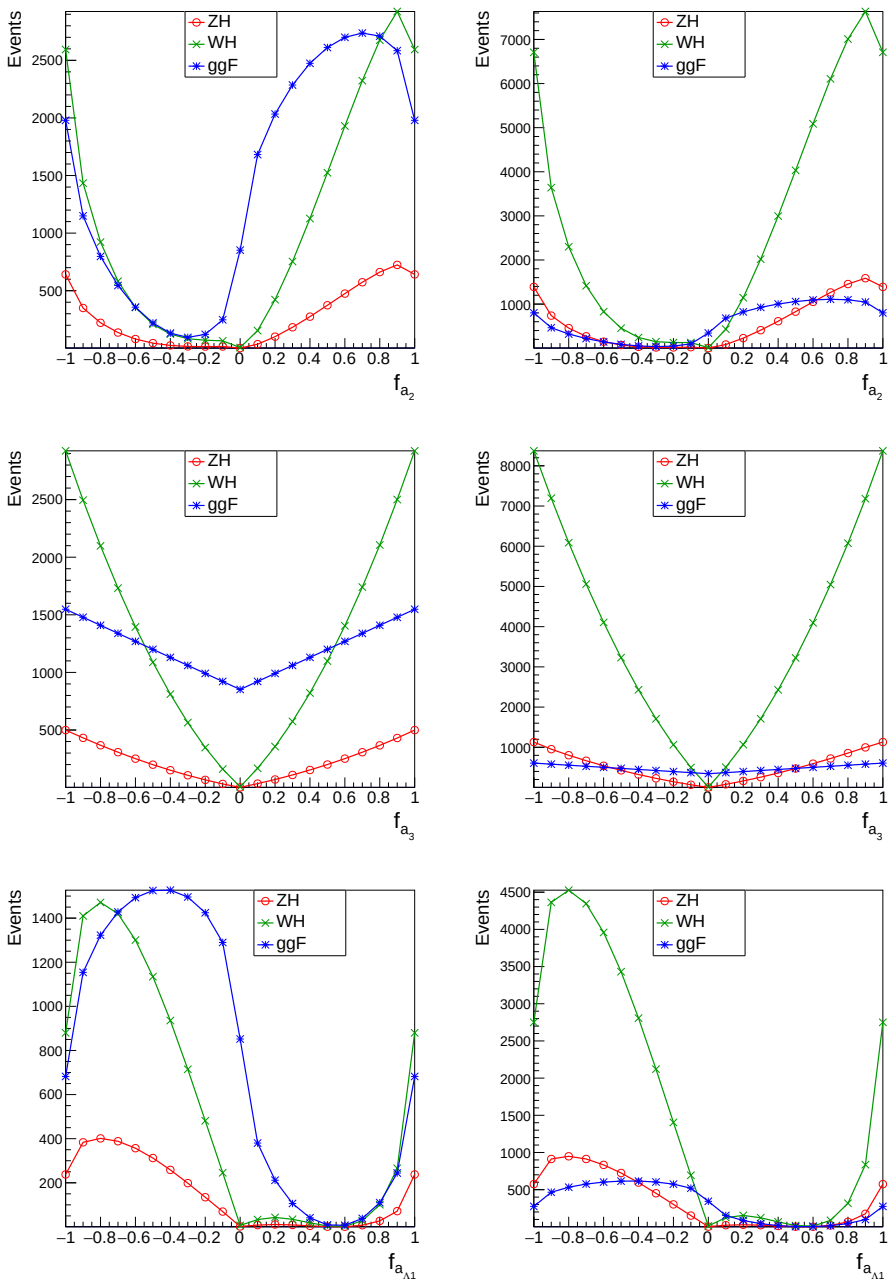

Figure 4.15: The expected number of signal events in the ggF + 0-jet (left) and ggF + 1-jet (right) channels for the spectrum of possible f_{a_2} (top), f_{a_3} (middle) and $f_{a_{\Lambda 1}}$ (bottom) values, assuming that the signal strength modifiers are equal to 1.

CHAPTER

FIVE

ANALYSIS STRATEGY

This analysis is performed in a kinematic phase space defined using reconstruction-level observables. Selected pp-collision events (both in MC simulation and data) are studied in five main Signal Regions (SRs) targeting Higgs boson production topology, further divided by the number of reconstructed hadronic jets in the final state. In particular, we define following signal regions:

- ggF + 0-jet,
- ggF + 1-jet,
- VBF + 2-jets,
- Resolved VH (i.e. with two resolved AK4 jets),
- Boosted VH (i.e. with one boosted AK8 V-jet),

where the categorisation criteria for these regions are mutually exclusive and will be discussed in Section 5.1. Selection requirements (often denoted as kinematic "cuts") are motivated by the preceding analysis measuring properties of the SM Higgs boson in the $H \to WW$ channel [130], henceforth denoted as "HWW Legacy". In HWW Legacy, the selected signal phase space was optimized for the extraction of the SM signal. Nevertheless, the anomalous nature of the couplings studied in this thesis implies that the same phase space is suitable for the extraction of an alternative signal. In that regard, any deviation from the HWW Legacy selection will be pointed out explicitly, otherwise, it should be understood as identical.

The employed analysis workflow follows the standard HEP procedure of an analysis in a "blind" regime. This means that observed distributions of events (i.e. actual data) in SRs were not examined until all details of this analysis were finalized and tested. This precaution is taken to prevent any possible bias caused by human factors. However, to assure the quality of physical object reconstruction and a basic level of understanding of the studied processes prior to "unblinding" (i.e. examining data), we are allowed to study observed distributions in dedicated Control Regions (CRs). The CRs are designed to target major background processes with minimal contamination by signal events. In particular, we employ selection criteria motivated by HWW Legacy analysis to target:

- Drell-Yan production (DY CR),

- Top quark production (Top CR),
- Nonresonant WW production (WW CR).

The CRs are especially useful to constrain relevant background normalisation in the final fit and will be further discussed in Section 5.1.3.

Assuming the same kinematic phase space to extract both SM and BSM signal contribution, it is imperative to define kinematic observables sensitive to subtle differences between alternative signal templates. Various Kinematic Discriminants (KDs) were studied and optimized for the best performance in each of the signal regions and will be discussed in Section 5.2. Because of my personal involvement, a more detailed description of the KDs will be provided in the case of 0-jet and 1-jet channels, while for the sake of completeness, a brief summary of the KDs in the 2-jet channels will be also presented.

5.1 Analysis phase space

All events entering the analysis SRs need to pass the base selection criteria targeting the $H \to WW^* \to e\mu$+MET decay topology. Events are further categorised according to the different Higgs production kinematics and jet activity in the final state. Orthogonality (i.e. exclusivity) of the $H \to WW^*$ phase space with other Higgs decay channels is also assured, targeting possible future combination of results. Alternative base selection and categorisation are used to define dedicated CRs, while the final state jet composition is preserved to mimic the jet activity in the SRs.

5.1.1 Base selection and orthogonality cuts

Candidate events entering the inclusive signal phase space need to have exactly one electron-muon ($e\mu$) pair in the final state. The study of the Anomalous Couplings (ACs) in the same-flavor topology (ee or $\mu\mu$) would not be currently possible because of the overwhelming background, predominantly from the DY same-flavor events. Because of this contamination, the sensitivity to measure anomalous effects would not be improved by including the analysis channels targeting the same-flavor topology. Therefore, the ee and $\mu\mu$ final states are not considered.

The sum of electric charges in the dilepton final state needs to be zero to ensure that the leptons originate in an electrically neutral Higgs boson decay. The charged lepton with the highest transverse momentum p_{T} is denoted as the "leading" lepton (index "lep1"), and the second in a row in terms of lepton p_{T} will be called the "sub-leading" lepton (index "lep2"). Events selected in this analysis are required to have the leading and sub-leading lepton above 25 GeV and 13 GeV threshold, respectively. The p_{T} threshold in the case of a sub-leading muon is lowered to 10 GeV for the 2016 dataset to take advantage of the lower threshold in the corresponding HLT path used during data-taking. Events containing additional leptons with $p_{\mathrm{T}} > 10$ GeV are not selected. The presence of neutrinos in the final state is addressed by $p_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV cut.

Complementary to studying the final state leptons one-by-one, it is practical to reconstruct the dilepton system and measure the magnitude of its transverse momentum p_{T}^{ll} and the invariant mass m_{ll} defined as

$$m_{\mathrm{ll}} = \sqrt{(E_{\mathrm{lep1}} + E_{\mathrm{lep2}})^2 - (\vec{p}_{\,\mathrm{lep1}} + \vec{p}_{\,\mathrm{lep2}})^2}, \tag{5.1}$$

where $E_{\mathrm{lep1/lep2}}$ and $\vec{p}_{\,\mathrm{lep1/lep2}}$ are the energies and momentum vectors for the leading and sub-leading leptons forming the dilepton system. The minimal requirements on the

Selection
2 leptons ($e\mu$) of opposite charge
$p_{\mathrm{T}}^{\mathrm{lep1}} > 25$ GeV
$p_{\mathrm{T}}^{\mathrm{lep2}} > 13$ GeV (10 GeV in 2016)
No additional lepton with $p_{\mathrm{T}} > 10$ GeV
$12 < m_{\mathrm{ll}} < 76.2$ or $m_{\mathrm{ll}} > 106.2$ GeV
$p_{\mathrm{T}}^{ll} > 30$ GeV
$p_{\mathrm{T}}^{\mathrm{miss}} > 20$ GeV
$m_{\mathrm{T}}^{\mathrm{lep2}} > 30$ GeV
$60 < m_{\mathrm{T}}^{\mathrm{H}} < 125$ GeV
b-jet veto

Table 5.1: Summary of the base and orthogonality selection criteria applied in the signal regions.

dilepton system are $m_{\mathrm{ll}} > 12$ GeV and $p_{\mathrm{T}}^{ll} > 30$ GeV. To further target the leptonic final state in the $H \to WW^*$ decay channel and to suppress possible background contribution in all affected signal regions, we place requirements on the invariant transverse masses $m_{\mathrm{T}}^{\mathrm{H}} > 60$ GeV and $m_{\mathrm{T}}^{\mathrm{lep2}} > 30$ GeV, where

$$m_{\mathrm{T}}^{\mathrm{H}} = \sqrt{2p_{\mathrm{T}}^{\mathrm{ll}} p_{\mathrm{T}}^{\mathrm{miss}} \left[1 - \cos\Delta\theta\left(\vec{p}_{\mathrm{T}}^{\,\mathrm{ll}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\right)\right]} \tag{5.2}$$

and

$$m_{\mathrm{T}}^{\mathrm{lep2}} = \sqrt{2p_{\mathrm{T}}^{\mathrm{lep2}} p_{\mathrm{T}}^{\mathrm{miss}} \left[1 - \cos\Delta\theta\left(\vec{p}_{\mathrm{T}}^{\,\mathrm{lep2}}, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\right)\right]}. \tag{5.3}$$

Here, the cosine of the angle $\Delta\theta$ between the vectors explicitly stated in the brackets is used. Finally, events with reconstructed b-jets above $p_{\mathrm{T}} > 20$ GeV threshold are vetoed to suppress the top quark-induced background more effectively. The requirements enlisted in this subsection might be inverted or altered to define kinematically clean control regions, as will be further discussed .

To avoid overlap with the off-shell $H^* \to ZZ \to 2l2\nu$ analysis [140], the region of $76.2 < m_{\mathrm{ll}} < 106.2$ GeV is excluded from our phase space. Orthogonality with the future off-shell $H^* \to WW$ study is also assured by including $m_{\mathrm{T}}^{\mathrm{H}} < 125$ GeV cut. Furthermore, the $m_{\mathrm{T}}^{\mathrm{H}} > 60$ requirement not only suppresses the DY background, it also serves as a necessary requirement to avoid overlap with the $H \to \tau\tau$ analysis [141]. Orthogonality cuts were applied on top of the base selection and were found to have a negligible effect on sensitivity to constrain ACs while allowing for future combination across different Higgs decay channels. For the sake of clarity, we summarise the aforementioned base and orthogonality selection criteria in Table 5.1.

Variable	ggF	VBF	Resolved VH	Boosted VH
N_{jet} (AK8 V-jets)	0	0	0	> 0
N_{jet} (AK4 resolved jets)	0 and 1	2	2	–
m_{jj}	–	> 120 GeV	$[60, 120]$ GeV	–
$\lvert\eta_{\text{jet}}\rvert$	–	< 4.7	< 2.4	< 2.4

Table 5.2: Summary of the ggF, VBF, and VH categorisation criteria used for the HVV coupling study.

5.1.2 Signal regions

Signal regions are designed to target different Higgs boson production processes based on the associated jet activity in the event. Table 5.2 summarises the categorisation requirements on the number of resolved AK4 jets, the number of AK8 V-jets, the dijet invariant mass, and the pseudorapidity range for the jets to be applied on top of the base and orthogonality criteria listed in Table 5.1.

As mentioned in Section 3.2.4, the AK4 (AK8) jets are expected to have transverse momentum above 30 GeV (200 GeV) threshold. LO Higgs boson production via the ggF process is typically not accompanied by more than one resolved jet. In that regard, the corresponding ggF SRs require zero or maximum of one AK4 jet in the final state. For the VBF and VH Resolved SRs, we require exactly two resolved AK4 jets with a dijet invariant mass of $m_{\text{jj}} > 120$ GeV and $60 < m_{\text{jj}} < 120$ GeV, respectively. It turns out that in some cases, two resolved AK4 jets from the hadronic V decay in VH topology appear highly collimated and can be reconstructed as one boosted AK8 jet. Further investigation has shown that the number of events showing VH Boosted topology is sensitive to the AC effects, thus including a dedicated Signal Region (SR) might increase the sensitivity of the VH production channel. Contrary to the HWW Legacy analysis, where this effect can be neglected, we define the VH Boosted SR by selecting events with at least one AK8 V-jet. To ensure the exclusivity of this category, a V-jet veto is applied to the rest of the SRs.

In practice, individual SRs might be further divided into kinematic sub-regions depending on the particular KDs used for the construction of analysis templates. The particular division will be pointed out explicitly in Section 5.2.

5.1.3 Control regions

The CRs are defined using the base and orthogonality selection criteria in Table 5.1 with some of the cuts on relevant observables altered as summarised in Table 5.3. In particular, background processes might be enhanced by cutting on mass variables, such as m_{ll}, m_{T}^{H} and $m_{\text{T}}^{\text{lep2}}$, the range of which varies depending on the mother particle inducing the targeted background process. For example, DY production in our phase space is mostly represented by the $Z \to \tau\tau \to e\mu + \text{MET}$ decay chain, thus, selecting the dilepton mass window of $40 < m_{\text{ll}} < 80$ GeV (some of the mass spectrum is taken by neutrinos) and $m_{\text{T}}^{\text{H}} < 60$ GeV was found to be a natural phase space for selecting DY events. In general, cuts are chosen to ensure the exclusivity of events (i.e. no event falls to multiple CRs/SRs). In that regard, a requirement of b-jet veto is also applied in the DY CR

Variable	DY CR	Top CR	WW CR
$m_{\ell\ell}$	$[40, 80]$ GeV	> 50 GeV	> 106.2 GeV
$m_{\mathrm{T}}^{\mathrm{H}}$	< 60 GeV	–	$[60, 125]$ GeV
$m_{\mathrm{T}}^{\mathrm{lep2}}$	–	> 30 GeV	> 30 GeV
b-jet veto	yes	no	yes
Inverse b-jet veto	no	yes	no

Table 5.3: Summary of the DY, Top and WW CR requirements. Additional sub-regions are defined according to jet multiplicity following the SR categorisation.

and WW CR, while an inverse of the cut on the b-tagging classifier (see Section 4.3.2) is used to define the inversed b-jet veto employed in the Top CR to enhance the number of events with top quark production.

The CRs in this analysis are used in a twofold way. Initially, various physical object distributions are plotted for both MC simulation and data to check for any discrepancies in object reconstruction or the definition of kinematic discriminants. Secondly, background normalisation in the analysis SRs might differ from reality. To correct possible data-to-MC discrepancies, rate parameters controlling the normalisation of dominant backgrounds are left freely floating in the fit on data, such as the correct background yield can be interpolated from dedicated CRs, where this number is constrained. For that reason, the CRs are further divided into sub-regions depending on jet multiplicity to partially match the final state composition in the SRs under study. In particular, we define 0-jet and 1-jet DY CRs and Top CRs to target the DY and top contamination in ggF signal channels. The normalisation of the nonresonant WW background in ggF channels is extracted directly from the fit of the SR phase space. This is possible thanks to the distinctive shape of WW contamination in this region. The normalisation of the dominant DY, top and nonresonant WW backgrounds in 2-jet and V-jet channels is extracted from the corresponding DY, Top and WW CRs with exactly 2 resolved jets and 1 or more V-jets, respectively.

In conclusion, let us stress that the variable shape in the CRs is important to validate the background description, however, it is not used in the final fit to constrain the ACs, where only the background yield is considered. Examples of 0-jet and 1-jet CR distributions are presented in Figures 5.1 – 5.4 for all 3 years of data-taking. The simulated distribution shapes (filled histograms) are reasonably consistent with the data points within the assigned statistical and systematic uncertainties on the MC (dubbed as "Syst." in the figures) and the statistical uncertainties on data (vertical bars in the figures). A normalisation effect of $\approx 10\%$ (15%) is present in the case of DY CR + 0-jet (DY CR + 1-jet) for the 2016 simulation, however, this is not meant to be a concern given that the DY normalisation is left floating in the fit. The agreement in the DY CRs was also largely improved by considering the data-driven method of estimating the DY background discussed in Section 4.3.1.

5.2 Kinematic discriminants

The AC contribution to the HVV vertex is, by definition, very subtle. The strategic choice of observables sensitive to these effects, also called Kinematic Discriminants (KDs), is a prerequisite for successfully constraining AC values. The particular definition of the KDs differs for each SR. For the sake of clarity and completeness, we present an effort in constructing and optimizing the KDs for all SR topologies. However, it is important to acknowledge the significant contributions solely made by my colleagues in developing and refining the KDs specifically for the VBF and VH channels.

Additionally, it is worth noting that the KD distributions in the analysis SRs, as displayed throughout this section, show estimated signal and background contributions corresponding to the expected yield prior to the final fit, also known as "prefit" templates. Data points were only added to the prefit distributions after this analysis was finalized and were never examined during the early analysis stage.

5.2.1 KD in VBF and VH channels

As briefly mentioned in Section 4.4.1, MELA-computed matrix elements (ME), if determined at the reconstruction-level, could be used to build discriminants suitable for the extraction of an alternative signal. Unfortunately, one of the most challenging features of the leptonic $H \to WW^*$ decay channel is the presence of neutrinos, preventing us from full event reconstruction. In that regard, we are unable to fully explore the AC contribution to the HVV decay vertex, even though we can still partially rely on reconstructible decay observables, such as the dilepton system invariant mass $m_{\ell\ell}$. On the other hand, the EWK-like production topology, typical for the VBF and VH signal, might be used to our advantage. The requirement of two resolved jets or one boosted V-jet accompanying Higgs boson production provides us with extra kinematic information necessary to explore the HVV production vertex using the MELA technique. In practice, the following ingredients are needed in order to compute the production ME:

- four-vectors for two resolved AK4 jets in VBF + 2-jets and Resolved VH phase space,
- four-vectors for two AK4 sub-jets used in construction of boosted AK8 V-jet in Boosted VH topology,
- the proxy Higgs boson four-vector.

The last ingredient in the list represents an approximative representation of the reconstructed Higgs boson using the kinematics of two charged leptons and missing transverse energy. In particular, the system of two neutrinos ($\nu\nu$) that would normally be used in Higgs boson reconstruction was estimated as follows:

- The p_x and p_y momentum components of the dineutrino system are approximated by the corresponding x and y components in the $p_{\mathrm{T}}^{\mathrm{miss}}$ vector,
- The p_z component of the dineutrino system is set to the z component of the dilepton system as these variables were found to be highly correlated at the generator-level,
- Finally, the dineutrino mass was assigned a value corresponding to the mean of the $m_{\nu\nu}$ distribution obtained at the generator-level.

The resulting $\nu\nu$ four-vector was then combined with that of the reconstructed dilepton system to create the proxy Higgs boson four-vector. Note that the MELA technique of probability density calculation for the production HVV vertex is largely based on the kinematics of the associated jets. As such, the reconstruction of the proxy Higgs four-vector has little effect on the final discriminants.

The application of the MELA technique in VBF and VH regions was motivated by its success in similar CMS analyses performed using Run-1 datasets in various decay channels [33] as well as in the $H \to ZZ$ [64] and $H \to \tau\tau$ [141] decay channels using datasets collected during the Run-2 period of data-taking. In this analysis, three types of MELA-based kinematic discriminants were found to have sufficient discrimination power. The discriminant $\mathcal{D}_{\text{sig}}$, defined as

$$\mathcal{D}_{\text{sig}} = \frac{\mathcal{P}_{\text{sig}}(\vec{\Omega})}{\mathcal{P}_{\text{sig}}(\vec{\Omega}) + \mathcal{P}_{\text{bkg}}(\vec{\Omega})}, \tag{5.4}$$

can be used to distinguish between different signal production mechanisms and background processes. The second type of discriminant, $\mathcal{D}_{\text{BSM}}$, differentiates between pure AC processes and pure SM contributions. It is defined as

$$\mathcal{D}_{\text{BSM}} = \frac{\mathcal{P}_{\text{BSM}}(\vec{\Omega})}{\mathcal{P}_{\text{BSM}}(\vec{\Omega}) + \mathcal{P}_{\text{SM}}(\vec{\Omega})}, \tag{5.5}$$

where BSM is generally standing for various AC scenarios considered in this analysis. Throughout this text, we have adopted the $\mathcal{D}_{\text{BSM}}$ notation emphasizing spin-parity properties for individual couplings, e.g. $a_2 \to \mathcal{D}_{0+}$, $a_3 \to \mathcal{D}_{0-}$, $a_{\Lambda 1} \to \mathcal{D}_{\Lambda 1}$ and $a_{\Lambda 1}^{Z\gamma} \to \mathcal{D}_{\Lambda 1}^{Z\gamma}$. Finally, we construct a discriminant designed to isolate SM-BSM interference distribution,

$$\mathcal{D}_{\text{int}} = \frac{\mathcal{P}^{\text{int}}_{\text{SM-BSM}}(\vec{\Omega})}{\mathcal{P}_{\text{BSM}}(\vec{\Omega}) + \mathcal{P}_{\text{SM}}(\vec{\Omega})}, \tag{5.6}$$

where "int" label denotes interference between two couplings. In the case of interference between the SM and a_3 couplings, we substitute the "int" label with the "CP" label to emphasize the sensitivity of the corresponding KD to distinguish between CP-even and CP-odd states. Similarly, as in the MELA technique for the sample re-weighting, the probability density terms $\mathcal{P}$ (now entering the aforementioned KD formulae) are normalised to give the same integrated cross-sections in the relevant phase space of each process. Such normalisation leads to an intuitive range of allowed KDs values. In the case of $\mathcal{D}_{\text{sig}}$ and $\mathcal{D}_{\text{BSM}}$, the allowed range is $[0, 1]$, while in the case of $\mathcal{D}_{\text{int}}$, the quasi-probabilistic nature of the discriminant allows for negative values, i.e. the assumed range is $[-1, 1]$.

A few general remarks on the KDs should be noted at this moment. In the VBF SR, $\mathcal{D}_{\text{VBF}}$ was constructed according to relation 5.4, where $\mathcal{P}_{\text{sig}}$ represents the probability density for the VBF production hypothesis and $\mathcal{P}_{\text{bkg}}$ was chosen to represent the probability density for the ggF NLO production with 2 associated jets. $\mathcal{D}_{\text{VBF}}$ was also shown to serve as a good discriminant between VBF signal and SM background processes. In the case of VH categories, the relevant VH production discriminant has not shown sufficient separation power, therefore it is not considered. The discriminant isolating SM-BSM interference is expected to show a certain asymmetry between the number of events falling into the positive or negative part of the $\mathcal{D}_{\text{CP}}$ spectrum, especially for events representing a mixed CP state. As such, we divide events in VBF and VH channels into

the "forward-backward" sub-categories, i.e. $\mathcal{D}_{\text{CP}} < 0$ and $\mathcal{D}_{\text{CP}} > 0$, to target the study of CP-odd couplings. Similarly, $\mathcal{D}_{\text{int}}$ is sensitive to the sign of interference between CP-even a_2 and SM couplings, therefore, the forward-backward categorisation is also considered. In this case, however, the value of $\mathcal{D}_{\text{int}} = 0.4$ is used to define the sub-categories in order to symmetrize the SM Higgs boson expectation. It was found out that $a_{\Lambda 1}$ and $a_{\Lambda 1}^{Z\gamma}$ interference discriminants are highly correlated to those of the corresponding $\mathcal{D}_{\text{BSM}}$. As such, they are not considered.

Finally, the Multidimensional Kinematic Discriminants (MKDs) are constructed from generally multiple MELA KDs as well as the m_{ll} distribution and are eventually used for the signal extraction fit. Depending on the supposed VBF or VH channel and analysis approach, particular form of the MKD is represented by distinctive template histogram with bins being a projection of the individual bins corresponding to KD and m_{ll} distributions. A detailed explanation of the binning scheme is in disposal throughout the following subsections and is summarised in Table 5.4. The presented template binning was optimized 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.

VBF and VH MKD in Approach 1

In the case of Approach 1, only one AC (a_2, a_3, $a_{\Lambda 1}$ or $a_{\Lambda 1}^{Z\gamma}$) is considered alongside the SM scenario. In that regard, MKDs are optimized to show superior sensitivity for that particular choice of the studied AC. The following binning structure was considered:

- VBF + 2-jets channel: Two bins of $\mathcal{D}_{\text{VBF}}$ corresponding to low and high purity are defined within the bin threshold of 0.75. Targeting extra sensitivity to the HVV decay vertex, two bins of dilepton invariant mass in the range $m_{\text{ll}} \in [12, 76.2]$ GeV are defined within the 45 GeV bin threshold. Eventually, 10 equidistant bins in $[0, 1]$ range are considered for the $\mathcal{D}_{\text{BSM}}$ KD, where one of $\mathcal{D}_{0+}$, $\mathcal{D}_{0-}$, $\mathcal{D}_{\Lambda 1}$ or $\mathcal{D}_{\Lambda 1}^{Z\gamma}$ is chosen depending on the studied coupling scenario. Let us remind you that $\mathcal{D}_{\text{CP}}$ and $\mathcal{D}_{\text{int}}$ were used to define VBF forward-backward sub-categories.

- VH Resolved and Boosted channels: Dilepton invariant mass is considered employing the same binning as in the VBF case. Besides that, four equidistant $\mathcal{D}_{\text{BSM}}$ bins are considered in the Resolved VH category, while three $\mathcal{D}_{\text{BSM}}$ bins are defined within the 0.6 and 0.8 threshold values – the first large bin is due to small signal statistics in this region. Forward-background categorisation is taken into account for the $\mathcal{D}_{\text{CP}}$ discriminant, while $\mathcal{D}_{\text{int}}$ was found to be highly correlated with $\mathcal{D}_{0+}$. As such, it is not considered.

Prefit signal and background template distributions, as produced with the Approach 1 binning scheme, are displayed in Figure 5.5 (a_3), Figure 5.6 (a_2), Figure 5.7 ($a_{\Lambda 1}$), and Figure 5.8 ($a_{\Lambda 1}^{Z\gamma}$). For the sake of clarity, only pure SM and pure BSM templates, i.e. no interference, were plotted. This way, it is convenient to compare the distinct shapes of MKD distributions.

VBF and VH MKD in Approach 2

In the case of Approach 2, we intend to extract all considered couplings (a_3, a_2 and $a_{\Lambda 1}$) simultaneously, therefore, the categorisation and binning schemes were adjusted. Only

Analysis	Category	Sub-category	Final discriminant
HVV Approach 1	VBF (a_3)	$\mathcal{D}_{\text{CP}}$	$[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{0-}]$
	VBF (a_2)	$\mathcal{D}_{\text{int}}$	$[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{0+}]$
	VBF ($a_{\Lambda 1}$)	-	$[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{\Lambda 1}]$
	VBF ($a_{\Lambda 1}^{Z\gamma}$)	-	$[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{\Lambda 1}^{Z\gamma}]$
	VH (a_3)	$\mathcal{D}_{\text{CP}}$	$[m_{\text{ll}}, \mathcal{D}_{0-}]$
	VH (a_2)	-	$[m_{\text{ll}}, \mathcal{D}_{0+}]$
	VH ($a_{\Lambda 1}$)	-	$[m_{\text{ll}}, \mathcal{D}_{\Lambda 1}]$
	VH ($a_{\Lambda 1}^{Z\gamma}$)	-	$[m_{\text{ll}}, \mathcal{D}_{\Lambda 1}^{Z\gamma}]$
	0- and 1-jet ggF	-	$[m_{\text{T}}^{\text{H}}, m_{\text{ll}}]$
HVV Approach 2	VBF	$\mathcal{D}_{\text{CP}}$, $\mathcal{D}_{\text{int}}$	$[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{0-}, \mathcal{D}_{0+}]$
	VH	$\mathcal{D}_{\text{CP}}$	$[m_{\text{ll}}, \mathcal{D}_{0-}, \mathcal{D}_{0+}]$
	0- and 1-jet ggF	-	$[m_{\text{T}}^{\text{H}}, m_{\text{ll}}]$

Table 5.4: The kinematic observables used for interference-based categorisation and for the final discriminants used in the fit to data to study the HVV anomalous couplings (AC). For each of the AC scenarios in Approach 1, we have a dedicated analysis in the VBF and VH categories. In Approach 2, we use one analysis to target all ACs simultaneously. The two-dimensional kinematic discriminant based on the decay observables, m_{ll} and m_{T}^{H}, was optimized to separate the SM signal and background, as well as the AC contribution in the ggF categories. The final discriminant bin thresholds are distict for the VH Resolved and VH Boosted channels.

one MKD is defined per each analysis channel, making it sensitive to all ACs at the same time. Of course, the sensitivity of such a general discriminant is largely dependent on signal statistics in individual bins. The following strategy was chosen:

- VBF + 2-jets channel: Four sub-categories using both $\mathcal{D}_{\text{CP}}$ and $\mathcal{D}_{\text{int}}$ discriminants are considered. $\mathcal{D}_{\text{VBF}}$ and m_{ll} are also used, employing the same binning as in Approach 1. The $\mathcal{D}_{\text{BSM}}$ discriminants targeting a_3, a_2 and $a_{\Lambda 1}$ couplings were all found to show similar performance. As such, we have decided to use $\mathcal{D}_{0-}$ and $\mathcal{D}_{0+}$ to cover both CP-odd and CP-even couplings, respectively. Initially, both KDs were represented by three bins in $[0, 1]$ range, as defined within the threshold of 0.1 and 0.9. Consequently, dedicated re-binning in the $[\mathcal{D}_{0-}, \mathcal{D}_{0+}]$ space was applied by merging bins showing an abundance of SM Higgs events or with low precision in the background prediction.

- VH Resolved and Boosted channels: A forward-backward scheme is used to define two sub-categories using the $\mathcal{D}_{\text{CP}}$ discriminant independently in both VH Resolved and Boosted channels. The Approach 1 binning scheme for the m_{ll} distribution was also used in this case. Again, $\mathcal{D}_{0-}$ and $\mathcal{D}_{0+}$ discriminants were implemented and are using three bins defined within the 0.25 and 0.75 thresholds in the VH

Resolved channel and two bins within the 0.8 threshold in the VH Boosted channel. Dedicated re-binning in $[\mathcal{D}_{0-}, \mathcal{D}_{0+}]$ space was also applied.

5.2.2 KD in ggF channels

The difficulty of the ggF SR lies in the limited kinematic information due to both the presence of missing transverse momentum and the lack of associated jets in the final state (as much as the LO ggF production is concerned). In such a case, the application of the MELA technique is fully out of picture.

Instead, various reconstruction-level observables describing the final state topology were studied for sensitivity to AC effects, with $m_{\ell\ell}$ and $m_{\mathrm{T}}^{\mathrm{H}}$ historically (e.g. analysis [33]) being suitable candidates. Arguably, the difference in azimuthal angles between the two charged final state leptons, $\Delta\phi_{ll}$, was found to be sensitive to AC effects. Nevertheless, a known issue with the reconstruction procedure for this variable, as observed exclusively in DY events (as seen in the trend-like behavior forming in Figure 5.10), has not yet been resolved. As such, we prefer to keep using the $m_{\ell\ell}$ distribution (together with the $m_{\mathrm{T}}^{\mathrm{H}}$ distribution), which is assumed to be highly correlated with $\Delta\phi_{ll}$ and does not exhibit reconstruction issues.

The limited decay kinematic phase space and subtle differences between SM and BSM signals in ggF channels do not allow us to construct KDs targeting individual anomalous couplings. Instead, one dedicated binning scheme in the $[m_{\mathrm{T}}^{\mathrm{H}}, m_{\ell\ell}]$ space was optimized to ensure both good separation between the SM signal and background, as well as between the different anomalous signal hypotheses. The binning optimization procedure was performed as follows:

1. As a starting point, the original binning scheme optimized for the separation of the SM signal and background in the HWW Legacy analysis was used. In particular, the HWW Legacy analysis employed a binning scheme defined as

   $$[m_{\mathrm{T}}^{\mathrm{H}}, m_{\ell\ell}] \equiv [\{60, 95, 110, 125\}, \{12, 17, 25, 30, 35, 40, 45, 65, 200\}],$$

   where the thresholds are stated in GeV units.

2. Two-dimensional distributions, such as those shown in Figures 5.12 for different ggF signal hypotheses, were examined to identify how the signal is spread across the $[m_{\mathrm{T}}^{\mathrm{H}}, m_{\ell\ell}]$ space. This observation led to the decision to make the low $m_{\mathrm{T}}^{\mathrm{H}}$ binning finer, but also to keep the original finely defined $m_{\ell\ell}$ binning.

3. It was also observed that VH events falling into the ggF SRs are showing a relatively long tail in the $m_{\ell\ell}$ distribution (see Figure 5.11). Splitting the overflow bin (i.e. the bin where all events with $m_{\ell\ell}$ higher than the last threshold are falling) into two parts showed an increase in sensitivity to extract anomalous signal.

4. A rough estimate of the new binning was then systematically tested for optimal bin thresholds by maximizing the sensitivity of the (expected) final fit in the ggF channels while shifting bin thresholds.

The final optimized binning scheme consists of nine $m_{\ell\ell}$ bins and six $m_{\mathrm{T}}^{\mathrm{H}}$ bins:

$$[m_{\mathrm{T}}^{\mathrm{H}}, m_{\ell\ell}] \equiv [\{60, 70, 80, 90, 100, 110, 125\}, \{12, 17, 25, 30, 35, 40, 45, 65, 100, 200\}].$$

Prefit signal and background template distributions for the ggF 0-jet and 1-jet regions separately are displayed in Figure 5.9. The same binning scheme is used for both Approach 1 and 2.

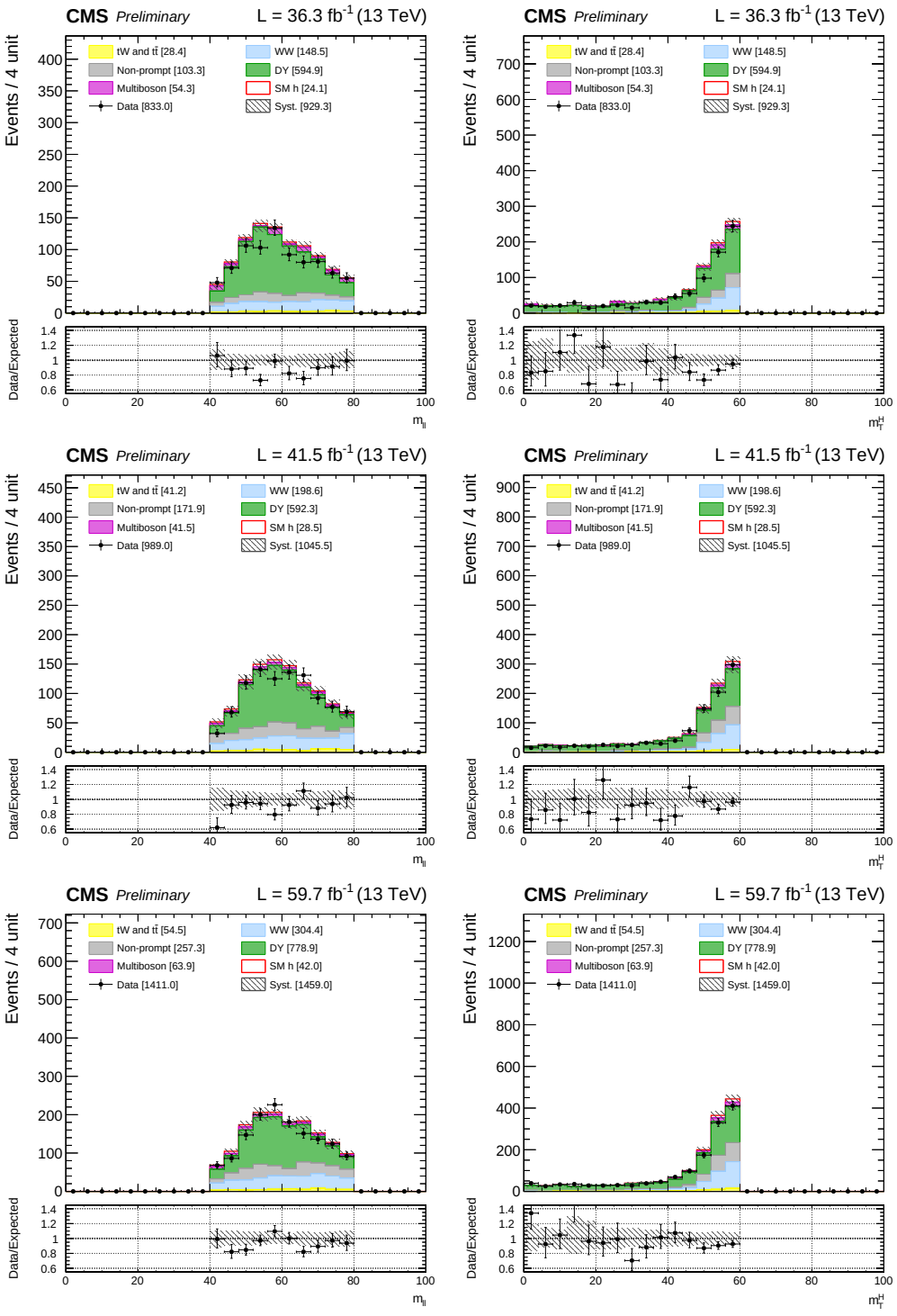

Figure 5.1: DY CR + 0-jet distributions for the 2016 (top row), 2017 (middle row) and 2018 (bottom row) datasets. Displayed observables are: the dilepton invariant mass m_{ll} (left) and the $m_{\mathrm{T}}^{\mathrm{H}}$ invariant mass (right).

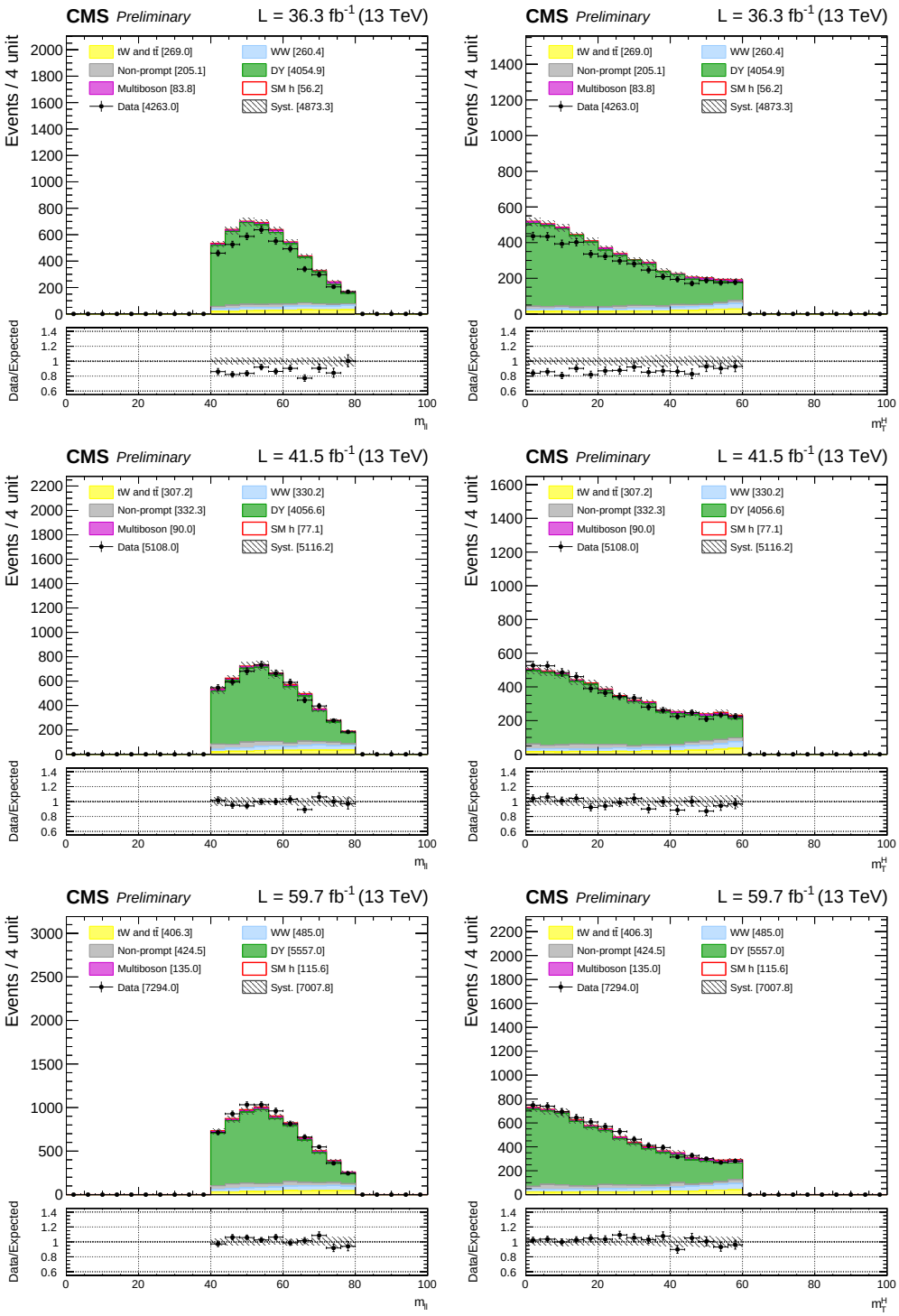

Figure 5.2: DY CR + 1-jet distributions for the 2016 (top row), 2017 (middle row) and 2018 (bottom row) datasets. Displayed observables are: the dilepton invariant mass m_{ll} (left) and the $m_{\mathrm{T}}^{\mathrm{H}}$ invariant mass (right).

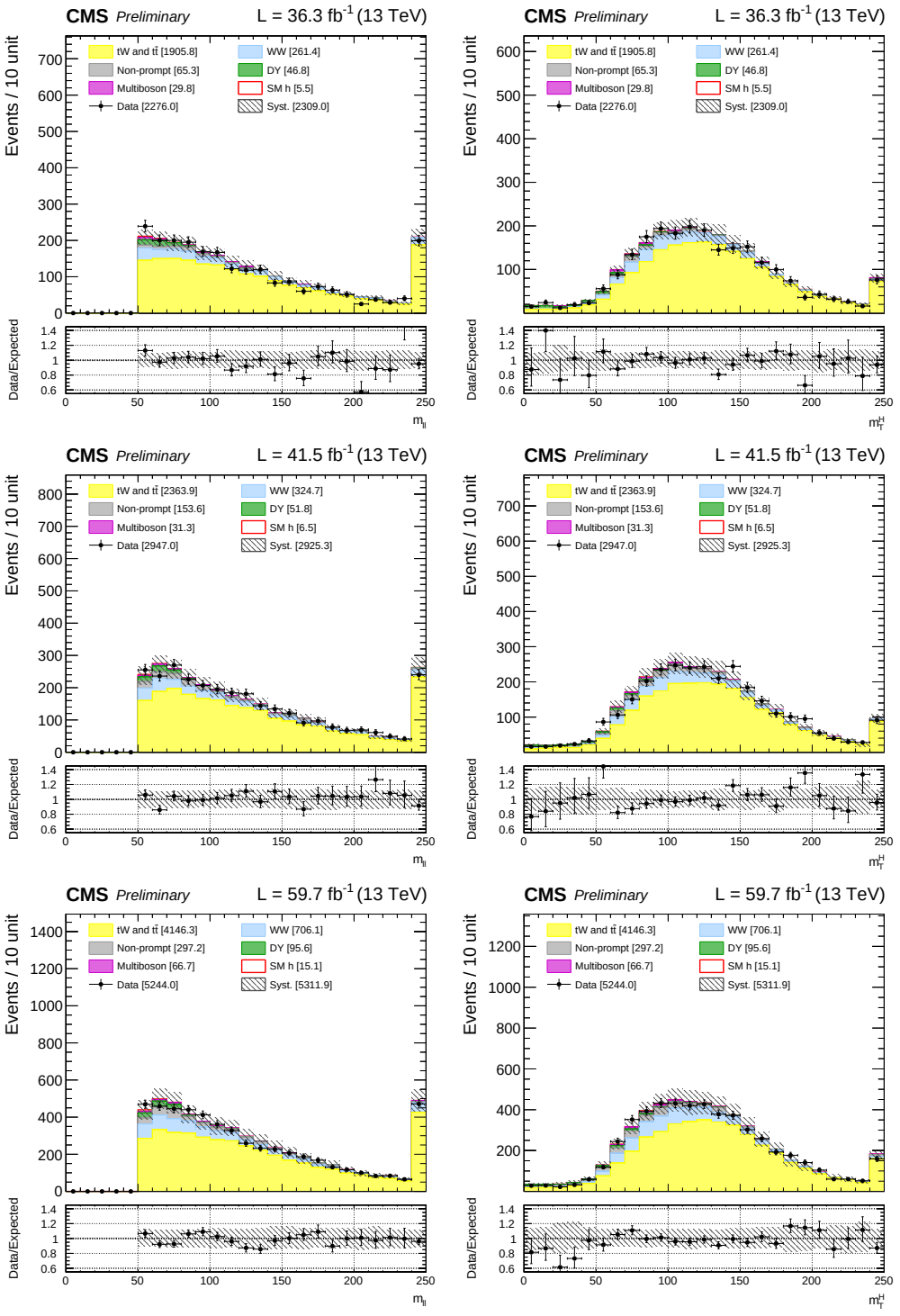

Figure 5.3: Top CR + 0-jet distributions for the 2016 (top row), 2017 (middle row) and 2018 (bottom row) datasets. Displayed observables are: the dilepton invariant mass m_{ll} (left) and the $m_{\mathrm{T}}^{\mathrm{H}}$ invariant mass (right).

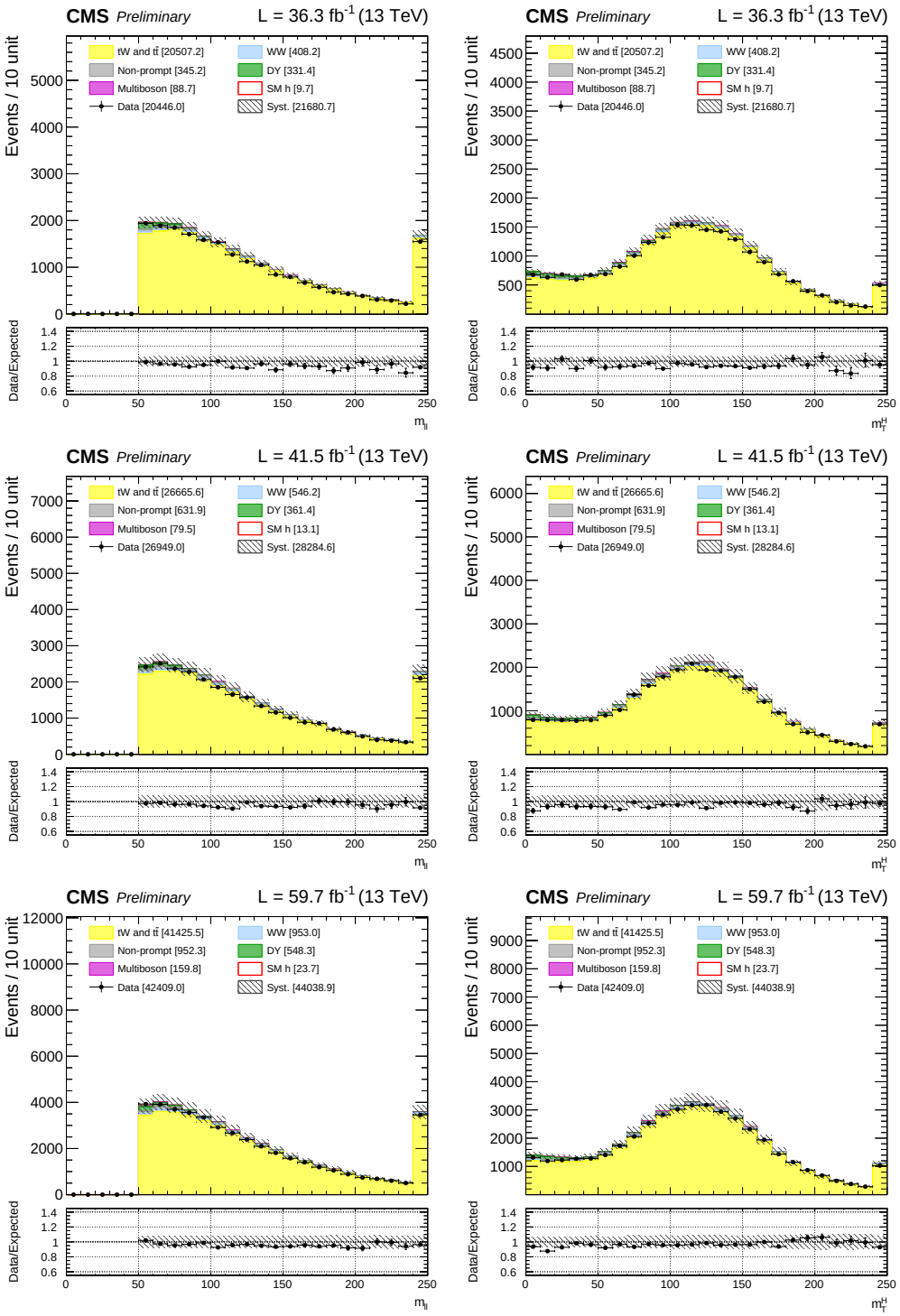

Figure 5.4: Top CR + 1-jet distributions for the 2016 (top row), 2017 (middle row) and 2018 (bottom row) datasets. Displayed observables are: the dilepton invariant mass m_{ll} (left) and the $m_{\mathrm{T}}^{\mathrm{H}}$ invariant mass (right).

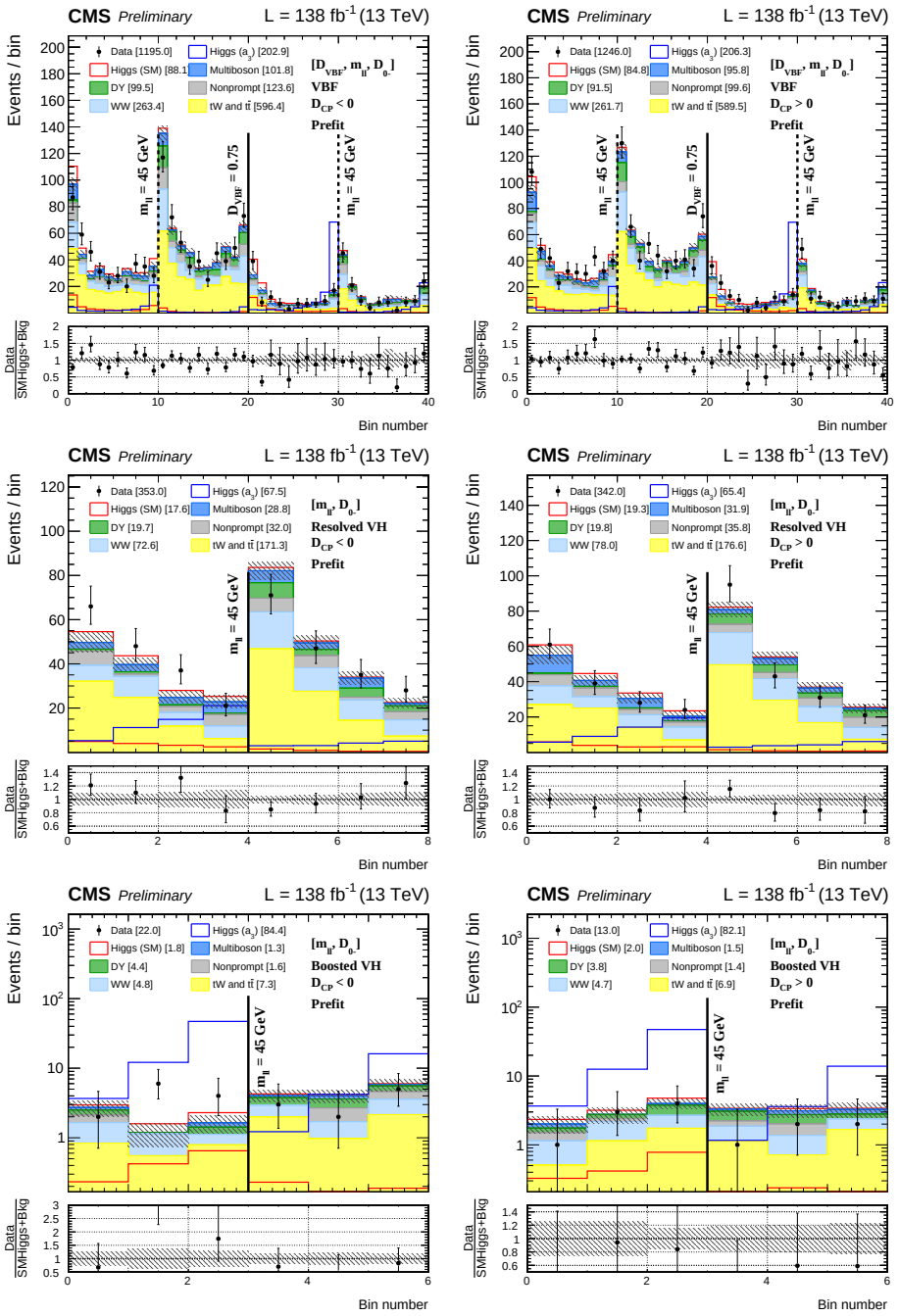

Figure 5.5: Prefit distributions of events in Approach 1 for $[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{0-}]$ in the VBF channel (top row) and for $[m_{\text{ll}}, \mathcal{D}_{0-}]$ in the VH Resolved (middle row) and VH Boosted (bottom row) channels. For each channel, the $\mathcal{D}_{\text{CP}} < 0$ (left) and $\mathcal{D}_{\text{CP}} > 0$ (right) subcategories are shown. The SM and a_3 HVV coupling signals are considered. Displayed error bars and bands represent the statistical uncertainty corresponding to the full Run-2 combination of the 2016-2018 datasets.

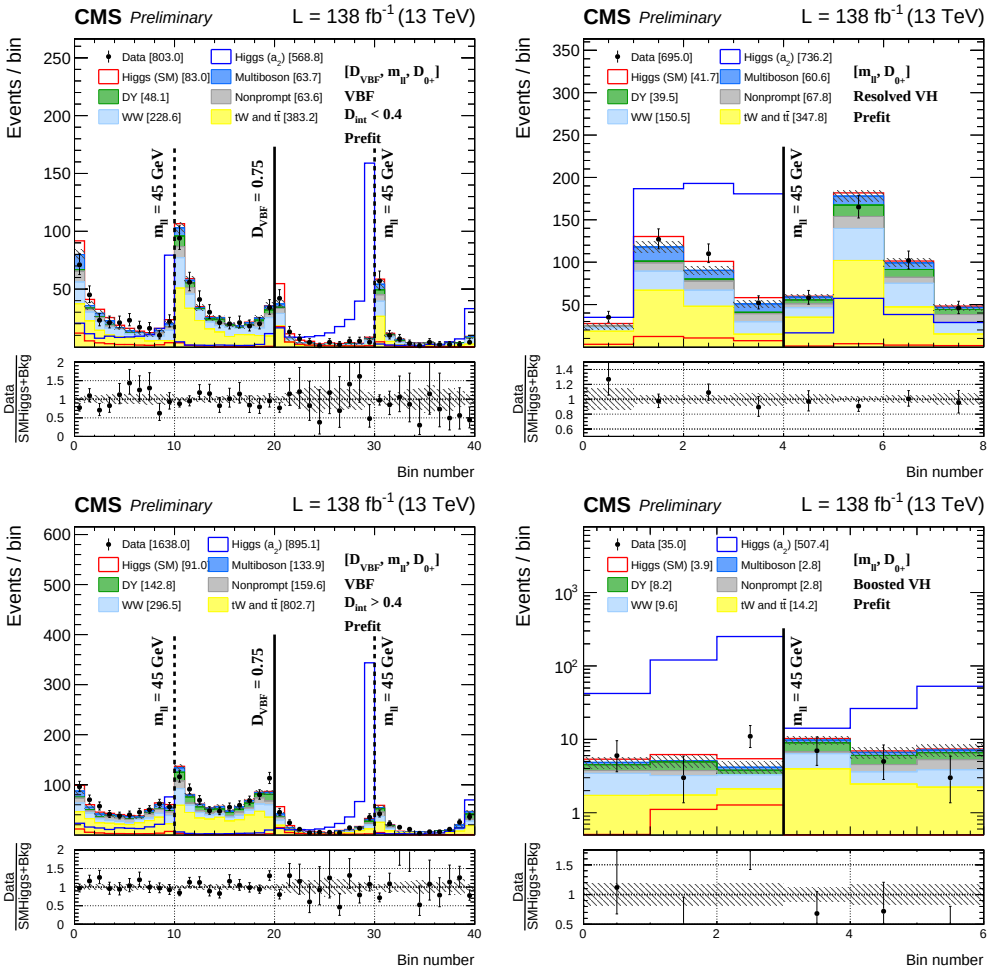

Figure 5.6: Prefit distributions of events in Approach 1 for [$\mathcal{D}_{\text{VBF}}$, m_{ll}, $\mathcal{D}_{0+}$] in the VBF channel (left) and for [m_{ll}, $\mathcal{D}_{0+}$] in the VH Resolved (right upper) and VH Boosted (right lower) channels. For the VBF channel, the $\mathcal{D}_{\text{int}} < 0.4$ (left upper) and $\mathcal{D}_{\text{int}} > 0.4$ (left lower) sub-categories are shown. The SM and a_2 HVV coupling signals are considered. Displayed error bars and bands represent the statistical uncertainty corresponding to the full Run-2 combination of the 2016-2018 datasets.

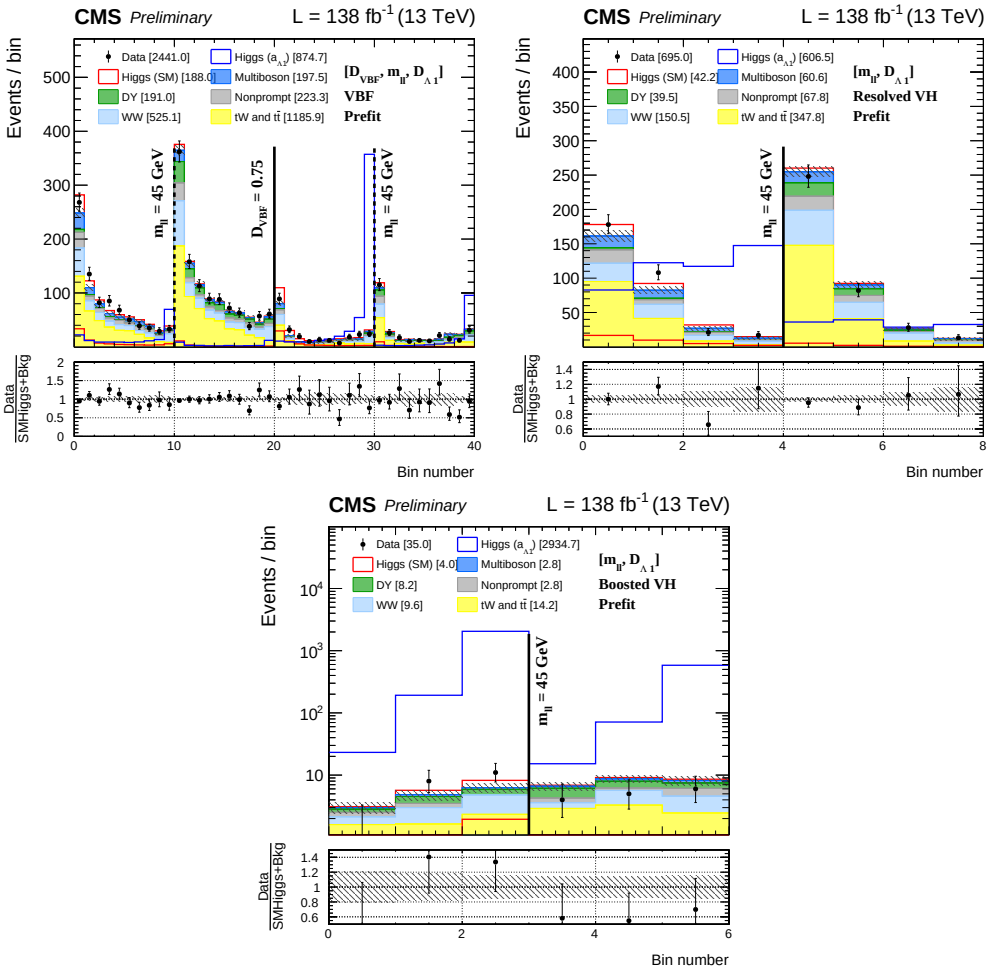

Figure 5.7: Prefit distributions of events in Approach 1 for $[\mathcal{D}_{\mathrm{VBF}}, m_{\mathrm{ll}}, \mathcal{D}_{\Lambda 1}]$ in the VBF channel (top left) and for $[m_{\mathrm{ll}}, \mathcal{D}_{\Lambda 1}]$ in the VH Resolved (top right) and VH Boosted (bottom) channels. The SM and $a_{\Lambda 1}$ HVV coupling signals are considered. Displayed error bars and bands represent the statistical uncertainty corresponding to the full Run-2 combination of the 2016-2018 datasets.

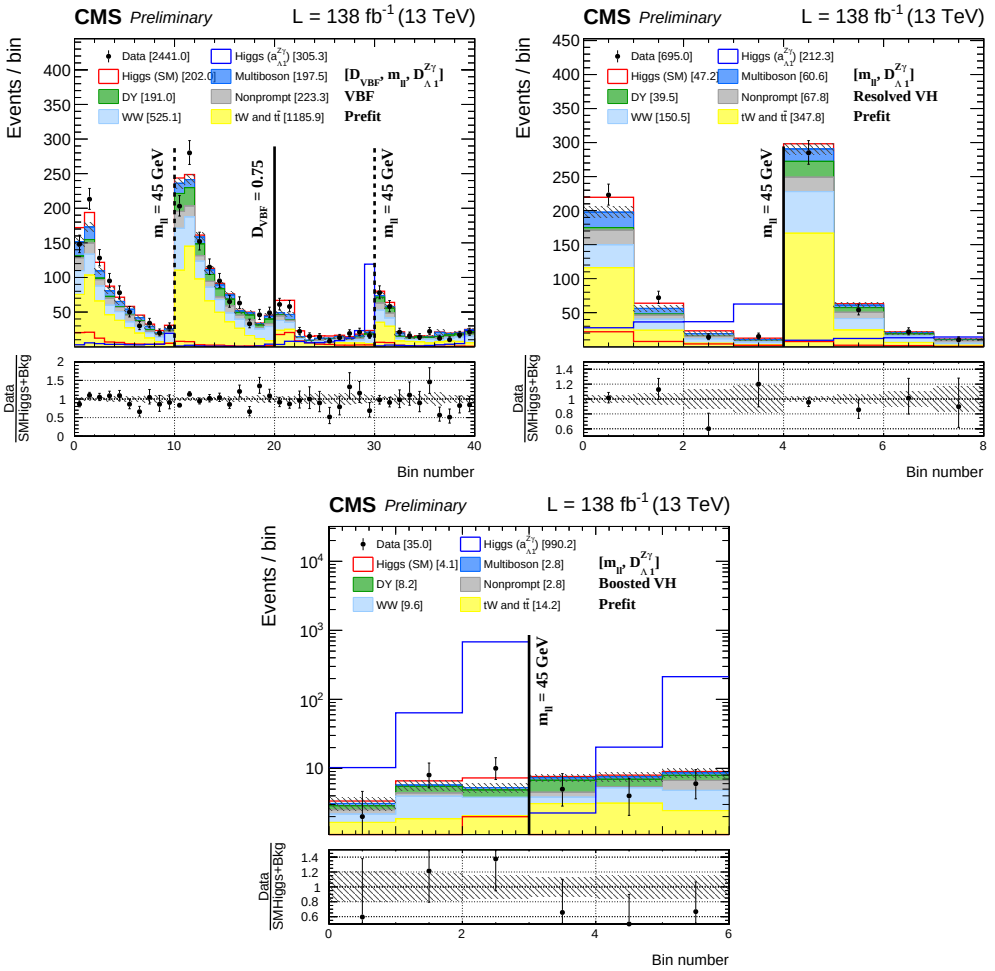

Figure 5.8: Prefit distributions of events in Approach 1 for $[\mathcal{D}_{\mathrm{VBF}}, m_{\mathrm{ll}}, \mathcal{D}^{Z\gamma}_{\Lambda 1}]$ in the VBF channel (top left) and for $[m_{\mathrm{ll}}, \mathcal{D}^{Z\gamma}_{\Lambda 1}]$ in the VH Resolved (top right) and VH Boosted (bottom) channels. The SM and $a^{Z\gamma}_{\Lambda 1}$ HVV coupling signals are considered. Displayed error bars and bands represent the statistical uncertainty corresponding to the full Run-2 combination of the 2016-2018 datasets.

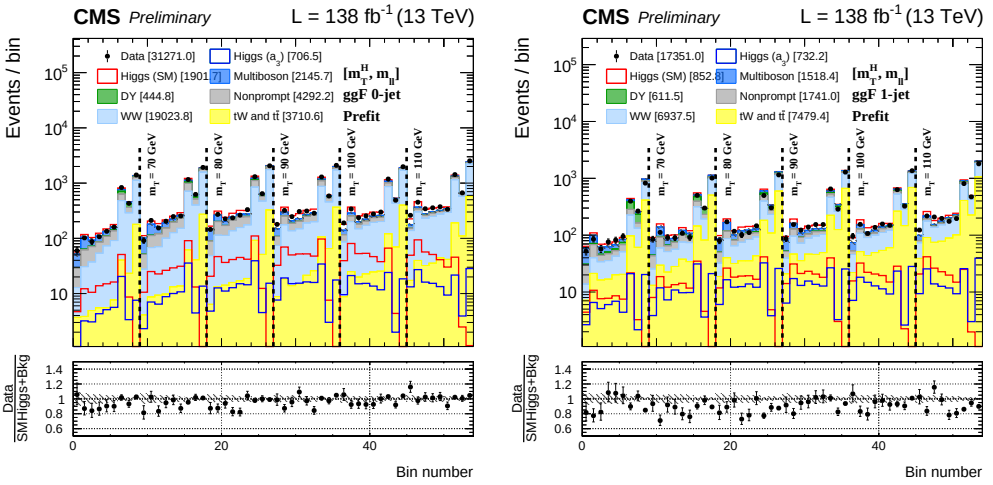

Figure 5.9: Prefit distributions of events for $[m_{\mathrm{T}}^{\mathrm{H}}, m_{\mathrm{ll}}]$ discriminant in the ggF + 0-jet channel (left) and ggF + 1-jet channel (right). The SM and a_3 HVV coupling signals are considered, however, the same binning scheme was optimized and implemented for signal extraction for the rest of anomalous couplings and both analysis approaches. Displayed error bars and bands represent the statistical uncertainty corresponding to the full Run-2 combination of the 2016-2018 datasets.

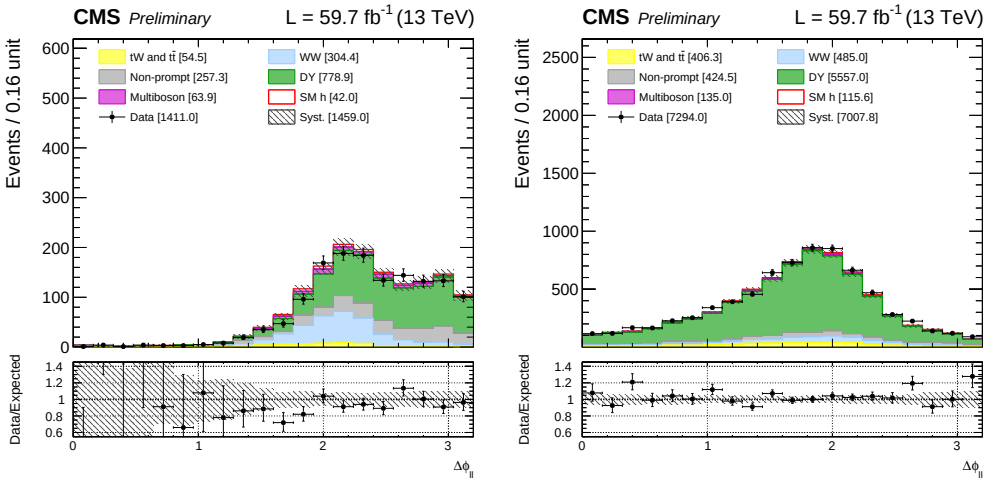

Figure 5.10: The distribution of the difference between the azimuth angles of two charged final state leptons, $\Delta\phi_{ll}$, was tested for possible application as a KDs in ggF + 0-jet (left) and ggF + 1-jet (right) channels. Reconstruction issues for this variable, visible as a positive slope trend in the data-to-MC ratio graph (especially for the 0-jet case), are not yet understood. As such, $\Delta\phi_{ll}$ was not considered in this analysis.

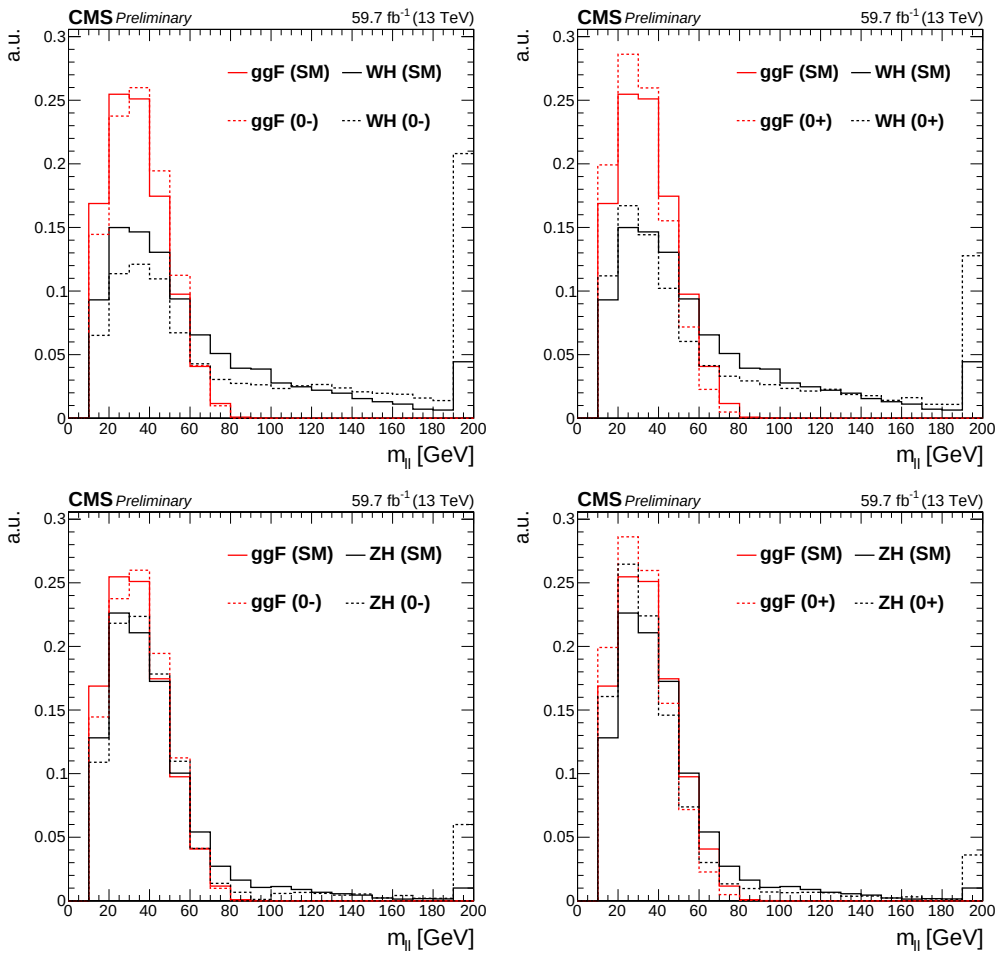

Figure 5.11: Normalized dilepton invariant mass m_{ll} distributions for WH (top row) and ZH (bottom row) Higgs boson production as compared to the ggF contribution. The displayed hypotheses correspond to that of the SM in comparison to pure a_3 (left) and a_2 (right) anomalous HVV couplings. The distributions are shown for the ggF + 1-jet SR in the 2018 MC simulation.

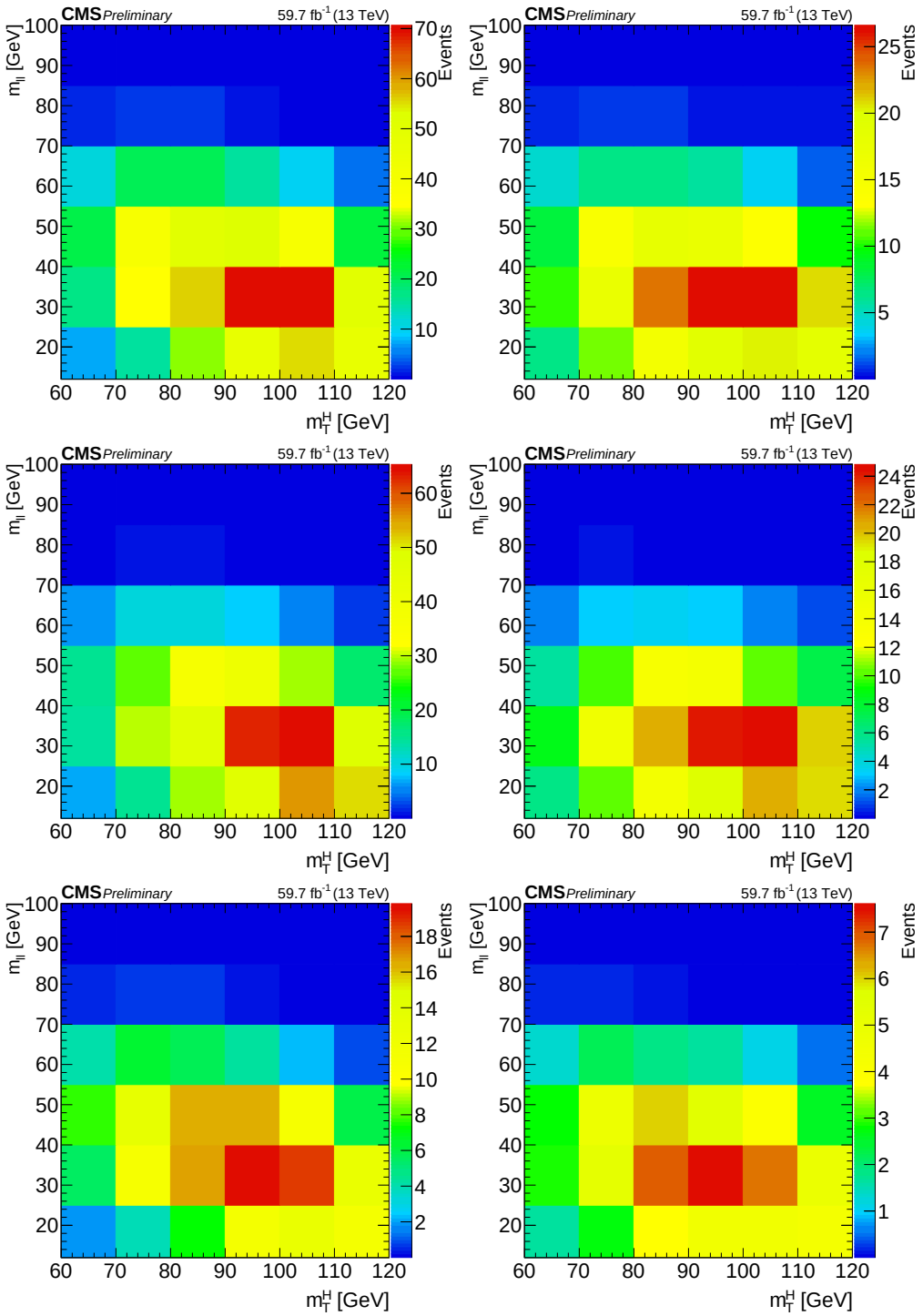

Figure 5.12: Two-dimensional $m_{\ell\ell}$ versus $m_{\mathrm{T}}^{\mathrm{H}}$ distributions of ggF events studied prior to the final binning employed in ggF channels. Displayed are the SM hypothesis (top row), a_2 hypothesis (middle row) and a_3 hypothesis (bottom row), for both the ggF + 0-jet (left) and ggF + 1-jet (right) categories in the 2018 MC simulation.

CHAPTER

SIX

STATISTICAL METHODS AND ANALYSIS UNCERTAINTIES

Data analysis in HEP stands on the principles of the repeatability of experiments. Such an ensemble of experimental results is perfectly suitable for statistical interpretation. One of the fundamental analysis tasks is the process of parameter estimation from distributions sensitive to the studied effects, also called "fitting". So far, we were using this term in a rather intuitive manner while we dedicate this chapter to a more detailed discussion.

The aim of this analysis is to provide an estimation of the true AC parameters, generally referred to as Parameters Of Interest (POI), and the corresponding estimation of their uncertainties expressed in terms of confidence intervals. The typical fitting method employed for POI estimation is the so called Maximum-Likelihood Method (MLM) and was also adopted in our case. In particular, we measure the expected and observed Profile-Likelihood (PL) scans as a function of possible AC values (translated into fractional cross-section). This method is briefly introduced in Section 6.1, while a detailed explanation can be found in [142] or any other HEP statistics textbook.

In the MLM, both signal and background events are entering the fitting procedure. Related statistical uncertainties driven by the limited number of events and systematic uncertainties are taken into account in the form of the so called Nuisance Parameters (NPs), which also enter the maximum-likelihood fit. The particular incorporation of statistical uncertainties, sources of systematic uncertainties, and the correlation scheme among analysis channels and datasets are discussed in Section 6.2. Additional focus is put on determining the parameters' impact on the analysis results.

Note that whenever we speak of observed results in this text, we refer to the fit of the estimated signal and background contribution to the actual data. On the other hand, to resolve any possible fitting issues appearing prior to the final fit on data, we perform the fit to the MC-generated "toy" datasets, also known as expected or Asimov fit on the pseudo-dataset [143].

6.1 Maximum-Likelihood Method

Let us assume we have measured N statistically independent quantities $\vec{x} = (x_1, ..., x_N)$, e.g. multidimensional kinematic discriminants (MKD) introduced in the previous chapter. Also, suppose that the measured quantities x_i, where $i = 1...N$, follow the probability

density $f(x_i;\vec{\Theta})$ with $\vec{\Theta} = \{\mu_{\text{ggF}}, \mu_{\text{EWK}}, f_{a_i}, \vec{\theta}\}$ summarily denoting the set of M parameters. Out of these, it is only f_{a_i} that are considered to be POI. In this analysis, signal strength modifiers μ_{ggF} and μ_{EWK} are profiled in the fit and are effectively treated as NPs, i.e. they are left freely floating in the fit. The rest of the NPs representing uncertainties is dubbed with the symbol $\vec{\theta}$ and will be discussed in Section 6.2. The joint probability density for all measured values $\vec{x}$ is then given by the likelihood function

$$\mathcal{L}(\vec{x};\vec{\Theta}) = \prod_{i=1}^{N} f(x_i;\vec{\Theta}). \tag{6.1}$$

The task of finding the maximum-likelihood estimate of the parameters $\vec{\Theta}$, denoted as $\hat{\vec{\Theta}}$, is now translated into the task of finding the global maximum of $\mathcal{L}(\vec{x};\vec{\Theta})$. In practice, it is even more convenient to work with the negative logarithm of the likelihood function,

$$-\ln\mathcal{L}(\vec{x};\vec{\Theta}) = -\sum_{i=1}^{N} \ln f(x_i;\vec{\Theta}), \tag{6.2}$$

as the process of minimization of the sum is algorithmically easier than the process of maximization of the product. The general problem of locating the global extremum in multiparameter space will be left for the expert literature [144]. In this analysis, we employ a statistical tool developed by the CMS Collaboration that is based on the principles of the MINUIT program [145].

So far we have narrowed down our description of the likelihood function to the most general case. In HEP, however, it is convenient and computationally efficient to perform all fits on binned datasets (i.e. histogram templates). Assuming that the number of events N is large (but finite), the probability density for the bins is given by a multinomial distribution. In that regard, the likelihood function can be written in the form

$$\mathcal{L}(\vec{n};\vec{\Theta}) = N! \prod_{i=1}^{B} \frac{\mathcal{P}_i(\vec{\Theta})^{n_i}}{n_i!}, \tag{6.3}$$

where B stands for the total number of bins in each analysis channel, and n_i (an element of $\vec{n}$) represents the number of observed data entries in bin i. Alternatively, n_i can represent the number of pseudo-data entries in the case of performing an Asimov fit. The quantity $\mathcal{P}_i$ stands for the expected probability for an event to fall into bin i and is related to the probability density via the relation

$$\mathcal{P}_i(\vec{\Theta}) = \int_{x_i^{\text{low}}}^{x_i^{\text{up}}} f(x_i;\vec{\Theta})\, dx_i \tag{6.4}$$

where x_i^{low} and x_i^{up} are the corresponding bin thresholds.

Another key feature of HEP experiments is their repeatability under identical conditions. When repeating a measurement, the observed rate $N = \sum_i n_i$ will fluctuate around its expected (true) value v according to the Poisson distribution (i.e. N itself is considered a Poisson-random variable). This fact can be incorporated into the likelihood function as a multiplicative term, i.e. we assume the extended likelihood function

$$\mathcal{L}(\vec{n}; v, \vec{\Theta}) = e^{-v}\frac{v^N}{N!} N! \prod_{i=1}^{B} \frac{\mathcal{P}_i(\vec{\Theta})^{n_i}}{n_i!} = \prod_{i=1}^{B} e^{-v_i}\frac{v_i^{n_i}}{n_i!}, \tag{6.5}$$

where $v = \sum v_i$ is the sum of expected events in each bin, such as we can write $v_i = \mathcal{P}_i v$. In the most simplified case, we could assume that $v_i = \mu s_i + b_i$, where μ is a signal strength modifier, and s_i (b_i) are numbers of expected signal (background) events in bin i. In our case, the physical model is more complicated and involves interference contributions according to the formulae in Equation 4.23 (AC scheme) and Equation 4.24 (SMEFT scheme). In conclusion, according to the right-hand side of formula 6.5, the extended likelihood function determines a product of Poisson probabilities of observing n_i events in bin i given that v_i events were expected.

Finally, we should not forget that $s_i(\vec{\theta})$, $b_i(\vec{\theta})$ or generally $v_i(\vec{\theta})$ all depend on the Nuisance Parameter (NP) values. As such, we further constrain the likelihood function by a multiplicative term $p(\tilde{\vec{\theta}}, \vec{\theta})$,

$$\mathcal{L}(\vec{n}; v, \vec{\Theta}) = p(\tilde{\vec{\theta}}, \vec{\theta}) \prod_{i=1}^{B} e^{-v_i} \frac{v_i^{n_i}}{n_i!}, \tag{6.6}$$

where p represents the Probability Density Function (PDF*) of the considered uncertainties, assuming some $\tilde{\vec{\theta}}$ parameter related to the best estimate of the nuisance (e.g. mean, median, etc.) and sometimes other shape-describing parameters such as PDF* width, which are both considered to be known. This way we can reflect our belief of what the true values of nuisance parameters $\vec{\theta}$ could be and how they are affecting the final results. Note that the particular form of the probability density p depends on whether the NP affects only signal and background normalisation or it introduces shape effects on reconstructed observables. In the former case, p is represented by a log-normal PDF*, while in the latter case, a Gaussian PDF* is used.

The recipe for constructing the likelihood function described in this section is part of the statistical procedure developed by the CMS and ATLAS Collaborations in the context of the LHC Higgs Combination Group [146]. In the following subsection, we will yet introduce a particular implementation of the MLM targeting the estimation of constraints on AC in the PL scans.

6.1.1 Profile-Likelihood scans

More than 1400 NPs enter the parameter phase space considered for the final fit over all channels in this analysis and all three datasets (2016-2018). Such a level of complexity is typical for HEP studies and is usually addressed by employing the Profile-Likelihood (PL) ratio that allows focusing on only a few POI, while the rest of the parameters are profiled (i.e. left floating in the fit). Assuming the fractional cross-sections f_{a_i} are representing all the POI in the model, the PL ratio is defined as

$$\lambda(f_{a_i}) = \frac{\mathcal{L}(\vec{x}; f_{a_i}, \hat{\tilde{\vec{\theta}}}(f_{a_i}))}{\mathcal{L}(\vec{x}; \hat{f_{a_i}}, \hat{\tilde{\vec{\theta}}})}. \tag{6.7}$$

In the numerator, the $\vec{\theta}$ parameters were fitted to their maximum-likelihood estimate $\hat{\tilde{\vec{\theta}}}$ as a function of f_{a_i} (i.e. $\hat{\tilde{\vec{\theta}}}$ is a value of $\vec{\theta}$ maximizing $\mathcal{L}$ for a given value of f_{a_i}). In the denominator, on the other hand, f_{a_i} was also estimated by the $\hat{f_{a_i}}$ value, and together with $\hat{\tilde{\vec{\theta}}}$, these values determine the global maximum of the likelihood function $\mathcal{L}$.

Furthermore, it can be shown that in the asymptotic limit of large samples where the likelihood function approaches a Gaussian distribution, the relation $-2\ln\lambda(f_{a_i})$ follows

the χ^2 distribution. As such, the asymmetric uncertainties $[f_{a_i}^{\text{down}}, f_{a_i}^{\text{up}}]$ on the best fit value $\hat{f_{a_i}}$ (also called confidence interval) can be obtained as f_{a_i} values corresponding to the intersection of the PL ratio with contours defined by the relation

$$-2\ln\lambda(f_{a_i}) = -2\left(\ln\mathcal{L}(\vec{x}; f_{a_i}, \hat{\hat{\vec{\theta}}}(f_{a_i})) - \ln\mathcal{L}(\vec{x}; \hat{f_{a_i}}, \hat{\hat{\vec{\theta}}})\right) \equiv -2\Delta\ln L \stackrel{!}{=} F_{\chi^2}^{-1}(\text{CL}, s), \tag{6.8}$$

where, in the first part of this formula, we have employed well-known logarithm properties and introduced the $\Delta\ln L$ notation appearing in PL scans throughout this text. On the right-hand side of this formula stands the inverse of the cumulative function F_{χ^2} of χ^2 distribution with $s = 1$ degrees of freedom. Traditionally, we would want to extract the 1σ (2σ) confidence interval corresponding to $\approx 68\%$ ($\approx 95\%$) of the Confidence Level (CL), i.e. using 1-dimensional contours defined by $F_{\chi^2}^{-1}(68\%, 1) = 1.00$ and $F_{\chi^2}^{-1}(95\%, 1) = 3.84$.

For the sake of readability of the final results (see the next chapter), we would introduce a set of preliminary PL scans performed at the expected level, i.e. the fit was done on a pseudo-dataset instead of real data. In Figure 6.1, expected PL scans corresponding to the combination of the 0− and 1−jet ggF channels, individually per the 2016-2018 datasets as well as in the Run-2 combination, are presented. The corresponding physical model assumed in the fit is following the AC scheme and Approach 1, i.e. the $a_i^{WW} = a_i^{ZZ}$ relations are assumed, and the SM coupling is considered together with exactly one of the anomalous couplings. The best (expected) maximum likelihood estimate, found as a global minimum of $-2\Delta\ln L$, for the studied $f_{a_i} \in [-1, 1]$ range is found at zero value for each of the AC scenarios accessible in ggF channels. This confirms the validity of the physical model, as the SM-like behavior of the PL scan shape (i.e. only the SM Higgs signal and background contribution) is expected for the fit on the pseudo-dataset. These scans obviously need to be repeated in the fit on real data, so that the observed AC constraints can be extracted. Two horizontal lines are representing the aforementioned contours, and their intersection with the PL scan defines the 1σ and 2σ confidence intervals. A typical PL scan shape is found in an example with the f_{a_3} fractional cross-section (see the middle row in Figure 6.1). Nevertheless, it is obvious that the expected scan in this case is never crossing the 2σ contour, which indicates insufficient sensitivity of fitting ggF channels alone and the necessity of combination with the rest of categories. Traditionally in HEP studies, CL=95% is needed to define regions in the POI range that can be reasonably excluded, otherwise, no conclusive interpretation of the results can be stated. Another important observation is that more statistics generally mean bigger sensitivity to constrain AC, as shown in the combination of Run-2 datasets compared to PL scans performed for individual years.

In the case of PL scans with f_{a_2} and $f_{a_{\Lambda 1}}$ (see the top and bottom rows in Figure 6.1), peculiar scan shapes were found with a sort of "peaks" (local maxima) forming around negative $f_{a_2} = -0.25$ and positive $f_{a_{\Lambda 1}} = 0.5$ fractional cross-section values, respectively. Further studies revealed that the actual shape of the signal distribution is a driving factor determining how the likelihood for a given signal hypothesis will look like. Let us consider an example with the ggF + 0-jetcategory, for which the most significant signal contribution (assuming the SM Higgs boson) comes from ggF production. As such, it is reasonable to assume that the ggF signal shape will also dictate the final shape for the PL scans in this channel. Figure 6.2 shows the ggF distributions of m_{ll} and m_{T}^{H} (variables used for the construction of MKD in ggF channels) for a number of different f_{a_2} and $f_{a_{\Lambda 1}}$ hypotheses. The biggest shape deviation from the SM case was indeed observed for hypotheses corresponding to the local maxima in the PL scan. Taking the mean of m_{ll} or m_{T}^{H} distributions as a function of f_{a_2} and $f_{a_{\Lambda 1}}$ can also be used as an

indicator of f_{a_i} values for which the interference effect plays the most significant role, as demonstrated in Figure 6.3. In constrast, changes in distribution shape for f_{a_3} values are relatively symmetric around zero, which is also reflected in the corresponding PL scans. Similar tests were also performed outside of scope of this thesis for the rest of the analysis channels and employed Kinematic Discriminant (KD) distributions. In each case, tests were leading to the observation that particular distribution shapes (determined by signals with a dominant contribution) also dictate the shape of the PL scans. In the case of the SMEFT scheme and Approach 2, similar local maxima in the PL shapes are observed, however, interference effects forming the MKD distribution shapes were found shifted in one or other direction, thus local maxima in the PL scan shapes are also shifted accordingly. Finally, let us note that the effect of the SM-BSM interference, leading to the peculiar scan shape behavior, is also evident in the event yield distributions shown in Figures 4.15. Within these pictures, for specific f_{a_i} values and a given process, the event yield is noticeably higher compared to the $f_{a_i} = 0$ scenario. This provides an early indication of the range of f_{a_i} values where our sensitivity to interference effects is particularly pronounced, in addition to our earlier discussion on observable distribution shapes.

6.2 Uncertainty sources

Both statistical and systematic uncertainties are modeled as nuisance parameters that are profiled in the maximum-likelihood fit. Any assumptions on uncertainty estimation and employed procedures are discussed in this section. Let us emphasize that the enlisted uncertainties follow the correlation scheme identical to the HWW Legacy analysis (i.e. we are following the recommendations of relevant POG and PAG experts) unless stated otherwise.

6.2.1 Statistical uncertainties

Suppose this analysis is a counting experiment and forget about any shape dependencies for a while. Repeating this experiment N times, we can assume that the uncertainty on the count n_{ij} in each bin i and each signal or background process j is determined by one standard deviation for a Poisson distribution, i.e. the absolute statistical per-bin uncertainty is given by $\text{err}_{ij} = \pm\sqrt{n_{ij}}$. For N becoming a large number, due to the principles of the central limit theorem, the uncertainty in each bin can be modeled by a Gaussian distribution. In particular, one would need to compute one Gaussian distribution per each process contributing to the bin and take $\pm 1\sigma$ variation of the corresponding nominal yield. However, considering all counts are statistically independent, the total uncertainty in the bin is given by a sum of all Gaussian distributions for each process, which itself is a Gaussian-distributed variable. As such, we only define the Poisson-like per-bin uncertainties when $N \leq 10$, while a single Gaussian distribution is used to define uncertainty for each bin in case that $N > 10$.

6.2.2 Systematic uncertainties

Systematic uncertainties which affect the normalization or shape of the fitted template distributions are taken into account. In cases where the shape effect was found negligible, only the change in normalization is assumed. We consider two major groups of systematic uncertainties – those related to imperfections of experimental methods and those

Uncertainty	Effect	Correlation
Luminosity	norm	partially correlated[1]
PU modeling	norm	uncorrelated
Trigger efficiency	shape	uncorrelated
Prefiring weight	shape	uncorrelated 2016-2017
Lepton ID eff.	shape	uncorrelated
Lepton p_{T} scale	norm	uncorrelated
Jet energy scale	norm	partially correlated[2]
Jet energy resolution	norm	uncorrelated 2017-2018
AK8 JMS and JMR	norm	uncorrelated
V-tagging SF	norm	uncorrelated
Unclustered MET energy scale	norm	uncorrelated
Jet PU ID scale	shape	uncorrelated
Fake rate (stat.+syst. origin)	shape	uncorrelated
Fake rate (30% jet composition)	norm	uncorrelated
B-tagging SF	shape	partially correlated[3]

Table 6.1: Correlation scheme for the experimental uncertainties between the 2016-2018 datasets as considered in this analysis. The type of uncertainty effect, either shape or normalization ("norm"), is also specified. Partial correlation corresponds to [1]LUM POG recipe, [2]JetMET POG recipe, [3]BTV POG recipe as referenced in the text.

arising from employed theoretical assumptions. The magnitude of (relative) uncertainty effect on the fitted distributions per individual uncertainty sources is also presented. Note that the quoted numbers reflect the average effect across datasets (unless specified) and were selected to represent relevant (i.e. statistically significant) signal/background contributions depending on affected signal region. A summary of the correlation scheme for experimental and theoretical uncertainties is shown in Table 6.1 and Table 6.2, respectively.

Experimental uncertainties

The following sources of experimental uncertainties and their correlation are considered in the fit:

- The total normalisation uncertainty associated with the measurement of the integrated luminosity for 2016 [147], 2017 [148] and 2018 [149] is 1.2%, 2.3% and 2.5%, respectively. Following the official recommendation of the Luminosity POG (LUM POG), these uncertainties are considered partially correlated among all three datasets, i.e. both uncorrelated and correlated uncertainty sources are considered.

Uncertainty	Effect	Correlation
PDF	norm	correlated
QCD higher order	norm (background)	correlated
	shape (signal)	
Parton shower	shape	correlated
Underlying event	norm	uncorrelated 2016
		correlated 2017-2018
WW resummation	shape	uncorrelated (also per cat.)
Single top/$t\bar{t}$ composition	shape	correlated
$V\gamma$, $V\gamma^*$ and VZ cross-sections	norm	correlated

Table 6.2: Correlation scheme for the theoretical uncertainties between the 2016-2018 datasets as considered in this analysis. The type of uncertainty effect, either shape or normalization ("norm"), is also specified.

Another source of the uncertainty arising from the "beam-beam" effects (for details see [148]) is considered correlated among the 2017 and 2018 datasets.

- The normalisation uncertainties originating in the PU modeling were considered for the main simulated background processes (DY, WW, top) as well as ggF and VBF SM/BSM signals and were determined by varying the total inelastic pp cross-section (69.2 mb [150, 151]) within the assigned 5% uncertainty. This uncertainty comes from the measurement of the inelastic pp cross-section and the discrepancy between the simulation and data efficiency of the primary vertex reconstruction.

- Uncertainties on the prefiring MC weights discussed in Section 4.2.3 were considered by varying the prefire rate by a capped maximum of 20%, which results in the shape effect on background and signal contribution of up to only 1%. Uncertainties are uncorrelated between the 2016 and 2017 datasets.

- The lepton trigger efficiency corrections are measured using the Tag&Probe method in a kinematic region rich in Z boson events. Corresponding trigger systematic uncertainty is then determined by variation of the tag lepton selection criteria and of the Z mass window as a function of lepton p_{T} and η. As such, this uncertainty source is affecting both the normalisation and shape of the signal and background distributions and is treated uncorrelated among datasets. The total normalisation uncertainty is found below 1%.

- The lepton identification and isolation efficiency is measured using the Tag&Probe method, and the corresponding systematic uncertainty is derived in the same fashion. In this case, the average shape effect is found about 1% for electrons and 2% for muons.

- Different detector subsystems introduce various systematic uncertainties related to lepton momentum and resolution measurements. Their magnitude, however, is

largely overcome by uncertainties arising from limited samples used for the estimation of data-to-MC lepton momentum scale factors. The magnitude of their effect on the background and signal yields ranges between $0.6 - 1.0\%$ for electron scale factors and is about 0.2% for muon scale factors. This effect is also reflected in the re-computation of affected kinematic observables and reconstruction of the PUPPI MET object. Lepton momentum scale uncertainties are treated as uncorrelated among the three datasets.

- The jet energy corrections (JEC) were introduced in Section 3.2.4. Different detector effects were assumed by the JetMET POG to model uncertainty on the corresponding jet energy scales by implementing eleven independent nuisance parameters, six of which are correlated among the three datasets. Their effects vary in the range of $1 - 10\%$, depending mainly on the jet multiplicity in the analysis phase space, and were translated to the computation of the PUPPI MET object and relevant kinematic observables. Another source of uncertainty arises from the jet energy resolution (JER) smearing applied to simulated samples, however, it was not translated to the PUPPI MET reconstruction given the incompatibility of algorithms used to model PUPPI MET and jet objects assumed in this analysis. The JER uncertainty effect varies between $1-5\%$ depending on the jet multiplicity and is uncorrelated between the 2017 and 2018 datasets. JEC and JER related uncertainties were considered for both AK4 and AK8 jet collections and were found to have a negligible effect on shape of the fitted templates. As such, only the normalisation effect was considered.

- Additionally, the Jet Mass Scale (JMS), Jet Mass Resolution (JMR) and V-tagging corrections were applied, and corresponding uncertainties considered for AK8 jets requested in VH Boosted channel. Their normalisation effect on the background and signal contribution in VH Boosted channel was found to be less than 8% and was treated as uncorrelated among the three datasets.

- Both the normalization and shape of the signal and background templates are affected by the jet PU ID uncertainty. The effect is smaller than 1%.

- The effect of the unclustered energy scale on the reconstruction of PUPPI MET, i.e. corrections due to PF candidates entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation other than jets and leptons, is considered. The corresponding normalisation uncertainty is estimated by varying PF candidate momenta and was found between $1 - 10\%$. This uncertainty is treated as uncorrelated among the datasets.

- The uncertainty related to the b-tagging efficiency is considered and was modeled by seventeen nuisance parameters, five of which can be traced to the theoretical origin, therefore, they are treated as correlated among the datasets. The remaining parameters reflect the statistical accuracy of the efficiency measurement and are considered as uncorrelated. The typical shape effect is of the order of 1% for most of the signal and background contributions with an exception of top production where b-tagging uncertainty effect can go up to 5%.

- Estimation of the nonprompt background is affected by the limited size of the datasets used for the misidentification rate measurements. It is also affected by the difference in the flavor composition of jets misidentified as leptons between the misidentification rate measurement region (enriched in QCD multijet events) and the signal phase space. Both sources are affecting the shape of the nonprompt

background distribution with an effect ranging between a few percent to about 10%, depending on the signal region, and are treated as nuisance parameters uncorrelated between electrons and muons and among the three datasets. A normalization uncertainty of 30% [130] is assigned to fully cover for any discrepancies with respect to data and is treated as uncorrelated among the datasets.

Theoretical uncertainties

Multiple theoretical uncertainties are considered and are correlated among datasets unless stated otherwise:

- The uncertainties related to the choice of PDF* and α_S were found to have a minor effect on the shape of the distributions. As such, only normalisation effects due to the event acceptance and cross-section are considered. Uncertainties of this kind are not defined for background processes, in which case the normalisation is constrained directly through data in dedicated CRs. In the case of signal processes (both SM and BSM), uncertainties were estimated by the LHC Higgs cross-section working group [32].

- The theoretical uncertainties arising from missing higher order corrections in cross-section computation are considered. Background simulation events are re-weighted to the alternative scenarios corresponding to renormalisation μ_R and factorisation μ_F scales varied by factors of 0.5 or 2, and the envelopes of the varied templates are taken as the one standard deviations. For background processes whose normalization is constrained through data in dedicated CRs, we consider only the shape effect of the uncertainties coming from the missing higher-order corrections. The WW nonresonant background has the uncertainties derived by varying μ_R, μ_F and the resummation scale. SM and BSM signal ggF and VBF processes have the effect of the missing higher order corrections on the overall cross-section decoupled into multiple sources according to the recipe described in [32].

- The PS modeling mainly affects jet multiplicity causing migration of events between categories resulting in corresponding template shape changes. Associated uncertainties are evaluated by re-weighting events with varied PS weights computed by PYTHIA 8.212.

- Uncertainties associated with modeling of the UE are evaluated by varying the respective UE tunes used in the MC generation. As such, the correlation scheme is adapted, and UE uncertainties are only kept correlated in 2017 and 2018 (sharing the same CP5 tune). The UE uncertainty is found to have a minor effect on the template shapes while affecting the normalization by about 1.5%.

- Several uncertainty sources are relevant only to selected background processes. The gluon-gluon induced nonresonant WW production was corrected to NLO accuracy while accounting for a 15% normalisation uncertainty [152]. The relative fraction between single top and $t\bar{t}$ production was corrected with a respective systematic uncertainty on the shape of an order of 8%. Additional process-specific (DY, VZ, Vγ, Vγ^*) uncertainties, related to the yield corrections to account for possible discrepancies between data and simulation, are assigned and are correlated among datasets.

6.2.3 Nuisance parameters and their impact on analysis

The impact of each (statistical and systematic) uncertainty on the fractional cross-sections f_{a_i}, defined as a shift in the best f_{a_i} estimate, $\Delta\hat{f_{a_i}}$, as it is induced by a particular NP varied up (red) and down (blue) by a standard deviation of 1σ, is shown in Figures 6.4 – 6.5, individually for the assumed AC hypothesis. For the brevity of this text, only the NPs' impacts corresponding to Approach 1 scenario are shown, nevertheless, similar studies were performed for Approach 2. Uncertainty sources are ordered from the most "impactful", i.e. showing maximal $\Delta\hat{f_{a_i}}$ shift, while we only display the first 30 NPs with the highest rank. Note that the naming convention used in the impact plots is somewhat cryptic, nevertheless, the main message arising from the shown quantities should be straightforward without an explicit knowledge of all abbreviations in use. For your convenience, however, we mention a few rules that might help with their identification:

- μ_{ggF}, dubbed as "muF", and μ_{EWK}, dubbed as "muV", are kept floating in the fit (the term "unconstrained" which is different from "not constrained" is used in impact plots),
- NPs consisting of "Topnorm", "DYttnorm" and "WWnorm" tags (grey font) stand for the rate parameters assuring proper background normalisation using dedicated Top, DY and WW CRs and are also left floating in the fit. 4.15
- NPs named as "prop_channel_dataset_bin(_process)" represent the Gaussian (black) or Poisson (green) statistical uncertainties in a given analysis channel, dataset, and particular bin. In the case of uncertainties driven by the Poisson distribution, uncertainty is estimated per individual processes contributing to that bin.
- "CMS_*"-tag is usually denoting the experimental NPs specific for the CMS detector; e.g. "CMS_eff_e_2018_hww2l2v_13TeV_of2j_vbf_hpip" stands for the systematic uncertainty on electron trigger efficiency in VBF (a_2) signal category, $\mathcal{D}_{\text{int}} > 0.4$ subcategory and 2018 dataset.
- NPs without "CMS_*"-tag are standing for theoretical uncertainties which might find their application beyond the CMS analysis framework; e.g. "PS_FSR" denotes uncertainty on PS modeling, in particular, the Final State Radiation (FSR) origin.

One can notice that both signal strength modifiers belong among the most impactful parameters for each AC scenario. As such, the corresponding interplay between the best-fit f_{a_i} value and postfit μ_{ggF} and μ_{EWK} values might be expected. The general purpose of the "impact plots" shown here is twofold:

1. to asses correlation between the systematic uncertainty source and the $\hat{f_{a_i}}$ best-fit estimate as can be concluded from the $\Delta\hat{f_{a_i}}$ values,
2. to make sure that individual NPs (with an exception of signal strength modifiers and the main background normalisation from dedicated CRs) were not (over)constrained in the fit on data.

The latter point can be studied from the left panel displaying the so called "pull" quantity, $\hat{p} = (\hat{\theta} - \theta_0)/\Delta\theta$, where $\hat{\theta}$ and θ_0 are the nominal postfit an prefit nuisance parameter values, respectively, and $\Delta\theta$ represents the assumed prefit uncertainty on θ_0. No high-ranked systematic uncertainty was found to be significantly pulled, i.e. $\hat{p}$ stays in the

$[-1, 1]$ region. Note that asymmetric error bars assigned to each $\hat{p}$ value, defined as the postfit uncertainty divided by the prefit uncertainty, should be significantly smaller than ± 1, for the systematic uncertainty to be constrained. This was checked to be false for most of the relevant high-ranked nuisance parameters. Several seldom exceptions, such as b-tagging uncertainties, were found to be constrained in data, nevertheless, it was verified in the Asimov fit that the same level of constraints is expected.

Let us mention one of the typical pathological situations we have encountered in the early "unblinding" stage, which might also occur for other analysers and the way we have solved it. It is possible to locate NPs that are severely under-constrained in the fit on data, i.e. errors assigned to $\hat{p}$ are much bigger than ± 1. In such a case, one can observe that deliberately fitting the "misbehaving" NP, for a fixed value of f_{a_i}, reveals a double minimum in the profile-likelihood ratio, e.g. see Figure 6.6 (left). Further investigation has shown that for particular signal regions, several low-statistic background contributions (e.g. MC DY samples in VBF SR) might induce unrealistically big fluctuations in the normalisation effect for the considered NP. In that regard, it is reasonable to assume that the uncertainty effect on the signal distribution, which is dominant in these SRs (e.g. VBF signal in VBF SR), would be a better estimate for the low-statistic background contribution. In these rare cases, we physically exchange the up and down variations on the low-statistic background normalisation for the estimated variations on the well-behaved signal normalisation. Indeed, the secondary local mimimum disappears after this procedure is applied, as confirmed in Figure 6.6 (right), and the NP is not under-constrained anymore. It is important to say that this somewhat technical procedure is justifiable only due to the negligible assumed background contribution (e.g. MC DY events in any of the SR).

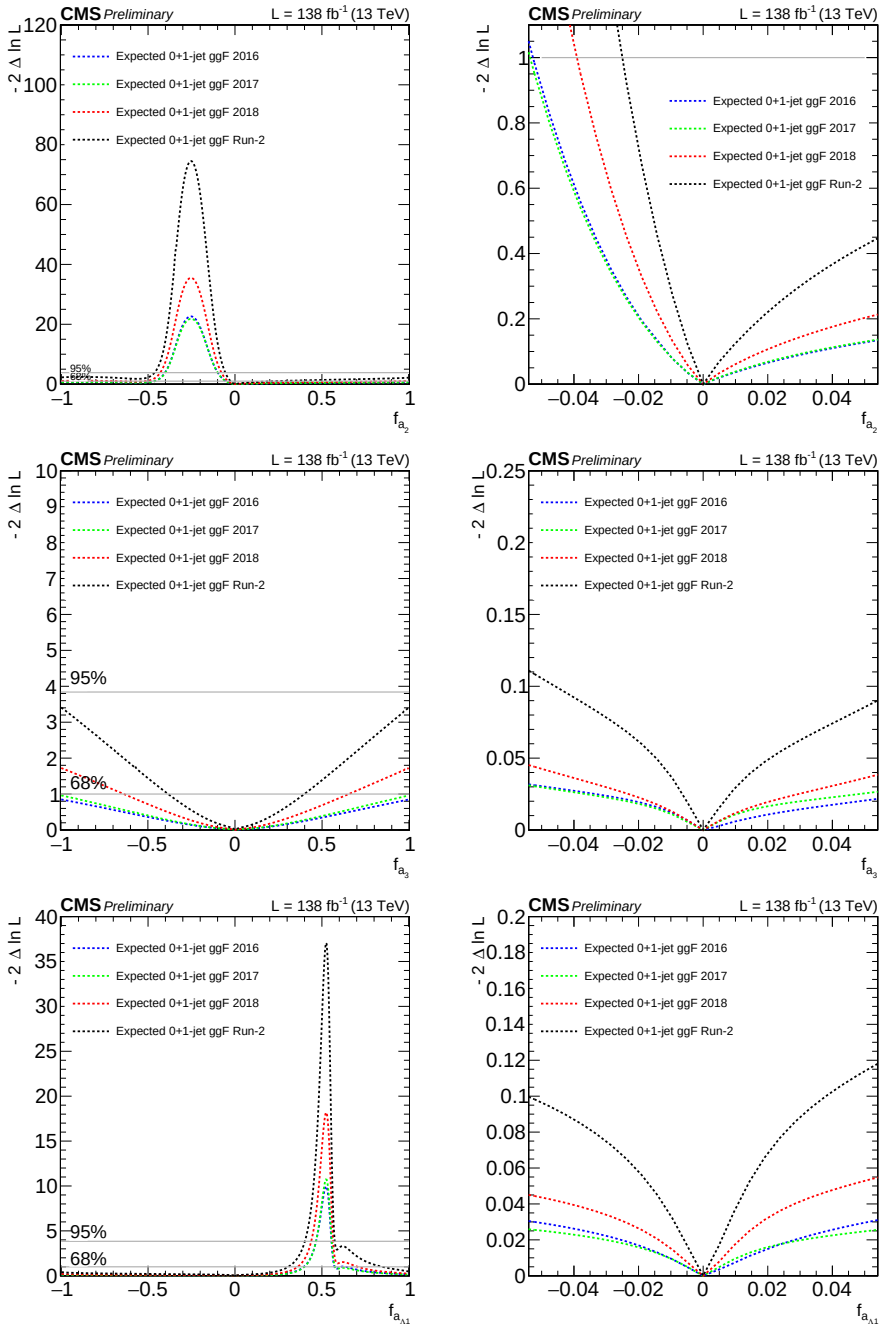

Figure 6.1: Expected profile-likelihood (PL) scans for f_{a_2} (top row), f_{a_3} (middle row) and $f_{a_{\Lambda 1}}$ (bottom row) are presented for the combination of 0− and 1−jet ggF channels in 2016 (blue), 2017 (green), 2018 (red) and in the full Run-2 combination (black). Horizontal lines (grey) represent the 1σ and 2σ contours corresponding to CL = 68% and CL = 95%, respectively. The second column displays the same PL scans but focused on a small region around the SM (zero) value of fractional cross-section.

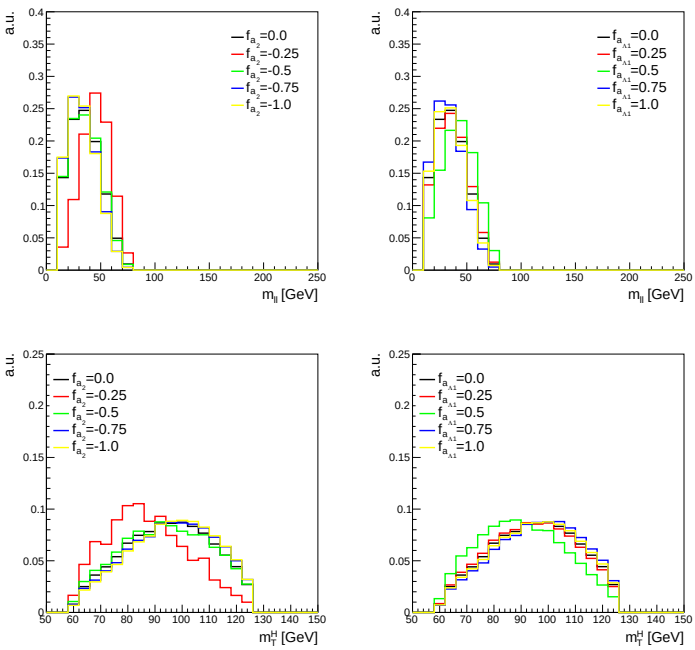

Figure 6.2: Shape of the ggF m_{ll} (top row) and $m_{\mathrm{T}}^{\mathrm{H}}$ (bottom row) distributions for various values of f_{a_2} (left column) and $f_{a_{\Lambda 1}}$ (right column) in the ggF + 0-jet SR.

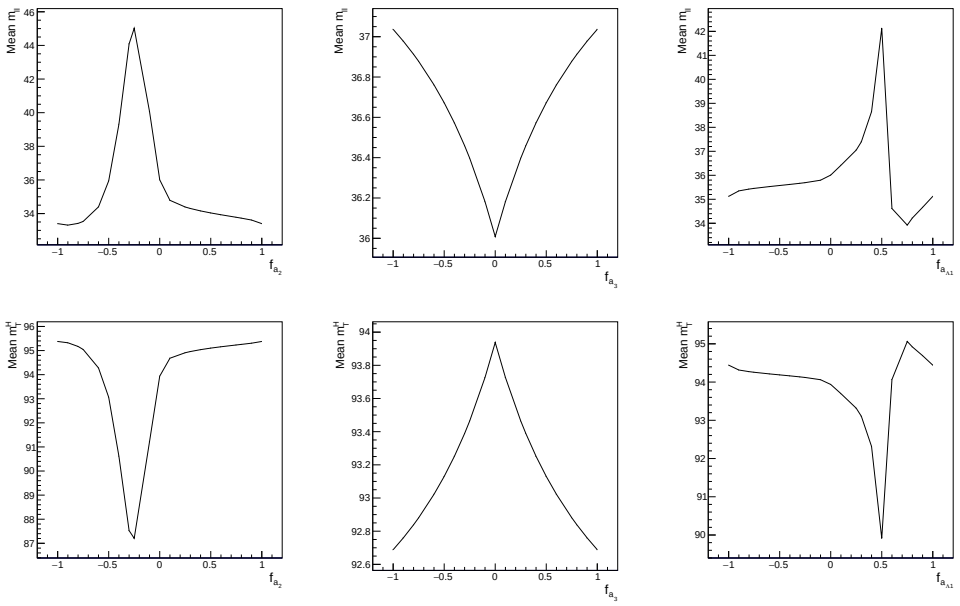

Figure 6.3: Mean of the ggF m_{ll} (top row) and $m_{\mathrm{T}}^{\mathrm{H}}$ (bottom row) distributions as a function of f_{a_2} (left), f_{a_3} (middle) and $f_{a_{\Lambda 1}}$ (right) in the ggF + 0-jet SR.

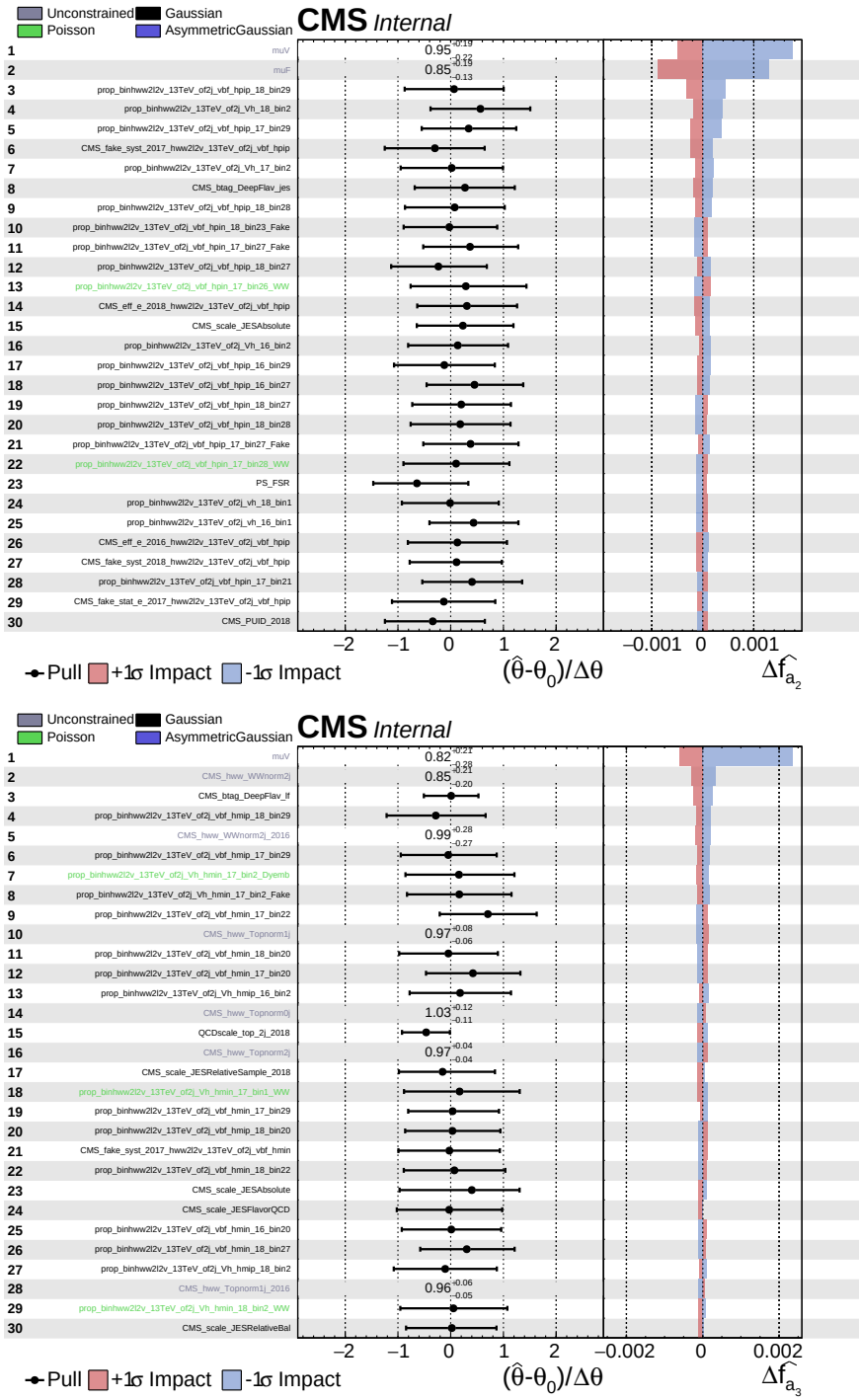

Figure 6.4: The first 30 high-ranked NPs listed in decreasing order reflecting on the observed impact on the best f_{a_2} (top) and f_{a_3} (bottom) estimates, respectively. The left panel is showing the observed pull from the prefit parameter value for each of the NPs and the corresponding uncertainty.

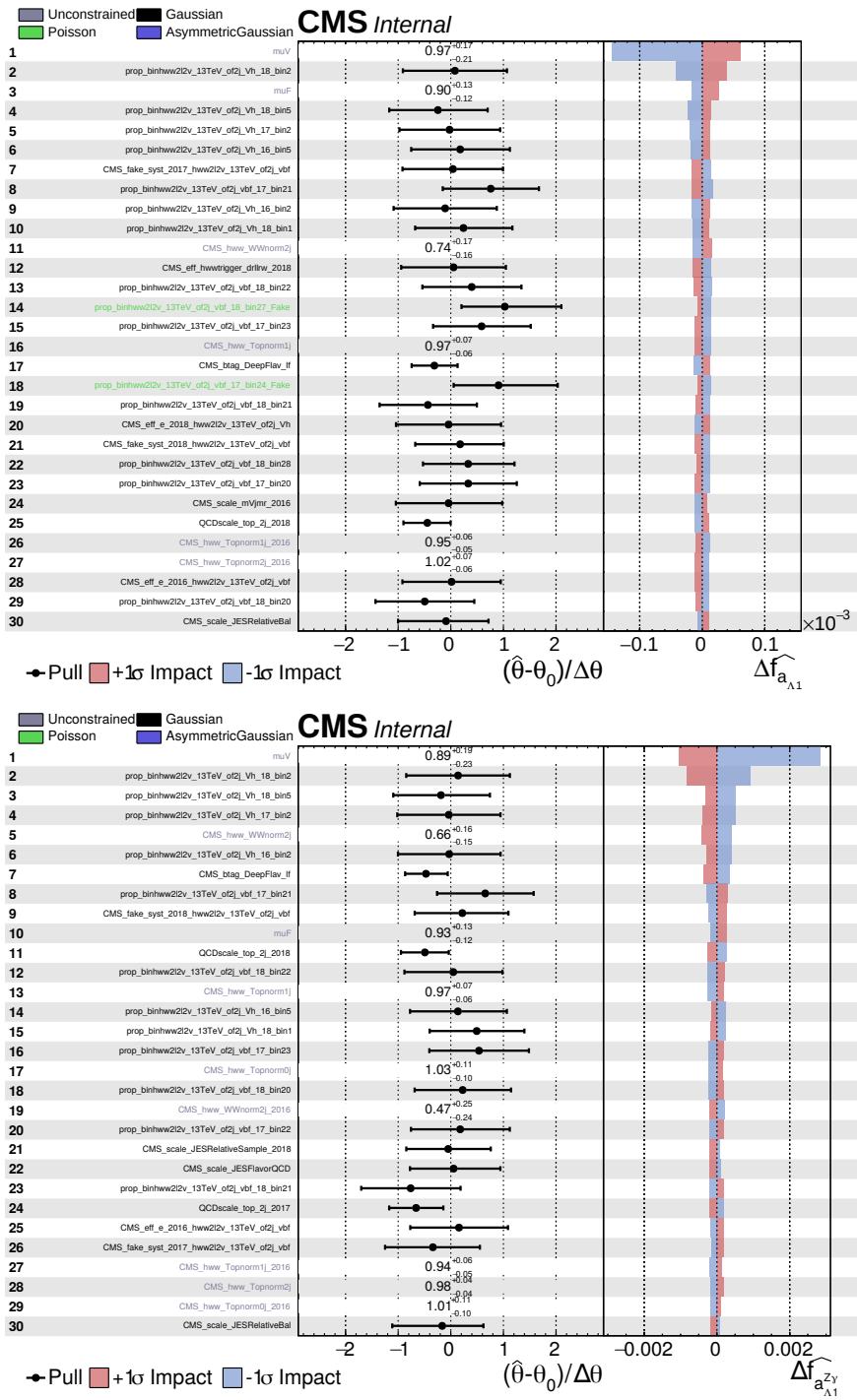

Figure 6.5: The first 30 high-ranked NPs listed in decreasing order reflecting on the observed impact on the best $f_{a_{\Lambda 1}}$ (top) and $f_{a^{Z\gamma}_{\Lambda 1}}$ (bottom) estimates, respectively. The left panel is showing the observed pull from the prefit parameter value for each of the NPs and the corresponding uncertainty.

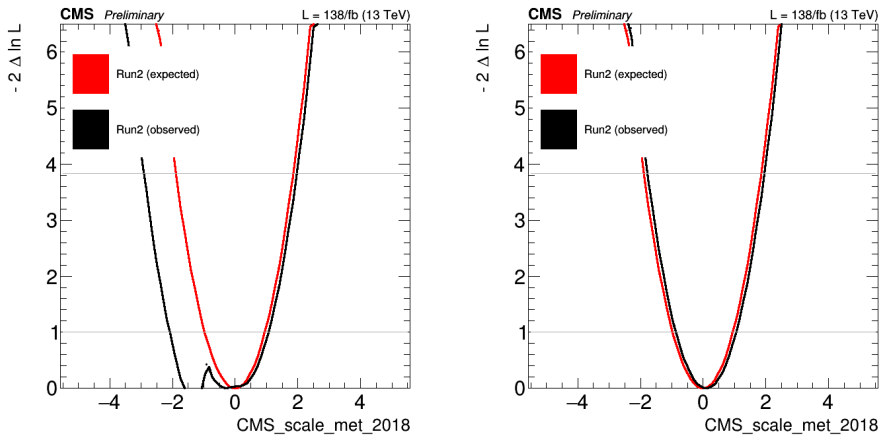

Figure 6.6: Profile-likelihood scan for the NP corresponding to the unclustered energy scale on the reconstruction of PUPPI MET in 2018. Fluctuations in the normalisation effect induced by corresponding uncertainty for low-statistic background samples in the fitted SR are reflected by a secondary local minimum forming in the left scan. This issue is solved (right scan) by assigning the realistic normalisation effect as described in the text. During the fit, f_{a_i} was kept fixed to the SM value.

CHAPTER

SEVEN

RESULTS

This chapter presents the results of an analysis studying the anomalous Higgs boson couplings (AC) from its production and decay, using the $H \to WW^*$ decay channel. Postfit distributions of the multidimensional kinematic discriminants (MKD) utilized for signal extraction from the data are presented in Section 7.1. Both the expected and observed profile-likelihood (PL) scans of the studied AC fractional cross-section for the combination of all analysis channels are shown in Section 7.2, and the corresponding 1σ and 2σ confidence intervals are subsequently extracted and reported in Section 7.3. The results presented here are considered an integral part of a supporting analysis [38] currently in preparation for publication, for which I assert shared main authorship.

7.1 Postfit distributions

For the sake of completeness, postfit MKD distributions are shown independently for Approach 1 (Figures 7.2 – 7.6) and Approach 2 (Figures 7.7 – 7.8), as defined in Section 5.2. In both approaches, the background is divided into the following groups of related processes:

- DY events,
- nonprompt background,
- nonresonant WW background,
- single top and $t\bar{t}$ production,
- and multiboson production.

These groups, along with the total signal contribution (SM+BSM+interference), dubbed as "Higgs boson" in the plots, are shown individually for each of the considered SRs (ggF, VBF and VH channels). As usual, the SM and one of the ACs were considered at the same time in the fit for Approach 1 (in the AC scheme). In the case of Approach 2 (in the SMEFT scheme), we considered the SM coupling together with a_2, a_3 and $a_{\Lambda 1}$ couplings in the fit. For visualisation purposes in Approach 2, the Higgs boson signal contribution consists of the SM signal, pure a_2 prediction and the interference of both. Similar distributions were plotted for the BSM contributions related to the remaining number of couplings. The same conclusions as with the a_2 example were confirmed and

are discussed below. Similarly, for ggF categories in Approach 1, we have arbitrarily selected the a_3 contribution to be displayed. This is due to the similarity of the final postfit distributions when assuming the contribution from the remaining coupling models. It is worth noting that both ggF distributions in Approach 1 and Approach 2 are almost identical, due to the same binning scheme employed and the effective suppression of any anomalous signal in the data fit. The assumed bin thresholds in the case of the distributions in Approach 1 are shown explicitly, however, this was not possible in the case of Approach 2 in order to retain the readability of the plots.

Upon close inspection, the postfit values of the total signal yield as well as the total signal shape were found to closely mimic those of the SM contribution. This resemblance is easily observed when comparing with the prefit distributionss (e.g. in Approach 1) presented in Figures 5.5 to 5.9. This suggests that any anomalous signal contribution (in both approaches) was strongly suppressed in the fit to the data. Any remaining discrepancies are addressed by the corresponding statistical and systematic uncertainties. In other words, the overall good agreement of the estimated signal and background contributions with the data was confirmed in accordance with the SM, and no obvious data excess beyond the scope of SM physics was observed.

7.2 Profile-likelihood scans

The purpose of the PL scans presented in this section is to visually define exclusion regions for the signal hypothesis concerning the corresponding values of the fractional cross-section, f_{a_i}, assumed in the scan. As explained in the previous chapter, both CL = 68% and CL = 95% confidence intervals (uncertainties on the best-fit value of f_{a_i}) can be extracted from the PL scans. The PL scans were conducted for the full combination of the considered signal and control regions (defined in Section 5.1), as well as for the entire combination of Run-2 datasets (2016-2018).

In Figure 7.9, the PL scans for fractional cross-section values, f_{a_i}, corresponding to the anomalous couplings a_2, a_3, $a_{\Lambda 1}$ and $a_{\Lambda 1}^{Z\gamma}$ in the AC scheme and Approach 1 are shown. The PL scans for the SMEFT scheme and Approach 2 are shown in Figure 7.10, individually for f_{a_2}, f_{a_3} and $f_{a_{\Lambda 1}}$. In these scans, the two remaining f_{a_i} fractional cross-sections considered in the SMEFT model were either allowed to float in the fit (dubbed as "float others") or were alternatively fixed at the SM (zero) value in the fit (dubbed as "fixed others"). Both the PL ratio for the fit on data (observed) and the Asimov fit (expected) are plotted.

In all cases, only the relatively small region around $f_{a_i} = 0$, i.e. the SM case, cannot be excluded. In other words, all the presented results are found to be in agreement with the SM expectations. Notably, significant interference effects around the $f_{a_2} = -0.25$ and $f_{a_{\Lambda 1}} = 0.5$ in Approach 1 as well as $f_{a_2} = -0.85$ and $f_{a_{\Lambda 1}} = 0.85$ in Approach 2 were observed and are discussed in the preceding chapter. It is worth mentioning that the observed $-2\Delta \ln L$ values for the f_{a_3} and $f_{a_{\Lambda 1}^{Z\gamma}}$ scans in Approach 1, and the f_{a_3} scan in Approach 2, are significantly smaller (over the large span of the fitted range) than that of the expected results. This can largely be attributed to the profiled value of μ_{EWK} corresponding to the best-fit of f_{a_i}. The lowest $\mu_{\text{EWK}} \approx 0.82$ value corresponds to the f_{a_3} fit in Approach 1, while the largest $\mu_{\text{EWK}} \approx 0.97$ value was found in the fit for $f_{a_{\Lambda 1}}$ in the same approach. In each case, the uncertainty on μ_{EWK} itselft is approximately 20%, demonstrating that all fitted values are consistent with both the SM and with each other.

The correlations (anti-correlations) between individual f_{a_i} parameters and signal

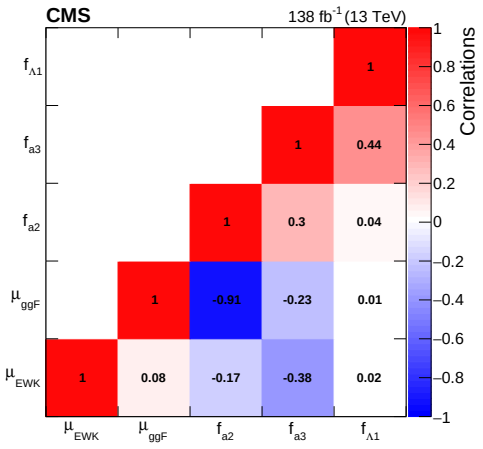

Figure 7.1: The observed correlation coefficients between the HVV anomalous coupling cross-section fractions and signal strength modifiers as obtained within the Approach 2.

strength modifiers as observed in the fit within the Approach 2 ("float others") are documented in Figure 7.1. The quoted correlation factors can be interpreted as a measure of how independently the individual parameters are affecting the final fit. In the example with a strong anti-correlation factor of -0.38 in the last row of the matrix, the central value of the μ_{EWK} affects the cenral value of the f_{a_3} in an anti-correlated way. In other words, when the f_{a_3} value is increased (decreased) in the fit on data, the corresponding value of μ_{EWK} is decreased (increased), and vice versa. Similar effects are observed also for the remaining pairs of studied parameters. Finally, let us note it is the interplay of all assumed parameters that determines a signal that best matches the data.

7.3 Constraints on the Higgs boson anomalous couplings

Constraints on the Higgs boson anomalous couplings, as expressed by means of the corresponding fractional cross-sections, are extracted from the PL scans outlined in the previous section. Specifically, both expected and observed best-fit f_{a_i} estimates are retrieved, together with 68% and 95% CL intervals. These outcomes are reported in Table 7.1 individually for Approach 1, Approach 2 with other f_{a_i} parameters floating, and Approach 2 with other f_{a_i} parameters fixed to the SM value. In general, all results were found to be in agreement with the SM expectations. The detailed discussion of the obtained results in the context of the previous or parallel analyses constraining the ACs assuming the alternative Higgs boson decay channels is presented in the following chapter.

Analysis	f_{a_i}	Observed (x10^{-3})		Expected (x10^{-3})	
		Best-fit with CL = 68%	CL = 95%	Best-fit with CL = 68%	CL = 95%
HVV Approach 1	f_{a_3}	0.9 [-2.7, 4.1]	[-553, 561]	0.0 [-0.7, 0.7]	[-2.8, 2.9]
	f_{a_2}	0.5 [-0.8, 3.5]	[-5.7, 12]	0.0 [-1.4, 1.3]	[-5.2, 6.1]
	$f_{a_{\Lambda 1}}$	-0.2 [-0.5, 0.0]	[-1.4, 0.7]	0.0 [-0.2, 0.5]	[-0.6,1.4]
	$f_{a_{\Lambda 1}}^{Z\gamma}$	3.0 [-11, 9.1]	[-55, 42]	0.0 [-5.0, 3.8]	[-14, 11]
HVV Approach 2 (fixed others)	f_{a_3}	0.84 [-0.83, 3.5]	[-7.6, 58.8]	0.0 [-0.8,1.1]	[-3.4,4.3]
	f_{a_2}	38 [-112.2, 129.3]	[-376.6, 430.0] $\cup$[$-$989.2, $-$826.3]	0.0 [-30.9,37.5]	[-126.1,136.8]
	$f_{a_{\Lambda 1}}$	-0.15 [-1.21, 0.16]	[-19.5, 118.5] $\cup$[909.9, 964.1]	0.0 [-0.4,0.4]	[-1.7,18.9]
HVV Approach 2 (float others)	f_{a_3}	0.34 [-0.69, 3.4]	[-8.0, 361.5]	0.0 [-1.0,1.2]	[-4.3,5.3]
	f_{a_2}	-1.0 [-104.1, 139.9]	[-986.4, 981.2]	0.0 [-31.1,39.8]	[-127.5,148.7]
	$f_{a_{\Lambda 1}}$	-0.1 [-1.08, 3.78] $\cup$[7.2, 20.7]	[-994.8, 993.9]	0.0 [-0.4,0.90]	[-1.9,21.4]

Table 7.1: Summary of constraints on the anomalous HVV coupling parameters with the best-fit values and allowed 68% and 95% CL (in square brackets) intervals. For Approach 1, each f_{a_i} is studied independently. For Approach 2, each f_{a_i} is shown separately with the other two fractional cross-sections either fixed to zero or left floating in the fit. In each case, the signal strength modifiers are treated as free parameters.

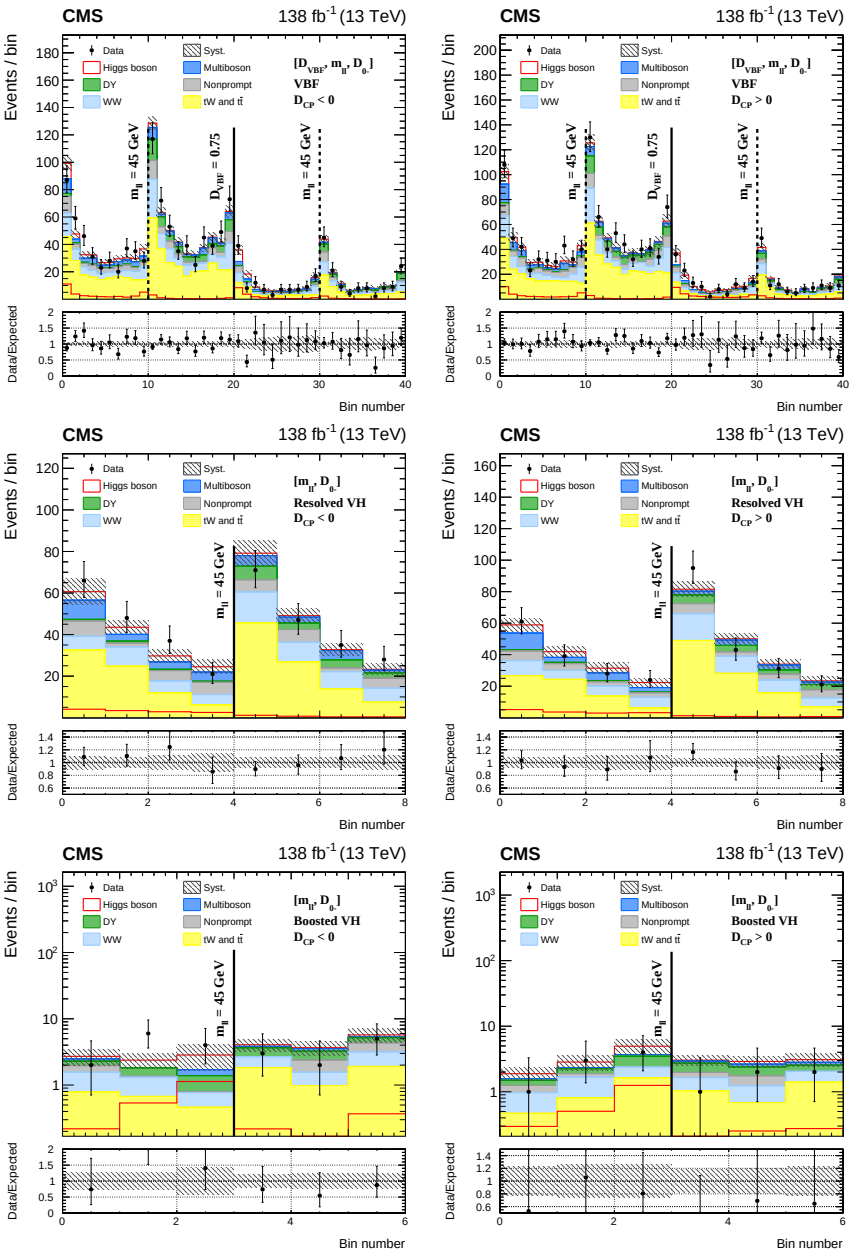

Figure 7.2: Event distributions in Approach 1 for [$\mathcal{D}_{\text{VBF}}$, m_{ll}, $\mathcal{D}_{0-}$] within the VBF channel (top row) and for [m_{ll}, $\mathcal{D}_{0-}$] within the VH Resolved (middle row) and VH Boosted (bottom row) channels. For each channel, the $\mathcal{D}_{\text{CP}} < 0$ (left) and $\mathcal{D}_{\text{CP}} > 0$ (right) sub-categories are shown. The predicted distributions for the Higgs boson signal and backgrounds are shown post data fitting. For the fit, the SM and a_3 HVV coupling contributions are considered. The shaded regions represent the total systematic uncertainty. Statistical error bars are plotted for data. The lower panel of each plot displays the ratio of observed event count to total prediction.

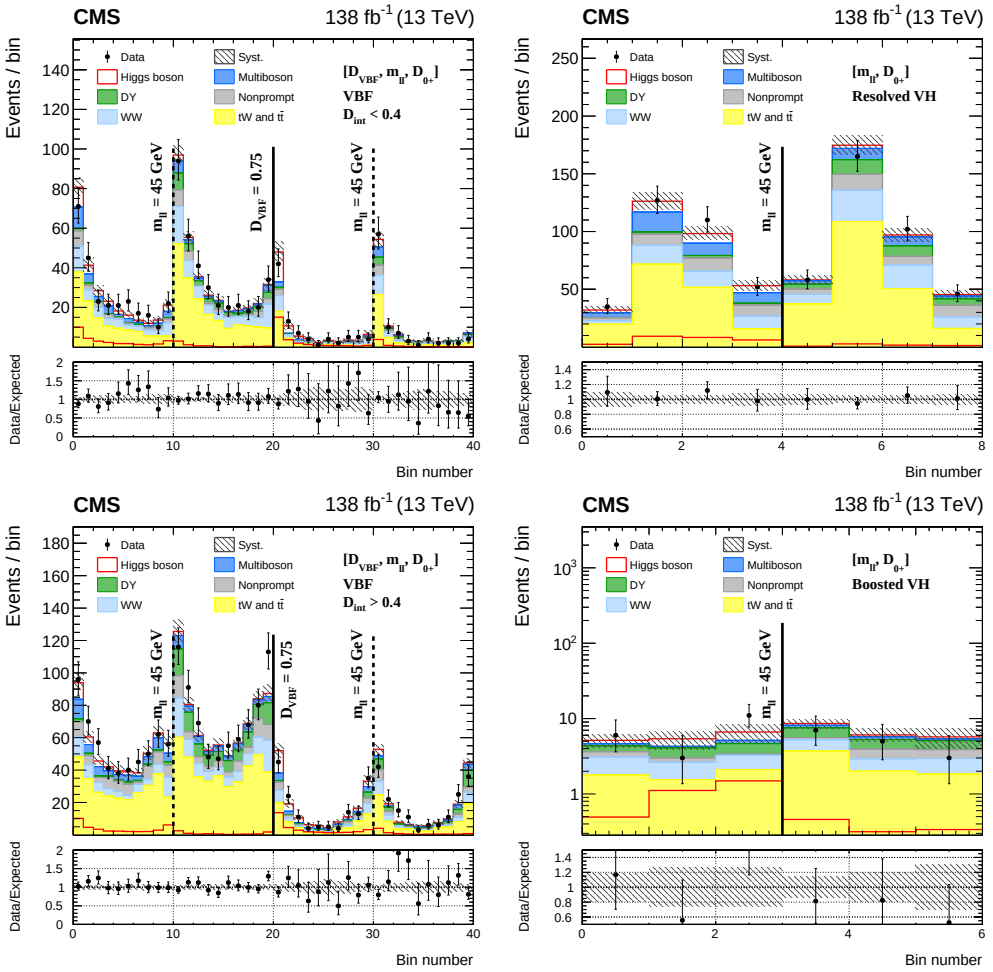

Figure 7.3: Event distributions in Approach 1 for $[\mathcal{D}_{\mathrm{VBF}}, m_{\mathrm{ll}}, \mathcal{D}_{0+}]$ within the VBF channel (left) and for $[m_{\mathrm{ll}}, \mathcal{D}_{0+}]$ within the VH Resolved (right upper) and VH Boosted (right lower) channels. For VBF channel, the $\mathcal{D}_{\mathrm{int}} < 0.4$ (left upper) and $\mathcal{D}_{\mathrm{int}} > 0.4$ (left lower) sub-categories are shown. The predicted distributions for the Higgs boson signal and backgrounds are shown post data fitting. For the fit, the SM and a_2 HVV coupling contributions are considered. The shaded regions represent the total systematic uncertainty. Statistical error bars are plotted for data. The lower panel of each plot displays the ratio of observed event count to total prediction.

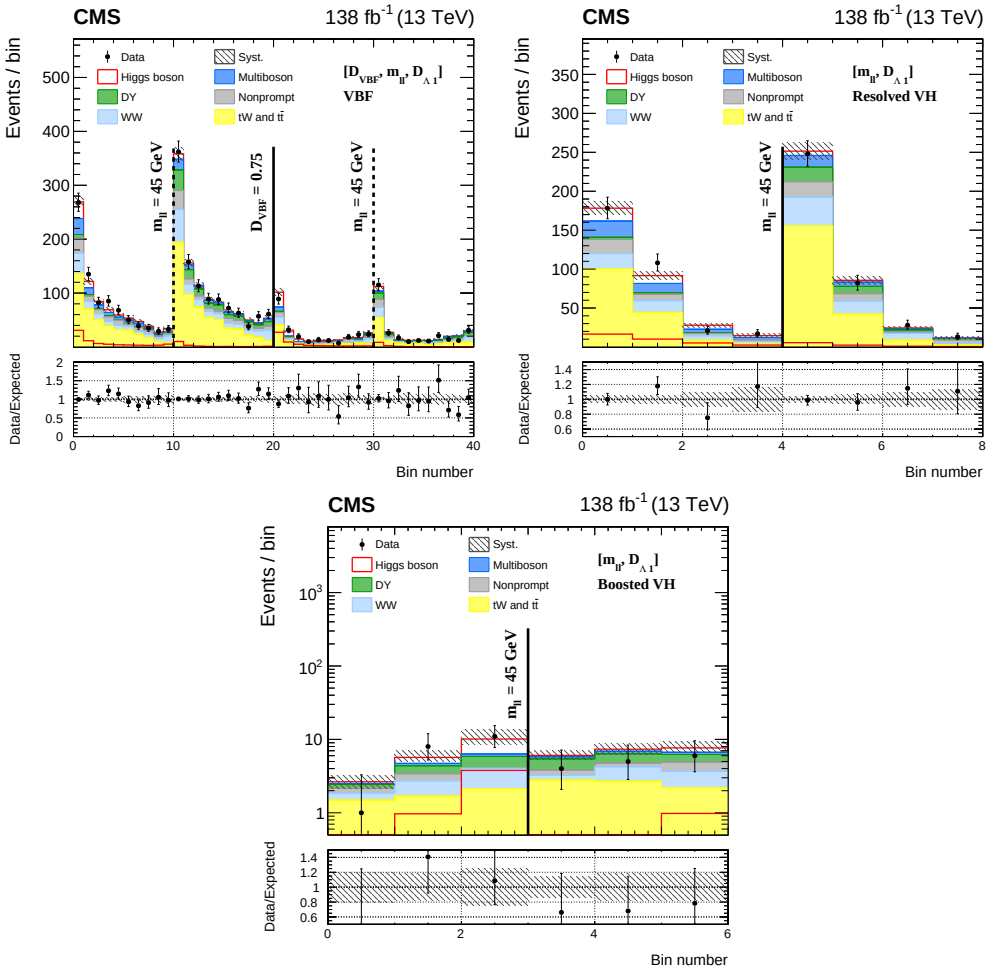

Figure 7.4: Event distributions in Approach 1 for $[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{\Lambda 1}]$ within the VBF channel (top left) and for $[m_{\text{ll}}, \mathcal{D}_{\Lambda 1}]$ within the VH Resolved (top right) and VH Boosted (bottom) channels. The predicted distributions for the Higgs boson signal and backgrounds are shown post data fitting. For the fit, the SM and $a_{\Lambda 1}$ HVV coupling signals are considered. The shaded regions represent the total systematic uncertainty. Statistical error bars are plotted for data. The lower panel of each plot displays the ratio of observed event count to total prediction.

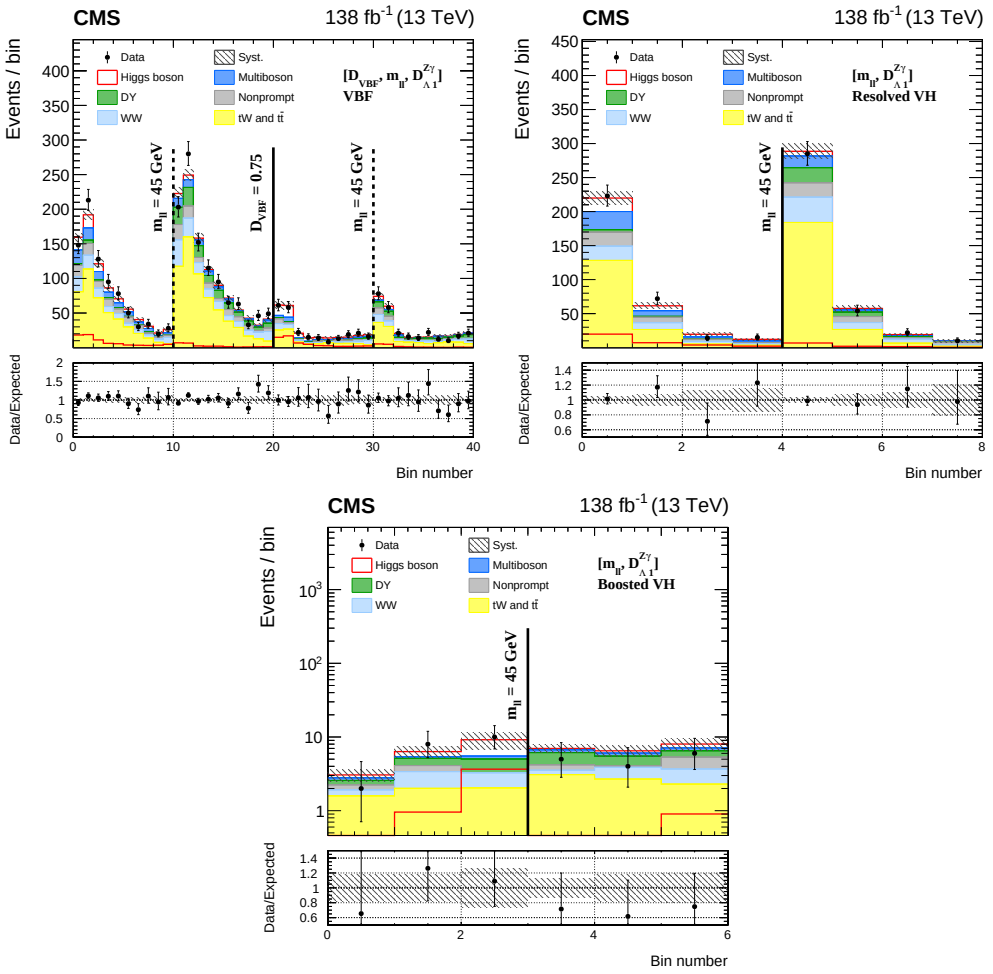

Figure 7.5: Event distributions in Approach 1 for $[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}^{Z\gamma}_{\Lambda 1}]$ within the VBF channel (top left) and for $[m_{\text{ll}}, \mathcal{D}^{Z\gamma}_{\Lambda 1}]$ within the VH Resolved (top right) and VH Boosted (bottom) channels. The predicted distributions for the Higgs boson signal and backgrounds are shown post data fitting. For the fit, the SM and $a^{Z\gamma}_{\Lambda 1}$ HVV coupling signals are considered. The shaded regions represent the total systematic uncertainty. Statistical error bars are plotted for data. The lower panel of each plot displays the ratio of observed event count to total prediction.

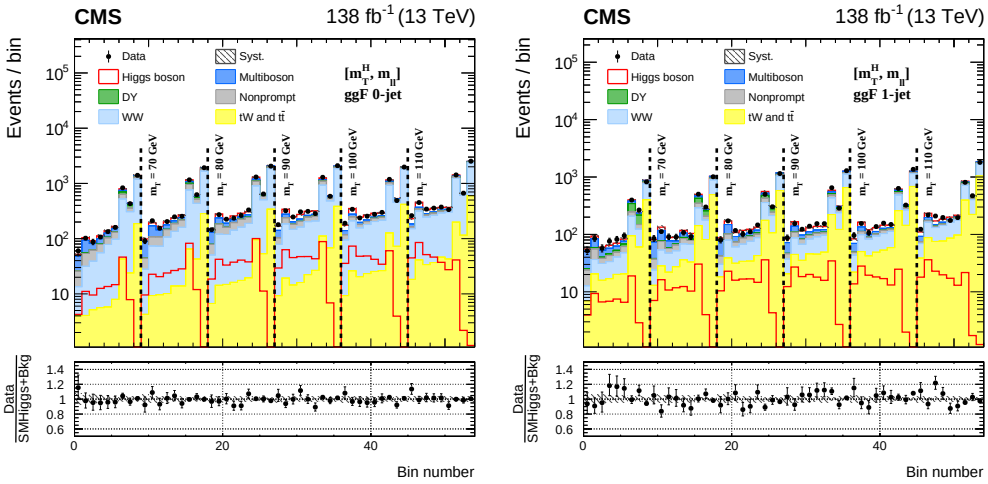

Figure 7.6: Event distributions in Approach 1 for $[m_{\mathrm{T}}^{\mathrm{H}}, m_{\mathrm{ll}}]$ discriminant within the ggF + 0-jet channel (left) and ggF + 1-jet channel (right). The predicted distributions for the Higgs boson signal and backgrounds are shown post data fitting. For the fit, the SM and a_3 HVV coupling signals are considered. The shaded regions represent the total systematic uncertainty. Statistical error bars are plotted for data. The lower panel of each plot displays the ratio of observed event count to total prediction.

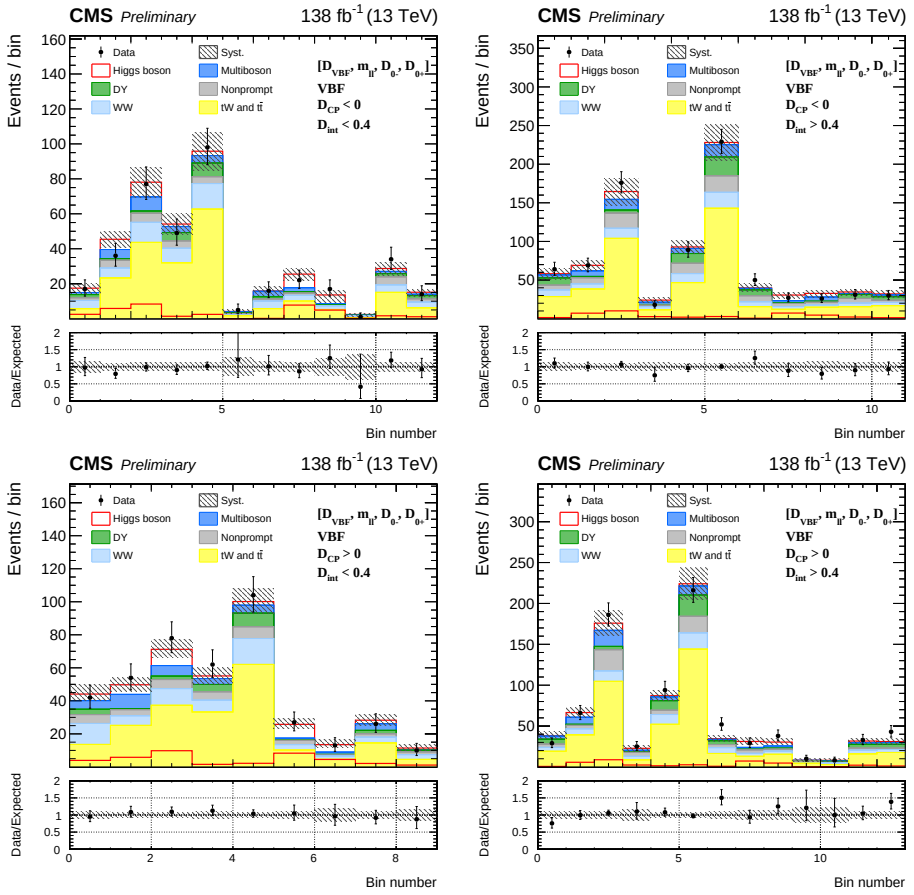

Figure 7.7: Event distributions in Approach 2 for $[\mathcal{D}_{\text{VBF}}, m_{\text{ll}}, \mathcal{D}_{0-}, \mathcal{D}_{0+}]$ within the VBF channel, divided into sub-categories based on the further selection: $\mathcal{D}_{\text{CP}} < 0$ (top row), $\mathcal{D}_{\text{CP}} > 0$ (bottom row), $\mathcal{D}_{\text{int}} < 0.4$ (left) and $\mathcal{D}_{\text{int}} > 0.4$ (right). The predicted distributions for the Higgs boson signal and backgrounds are shown post data fitting. For visualisation purposes, only the contributions corresponding to the a_2 coupling, the SM coupling and their interference enter the Higgs signal. The shaded regions represent the total systematic uncertainty. Statistical error bars are plotted for data. The lower panel of each plot displays the ratio of observed event count to total prediction.

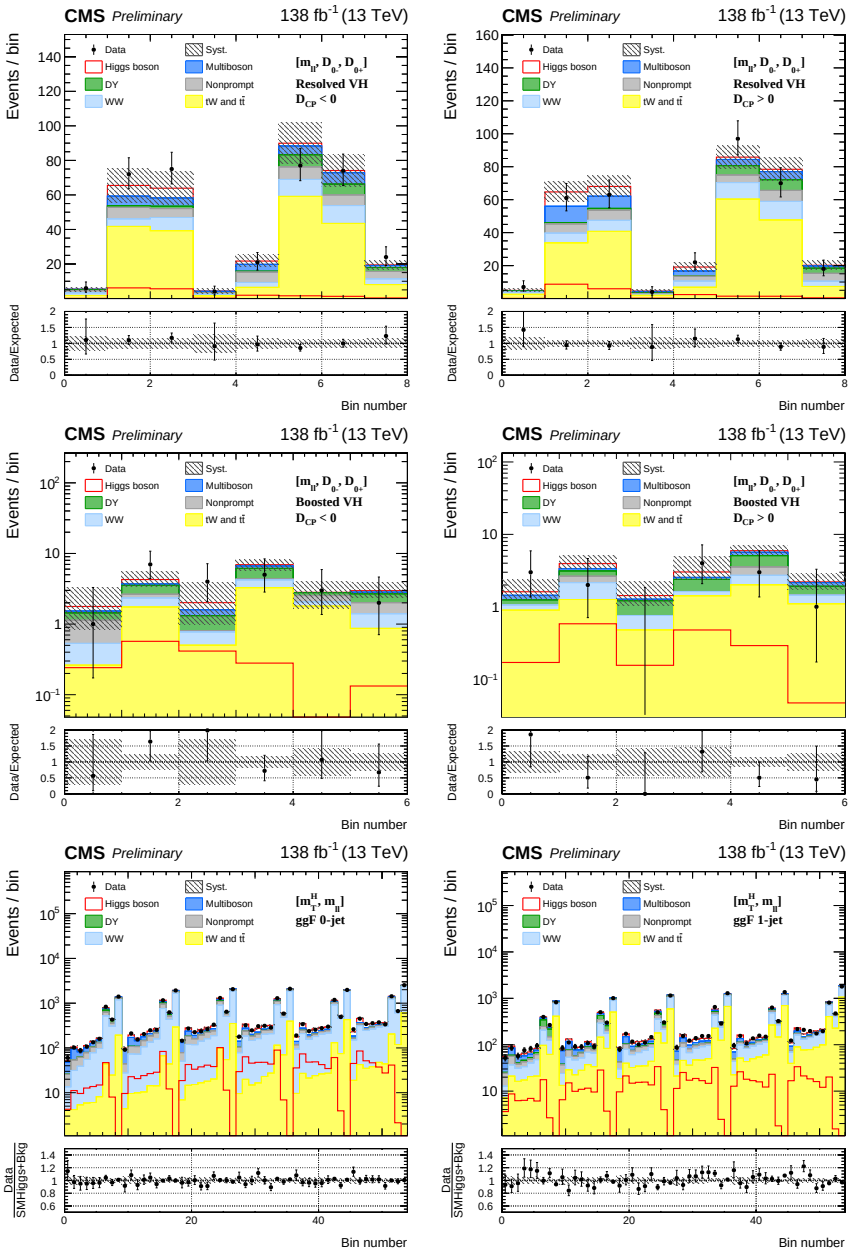

Figure 7.8: Event distributions in Approach 2 for $[m_{\ell\ell}, \mathcal{D}_{0-}, \mathcal{D}_{0+}]$ within the VH Resolved (top row) and VH Boosted (middle row) channels, and for $[m_{\mathrm{T}}^{\mathrm{H}}, m_{\ell\ell}]$ within the 0-jet (bottom left) and 1-jet ggF (bottom right) channels. VH categories are further divided into sub-categories based on the selection: $\mathcal{D}_{\mathrm{CP}} < 0$ (left) and $\mathcal{D}_{\mathrm{CP}} > 0$ (right). The predicted distributions for the Higgs boson signal and backgrounds are shown post data fitting. For visualisation purposes, only the contributions corresponding to the a_2 coupling, the SM coupling and their interference enter the Higgs signal. The shaded regions represent the total systematic uncertainty. Statistical error bars are plotted for data. The lower panel of each plot displays the ratio of observed event count to total prediction.

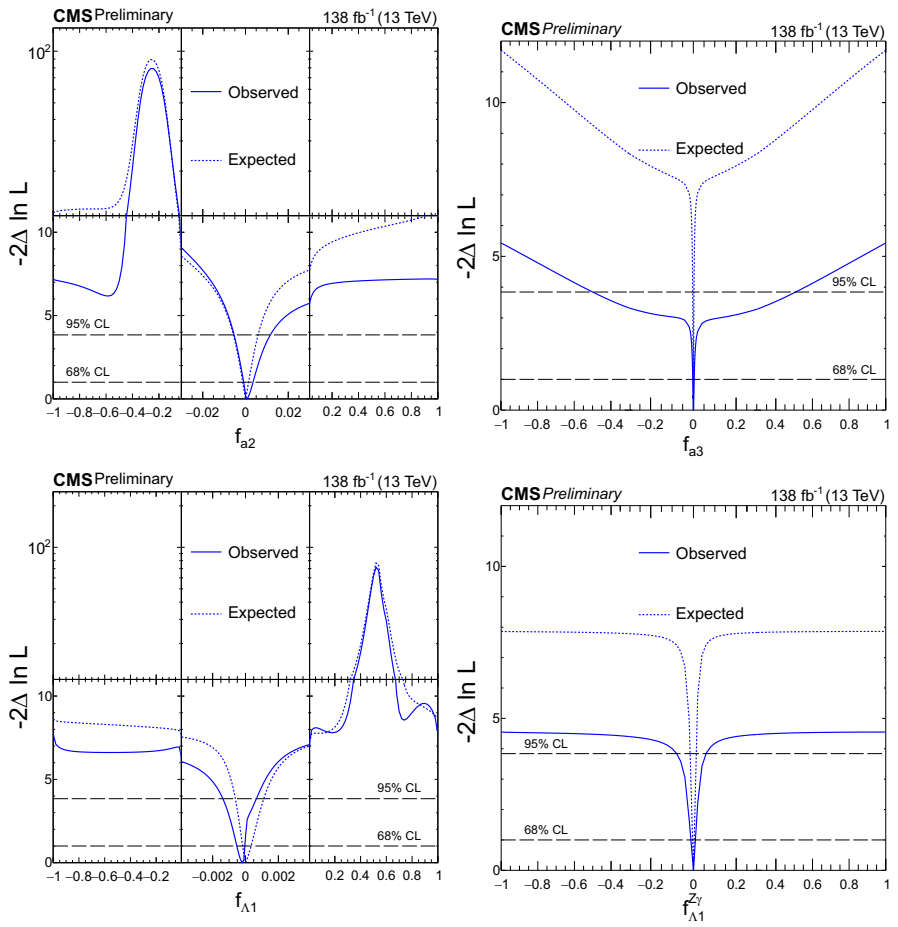

Figure 7.9: Expected (dashed) and observed (solid) profile-likelihood (PL) scans as a function of f_{a_2} (top left), f_{a_3} (top right) and $f_{a_{\Lambda 1}}$ (bottom left) and $f_{a_{\Lambda 1}^{Z\gamma}}$ (bottom right) using Approach 1 are shown for the combination of Run-2 datasets. In each case, signal strength modifiers were kept profiled in the fit. Intersection of the PL scans with CL = 68% and CL = 95% contours (horizontal lines) are marking the corresponding 1σ and 2σ confidence intervals.

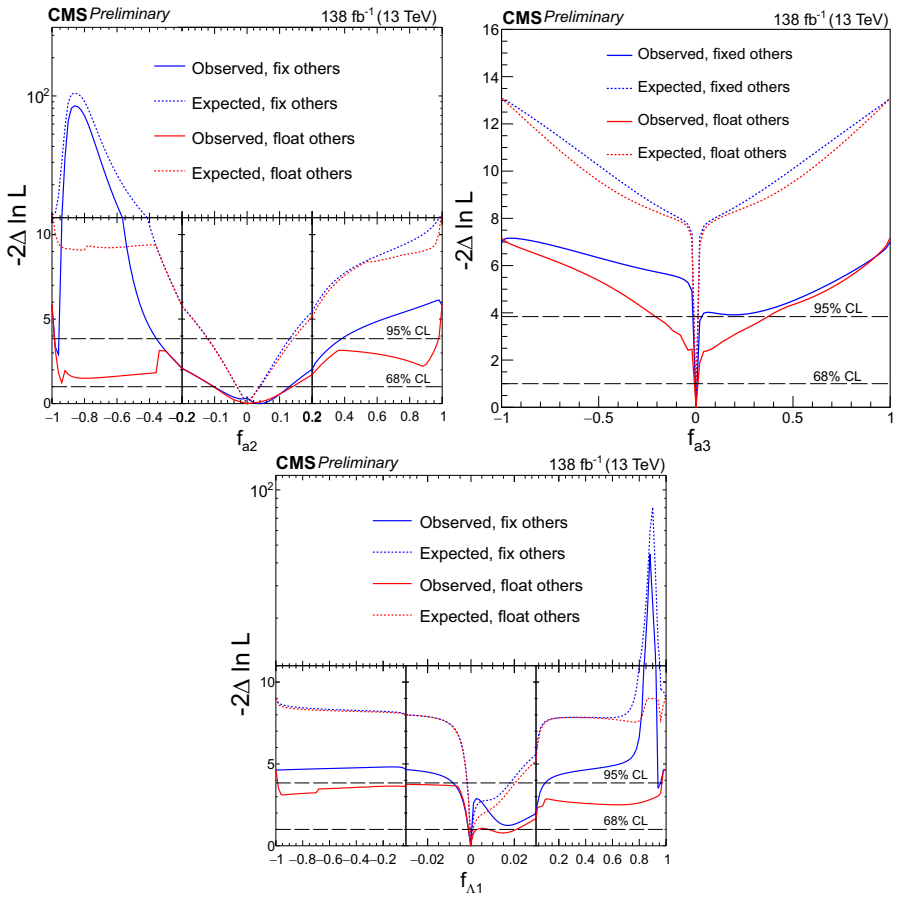

Figure 7.10: Expected (dashed) and observed (solid) profile-likelihood (PL) scans as a function of f_{a_2} (top left), f_{a_3} (top right), $f_{a_{\Lambda 1}}$ (bottom) using Approach 2 are shown for the combination of Run-2 datasets. Results are presented in a twofold way: other two f_{a_i} parameters considered in the fit are kept floating (red curves) or fixed to the SM (zero) value (blue curves). In each case, signal strength modifiers were kept profiled in the fit. Intersection of the PL scans with CL = 68% and CL = 95% contours (horizontal lines) are marking the corresponding 1σ and 2σ confidence intervals.

CHAPTER

EIGHT

DISCUSSION AND OUTLOOK

This thesis aims to present a detailed description of an analysis placing constraints on anomalous Higgs boson couplings (ACs) from its production and decay using the (on-shell) $H \to WW^*$ channel with a fully leptonic (different flavor $e\mu$) final state. Final results correspond to an integrated luminosity of 138 fb^{-1} obtained in a combination of the 2016-2018 datasets collected during the Run-2 period of data-taking at $\sqrt{s} = 13$ TeV by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at CERN.

The analysis was initially performed independently for ggF, VBF, VH Resolved and VH Boosted signal regions (SRs) as well as separately for all three datasets. During this process, analysers had the chance to optimize the selection cuts, define the set of observables used for the construction of kinematic discriminants, and optimize their binning scheme. Furthermore, the limited version of the results, such as Figures 6.1, served as a hint at the possibly achievable overall sensitivity to constrain ACs. The final constraints were then extracted in the combination of datasets and channels, as visualized in the Profile Likelihood (PL) scans in Figure 7.9 and Figure 7.10. It needs to be stressed out, however, this analysis was from the beginning aiming towards the overall combination of channels (e.g. by assuming a general physical model), therefore, these limited results, if shown explicitly in this thesis, should not be considered as final in any sense. On the other hand, results corresponding to the full combination [38] are currently considered for publication. Several remarks concerning the limited results as compared to the final results are now in place.

In general, the inclusion of more data (e.g. by merging datasets) was shown to significantly improve the final results, while the 2018 dataset on its own (with the largest fraction of the total integrated luminosity) is shown to outperform the analyses with the 2016 or 2017 data. It was also shown that the ggF + 0-jet and ggF + 1-jet channels on their own (neither in a $0 + 1$ jet combination) are not enough to set 95% CL intervals for the ACs accessible in these channels, and that the combination with VBF and VH channels is necessary. As a major part of another PhD studies (fully out of scope of this thesis), it was confirmed that adding the VH Boosted channel significantly improves the combination of VBF and VH Resolved channels and consequently improves the overall combination. On the other hand, the sensitivity of constraining ACs in the combination of VBF and VH channels alone is suboptimal and gets improved after including the ggF channels. Before we proceed to the discussion of results in the context of similar analyses, let us present a short summary of the content presented in this thesis as follows.

Chapter 1 is dedicated to the overview of Standard Model (SM) physics and can serve as an introduction (or reminder) to the world of High Energy Physics (HEP) experiments. The last few sections in this chapter are targeting the Effective Field Theory (EFT) extension of the SM (also referred to as Beyond the SM or BSM) with a particular focus on possible scenarios of various ACs contributing to the HVV interaction vertex. The AC and SMEFT schemes were introduced as two equivalent concepts in the interpretation of the final results, assuming distinct relations between anomalous couplings, i.e. $a_i^{WW} = a_i^{ZZ}$ is considered for the AC scheme, and $SU(2)\times U(1)$ symmetry relations are considered for the SMEFT scheme.

This analysis would never be possible without the dedication of countless people involved in LHC operations and the CMS experiment. A short summary of the main detector features, components, and their sub-detection systems is presented in Chapter 2. The general principles of detecting particles of various properties were also introduced. Knowledge of the detector sub-systems has become handy in Chapter 3, which is focused on two important aspects in all HEP analyses: Monte Carlo (MC) simulation of events and physical object reconstruction for both simulation and data. Particular focus was put on the latter, as all measurable objects involved in this analysis have received special treatment tuned especially for the purpose of this analysis (and analyses sharing the similar phase space). The topic of jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is presented in more detail, given my personal contribution as the HWW Physics Analysis Group (PAG) JetMET contact during the duration of my PhD studies.

Two analysis approaches, dubbed as Approach 1 and Approach 2, are introduced later on in Chapter 4 and are closely related to the choice of the anomalous coupling scheme, i.e. Approach 1 is employed in the case of AC scheme and Approach 2 in the case of SMEFT scheme. An important distinction is a way couplings are treated in the physical model entering the final fit, e.g. exactly one AC at a time (together with the SM coupling) is considered in Approach 1, while the simultaneous fit of all ACs is attempted in Approach 2. This chapter also focuses on various techniques in the estimation of the background contribution and describes the construction of the signal templates including the SM-BSM interference effects, by employing the MELA re-weighting procedure. The latter aspect was particularly involved and required a good amount of cross-checks, namely to understand fluctuating or negative templates and their actual incorporation into the physical model used in the fit.

Chapter 5 speaks about the overall analysis strategy involving both the description of the analysis phase space and the construction of the kinematic discriminants (KDs) used in particular analysis channels and for both approaches. The former topic introduces the base kinematic selection, requirements needed to ensure orthogonality with other analyses, and categorisation criteria used to define analysis signal regions (targeting different Higgs boson production processes) as well as control regions (used to constrain the background contribution from data). The latter topic focuses on the construction and optimisation of particular KDs, e.g. a two-dimensional m_{ll} vs. $m_{\mathrm{T}}^{\mathrm{H}}$ discriminant is formed in ggF channels, while two and more-dimensional discriminants are formed in VBF and VH channels, predominantly relying on MELA KDs, depending on the availability of kinematic information and sensitivity of particular observables to anomalous effects.

Inclusion of the physical model in the Profile Likelihood (PL) fit procedure is discussed in Chapter 6. The particular shapes of the PL scans are explained in the context of SM-BSM interference effects occurring at $f_{a_2} = -0.25$ and $f_{a_{\Lambda 1}} = 0.5$ values for the AC scheme (Approach 1) as well as at $f_{a_2} = -0.85$ and $f_{a_{\Lambda 1}} = 0.85$ values for the SMEFT scheme (Approach 2). In both approaches, fitted discriminant templates were showing

distinct distribution shapes when compared to other values of f_{a_i}, thus directly affecting the observed PL scan shapes. A detailed discussion of the statistical and systematic uncertainties entering the PL fit in the form of nuisance parameters is also included, and their impacts on the final results are shown. It was observed that the signal strength modifiers, μ_{ggF} and μ_{EWK}, actually belong among the most "impactful" nuisance parameters. Besides those, various parameters controlling background normalisation from dedicated control regions were found to have an effect on the final f_{a_i} estimate. An example is a parameter controlling nonresonant WW normalisation measured in 2-jet channels in the case of the f_{a_3} PL scan. To name a few nuisance parameters directly related to systematic uncertainties, we observe some high-ranked impact of b-tagging uncertainties, uncertainties related to the estimation of lepton misidentification (fake) rate, uncertainties arising from higher-order QCD corrections, or uncertainties on lepton ID efficiency. Some of the individual analysis bins also appear to have high impact, given the assigned statistical uncertainty, nevertheless, this is expected considering the huge number of bins entering the total fit.

Note that this analysis was initially performed in a blind regime, i.e. observed distributions of events in SRs were not examined until all details of this analysis were tested and finalized. Once approved, the "unblinding" procedure was performed in a controlled manner, including the study of the validity of the fit, determining the impacts of nuisance parameters, the study of data distributions before and after the fit, and finally performing PL scans and extracting AC constraints and corresponding confidence intervals. Eventually, most of these items are presented in Chapter 7 individually for both approaches. The same f_{a_i} symbols are used in both cases, however, constraints on the a_i coupling in Approach 1 should be considered distinct from the quoted constraints on the a_i coupling in Approach 2. As a matter of fact, results in Approach 2 are yet shown in a two-fold way: all f_{a_i} parameters are allowed to float simultaneously in the fit or others than the fitted f_{a_i} parameter are kept fixed to the SM value. All results were found to be in agreement with the expected SM values, i.e. all the best f_{a_i} estimates as obtained in the fit on real data were found to be very close to zero. This fact is also reflected in the postfit kinematic distributions showing the total signal contribution very close to the expected SM signal while no significant excess of data was found. Special attention was paid to the f_{a_3} PL scans, however, no hints for the CP violating effects in the HVV vertex could be confirmed (neither fully excluded). The CL intervals were also extracted, and dedicated discussion will follow in separate sections as we compare our results with previous or parallel studies. My personal contribution to this analysis and other professional activities are summarised in the last section of this chapter.

Comparison with the Run-1 results in HWW channel

The study of anomalous Higgs boson couplings was previously performed (fully out of the scope of this thesis) in $H \to ZZ^*, Z\gamma^*, \gamma^*\gamma^* \to 4l$, $H \to WW^* \to 2l2\nu$ and $H \to \gamma\gamma$ decay channels [33] using datasets collected by the CMS during the Run-1 period of data-taking corresponding to the integrated luminosity of 19.7 fb^{-1} ($\sqrt{s} = 8$ TeV) and 5.1 fb^{-1} ($\sqrt{s} = 7$ TeV). This analysis was especially useful in providing the ultimate exclusion intervals for various exotic spin-parity Higgs boson scenarios, nevertheless, constraints on spin-0 anomalous couplings were also presented. In particular, analysers studying the $H \to WW^* \to 2l2\nu$ decay channel have set the following best-fit values and (observed) 95% CL intervals: $f_{a_2}^{\text{Run-1}} \approx -0.02[-1.00, -0.54] \cup [-0.29, 1.00]$, $f_{a_3}^{\text{Run-1}} \approx -0.03[-1.00, 1.00]$, $f_{a_{\Lambda 1}}^{\text{Run-1}} \approx 0.21[-1.00, 1.00]$ with ranges truncated at the

physical boundaries of $f_{a_i} = \pm 1$. The quoted Run-1 results would correspond to our Approach 1 and AC scheme, however, the considered categories were only those targeting the ggF process, neglecting the combination with VBF or VH channels. As such, no direct comparison is possible, on the other hand, an improvement of roughly three orders in the estimation of the best-fit f_{a_i} values with Run-2 results is notable. Furthermore, it was shown that combining $H \to WW^*$ and $H \to ZZ^*$ channels leads to a sizable improvement in constraining ACs and served as the prime motivation for performing our analysis using Run-2 datasets.

Comparison with the HZZ Run-2 analysis

$H \to ZZ^* \to 4l$ (HZZ) channel is particularly important among the studies constraining anomalous couplings as it is sensitive to anomalous effects in Higgs boson production and decay (similar to the $H \to WW^* \to 2l2\nu$ channel), while also operating in a fully reconstructible final state (without the presence of neutrinos). Taking into account complications arising from dealing with $E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state, the constraints displayed in Table 7.1 are found to be competitive with those reported in [153]. In particular, we can compare two sets of results – AC scheme with Approach 1 and the SMEFT scheme in Approach 2 with all f_{a_i} parameters floating. In the former case, analysers have set the following best-fit values and (observed) 95% CL intervals:

- $f_{a_3}^{\mathrm{HZZ}} \approx 0.04[-0.02, 1.68] \times 10^{-3}$,
- $f_{a_2}^{\mathrm{HZZ}} \approx 0.20[-0.78, 3.68] \times 10^{-3}$,
- $f_{a_{\Lambda 1}}^{\mathrm{HZZ}} \approx 0.04[-0.14, 0.60] \times 10^{-3}$,
- $f_{a_{\Lambda 1}^{Z\gamma}}^{\mathrm{HZZ}} \approx -0.01[-3.87, 3.01] \times 10^{-3}$,

while in the latter case, the best-fit values and (observed) 95% CL intervals are:

- $f_{a_3}^{\mathrm{HZZ}} \approx 0.05[-0.72, 2.18] \times 10^{-3}$,
- $f_{a_2}^{\mathrm{HZZ}} \approx -0.02[-6.94, 5.36] \times 10^{-3}$,
- $f_{a_{\Lambda 1}}^{\mathrm{HZZ}} \approx 0.12[-1.84, 4.43] \times 10^{-3}$.

In both cases, the quoted CL intervals are found to be a subset of the CL intervals reported in this thesis, i.e. sensitivity of the HZZ channel to constrain ACs is superior to that of the HWW channel. Further improvement in constraining the ACs is expected once the combination of these channels is performed. It is worth mentioning that both compared analyses have employed the MELA technique to perform MC matrix element re-weighting. Both studies also used MELA kinematic discriminants sensitive to anomalous effects to the extent allowed by different analysis phase space.

Comparison with the H$\tau\tau$ Run-2 analysis

It is possible to obtain constraints on ACs via the $H \to \tau\tau$ (H$\tau\tau$) decay channel. One of the advantages of looking into this phase space is the relatively high significance of the signal events in the VBF channel as compared to other channels. On the other hand, the presence of neutrinos in leptonic τ lepton decay causes complications comparable to the HWW analysis. Initially, analysers attempted to measure anomalous effects in the HVV

vertex with early 2016 data and combined these results with the HZZ channel [141]. Later on, a similar analysis targeting H$\tau\tau$ channel was performed in combination of Run-2 data [154]. The analysers presented the following best-fit f_{a_i} values and 95% CL intervals corresponding to the AC scheme and Approach 1:

- $f_{a_3}^{\mathrm{H}\tau\tau} \approx 0.28[-0.01, 1.30] \times 10^{-3}$,
- $f_{a_2}^{\mathrm{H}\tau\tau} \approx 1.1[-3.4, 3.2] \times 10^{-3}$,
- $f_{a_{\Lambda 1}}^{\mathrm{H}\tau\tau} \approx -0.12[-0.34, 0.01] \times 10^{-3}$,
- $f_{a_{\Lambda 1}^{Z\gamma}}^{\mathrm{H}\tau\tau} \approx 2.5[-3.6, 6.6] \times 10^{-3}$.

Constraints corresponding to the CP-odd a_3 coupling were also presented in the SMEFT scheme and Approach 2. In particular, they found a $f_{a_3} \approx 0.40 \times 10^{-3}$ best-fit value and a corresponding 95% CL interval of $[-0.01, 1.90]$. These results are again competitive with the results presented in this thesis and those obtained in the HZZ decay channel. MELA re-weighting and MELA KDs were also found to provide superior results.

Available analysis extensions

At the time of writing this thesis, the related publication [38] supporting the AC analysis in the $H \to WW^* \to 2l2\nu$ decay channel was in the process of internal CMS approval. As such, additional studies were performed but did not fit into the scope of this thesis. Among those, the reader can expect the following extra results to get published:

- constraints on the CP-odd a_3^{ggH} coupling indirectly affecting the Hgg vertex as measured in the dedicated ggF + 2-jet SR,
- constraints on ACs translated to the language of Wilson coefficients (entering EFT Lagrangian terms) as rotated to both the Warsaw and Higgs bases.

Future plans and prospects

Furthermore, the results presented in this thesis and the supporting publication [38] are expected to get included in a combination of all Higgs decay channels sensitive to anomalous effects in the HVV vertex, e.g. HWW, HZZ, H$\tau\tau$, HZγ, H$\gamma\gamma$ and possibly other channels, including those assuming an off-shell configuration. Such an analysis was not yet performed with Run-2 data, while it is expected that it would lead to the ultimate set of AC constraints. Given the possibility of ultimate Run-2 constraints not yet fully excluding a non-zero anomalous contribution to the HVV vertex, follow-up analyses with Run-3 data might be considered.

All of the aforementioned analyses constraining the ACs in the HVV vertex are based on the inclusion of the EFT information straight from the beginning of the analysis chain. The various EFT models are affecting the event kinematics and rate in different manners. The kinematic discriminants have their shape influenced by the choice of the assumed anomalous coupling, and the anomalous terms are explicitly entering the formula for the scattering amplitude. In other words, the analysis is optimized for the extraction of the particular selection of anomalous couplings in a model-dependent way. Alternatively, it is possible to measure the (differential) cross-section for a particular process (in our case,

Higgs boson production followed by $H \to WW^*$ decay), while suppressing the model-dependence of the final results. This is typically achieved by means of the fiducial cross-section measurement [155] (only partially model-independent) or by employing more sophisticated methods, e.g. the adversarial deep neural network [156]. Once this is done, any deviation from the SM cross-section expectation can be found by assuming the proper cross-section parametrisation with the AC effects included in the physical model used in the fit. Up to this date, this approach has not been adopted by the CMS Collaboration, however, it is favored in similar studies performed with data collected by the ATLAS detector. As such, this method could serve as a complementary measurement opening further possibilities for comparison with a wider range of analyses and across theoretical models.

Another important property describing the Higgs boson is its decay width, the value of which is theoretically predicted by the SM as $\Gamma_{\mathrm{H}} = 4.1$ MeV [157]. Because of the limited mass resolution of 1 GeV currently achievable by the CMS and ATLAS detectors, it is not possible to experimentally confirm this value by means of direct measurements. On the other hand, comparing the Higgs boson production and decay cross-section measurements for both on-shell ($\sigma^{\text{on-shell}}_{pp\to H\to VV} \approx \mu$) and off-shell ($\sigma^{\text{off-shell}}_{pp\to H\to VV} \approx \mu \times \Gamma_{\mathrm{H}}$) productions modes, offers the possibility to indirectly extract the Γ_{H} value. As a matter of fact, around 10% of all $pp \to H \to VV$ events in the SM is with the Higgs boson produced off-shell. Such a sensitivity is sufficient for the standalone measurement, as already demonstrated in the off-shell HZZ study [140] with Run-2 datasets. A similar measurement can now also be performed in the HWW off-shell channel. In addition to the extraction of the decay width, it would also be possible to probe sensitivity to constrain the anomalous couplings.

Personal contribution

Finally, let me summarise my personal contribution, both directly affecting the AC analysis presented in the core of this thesis, as well as the secondary activities performed throughout the scope of my PhD studies.

Constraining the ACs in the HWW decay channel was a group effort which has resulted in two PhD theses (including this one) and a corresponding publication [38] (currently in the approval process). As previously stated, the main analysis workflow was divided according to the main SRs, with my responsibilities primarily including analyses in the 0- and 1-jet ggF channels, performing a combination across datasets and channels, as well as performing analysis unblinding in the process of extracting observed results. The combination and unblinding steps were also performed independently by other analysers to prevent possible human error. Among the crucial analysis steps to which I have contributed are: analysis phase space definition, physical object reconstruction (primarily jets and $E_{\mathrm{T}}^{\mathrm{miss}}$), data-to-MC agreement studies in dedicated control regions, signal template construction and template studies in the ggF channels, definition and optimization of kinematic discriminants in the ggF channels, implementation of systematic uncertainties, physical model definition, fitting and fit diagnostics, unblinding and analysis documentation. Various personal contributions are scattered throughout this thesis, starting from Chapter 3. The true nature of any HEP analysis effort, however, involves a good deal of teamwork and cooperation. As such, overlap between analysis channels could not be avoided, and even more importantly, this analysis would not be possible without it. On the other hand, full credit goes to my fellow analysers performing individual studies in VBF and VH channels, especially dealing with definition of MELA

kinematic discriminants and template binning strategy both in AC and SMEFT schemes in 2-jet channels (mostly Section 5.2.1). Another notable contribution out of scope of my thesis is the development of a smoothing procedure for the SMEFT PL scans and the inclusion of the VH Boosted channel. Not any less important, several analysis aspects, such as MC simulation of common background samples, as well as original Higgs signal samples, numerous tasks concerning LHC and CMS operations during data-taking, and the development of various analysis software tools were all outsourced among the wide range of scientists.

The Higgs boson AC analysis has been performed in close cooperation with the HWW subgroup of the Higgs boson PAG. Throughout the period of my PhD studies I have been performing tasks of the HWW JetMET contact. My responsibilities included: keeping track of various JetMET POG recommendations and assuring their application in all analyses performed by the HWW subgroup, performing various jet and $E_{\mathrm{T}}^{\mathrm{miss}}$ oriented studies – some of which were reported in Section 3.2.4 and Section 3.2.5 as well as developing software tools necessary for the application of JetMET-related corrections and systematic uncertainties. My work on jet object definition in the scope of HWW analysis framework was accounted for in the publication [158], which was particularly sensitive to the proper definition of jet multiplicity and jet reconstruction.

About 20% of my time during doctoral studies, I have been involved in the CMS Tracker Alignment efforts. My responsibilities involved determining the performance of the CMS Tracker detector under various conditions defining the detector alignment (see more in Appendix A), and SW development of the corresponding validation tools. My work has largely contributed to the public technical report of the CMS Tracker performance during the Run-2 period of data-taking [74]. Furthermore, I have been leading the development efforts towards the most up-to-date alignment validation software currently in use.

Another 10% of my professional time I have taken part in the prototyping activities conducted at the Belgium 2S-module assembly center located in Brussels. Assembled silicon-based 2S-modules are planned for installation in the CMS Tracker as a part of the planned CMS Phase-2 upgrade towards the High-Luminosity LHC (HL-LHC) upgrade. My work included the development of a testing box – a semi-automated apparatus capable of performing high-voltage (isolation) and I-V measurements for 2S silicon sensors/-modules at various stages throughout the assembly steps during the prototyping stage. Another important aspect of my contribution was developing the testing procedure itself given my experience acquired during an early testing campaign for the first 5 functional 2S-module prototypes. More about 2S-modules and details of my work is presented in Appendix B.

APPENDIX

A

THE CMS TRACKER ALIGNMENT & VALIDATION

The CMS Tracker is the innermost subdetector system of the CMS and holds the distinction of being the largest silicon-based tracking detector ever built. It was designed to provide measurements of charged particles trajectories, also known as tracks, with an ultimate hit resolution of an order of $\mathcal{O}(10\ \mu\text{m})$. Despite this design, limitations arising from the mechanical installation of the CMS Tracker result in a suboptimal resolution of about $\mathcal{O}(100\ \mu\text{m})$, thus preventing us from exploiting the full detector potential. To address this issue, scientists developed various algorithms, commonly denoted as tracker alignment, providing corrections on the position, orientation and surface deformation for all silicon sensors forming detector's volume. Tracker alignment conditions are essentially expressed by a set of alignment parameters that determine the properties of the sensors and/or the High-Level Structures (HLS), such as ladders forming half-shells in Tracker Barrel Pixel (Barrel Pixel (BPIX)) or half-cylinders in Tracker Forward Pixel (Forward Pixel (FPIX)).

Accurate and timely derivation of alignment parameters is of utmost importance to account for unavoidable movements inside the detector. Among the most common sources of misalignment are

- magnet voltage ramp-ups and ramp-downs $\approx \mathcal{O}(1\ \text{mm})$,
- cooling cycles during detector maintenance $\approx \mathcal{O}(10\ \mu\text{m})$,
- sensor ageing due to the high radiation levels $\approx \mathcal{O}(1\ \mu\text{m})$,

where we have indicated the order of the shift magnitude. The first two types of movement occur several times during the year, typically as a side effect of the scheduled detector maintenance. However, the accidental rapid magnet ramp-downs should also be taken into account. The last point concerns changes in the mobility of the charge carriers (in the silicon) under the influence of the large radiation dose. As a result, the lateral drift of the charged carriers in the presence of the magnetic field is affected, which leads to the shift of the measured hit position over time. Dedicated calibration needs to be performed to account for this phenomenon, while the remaining discrepancies are removed by means of the alignment procedure.

For visualisation purpose, Figure A.1 shows a sketch of a charged particle trajectory as obtained using reconstruction algorithms introduced in Section 3.2.1 for a case with misaligned modules and for modules geometry after corrections. Note that a quantity denoted as a "residual" between the fitted track and the measured detector hits is one of the

Figure A.1: Schematic view of a charged particle traversing detector layers (dotted blue line). Fitted particle trajectory (solid blue line) is shown in situation with misaligned tracker modules (left) as compared to corrected tracker geometry (right). Residual distance between predicted trajectory and actual detector hits (solid red line) is minimal in case of perfectly aligned tracker. Source [163].

most important characteristic features used to assess the accuracy of track reconstruction compared to reality.

In Section A.1, we will briefly describe basic principles of the track-based alignment and introduce two algorithms, Mille-Pede-II [159] and HipPy [160], as particular solutions to this problem. Section A.2 will focus on Distributions of Median Residuals (DMR) and Primary Vertex (PV) validation, two out of many methods employed to evaluate detector performance at various stages of tracker alignment. The presented results are a small subset of the overall alignment and validation efforts [74] performed by our team.

For the sake of completeness, my additional contributions to the development of common alignment and validation tools include:

- The implementation of parallel deployment of HipPy campaigns using HTCondor Dagman [161],
- The implementation of luminosity-weighted variation on DMR and PV validation methods discussed in Section A.2,
- The development and integration of the new "All-in-one" offline validation tool into the CMS software [162].

A.1 Tracker Alignment

The alignment procedure can be essentially narrowed down to the problem of minimisation for a χ^2 function defined as

$$\chi^2(\vec{p},\vec{q}) = \sum_{j}^{\text{tracks}} \sum_{i}^{\text{hits}} \left(\frac{m_{ij} - f_{ij}(\vec{p},\vec{q}_j)}{\sigma_{ij}^m} \right)^2, \tag{A.1}$$

where $\vec{p}$ is the complete set of all assumed alignment parameters describing the detector's geometry, $\vec{q}$ stands for the complete set of all track parameters (e.g. curvature), $m_{ij} \pm \sigma_{ij}$

represents the measurement (e.g. hits) with the corresponding uncertainty (e.g. local hit resolution) and f_{ij} stands for a prediction (i.e. trajectory obtained in a fit assuming $\vec{p}$ and $\vec{q}$ values). Note that the size of the parameter set $\vec{p}$ largely depends on the requested granularity of alignment, ranging from several hundreds of parameters for alignment with higher level detector structures (HLS) to several hundreds of thousands for alignment at the module level. Moreover, the number of parameters entering the χ^2 function can reach millions when also accounting for all track parameters $\vec{q}$. Such a huge parameter space might cause severe computational problems and needs to be optimized for real time application.

In practice, χ^2 minimisation is performed assuming an approximative set of alignment and track parameters, $\vec{p_0}$ and $\vec{q_0}$, and by linearising the prediction term f_{ij}, i.e.

$$f_{ij}(\vec{p},\vec{q}) \rightarrow f_{ij}(\vec{p_0},\vec{q_{0,j}}) + \Delta\vec{p}\frac{\partial f_{ij}}{\partial \vec{p}}(\vec{p_0},\vec{q_{0,j}}) + \Delta\vec{q_j}\frac{\partial f_{ij}}{\partial \vec{q_j}}(\vec{p_0},\vec{q_{0,j}}). \tag{A.2}$$

After some manipulation [159], Equation A.1 can be rewritten as the reduced system of hundreds of thousands linear equations which are further treated via matrix inversion procedure. At this point, two particular solutions are currently available:

- Mille-Pede-II algorithm: Simultaneously estimates all track and module parameters in one global fit. In this case, full matrix inversion is performed.

- HipPy algorithm: The position and orientation of all individual sensors are determined independently from the other sensors, i.e. partial matrix inversion is performed, neglecting the matrix blocks relating the alignment parameters to the track parameters (and consequently to the other modules). The local nature of the HipPy algorithm requires an iterative approach in order to resolve the correlations between the modules. This necessitates the re-running of the CPU-expensive track-fitting algorithm multiple times.

Thanks to the sufficient CMS computational resources, the χ^2 minimisation task, either solved using the Mille-Pede-II or HipPy algorithm, is now achievable in a reasonable time frame.

During data-taking, the altering environment and detector conditions require running multiple alignment campaigns both online with limited degrees of freedom (e.g. automated alignment) and offline shortly after data-taking. Both approaches typically employ alignment on higher-level detector structures to mitigate the effect of limited statistics and time constraints. To achieve a more advanced level of alignment suitable for physics analyses, it is necessary to wait until the CMS Tracker records a sufficient number of events and tracks. As such, more dedicated alignment is typically performed in the middle or at the End of Year (EOY) when this can be achieved. Ultimately, the alignment parameters are determined also after the completion of data-taking (e.g. after Run-2), dubbed as Legacy reprocessing, often with superior granularity (i.e. alignment at the module level).

In this appendix, the Legacy Run-2 reprocessing (green markers and lines) will be compared to the alignment after the EOY re-reconstruction (red) and to the alignment conditions as obtained during data-taking (blue). Principles of alignment validation will be presented in the following section.

A.2 Monitoring Tracker Performance

Several validation methods were developed to monitor the CMS Tracker's performance, e.g. after a new set of alignment parameters is derived. Generally, the validation results should provide a quantitative description of the detector's status and its readiness to properly measure physical objects used in physics analysis. Depending on the particular aspect of the CMS Tracker's functionality under scrutiny and the nature of the test, we differentiate between several validation types. Among those extensively used during Run-2 data-taking and Legacy reprocessing are:

- Validation of tracking performance – based on the measurement of track-hit residuals (i.e. the distance between the re-fitted track and the measured hit in local coordinates). A particular representation of this test involves calculating Distributions of Median Residuals (DMR) and will be discussed later.
- Validation of vertexing performance – based on the measurement of track-vertex residuals (i.e. longitudinal d_z and transversal d_{xy} impact parameters), also known as Primary Vertex (PV) validation, will also be discussed later.
- Reconstruction of dimuon invariant mass – based on the idea that the reconstructed $Z \to \mu\mu$ invariant mass, $m_{\mu\mu}$, should be minimally dependent on the actual position of muons in the detector. This method is particularly useful for the detection of systematic distortions of assumed module positions as compared to reality. To be more specific, a subset of such systematic distortions, e.g. global detector movements, results in changes for which $\Delta\chi^2 \approx 0$. In other words, to account for a pathological situation in which a new set of reconstructed tracks is valid even after a global detector shift has occurred is detectable by means of observable differences in the $m_{\mu\mu}$ distribution.
- Reconstruction of cosmic ray muon tracks – based on an idea that two halves of cosmic muon tracks traversing the detector volume should share the same parameters at the point closest to the nominal beamline. This method is useful in estimating track parameter resolution and, similar to $Z \to \mu\mu$ validation, it helps in detecting systematic distortions.
- Detector overlap validation – takes advantage of detector layers with two overlapping modules placed very close to each other. The difference of track-hit residuals between overlapping modules should be minimal for a perfectly aligned detector.

For the brevity of this text, let us introduce DMR and PV validation methods in more detail, while other methods are left for discussion in [74].

DMR validation

Track-hit residuals measured in local x (y) coordinates, $x'_{\text{pred}} - x'_{\text{hit}}$ ($y'_{\text{pred}} - y'_{\text{hit}}$), are an excellent figure of merit in quantifying tracking performance in data (or simulation). To avoid possible bias with the measurement, all tracks are re-fitted after removing a hit under consideration. Eventually, the median of all track-hit residuals is found for each detector module, and the results are plotted individually for different CMS Tracker subdetectors (i.e. FPIX, BPIX, Tracker Inner Barrel (TIB), Tracker Inner Disks (TID), Tracker Outer Barrel (TOB), Tracker EndCaps (TEC)). In the case of an ideally aligned tracker, the distributions of median residuals (DMR) tend to be centered at zero value,

while any shifts in a distribution peak might hint at misaligned detector conditions. The width of the DMR determines local alignment precision as it is sensitive to the residual random misalignment of the modules, and it depends on the number of tracks used in validation. As such, any DMR comparison, e.g. for different alignment conditions, was performed by ensuring the same number of tracks entering validation.

An example of DMR for BPIX and TID, as obtained for the single muon dataset over the scope of the Run-2 period of data-taking, is displayed in Figure A.2. Initially, the DMR are obtained per smaller portions of the data-taking period, also denoted as Intervals of Validity (IOVs), for which we can assume stable alignment conditions (e.g. periods with an unchanged magnetic field). Consequently, distributions are averaged over all IOVs weighted by the corresponding integrated luminosity delivered to the CMS interaction point by LHC. Distributions for validation on MC simulation are normalised to match the total delivered integrated luminosity per assumed period in data. The particular choice of the single muon HLT path is to ensure a similar topology of events in both MC simulation and data. The validity of the MC-to-data comparison was also ensured by requiring exactly 100 hits per module in the simulation and over the entire range of IOVs considered for validation. Because of the deteriorating status of silicon modules over time, it was not possible to assure this condition for the entirety of the Run-2 period. In particular, pixel detector (strip detector) modules with fewer than 100 hits per Interval of Validity (IOV) over the period of data-taking corresponding to an integrated luminosity of more than 2 (7) fb^{-1} were excluded from this study. However, about 91% of all functional modules have passed this condition. In all cases, Run-2 Legacy reprocessing has shown superior detector performance, as concluded from the comparison of the DMR mean, μ, and width, σ, extracted using a Gaussian fit.

Instead of quantifying the "luminosity-weighted" DMR performance during the Run-2 period, it is useful to visualise how tracking performance evolves over time. An example of the mean of the DMR distribution, μ, as a function of the delivered integrated luminosity is shown in Figure A.3. An improvement in performance stability is shown for alignment with Run-2 Legacy reprocessing as values of μ are shifted closer to zero. In this case, DMR were evaluated with a dataset collected using an inclusive L1 trigger as, due to its higher rate, it assures a larger number of tracks.

PV validation

The ability of the CMS Tracker (predominantly BPIX layers) to reconstruct primary vertices (i.e. vertices belonging to the highest p_{T} tracks) is strongly affected by the quality of the alignment. In particular, the measurement of track-vertex residuals, also denoted as impact parameters, is perfectly suitable for the evaluation of detector's vertexing performance. To avoid possible bias with the measurement, all PV positions are first reconstructed excluding the track under consideration. Additionally, dedicated clustering algorithms [164, 165] are applied to mitigate PU effects. In the case of the perfectly aligned detector, distributions of the mean values for the unbiased impact parameters (usually displayed as a function of the track's angular disposition) are expected to get flat (i.e. centered around zero at y-axis). In practice, any random misalignment of the modules is expected to affect only the resolution of PV residuals, making the width of their distribution larger but preserving the mean. On the other hand, any systematic movements are expected to bias the mean of PV residuals in a way defined by the particular character of misalignment.

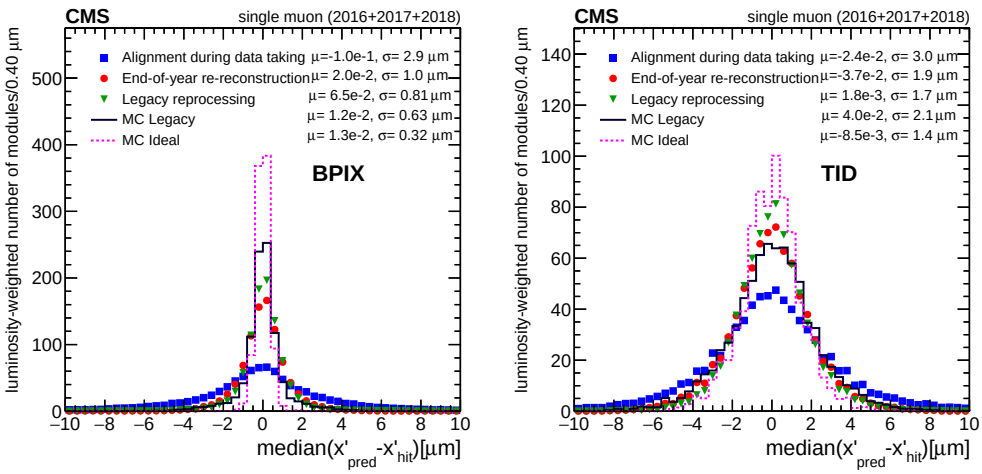

Figure A.2: The luminosity-weighted distributions of median track-hit residuals (DMR) measured in local x (x') coordinates are shown for the CMS Tracker BPIX (left) and TID (right) subdetectors. The DMR are displayed for three different geometries in data and compared with the design MC scenario (magenta) and the realistic MC scenario (black). The quoted means, μ, and standard deviations, σ, are extracted using Gaussian fits to the distributions.

An example of the mean of PV residuals in the transverse plane, $\langle d_{xy} \rangle$, and along the beam axis, $\langle d_z \rangle$, is shown as a function of track pseudorapidity, η, in Figure A.4. Distributions are averaged over all considered data-taking runs (i.e. periods of data-taking shorter or equal to IOVs) and weighted with corresponding delivered integrated luminosity. Only tracks with $p_{\mathrm{T}} > 3$ GeV were considered for this study. Tracker performance after Run-2 Legacy reprocessing is superior to previous alignment campaigns, as mostly visible for the $\langle d_{xy} \rangle$ (η) distribution. Residual discrepancies from an ideal flat scenario in data for high-η values can be explained by limited tracking performance in this region. PV validation in data is also compared to the realistic MC scenario evaluated in simulated inclusive L1 trigger samples where the affect of radiation and limited alignment precision in data (mostly visible for the high-η region) was taken into account.

Similar to the case with DMR validation, it is useful to visualise how the vertexing performance evolves over time. Figure A.5 displays an average of PV residual distributions on a run-by-run basis as a function of delivered integrated luminosity. Apart from a very short period in the beginning of 2017 data-taking corresponding to the commissioning of new Phase-1 Pixel detector, we observe an improvement in the mean of this distribution. Particular improvement is observed for the alignment with Run-2 Legacy reprocessing.

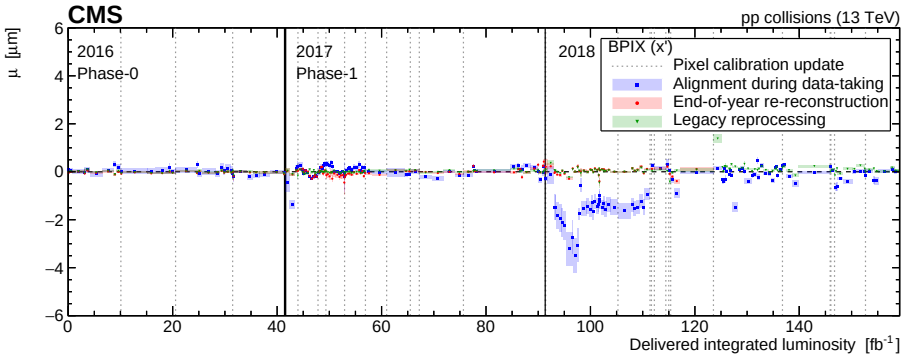

Figure A.3: The DMR trend for the Run-2 period of data-taking (2016-2018) as a function of the delivered integrated luminosity, evaluated with a sample of data recorded by the inclusive L1 trigger. The displayed quantity corresponds to the mean value of the distribution of median residuals for the local x (x') coordinate in the CMS Tracker BPIX subdetector. The shaded band shows one standard deviation from the Gaussian fit of the corresponding DMR. No EOY alignment was performed for the last 33 fb^{-1} of delivered integrated luminosity.

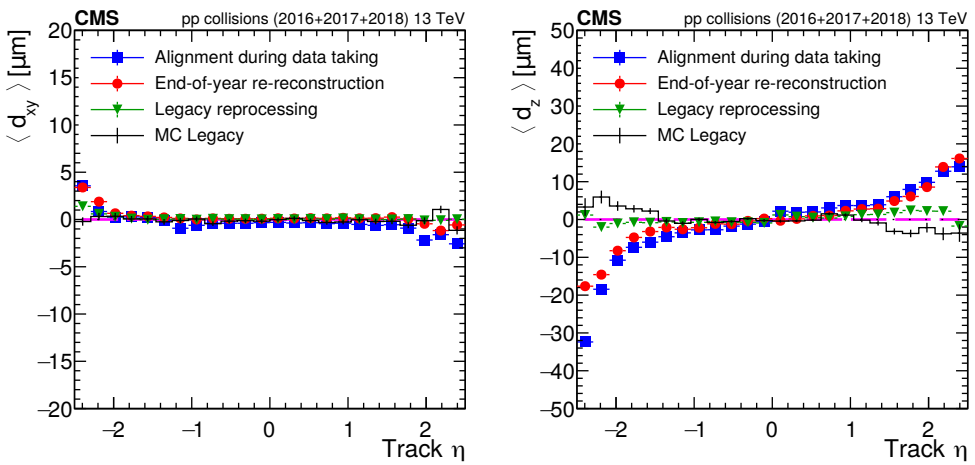

Figure A.4: The luminosity-weighted distributions of track-vertex residuals measured in the transverse plane (left) and along the beam axis (right) are shown as a function of track pseudorapidity. The distributions are displayed for three different geometries in data and compared with the realistic MC scenario (black). The statistical uncertainties due to the limited number of tracks are denoted by vertical error bars (smaller than the marker size in the case of data).

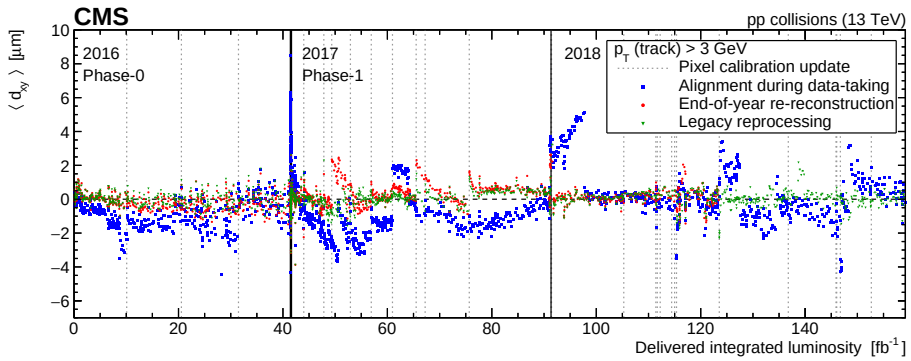

Figure A.5: Mean of the impact parameter trend measured in the transverse plane, d_{xy}, for the Run-2 period of data-taking (2016-2018) as a function of the delivered integrated luminosity. Only tracks with $p_{\mathrm{T}} > 3$ GeV are considered. No EOY alignment was performed for the last 33 fb^{-1} of delivered integrated luminosity.

APPENDIX

B

THE CMS TRACKER PHASE-2 UPGRADE

The High-Luminosity LHC (HL-LHC) is planned to start operations after the LHC Long Shutdown 3 (LS-3) in 2029 [166], as depicted by the timeline in Figure B.1. It is designed for colliding proton beams at a center of mass energy of 14 TeV, enabling the CMS and ATLAS experiments to collect about 3000 fb^{-1} of integrated luminosity each over an anticipated 10-year lifespan. If this projection is achieved, the recorded amount would surpass the integrated luminosity gathered during the Run-2 period by more than 18×, and would be approximately 10× the estimated integrated luminosity to be collected during the Run-3 data-taking. The unprecedented number of about $140 - 200$ proton-proton collisions per bunch crossing (i.e. pile-up) and significantly increased radiation damage give rise to challenging expectations for the CMS detector components. To address these concerns, the Phase-2 CMS detector upgrade [167] has been planned, with some of the new key features including superior detector granularity, radiation hardness, and improved readout capabilities. Most of the Phase-2 detector installation work is scheduled for the LS-3.

The HL-LHC opens a broad spectrum of physics domains reachable with the new upgrade, summarily reported in [168]. Some of the possible studies include:

- the high-precision measurements of the Higgs boson properties, production cross-sections and decay rates,
- other SM precision studies targeting measurements of the global parameters, such as heavy quark or vector boson masses,
- study of rare processes, such as the vector boson scattering or b-meson decays,
- the BSM searches for the dark matter candidates, leptoquarks, heavy scalar particles, etc.

Most of the aforementioned research areas would also benefit from the extended pseudorapidity range in the CMS forward region, introducing additional constraints on the final Phase-2 design. In practice, almost all of the CMS sub-detectors were redesigned to match these and preceding requirements (to cope with higher interaction rates).

The CMS Phase-2 Tracker will comprise the inner [170] and outer [171] subsystems. The inner part will consist of approximately 6× smaller pixelated sensors compared to those of the Phase-1 Pixel Tracker. Both pixel and strip sensor technologies will be employed in constructing the modules installed in the outer part of the tracker. The

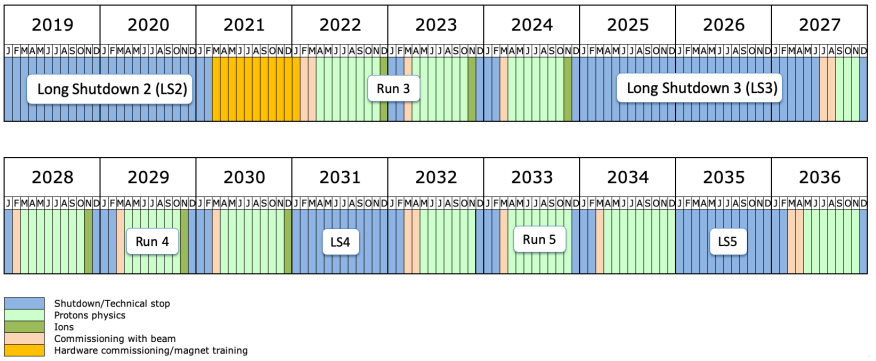

Figure B.1: The LHC schedule up to 2036. Majority of the Long Shutdown 3 period is dedicated to the High-Luminosity LHC upgrade. Source [169].

new CMS Tracker is anticipated to extend the pseudorapidity range up to the $|\eta| = 4$. Another innovation is the CMS Outer Tracker's capacity to provide tracking information to the L1 trigger, thereby aiding in maintaining sustainable trigger rates while ensuring the same efficiency in selecting signal events.

A new calorimetry system featuring finer longitudinal and lateral granularity will provide an improved shower identification and separation within the pseudorapidity range of the $1.4 < |\eta| < 3$. The calorimetry system in the end-cap region of the CMS detector will undergo a complete replacement with the silicon detector technology [172], while the barrel region will receive electronics upgrade to accommodate enhanced trigger requirements [173].

The original Muon System (MS) will be complemented by a new version of Resistive Plate Chambers (RPC) [174], improving detector acceptance and providing extra redundancy. Some of the muon stations will also be equiped by the Gas Electron Multiplier (GEM) [79] technology, helping to extend the MS coverage up to the $|\eta| = 2.8$. Furthermore, all the MS sub-detectors will receive the readout electronics upgrade to satisfy requirements for improved accuracy in measuring and identifying muons.

In addition, two entirely new detector systems will be introduced. For the identification of particles coming from interesting vertices, time of flight information will be provided by a Minimum Ionizing Particle Timing Detector (MTD) [175]. This information is particularly useful in selecting particles coming from the primary collision and identifying those originating in pile-up. Second addition is a new luminometer [176], a device designed to measure luminosity with improved accuracy.

Numerous research centers and laboratories across the globe have pledged their resources, person-power, and time to ensure the timely completion of the CMS Phase-2 upgrade prior the HL-LHC commissioning. Among these institutions, the Belgian consortium of universities (henceforth acronymed as "Belgian consortium"), consisting of Vrije Universiteit Brussel (VUB), Université Libre de Bruxelles (ULB), Université Catholique de Louvain (UCL), Universiteit Gent (UG) and Universiteit Antwerpen (UA), is actively contributing to the development and construction of the CMS Phase-2 Outer Tracker. The particular initiative of the Belgian consortium includes establishing a dedicated assembly center for the so called 2S-modules. Further details regarding the design of the 2S-modules are discussed in Section B.1. Both prototyping and full-scale production assembly necessitate comprehensive testing of module functionality. My personal contribution to this project, as summarised in Section B.2 and B.3, involved an implementation

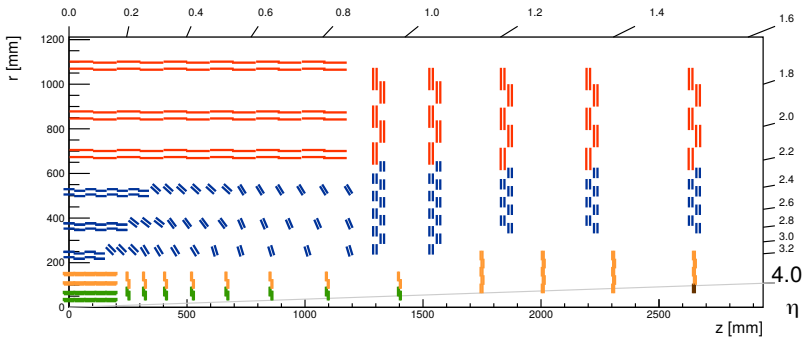

Figure B.2: One quarter of the CMS Phase-2 Tracker layout in the $r - z$ plane. The CMS Inner Tracker layers and discs are shown in green and orange. In the CMS Outer Tracker, PS-modules are shown in blue and 2S-modules are depicted in red. Source [171].

of the software control for the test apparatus capable of performing $\mathrm{I} - \mathrm{V}$ tests for the 2S-modules at various stages during the assembly. I have also participated in conducting these tests during the early prototyping, thus contributing to refining the testing procedures for the anticipated accelerated production phase.

B.1 A new tracker design

The layout of the planned CMS Phase-2 Tracker is depicted in Figure B.2. Conventionally, it comprises a barrel section, constructed from concentric cylidrical layers, and two end-cap sections on each side of the barrel, composed of discs.

The initial four innermost barrel layers constitute the CMS Inner Tracker (IT) barrel section. Thanks to advancements in pixel sensor technology, as detailed in [170], the IT is expected to provide three-dimensional hit measurements (i.e. including the z-coordinate) with an improved precision. The subsequent six barrel layers form the CMS Outer Tracker (OT) barrel section. The OT barrel layers are made of modules, each consisting of two silicon sensors separated only by a few millimeters. This design aims to improve the capability of the CMS Tracker to differentiate between high and low-p_{T} tracks as further discussed below.

Two types of the two-layer modules are considered for the installation in the CMS Outer Tracker. The first three barrel layers closer to the CMS center will comprise modules made of one pixelated sensor and one micro-strip sensor, denoted as PS-modules. The PS-module design is deferred for discussion in the technical report [171]. The subsequent three barrel layers will consist of modules made of two micro-strip sensors, also known as 2S-modules. Similar arrangement of the modules in the radial direction from the beamline is planned for the end-cap sections of the CMS Phase-2 Tracker, as shown in Figure B.2.

A sketch of the 2S-module layout is presented in Figure B.3. The readout for these modules is facilitated by the CMS Binary Chip (CBC) with 254 readout channels connected to the strips on the top and bottom sensors. A total of 8 CBCs are mounted onto each of the two front-end hybrids, which are laminated on a baseplate made from Carbon Fibre Reinforced Polymer (CFRP). These CFRP baseplates are then glued to the supporting structure made of the Carbon Fibre Reinforced Aluminium (Al-CF), which also holds two micro-strip silicon sensors (n-type strips in p-type silicon substrate) and pro-

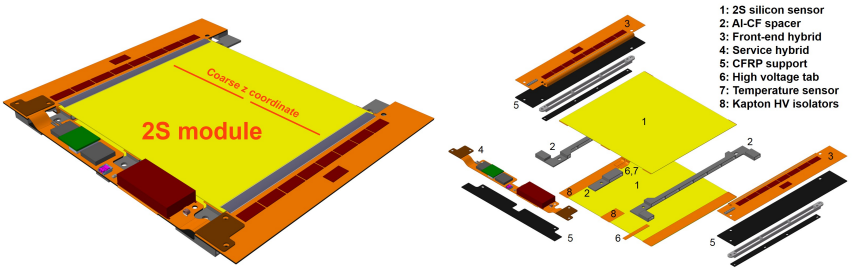

Figure B.3: The sketch of the assembled 2S-module (left) and with all of its components superimposed (right). The 2S modules are intended for installation in the CMS Outer Tracker. Source [167].

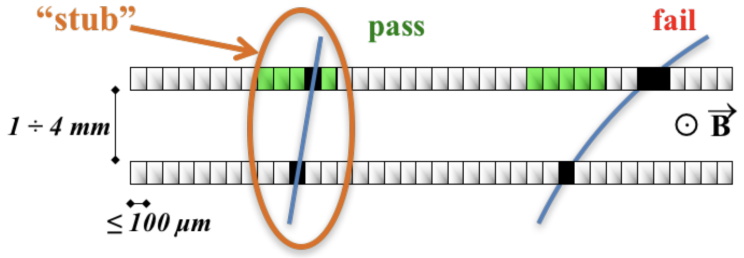

Figure B.4: Sketch of the "stub" formation in the new tracker module with two parallel sensors. In the presence of the magnetic field, the low-p_{T} tracks bend significantly more than the high-p_{T} tracks (in the plane orthogonal to the magnetic field). Two hits are matched only if they fit into the predefined window depending on the track momentum threshold. Source [167].

vides a necessary space gap between them. The single service hybrid is attached to the third side of the 2S-module, following the same integration procedure as with the front-end hybrids. It provides the necessary interface for module power supply, output data serialisation, and consequent opto-electrical conversion. The kapton High-Voltage (HV) isolation stripes, glued to the backside of the sensors, ensure electrical isolation between the Al-CF spacers and the sensors themselves.

Assuming the presence of the CMS magnetic field of 3.8 T, the trajectories of charged particles will be bent in the transverse plane, i.e. the plane orthogonal to the direction of the beam. The curvature of these tracks depends on the particles' momenta. As such, hits recorded in the bottom sensor can only be matched to the hits in the top sensor if they are localised within a predefined window. Once two hits are matched, they form a short segment of the full track (see Figure B.4), denoted as a stub. The stub-finding algorithm is performed throughout the detector layers, and these are then matched in the reconstruction of the global track information at the level of the L1 trigger using a farm of the FPGAs (i.e. programmable integrated circuits).

At various stages throughout the assembly process, the 2S-modules need to undergo various tests to assess their functionality and performance. Naturally, the number of stages requiring testing is reduced towards the full-scale production phase. Some of the tests performed at the Belgian consortium assembly center in Brussels will be discussed in the following section.

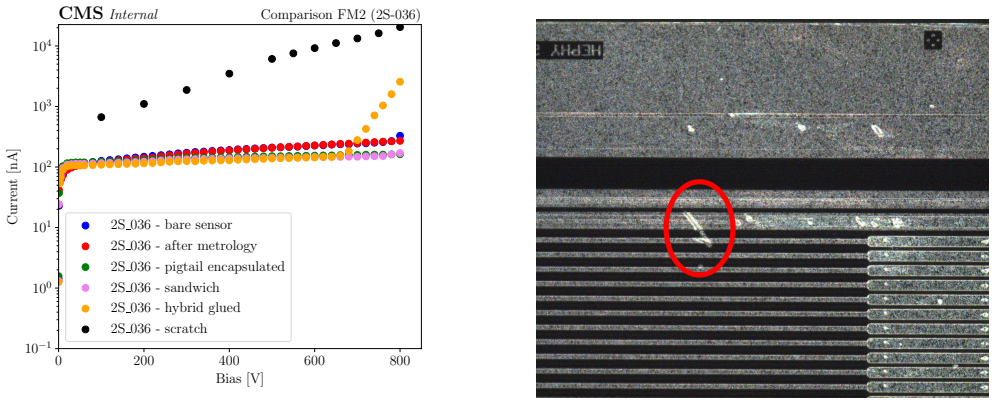

Figure B.5: The I − V diagnostics for the 2S sensor number 036 at various stages of assembly of Functional Module 2 (left) and the accidental scratch on this sensor highlighted in the red circle (right). An electrical short between the bias ring and the topmost strip is the most probable cause of the premature current breakdown.

B.2 Monitoring the 2S-modules during the prototyping

One of the key health indicators for a silicon sensor is the level of noise obscuring the measured signal from incoming high-energy particles. The lower the noise, the better the performance of the silicon sensors.

Some possible noise sources include disturbances caused by readout electronics, thermal noise (i.e. the random motion of electrons within the sensor's material), and various environmental effects. Another important characteristic of the sensor that could negatively influence the measured noise levels is leakage current. This refers to the residual current flowing through the silicon detector as a voltage bias is applied to the sensor's terminals. Its origin can vary from small imperfections in the silicon crystal lattice to surface contamination, and more serious issues like electrical shorts between strips in the sensor's active area caused by mechanical damage.

Leakage current is measured as a function of the applied bias, also known as Current − Voltage or I − V diagnostics. Performing this test after all the significant steps in the assembly process during the prototyping phase served as the first indicator for possible issues that might have occurred. A series of I − V measurements, as performed for the 2S sensor prototype number 036 during the assembly of Functional Module 2 (FM2), is shown in chronological order (top to bottom) in Figure B.5 (left). Note that in our case with n-in-p sensor type, we are interested in the reverse bias region (i.e. negative voltage is applied to the n-type side of the sensor and positive voltage is applied to the p-type side).

A typical example of the I − V curve for a well-behaved 2S silicon sensor can be seen in the first four assembly steps (in the plot marked as blue, red, green and pink). The whole curve can be categorised into three characteristic regions:

- Bias < X, where X varies with the sensor's design – in our case, it is roughly 50 V. The leakage current rises due to the small amount of charge carriers present in the sensor's depletion zone (mostly thermally generated).

- X < Bias < Y, where in our case, Y is at least 800 V. Electron-hole pairs in the sensor's depletion zone recombine at a rate comparable to the rate at which new charge carriers are created. As such, the leakage current remains almost constant. Typically, the bias point slightly above the value of X is chosen as the operational bias voltage for the sensor.

- Bias > Y. As the bias voltage rises, the electric field becomes strong enough to increase the kinetic energy of the charge carriers to high values. These energetic carriers might induce additional electron-hole production, e.g. through ionisation. This phenomenon eventually leads to the avalanche breakdown, causing the exponential growth of the leakage current.

The 2S silicon sensors were designed to withstand the reverse bias voltage up to 800 V without experiencing breakdown. Specifically, the leakage current at 800 V must remain below $3/2\times$ the leakage current at 600 V for the sensor to be in the best condition. On the other hand, unintended mechanical damage, such as the accidental scratch shown in Figure B.5 (right), can lead to premature sensor breakdown, as demonstrated in Figure B.5 (left) in black. Some level of premature breakdown at ≈ 700 V can also be seen in the measurement after the service hybrids were glued (orange), however, no particular cause of this breakdown has been localized.

Performing $\mathrm{I}-\mathrm{V}$ diagnostics during the initial prototyping phase, encompassing the assembly of the first five functional modules, has helped in recognising critical (or rather not so critical) operations in the assembly process, led to improved procedures in physically handling sensors and fine-tuned the testing process itself. The testing apparatus used to perform $\mathrm{I}-\mathrm{V}$ tests was developed at the Belgian consortium assembly center and will be presented in the following section.

B.3 The testing apparatus

A photo of the Test-box apparatus as placed on top of the anti-vibration table within the Belgian consortium clean-room is shown in Figure B.6 on the left. An example of the FM2 prepared for $\mathrm{I}-\mathrm{V}$ diagnostics is presented on the right within the same figure. The Test-box setup can be described as follows:

- The commercial multimeter Keithley 2470 is used to measure leakage current across the bias range of $[-800, 0]$ V. Additionally, this multimeter serves as the primary HV source. Safety measures have been implemented: an interlock connecting to the mechanical lock on the Test-box lid ensures that the Keithley cannot activate the source output while the box is open. This is assured both at the firmware and software level.

- The external HV source unit NHQ201 is kept as a back-up in case of Keithley malfunctions.

- While not visible in the photo, a brass pad establishes a conductive link to the sensor's rear during the initial assembly phase when the sensor is not yet secured in a support frame. Subsequently, as assembly progresses, four columns at each corner around the pad provide mounting points for the frame. The pad's vertical position is controlled by the motorized motion controler ESP100.

- To enhance stability and maintain the sensor's fixed position, localized vacuum is generated between the sensor and the pad using the vacuum pump.

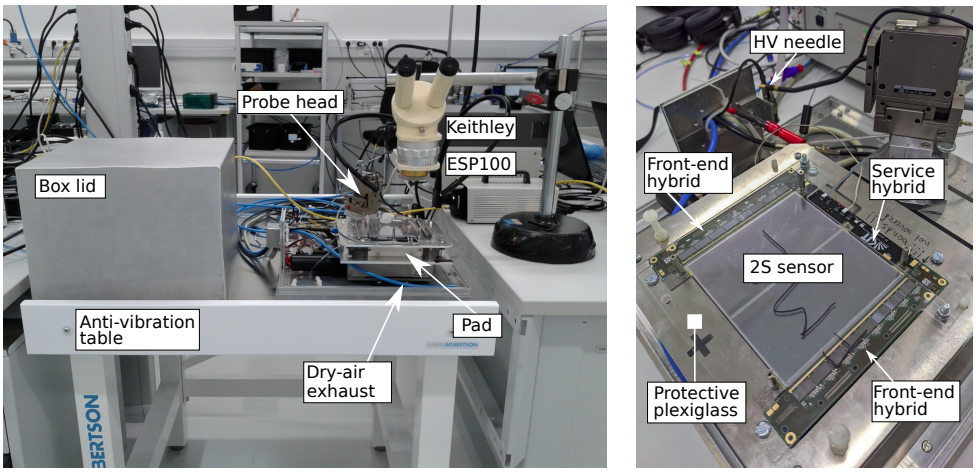

Figure B.6: The photo of the Test-box setup situated on top of the anti-vibration table (left) and a detail on the Functional Module 2 prepared for the $I-V$ diagnostics (right).

- Before the service hybrid is integrated into the module, a negative voltage bias is applied to the sensor's bias ring through a sharp needle securely positioned in the probe head. Adjustments to the needle's xyz-coordinates are made using sensitive springs manipulated by the operator's hand. The needle's head position is monitored with a microscope. Once the service hybrid is in place, voltage is conveyed to the sensor via integrated HV terminals.

- Given the extreme sensitivity of silicon detectors to light, the $I-V$ test is conducted under complete darkness ensured by the box lid. The ambient light sensor VEML7700 is used to monitor lighting levels.

- Maintaining the proper module functionality requires that silicon sensors be operated and stored in an environment with low Relative Humidity (RH). In particular, the RH level of $< 11\%$ is required for a time when bias voltage is applied. The room's air temperature remains consistent during measurement. Environmental conditions both before and during measurement are monitored using the HTS221 (humidity) and AT30SE75x (temperature) probes.

The particular measurement procedure for $I-V$ diagnostics depends on the stage of the module in the assembly process. Measurement is designed for a minimal number of operator interventions, however, certain manipulation with sensors/modules is required. These actions include: optical inspection (using a microscope) before and after the measurement, placement of the sensor onto the brass pad or mounting the sensor's support frame, establishing conductive contact with a tip of the needle or connecting the HV to integrated terminals, closing the Test-box and initiating automated measurements, opening/closing the dry air valve and turning on/off the vacuum pump, removing the module/sensor from the box.

On the other hand, dedicated Python-based Software (SW) developed in house, referred to as "mkMeasure", takes control over remaining steps. The following sequence of SW routines outlines the automated aspects of the measurement procedure:

1. A two-way connection is established with all involved devices within this setup.

Supported communication protocols include RS232, I2C, USB and GPIB. Initialization commands are sent to these devices.

2. The ESP100 receives command to send the brass pad to the operational height. The mkMeasure waits for the operator's confirmation of successful positioning and secure attachment of the sensor/module, along with the establishment of bias connections, e.g. connecting needle to the bias ring.

3. The connection is confirmed from the KEITHLEY readings (i.e. negligible but distinguishable current value is measured for an unbiased closed circuit as compared to an open one).

4. Upon confirmed connection, the ESP100 receives command to lower the brass pad to the scratch-safe position. The operator is prompted to close the Test-box lid.

5. The mkMeasure monitors the interlock status (signaling the box is closed) and required environmental conditions (light level, relative humidity and temperature).

6. Once the required environmental conditions are met, the brass pad is raised to the operational position, and needle connection is re-established under vibration-free conditions. The HV source is turned on.

7. One by one, the HV source output is set to the predefined bias values. The mkMeasure observes the voltage ramp-up time, then command is sent to the KEITHLEY to readout leakage current. In the event of current breakdown, measurement is aborted either at the firmware, software or operator intervention level (whichever is triggered first). Furthermore, operator can terminate measurement at any moment during the mkMeasure run-time by pressing designated key combination (CTRL+C).

8. At the end of measurement or upon receiving an abort signal, the HV source is turned off and brass pad returns to its home position. Communication channels with all involved devices are closed.

9. Collected data (including bias, current and environmental readings) are transformed to the desired format, suitable for database upload, and visualised for immediate inspection.

One notable advantage of the mkMeasure's code architecture, as visualised in the simplified scheme in Figure B.7, lies in its modularity. The "mother class", Device.py, defines all measurement types, e.g. $I-V$ test, constant bias, standalone environmental readings, etc., using universal command aliases tailored specifically for the mkMeasure SW. These are consequently translated using the dictionary of device-specific commands provided by a manufacturer. Versatility of such a code architecture was tested on the fly when an older KEITHLEY model was exchanged for the KEITHLEY 2470, and on multiple occasions when the firmware version operating the environmental probes was updated. In all cases, changes to the code were minimal and only required an exchange of the device-specific dictionary of commands. The modularity of the mkMeasure is also reflected in a possibility to incorporate new communication protocols maintained by the SerialConnector.py and SocketConnector.py classes. The mkMeasure deployment has been tested on CentOS, Kubuntu and Ubuntu Operating Systems (OS) and is not intended for deployment on other than Linux-based OS. The mkMeasure is an open source project under MIT license with a full license terms specified at project's repository [177].

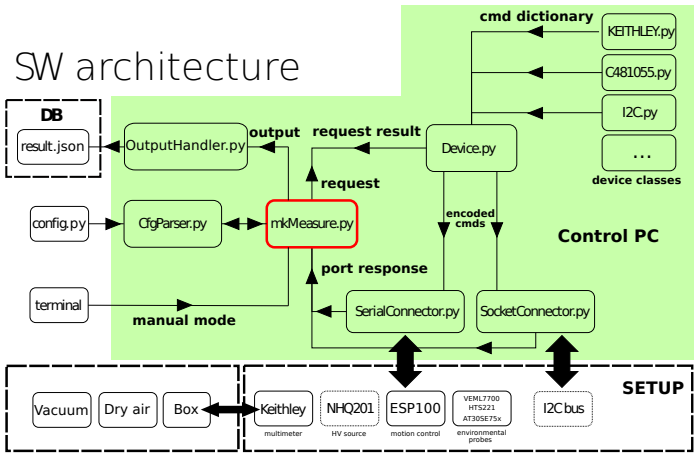

Figure B.7: The simplified scheme of the mkMeasure software architecture in the scope of the Test-box setup.

LIST OF ABBREVIATIONS

AC Anomalous Coupling

ACs Anomalous Couplings

AK Anti-kT

ALICE A Large Ion Collider Experiment

ATLAS A Toroidal LHC ApparatuS

Al-CF Carbon Fibre Reinforced Aluminium

BDT Boosted Decision Tree

BEH Brout-Englert-Higgs

BOOSTER Circular Proton Synchrotron Booster

BPIX Barrel Pixel

BSM Beyond the Standard Model

BTV POG b-tag & vertexing POG

CBC CMS Binary Chip

CBCs CMS Binary Chips

CFRP Carbon Fibre Reinforced Polymer

CHS Charged Hadron Subtraction

CKF Combinatorial Kalman Filter

CL Confidence Level

CMS Compact Muon Solenoid

CMSSW CMS SoftWare

CR Control Region

CRs Control Regions

CSC Cathode Strip Chamber

DMR Distributions of Median Residuals

DT Drift Tubes

DY Drell-Yan

EB ECAL Barrel

EC ECAL Endcaps

ECAL Electromagnetic Calorimeter

EFT Effective Field Theory

EGM POG Electron-Gamma POG

EM Electromagnetic

EOY End of Year

ES Preshower detector

EWK Electroweak

FM2 Functional Module 2

FPIX Forward Pixel

FSR Final State Radiation

GEM Gas Electron Multiplier

ggF Gluon-gluon fusion

GSF Gaussian Sum Filter

GWS (model) proposed by Glashow, Weinberg, and Salam that unifies electromagnetic and weak interactions

HB HCAL Barrel

HCAL Hadron Calorimeter

HE HCAL Endcap

HEP High Energy Physics

HF HCAL Forward Calorimeter

HL-LHC High-Luminosity LHC

HLS High-Level Structures

HLT High Level Trigger

HO HCAL Outer Barrel

HV High-Voltage

HVV Higgs boson production and decay vertices with vector bosons V

ID Identification

IOV Interval of Validity

IOVs Intervals of Validity

IP Interaction Point

ISO Isolation

ISR Initial State Radiation

IT Inner Tracker

JEC Jet Energy Corrections

JER Jet Energy Resolution

JMR Jet Mass Resolution

JMS Jet Mass Scale

JetMET POG Jet and Missing E_{T} POG

KD Kinematic Discriminant

KDs Kinematic Discriminants

L1 Level-1 trigger

LHC Large Hadron Collider

LHCb LHCb detector

LO Leading Order

LS Long Shutdown

LUM POG Luminosity POG

MC Monte Carlo

ME Matrix Elements

MELA Matrix Element Likelihood Approach

MET Missing Transverse Momentum

MKD Multidimensional Kinematic Discriminant

MKDs Multidimensional Kinematic Discriminants

MLM Maximum-Likelihood Method

MPI Multiple-Parton Interaction

MS Muon System

MSB MS Barrel

MSE MS Endcap

MTD Minimum Ionizing Particle Timing Detector

MVA Multivariate Analysis

NLO Next-to-Leading Order

NNLO Next-to-NLO

NP Nuisance Parameter

NPs Nuisance Parameters

OS Operating Systems

OT Outer Tracker

PAG Physics Analysis Group

PAGs Physics Analysis Groups

PDF Parton Distribution Functions

PDF* Probability Density Function

PF Particle Flow

PF MET PF Missing E_{T}

PL Profile-Likelihood

POG Physics Object Group

POGs Physics Object Groups

POI Parameters Of Interest

pp proton-proton

pQCD Perturbative QCD

PS Proton Synchrotron

PS Parton Showers

PU Pile-up

PUPPI Pile-up Per Particle Identification

PUPPI MET PUPPI Missing E_{T}

PV Primary interaction Vertex

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

QFT Quantum Field Theory

QM Quantum Mechanics

RF Radio Frequency

RH Relative Humidity

RMS Root Mean Square

RPC Resistive Plate Chambers

SC Supercluster

SCs Superclusters

SF Scale Factors

SM Standard Model

SMEFT Standard Model Effective Field Theory

SPR Special Relativity

SPS Super Proton Synchrotron

SR Signal Region

SRs Signal Regions

SW Software

T0 Tier-0

TEC Tracker EndCaps

TIB Tracker Inner Barrel

TID Tracker Inner Disks

TOB Tracker Outer Barrel

UE Underlying Event

VBF Vector boson fusion

VH Vector-boson-associated production

WH Vector-boson-associated production for $V = W$

YETS Year-End Technical Stops

ZH Vector-boson-associated production for $V = Z$

BIBLIOGRAPHY

[1] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[2] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[3] R. A. Millikan. Einstein's photoelectric equation and contact electromotive force. *Phys. Rev.*, 7:18–32, Jan 1916.

[4] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008.

[5] T. Davídek and R. Leitner. *Elementární částice od prvních objevů po současné experimenty*. Physics textbook. Matfyzpress, Prague, Czech Republic, 2012.

[6] C. Galbraith, D. Burgard. Standard Model infographics. `https://cds.cern.ch/record/1473657/files/SMinfographic_image.png?subformat=`, 2012. General Photo. Accessed: 2022-12-20.

[7] J. Chadwick. Possible existence of a neutron. *Nature*, 129:312, 1932.

[8] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire. Detection of the Free Neutrino: a Confirmation. *Science*, 124:103 – 104, 1956.

[9] C. D. Anderson and S. H. Neddermeyer. Cloud chamber observations of cosmic rays at 4300 meters elevation and near sea-level. *Phys. Rev.*, 50:263–271, Aug 1936.

[10] Martin L. Perl. The Discovery of The Tau Lepton. *NATO Sci. Ser. B*, 352:277–302, 1996.

[11] G. Danby, J-M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger. Observation of high-energy neutrino reactions and the existence of two kinds of neutrinos. *Phys. Rev. Lett.*, 9:36–44, Jul 1962.

[12] DONUT Collaboration. Observation of tau neutrino interactions. *Physics Letters B*, 504(3):218–224, Apr 2001.

[13] The Super-Kamiokande Collaboration. Measurements of the solar neutrino flux from super-kamiokande's first 300 days. *Physical Review Letters*, 81(6):1158–1162, Aug 1998.

[14] SNO Collaboration. Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory. *Phys. Rev. Lett.*, 89:011301, Jun 2002.

[15] LHCb Collaboration. Observation of an exotic narrow doubly charmed tetraquark. *Nature Physics*, 18:751–754, Jul 2022.

[16] LHCb Collaboration. Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays. *Physical Review Letters*, 115(7), Aug 2015.

[17] Elliott D. Bloom et al. High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees. *Phys. Rev. Lett.*, 23:930–934, 1969.

[18] Martin Breidenbach, Jerome I. Friedman, Henry W. Kendall, Elliott D. Bloom, D. H. Coward, H. C. DeStaebler, J. Drees, Luke W. Mo, and Richard E. Taylor. Observed behavior of highly inelastic electron-proton scattering. *Phys. Rev. Lett.*, 23:935–939, 1969.

[19] M Gell-Mann. The eightfold way: A theory of strong interaction symmetry. 3 1961.

[20] G Zweig. An SU3 model for strong interaction symmetry and its breaking; Version 2. 1964.

[21] Luciano Maiani. The GIM Mechanism: origin, predictions and recent uses. 2013. Comments: Opening Talk, Rencontres de Moriond, EW Interactions and Unified Theories, La Thuile, Valle d'Aosta, Italia, 2-9 March, 2013.

[22] J. J. Aubert et al. Experimental Observation of a Heavy Particle J. *Phys. Rev. Lett.*, 33:1404–1406, 1974.

[23] J. E. Augustin et al. Discovery of a Narrow Resonance in e^+e^- Annihilation. *Phys. Rev. Lett.*, 33:1406–1408, 1974.

[24] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.

[25] CDF Collaboration. Observation of top quark production in $p - \bar{p}$ collisions with the collider detector at fermilab. *Physical Review Letters*, 74(14):2626–2631, Apr 1995.

[26] D0 Collaboration. Observation of the top quark. *Physical Review Letters*, 74(14):2632–2637, Apr 1995.

[27] Arthur H. Compton. A quantum theory of the scattering of x-rays by light elements. *Phys. Rev.*, 21:483–502, May 1923.

[28] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV. *Phys. Lett. B*, 122:103–116, 1983.

[29] M. Banner et al. Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider. *Phys. Lett. B*, 122:476–485, 1983.

[30] TASSO Collaboration. Evidence for planar events in e+e− annihilation at high energies. *Physics Letters B*, 86(2):243–249, 1979.

[31] Particle Data Group et al. Review of particle physics. *Progress of Theoretical and Experimental Physics*, 2022(8), 08 2022. 083C01.

[32] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 10 2016.

[33] CMS Collaboration. Constraints on the spin-parity and anomalous hvv couplings of the higgs boson in proton collisions at 7 and 8 TeV. *Physical Review D*, 92(1), Jul 2015.

[34] ATLAS Collaboration. Determination of spin and parity of the higgs boson in the $ww^* \to e\nu\mu\nu$ decay channel with the ATLAS detector. *The European Physical Journal C*, 75(5), May 2015.

[35] Feynman Richard P. Robert B. Leighton and Matthew L. Sands. *The Feynman Lectures on Physics. New millennium ed.* Physics textbook. Basic Books a member of the Perseus Books Group, New York, NY, 2011.

[36] Mark Thomson. *Modern particle physics.* Physics textbook. Cambridge University Press, New York, NY, 2013.

[37] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in β Decay. *Phys. Rev.*, 105:1413–1414, 1957.

[38] CMS Collaboration. Constraints on anomalous Higgs boson couplings from its production and decay in the WW channel. Technical report, CERN, Geneva, 2023. `http://cds.cern.ch/record/2871687`.

[39] Jiri Horejsi. *Fundamentals of electroweak theory.* Physics textbook. Karolinum Press, Prague, Czech Republic, 2022.

[40] Chen-Ning Yang and Robert L. Mills. Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Phys. Rev.*, 96:191–195, 1954.

[41] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961.

[42] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[43] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968.

[44] J. Goldstone. Field Theories with Superconductor Solutions. *Nuovo Cim.*, 19:154–164, 1961.

[45] M. Baker and S. L. Glashow. Spontaneous breakdown of elementary particle symmetries. *Phys. Rev.*, 128:2462–2471, Dec 1962.

[46] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[47] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[48] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964.

[49] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[50] Peter W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.*, 145:1156–1163, 1966.

[51] T. W. B. Kibble. Symmetry breaking in non-abelian gauge theories. *Phys. Rev.*, 155:1554–1561, Mar 1967.

[52] John Ellis. Physics beyond the standard model. *Nuclear Physics A*, 827(1-4):187c–198c, Aug 2009.

[53] James D. Wells. *Effective Theories of Classical Gravity*, pages 15–42. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[54] Aneesh V. Manohar. Effective field theories. In *Perturbative and Nonperturbative Aspects of Quantum Field Theory*, pages 311–362. Springer Berlin Heidelberg.

[55] Elizabeth E. Jenkins, Aneesh V. Manohar, and Michael Trott. Naive dimensional analysis counting of gauge theory amplitudes and anomalous dimensions. *Physics Letters B*, 726(4-5):697–702, Nov 2013.

[56] A. Pich. Effective field theory. 1998.

[57] Gerard 't Hooft. Dimensional regularization and the renormalization group. *Nucl. Phys. B*, 61:455–468, 1973.

[58] William A. Bardeen, A. J. Buras, D. W. Duke, and T. Muta. Deep-inelastic scattering beyond the leading order in asymptotically free gauge theories. *Phys. Rev. D*, 18:3998–4017, Dec 1978.

[59] Steven Weinberg. Baryon and Lepton Nonconserving Processes. *Phys. Rev. Lett.*, 43:1566–1570, 1979.

[60] Landon Lehman. Extending the standard model effective field theory with the complete set of dimension-7 operators. *Physical Review D*, 90(12), Dec 2014.

[61] B. Grzadkowski, M. Iskrzyński, M. Misiak, and J. Rosiek. Dimension-six terms in the standard model lagrangian. *Journal of High Energy Physics*, 2010(10), Oct 2010.

[62] Rick S. Gupta, Alex Pomarol, and Francesco Riva. Bsm primary effects. 2014.

[63] Adam Falkowski and Adam Falkowski. Higgs Basis: Proposal for an EFT basis choice for LHC HXSWG. 2015.

[64] CMS Collaboration. Constraints on anomalous higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state. *Phys. Rev. D*, 104:052004, Sep 2021.

[65] E. Mobs. The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018. `https://cds.cern.ch/record/2636343`, 2018. General Photo. Accessed: 2022-12-20.

[66] L. Evans and P. Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, Aug 2008.

[67] R. Steerenberg et al. Operation and performance of the CERN Large Hadron Collider during proton Run 2. page MOPMP031, 2019.

[68] I. Neutelings. CMS coordinate system with a cylindrical detector. `https://tikz.net/axis3d_cms/`, 2021. General Photo. Accessed: 2022-12-20.

[69] T. Sakuma and T. McCauley. Detector and event visualization with sketchup at the cms experiment. *Journal of Physics: Conference Series*, 513(2):022032, Jun 2014.

[70] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. 2006.

[71] L. Viliani. CMS tracker performance and readiness for LHC Run II. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 824:67–69, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[72] B. Akgün. Performance of the CMS phase 1 pixel detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 924:262–264, 2019. 11th International Hiroshima Symposium on Development and Application of Semiconductor Tracking Detectors.

[73] The Tracker Group of the CMS Collaboration. The CMS phase-1 pixel detector upgrade. *Journal of Instrumentation*, 16(02):P02027, Feb 2021.

[74] CMS Collaboration. Strategies and performance of the CMS silicon tracker alignment during LHC Run 2. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1037:166795, 2022.

[75] A. Benaglia. The CMS ECAL performance with examples. *JINST*, 9:C02008, 2014.

[76] CMS Collaboration. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 15(05):P05002, May 2020.

[77] CMS Collaboration. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. *Journal of Instrumentation*, 8(11):P11002–P11002, Nov 2013.

[78] N. Pozzobon. The CMS Muon System: performance during the LHC Run-2. *Journal of Instrumentation*, 14(11):C11031, Nov 2019.

[79] A. Colaleo, A. Safonov, A. Sharma, and M. Tytgat. CMS Technical Design Report for the Muon Endcap GEM Upgrade. Technical report, 2015.

[80] M. Tosi. The CMS trigger in Run 2. *PoS*, EPS-HEP2017:523, 2017.

[81] CMS Collaboration. Commissioning of CMS HLT. *Journal of Instrumentation*, 5:T03005, 03 2010.

[82] Andy Buckley, Jonathan Butterworth, Stefan Gieseke, David Grellscheid, Stefan Höche, Hendrik Hoeth, Frank Krauss, Leif Lönnblad, Emily Nurse, Peter Richardson, Steffen Schumann, Michael H. Seymour, Torbjörn Sjöstrand, Peter Skands, and Bryan Webber. General-purpose event generators for LHC physics. *Physics Reports*, 504(5):145–233, 2011.

[83] CMS Collaboration. CMS SoftWare. `https://github.com/cms-sw/cmssw`, 2022. Software. Accessed: 2022-12-20.

[84] T Gleisberg, S Höche, F Krauss, M Schönherr, S Schumann, F Siegert, and J Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(02):007–007, Feb 2009.

[85] T. Kinoshita. Mass singularities of Feynman amplitudes. *J. Math. Phys.*, 3:650–677, 1962.

[86] T. D. Lee and M. Nauenberg. Degenerate Systems and Mass Singularities. *Phys. Rev.*, 133:B1549–B1562, 1964.

[87] John C. Collins and Davison E. Soper. Parton Distribution and Decay Functions. *Nucl. Phys. B*, 194:445–492, 1982.

[88] Richard D. Ball, , Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, and Maria Ubiali. Parton distributions for the LHC run II. *Journal of High Energy Physics*, 2015(4), Apr 2015.

[89] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Patrick Groth-Merrild, Alberto Guffanti, Nathan P. Hartland, Zahari Kassabov, José I. Latorre, Emanuele R. Nocera, Juan Rojo, Luca Rottoli, Emma Slade, and Maria Ubiali. Parton distributions from high-precision collider data. *The European Physical Journal C*, 77(10), Oct 2017.

[90] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *Journal of High Energy Physics*, 2004(11):040, Dec 2004.

[91] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007(11):070, Nov 2007.

[92] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6), Jun 2010.

[93] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), Jul 2014.

[94] Sara Bolognesi, Yanyan Gao, Andrei V. Gritsan, Kirill Melnikov, Markus Schulze, Nhan V. Tran, and Andrew Whitbeck. Spin and parity of a single-produced resonance at the LHC. *Physical Review D*, 86(9), Nov 2012.

[95] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2):31–145, 1983.

[96] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.

[97] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018–P09018, Sep 2020.

[98] S. Agostinelli et al. GEANT4–a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[99] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. 4 2009.

[100] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987.

[101] Susanna Cucciarelli, Marcin Konecki, Danek Kotlinski, and Teddy Todorov. Track reconstruction, primary vertex finding and seed generation with the Pixel Detector. 2006.

[102] W. Adam, B. Mangano, T. Speer, and T. Todorov. Track reconstruction in the CMS tracker. 12 2005.

[103] Wolfgang Waltenberger, Rudolf Frühwirth, and Pascal Vanlaer. Adaptive vertex fitting. *Journal of Physics G: Nuclear and Particle Physics*, 34(12):N343, nov 2007.

[104] W Erdmann. Vertex reconstruction at the CMS experiment. *Journal of Physics: Conference Series*, 110(9):092009, May 2008.

[105] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, Oct 2017.

[106] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, Sep 1975.

[107] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 13(06):P06015–P06015, Jun 2018.

[108] 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. *The European Physical Journal C*, 81(06):378, Apr 2021.

[109] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9–N20, Jul 2005.

[110] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, May 2021.

[111] Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(08):P08010–P08010, Aug 2015.

[112] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, Apr 2008.

[113] CMS Collaboration. Pileup mitigation at cms in 13 tev data. *Journal of Instrumentation*, 15(09):P09018, Sep 2020.

[114] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, Feb 2017.

[115] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, nov 2011.

[116] 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.

[117] CMS Collaboration. Identification techniques for highly boosted w bosons that decay into hadrons. *Journal of High Energy Physics*, 2014(12), Dec 2014.

[118] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft drop. *Journal of High Energy Physics*, 2014(5), May 2014.

[119] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018–P09018, Sep 2020.

[120] T. P. Wangler and K. R. Crandall. Beam halo in proton linac beams. *eConf*, C000821:TU202, 2000.

[121] Common analysis object definitions and trigger efficiencies for the $H \to WW$ legacy analysis with full Run-II data. *CMS Analysis Note*, CMS-AN-2019/125, 2020.

[122] Yanyan Gao, Andrei V. Gritsan, Zijin Guo, Kirill Melnikov, Markus Schulze, and Nhan V. Tran. Spin determination of single-produced resonances at hadron colliders. *Physical Review D*, 81(7), Apr 2010.

[123] Ian Anderson, Sara Bolognesi, Fabrizio Caola, Yanyan Gao, Andrei V. Gritsan, Christopher B. Martin, Kirill Melnikov, Markus Schulze, Nhan V. Tran, Andrew Whitbeck, and Yaofu Zhou. Constraining anomalous hvv interactions at proton and lepton colliders. *Physical Review D*, 89(3), Feb 2014.

[124] Andrei V. Gritsan, Raoul Röntsch, Markus Schulze, and Meng Xiao. Constraining anomalous higgs boson couplings to the heavy-flavor fermions using matrix element techniques. *Physical Review D*, 94(5), Sep 2016.

[125] CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *The European Physical Journal C*, 76(3), Mar 2016.

[126] 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.

[127] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. Finite quark-mass effects in the nnlops powheg+minlo higgs generator. *Journal of High Energy Physics*, 2015(140), 2015.

[128] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. Hw±/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(83), 2013.

[129] CMS Collaboration. A measurement of the higgs boson mass in the diphoton decay channel. *Physics Letters B*, 805:135425, Jun 2020.

[130] CMS Collaboration. Measurements of properties of the higgs boson decaying to a w boson pair in pp collisions at 13 tev. *Physics Letters B*, 791:96–129, Apr 2019.

[131] J. M. Campbell and R. K. Ellis. Update on vector boson pair production at hadron colliders. *Physical Review D*, 60(11), Nov 1999.

[132] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *Journal of High Energy Physics*, 2011(7), Jul 2011.

[133] John M. Campbell, R. Keith Ellis, and Walter T. Giele. A multi-threaded version of mcfm. *The European Physical Journal C*, 75(6), 2015.

[134] Fabrizio Caola, Matthew Dowling, Kirill Melnikov, Raoul Röntsch, and Lorenzo Tancredi. QCD corrections to vector boson pair production in gluon fusion including interference effects with off-shell higgs at the LHC. *Journal of High Energy Physics*, 2016(7), Jul 2016.

[135] CMS Collaboration. An embedding technique to determine backgrounds in proton-proton collision data. *Journal of Instrumentation*, 14(06):P06032, Jun 2019.

[136] Michal Czakon, David Heymes, Alexander Mitov, Davide Pagani, Ioannis Tsinikos, and Marco Zaro. Top-pair production at the LHC through NNLO QCD and NLO EW. *Journal of High Energy Physics*, 2017(10), Oct 2017.

[137] Patrick Meade, Harikrishnan Ramani, and Mao Zeng. Transverse momentum resummation effects in w^+w^- measurements. *Physical Review D*, 90(11), Dec 2014.

[138] Prerit Jaiswal and Takemichi Okui. Explanation of the ww excess at the LHC by jet-veto resummation. *Physical Review D*, 90(7), Oct 2014.

[139] Yi-Song Lu, You-Kai Wang, and Xiang-Yuan You. Study of hzz anomalous couplings by angular differential cross sections, 2022.

[140] CMS Collaboration. Measurement of the higgs boson width and evidence of its off-shell contributions to zz production. *Nature Physics*, 18(11):1329–1334, November 2022. DOI: 10.1038/s41567-022-01682-0.

[141] CMS Collaboration. Constraints on anomalous hvv couplings from the production of higgs bosons decaying to τ lepton pairs. *Physical Review D*, 100(11), Dec 2019.

[142] Olaf Behnke, Kevin Kroninger, Gregory Schott, and Thomas Schorner-Sadenius. *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*. Wiley-VCH, Weinheim, Germany, 1st edition, 2013.

[143] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb 2011.

[144] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, 2e edition, 2006.

[145] F. James and M. Roos. Minuit - a system for function minimization and analysis of the parameter errors and correlations. *Computer Physics Communications*, 10(6):343–367, 1975.

[146] Procedure for the LHC Higgs boson search combination in Summer 2011. 8 2011.

[147] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *Eur. Phys. J. C*, 81(9):800, 2021.

[148] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. 2018.

[149] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. 2019.

[150] ATLAS Collaboration. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the atlas detector at the lhc. *Phys. Rev. Lett.*, 117:182002, Oct 2016.

[151] CMS Collaboration. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2018(7), Jul 2018.

[152] Giampiero Passarino. Higgs CAT. *The European Physical Journal C*, 74(5), May 2014.

[153] 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.

[154] 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.

[155] 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, 2021.

[156] Benedetta Camaiani, Roberto Seidita, Lucio Anderlini, Rudy Ceccarelli, Vitaliano Ciulli, Piergiulio Lenzi, Mattia Lizzo, and Lorenzo Viliani. Model independent measurements of standard model cross sections with domain adaptation. *The European Physical Journal C*, 82(10):921, 10 2022.

[157] 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.

[158] 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.

[159] Volker Blobel and Claus Kleinwort. A New method for the high precision alignment of track detectors. In *Conference on Advanced Statistical Techniques in Particle Physics*, 6 2002.

[160] V. Karimaki, T. Lampen, and F. P. Schilling. The HIP algorithm for track based alignment and its application to the CMS pixel detector. 1 2006.

[161] CMS Collaboration. TkAlignment HipPy condor submission. `https://github.com/cms-sw/cmssw/pull/27092`, 2019. Software. Accessed: 2023-07-19.

[162] CMS Collaboration. Introduction of New-All-In-One tool for Offline Validation of TrkAlignment. `https://github.com/cms-sw/cmssw/pull/38304/`, 2022. Software. Accessed: 2023-07-19.

[163] Nazar Bartosik. *Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s}$ = 8 TeV as test of QCD and background tt+Higgs production.* PhD thesis, U. Hamburg, Dept. Phys., Hamburg, 2015.

[164] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009–P10009, Oct 2014.

[165] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, 1998.

[166] O. et al. Aberle. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report.* CERN Yellow Reports: Monographs. CERN, Geneva, 2020.

[167] CMS Collaboration. The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva, 2017.

[168] Andrea Dainese, Michelangelo Mangano, Andreas B Meyer, Aleandro Nisati, Gavin Salam, and Mika Anton Vesterinen. Report on the Physics at the HL-LHC, and Perspectives for the HE-LHC. Technical report, Geneva, Switzerland, 2019.

[169] CMS Collaboration. The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger. Technical report, CERN, Geneva, 2021. This is the final version of the document, approved by the LHCC.

[170] Stella Orfanelli. The Phase 2 Upgrade of the CMS Inner Tracker. Technical report, CERN, Geneva, 2020.

[171] Suvankar Roy Chowdhury. The Phase-2 Upgrade of the CMS Outer Tracker. Technical report, CERN, Geneva, 2020.

[172] The Phase-2 Upgrade of the CMS Endcap Calorimeter. Technical report, CERN, Geneva, 2017.

[173] The Phase-2 Upgrade of the CMS Barrel Calorimeters. Technical report, CERN, Geneva, 2017. This is the final version, approved by the LHCC.

[174] The Phase-2 Upgrade of the CMS Muon Detectors. Technical report, CERN, Geneva, 2017. This is the final version, approved by the LHCC.

[175] Collaboration CMS. A MIP Timing Detector for the CMS Phase-2 Upgrade. Technical report, CERN, Geneva, 2019.

[176] CMS Collaboration. The Phase-2 Upgrade of the CMS Beam Radiation Instrumentation and Luminosity Detectors. Technical report, CERN, Geneva, 2021. This is the final version, approved by the LHCC.

[177] Tomáš Kello. mkMeasure: Multi-device control console application. `https://github.com/TomasKello/mkMeasure/blob/main/LICENSE.md`, 2023. Software. Accessed: 2023-08-18.