Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN FISICA

Ciclo XXXIII

Settore Concorsuale: 02/A1

Settore Scientifico Disciplinare: FIS/01

Precision measurements of the Higgs boson properties: from the $H \to ZZ^* \to 4\ell$ analysis with CMS at the LHC to the future large lepton colliders

Presentata da:
Elisa Fontanesi

Coordinatore Dottorato
Prof. Michele Cicoli

Supervisore
Prof. Sylvie Braibant

Co-supervisore
Prof. Toni Šćulac

Esame finale anno 2021

Abstract

The discussion of this thesis is dedicated to the precision measurements of the Higgs (H) boson properties, spanning from the study of the H boson decaying into four leptons with the CMS experiment at the Large Hadron Collider (LHC) to the future investigation of the H boson sector at large lepton colliders. Precision measurements are intended to be a powerful experimental tool to further verify the standard model (SM) but also search for possible small deviations from the SM prediction induced by physics beyond the SM.
The $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ ($\ell = \mathrm{e}, \mu$) decay channel offers an optimal way to study the H boson profile, playing a central role since the time of the discovery. Properties of the H boson are measured exploiting the total amount of data produced in proton-proton collisions at the LHC during Run 2. The dataset collected by the CMS experiment at a center-of-mass energy of 13 TeV corresponds to an integrated luminosity of $137\ \mathrm{fb}^{-1}$. The inclusive signal strength modifier μ, defined as the ratio of the H boson rate in the 4ℓ channel to the SM expectation, is measured to be $\mu = 0.94 \pm 0.07\ (\mathrm{stat})\ ^{+0.09}_{-0.08}\ (\mathrm{syst})$ with the H boson mass fixed to $m_{\mathrm{H}} = 125.38\ \mathrm{GeV}$, the most precise measurement currently available. The signal strength modifiers for the main H boson production modes are also reported. In addition, results within the framework of the Simplified Template Cross Sections are produced to explore the different H boson production mechanisms in specific kinematic regions of the phase space. The inclusive fiducial cross section for the $\mathrm{H} \to 4\ell$ process is measured to be $\sigma_{\mathrm{fid.}} = 2.84\ ^{+0.23}_{-0.22}\ (\mathrm{stat})\ ^{+0.26}_{-0.21}\ (\mathrm{syst})\ \mathrm{fb}$ for $m_{\mathrm{H}} = 125.38\ \mathrm{GeV}$ with respect to the corresponding SM expectation of $2.85\ \pm\ 0.15\ \mathrm{fb}$. Finally, the differential cross sections as a function of the transverse momentum and rapidity of the four-lepton system, the transverse momentum of the leading jet, and the jet multiplicity are shown. All results are found to be in agreement with the SM predictions.
Along the way, the perspectives offered by future colliders in the exploration of the Higgs sector and an overview of the different projects of future machines and experiments are also presented. A large focus is reserved for the ongoing development of the design of a detector proposal called IDEA (Innovative Detector for Electron-positron Accelerators), specifically intended for the future H boson factory colliders. The IDEA detector concept has been accepted and included in both the Conceptual Design Reports of the the electron-positron Future Circular Collider (FCC-ee) and Circular Electron-Positron Collider (CEPC) projects. The first test beam of a full slice of the IDEA detector is described. Moreover, the current status of the IDEA simulation is presented, focusing on both the full (GEANT4-based) and fast (DELPHES-based) simulations, showing promising prospects for high precision H boson properties measurements.

Acknowledgements

While I was writing this thesis, I realized how many people I had chance to meet in the context of my work and how much I received from most of them along my three-year Ph.D. journey. This is why I would like to give thanks to some of them.

I would like to start with my colleagues from the CMS Collaboration, especially from the CJLST group. It was a pleasure to work with you in an extremely favourable environment where I had the opportunity to learn from scratch most of the scientific aspects discussed in this thesis. I "accidentally" joined you from Bologna after a Skype call with Roberto and you welcomed me from the first time. So thank you for this chance, Roberto, and also for your support in different occasions. A special thanks goes to Toni and Matteo, who I now consider my friends. I really hope that we can still work together in the future. Toni, you have invested so much time helping me to understand the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis, starting from the very basic aspects during a bus journey in Mexico and becoming my supervisor in the end. I really enjoyed working with you. Thank you for your optimism and availability, for having selflessly shared your knowledge and experience with patience and energy. You pushed me to the forefront whenever you could, always having faith in me. Matteo, thank you for your hard work, your "organized disorder", your humor and your enthusiasm, for having dedicated time for free to all my questions, for being a brilliant mind always ready to discuss physics and to share skills and for never failing to get a laugh. I would also like to mention Christophe, Giacomo, Marko, Nicola, and my colleagues of the Johns Hopkins University Andrei, Ulascan, Meng, Savvas, Jeff, and Lucas: thank you for our weekly meetings with a lot of physics discussions, for your dedication and collaboration. Besides CJLST people, I want to thank my BBF colleagues and, finally, Maria, for her generous support in various moments.

I would like to express my gratitude to all my colleagues from Bologna, starting with Sylvie and Paolo. You supervised my work constantly supporting me and offering me a lot of golden occasions to improve my scientific experience. Thank you for your help and advice throughout these years. A special mention for Carlo and your special help with the muon physics at CMS during this work of thesis, for Federica and Daniele and the time spent together in Geneva, for Fabrizio, Luigi, Alessandra, Sergio, and Francesca. Finally, I would like to express my gratitude to my beloved friends of the D040 office: Lisa, Fabio, and Leonardo. You have been with me in every single moment of my journey. We shared great moments together, not only in our research but also in the everyday life.

Coming to my INFN collegues who are working on the IDEA project, I would like to thank in particular Lorenzo, Massimiliano, Riccardo, and Gianfranco. We have met during exciting test beam days at CERN and this work together definitely made us friends. Thank you for teaching me a lot about the world of hardware and detectors: I have never seen an experimental area nor touched a detector before! Thanks also to Patrizia. Franco, Roberto, Massimo, Gabriella, Gianfranco, Francesco, and Gianni. Finally, a big thank to Michele for his guidance and help in solving problems while working with Delphes.

Thanks to my work at CERN over the last year and a half I met a lot of special friends who made my stay far from home very enjoyable. Please forgive me if I am missing someone! I would like to thank my friends from the volleyball team, in particular Eleonora, Gaia, and Alice, for being so bright people; my friends from the CERN lunches, especially Alberto, Federico, Nico, Dario, Chiara, Tommaso, who closely stood by me during the final rush of this work; my friends from the CERN teas, coffees, and breaks: Alessandra, Angela, Davide, Luca, Nadir, Marco, Simranjit, and Deborah; special people who shared with me the Sunday morning volunteering, in particular Simone, Federica, Gabriele, and Lucia.

And I would like to conclude with my housemates and friends of the special Pregnin house: you made me feel at home and you took care of me with your unconditional help and support. You have always found time to answer my stressing questions about physics but also to take my mind off science: Luca with his jazz music, piano, snowboard, oysters, and artichokes, and the charm, passion, and positivity that he radiates; Raffaele with his funny and frequent complaints about everything and his attention to everybody; Matteo with his crazy sports, his love for the mountains, and special cooking, hard like a stone but only as a first appearance; Enrico with his incredible laziness and his love for tv series, good food, and sofas; Andrea with his frequent and sincere laughs and passion for pizza and for the heaviest foods that can be imagined; Marco with his films and music and his dedication to the physics and the house; Lorena, with her welcoming approach to everyone, her special sweets, and her exceptional sensibility; Stefania, with a myriad of interests and her cheery attitude towards trash, her sweatness, and funny tendency to organize jokes. And I need to add the last arrivals in the house who soon became an important part of the group: Marta, with her love for fitness and healthy dinners and her contagious optimism, and Isabella, with her bike, her pictures, her passion for nature, and her uncontainable lovely personality. Together with you, I have learned how beautiful and stimulating it is to be different from one another and my personality grew a bit stronger.

And at the very end, I wish to thank my mum Patrizia, my brothers Stefano and Michele, my grandma Anna, and all my friends from Bologna, who are like my family, my home from home: Giovanni, Sofia and Sofia, Pepi, Erica and Matteo, Giacomo and Stefano, Maria, Giuseppe and Matilde and their wonderful family. I would like to mention also Marti, Giacomo and Maria, Luca and Pippo, Fiorenza, Agnese, Laura, Martina and Luigi, Virginia and Matteo, Francesca and Simone, Angela and Dario: even though I was far away, you are always there for me to share stories and, in general, our life. A special thank to my friends who shared with me last summer holiday in the amazing and unforgettable Carloforte, where the writing of this thesis had the first significant boost: Francesco, Enrica, Sara, Giuseppe, Matteo, Lucia, Tommaso, Alessandra, Tommaso, Giovanni, Stefano, and Sofia.
All of you, guys, have always supported me with curiosity, enthusiasm, joy, love, and wonder my journey in physics, being like my safe anchor.

Contents

Dedicated to my daddy Domenico,
who always supported my journey in physics
loving it more than anyone else

"Not all of us can do great things. But we can do small things with great love."

Introduction

The quest of describing the fundamental structure of matter and the history of the universe is a long and fascinating story. In the late 1960s, particle physicists organized their knowledge about the elementary constituents of matter and their fundamental interactions in an elegant theoretical framework called the standard model (SM).
The SM is totally based on the concepts of invariance and symmetry groups. The model describes the nature in terms of a few basic blocks, the elementary particles, governed by three out of the four fundamental forces, omitting gravity. The effect of gravity is so weak as to be negligible at the subatomic scale of particles. Since its original formulation, the SM has been tested in its predictions and limits to a remarkable accuracy exploiting different generations of experiments. Its great predictive power has been successfully demonstrated by several experimental confirmations in a wide range of phenomena, from the discovery of the $W^{\pm}$ and Z bosons by Carlo Rubbia and Simon van der Meer in 1983 at the CERN super proton synchrotron (SPS) to the observation of the top quark by the CDF and DØ Collaborations at the FNAL Tevatron. However, the postulated existence of a hidden symmetry of the electromagnetic and weak interactions spontaneously broken to provide all particles with a mass remained a mistery for many years.

The existence of a non-vanishing Higgs field in the vacuum responsible for the generation of the masses of elementary particles is postulated by the SM of particle physics. The excitation of this field is known as the Higgs (H) boson, which represents a unique particle related to profound questions about the fundamental laws of nature. The almost fifty-year long physics hunt to probe the mechanism that allows for all particles in the universe to obtain their mass reached a historical turning point on the 4^{th} July 2012 with the first observation of the H boson by the ATLAS and CMS Collaborations at the CERN LHC proton-proton collider. The H boson discovery is a milestone in the history of particle physics from both a theoretical and experimental point of view. It enshrined the complete affirmation of the SM theory and the extraordinary success of the accelerators, the experiments, and the Grid computing at CERN. It ended up with a Nobel Prize in physics, that was awarded jointly to P. W. Higgs and F. Englert in 2013. Soon after, further tests of the properties of the H boson were performed on the available Run 1 dataset at a center-of-mass energy of 7-8 TeV to confirm that it was a SM H boson with properties consistent with a minimal scalar sector.

The landscape of particle physics has been dramatically changed after the discovery. Since then, the precise measurement of the H boson properties, such as its width and the structure of its couplings to the other SM particles, represented the primary goal of the physics program of the ATLAS and CMS experiments. The essential role of the H boson in the acquisition of the mass of elementary particles has been established, but the observed pattern of masses is still an enigma. Actually, the SM fails to answer some crucial questions about the nature of the universe. Each answer implies more questions and nothing is ever completely done. This is the path of science.

The work presented in this document is located at this point of the story. The aim of this thesis is to dig into all aspects mentioned up to now having as a starting point the two pillars of my journey as a Ph.D. student: the study of the H boson properties in the four-lepton final state with the CMS experiment at the LHC and the study of the future of the Higgs sector at large lepton colliders.

During my research activity in the CMS Collaboration, I had the great opportunity to work on one of the most relevant final states to study the H boson properties. The golden channel $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ (ℓ = e, μ) is characterized by a large signal to background ratio and its final state particles can be reconstructed with an excellent resolution. I joined the CMS group dedicated to the study of the $\mathrm{H} \to 4\ell$ process at the time of the effort to analyze the total amount of data collected during the LHC Run 2 (2016-2018) at a center-of-mass energy of 13 TeV. I had a chance to contribute to many key areas of the analysis, for example performing studies on the physics objects needed as well as on the estimation and modelling of the reducible background.
With the full Run 2 LHC data, we are definitely entering the era of the H boson precision measurements. The discussion of this thesis will present an extensive set of measurements of the H boson properties which are going to be part of the most precise measurements of the last discovered elementary particle. Significant improvements in many aspects of the analysis have been introduced with respect to the previously reported measurements by the CMS Collaboration. Preliminary results were released during spring 2019 and the final ones will be published in a journal soon. At the time of concluding the writing of this thesis, the final steps of the review of the analysis are ongoing. I had the opportunity to follow them in detail, contributing to the preparation of the conclusive checks and results required before the internal approval and publication of the work. My contribution was recognized by the group with the invitation to present some of the latest results on the differential H boson cross section measurements by the ATLAS and CMS Collaborations at the Higgs2020 conference in October 2020 at the end of my Ph.D. activity.

Given that the precision study of the Higgs sector provides a powerful experimental tool to investigate the open questions about the physical laws that regulate our universe, it represents the primary target of any future collider facility post the LHC era. In the context of the animated discussion about the future of the high energy physics and the kind of future accelerators needed, a renewed interest for the electron-positron physics emerged in the particle physics community. Indeed, it may offer an unprecedented sensitivity to tests of the SM and signs of new physics appearing in the form of small deviations from the SM predictions.
Along the way, my interests brought me to study the perspectives offered by future colliders in the exploration of the Higgs sector, focusing on the large lepton colliders and on the development of a specific detector proposal called IDEA (Innovative Detector for Electron-positron Accelerators). I contributed to different aspects of the development of the IDEA detector concept, both from the hardware and the software point of view, in particular during the first part of my activity as a Ph.D. student. I was awarded with the possibility to present the status of the software simulation of the IDEA detector on behalf of the IDEA Proto-Collaboration in a conference which took place in Hong Kong in January 2020.

The work of this thesis is carried out on two main topics, enlighting the recent past and present status of the precision measurements of the H boson and the future of the particle physics research in the Higgs sector.
Chapter 1 is devoted to characterization of the H boson in the SM of elementary particles, including a discussion of the phenomenology of the H boson at the LHC, its discovery, and its experimental picture currently available. Moreover, an overview of the possible BSM effects in the H sector is provided and the role of precision measurements at future colliders

is presented.
Chapter 2 introduces the experimental context with a description of the LHC machine and the CMS experiment, reserving a special focus to the different proposals of new colliders and experiments for the future after the LHC.
The subsequent chapters describe my personal contribution to the subjects mentioned previously and summarize my activity during the three years of my Ph.D. journey. The main body of the thesis is dedicated to the $\mathrm{H} \to 4\ell$ analysis performed with the total amount of data collected during the LHC Run 2, trying to emphasize numerous improvements introduced after Run 1.
Chapter 3, 4, and 5 include the areas in which I have devoted most of my time in the $\mathrm{H} \to 4\ell$ analysis. Chapter 3 describes the reconstruction and identification algorithms used in this analysis for leptons, photons, and jets. The event selection and categorization of the data events are discussed in Chapter 4, together with the physics observables exploited to improve the separation between signal and backgrounds and to categorize events. The signal extraction and the background estimation are presented in Chapter 5. A presentation of the Monte Carlo simulated samples, generated using the state-of-the-art theoretical calculations for both the SM H boson signals and relevant background processes, is also provided. A special focus is put on the estimate of the reducible background for the $\mathrm{H} \to 4\ell$ process, that represents one of my main novel contributions to the analysis. It is followed by Chapter 6, which presents the measurements performed in the $\mathrm{H} \to 4\ell$ final state.
Finally, Chapter 7 discusses the future exploration of the Higgs sector and the status of the IDEA detector project.

Chapter 1

The Higgs boson in the standard model of elementary particles

As a starting point of this manuscript, which is going to discuss a mixture of aspects of the history, the present, and the future of the particle physics research, many questions addressing the peculiar effort in the field of high energy physics arise. What is the matter of our universe made of? How do elementary particles interact with each other? How are elementary particles produced and how can they be detected? What kind of information can be extracted by experimental discoveries and results?

The rules governing our daily world come from classical mechanics, but they need to be modified in order to describe very fast objects or physical systems at the smallest scales, resulting in special relativity and quantum mechanics, respectively. These two aspects can be coherently summarized in a unique formal and conceptual framework, the *Quantum Field Theory* (QFT), which represents the basis of the elementary particle physics. According to this mathematical model, the most fundamental entities are fields associated to particles. Indeed, particles are interpreted as a fluctuation in the corresponding field and the interactions between them as an exchange of a mediator virtual particle.
The formation of the standard model (SM) results from a long sequence of experimentations and brilliant ideas developed both in the theoretical and experimental particle physics. In the late 1960s all the known elementary constituents of matter, the particles, and the fundamental forces through which they interact were included in the SM, except for gravity. Science's path is often seen as a quest for unification in our understanding of the physical universe, attempting to provide a theoretical explanation of all the experimental phenomena.

In this first chapter, the minimal theoretical background of the SM of elementary particles as a renormalizable QFT and the inclusion of a scalar sector to justify the origin of the particle masses through the spontaneous symmetry breaking mechanism is provided. Then, the discussion will be extended to the phenomenology and experimental confirmation of the predicted Higgs (H) boson as the actor of this complex scalar field and as the most relevant missing piece of the SM, moving to the rich physics program oriented to the detailed study of the H boson.

1.1 Brout-Englert-Higgs mechanism in the standard model

The current section is going to describe the fundamental particles of the SM and introduces the concept of gauge invariance [1–4], in order to explain how the general principle of *local gauge invariance* is intimately connected to the interacting field theories and the fundamental interactions at the subatomic scale.

1.1.1 Elementary particles and fundamental interactions

A point-like physical system representing a structureless unit of matter is defined as a "fundamental particle", meaning that its substructure is impossible to be explored.

> "*We will not be able to give a final answer to the question of which particles are elementary until we have a final theory of force and matter.*"
> (S. Weinberg, 1996)

According to the current view of the SM theory, the whole universe is made out of twelve fundamental particles and their corresponding antiparticles, characterized by the same spin, mass, and mean lifetime as particles but having opposite charge. In addition, the mediators of interactions are included in the SM description of the elementary particles.
A schematic description of all the particles encapsulated in the SM is shown in Fig. 1.1.

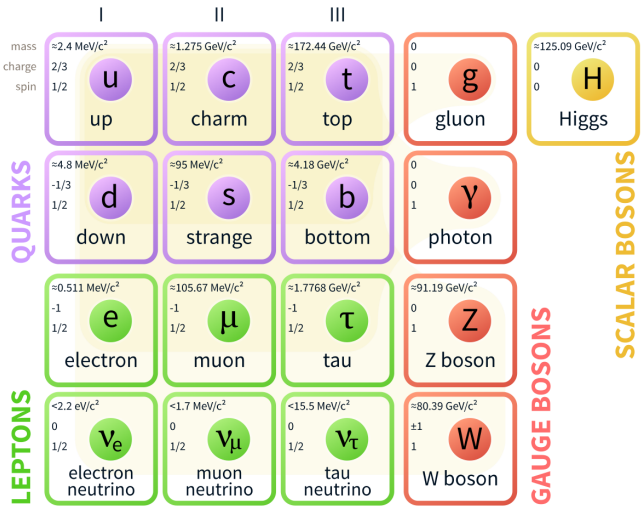


FIGURE 1.1: Fundamental particles of the standard model.

The organization of all fundamental particles within the SM reflects two remarkable ideas suggested during the 1950s and 1960s and developed later. From 1954, the Yang-Mills theories introduced the gauge symmetry idea. Then, the quark model was proposed in 1964 independently by M. Gell-Mann and G. Zweig.
The description of the matter is based on two groups of elementary particles according to their spin (s).

* *Fermions* have half-integer spin ($s = \frac{1}{2}$) and obey the Pauli exclusion principle. They are described by the Fermi-Dirac statistics and are classified into six leptons and six quarks. The projection of the spin along the particle momentum is defined as chirality, that is a relativistic invariant for fermions without mass. The two chiral components of a field, left and right, are expressed using the $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ matrix with the operators $\frac{1-\gamma^5}{2}$ and $\frac{1+\gamma^5}{2}$, respectively. The interaction between fermions and vector bosons preserves the chirality in the massless limit, where the chirality corresponds to the helicity.

* *Gauge bosons* are responsible for interactions between particles according to the QFT, that describes interactions as an exchange of a quantum of the associated field. They have integer spin ($s = 1$) and they follow the Bose-Einstein statistics.

Leptons include the electron (e), the muon (μ), and the tau (τ), with increasing mass values of about 0.5 MeV/c^2, 0.1 GeV/c^2, and 1.8 GeV/c^2, and their associated neutrinos (ν_e, ν_μ, ν_τ)[1]. The former carry a single unit of negative electric charge and interact in a weak or electromagnetic way, while the latter are electrically neutral and are sensitive only to the weak interaction. They are divided into three families according to their weak interaction properties, resulting in three weak-isospin doublets:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix} \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}. \tag{1.1}$$

The conservation of the lepton flavor is verified for each family, except for phenomena of neutrino oscillations, where only the total lepton number is conserved. A lepton quantum number (L) is associated to leptons, which is additive and assumes values 1 and -1 for particles and antiparticles, respectively. Analogously, a baryon quantum number (B) describes quarks, with a value equal to $\frac{1}{3}$ for quarks and -$\frac{1}{3}$ for antiquarks. The information about the quark mixing due to the flavor-changing weak interaction is contained in the Cabibbo–Kobayashi–Maskawa (CKM) matrix [5].

Quarks occur in six different flavors: up (u), charm (c), and top (t), carrying positive charge of $\frac{2}{3}$ units; down (d), strange (s), and bottom (b), carrying a negative charge of $\frac{1}{3}$ units. They are also organized into three generations resulting in the following isospin doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} c \\ s \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}. \tag{1.2}$$

The strong interaction, referred to as the color force, accounts for three different states of charge. Thus, each quark occurs in three colors. Contrary to leptons, quarks are observed only as bound states of three quarks (antiquarks), the baryons (antibaryons), having half-integer spin, or quark-antiquark pairs, the mesons, with an integer spin. They do not exist in free state because the color force does not drop off with distance. They are always confined to hadrons, which are colorless combinations. The top quark represents an exception, because the formation of bound states is prevented by its particularly short lifetime.
All these objects are considered as fundamental particles, but most of them are unstable. Only electrons, neutrinos, photons and, in the same way, quarks of the first generation constituting the ordinary matter are stable particles.

If a theory involves more than one state for particles, the associated fields need different states as well. According to the gauge theories (details given in Section 1.1.2), N possible states for a particle imply that there are N^2-1 corresponding gauge fields out of which N^2-N charged fields and (N-1) neutral fields, where neutral indicates that the multiplet represents the same physical state under a specific transformation. Electromagnetism has a single internal charge, the electric charge: the photon (γ) is the mediator of this interaction, which is neutral because of the conservation of electric charge. The strong interactions are characterized by three different states related to the color charge and are mediated by eight gluons (g). The mediators for weak forces are three vector bosons, because weak interactions are described by two different states related to the isospin charge: two charged vector bosons, $W^{\pm}$, and one neutral vector boson, Z^0, hereafter usually referred to as W and Z bosons. As anticipated before, the SM is not able to give a description of the gravitational force, but the effect of gravity is totally negligible at the scale of subatomic particles. The corresponding force-carrying particle of gravity is supposed to be the graviton (G). The photons and gluons are massless, while W and Z are massive particles, although they are supposed not to have a mass from the theory.

[1]Hereafter, units will be reported assuming speed of light c = 1.

Therefore, an external mechanism is needed to give mass to these mediators (see Section 1.1.4).

The theoretical solution was proposed in 1964 by P. Higgs, F. Englert, and R. Brout [6–9] and by G. Guralnik, C. R. Hagen, and T. W. B. Kibble [10, 11], predicting the existence of a new scalar boson with spin zero in the SM: the H boson.

1.1.2 Gauge theory behind the standard model

Our best understanding and description of the fundamental structure of matter is deeply related to symmetry principles. A symmetry is a mathematical or physical feature of a system, which alludes to an invariant system with respect to some transformations. The symmetries and corresponding transformations could be classified as continuous or discrete, geometrical or internal, and global or local. The Noether's theorem connects exact symmetries and conservation laws or selection rules: every continuous symmetry of the Lagrangian corresponds to a conserved quantity. Invariance principles play an essential role in particle physics, because they guide the construction of SM theories and are intimately connected to gauge field theories. The requirement of a local gauge invariance implies the existence of interactions and gives rise to interaction field theories.

> "*We wish to explore the possibility of requiring all interactions to be invariant under independent rotations of the isotopic spin at all space-time points.*"
> (C. N. Yang and R. Mills, 1954)

A generic gauge theory relies on a symmetry group of transformations in an abstract space, $\mathcal{G}$, where a gauge invariant action $S = \int \mathcal{L} d^4x$ is defined so that the Lagrangian itself is invariant and the transformations of the fields are linear under the action of the gauge symmetry:

$$\psi'^{i}(x) = U^{ij}(x)\psi^{j}(x). \tag{1.3}$$

Global gauge invariances are common in particle physics and do not depend on space-time. A global gauge invariance can be generalized to a local one, as suggested by C. N. Yang and R. Mills, by choosing the phase arbitrarily in the whole space-time, so that the transformation varies from place to place in space-time. Hence, a covariant derivative $\mathcal{D}_\mu$ has to be defined to replace the derivative δ_μ:

$$\mathcal{D}_\mu \psi^i(x) = \delta_\mu \psi^i(x) - igt_a^{ij} A_\mu^a(x)\psi^j(x), \tag{1.4}$$

where $A_\mu^a(x)$ indicate the gauge vector fields, t_a^{ij} are the generators of the symmetry group $\mathcal{G}$ in the representation of the fields ψ^i, and g refers to the charge. The charge is unique for a simple symmetry group. However, in case $\mathcal{G}$ consists of the product of different groups $\mathcal{G}_i$, each group has its own charge g_i.

The fundamental interactions in the SM are described by three dynamical theories: *Quantum Electrodynamics* (QED) for the electromagnetic force, *Quantum Cromodynamics* (QCD) for the strong interaction, and *Electroweak theory* (EWT) for the electroweak interaction. The mathematical description of the corresponding symmetries relies on the group theory. The gauge group of the SM, based on the gauge principle of local phase invariance, combines three different symmetry groups:

$$\mathrm{SU(3)_C} \otimes \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}.$$

The single elements of the product are:

* the one dimensional group $\mathrm{U(1)_Y}$ of electrodynamics;

- the group $SU(2)_L$ of isotopic spin conservation, representing a local symmetry which governs the weak interactions of quarks and leptons;
- the three dimensional group $SU(3)_C$, which governs the strong interactions of gluons and quarks, binding them into hadrons and providing stability to the nucleus overcoming electric repulsion.

On the one hand, the electrodynamics is described by an abelian group, since the phase factors commute with each other. On the other hand, weak and strong interactions involve different states, which generally do not commute with each other: the associated symmetry is called a nonabelian symmetry and the phase factors become matrices. In the following, the formalism of a local gauge symmetry is used to provide a description of the strong and electroweak interactions.

Strong interaction The $SU(3)_C$ nonabelian group accounts for eight Gell-Mann matrices λ_a as generators, corresponding to eight massless gluons which can have self-interactions, and three degrees of freedom of the associated quantum number, the color, which are blue, red, and green. Quarks are described through a $3 \times \bar{3}$ representation of the group composed of two terms: the gauge invariant mass term and the covariant vector coupled to gluons.

The Dirac equation for a free fermion can be given as

$$(i\gamma^\mu \partial_\mu - m)\psi(x) = 0, \tag{1.5}$$

driven from the Lagrangian density:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial\mu - m)\psi. \tag{1.6}$$

The local gauge invariance induces a linear transformation where the space-time derivative has to be redefined according to Eq. 1.4:

$$\begin{aligned} &\psi(x) \rightarrow e^{i\alpha_a(x)\frac{\lambda_a}{2}}\psi(x), \\ &\mathcal{D}_\mu = \delta_\mu + ig_s\frac{\lambda_a}{2}G^a_\mu. \end{aligned} \tag{1.7}$$

The gluons are represented by eight gauge vector fields G^a_μ which have to satisfy the gauge invariance, as well. Consequently they transform as

$$G^a_\mu \rightarrow G^a_\mu + \alpha^b(x)f^{abc}G^c_\mu + \frac{1}{g_s}\delta_\mu\alpha^a(x), \tag{1.8}$$

with f^{abc} denoting the structure constants of the group, derived from the commutation rules $\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2}\right] = if^{abc}\frac{\lambda_c}{2}$. Considering a summation of all quark fields in the expression, the final QCD Lagrangian invariant under gauge transformations looks like:

$$\mathcal{L}_{\mathrm{QCD}} = \bar{\psi}(i\gamma^\mu\delta_\mu - m)\psi - g_s\bar{\psi}\gamma^\mu\frac{\lambda_a}{2}\psi G^a_\mu - \frac{1}{4}G^{\mu\nu}_a G^a_{\mu\nu}, \tag{1.9}$$

with the first term corresponding to Eq. 1.6, the second and third term representing the interaction between quarks and gluons and the trilinear and quadrilinear self-interactions of the gluon fields, respectively.

Electroweak interaction A clear analogy between electromagnetic and weak interactions was explained in 1961 by S. Glashow, who proposed a structure based on a single group symmetry $SU(2) \otimes U(1)$. The electrical charge of the abelian U(1) group is associated

with another generator, the weak hypercharge (Y), through the Gell-Mann and Nishijima relation:

$$Q = I_3 + \frac{1}{2}\,Y. \tag{1.10}$$

The I_3 generator, representing the third component of the weak isospin and being $\pm\frac{1}{2}$ for doublets and zero for singlets, is a particularly convenient choice because it applies only to left-handed (L) fermions and the SM is a theory with a defined chirality[2]. The $\mathrm{U(1)_Y}$ local gauge invariance leads to a single gauge field, denoted as B_μ, which interacts with both left-handed and right-handed particles. The weak isospin is related to a triplet of gauge bosons ($\vec{\mathrm{W}}$ or W^l) as generators. The representation of the fermions are a doublet L with $I_3 = \frac{1}{2}$, which indicates a left-handed chirality, and two singlets with $I_3 = 0$, which corresponds to right-handed fermions that do not interact with the $\vec{\mathrm{W}}$ bosons, as presented below. The neutrino fields, the corresponding lepton fields, and the associated up and down quark fields are represented by ψ and ψ', respectively.

$$\begin{aligned} L &\equiv \frac{1-\gamma^5}{2}\begin{pmatrix}\psi\\ \psi'\end{pmatrix} = \begin{pmatrix}\psi_L\\ \psi'_L\end{pmatrix} \\ \psi_R &\equiv \frac{1+\gamma^5}{2}\psi \\ \psi'_R &\equiv \frac{1+\gamma^5}{2}\psi' \end{aligned} \tag{1.11}$$

A difference in the fermion chirality implies different fields and, as a consequence, different interactions. For this reason, the explicit presence of the fermion mass in the Lagrangian is forbidden. The Lagrangian for the electroweak theory can be written as $\mathcal{L}_{\mathrm{EW}} = \mathcal{L}_{\mathrm{gauge}} + \mathcal{L}_{\mathrm{leptons}} + \mathcal{L}_{\mathrm{quarks}}$. In the following, a focus on the leptonic terms of the Lagrangian may serve as an example. Starting from the free-field Dirac Lagrangian density in Eq. 1.6, the Lagrangian density of the electroweak interaction can be written as

$$\mathcal{L}_{\mathrm{EW}} = -\frac{1}{4}W^l_{\mu\nu}W^{\mu\nu l} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{L}i\gamma^\mu\delta_\mu L + \bar{\psi}'_R i\gamma^\mu\delta_\mu\psi'_R. \tag{1.12}$$

Requiring local gauge invariance under a $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ transformation, we need to introduce four gauge fields ($W^1_\mu, W^2_\mu, W^3_\mu, B_\mu$) and to replace the derivative ∂_μ with the covariant derivative:

$$D_\mu = \partial_\mu + \frac{ig'}{2}B_\mu Y + \frac{ig}{2}\vec{\tau}\,\vec{W}_\mu, \tag{1.13}$$

where g is the coupling constant of the weak-isospin group $\mathrm{SU(2)_L}$, $\frac{g'}{2}$ is the coupling constant for the weak-hypercharge group $\mathrm{U(1)_Y}$, and $\vec{\tau}$ are the generators of the group called Pauli matrices. The four gauge fields can be expressed as

$$\begin{aligned} W^l_{\mu\nu} &= \partial_\nu W^l_\mu - \partial_\mu W^l_\nu + g\varepsilon_{jkl}W^j_\mu W^k_\mu, \\ B_{\mu\nu} &= \partial_\nu B_\mu - \partial_\mu B_\nu. \end{aligned} \tag{1.14}$$

The kinetic term for the gauge fields and the matter term appear as:

$$\begin{aligned} \mathcal{L}_{\mathrm{gauge}} &= -\frac{1}{4}W^l_{\mu\nu}W^{\mu\nu l} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \\ \mathcal{L}_{\mathrm{leptons}} &= \bar{\psi'_R}\,i\,\gamma^\mu\,(\partial_\mu + \frac{ig'}{2}B_\mu Y)\psi'_R + \bar{L}\,i\,\gamma^\mu(\partial_\mu + \frac{ig'}{2}\,B_\mu\,Y + \frac{ig}{2}\,\vec{\tau}\,\vec{W}_\mu)\,L. \end{aligned} \tag{1.15}$$

[2]Note that both left-handed and lepton number are represented by L. In this theory of the SM the L stands for left-handed only.

The index l related to fields W^l and the extra term with the Levi-Civita tensor ϵ_{jkl} underline that generators do not commutate with each other because of the nonabelian nature of the symmetry group $SU(2)_L$. A complex structure of gauge bosons self-interactions is predicted by the theory, including both cubic (ZWW, γWW) and quartic (ZZWW, γZWW, WWWW, $\gamma\gamma$WW) vertices.
The interaction with matter particles described by $\mathcal{L}_{\text{leptons}}$ is mediated by the exchange of massless gauge bosons. Quadratic terms for gauge fields are not present in the expression. The inclusion "a posteriori" of a mass term in the description would destroy the Lagrangian invariance.

The gauge principles applied to Yang-Mills theories leads inevitably to a model of interactions mediated by massless bosons and it was supposed that any massless gauge bosons would surely have been detected. However, contrary to theoretical predictions, experimentalists observed the existence of massive vector bosons, carriers of the weak interaction. Consequently, from the beginning the main obstacle to the application of the Yang–Mills approach to the weak interaction theory was the problem of the mass. Moreover, gauge theories were compatible only with exact symmetries, but the physical world manifests a large fraction of symmetry principles which hold only approximately.
In the early 1960s, physicists thought about introducing a dynamical mechanism to give a nonnull mass to gauge fields, A turning point of the problem was the idea of introducing a spontaneously broken symmetry in the gauge theory. There might be symmetries of the Lagrangian that are not symmetries of the vacuum, known as the ground state. One of the main themes of the modern physics is the study of how symmetries of the Lagrangian can be broken.

"*While we would like to believe that the fundamental laws of Nature are symmetric, a completely symmetric world would be rather dull, and as a matter of fact, the real world is not perfectly symmetric.*"
(A. Zee, 2010)

1.1.3 Spontaneous symmetry breaking

Sometimes in nature it could happen that some physical laws are characterized by hidden symmetries because the ground state is not invariant under corresponding transformations. This translates to a fact that the Lagrangian density is invariant under a certain symmetry ($\partial\mathcal{L} = 0$). However, the physical vacuum is not unique and, consequently, not invariant under the symmetry transformations. This situation is usually named as spontaneous symmetry breaking (SSB).
In the following, this reasoning is applied to the QFT, starting from a minimalistic model which considers a real scalar field and extending the discussion to the case of interest related to the $SU(2)_L \otimes U(1)_Y$ gauge symmetry.

The Lagrangian of a generic scalar field and its potential can be expressed as:

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}(\partial_\mu\phi)^2 - U(\phi), \\ U(\phi) &= \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4, \end{aligned} \tag{1.16}$$

where the energy can be modified only by spatial variation in ϕ and λ is positive to ensure an absolute minimum in the Lagrangian. The value assumed by ϕ in the vacuum is known as the vacuum expectation value (VEV) of ϕ and is denoted as $<\phi>$. We look for minima of the potential energy $U(\phi)$.
In a one dimensional space, it is possible to see clearly that two cases could be distinguished depending on the sign of μ^2:

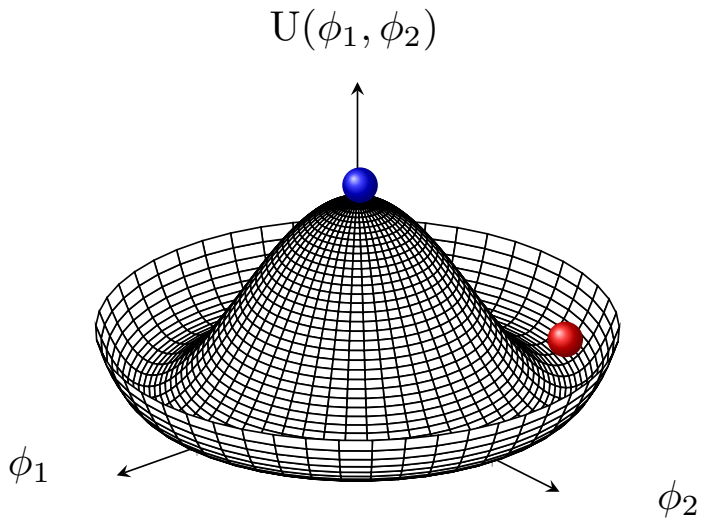


FIGURE 1.2: A mexican hat potential typical of a two scalar fields theory, with minima corresponding to the $\mu^2 > 0$ and $\mu^2 < 0$ scenarios indicated in blue and red, respectively.

- if the parameter $\mu^2 > 0$, the potential has an unique minimum at $\phi = 0$, coincident with the vacuum state, and the symmetry is manifest;
- else if the parameter $\mu^2 < 0$, the potential minima are located in two degenerate lowest energy states, $\phi_0 = \pm v = \pm\sqrt{-\frac{\mu^2}{\lambda}}$, either of which may be chosen to be the vacuum. This is a spontaneously broken symmetry, representing our case of interest. The choice of the positive value leads to the relation $\mu^2 = -\lambda v^2$.

In a two dimensional space, corresponding to a continuous symmetry, the shape of the potential is similar to a mexican-hat, as shown in Fig. 1.2. If $\mu^2 < 0$ and the positive v is chosen, its minimum represents an infinite number of vacua satisfying:

$$\sqrt{\phi_1^2 + \phi_2^2} = \sqrt{-\frac{\mu^2}{\lambda}} = v. \tag{1.17}$$

The dynamics determined by $\mathcal{L}$ implies a degenerate set of vacuum states that are noninvariant under the symmetry. The continuum of infinite vacuum states may be distinguished only by the direction of $\vec{\phi}$ in a specific vacuum. We choose a particular direction of $\vec{\phi}$ to have

$$\phi_1 = v = \sqrt{-\frac{\mu^2}{\lambda}} \quad \text{and} \quad \phi_2 = 0. \tag{1.18}$$

Consequently, considering fluctuations around the chosen field configuration, the vacuum can be described as:

$$\begin{aligned} \phi_1 &= v + \phi_1', \\ \phi_2 &= \phi_2', \\ \phi_0 &= \frac{1}{\sqrt{2}}(\phi_1' + v + i\phi_2'). \end{aligned} \tag{1.19}$$

Expanding $\mathcal{L}$ with $\mu^2 = -\lambda v^2$, we obtain:

$$\mathcal{L} = \frac{1}{2}[(\partial_\mu \phi_1')^2 + (\partial_\mu \phi_2')^2] + \frac{1}{2}\lambda v^2[(\phi_1' + v)^2 + \phi_2'^2] - \frac{\lambda}{4}[(\phi_1' + v)^2 + \phi_2'^2]. \tag{1.20}$$

Without considering the constant and higher order terms, the Lagrangian appears in the form

$$\mathcal{L}(\phi_1', \phi_2') = \frac{1}{2}(\partial_\mu \phi_1')^2 - (\lambda v^2)\phi_1'^2 + (\partial_\mu \phi_2')^2 + 0 \cdot \phi_2'^2 + ..., \tag{1.21}$$

where the quadratic term $\phi_1'^2$ reveals that the field ϕ_1' has assumed a mass equal to $\sqrt{-2\mu^2}$, while the term $\phi_2'^2$ is absent. As a consequence of the SSB applied to a continuous symmetry, a massless scalar particle has emerged. This phenomenon is explained by the Goldstone theorem [12], stating that for every spontaneously broken symmetry there must be a massless and spinless particle referred to as Nambu-Goldstone (NG) boson.

According to this theorem, if there exists a conserved operator Q_i which causes a transformation on the component A_j of a field,

$$[Q_i, A_j(x)] = \sum_k t_{ijk}\, A_k(x), \tag{1.22}$$

where t_{ijk} is the representative matrix of the transformation, and if it is possible to break the symmetry so that $\sum_k t_{ijk} \times < 0|A_k|0 > \neq 0$, therefore $A_j(x)$ has a massless particle in its spectrum.

The NG bosons result from the breakdown of a global symmetry, which is generalized to a local one in the Yang-Mills theory. Because of the gauge invariance the problem seems to be more difficult. Nevertheless, a special interplay between the gauge particles of the Yang-Mills theory and NG bosons was supposed to be possible, endowing the former with mass and removing the latter from the spectrum.
The Brout-Englert-Higgs (BEH) mechanism provides an elegant solution which is able to give rise to this mutual coupling, becoming a central ingredient in the current understanding of the electroweak interaction.

1.1.4 The BEH mechanism and the Higgs boson

The theories presented by P. Higgs, F. Englert, and R. Brout [6–9] and by G. Guralnik, C. R. Hagen, and T. W. B. Kibble [10, 11] in 1964 demonstrated that massive mediators of the electroweak force could acquire mass thanks to the SSB mechanism and the introduction of a scalar field responsible for that, the Higgs field. If the vacuum is degenerate, a mass term emerges in the Lagrangian without violating the invariance as a consequence of the broken symmetry and of the interaction with the H boson. Moreover, the Yukawa interactions between the H scalar field and fermions determine the masses of the quarks and leptons. Instead, the mass of the H is a free parameter of the model which can be determined experimentally.

The symmetry breaking within the $SU(2)_L \otimes U(1)_Y$ group is realized by combining two real scalar fields into a complex scalar doublet and adding the scalar field in the Lagrangian. Firstly, the effect of the BEH mechanism is shown for the simple case of a U(1) gauge boson A_μ and a charged scalar particle represented by the complex field ϕ. Then, a generalization to the case of the electroweak Lagrangian is discussed.
Starting from a Lagrangian density $\mathcal{L}(\phi, \partial_\mu \phi)$ similar to the example of scalar fields reported in Eq. 1.16 and substituting ∂_μ with the covariant derivative $D_\mu \phi = (\partial_\mu - ieA_\mu)\phi$, a complex field is included as the combination of two fields ϕ_1 and ϕ_2. The Lagrangian can

be expressed as:

$$\begin{aligned}
&\phi = \phi_1 + i\phi_2, \\
&\phi^*\phi = \phi_1^2 + \phi_2^2, \\
&\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + (D^\mu\phi)^\dagger(D_\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2,
\end{aligned} \tag{1.23}$$

where the first term represents the kinetic term for the gauge field A_μ associated to the photon. Local U(1) gauge invariance implies that the Lagrangian is invariant under $\phi' \to e^{i\alpha(x)}\phi$ with A_μ transforming like $A'_\mu = A_\mu + 1/e\partial_\mu\alpha$. By comparing the case of a two dimensional scalar field presented in the previous section, an additional term in the kinetic part appears because of the local gauge invariance. Considering a specific configuration of the vacuum and two shifted fields now denoted as η and ξ, we look at perturbations around this minimum:

$$\begin{aligned}
\phi_1 &= v + \eta, \\
\phi_2 &= \xi, \\
\phi_0 &= \frac{1}{\sqrt{2}}(\eta + v + i\xi).
\end{aligned} \tag{1.24}$$

Starting from a kinetic term like:

$$\mathcal{L} = (D^\mu\phi)^\dagger(D_\mu\phi) = (\partial^\mu + ieA^\mu)\phi^*(\partial_\mu + ieA_\mu)\phi, \tag{1.25}$$

the full Lagrangian (excluding higher order terms) appears as:

$$\mathcal{L}(\eta,\xi) = \frac{1}{2}(\partial_\mu\eta)^2 - (\lambda v^2)\eta^2 + \frac{1}{2}(\partial_\mu\xi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}e^2v^2A_\mu^2 - evA_\mu(\partial^\mu\xi) + \dots \tag{1.26}$$

The effect of the SSB in the Lagrangian is similar to the result shown in the previous section. Following the order of the expression, there is a massive field η and a massless field ξ. Moreover, the vector field A_μ has acquired mass. Given that A_μ and ϕ vary simultaneously, part of the expression can be rewritten exploiting the gauge invariance under a specific choice called the *unitary gauge*:

$$\frac{1}{2}(\partial_\mu\xi)^2 + \frac{1}{2}e^2v^2A_\mu^2 - evA_\mu(\partial^\mu\xi) = \frac{1}{2}e^2v^2\Big[A_\mu - \frac{1}{ev}(\partial_\mu\xi)\Big]^2 = +\frac{1}{2}e^2v^2A_\mu'^2. \tag{1.27}$$

Setting $\alpha = \frac{-\xi}{v}$, the field ϕ transforms accordingly:

$$\phi' \to e^{-i-\xi/v}\phi = e^{-i-\xi/v}\frac{1}{\sqrt{2}}(\eta+v+i\xi) = e^{-i-\xi/v}\frac{1}{\sqrt{2}}(\eta+v)e^{+i-\xi/v} = \frac{1}{\sqrt{2}}(v+h). \tag{1.28}$$

The field ξ, which would have originated a NG boson, disappears. It is incorporated by the gauge field A_μ providing it with a mass $m = ev$. It is known that a massive vector boson has got three degrees of freedom corresponding to three states of spin, while a massless vector boson has got only two degrees of freedom related to the two states of the photon's helicity. Because of the BEH mechanism, two degrees of freedom of the massless gauge field and a single degree of freedom of the NG boson combine themselves producing a longitudinal degree of freedom of the gauge field A_μ, which becomes massive. Expanding $(v+h)^2$, the full Lagrangian in the unitary gauge accounts for a massive scalar particle h (the H boson), a massive gauge field A_μ corresponding to the photon, H interactions with the gauge fields, and H self-interactions with both cubic and quartic vertices:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 - \lambda v^2h^2 + \frac{1}{2}e^2v^2A_\mu^2 + e^2vA_\mu^2h + +\frac{1}{2}e^2A_\mu^2h^2 - \lambda vh^3 - \frac{1}{4}\lambda h^4. \tag{1.29}$$

The same approach can be now extended to the $SU(2)_L \otimes U(1)_Y$ symmetry. The electroweak unification theory was presented in 1967-1968 by S. Weinberg and A. Salam, who included both SSB and BEH mechanisms in a nonabelian gauge theory. The way to spontaneously break gauge symmetry represented a challenging question among the particle physicists. The answer suggested by the SM is the existence of a scalar field whose interactions select a vacuum state in which electroweak symmetry is hidden.
Starting from the description of the electroweak interaction given in Eq. 1.15, a proper scalar field ϕ is added to the Lagrangian:

$$\begin{aligned} &\mathcal{L}_{\text{scalar}} = (D_\mu \phi)^\dagger (D_\mu \phi) - U(\phi), \\ &U(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \end{aligned} \tag{1.30}$$

where μ^2 is negative and D_μ represents the covariant derivative reported in Eq. 1.13 for the isospin doublet ϕ. It is a complex doublet of two scalar fields, ϕ^+ and ϕ^0, which denote charged and neutral fields, respectively:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad \text{where} \quad \phi^+ = \frac{\phi_1 + i\phi_2}{\sqrt{2}} \quad \text{and} \quad \phi^0 = \frac{\phi_3 + i\phi_4}{\sqrt{2}}. \tag{1.31}$$

Focusing only on this scalar part of the Lagrangian, it is possible to show how the gauge bosons acquire their mass as a consequence of the BEH mechanism. The physical vacuum state is selected as the configuration

$$\phi \to \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.32}$$

Consequently, all four electroweak generators (the Pauli matrices $\vec{\tau}$ and the hypercharge Y) are broken. However, the linear combination corresponding to the electric charge Q is unchanged. This fact implies that the vacuum state is not invariant for symmetry groups $SU(2)_L$ and $U(1)_Y$, but the invariance is conserved in the case of $U(1)_{EM}$. Therefore, three mediators of weak interaction acquire mass but the photon remains massless.

At this point, since ϕ is a doublet of $SU(2)_L$, the expansion of the field around the potential minimum is given by:

$$\phi(x) = e^{i\vec{\tau}\cdot\vec{\theta}(x)/v} \begin{pmatrix} 0 \\ \frac{v+h(x)}{\sqrt{2}} \end{pmatrix}, \tag{1.33}$$

where $\vec{\tau}$ are SU(2) generators, $\vec{\theta}(x)$ represents an arbitrary phase, and $h(x)$ a scalar field. As a consequence of the field expansion around the minimum, a scalar massive field $h(x)$ and three massless fields $\vec{\theta}(x)$ are present, as expected from the Goldostone theorem.
Following the same procedure presented above in the case of the local U(1) gauge invariance, the choice of the unitary gauge within a $SU(2)_L$ transformation allows to remove these massless bosons. The NG bosons are incorporated as additional degrees of freedom of the three weak gauge bosons W_μ^l in Eq. 1.14, corresponding to their longitudinal polarization and providing them with a mass.

The mediators of the weak interactions, the charge bosons $W^\pm$ and the neutral boson Z^0, and the photon field A_μ emerge from a linear combination of W_μ^l and B_μ. Redefining the quantities W_μ^1 and W_μ^2 as

$$W^\pm = \frac{W_\mu^1 \mp i W_\mu^2}{\sqrt{2}}, \tag{1.34}$$

in the Lagrangian an additional term appears in the following form:

$$\frac{g^2v^2}{4}W_\mu^+W^{-\,\mu}+\frac{v^2}{8}(gW_\mu^3-g'B_\mu)^2. \tag{1.35}$$

Then, the linear combination $(gW_\mu^3-g'B_\mu)$ becomes massive, while the orthogonal one remains massless and represents the photon. Hence,

$$Z_\mu=\frac{gW_\mu^3-g'B_\mu}{\sqrt{g^2+g'^2}}=W_\mu^3\,\cos\theta_W-B_\mu\,\sin\theta_W, \tag{1.36}$$

$$A_\mu=\frac{gW_\mu^3+g'B_\mu}{\sqrt{g^2+g'^2}}=W_\mu^3\,\cos\theta_W+B_\mu\,\sin\theta_W, \tag{1.37}$$

where θ_W is the Weinberg angle, which represents the weak mixing angle of the two coupling constants g and g' in the Weinberg-Salam theory of the electroweak interaction:

$$g'=g\tan\theta_W. \tag{1.38}$$

The mass of the weak gauge bosons can be expressed as:

$$\begin{aligned} m_{\mathrm{Z}}^2&=\frac{v^2\,(g^2+g'^2)}{4},\\ m_{\mathrm{W}}^2&=\frac{g^2v^2}{4}=m_{\mathrm{Z}}^2\cos^2\theta_{\mathrm{W}}. \end{aligned} \tag{1.39}$$

In conclusion, the BEH mechanism is able to give a mass to the electroweak vector bosons and the full Higgs Lagrangian is

$$\begin{aligned} \mathcal{L}_{\mathrm{Higgs}}=&\frac{1}{2}\left[\partial_\mu h\partial^\mu h-(2\lambda v^2)h^2\right]+\\ &\frac{\mu^2}{v}h^3+\frac{\mu^2}{8v^2}h^4+\\ &\frac{1}{2}\left[m_{\mathrm{W}}^2(W_\mu^+)^\dagger W_+^\mu+m_{\mathrm{W}}^2(W_\mu^-)^\dagger W_-^\mu+m_{\mathrm{Z}}^2Z_\mu Z^\mu\right]\left(1+\frac{h}{v}\right)^2. \end{aligned} \tag{1.40}$$

Looking at the mass term μ associated to h, the Higgs boson mass is given by $m_{\mathrm{H}}=\sqrt{2\lambda v^2}$. On the one hand, the VEV v, representing the energy scale of the electroweak symmetry breaking, is known because it can be obtained from the Fermi constant G_F:

$$v=\sqrt{\frac{1}{\sqrt{2}G_F}}\approx 246\,\mathrm{GeV}. \tag{1.41}$$

On the other hand, λ is a free parameter. Therefore, the Higgs mass is not predicted in the SM. The cubic and quartic self-interactions of the H boson as predicted by the SM are visible in the second line of Eq. 1.40. The H boson potential can be rewritten as

$$V(\mathrm{H})=\frac{1}{2}m_{\mathrm{H}}^2H^2+\lambda_{\mathrm{HHH}}v\mathrm{H}^3+\frac{1}{4}\,\lambda_{\mathrm{HHHH}}\mathrm{H}^4-\frac{\lambda}{4}v^4. \tag{1.42}$$

where, according to the SM, the mass and the self-couplings of the H boson depend only on λ and v:

$$\lambda_{\mathrm{HHH}}=\lambda_{\mathrm{HHHH}}=\lambda=\frac{m_{\mathrm{H}}^2}{2v^2}. \tag{1.43}$$

Moreover, the H interactions with the weak bosons can be found in the third line: HZZ and HWW interactions originate from the $2\,h/v$ term and HZZZ and HWWW interactions come from the h^2/v^2 term.

The same mechanism is responsible for generating the fermion masses leading to an extension of the SM Lagrangian to include also gauge-invariant Yukawa terms:

$$\mathcal{L}_{\text{Yukawa}} = \sum_{\iota=1}^{3} -m_f \bar{\psi}\psi\Big(1+\frac{h}{v}\Big) + \sum_{\iota=1}^{3} -m_{f'} \bar{\psi}'\psi'\Big(1+\frac{h}{v}\Big), \tag{1.44}$$

where the two parts indicate up and down components of the fermion doublets, respectively. The intensity of the interaction is related to the fermion masses.

Finally, in 1971 M. Veltman and G. 't Hooft demonstrated that the theory was renormalizable. This definitive success made the electroweak theory the most impressive application of the SSB and the BEH mechanism, providing a consistent and predictive model later confirmed by experimental results, that will be described in the next section.

> "*An optimist might say that we are on the road to the first truly unified theory of the fundamental interactions. All of these marvellous developments are based upon the ideas of spontaneous symmetry breakdown and gauge fields.*"
> (S. Coleman, 1973)

1.2 Phenomenology of the Higgs boson at the LHC

The SM theory provides a complete and coherent description of phenomena which can be proved by the experimental observations at collider experiments down to scales three order of magnitude smaller than the atomic nucleus. The SM has been further confirmed by the discovery of the H boson by the ATLAS and CMS Collaborations in July 2012 [13–15]. As an introduction to the observation of this new scalar boson with a mass around 125 GeV and the summary of the experimental knowledge currently available about the H boson, a description of the main features of the H boson production and decay channels in proton-proton (p-p) collisions is provided, followed by a discussion of the framework used to interpret the H boson measurements at the LHC. The concepts and notations presented here will constitute an essential basis to understand the details of the $\mathrm{H} \to 4\ell$ analysis that will be treated in the central part of the thesis.

1.2.1 Production and decay of the Higgs boson

According to the SM, there are four main mechanisms contributing to the H boson production at hadron colliders: gluon fusion (ggF), vector boson fusion (VBF), associated production with a vector boson (VH, where V = Z, W), and associated production with a top quark pair ($\mathrm{t\bar{t}H}$). The respective Feynman diagrams are reported in Fig. 1.3. The corresponding cross sections as a function of the center-of-mass energy ($\sqrt{s}$) are shown in Fig. 1.4 considering a SM H boson with a mass approximately of 125 GeV.

Gluon fusion The ggF production mode has the highest cross section due to the high density of gluons in p-p collisions at the LHC center-of-mass energies. Subsequently, it is the dominant one contributing to the H boson production in the whole mass range. It happens through an intermediate heavy quark loop, with the largest contribution of about 90% arising from the top quark, which is characterized by a mass scale comparable to the H boson mass. Then, a sizeable contribution comes from the bottom quark, which is around 5-10% of the total cross section. The Leading-Order (LO) amplitude of the process is zero

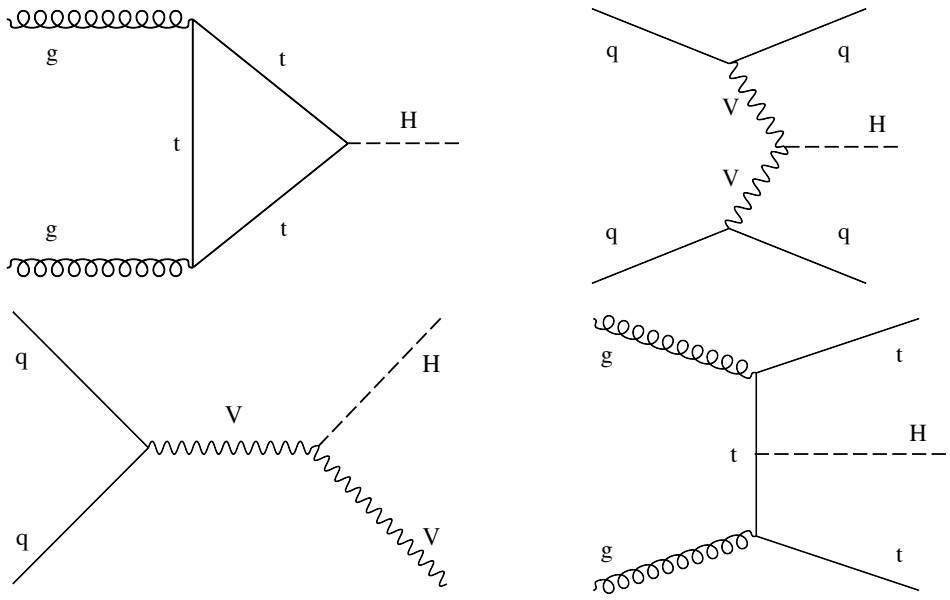


FIGURE 1.3: Higgs boson production through gluon fusion (top left) and vector boson fusion (top right); associated production with a vector boson (bottom left) and a top quark pair (bottom right).

for massless quarks and saturates for increasing values of the ratio m_q/m_{H}, leading to a proportionality between the cross section and the square of the number of heavy quarks. Considering higher orders, QCD corrections have a significant impact: they are positive and large, leading to an increase of the cross section by a factor around 50% and 30-40% accounting for Next-to-Leading-Order (NLO) and Next-to-Next-to-Leading-Order (NNLO) corrections, respectively.

Vector boson fusion The VBF process is about ten times rarer than ggF, representing the second dominating contribution to the H boson production at the LHC. Two fermions take part in a weak process which leads them to exchange a vector boson, W or Z. The fusion of these two vector bosons produces the H boson. The main feature of the process is a clear experimental signature of the signal, which is associated to a pair of hadronic jets (defined in Section 2.2.2) in the very forward regions at small angles with respect to the beam axis. Subsequently, a good separation from background events can be obtained. QCD radiation is significantly suppressed in the region between the pair of jets because of the absence of any quark or gluon line connecting the two outgoing quarks, so that higher order QCD corrections are found to be small. Moreover, uncertainties associated to both LO and NLO cross sections are found to be small as well [16].

Associated production with a vector boson The VH production mode, referred to as Higgs-strahlung, is the third dominating mode at the LHC and it is two times less frequent than VBF. The H boson is irradiated from a vector boson V leading to a final state with the H boson associated to a W or Z boson. A virtual boson is produced during the interaction and is represented by the internal line in the Feynman diagram. It is an important production mechanism because it provides unique information on the VH coupling. For it to be experimentally viable, the best scenario is with the H and V bosons decaying into a bottom quark pair and a lepton pair, respectively. In fact, the $\mathrm{H} \to \mathrm{b\bar{b}}$ channel has an advantageous branching ratio ($\mathcal{B}$) and leptons and neutrinos from the vector boson help in selecting events with respect to the fully hadronic final state and dominant QCD background.

Associated production with a top quark pair The $\mathrm{t\bar{t}H}$ production mechanism is characterized by a final state including the H and two top quarks, where a top quark pair is originated by two gluons and then the H boson is produced by the fusion of a top and an antitop. Although it represents the smallest contribution to the H boson production, it is an important process because it gives access to a direct measurement of the Yukawa coupling between the H boson and the top quark (y_t), complementary to the indirect determination from ggF. However, a proper reconstruction is challenging because of the small cross section and the large number of hadronic jets in the final state.

Rare production modes In the context of the $\mathrm{H} \to 4\ell$ analysis, two additional production modes are considered, where the H boson is produced in association with a pair of bottom quarks ($\mathrm{b\bar{b}H}$) or a single top quark (tHq and tHW, indicated as tH). The cross section of the $\mathrm{b\bar{b}H}$ process is comparable to the $\mathrm{t\bar{t}H}$ process, while the one of the tH production mode is one order of magnitude smaller. However, the latter is particularly interesting because it permits the determination of the sign of y_t.

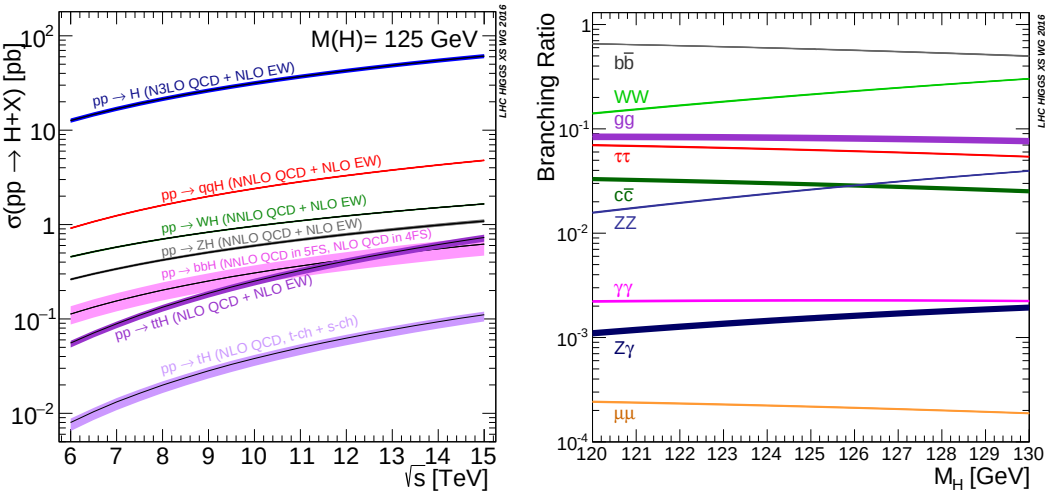


FIGURE 1.4: Production cross sections of the SM H boson as a function of $\sqrt{s}$ (left) and $\mathcal{B}$ of its most relevant decay channels in the mass region near 125 GeV as a function of mass (right) [17].

The unique nature of the H boson, which can decay in a variety of channels coupling to all SM massive particles and to massless particles via intermediate loops, provides an extraordinary occasion to investigate its couplings to many SM particles. In the SM, the H boson can decay into pairs of heavy fermions through the Yukawa coupling and pairs of gauge bosons. The interaction vertices are shown in Fig. 1.5. As mentioned in Section 1.1, the H boson coupling to fermions is proportional to the fermion masses:

$$g_{f\bar{f}H} = \sqrt{\sqrt{2}\,G_F}\; m_f, \tag{1.45}$$

while vector boson couplings, characterized by cubic or quartic vertices, are linearly dependent on the square of boson masses:

$$g_{VVH} = 2\sqrt{\sqrt{2}\,G_F}\; m_V^2. \tag{1.46}$$

Loop-induced decays into pairs of gluons or photons can also occur in the low mass region. The decay into a pair of gluons happens through a quark loop in the same way presented above for the gluon fusion production, while the $\mathrm{H} \to \gamma\gamma$ decay occurs through a fermion

loop, which is dominated by the contribution of the top quark, or a W boson loop with two WWγ vertices. Similar loops can also lead to Zγ final states.

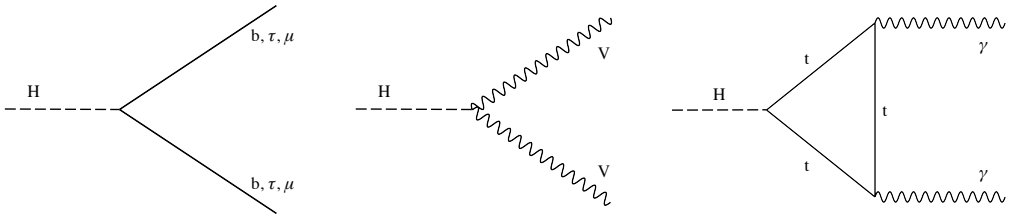


FIGURE 1.5: Interaction vertex between the H boson and a pair of fermions (left), a pair of vector bosons (middle), and a pair of photons through a top loop (right).

The total decay width (Γ_{H}) and mean lifetime of the H, as well as the relative $\mathcal{B}$ of its possible decay modes, are fully determined by the value of its mass (m_{H}). The $\mathcal{B}$ is determined by the partial width of the decay:

$$\mathcal{B}(\mathrm{H} \to \mathrm{X}) = \frac{\Gamma(\mathrm{H} \to \mathrm{X})}{\sum_{\mathrm{Y}} \Gamma(\mathrm{H} \to \mathrm{Y})} \,. \tag{1.47}$$

Table 1.1 lists some of the main decay channels of the H boson, also shown in Fig. 1.4, reporting $\mathcal{B}$ and associated uncertainties for a SM H boson with m_{H} approximately of 125 GeV. These uncertainties are due to the missing higher order corrections in the theoretical calculations on the one hand and the errors on the SM input parameters on the other hand, in particular the fermion masses and the QCD gauge coupling α_S.

Decay channel	$\mathcal{B}$	THU [%]	PU (m_q) [%]	PU (α_s) [%]
$\mathrm{H} \to \mathrm{b\bar{b}}$	5.81×10^{-1}	$+0.65$ -0.65	$+0.72$ -0.74	$+0.77$ -0.79
$\mathrm{H} \to \mathrm{W^+W^-}$	2.15×10^{-1}	$+0.99$ -0.99	$+0.98$ -0.98	$+0.64$ -0.62
$\mathrm{H} \to \mathrm{gg}$	8.18×10^{-2}	$+3.40$ -3.41	$+1.12$ -1.14	$+3.70$ -3.59
$\mathrm{H} \to \tau^+\tau^-$	6.26×10^{-2}	$+1.17$ -1.16	$+0.98$ -0.98	$+0.62$ -0.60
$\mathrm{H} \to \mathrm{c\bar{c}}$	2.88×10^{-2}	$+1.20$ -1.20	$+5.27$ -0.94	$+1.26$ -1.25
$\mathrm{H} \to \mathrm{ZZ}$	2.64×10^{-2}	$+0.99$ -0.99	$+0.98$ -0.98	$+0.64$ -0.62
$\mathrm{H} \to \gamma\gamma$	2.27×10^{-3}	$+1.73$ -1.72	$+0.97$ -0.94	$+0.66$ -0.61
$\mathrm{H} \to \mathrm{Z}\gamma$	1.54×10^{-3}	$+5.71$ -5.71	$+0.91$ -1.00	$+0.58$ -0.64
$\mathrm{H} \to \mu^+\mu^-$	2.17×10^{-4}	$+1.23$ -1.23	$+0.97$ -0.99	$+0.60$ -0.64

TABLE 1.1: Branching fraction for the main decays of the SM H boson assuming a mass around the peak, m_{H} = 125.09 GeV. The associated theoretical uncertainties (THU) and parametric uncertainties from the quark masses and the QCD gauge coupling, indicated as PU (m_q) and PU (α_s), are expressed in percentage [16].

The role of each decay mode for the discovery and measurement of the H boson properties is related not only to the $\mathcal{B}$ of the process, but also on the feasibility of an efficient extraction of signal events while rejecting associated backgrounds. The $H \to ZZ$ (2.6%) and $H \to \gamma\gamma$ (0.2%) decay modes are among the rarest decay channels but they played a leading role in the discovery of the H boson thanks to the high signal to background ratio and the excellent resolution on the invariant mass. The dominant decay mode is the $H \to b\bar{b}$ process, with a $\mathcal{B}$ of 58%, but its search is very challenging because of the significant QCD background. It is followed by the decay into a pair of W bosons (21%), which is studied selecting a pair of quarks or a charged lepton with the associated neutrino in the final state. The latter allows a larger sensitivity in the H boson searches thanks to accurate lepton identification and missing transverse energy measurement, in contrast with the high QCD background of the former. Considering H boson decays to fermions, the $H \to \tau^+\tau^-$ channel is the second most probable one (6%). It is probed by exploiting various decay products of the τ leptons. To probe the couplings to the second fermion generation, $H \to \mu^+\mu^-$ and $H \to c\bar{c}$, is particularly hard at the LHC, because of their small production rate and the challenging experimental signatures. Finally, two gluons or $Z\gamma$ decay channels, like the diphoton final state, are induced by a vector boson or quark loop, so that they offer a useful mean to investigate the effective coupling of the H boson to these SM particles.

1.2.2 Higgs couplings within the κ-framework

Up to now, the LHC measurements of H boson couplings via exclusive production and decay mechanisms, as presented in the previous section, have been mostly interpreted within the κ-framework [18, 19]. It represents a convenient parametrization of effects due to possible deviations from the SM induced by new physics on single H boson production and decay. Based on a set of on-shell H boson observables and assuming that the H boson couplings to the SM particles have the same helicity structure as in the SM, it is able to capture the dominant effects of various new physics scenarios which are currently well established. However, it is affected by some limitations and biases, so that it is not able to provide a systematic description of new physics. For example, it does not allow to correlate or combine measurements of different processes at different center-of-mass energies with a description of the energy dependence or to compare processes with different particle multiplicities. For this reason, an alternative approach based on the Effective Field Theory (EFT) is often adopted in recent searches and specifically in view of studies at future particle colliders (see Section 2.3.1). This approach aims to include new H boson couplings with different helicity structures, energy dependence or number of particles in the model, potentially originated by new heavy degrees of freedom. The EFT approach represents a natural extension of the κ-framework. Predictions can be systematically improved including both higher loop corrections in the SM couplings and new physics corrections encoded in operators of even higher dimensions. Moreover, polarization- and angular-dependent observables can be exploited. Finally, kinematical regions relevant for colliders operating far above the weak scale can be fully explored within an EFT analysis. For a detailed discussion of the EFT interpretation of H boson measurements see Ref. [20, 21]. In the following, an overview of the κ-framework is provided because it represents the model behind the $H \to ZZ^* \to 4\ell$ measurements presented in Chapter 6 of this thesis.

In the context of the SM, H boson couplings are uniquely fixed in terms of the Fermi constant and the masses of the particles. The parametrization provided by the κ-framework compares the experimental measurement of the H boson production and decay modes to the SM expectation, without requiring any BSM computation. Under the assumption of no BSM contributions to the total H boson width, the expression of the cross section times

branching ratio ($\sigma\mathcal{B}$) for a production mode i and decay f is:

$$(\sigma\mathcal{B})(i \to \mathrm{H} \to f) = \frac{\sigma_i(\vec{\kappa}) \cdot \Gamma^f(\vec{\kappa})}{\Gamma_\mathrm{H}(\vec{k})}, \tag{1.48}$$

where Γ^f indicates the partial decay width into the final state f and $\vec{\kappa}$ refers to a series of H boson coupling strength modifier parameters. Each κ is introduced to express each of the components of Eq. 1.48 as the SM expectation multiplied by the square of a coupling strength modifier for the corresponding process at LO:

$$\kappa_i^2 = \frac{\sigma_i}{\sigma_i^\mathrm{SM}} \text{ or } \kappa_i^2 = \frac{\Gamma_i}{\Gamma_i^\mathrm{SM}}. \tag{1.49}$$

Consequently, Eq. 1.48 can be expressed as:

$$(\sigma\mathcal{B})(i \to \mathrm{H} \to f) = \frac{\sigma_i^\mathrm{SM}\kappa_i^2 \cdot \Gamma_f^\mathrm{SM}\kappa_f^2}{\Gamma_\mathrm{H}^\mathrm{SM}\kappa_\mathrm{H}^2}, \tag{1.50}$$

resulting in another quantity largely used in the H boson analysis, the signal strength modifier (μ), which is the rate relative to the SM expectation for a given process:

$$\mu_i^f \equiv \frac{\sigma\mathcal{B}}{\sigma_\mathrm{SM}\mathcal{B}_\mathrm{SM}} = \frac{\kappa_i^2 \cdot \kappa_f^2}{\kappa_\mathrm{H}^2}. \tag{1.51}$$

When all κ_i are set to 1, the SM is reproduced. The SM H boson width can also be expressed in a way that accounts for modifications κ_i of the SM H boson coupling strengths:

$$\kappa_\mathrm{H}^2 \equiv \sum_j \frac{\kappa_j^2\Gamma_j^\mathrm{SM}}{\Gamma_\mathrm{H}^\mathrm{SM}}. \tag{1.52}$$

Finally, an extension of the κ-framework allows to also consider the existence of H boson decays to BSM particles with a branching fraction $\mathcal{B}_\mathrm{BSM}$, which implies the increase of Γ_H by a factor 1/(1 - $\mathcal{B}_\mathrm{BSM}$). Two classes of H boson decays to BSM particles can be identified: decays into invisible particles with $\mathcal{B}_\mathrm{inv}$ and decays into all other untagged particles with $\mathcal{B}_\mathrm{unt}$. Therefore, the total H boson width can be expressed as:

$$\Gamma_\mathrm{H} = \frac{\Gamma_\mathrm{H}^\mathrm{SM} \cdot \kappa_\mathrm{H}^2}{1 - (\mathcal{B}_\mathrm{inv} + \mathcal{B}_\mathrm{unt})}. \tag{1.53}$$

In case of colliders able to directly measure the H boson width, $\mathcal{B}_\mathrm{unt}$ can be constrained together with κ_i and $\mathcal{B}_\mathrm{inv}$ from a joint fit to the data.

1.2.3 Experimental picture of the Higgs boson

Given that the SM theory was not able to predict the mass of the H boson (see Section 1.1), the particle had to be searched for in the widest possible mass range. During the initial searches at particle colliders, the H boson eluded detection and lower limits on its mass were set. The direct searches at the Large Electron-Positron collider (LEP) [22, 23] at CERN allowed to determine a lower bound on m_H by the final year of operation (2000). Data was collected at energies between $\sqrt{s}$ = [189, 209] GeV corresponding to an integrated luminosity of 2461 pb^{-1}, of which 536 pb^{-1} were accumulated at $\sqrt{s}$ > 206 GeV. A signal with $m_\mathrm{H} \leqslant 114.4$ GeV was excluded at a confidence level (CL) of 95% [24–26]. Then, the searches performed at Tevatron, the $p\bar{p}$ collider at Fermilab in the United States, excluded

the mass window m_{H} = [158, 173] GeV at 95% CL using the combination of the data analyzed at CDF and D0, corresponding to an integrated luminosity of 8.2 fb^{-1} [27]. Its operation ended in September 2011, when the Large Hadron Collider (LHC) at the CERN laboratories got started with the specific purpose of observing the Higgs boson. Thanks to its high energy, the LHC offered a greater capability to explore the remaining mass ranges.

The discovery The search for the H boson at the LHC was leaded by the ATLAS and CMS Collaborations and covered a large mass spectrum between 100 GeV and 1 TeV and different decay modes with the H boson decaying into a pair of vector bosons, a pair of photons, a pair of bottom quarks, or a pair of τ leptons. The observation of a new particle in the mass region around 125 GeV, compatible with the SM predictions, was announced on the 4^{th} July 2012 [13–15]. Both experiments observed a new scalar boson with $m_{\mathrm{H}} \sim 125$ GeV, combining data collected during the years 2011 and 2012 at energies $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, with a significance exceeding 5 standard deviations (σ) [24], as reported in Table 1.2.

The most significant evidence, shown in Fig. 1.6, was given by the $H \to \gamma\gamma$ and the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ ($\ell = e, \mu$) decay channels, which guaranteed the best mass resolution and allowed a good sensitivity for a large spectrum of masses. The H boson mass was estimated using different decay modes, among which the above mentioned ones were the most relevant. Table 1.2 shows measurements presented by CMS and ATLAS experiments on the date of the announcement of the Higgs discovery.

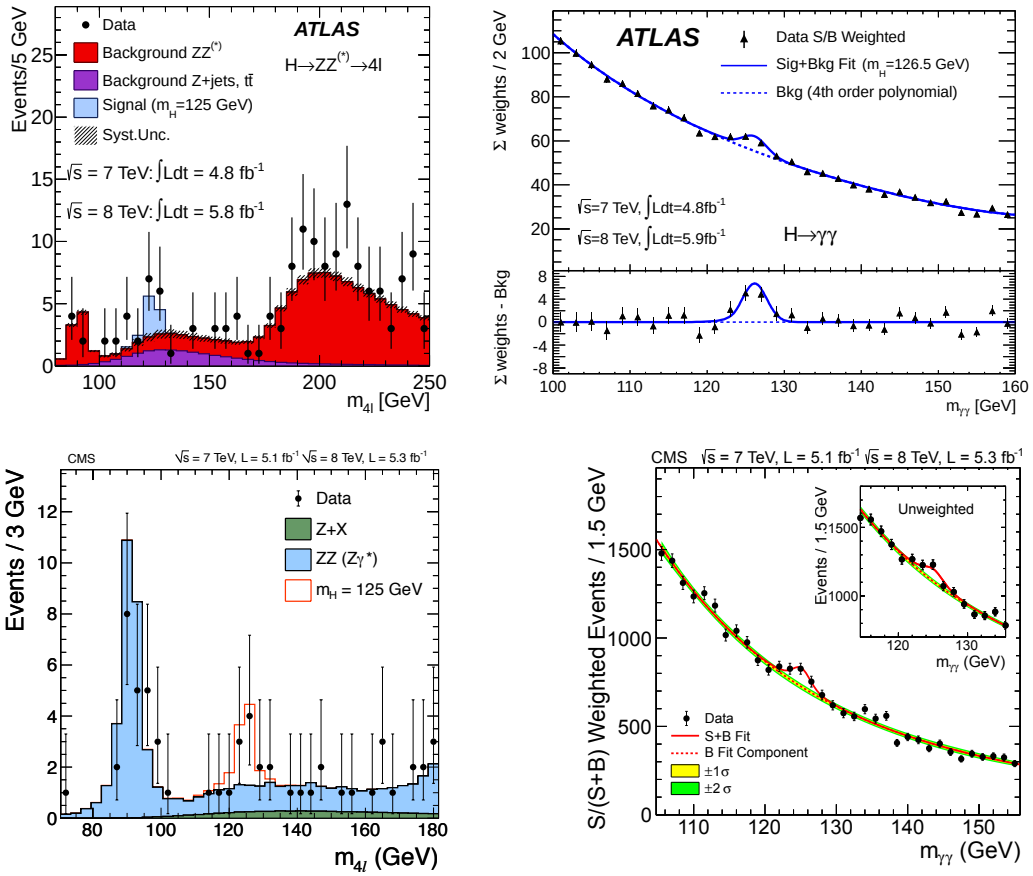


FIGURE 1.6: Invariant mass distribution of four-lepton (left) and diphoton (right) candidates selected by the ATLAS (top) and CMS (bottom) Collaboration using the combined 7 TeV and 8 TeV datasets [13–15].

Collaboration	Mass	Observed significance
ATLAS	$m_H = 126 \pm 0.4$ (stat) ± 0.4 (syst) GeV	5.9σ
CMS	$m_H = 125 \pm 0.4$ (stat) ± 0.5 (syst) GeV	5.0σ

TABLE 1.2: Observation significance of the H boson and mass measurement by the ATLAS [13] and CMS [14, 15] Collaborations in 2012.

"*It is somewhat surreal to find that work we did nearly fifty years ago is once again at the center of attention. This is a triumph for the standard model of particle physics, but even more for the experimenters. The achievement of the two great experimental collaborations reported here is quite magnificent.*"
(T. Kibble, C. R Hagen, G. Guralnik, 2012)

The exploration with Run 1 and Run 2 data Since the discovery of the H boson, more accurate measurements have been performed to validate and confirm the hypothesis that the new particle was effectively the predicted scalar particle of the SM. A spin-parity $J^P = 0^+$ has been demonstrated [28] and so far all coupling strengths and decay rates are found to be compatible with the SM expectations from the combination of different decay channels using Run 1 [29–31] and Run 2 data [32, 33]. Coupling modifiers (see Section 1.2.2) can be defined separately for each particle and then extracted with a global likelihood fit for fermions and weak bosons from the κ-framework model. The most updated measurement of the coupling modifiers performed after the recent evidence of the H boson decaying to a pair of muons [34] is shown in Fig. 1.7 with the comparison to their corresponding prediction from the SM.

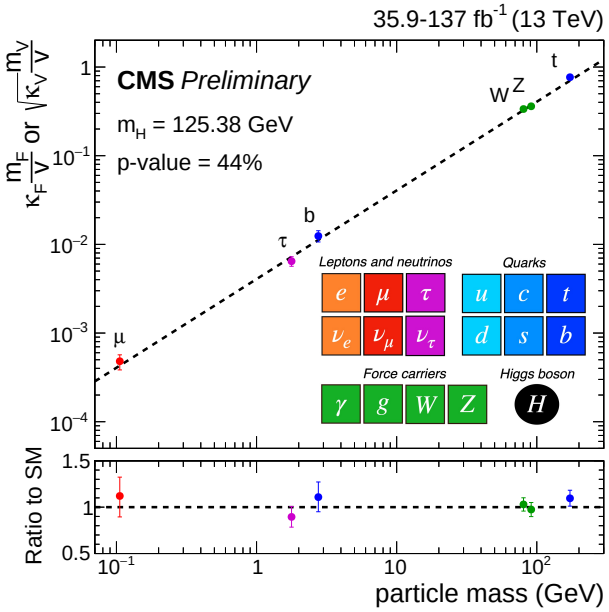


FIGURE 1.7: Measured values of the coupling modifiers as a function of the particle mass [35]. The green points represent the coupling modifiers for the interactions between the H and vector bosons, while the red, magenta, and blue points refer to the couplings with μ, τ, and quarks of the third generation, respectively. The associated error bars represent 68% CL intervals for the measured parameters. The lower panel shows the ratios of the measured coupling modifiers values to their SM predictions.

Thanks to the unprecedented possibilities given by the Run 2 dataset, the experimental research has been geared toward probing all main production and decay modes of the H boson and its couplings with SM particles. All four main production mechanisms have been observed: the H boson production in association with a top quark pair is the last observation among the production modes [36, 37]. In the landscape of the H boson decays to fermions, important steps are represented by the observation of the H boson decay into a pair of τ particles [38, 39], the measurement of the Yukawa coupling with the bottom quark [40, 41], and the new observation of the H boson decay into a pair of muons presented by the CMS experiment this year [34]. The precision on the H boson couplings achieved with a fraction of the Run 2 data is shown in Fig. 1.8. A combination of results based on the full Run 2 dataset will be prepared as soon as all the single analyses similar to the one presented in this thesis will be completed.

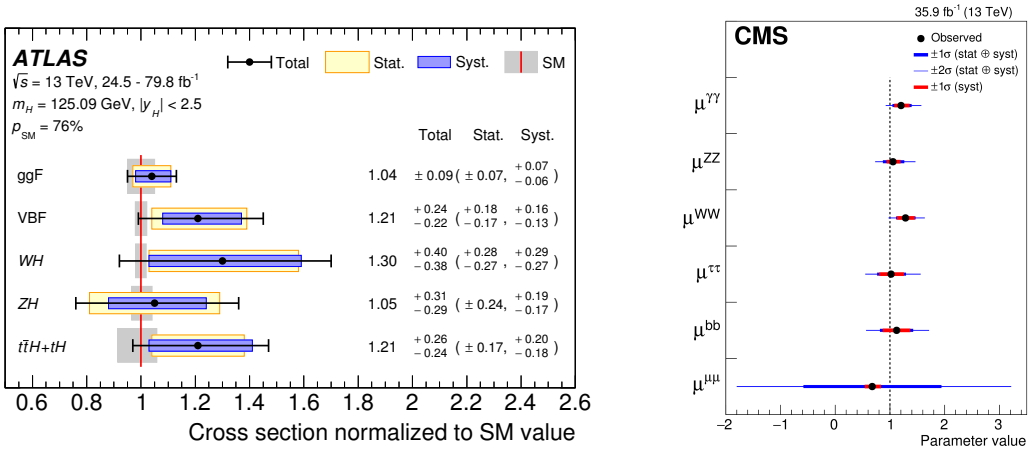


FIGURE 1.8: Measurements of the combined signal strengths per production (left) and decay (right) mode of the H boson performed by the ATLAS [32] and CMS [33] Collaborations, respectively, using a part of Run 2 data.

The current picture of the SM particles and interactions is accurate and almost completely probed, but a significant piece is still missing: the self-interaction of the H boson, predicted by the SM and totally determined from m_{H} and v, as shown in Eq. 1.43. It is responsible for the generation of the H boson mass. The production of a pair of H bosons (HH) can directly probe the H self-interaction, but it is an extremely rare process in the SM (see Section 1.3.3). For this reason, the measurement of the H boson self-coupling is of primary interest in the current searches at the LHC and in the prospects at the HL-LHC, but also in the context of the future colliders. It is a crucial test of the validity of the SM, also representing a relevant way to look for effects induced by BSM physics. The elegance and success of the SM description of the electroweak and strong interactions between elementary particles leaves the door open to some other mysteries of our universe. An overview of the possible BSM effects in the Higgs sector is presented in the next section together with the discussion of the level of precision achievable at future particle colliders in the measurements of the Higgs properties.

1.3 Precision measurements at future colliders: possible BSM effects in the H boson sector

The monumental discovery of the H boson represented a turning point for the research in high energy particle physics. Then, remarkable studies of the H boson properties have been

performed at the LHC in the last years and are still ongoing. Currently, the investigation of the H boson sector through precision measurements and direct searches for new phenomena at future collider is among the most relevant aspects of their scientific programs. In fact, many open questions about the fundamental laws of our universe both from an experimental and theoretical point of view remain unsolved. For example, the baryon asymmetry causing the abundance of matter with respect to antimatter in the universe, the nature of dark matter and the evidence of nonzero neutrino masses, the hierarchy problem, or the stability of the H boson mass upon quantum corrections. Direct searches for new physics are the most powerful tools for discoveries. However, the long search for particles predicted by the theory of the supersymmetry (SUSY) and for dark matter (DM) candidates has not provided results. For this reason, H boson precision measurements have become one of the hottest topics in particle physics. They represent one of the most powerful and broadly efficient methods to probe the H boson properties looking for possible deviations in specific phase spaces with such a precision to be sensitive to new physics. A discussion of some measurements particularly suitable in this context is presented in the following, considering the coupling of the H boson to SM elementary particles, together with the study of H boson rare decays and the sensitivity to H boson CP, the H boson mass and width determination, and the sensitivity to the H boson self-coupling via single H and H H production measurements.

This section aims to provide an assessment of the potential of future colliding beam facilities to explore the H boson, assuming as a reference the results expected at the completion of the HL-LHC program. Quantitative results on different aspects of the H boson physics for future collider projects of sufficient maturity (see Section 2.3.1) have been produced using uniform approaches and methodologies and are partially presented here. All details can be found in Ref. [21].
In this context, at the beginning of my Ph.D., I contributed to the study of the prospects on the H H production at future hadronic colliders (see Section 1.3.3): FCC-hh and HL-LHC. I had the opportunity to present these results during the poster session of the LHCP conference in Bologna (2018) [42] and Puebla (2019) [43], respectively, receiving the best poster award of the LHCP 2019 conference. However, these studies are not part of the discussion of this thesis. Details can be found in Ref. [44].

1.3.1 Higgs boson couplings to fermions and vector bosons

The exploration of the H boson couplings to fermions and bosons at the end of the LHC operation is expected to achieve an overall precision at the level of 10% and down to 5% in the most accurate measurements, while a precision of the few per-cent level is expected after the subsequent high luminosity operation of the LHC (HL-LHC) [45, 46]. In the final stage of the future colliders, the measurements of most of the H boson couplings will reach a few per-mille precision, as shown in Fig. 1.9 with corresponding numerical values reported in Table 1.3.

The extraction of the H boson couplings relies on a simultaneous fit of all the measurements of $\sigma\mathcal{B}$ and their comparison to the SM values. The effective couplings measured within the EFT framework, defined as

$$g^2_{HX} \equiv \frac{\Gamma_{\mathrm{H}\to\mathrm{X}}}{\Gamma^{\mathrm{SM}}_{\mathrm{H}\to\mathrm{X}}}, \tag{1.54}$$

are also shown in Fig. 1.10. A global fit including also diboson and electroweak observables is performed to extract these coefficients.

A significant interest on rare decays of the H boson, such as the coupling to second and first generation fermions, is due to the fact that they give access to H boson couplings

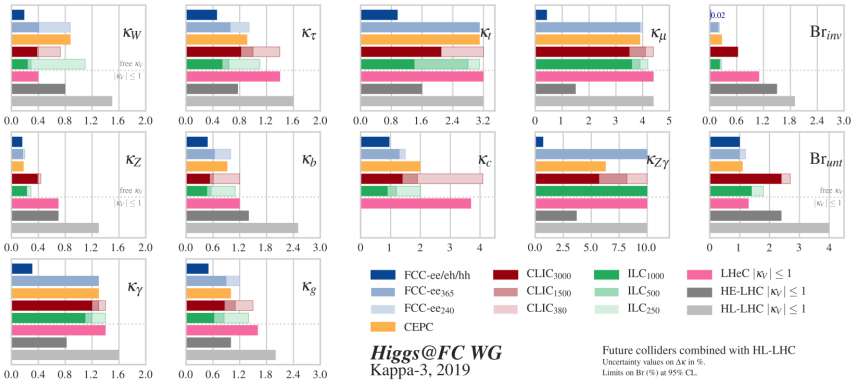


FIGURE 1.9: Expected relative precision (%) of the κ parameters for various future accelerators beyond the LHC under the assumption that BSM H boson decays exist [21].

that are expected to be small in the SM and have not yet been directly probed. The hypothesis that the same H boson doublet originates the mass of the lighter states of the SM can be tested by a precise determination of these couplings. Moreover, the sensitivity to BSM physics at higher scales is significantly enhanced for processes that are extremely suppressed in the SM, such as the search for flavor-changing neutral interactions. Finally, there are some rare signatures, for example the decay of the H boson to invisible particles, which are used to constrain models including DM candidates.

As particularly evident in the context of the $\mathrm{H} \to 4\ell$ analysis, the increasing precision in the measurement of the H boson coupling can provide interesting hints of BSM physics. On the one hand, probing the H boson at high p_T enhances the sensitivity to new physics. On the other hand, an anomalous tensor structures of H boson interactions with fermions (Hff) and vector bosons (HVV) may arise from BSM effects.
Studies performed after the H boson discovery set constraints on the anomalous couplings

HL-LHC +	ILC			CLIC			Circular colliders			
	250	500	1000	380	1500	3000	CEPC	FCC-ee$_{240}$	FCC-ee$_{365}$	FCC
κ_W	1.0	0.29	0.24	0.73	0.40	0.38	0.88	0.88	0.41	0.19
κ_Z	0.29	0.22	0.23	0.44	0.40	0.39	0.18	0.20	0.17	0.16
κ_g	1.4	0.85	0.63	1.5	1.1	0.86	1.0	1.2	0.9	0.5
κ_γ	1.4	1.2	1.1	1.4*	1.3	1.2	1.3	1.3	1.3	0.31
$\kappa_{Z\gamma}$	10.*	10.*	10.*	10.*	8.2	5.7	6.3	10.*	10.*	0.7
κ_c	2.0	1.2	0.9	4.1	1.9	1.4	2.0	1.5	1.3	0.96
κ_t	3.1	2.8	1.4	3.2	2.1	2.1	3.1	3.1	3.1	0.96
κ_b	1.1	0.56	0.47	1.2	0.61	0.53	0.92	1.0	0.64	0.48
κ_μ	4.2	3.9	3.6	4.4*	4.1	3.5	3.9	4.0	3.9	0.43
κ_τ	1.1	0.64	0.54	1.4	1.0	0.82	0.91	0.94	0.66	0.46
$\mathcal{B}_\mathrm{inv}$	0.26	0.23	0.22	0.63	0.62	0.62	0.27	0.22	0.19	0.024
$\mathcal{B}_\mathrm{unt}$	1.8	1.4	1.4	2.7	2.4	2.4	1.1	1.2	1.	1.

TABLE 1.3: Expected relative precision (%) of the κ parameters for various future accelerators beyond the LHC operating at different center-of-mass energy values. Results are provided under the assumption that BSM H boson decays exist. The upper limits (<%) at 95% CL on the $\mathcal{B}$ of the H boson decays to invisible ($\mathcal{B}_\mathrm{inv}$) and untagged ($\mathcal{B}_\mathrm{unt}$) particles are also provided. All scenarios take into account the combination with the expected results at the end of HL-LHC. The asterisk ($*$) indicates measurements dominated by the HL-LHC because of the lack on analysis inputs [21].

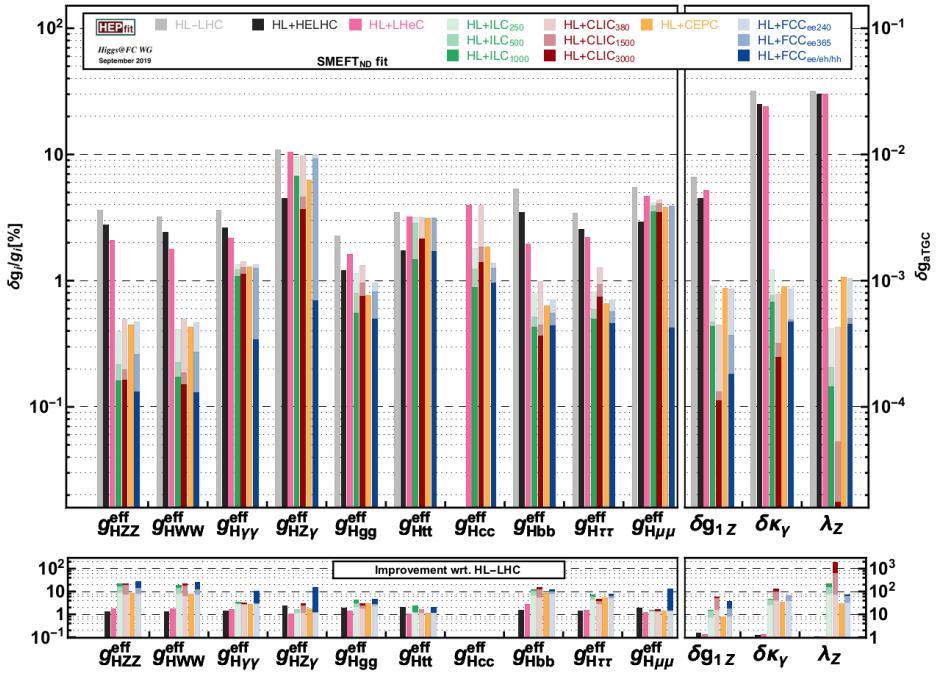


FIGURE 1.10: Expected sensitivity to deviations in the effective H boson couplings from a global fit based on the EFT approach at 68% CL for various future accelerators beyond the LHC [21].

and spin-parity of the H, which is predicted to have CP-even ($J^{PC} = 0^{++}$) and flavor diagonal interactions. However, small anomalous couplings to a pair of EW gauge bosons are still allowed and possible effects of CP violation in Hff couplings, corresponding to the detection of nonzero CP-odd components in the H boson interactions with the SM particles, may appear at the tree level. In addition, the measurement of the effective Hgg couplings is interesting because sensitivity to the CP-odd Hgg interaction derives from ggF processes at the inclusive level. Anomalous couplings to fermions, bosons, or new particles and Htt coupling with CP-odd can contribute to the loop and provide signs of CP violation.
The $\mathrm{H} \to 4\ell$ analysis is particularly sensitive to the study of CP-violation and anomalous couplings, which is able to provide the simultaneous measurement of five independent HVV couplings, two Hgg, and two Htt couplings exploiting the single $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$ decay channel, as reported in the most recent CMS results in Ref. [47].

1.3.2 Higgs boson mass and width

The most precise measurement of the H boson mass currently available comes from the CMS combination of the 2016 results of the $\mathrm{H} \to 4\ell$ and $\mathrm{H} \to \gamma\gamma$ analyses with an earlier measurement of m_{H} based on the 2011 and 2012 data [30, 48, 49]:

$$125.38 \pm 0.14 \ \text{GeV}.$$

A significant improvement in the precision of the mass measurement is expected at future accelerators. In the context of HL-LHC, the mass measurement is performed in the $\mathrm{H} \to 4\ell$ decay channel, indicating an achievable precision down to 10-20 MeV. At lepton colliders operating at a center-of-mass energy between [240,350] GeV, the mass measurement relies on the recoil mass method, where the H boson decays into a pair of electrons or muons in ZH events. In this case, systematic uncertainties are expected to be negligible. An overview

of the expected precision δm_{H} on the mass measurement for various future accelerator scenarios is shown in Table 1.4, with the corresponding impact on the $\mathrm{H} \to \mathrm{ZZ}^*$ partial decay width, evaluated as $1.2\% \cdot (\delta m_{\mathrm{H}}/100)$ MeV [50].

Collider scenario	Strategy	δm_{H} (MeV)	$\delta(\Gamma_{\mathrm{ZZ}^*})$ [%]
HL-LHC	m_{ZZ}	10-20	0.12-0.24
ILC_{250}	ZH recoil	14	0.17
CLIC_{380}	ZH recoil	78	0.94
CLIC_{3000}	$m_{b\bar{b}}$ in $\mathrm{H}\to\nu\nu$	23	0.28
FCC-ee	ZH recoil	11	0.13
CEPC	ZH recoil	5.9	0.07

TABLE 1.4: Expected precision of the H boson mass measurement for various future accelerators beyond the LHC with the impact of the m_{H} uncertainty on the partial decay width of the $\mathrm{H}\to\mathrm{ZZ}^*$ process [21].

The measurement of the total width of the H boson is crucial in the search for BSM manifestations. Possible deviations in the measured width from the SM expectation can arise if the H boson has new BSM decay channels or the known channels present non-SM rates. As a consequence, the precise measurement of the H boson width complements the measurement of the H boson couplings to known SM particles. Similarly, it can provide significant information on possible H boson decays to invisible or undetected particles. Moreover, joint constraints can be set on Γ_{H} as well as on parameters that express HVV anomalous couplings.
Indirect measurements of the H boson width are possible at hadron colliders because the width affects the physics that we can observe. The best results come from an indirect approach which exploits the relative measurement of the off-shell and on-shell H boson production and profits from the fact that the SM H boson width is proportional to the ratio of the event yields in the off-shell and on-shell regions. An additional theoretical assumption is needed if untagged decays are allowed, for example requiring $|\kappa_V| \leq 1$ or including constraint from off-shell H boson measurements.
At lepton colliders, the recoil mass method can be exploited to obtain a precise ZH cross section measurement in a model independent way, having knowledge of the total initial energy of the event. Then, it can be used in combination with measurements of exclusive cross sections of H boson decays to retrieve Γ_{H}, for example:

$$\frac{\sigma(e^+e^- \to \mathrm{ZH})}{\mathcal{B}_{(\mathrm{H} \to \mathrm{ZZ}^*)}} = \frac{\sigma(e^+e^- \to \mathrm{ZH})}{\Gamma(\mathrm{H} \to \mathrm{ZZ}^*)/\Gamma_{\mathrm{H}}} \simeq \left[\frac{\sigma(e^+e^- \to \mathrm{ZH})}{\Gamma(\mathrm{H} \to \mathrm{ZZ}^*)}\right] \times \Gamma_{\mathrm{H}}. \tag{1.55}$$

The main limitation is represented by the measurement of the $\mathcal{B}(\mathrm{H} \to \mathrm{ZZ}^*)$, since it has not been measured with an optimal level of accuracy.

1.3.3 Higgs boson self-coupling

Currently one of the main goals in the physics program at the LHC experiments and at all future colliders is the measurement of the H boson self-coupling (see Eq. 1.43) or Higgs trilinear coupling (λ_{HHH}). On the one hand, it represents an important test of the SM electroweak symmetry breaking sector. It is directly related to the structure of the H boson scalar potential and the shape of the field potential is linked to many open questions of particle physics and cosmology, such as the metastability of our universe or its implications in the baryogenesis. On the other hand, any possible deviation in the H boson self-coupling due to BSM effects could open the door to new physics searches.

The only way to directly access and measure λ_{HHH} is the measurement of the H boson pair production at hadron colliders such as LHC. In this case, the dominant nonresonant H H production mode proceeds through top quark loop diagrams in the gluon fusion channel, as shown in Fig. 1.11. This contribution is affected by a destructive interference between the box and the triangle diagrams. Consequently, the total SM production rate of the H H process is really small. The corresponding cross sections at a center-of-mass energy of 13 TeV and 14 TeV are 31.05 fb and 36.69 fb, and they are expected to increase by a factor 40 at FCC-hh at a center-of-mass energy of 100 TeV [51].

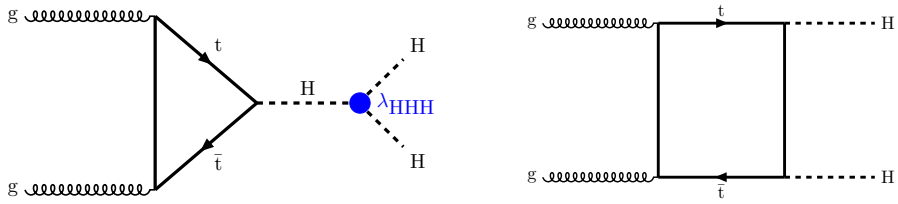


FIGURE 1.11: LO Feynman diagrams for the HH production in the SM through the Higgs boson self-coupling (left), with the coupling λ_{HHH} in blue, and the top box diagram (right) at hadron colliders.

The most recent prospects on the HH production at the HL-LHC with the CMS experiment have been obtained combining five decay channels ($\mathrm{b\bar{b}b\bar{b}}$, $\mathrm{b\bar{b}\tau\tau}$, $\mathrm{b\bar{b}\gamma\gamma}$, $\mathrm{b\bar{b}WW}$, $\mathrm{b\bar{b}ZZ}$) and are summarized in Fig. 1.12. My early contribution to the CMS experiment was the study of the $\mathrm{HH} \rightarrow \mathrm{b\bar{b}ZZ}(4\ell)$ process at the HL-LHC, which took part at this combination. For this reason, I would like to report these results.
The statistical combination results in a 95% CL upper limit on the SM HH cross section of 0.77 times the SM prediction. The μ=0 scenario, indicating the absence of a SM HH signal, is excluded at the 99% CL, which corresponds to a significance of 2.6σ. Prospects for the measurement of the trilinear H boson coupling have been also evaluated. The expected 68% and 95% CL intervals for the self-coupling modifier $\kappa_\lambda = \lambda_{\mathrm{HHH}}/\lambda^{\mathrm{SM}}_{\mathrm{HHH}}$ are [0.35, 1.9] and [-0.18, 3.6], respectively.
Considering the combination of the results produced by the ATLAS and CMS experiments,

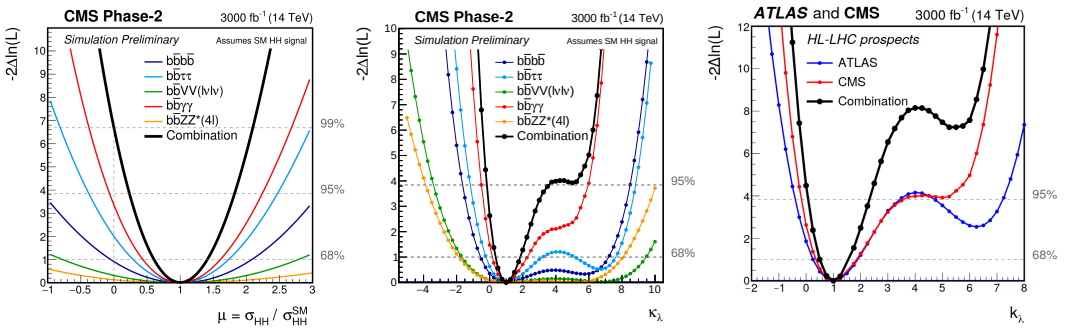


FIGURE 1.12: Scan of the negative log-likelihood as a function of the signal strength μ (left) and the self-coupling modifier κ_λ (middle) assuming SM HH signal in the five decay channels analyzed and their combination [20]. The negative log-likelihood as a function of κ_λ is also shown for the combination of the ATLAS and CMS results (right) [20].

the expected significance on the HH production at the end of the HL-LHC is 4.0σ and constraints on the Higgs self-coupling of $[0.5, 1.5]$ at 68% CL are set, corresponding to a precision of 50% on the measurement. The second minimum in the likelihood lineshape due to the degeneracy in the total number of HH signal events is excluded at 99.4% CL thanks to the differential information of m_{HH}. The results are very promising in terms of a possible observation of the process in the following years.

Also in the context of lepton colliders, the trilinear coupling contributes to the HH production process via $e^+e^- \to \mathrm{ZHH}$ (double Higgs-strahlung) at low energies or via VBF in the $e^+e^- \to \mathrm{H\,H}\,\nu_e\bar{\nu}_e$ for center-of-mass energy values above 1 TeV. Diagrams for the LO contribution to HH production at lepton colliders are presented in Fig. 1.13.

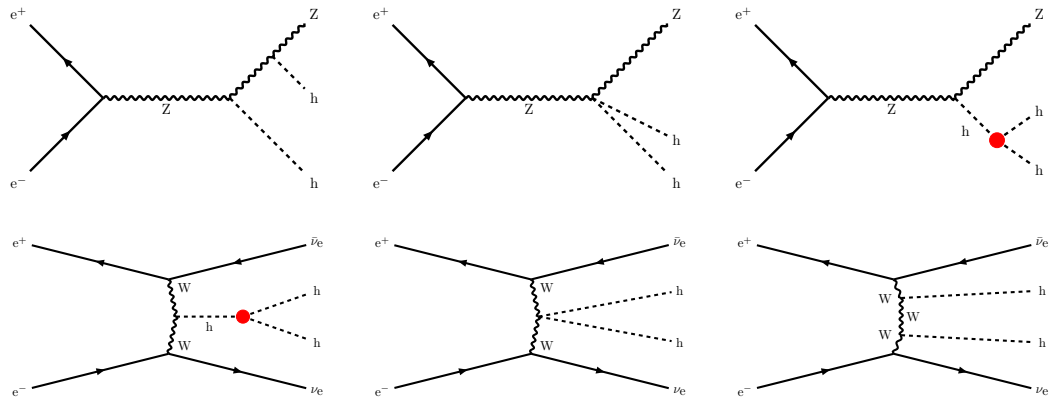


FIGURE 1.13: LO Feynman diagrams for the HH production process via $e^+e^- \to \mathrm{ZHH}$ (top) and via VBF in the $e^+e^- \to \mathrm{HH}\nu_e\bar{\nu}_e$ (bottom) at lepton colliders. The coupling λ_{HHH} is shown in red.

However, at low energy colliders below the HH production threshold, the most relevant approach is to exploit single H boson processes, where deviations in the H boson self-interactions can produce measurable effects at NLO (Fig. 1.14).
The trilinear coupling does not enter single H processes at LO but it affects both H boson production and decay at higher orders. The NLO corrections to an observable Σ can be expressed as

$$\Sigma_{\mathrm{NLO}} = \mathrm{Z_H} \cdot \Sigma_{\mathrm{LO}}(1 + \kappa_\lambda C_1), \tag{1.56}$$

with $\mathrm{Z_H}$ representing an universal coefficient and C_1 is a process dependent coefficient varying with the center-of-mass energy. The impact of a deviation $\delta\kappa_\lambda$ from the SM value is

$$\delta\Sigma \equiv \frac{\Sigma_{\mathrm{NLO}}}{\Sigma_{\mathrm{NLO}}(\kappa_\lambda = 1)} - 1 \simeq (C_1 + 2\,\delta\mathrm{Z_H})\,\delta\kappa_\lambda + \delta\mathrm{Z_H}\,\delta\kappa_\lambda^2, \tag{1.57}$$

with $\delta Z_H \sim -0.00154$. For example, considering the ZH production cross section at NLO as observable of interest and a 240 GeV lepton collider, the variation induced by a modified H boson cubic coupling amounts to

$$\sigma_{\mathrm{ZH}}^{\mathrm{NLO}} \simeq \sigma_{\mathrm{ZH}}^{\mathrm{NLO\ SM}}(1 + 0.014\,\delta\kappa_\lambda). \tag{1.58}$$

Therefore, an accuracy below 1% is required in the ZH cross section measurement in order to be competitive with constraints achievable at the HL-LHC, but this is for example expected in the context of an e^+e^- Higgs factory operating at 240 GeV.

The expected sensitivity on the self-coupling of the H at 68% CL accounting for the combination with the HL-LHC results is reported in Fig. 1.15. On the one hand, from a

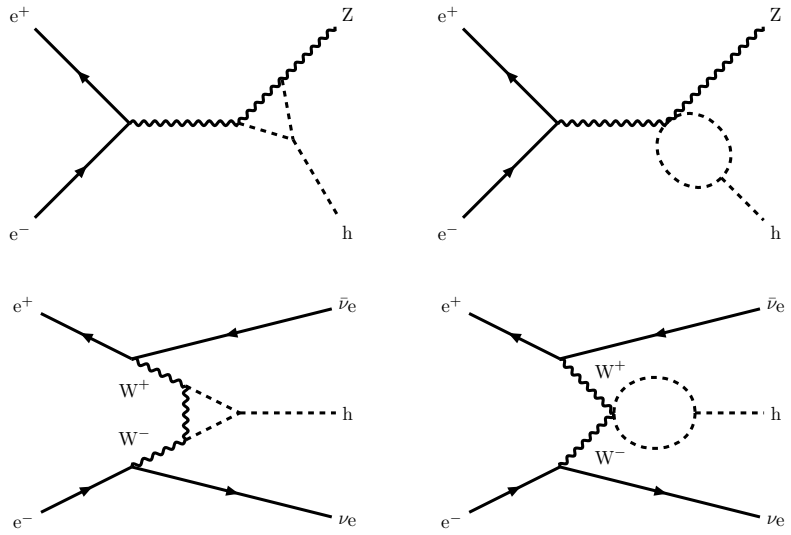


FIGURE 1.14: NLO Feynman diagrams for single H processes including the H boson self-coupling: ZH Higgs-strahlung process (top) and VBF processes (bottom).

direct measurement of the HH production at FCC-hh, the H boson trilinear coupling will be measured with a precision below the 5% (around 3% according to the latest prospects provided in Ref. [52]) at 68% CL. On the other hand, performing a global analysis of single H processes at higher order and considering also all possible deformations of the single H boson couplings in the context of a multi-parameter EFT likelihood fit, a precision on λ_{HHH} of 33% (24%) can be achieved at FCC-ee in the scenario with two (four) interaction points (see Section 2.3.2).

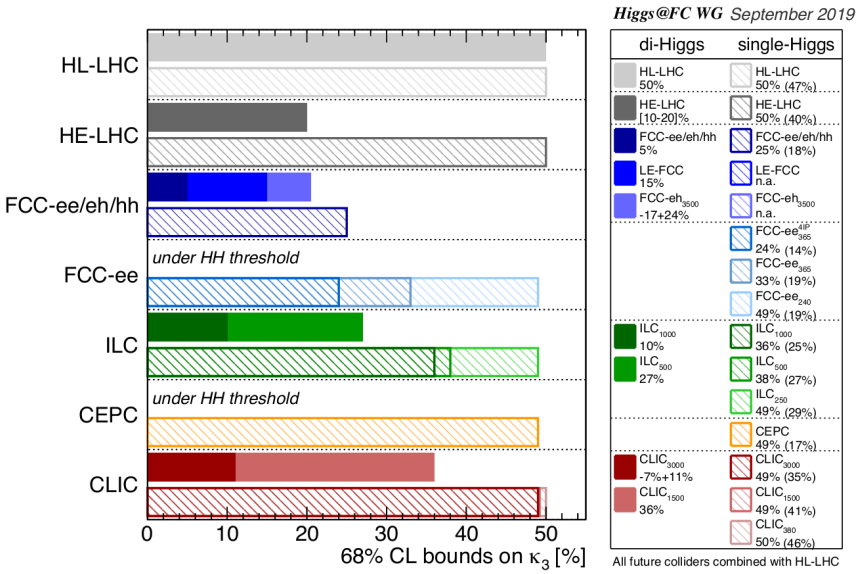


FIGURE 1.15: Sensitivity on the H boson trilinear coupling at 68% CL for various future accelerators beyond the LHC [21].

Chapter 2

The experimental apparatus

The goal of this chapter is to emphasize the complexity of the machines and of each kind of apparatus needed to perform the experimental research on the fundamental questions presented in the first chapter. An impressive sentence by A. Turing [1] (1950) can be taken as a good way to approach the following discussion:

Machines take me by surprise with great frequency.

The beginning of the pathway to any kind of investigation in high energy physics (HEP) is represented by a set of particles. The H boson candidates are searched among these particles, which are part of complex events produced in high energy collisions and recorded by a detector. It is clear that the availability and quality of our collected data constitute the essential point to achieve high precision in the final results. Thus, the effort invested to project and build highly performant machines and experiments is huge and the timeline usually spans over several decades.

In the first part, a brief introduction to the Large Hadron Collider (LHC) [53, 54], a milestone of the Conseil Européen pour la Recherche Nucléaire (CERN) history, will be provided. The central topic of this thesis is carried out within the Compact Muon Solenoid (CMS) experiment at the LHC, so the second section will describe how the CMS detector was built and is working. To conclude, a special focus is dedicated to the different proposals of new colliders and experiments for the future of HEP after the LHC era. The discussion about the kind of future accelerators needed at the end of the LHC operation is very animated at the present time and its outcome will shape the activities of the particle physics research for decades. As a young scientist, during my Ph.D. I have been involved in different kind of studies needed as a fundamental input to the ongoing discussion.

2.1 The Large Hadron Collider at CERN

The quest for the H boson at CERN started in the late 1980s with the Large Electron-Positron (LEP) collider, which occupied the tunnel that now houses the LHC until 2000, placing the laboratory at the forefront of the research in HEP and of the technology development. In parallel, searches for the H boson were performed by the experiments at the Tevatron collider at Fermilab in the United States during the 1990s. During the early 1980s, a committee of physicists at CERN was focusing on a joint project for the next big particle accelerator, involving the United States and Europe, and started thinking about the LHC. Then, the CERN Council approved the LHC project in December 1994 and the LHC Technical Design Report (TDR) was published already in October 1995 [55], containing a description of the accelerator's architecture and the basis for the subsequent detailed development of the proposal. Because of financial reasons, it was initially decided that the project would have been

[1] Alan Mathison Turing (23 June 1912 - 7 June 1954) was an English mathematician and computer scientist who highly influenced the development of theoretical computer science.

set up in two stages. However, as a result of diplomatic activities and intense negotiations which ensured significant contributions from the non-member states, the Council approved the construction of the machine in a single step. At that time, nineteen nations belonged to CERN. Thanks also to the contributions from Japan, USA, India, and other non-member states, four experiments received official approval and construction work started: ALICE, ATLAS, CMS, and LHCb. Later, two smaller experiments were added: TOTEM (located near CMS) and LHCf (near ATLAS).

The LHC is a gigantic machine built at the CERN laboratory near Geneva, spanning the border between Switzerland and France. It is installed in a 26.7 km circular tunnel at a mean depth of 100 m. One of its main goal is the study of rare events produced in proton-proton (p-p) collisions at a center-of-mass energy up to 14 TeV, in order to investigate the smallest fundamental building blocks that constitute our universe. It is designed to provide also proton-lead (p-Pb) and lead-lead (Pb-Pb) collisions.
The LHC is a double ring low temperature superconducting collider, able to accelerate two beams of hadrons (protons) or ions travelling in opposite directions at speeds very close to the speed of light. The beams collide in four points of the two rings where the main LHC experiments are located. A complex chain of smaller particle accelerators is used to boost the particles to their final energies and provide beams to a whole set of smaller experiments. The CERN accelerator complex is shown in Fig. 2.1.
The progressive acceleration of protons and beam injection in the LHC ring occurs in different steps.

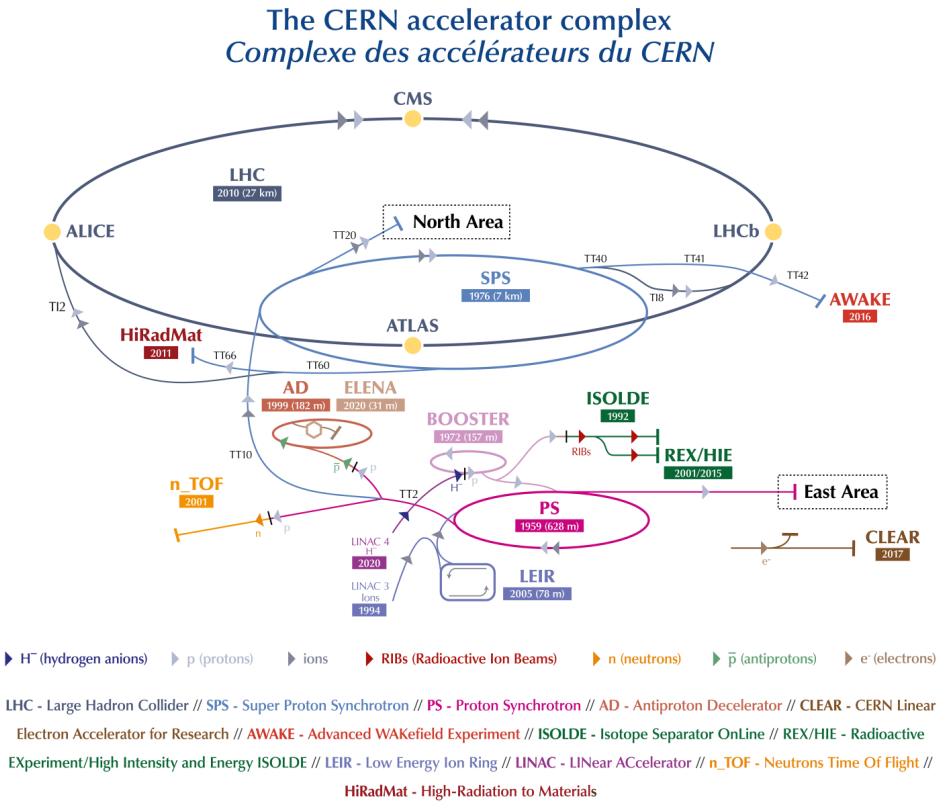


FIGURE 2.1: Schematic layout of the CERN particle accelerators chain and of the experiments around the circumference of the LHC, which is the last in the accelerator chain (dark blue line) [56].

1. Protons are obtained by stripping orbiting electrons from hydrogen atoms;
2. a linear accelerator (LINAC2) carries out the initial proton acceleration up to 50 MeV;
3. protons are then injected into the Proton Synchrotron Booster (PSB), where the beam is accelerated to 1.4 GeV;
4. protons are sent to the Proton Synchrotron (PS) and accelerated up to 25 GeV;
5. at this point, protons are sent to the Super Proton Synchrotron (SPS), where they reach an energy of 450 GeV;
6. protons are finally transferred in bunches to the two adjacent parallel beam pipes of the LHC, where they circulate for many hours around the ring in well defined bunches, one beam in the clockwise direction and the second one in an anticlockwise direction.

2.1.1 The LHC machine

The shape of the LHC collider is not a perfect circle. It is organized in eight 2.45 km long arcs, containing 154 dipole bending magnets per arch, and eight straight sections approximately 528 m long with a transition region at its end called *insertion*. The layout of each straight section depends on the specific use of the insertion, which can be oriented to the beam collisions within an experiment, the injection, the beam dumping, or the beam cleaning. The part of the machine between two insertion points is defined as a *sector*. Each sector is provided with an independent powering. To build the LHC, the most technologically advanced superconducting electromagnetic dipoles and radiofrequency (RF) cavities have been developed. The installation of the magnets was carried out sector by sector and all the dipoles of a sector are connected in series and placed in the same continuous cryostat. A particle accelerator relies on three different elements:

Vacuum The main LHC vacuum systems are the beam vacuum, the insulation vacuum for cryomagnets, and the insulation vacuum for the helium distribution line. The presence of three separate systems with different requirements with respect to the classical vacuum systems represents a special feature of the LHC. The beam vacuum is an ultrahigh vacuum with a pressure of 10^{-7} Pa in the beam pipe at a cryogenic temperature of 5 K. The pressure is lower than 10^{-9} Pa close to the interaction points (IPs)[2], where collisions take place. It avoids collisions with gas molecules allowing proper beam lifetime and low background to the experiments. The insulation vacuum for the powerful superconducting magnets is the largest volume to be pumped in the LHC ($\sim 9000\,\mathrm{m}^3$) and is cooled down with liquid helium to 1.9 K. It ensures a good thermal insulation of the cooling system in order to maintain the low temperatures. Finally, the insulation vacuum for the helium distribution line acts similarly as a thermal insulator to reduce the amount of heat coming from the surrounding room-temperature environment into the cryogenic parts.

The magnet system A strong magnetic field is required in order to maintain particles on a circular trajectory. The two beams are curved thanks to opposite magnetic fields. To bend protons in the LHC accelerator, a magnetic field B is needed with an intensity given by the equation:

$$p\ [\mathrm{TeV}] = 0.3\, B\ [\mathrm{T}]\ r\ [\mathrm{km}], \tag{2.1}$$

[2]Note that IP will be used in place of *impact parameter* in the next chapters, while here it always refers to the interaction point.

where p is the momentum of the beam particle and r is the radius of the LHC ring. The LHC presents separate magnetic fields and vacuum chambers in the eight arcs, alternated to straight sections which are at the insertion regions where the experimental detectors are located. The maximum energy that can be achieved in a collider is directly proportional to the strength of the dipole field and the radius of the accelerator. The LHC is equipped with a large variety of magnets, from dipoles to decapoles, and each family of magnets undertakes a specific task.

* Dipoles are the most numerous (1232) and provide the very high field of 8.3 T, thanks to a high current which flows in them. For example, a current of 11000 A is necessary to bend the 7 TeV beams around the 26.7 km ring.
* Quadrupoles (392) help to focus the beam squeezing it either vertically or horizontally in order to maximize the chance of two protons colliding head-on.
* High order multipole magnets contribute to correct for possible imperfections of the magnetic field in the main ring magnets (dipoles and quadrupoles) and in the interaction region magnets.

The superconducting technology is based on niobium-titanium (NbTi) cables with an operating temperature of 1.9 K (–271.3°C) and makes the LHC unique among superconducting synchrotrons. The LHC cryogenic system is based on the use of superfluid helium, which is pumped into the magnet system. The amount of liquid helium needed for cooling the LHC is so huge that CERN has its own production! It has a very high thermal conductivity, ideal for the refrigeration and stabilization of large superconducting systems. The cooling of the elements is a fundamental activity in order to increase the performance of NbTi and maximize the field strength of the superconducting magnets.

Cavities The primary purpose of the cavities is to accelerate charged particles. The LHC uses eight RF cavities per beam. They are metallic chambers that contain an electromagnetic field and operate at 4.5 K, arranged in groups of four in cryomodules with two cryomodules per beam. The RF fields are exploited to accelerate beams and to squeeze proton bunches as compact as possible in order to make head-on collisions easier and to guarantee high luminosity (see below) at the collision points.

As a consequence of the acceleration scheme based on RF fields used in modern colliders, the particle beams are characterized by a bunch structure. During a RF cycle, particles are only accelerated passing through an accelerating cavity, when the RF field presents the correct orientation. In the LHC, under nominal operating conditions, each proton beam consists of 2808 bunches and each bunch contains as many as 10^{11} protons. Bunches of particles measure a few centimeters long and a millimeter wide when they are far from a collision point. However, the bunch size is not constant around the ring: circulating around the ring, each bunch gets squeezed and expanded. By squeezing its length as much as possible around the collision points, it is possible to increase the probability of a p-p collision.
The number of bunches affects significantly the *instantaneous luminosity* ($\mathcal{L}$) in a machine. The instantaneous luminosity depends on the properties of the beams:

$$\mathcal{L} = f\gamma_r \frac{n_b N_b^2}{4\pi\epsilon_n \beta^*} F, \tag{2.2}$$

where in the numerator n_b and N_b indicate the number of bunches and the number of particles contained in each bunch, respectively, f represents the bunch crossing (BX) frequency, γ_r denotes the relativistic Lorentz factor of the protons, while the two terms in the denominator represent the transverse emittance (ϵ_n) and the focal length (β^*) at the collision point,

called also β function, describing the shape and focus of the beam, respectively. Finally, the geometric reduction of the instantaneous luminosity is taken into account in the factor F, which depends on the transverse and longitudinal dimensions of the beams at the IP and on the beam crossing angle θ_c as:

$$F = \left(1 + \theta_c \frac{\sigma_z}{2\sigma_{xy}}\right)^{-1}. \tag{2.3}$$

During the Run 2 operation, the bunch spacing at the LHC was 25 ns (half of the bunch crossing interval used during Run 1), which corresponds to a frequency of 40 MHz, resulting in about 40 million BX per second. For every BX about twenty collisions between 200 billion particles take place in the LHC. The luminosity constitutes the coefficient of proportionality between the number of events R produced per second, called the *event rate*, and the production cross section of the physical process in question (σ_{proc}):

$$R = \mathcal{L}\,\sigma_{proc}. \tag{2.4}$$

The total amount of collisions produced in a given time interval is called *integrated luminosity* and it is defined as $L = \int \mathcal{L}\,dt$.

2.1.2 Proton-proton physics at the LHC

To understand the requirements of the LHC and the solutions adopted in the development of the machine and of the experiments, a brief discussion about the physics expected to be studied in p-p collisions at a circular collider is needed.

The proton structure and its dynamics are determined by the strong interaction. They are described by the quantum chromodynamics (the QCD theory), which depicts the proton as a composite system made of elementary particles. The proton structure is a manifestation of the strong confining force and accounts for three valence quarks and a sea of quarks, antiquarks, and gluons, called *partons*. The proton internal structure can be modelled as a parton beam in which each parton carries a fraction x of the total momentum, as shown in Fig. 2.2:

$$p_{parton} = x\,p_{proton},$$

where the spatial distribution of partons as a function of x is called *parton distribution function*.

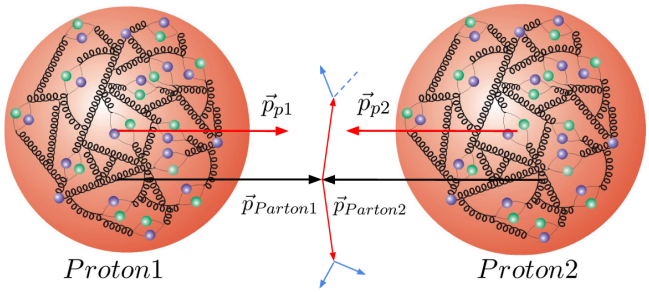


FIGURE 2.2: The elementary structure of the proton.

Defining the proton's momentum fractions carried by the two interacting partons, x_a and x_b, the effective energy available in a collision is related to the nominal center-of-mass energy by the expression:

$$\sqrt{\tilde{s}} = x_a x_b \sqrt{s}.$$

Since the interactions occur between partons, there are a lot of particles that accompany a hard scattering in a single p-p interaction, referred to as *underlying events* (UE). The modelling of the UE is important and it represents a significant ingredient for the correct identification of isolated physics objects and the estimare of their energy. The main uncertainties associated to the UE at the LHC come from the parton density functions, the initial and final state radiation from the gluons, and the modelling of the multiple partonic interactions.
Moreover, there is a nonzero probability of multiple soft interactions in a single collision, called *minimum bias* (MB). From the experimental point of view, it is dominated by low p_{T} QCD processes with low transverse energy and low multiplicity and it corresponds to a nonsingle diffractive inelastic interaction. The majority of the particles produced in p-p collisions arises from soft interactions, which require nonperturbative phenomenological models and represent a major background at higher luminosities.

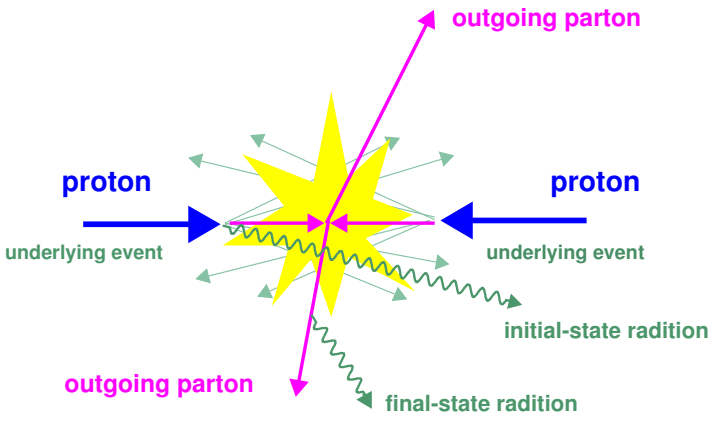


FIGURE 2.3: Illustration of a typical hard scattering event between two partons including initial and final state radiation and beam-beam remnants.

Finally, given that collisions take place every 25 ns, all particles produced in the first BX have no time to reach the external surface of the detector before the new particles leave the IP. This effect is referred to as *out-of-time pileup*. Moreover, multiple proton interactions can take place within each BX producing events which overlap the hard scatter interaction, hereafter denoted as *in-time pileup* or only *pileup*.

One of the main technical challenges of the experiments in similar data-taking conditions is to maintain the required detector efficiency and accuracy in the presence of extremely high background levels. For this reason, high granularity, fast response, and precision in the measurement of the energy and momentum of the final state particles represent crucial features of the LHC experiments, together with high resistence to the radiation. Then, other challenging aspects of the high energy physics at p-p colliders have to be considered, for example how to handle the enormous volume of data collected and to expeditiously analyze them in order to extract physics results.

2.1.3 Operation of the LHC

The LHC started operating in 2008, when the first beam was circulated through the collider (10^{th} September). However, on the 19^{th} September 2008, a magnet quenching occurred in about one-hundred bending magnets in sectors 3 and 4. An electrical fault led to a loss of approximately six tonnes of liquid helium, which was vented into the tunnel. Because of the expansion of the escaping vapour, fifty-three superconducting magnets and their mountings were damaged and the vacuum pipe contaminated loosing vacuum conditions.

Run 1 The first operational run started on 23^{rd} November 2009 at a center-of-mass energy of $\sqrt{s}$ = 900 GeV. Then, during the early part of 2010, the energy of the beams were ramped up to 3.5 TeV per beam. The LHC set a new record for high energy collisions by colliding proton beams at a center-of-energy of 7 TeV on 30^{th} March 2010. The 2012 run started at a beam energy of 4 TeV, corresponding to a collision energy of 8 TeV. The luminosity was also increased to target 15 fb^{-1} of data, three times larger than the total integrated luminosity collected in 2010-2011 at 7 TeV. During this data-taking period, collisions were characterized by an average number of overlapping interactions per bunch crossing of twenty-one (nine) at 8 TeV (7 TeV). Both the increased energy and the increased luminosity were expected to significantly extend the opportunities to explore new physics. The LHC remained in operation until February 2013, running continuously for three years and delivering a luminosity of around 30 fb^{-1}. Data corresponding to an integrated luminosity of about 20 fb^{-1} was collected by the CMS experiment and the discovery of the H boson by the ATLAS and CMS Collaborations dates back to this period. After that, a long shutdown (LS) period occured in order to increase the luminosity and the center-of–mass energy of the LHC up to the original design conditions and to perform ordinary and essential maintenance.

Run 2 After a two year LS, the second operational run started in spring 2015 and lasted until 2018. The LHC reached for the first time a p-p collision energy of $\sqrt{s}$ = 13 TeV in 2016, accelerating each beam at the unprecedented energy of 6.5 TeV. A summary of the beam and machine parameters during Run 2 compared to the LHC design values is reported in Table 2.1.
The total number of collisions in 2016 exceeded the number from the whole Run 1 at a higher energy per collision. The p-p run was followed by four weeks of proton-lead collisions. During 2017, the luminosity was further increased and reached twice the design value. Finally, the LHC operation in 2018 began on the 17^{th} April and stopped on the 3^{rd} December, including four weeks of Pb-Pb collisions. During the last steps of the intensity ramp-up, the average peak luminosity reached by the ATLAS and CMS experiments surpassed the record peak luminosity of 2017 being close to 2.1×10^{34} $\text{cm}^{-2}\text{s}^{-1}$, as shown in Fig. 2.4.
An overview of the integrated luminosity of data collected by the CMS experiment during Run 1 and Run 2 is shown in Fig. 2.5. During Run 2, collisions were characterized by a center-of-mass energy $\sqrt{s}$ = 13 TeV and an average of thirty-four pileup interactions.

The successful Run 2 data-taking period allowed to perform important precision measurements for constraining the SM as a stable theory. For example, the masses of the H and W boson are currently known with great precision, some new H boson couplings were observed (see Section 1.2.3), and the investigation on the CP-violation was provided with improved measurements of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. However, since direct searches for new physics have not revealed anything new, the road to precision measurements looking for BSM effects is becoming much more important.

Parameter	Design	2018	2017	2016	2015
Energy (TeV)	7.0	6.5	6.5	6.5	6.5
Number of bunches	2808	2556	2556-1868	2220	2244
Number of bunches per train	288	144	144-128	96	144
Max. stored energy (MJ)	362	312	315	280	280
β^* (cm)	55	30→27→25	40→30	40	80
Bunch population, $N_b(10^{11})$p	1.15	1.1	1.25	1.25	1.2
Normalized emittance (μm)	3.75	1.8-2.2	1.8-2.2	1.8-2	2.6-3.5
Peak luminosity (10^{34}cm^{-2}s^{-1})	1.0	2.1	2.0	1.5	<0.6
Half crossing angle (μrad)	142.5	160→130	150→120	185→140	185

TABLE 2.1: Comparison between the beam and machine parameters during Run 2 and the LHC design values.

2.2 Overview of the CMS experiment

The CMS Collaboration brings together members of the particle physics community from around two-hundred institutes and universities and from more than forty countries across the globe, counting over four thousand particle physicists, engineers, computer scientists, technicians, and students. Since the beginning, collaborators from all over the world helped to design and fabricate components of the CMS detector, which were brought to CERN for the final assembly. Since then, the collaboration operates to collect data from the experiment, taking care of the mainteinance and upgrade of the detector. Finally, data collected by the CMS detector are shared with several computing centres via the World Wide LHC Computing Grid to be analyzed by the CMS groups from all around the world.

As it was explicitely stated in the letter of intent presented by the CMS Collaboration in 1992 [58], CMS was designed as a general purpose detector able to run at the highest luminosity at the LHC and it was optimized for the SM H boson search over a mass range from 90 GeV to 1 TeV. However, its versatility allows to cover many possible signatures from alternative electroweak symmetry breaking mechanisms and a large range of phenomena at the LHC energies. The aim is to scrutinize different unproven models of the elementary structure of the matter. It is also well adapted for the study of the top, beauty, and tau physics at lower luminosities and it is able to cope with several aspects of the heavy ion physics program. It was conceived to study p-p (and Pb-Pb) collisions at a center-of-mass energy of 14 TeV (5.5 TeV) and at luminosities up to 10^{34}cm^{-2}s^{-1} (10^{27}cm^{-2}s^{-1}).
According to the experiment's goals, it was chosen to identify and measure muons, photons,

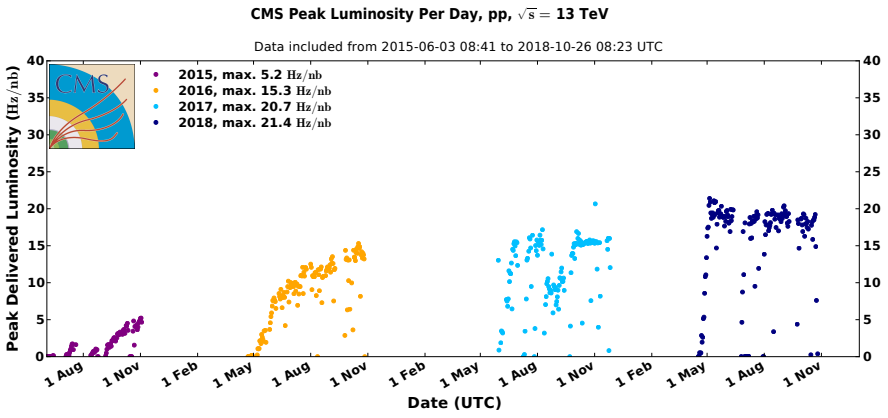


FIGURE 2.4: Peak luminosity versus time for 2010-2012 and 2015-2018 delivered to CMS during stable beams for p-p collisions at 13 TeV [57].

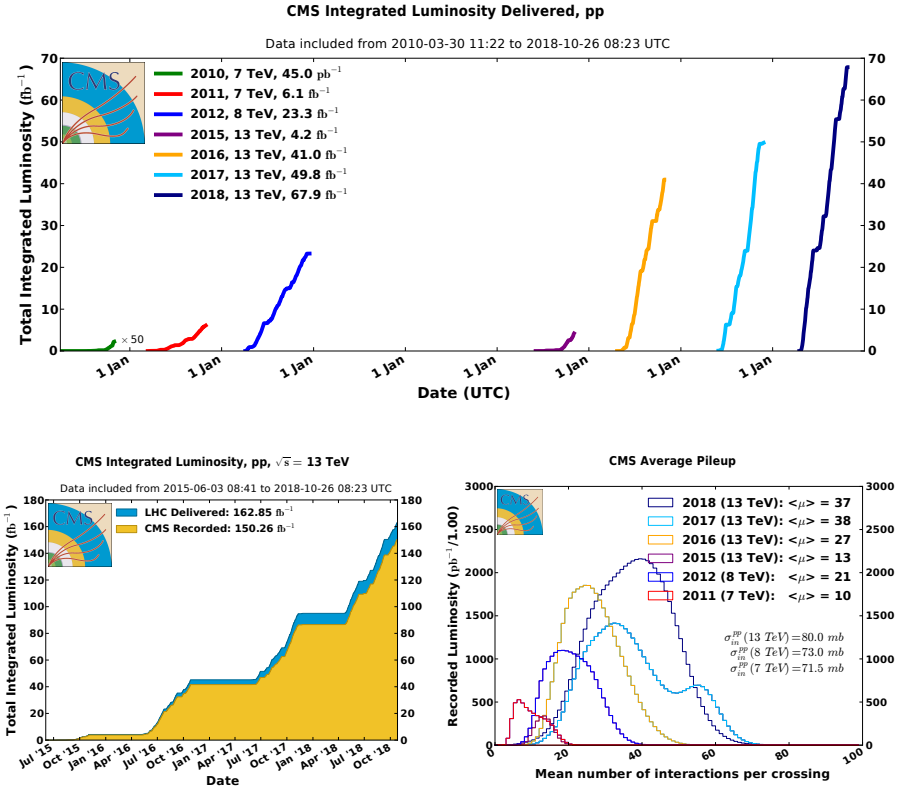


FIGURE 2.5: Integrated luminosity of data collected by the CMS experiment per year during Run 1 and Run 2 (top). The delivered and recorded luminosity cumulative over all years during stable beams for p-p collisions at a center-of-mass energy of 13 TeV (bottom left) and the distribution of the average pileup with the minimum bias cross sections (bottom right) are also shown [57].

and electrons with high precision as the main design considerations. Therefore, the design focused on a highly performant muon system, a high quality electromagnetic calorimeter, a robust and powerful central tracking system to achieve excellent charged particle reconstruction and detailed vertex reconstruction, a hadron calorimeter characterized by good hermiticity and sufficient energy resolution, and a strong magnetic field that allows great muon momentum resolution. Thanks to these features, the CMS experiment provided the most precise H boson mass measurement up to now!

2.2.1 The coordinate frame

To describe the detector and the collision events, the origin of the coordinate system [59] is centered at the nominal collision point inside the experiment. It is situated on the longitudinal z-axis along the beam direction toward the Jura mountains from LHC Point 5 in Cessy (France). The x-y plane of the right-handed Cartesian coordinate frame, shown in Fig. 2.6, identifies the transverse plane. The x-axis points radially inward toward the center of the LHC ring and the y-axis points vertically upward. Considering the cylindrical structure of the CMS detector, a polar system is often convenient to describe the four momentum of particles using their energy and two angles. The azimuthal angle ϕ is formed with respect to the positive x axis in the transverse plane, while the polar angle θ is measured from the z-axis in the (r-z) plane, where r denotes the radial coordinate.

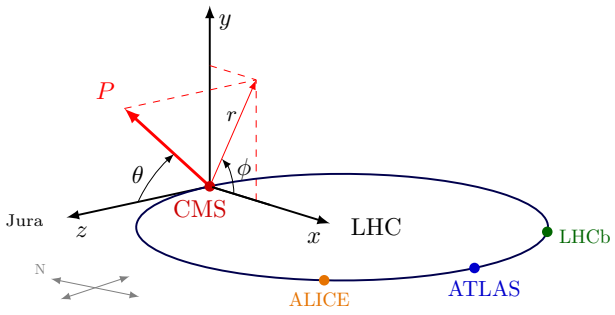


FIGURE 2.6: Coordinate system of the CMS experiment.

Instead of using the polar angle θ, it is convenient to construct a quantity with better transformation properties under Lorentz boosts along the beam axis, called *rapidity*, defined as:

$$y = \frac{1}{2} \log \frac{E + p_z}{E - p_z}. \tag{2.5}$$

Rapidity differences remain invariant under boosts along the z-axis. It goes to zero for particles with only a transverse component of the momentum, while it goes to $\pm\infty$ for particles parallel to the z-axis. However, this quantity does not provide an intuitive interpretation, so an analogue quantity is introduced, the *pseudorapidity*:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.6}$$

Small values of the θ angle correspond to large values of η. If $m \ll E$, these two quantities are more or less equal.
Interesting events concentrates in the region at large η values of the transverse plane, far from the beam pipe. Thus, the energy and momentum measured in the transverse plane with respect to the beam direction from the x and y components are useful quantities to describe events generated during a collision and are denoted by E_T and p_T. The transverse energy could be defined as:

$$E_T = E \sin\theta, \tag{2.7}$$

while the transverse momentum is the projection of the momentum of a particle onto the transverse plane. It is Lorentz invariant under boosts along the z-axis, so it is possible to assume that it is conserved in the transverse plane after the parton interaction:

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{2.8}$$

The imbalance of energy measured in the x-y is referred to as $p_T^{miss} = -\sum_i \vec{p_T}$:

$$|p_T^{miss}| = \sqrt{\left(\sum_i p_x^i\right)^2 + \left(\sum_i p_y^i\right)^2}. \tag{2.9}$$

Finally, to describe particles in terms of η, ϕ, and z, another quantity is defined, ΔR, which provides a distance measurement in the η-ϕ plane:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}, \tag{2.10}$$

where $\Delta\eta$ and $\Delta\phi$ represent the spatial separation of two objects in terms of pseudorapidity and azimuthal angle, respectively.

2.2.2 Structure of the detector

The CMS apparatus is located in the underground cavern of LHC Point 5 and is instrumented with different subdetectors. It features a cylindrical structure with a diameter of 14.6 m and a length of 21.6 m for an overall weight of 12500 tons. A large, high field superconducting solenoid sits at the core of the CMS system. The core of the magnet coil can accommodate the silicon-based inner tracking system and the calorimetry. Each subdetector is composed of a central section, called *barrel*, and two forward regions, the *endcaps*. The pseudorapidity coverage provided by the barrel and endcap detectors is extended by the forward calorimetry up to $|\eta| < 5$ assuring full geometric coverage. The steel flux-return yoke outside the solenoid is integrated with four stations of gas-ionization detectors to measure muons. It covers most of the 4π solid angle and ensures very good hermeticity and robustness. A quick description of each part of the CMS detector will be provided in the following. A perspective view of the CMS overall layout is shown in Fig. 2.7. A full description of the CMS detector can be found in Ref. [59].

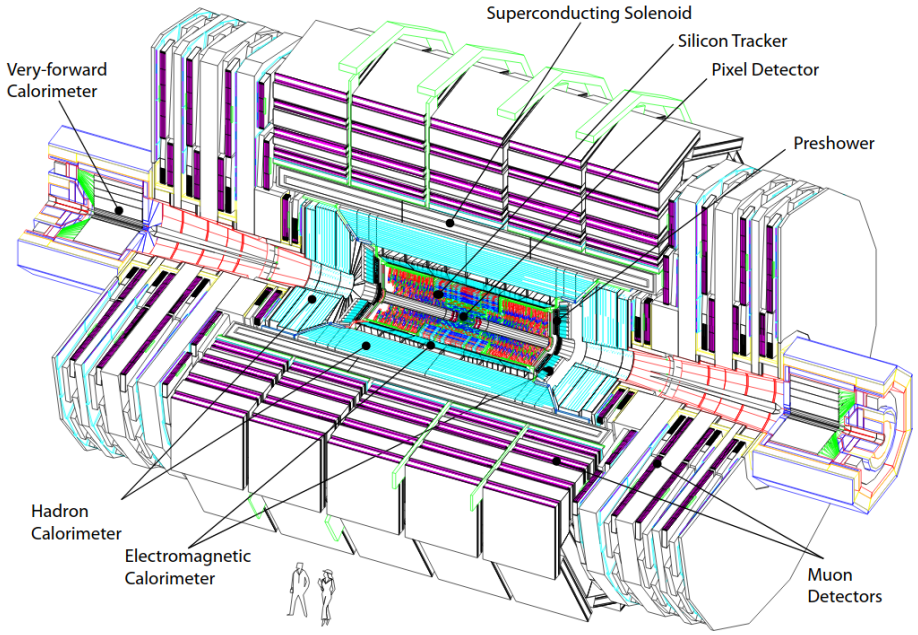


FIGURE 2.7: Perspective view of the CMS detector [59].

Magnet

A strong magnetic field is mandatory for a compact detector based on a single long solenoid, used to bend the paths of charged particles from collisions in the LHC and measure their momentum ($\vec{p}$) with a good resolution using the tracking subdetectors. The 6 m diameter high superconducting magnet [60] at the core of CMS provides a 3.8 T magnetic field and operates at a temperature of 4.5 K. It is made of NbTi cables wrapped with copper and enclosed in a 12000 tons iron yoke which is part of the outer muon detecting system. Consequently, the p_x, p_y, and p_z of the final state particles can also be measured with the muon detectors placed inside in the iron structure which surrounds the solenoid. A difficult task in hadron colliders is to obtain an efficient triggering on muons.

The design of the CMS solenoid mainly took into account the issue of stability. For this purpose, a high purity aluminium stabilized conductor, able to sustain higher rate current and magnetic forces, and a monolithic four-layer winding structure, capable of sustaining

all the induced magnetic forces by itself, was chosen. An indirect thermosiphon cooling keeps the magnet at a temperature of about 4 K and maintains the superconducting mode. Some of its main parameters are given in Table 2.2.
Compared to a toroid, the main advantage is related to the smaller overall size. A strong bending power can be obtained for a solenoid with modest dimensions, ensuring an efficient trigger, which depends on the vertex determination. Moreover with the field parallel to the beams, the bending of the track starts at the primary vertex ($R = 0$) and then continues in the transverse plane. The transverse position of the vertex can be determined to an accuracy better than 20 μm thanks to the small dimensions of the beam.

Nominal field	4T
Inner bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA

TABLE 2.2: Main features of the CMS superconducting solenoid.

The iron yoke plays the role of an absorber and is used to return the flux with an intensity of about 2 T. It is divided into two main components.

* *The barrel yoke* is the cylinder surrounding the superconducting coil and is divided into five 2.9 m long sections along the beam axis. Each of them is made up of three iron layers. The central barrel ring is fixed around the IP and supports the superconducting coil, while the other four sections can slide on rails along the beam direction. Most of the muon chambers are inserted in the iron yoke.

* *The endcap yoke* is composed of three independent rings which magnetically close the barrel yoke. The high magnetic field has a direct impact in terms of forces on the two endcap yokes, due to the axial magnetic attraction. To reinforce their structure, the first and second disks supporting the electromagnetic and hadronic calorimeter endcaps are composed of two 300 mm thick iron plates.

The vacuum tank is made of stainless steel and accomodates the superconducting coil. The outer shell of the vacuum tank is attached to the inner part of the central barrel ring, while the inner shell supports all the barrel subdetectors via a system of horizontal welded rails.

Tracker

The inner tracking system of CMS [61] is the innermost detector part of the experiment and is based on silicon pixel and fine-grained microstrip detectors. It aims to provide good particle momentum resolution and efficient reconstruction of the primary and secondary vertices. Exploiting the uniform magnetic field within the tracking detector volume, the combination of the information on the position of charged particles within each silicon detector, called *hits*, provides the measurement of their momentum and charge. To address the problem of pattern recognition at high luminosities, tracking detectors with small cell sizes are required. The CMS tracker is thin and finely segmented and is equipped with fast readout on-board electronics. Moreover, it is important to reconstruct high p_{T} isolated tracks produced in the central rapidity region with a transverse momentum resolution better than $\delta p_{\mathrm{T}}/p_{\mathrm{T}} = (15 \cdot p_{\mathrm{T}} \oplus 0.5)\%$. In fact, to identify isolated leptons it is fundamental to reconstruct all tracks with p_{T} greater than 2 GeV in the barrel region and then check that all interesting particles in an event are consistent with a common vertex.

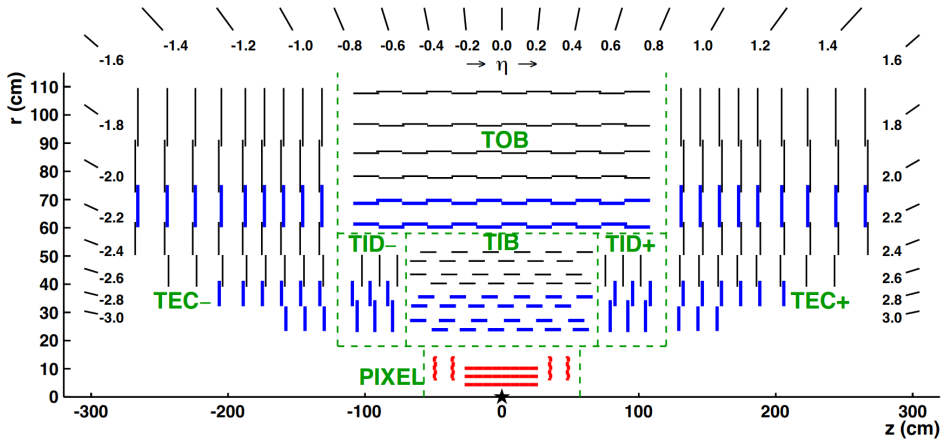


FIGURE 2.8: Schematic view of the CMS tracking system in the r-z plane. Each tracker subsystem is indicated with green dashed lines. Strip modules providing 2D hits (thin black lines) and permitting the 3D reconstruction of hit positions (thick blue lines) are highlighted.

The tracking volume is given by a cylinder 5.4 m long with a diameter of 2.4 m and is entirely made of silicon detectors. It covers the pseudorapidity region up to $|\eta| < 2.5$ with a radius smaller than 1.2 m and $|z| < 2.7$ m and represents the largest silicon tracker ever built. The total number of sensor elements amounts to 9.3 million strips and 66 millions pixels. They are arranged in concentric cylindrical volumes on a rigid carbon fiber structure in the center of the experiment. It features a central barrel region and two endcap disks of silicon pixel detectors which close the interaction region at higher values of pseudorapidity. To solve stringent resolution and granularity requirements in the high, medium, and lower particle density regions at decreaing distance from the IP, two different detector technologies have been adopted. They are both characterized by a fast response and resistence in harsh radiation environment. A schematic view of the CMS tracking system is shown in Fig. 2.8.

* *Silicon pixel detectors* are placed in the inner region below 20 cm, close to the IP, to allow the measurement of the impact parameter of charged particle tracks, as well as the position of secondary vertices.
  The CMS pixel system was originally constituted by three barrel layers placed between 4.4 cm and 10.2 cm and two endcap disks made of 24 blades in a turbine-like shape for each side. Modular detector units constitute the layers, providing very low occupancy, high resolution, and precise vertex reconstruction. These modules consist of a thin, segmented sensor plate with highly integrated readout chips. An upgrade of the silicon pixel detector was performed in 2017 [62]. It now features four layers in the barrel and three disks in the endcaps to provide an additional point in both regions. The innermost disk has been installed closer to the nominal IP at a distance of 3 cm for the barrel detector. The material budget has also been largely reduced by 40% and 80% of the weight of the previous detectors in the barrel and endcap detectors, respectively. In the barrel, hit resolutions at the 10 μm level can be obtained in the r-ϕ direction, while a spatial resolution of 15 μm is achieved for inclined tracks along the z-direction. Lower resolutions of 15 μm and 20 μm, respectively, characterize the endcap regions. The pixel detector upgrade explains some of the improvements in the $\mathrm{H} \to 4\ell$ object selection observed in 2017 and 2018 data with respect to 2016, as discussed in the next chapters.

* *Silicon microstrip detectors* are organized in ten layers about 10 cm long in the outer region between 20 cm and 120 cm. On the one hand, there are four inner barrel layers (TIB) and two inner endcaps (TID), each composed of three small rings. On the other hand, six concentric layers form the outer barrel (TOB) and two endcaps (TEC) close off the tracker.
  This part of the detector operates at a temperature of -15°C to reduce the damage caused by ionizing radiation and is composed of 15200 highly sensitive modules consisting of three elements: a set of sensors characterized by fast response and good spatial resolution, a mechanical support structure, and readout electronics. More than 24000 silicon sensors and 10 million detector strips read by 80000 microelectronic chips result in a silicon area of about 200 m^2 sensitive to the passage of charged particles from p-p collisions. Two different strip pitches are used according to the distance from the IP. Silicon microstrips allow to reduce the number of readout channels maintaining a good resolution and provide the required high granularity and precision keeping the cell occupancies below 1%. A spatial resolution of 40-60 μm is provided in the r-ϕ transverse plane and of 500 μm along the z-axis.

Electromagnetic calorimeter

The search for the H boson and for a large variety of SM and other new physics processes at the LHC strongly relies on the information from the calorimeters. The measurement provided by the tracking system and the calorimetry in CMS is complementary. On the one hand. , the tracker can identify only charged particles with an increasing precision of the momentum measurement as the p_{T} decreases, due to the larger curvature of the trajectory. On the other hand, the calorimeters can measure both charged and neutral particles with a resolution which improves in corrispondence of the increase of the particle energy.

A fully active scintillating crystal electromagnetic calorimeter (ECAL) [63] is used for the identification of photons and electrons and for the precise measurement of the energy of incident final state particles. Most of the energy from electrons or photons is deposited within the homogeneous crystal volume of the calorimeter, constituted of 75848 lead tungstate ($PbWO_4$) crystals. The energy measurement is based on the conversion of the incident particles to an electromagnetic shower, that interacts with the ECAL material producing scintillation light. About 80% of the light is emitted in 25 ns, which is an ideal feature considering the proton bunch spacing and the high instantaneous luminosity of the LHC collisions. Crystals provide important physical and optical characteristics: compactness, high granularity, excellent time resolution, and a fast scintillation response. The mechanical structure within the active volume of the calorimeter is minimized: high density crystals ($\rho = 8.28$ g/cm^3) with a short Molière radius (2.2 cm) and small radiation length ($X_0 = 0.89$ cm) result in a very compact electromagnetic calorimeter system with a length of approximately 25 X_0. It is able to contain the electromagnetic showers in an optimal way, ensuring an excellent energy resolution.

Two regions can be distinguished: the cylindrical barrel (EB) up to $|\eta| < 1.479$ and two endcap disks (EE) which provide a pseudorapidity coverage up to $|\eta|$ = 3.0. Moreover, a preshower system is installed for the discrimination of single photons from $\pi^0 \rightarrow \gamma\gamma$ decays just in front of the barrel (SB) and the endcap in the region $1.65 < |\eta| < 2.6$ (SE).
A schematic view of the ECAL is shown in Fig.2.9.
The barrel region is further divided into two parts, each consisting of eighteen supermodules. Every ECAL supermodule is made of 1700 crystals arranged in an η-ϕ geometry which accounts for twenty crystals along the ϕ direction and eighty-five crystals in η, covering 20° in the ϕ direction. Supermodules are further split into four modules, where the first one contains twenty crystals in ϕ and twenty-five in η, while all other modules contain twenty

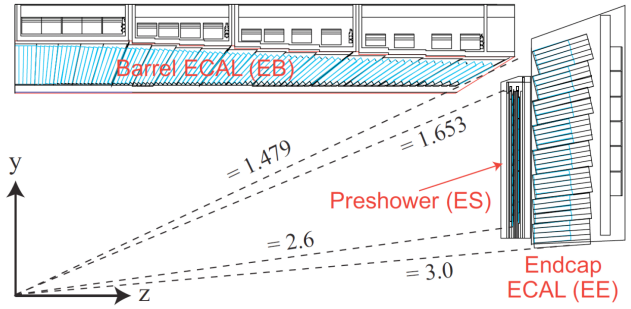


FIGURE 2.9: Schematic view of one quarter of the CMS ECAL during the LHC Run 2 operation.

crystals along both ϕ and η coordinate. Crystals are placed with a small 3° angle with respect to the direction from the nominal IP, in order to avoid gaps aligned with particle trajectories. However, some gaps which are referred to as *cracks* remain between modules making the energy reconstruction more complicated, especially at η = 0 and at the barrel-endcap transition region. The endcap structure is more complex and it is not planar in η-ϕ. It consists of four dees, each containing 3662 crystal complexes in x-y geometry, grouped into 5×5 supercrystals. The sampling preshower is composed of two layers of lead absorber in which the electromagnetic shower is initiated with a subsequent layer of 2 mm wide silicon strips to measure the deposited energy and the transverse profile of the shower shape.

The scintillation light produced through bremsstrahlung and pair production is collected and amplified by silicon avalanche photodiodes (APDs) in the barrel region, since they can provide gain in the presence of the high transverse magnetic field. In the endcap region, vacuum phototriodes (VPTs) are used to compensate the relative low light yield of $PbWO_4$, corresponding to about thirty photons per MeV of deposited energy. Here the photodetectors suffer from a much higher integrated radiation dose (50 kGy) and neutron flux (7×10^{14} n/cm^2). The large doses of radiation tend to modify the transparency of the crystals. For this reason, cycles of transparency reduction and recovery correspond to the collisions and refill operations of the LHC. This effect can be corrected by injecting 440 nm laser light in each crystal to derive time dependent correction factors which are then applied to the response. The energy resolution of the ECAL in the mass range [25, 500] GeV can be parametrized as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{\sigma_n}{E}\right)^2 + (c)^2,$$

where

* E is the particle energy in GeV;
* a is a stochastic term which considers a photostatistics contribution and fluctuations on the lateral containment of the shower or on deposits of energy;
* σ_n is the noise, which contains three different contributions: preamplifier noise, digitization noise, and pileup noise;
* c is a constant term reflecting energy leakage of the crystals, nonuniformity of the longitudinal light collection, and calibration errors.

Typical energy resolutions are measured with electrons [64]. The excellent performance can be parametrized by a stochastic term of 2.8%, a noise term of 12%, and a constant term of 0.3%:

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\%.$$

The calibration of the ECAL response determines both the absolute energy scale and the channel-to-channel intercalibration, compensating in particular for the intrinsic crystal light yield variations (around 15%) and the spread in the EE phototriodes (around 25%). In situ measurements based on collision events are used to complement the original calibration derived from laboratory studies and cosmic rays exposures of crystals.

Hadronic calorimeter

The ECAL is surrounded by a brass scintillator sampling hadronic calorimeter (HCAL) [65], which allows for the identification of hadrons. It is designed to absorb hadrons that escape from the ECAL volume and measure the energy of hadron showers induced in the HCAL. It is essential for the reconstruction of *jets*, which are narrow cones of hadrons and other particles produced by the hadronization of a quark or gluon.
The intrinsic resolution of the hadron energy is limited by various effects. Noninteractive particles such as neutrinos can be produced in nuclear and hadronic interactions and non-Poissonian effects can occur. In addition, $\pi^0 \to \gamma\gamma$ decays can produce an additional electromagnetic component in ECAL. For this reason, the usage of the particle flow (PF) reconstruction techniques (see Section 2.2.4) is crucial to improve the hadronic resolution offline. Moreover, the HCAL hermeticity and geometrical coverage allows to compute the imbalance in the transverse momentum sum of the event, which is an essential element at the level of the PF reconstruction.

The HCAL is divided into barrel (HB) and endcap (HE) regions and accounts for two forward calorimeters (HF) around the beam pipe located at $|z| = 11.2$ m very close to the beam axis. Thanks to the HF, a coverage up to $|\eta| < 5.2$ is obtained. A schematic view of the HCAL components is shown in Fig. 2.10. In addition, to obtain a better energy resolution of the barrel calorimeters, an outer calorimeter (HO) is placed outside the magnet coil extending the total interaction length (λ) to about 11 λ. The HCAL consists of a barrel part up to $|\eta| < 1.3$ and an endcap part in the region $1.3 < |\eta| < 3.0$, containing more than 2500 channels respectively. The barrel part is installed between the ECAL and the coil from R = 1.9 m to R= 2.95 m, corresponding to a total of 7 λ at η = 0 (10λ_0 at $\eta \geq 1$). Brass absorber plates are alternated to the active medium consisting of plastic scintillator tiles, coupled to an hybrid photodiodes (HPDs) readout. A segmentation of 0.087×0.087 and 0.17×0.17 in the transverse dimension $\Delta\eta \times \Delta\phi$ is used in HB and HE, respectively. It is necessary for the separation of nearby jets, the determination of their direction, and an adequate mass resolution.

The hermeticity of the detector is guaranteed by the HF, which provides full geometric coverage for the measurement of the transverse energy in the event in a very high rate environment. They are made of quartz fibers as active element, predominantly sensitive to Cherenkov light from neutral pions, embedded in a steel absorber. This design allows for a very localized response to hadronic showers. Finally, the Cherenkov light emitted in the quartz fibers is detected by photomultiplier tubes (PMTs). Fibers of two different lengths are installed to estimate the electromagnetic and hadronic components of the shower. The imperfect containment of the hadronic shower dominates the overall HCAL performance, resulting in a resolution sampling term of about 110% and a constant term of 9% according to the measurements performed in pion test beams [66]:

$$\frac{\sigma}{E} \simeq \frac{110\%}{\sqrt{E}} \oplus 9\%.$$

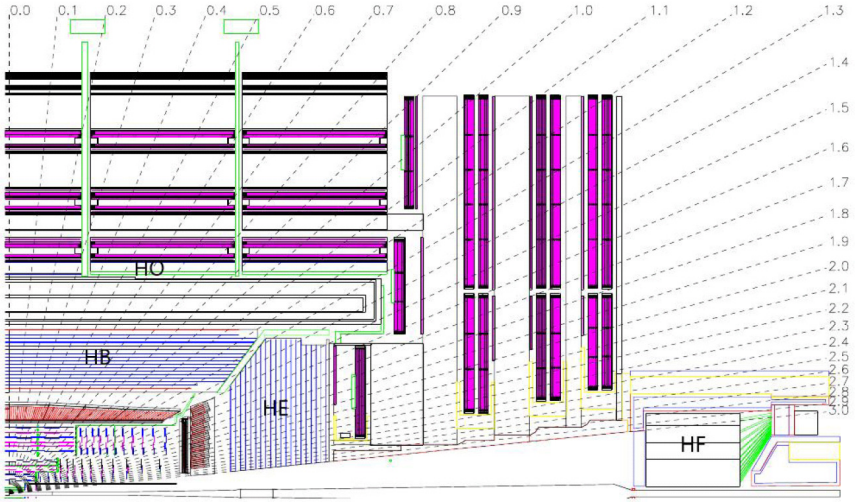


FIGURE 2.10: Schematic view of one quarter of the CMS HCAL during LHC Run 2 operation [59].

Muon system

According to the goals and the design of the experiment, an efficient detection of muons is one of the most relevant tasks of the CMS detector. Many signal signatures with muons in the final state rely on the muon detection system for the discrimination from the very high background rate expected at the LHC. Muons are often indicators of interesting physics because they are produced via electroweak interaction, where a W, Z, or H boson is involved, or eventually some new, unknown particles.
The CMS muon spectrometer [67] is designed to ensure the capability of reconstructing the momentum and charge of muons over the the entire kinematic range of the LHC. The robustness of the CMS muon identification system is provided by two independent measurements of the muon momentum in two different regions of the detector: inside the inner tracking volume and after the coil in the muon system. Since muons produced in collisions at the LHC have little interaction with matter and minimal energy loss rates caused by ionization in the detector, they are the only charged particles able to cross all the inner detector layers and reach the muon system. Consequently, the outer part of the detector is particularly relevant because it allows an efficient muon identification and momentum resolution at high p_{T}, an optimal charge determination, and a robust standalone muon trigger, while maximizing the ϕ acceptance for muons.

The CMS is characterized by a multilayer muon spectrometer, in which the muon detectors are organized in four stations interleaved with magnet return yokes to ensure robustness. Three different technologies of gaseous detectors are exploited, according to the expected background rates and uniformity of the magnetic field: Drift Tubes (DTs) in the central region $|\eta| < 1.2$ and Cathode Strip Chambers (CSCs) in the endcap region $0.9 < |\eta| < 2.4$, complemented by Resistive Plate Chambers (RPCs) up to $|\eta| < 1.6$ to ensure redundancy and improved trigger capabilities. In general, they exploit the ionization of electrons produced by the passage of charged particles in the gas. The main features of each kind of detector are summarized below. The overall organization of the muon chambers in the Run 2 configuration is illustrated in Fig. 2.11.

- *Drift Tubes* are tracking detectors which cover the muon spectrometer barrel providing a precise measurement in the bending plane and presenting good trigger capabilities. They can trigger on the muon p_{T} with good efficiency and high background rejection without relying on the rest of the detector. The Level 1 trigger p_{T} resolution

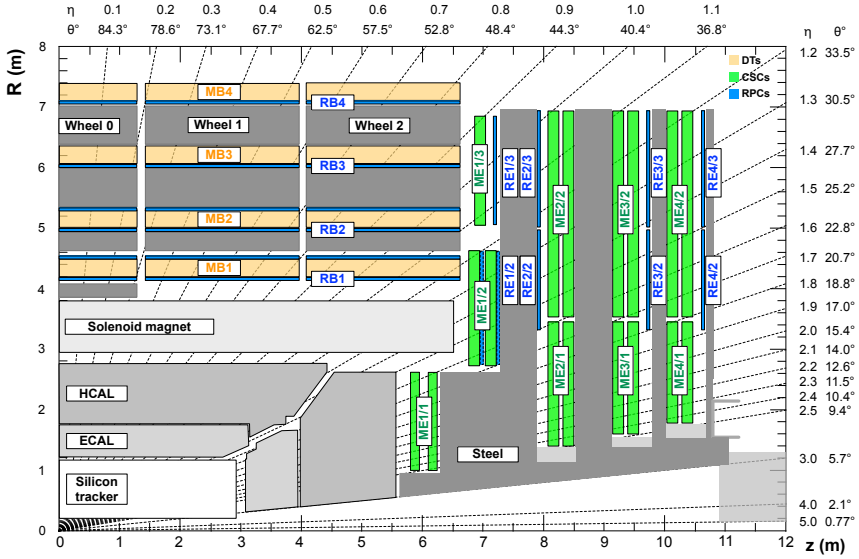


FIGURE 2.11: Quadrant of the CMS muon detector in its Run 2 configuration. DTs, CSCs, and RPCs subsystems are indicated in orange, green, and blue, respectively. In the Muon Barrel (MB) there are four stations of DT and RPC, divided into five wheels along the z axis, and in the Muon Endcap (ME) CSC and RPC are organized in four disks transverse to the beam axis [68].

is about 15% in the barrel and 25% in the endcap (see Section 2.2.3). The full DT system consists of 250 rectangular chambers (2×2.5 m^2) divided into five wheels along the z direction and arranged in groups of four stations. Given that most of the return flux of the magnetic field is contained in the iron yoke, chambers with standard rectangular drift cells perform adequately. Each wheel is divided into twelve contiguous sectors in ϕ and each sector covers around 30°. Each chamber is composed of three superlayers (SLs) and each SL consists of four layers of drift cells. A cell is a rectangular drift tube (4.2×1.3 cm^2) composed of an anode wire and two cathode strips, filled with a mixture of 85%-15% Ar-CO_2. The position and angle of traversing muons are measured from the time needed by the electrons to drift toward the anode wires when ionizing the gas. In each group of SLs, two of them have wires parallel to beam axis, which provide a measurement of the perpendicular coordinate in the r-ϕ plane, and the other one has wires normal to the beam axis and measures the coordinate along the direction parallel to the beam in the r-z plane. Thanks to a staggered layout, where the layers are not aligned with each others but they are placed with an offset by half a cell width with respect to the previous one, a time resolution lower than 3 ns and an efficiency of 99.8% are achieved. Muon positions are measured providing two spatial coordinates with a precision of 100 μm in the r-ϕ plane and better than 1 mrad along the θ-direction.

∗ *Cathode Strip Chambers* are employed in the endcap region to cope with high particle rates and a nonuniform magnetic field. In each endcap, trapezoidal detector chambers are organized into four stations supported by three vertical steel disks for a total of 468 chambers. CSCs are Multiwire Proportional Chambers (MWPCs) characterized by a short drift length, that work in avalanche mode with a finely segmented cathode strip readout. The cathode strips of each chamber run radially outward and provide a precision measurement in the r-ϕ bending plane, while the anode wires provide a radial measurement. They consist of arrays of positively charged anode wires crossed perpendicularly with negatively charged copper cathode strips within a gas volume.

When a muon crosses the cell, it ionizes the gas producing some charge. The charge then drifts toward the wire and gets multiplied, inducing a charge pulse in the strips. A mixture of Ar (40%), CO_2 (50%), and CF_4 (10%) is used. In this way, two position coordinates for each passing particle are obtained, ensuring a good spatial resolution (50-50 μm) and a precise time information of about 3 ns.

* *Resistive Plates Chambers* are fast gaseous detectors coupled to DTs and CSCs in both the barrel and endcap regions to provide a complementary muon trigger system, helping to resolve ambiguities when making tracks from multiple hits in a chamber. The barrel region accounts for six layers of RPCs, two in each of the first two stations and one in each of the last two stations. The redundancy in the first two stations allows the trigger algorithm to work even for low p_{T} tracks which may not to reach the outer two stations. In the endcap region, a plane of RPCs is placed in each of the first three stations. The coincidences between stations can be used to reduce the backgrounds, to improve the time resolution for BX identification, and to achieve a good p_{T} resolution. They consist of double gap chambers made of a high resistivity plastic material and filled with a mixture of $C_2H_2F_4$ (96.2%), i-C_4H_{10} (3.5%), and SF_6 (0.3%). Four planes, two positively charged anodes and two negatively charged cathodes, are alternated to form two thick gas gaps of 2 mm used in avalanche mode to cope with high rates of particles. The plates and the electrodes are made of bakelite and graphite, respectively, while the central region features rectangular aluminium strips which collect the signal produced when the particle passes. The RPCs combine a good time resolution of about 2 ns with a spatial resolution of about 1 cm, depending on the strip width, cluster size, and detector alignment. Moreover, they provide timing information for the muon trigger with a fast response and present an average detector efficiency of 95%. The matching between track segments in DTs and CSCs and any RPC cluster provides a measurement both for the efficiency and for the spatial resolution.

2.2.3 Trigger system

Among the huge number of p-p collisions produced each second during the LHC operation, corresponding to a collision frequency of 40 MHz, events of potential physics interest are selected using a two tiered trigger system [69] which leads to a drastic rate reduction and rejection of most of the low energy processes. The start of the physics event selection process is represented by the trigger selection, based on different technologies according to the constraints on the processing time and on the volume of data to be analyzed.

The Level 1 (L1) step is realized at the hardware level and is composed of custom hardware processors, which perform a rough reconstruction from the calorimeter system or the muon detector information. It exploits coarsely segmented data coming from the calorimeters and the muon detectors, while holding the high resolution data in pipelined memories in the front-end electronics, as shown in Fig. 2.12. The energy deposits from the calorimeters are grouped into *trigger towers*, which are the combination of calorimeter readout units into objects representing jets, electrons, photons, and taus, and are used to compute energy sums. In a similar way, hits in the DT, CSC, and RPC chambers are combined to reconstruct *track segments* or *hit patterns*. The local triggers, also called *Trigger Primitive Generators* (TPG), are based on these two classes of objects.
The goal of the L1 is to reduce the number of events to a rate of 100 kHz, that represents the upper limit imposed by the CMS readout electronics. It aims to select events containing detector signals consistent with a muon, electron, photon, τ lepton, jet, or MET, with trigger thresholds chosen according to the LHC instantaneous luminosity during different

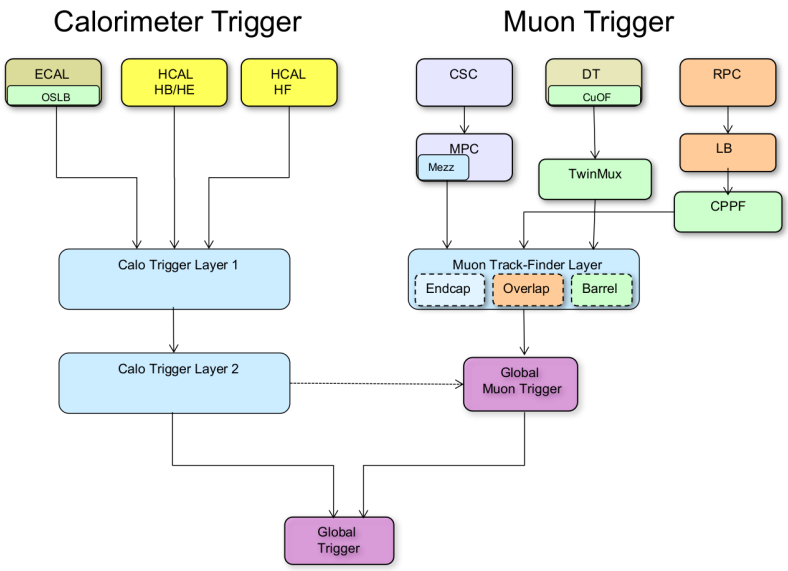


FIGURE 2.12: Overview of the upgraded L1 trigger system for the Run 2 [70].

data-taking periods in order to keep an output rate of about 100 kHz. The information provided by the ECAL, HCAL, and HF subsystems and by the DT, RPC, and CSC subdetectors are processed in the calorimeter trigger and in the muon trigger, respectively, and then combined by the *micro global trigger* (μGT), that accepts or rejects the event. The decision of accepting or not an event has to respect a fixed latency of 3.8 μs, corresponding to the time available for data processing. Given that the tracking detector is not used for the L1 trigger, the experimental signature of electrons and photons is very similar, so that they are both reconstructed as an e/γ object. Both the L1 calorimeter and muon trigger systems have been upgraded for the LHC Run 2 [70], reducing the trigger rate and improving the trigger efficiency for a wide variety of physics processes. All electronic boards of the system have been replaced in order to allow more sophisticated algorithms to be run online and to ensure high performance under the higher instantaneous luminosity and harsher pileup conditions. At the end, selected events are delivered to the *Data Acquisition System* (DAQ), which reads data from various subsystems for offline storage.

The second stage, referred to as *high level trigger* (HLT), represents a software-based system implemented in a farm of about one thousand commercial processors and 22000 CPU cores, having access to the complete detector information at the full granularity. It can perform complex calculations optimized for fast processing (220 μs), aiming to reduce the event rate below 1 kHz. The HLT menu in CMS has a modular structure accounting for more than 400 different HLT paths for the LHC Run 2 data-taking. Each HLT path is a sequence of reconstruction and filtering modules organized in blocks of increased complexity which reproduces the offline selection for a certain physics object, for combinations of different objects, or even for more sophisticated selections used in the physics analyses. Each physics analysis uses specific paths, so that the topic will be not examined in depth here. Details on the dedicated HLT paths adopted in the $H \to ZZ^* \to 4\ell$ analysis will be presented in Chapter 4. Before the beginning of Run 2, the HLT algorithms were substantially improved, in particular adopting new approaches for the online track reconstruction to drastically reduce the computing time while enhancing the performance.

At the end of the chain, events are recorded on the tapes of the CERN Tier-0 and then distributed for the subsequent analyses according to a hierarchical model which is organized in four levels.

2.2.4 Global event reconstruction and Particle Flow algorithm

After having presented how the CMS experiment detects particles and records events, this conclusive section is going to summarize the identification and reconstruction of physics object candidates emerging from each collision event via a particle flow approach [71].

The PF technique relies on the optimized combination of information from the CMS subdetectors, which are often redundant, in order to give a global and coherent description of the event in the form of reconstructed particle candidates and improve the reconstruction of final state objects. Practically, events collected by the CMS detector are centrally processed using a common set of reconstruction algorithms referred to as *event reconstruction* which start from raw data and provide as output a collection of detected particles with associated properties such as momentum and charge.
Firstly, the individual particles are classified into mutually exclusive types: muons, electrons, photons, charged hadrons, and neutral hadrons. The signatures of the main categories of particles passing through the CMS detector are illustarted in Fig. 2.13.

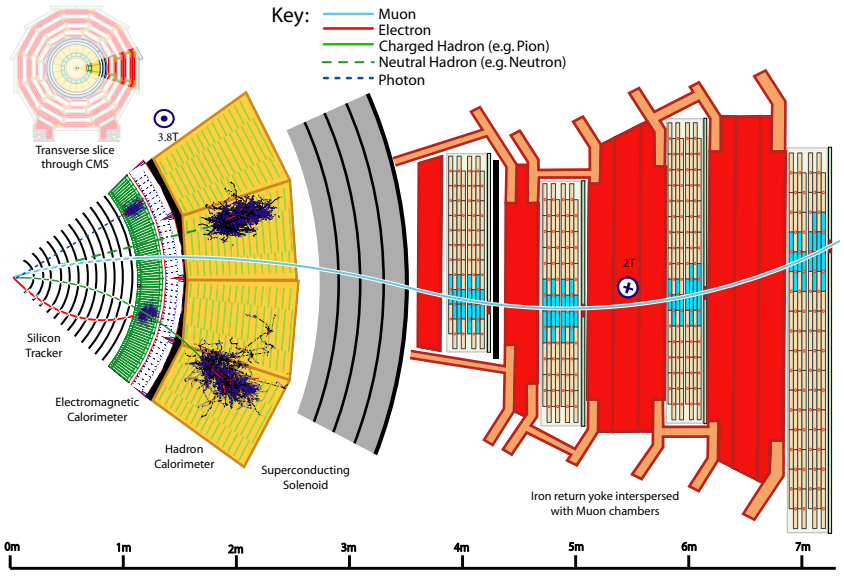


FIGURE 2.13: Sketch of the particle interactions in a transverse slice of CMS from the beam IP to the muon detector [71].

They are optimally measured by the CMS detector, thanks to the high granularity of the tracking system and ECAL, the high magnetic field of 3.8 T, the hermeticity of the HCAL and HF, and the excellent resolution of muon system. The track trajectory is reconstructed from the hits in the tracker and then linked to deposits in the ECAL only (electrons) or in the HCAL as well (charged hadrons), while the passage of a photon is identified as ECAL energy clusters not matched to the extrapolation of any charged particle trajectory to the ECAL. Muons are identified as tracks present in the central tracker and consistent with either tracks or several hits in the muon system, potentially associated with calorimeter deposits. Charged and neutral hadrons create hadronic showers in the ECAL with a subsequent absorption in the HCAL with the corresponding clusters being used to estimate their energies and directions. An indirect measurement of the presence of noninteracting, uncharged particles provided by the calorimeters is crucial to compute the missing energy which could be a signature of new particles and phenomena. The primary p-p interaction vertex is identified as the one with the largest value of summed physics object p_{T}.

Secondly, higher level physics objects such as jets, τ leptons, MET, and lepton isolation quantities can be built starting from the PF candidates.

* All particles are clustered into jets with the $anti\text{-}k_T$ algorithm, as detailed in Section 3.5.1. Jet momentum and spatial resolution are good thanks to the excellent ECAL granularity and the high quality of tracking detectors.
* The τ lepton, characterized by a very short lifetime, can be identified thanks to its hadronic decays through the reconstruction of the intermediate resonances.
* The missing transverse energy vector is reconstructed as the opposite of the transverse momentum sum of all final state particles reconstructed in the detector.

Reconstruction algorithms specific for the most relevant physics objects in the context of the $H \to ZZ^* \to 4\ell$ analysis will be discussed more in detail in Chapter 3.

2.3 The future of particle accelerators

The LHC is the most powerful hadron collider ever built, representing the frontier of the high energy physics research. In the coming decade, the LHC, including its high luminosity upgrade [20, 72], and the experiments currently operating in the CERN area will remain the primary tool in the world for exploring the high energy frontier. However, according to the current understanding, it seems that the direct observation of BSM particles contained in SUSY models, composite models, or models with extra dimensions is beyond the reach of the LHC and the indirect discovery is not allowed by the precision achievable with measurements at the LHC. For this reason, the CERN Council established the European Strategy Group (ESG) in the context of the European Strategy for Particle Physics (ESPP). It is the scientific process which aims to make a decision for the future of the field. It started in 2006 and it was first updated in 2013 to reflect the evolution of the field after the H boson discovery at the LHC.

The goal of the ESG is the coordination of the process of proposal and discussion about the new accelerator facilities needed at the end of the LHC lifespan in order to explore the Higgs sector and the existing mysteries of the fundamental physics laws. A summary of all alternatives which have been investigated and proposed will be presented in this section, putting some emphasis in the second part of the discussion on the project of an electron-positron accelerator operating as a H boson and electroweak factory. According to the last update of the European Strategy in June 2020, there are strong scientific arguments to favor such a collider as a priority [73].

2.3.1 Overview of the proposals

In the last years, the potential of different future colliding beam facilities has been investigated in detail. Two approaches are considered. On the one hand, to increase the achievable precision of the measurements related to the H boson and electroweak processes involving W and Z bosons or top quark. On the other hand, to operate at higher center-of-mass energues in order to access higher mass states from new physics. In this section, the following proposals of new high luminosity energy frontier particle accelerators projects beyond the HL-LHC are discussed and compared.

* **Large circular leptonic colliders** with a circumference of about 100 km, like the electron-positron Future Circular Collider (FCC-ee) [74, 75] and the Circular Electron Positron Collider (CEPC) [76, 77], are designed to run at the Z peak of 91.2 GeV and then at the WW/ZH production thresholds of 160 GeV and 240 GeV, respectively. FCC-ee also plans to collect data at the $t\bar{t}$ peak around 365 GeV.

* **Linear colliders**, like the International Linear Collider (ILC) [78, 79] and the Compact Linear Collider (CLIC) [80, 81], are supposed to operate at higher energies. For the former, a first stage at a 250 GeV center-of-mass energy is assumed while the baseline in the TDR aimed for 500 GeV. For the latter, three energy stages at 0.5 TeV, 1.4 TeV, and 3 TeV, respectively, are taken into account.
* **Large proton-proton colliders** in a circular tunnel of about 100 km in circumference, like the Future Circular Collider (FCC-hh) [82] and the Super Proton Proton Collider (SPPC) [83, 84], are designed to run at a center-of-mass energy of 100 TeV and 75 TeV (up to 125-150 TeV), respectively.
* The project of a muon collider is not discussed here since it is in the less advanced stage.

Circular lepton colliders can reach higher luminosities but lower center-of-mass energies, while linear lepton colliders can ensure higher energies but are not really competitive with circular colliders for luminosity at lower energies. The data-taking programs of the circular colliders call for 5 ab^{-1} of integrated luminosity in three years of operation at $\sqrt{s}$ = 240 GeV for FCC-ee and 5 ab^{-1} in seven years at $\sqrt{s}$ = 240 GeV for CEPC. ILC is expected to collect 2 ab^{-1} of data in fifteen years of operation at $\sqrt{s}$ = 250 GeV, while CLIC will collect 0.5 ab^{-1} in seven years of operation at $\sqrt{s}$ = 380 GeV. This demonstrates again the large advantage in terms of integrated luminosity of the circular lepton colliders. FCC-ee also plans to collect data at the $t\bar{t}$ peak, with an integrated luminosity of 1.5 ab^{-1}, making this machine the only one that can do extremely high precision measurements of all the electroweak sector. The baseline luminosities expected to be delivered at the CEPC, CLIC, FCC-ee, and ILC center-of-mass energies are illustrated in Fig. 2.14. Some proposals to further improve the luminosity are under study, so these reference values can be found slightly different in the future. The main limit for e^+e^- circular colliders is given by the energy loss by synchrotron radiation, which can be expressed per turn as:

$$\frac{\Delta E}{\text{Turn}} \sim \frac{\gamma^4}{r} \tag{2.11}$$

where $\gamma = {}^{E}/_{m}$, m, and E refer to the mass and energy of a particle, and r is the radius of the circular accelerator.

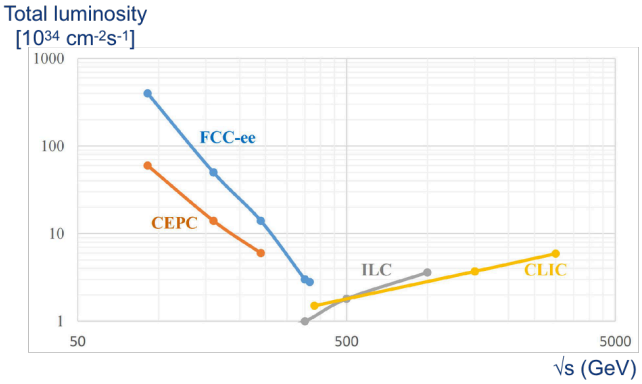


FIGURE 2.14: Maximum instantaneous luminosities of proposed future colliders. The center-of-mass energy and the instantaneous luminosity are reported on the horizontal and on the vertical axis, respectively. At energies below $\sqrt{s}$ = 365 GeV circular colliders have a significant advantage over linear colliders.

For this reason, in the last twenty years a large effort has been oriented to the development of linear accelerators, aiming to produce center-of-mass energies between the Z pole and around 1 TeV to cover the full SM phenomenology. In this context, an essential asset for precision tests of a chiral theory such as the SM is represented by polarized beams. Both ILC and CLIC are able to achieve a degree of polarization at the IP of around 80%.

The maximum energy reach of the linear accelerators is still not competitive with hadron colliders, which allows for investigating larger energy ranges enhancing the discovery potential. The most challenging aspect of large hadron colliders is the development of high field dipole magnets. Magnets used in the FCC-hh project have to provide a magnetic field of 16 T over a length of about 13.5 m, while in the SPPC values between 12 T (initial design) and 24 T at a length of 15 m have to be ensured. The current targeted densities at the highest values are around 1500 A/mm^2. This requires the application of Nb_3Sn as conductor material or high temperature superconducting wires instead of the commonly used NbTi conductors, because it allows for higher current densities. However, the precision of p-p machines is limited by the huge background coming from the QCD induced reactions due to the composite nature of the initial state hadrons (see Section 2.1.2). Consequently, the parallel operation of a precision machine is very useful for a deeper investigation of the SM and BSM physics.

In the following, a discussion of the detector requirements for electron-positron circular colliders is presented, starting from a short description of the concept of the FCC-ee machine, which is very similar to the CEPC project. All details about the preparation of a construction project, the design, and the technical aspects of the machines can be found in the Conceptual Design Report (CDR) of the FCC-ee [75] and CEPC [76], respectively.

2.3.2 Detector requirements at electron-positron circular colliders

Considering the unique nature of the H boson, an electron-positron collider able to copiously produce H bosons in a very clean environment would lead to a dramatic progress in the investigation of the Higgs sector and in mapping the interactions of the H boson with other particles. The strategic guideline reported in occasion of the 2013 update of the ESPP clearly stated [85]:

"There is a strong scientific case for an electron-positron collider, complementary to the LHC, that can study the properties of the Higgs boson and other particles with unprecedented precision and whose energy can be upgraded."

For this reason, the current vision of the particle physics community, reinforced after the outcome of the 2020 update of the ESPP, is to prepare a Higgs factory followed by a future hadron collider sensitive to energy scales an order of magnitude higher than those of the LHC. Currently two compelling proposals of infrastructures based on a high luminosity e^+e^- circular collider are considered: the FCC-ee, designed and supported by the CERN community in Europe, and the CEPC in China. They are supposed to be a precision instrument for the in-depth investigation of nature at the smallest scales, optimized to study the Z, W, and H bosons and, in the case of FCC-ee, also the top quark, with samples of 5×10^{12} Z bosons, 10^8 WW pairs, 10^6 H bosons, and 10^6 top quark pairs. Both colliders will be implemented in subsequent stages spanning the entire energy range.

In both projects, the machine will be installed in a tunnel of about 100 km and most of the infrastructure, such as the underground structure and surface sites, the electrical distribution, cooling, ventilation, and RF systems, will be used for a subsequent higher energy hadron collider. The large radius of the machine aims to remedy at least partially the energy loss issue due to synchrotron radiation mentioned above as the main shortcoming of

circular colliders. Moreover, they will naturally benefit from the progress in the acceleration gradient of cavities to compensate for the energy loss. They deliver the highest rates in a clean, well defined, and precisely predictable environment, offering extreme statistical precision and leading to accurate measurements of the SM particle properties, sensitivity to rare processes detection and to the observation of tiny violations of established symmetries. Moreover, they provide high precision center-of-mass energy calibration at the 100 keV level at the Z and WW energies, a peculiar feature of circular colliders.

The design goal is the maximization of the luminosity for each energy. The project foresees the ring installation in a quasi circular tunnel composed of arc segments interleaved with straight sections. A double ring collider configuration is used as shown in Fig. 2.15: two beam lines cross at two IPs with a large horizontal crossing angle of 30 mrad. A scenario accounting for four IPs is under study.

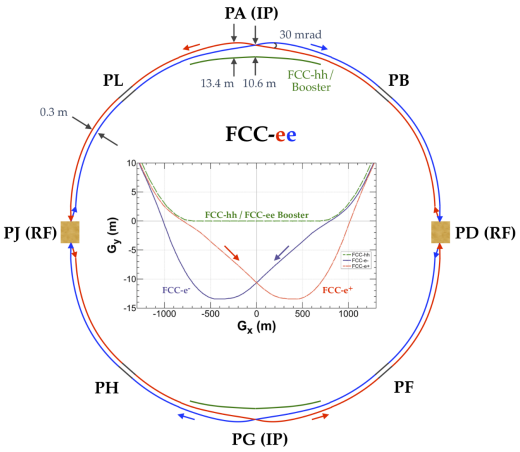


FIGURE 2.15: Schematic layout of the FCC-ee collider rings. The plot in the middle shows the two beams trajectories at the IP [75].

Beams are only mildly bent before the IP to minimize synchrotron radiation into detector volumes. In addition, profiting from the crossing angle, a crab waist collision scheme is adopted to enable an extremely small vertical beta function β_y^* at the IP. The length of the free area around the IP and the strength of the detector solenoid are kept constant at 2.2 m and 2 T, respectively, for all energies. The local energy of the beam can be affected by a deviation of $\pm 1.2\%$ between the entrance and the exit of the RF sections, resulting in a deviation of the orbit due to the horizontal dispersion in the arc and associated optical distortion which creates an unstable region. There are two main motivations for a double ring collider. On the one hand, the double ring scheme allows a large number of bunches. To reach the desired luminosity at low energies (especially at the Z peak), more than 16000 bunches must be stored in the beam so it is essential to avoid parasitic collisions. On the other hand, at the highest energy ($\mathrm{t\bar{t}}$ production) the optimal number of bunches is reduced to about 30, but the double ring scheme allows to adjust the strength of each magnet taking into account the energy loss due to synchrotron radiation in order to restore the ideal orbit of the beam and optics almost completely. This scheme is called *tapering* and it can not be applied to the e^+e^- beams simultaneously in case of a single ring. The critical energy of the synchrotron radiation of the incoming beams toward the IP is kept below 100 keV and the synchrotron radiation power is limited to 50 MW per beam at all beam energies. The colliding beams have to be provided by a top-up injection scheme in order to maintain the stored beam current and the luminosity at the highest level throughout the physics run, otherwise the integrated luminosity would be more than an order of magnitude lower. The lifetimes of the beams is particularly short (between 20 and 70 minutes) in order to mantain

the desired high luminosities. Moreover, the energy spread and the beam lifetime are also affected by *beamstrahlung*, which is a special type of synchrotron radiation, emitted during the collision due to the field of the opposite bunch. Thus, a booster ring is required in the collider tunnel where particles are accelerated to their nominal energy before being injected into the main ring.

An advantage of circular colliders is the possibility to deliver collisions to multiple IPs. Two complementary designs of a general purpose detector are currently under study for the FCC-ee and CEPC projects and they have been included and documented in the respective CDRs. The structure of both the experiments is outlined in Fig. 2.16.

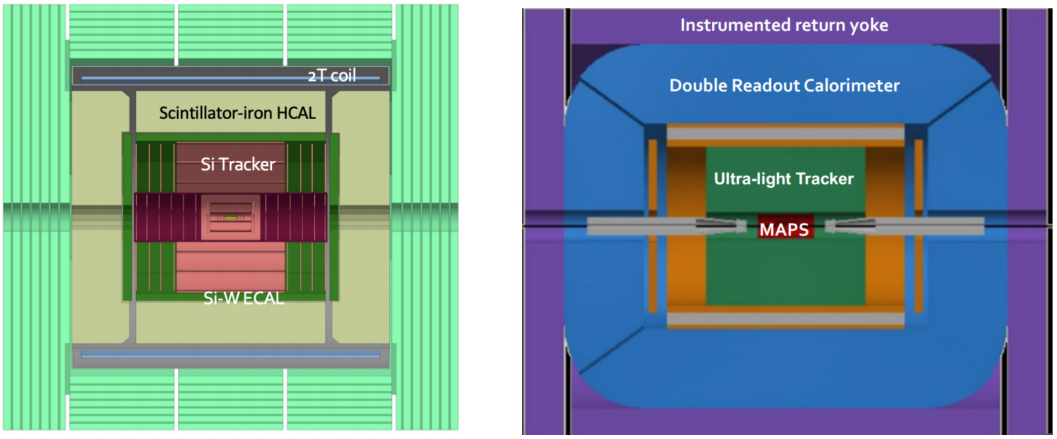


FIGURE 2.16: Schematic view of the overall structure of the CLD (left) and IDEA (right) detectors.

* **CLD** (CLIC-Like Detector) is a detector model inspired by detectors for CLIC and ILC and optimized for FCC-ee and CEPC conditions, accounting for a silicon tracker and a 3D-imaging highly granular calorimeter, surrounded by a conventional superconducting solenoid coil.
* **IDEA** (Innovative Detector for Electron-positron Accelerators) is composed of a silicon pixel vertex detector, an ultra light drift chamber (DCH) with excellent particle identification capabilities, a silicon wrapper that surrounds the central tracker, an ultra light superconducting solenoid coil followed by a preshower and then a dual-readout (DR) calorimeter for very good hadronic and jet energy resolution. It exploits the μ-RWELL technology for the preshower and the muon spectrometer, that, as usual, is the last subdetector of the experimental apparatus. The IDEA detector concept is going to be presented in the next section as part of the thesis discussion.

The collider performance and the layout of the interaction region, presented above, lead to some detector requirements which have to be taken into account in the development of the detector concept. First of all, the detector solenoid magnetic field is limited to 2 T because of the beam crossing angle, aiming to avoid a significant impact on the luminosity. Then, the interaction rates (up to 100 kHz at the Z pole) imply strict constraints on the event size and readout speed. To minimize the occupancy from beam-induced backgrounds, an extremely light tracker is required. The precise measurement of the significant center-of-mass energy spread, which ranges from 90 MeV at the Z pole up to 500 MeV at the highest energies, requires an angular resolution better than 100 μrad for muons. Finally, the luminometer has to ensure a precision better than 10^{-4} on the luminosity measurement, being located only 1 m away from the IP.

In addition to the constraints imposed by the machine, the statistical precision and the physics programme at different energies offered by e^+e^- circular colliders (see Section 7.1)

have to be taken into account to build an adequate detector. To study the physics at the Z and H pole in the best way, an optimal momentum resolution is required, aiming for $\Delta p_{\mathrm{T}}/p_{\mathrm{T}}^2 \sim$ 2-4$\times 10^{-5}$ (GeV^{-1}), together with an excellent electromagnetic and hadronic energy resolution $\sigma(E)/E$ around $10\%/\sqrt{E}$ and $30\%/\sqrt{E}$, respectively. These aspects allow to reach good jet energy resolution and efficiency in the building of the jet kinematic variables and in the flavor tagging. Considering the study of the τ polarization, high granularity and energy resolution of the calorimeter are required, ensuring PF capability and excellent separation of b and c quarks over a large p range. Moreover, the investigation of heavy flavor physics needs a good particle identification (PID) capability and high vertex reconstruction efficiency for flavor tagging and decay length measurements. Finally, to study top physics, the missing energy resolution is important and is mainly related to the calorimeter angular resolution.

The IDEA layout has been designed to assess all these detector requirements in the best possible way.

2.3.3 Concept of the IDEA detector

As a result of many years of Research and Development (R&D) activities, the IDEA detector concept has been designed exploting innovative detector technologies. It is geometrically composed of a cylindrical barrel closed hermetically at the two ends by endcaps. A scheme of the proposed layout for the IDEA proto-experiment is shown in Fig. 2.17. Starting from the primary vertex, the first detector is the central tracker, composed of a silicon pixel vertex detector followed by a large central drift chamber and a silicon wrapper. The central tracker is housed inside a thin, about 1 X_0 thick, solenoidal coil, that provides a 2 T magnetic field. The central tracker guarantees very good momentum resolution coupled with an excellent particle identification. After the solenoid there are a preshower detector and a dual-readout calorimeter that provide excellent energy resolution and particle identification, followed by the muon detection system, housed in the iron return yoke, based on an innovative type of Micropattern Gaseous Detector (MPGD), the μ-RWELL detector. The subsequent presentation of each subdetector is focused on the strong and innovative points of the project.

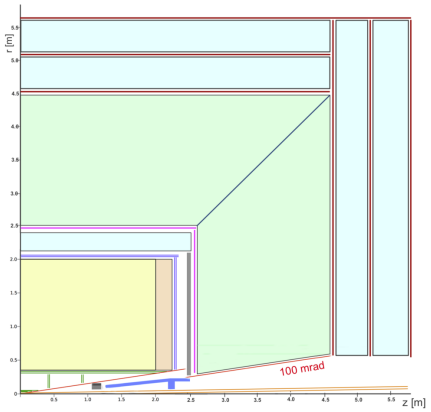


FIGURE 2.17: Longitudinal view of the IDEA detector layout with the drift chamber (yellow), surrounded by silicon pixels (green) and silicon strips (blue), the coil and the instrumented return yoke (light blue), the preshower layers (purple), and the DR calorimeter (green). The barrel region extends up to 45°, corresponding to η = 0.88, and the coverage of the endcap region reaches 100 mrad.

Magnet

The IDEA baseline design relies on a very thin and light superconducting solenoid of 2 T. It features an inner radius of 2.1 m and an outer radius of 2.4 m and is placed in front of the DR calorimeter. The fact that it is positioned between the tracker and the calorimeter system represents a specific feature of IDEA. The material thickness of the magnet, composed of the cold mass and the cryostat, should be as small as possible, both in radiation and geometrical lengths. The stored energy is reduced by a factor four by positioning the coil inside the calorimeter and the cost is halved. According to the current baseline design, the cryostat is made of an optimized conventional material that can provide a thickness of 0.74 X_0, corresponding to 30 cm, at normal incidence, of which 0.46 X_0 is for the cold mass and 0.28 X_0 for the cryostat, resulting in 0.16 λ_0 at 90°. The magnet is equipped with a thin iron return yoke which houses the muon system and protects the beams from emittance blow-up.

Tracking system

The IDEA tracking system is immersed in the 2 T magnetic field provided by the solenoid. It is based on a Monolithic Active Pixel Sensors (MAPS) silicon tracker in the innermost part of the detector, surrounding the beam pipe at a radius of 1.5 cm, and a ultra light highly granular drift chamber, which allows to reduce the material budget. Moreover, a silicon wrapper is placed around the DCH.

The IDEA vertex detector is composed of very light detectors corresponding to a material budget of 0.3% (1.0%) X_0 per innermost (outermost) layer. It is expected to have an excellent space resolution of about 5 μm, high efficiency at low power, and low dark noise rate. The DCH is a unique cylindrical 4 m long volume which extends from an inner radius of 0.35 m to an outer radius of 2 m. It consists of 112 coaxial layers, at alternating sign stereo angles, arranged in twenty-four identical azimuthal sectors. A conservative space resolution of $\sigma_{xy} \simeq 100$ μm is expected. It is composed of 56448 drift cells with an approximately square cell size varying between 12.0 mm and 13.5 mm. A very light gas mixture of 90% He and 10% $i\mathrm{C_4H_{10}}$ (isobutane) is used for the chamber operation, resulting in a drift time of about 350 ns. All the hits in a cell coherent with the maximum drift time are grouped together to create a hit. The number of ionization clusters generated by a minimum ionising particle (MIP) is about 12.5 cm^{-1}. Cluster counting and timing techniques can be exploited to improve both spatial resolution (σ_{xy} < 100 μm) and particle identification capabilities ($\simeq$ 2%). The high transparency in terms of radiation lengths represents the main peculiarity of the DCH. The total amount of material in the radial direction toward the barrel calorimeter is of the order of 1.6% X_0, while it is about 5.0% X_0 in the forward direction, with the largest percentage (75%) located in the end plates which are instrumented with the front-end electronics. The DCH has been designed to provide good tracking, high precision momentum measurement, and excellent particle identification by cluster counting. It is able to measure dE/dx for each track. The presence of a layer of silicon microstrip detectors in the outer part of the DCH provides an additional accurate space point and allows to define the tracker acceptance.

Preshower

A preshower detector with high space resolution is placed between the magnet and the calorimeter system in the barrel region and between the drift chamber and the endcap calorimeter in the forward region in order to provide a precise measurement of the position of showers initiated before the calorimeter. The magnet coil works as an absorber of about 1 X_0 and is followed by one layer of μ-RWELL chambers [86–88] with a strip pitch

of 400 μm. The current prototypes of μ-RWELLs are able to give information about a single space coordinate, x or y, but the development of chambers with a bidimensional readout is ongoing (some details in Chapter 7). The μ-RWELL detectors define the tracker acceptance volume with precision and improve the tracking resolution thanks to an excellent space resolution down to 40 μm. The optimization of the preshower system and the evaluation of its performance is in progress. Some of these studies are presented in Chapter 7 together with a brief description of the μ-RWELL technology, as one of the main targets of my personal contribution to the ongoing development of the IDEA detector.

Dual-readout calorimeter

Dual-readout calorimetry represents currently a mature and well-known technology that ensures optimal electromagnetic and hadronic resolution in the same detector. This calorimeter concept has been extensively tested on prototypes by the DREAM/RD52 collaboration [89–91] and is now moving toward a technology design study in order to be realistically usable within an experiment. The main advantage of this calorimetric technique is to overcome the noncompensation limit for hadronic shower energy measurements thanks to the fact that scintillation and Cherenkov light are simultaneously detected. Scintillating photons (S) are proportional to the deposition of energy by all ionizing particles in the volume of the calorimeter, while Cherenkov photons (C) provide a signal almost exclusively related to the electromagnetic component of the shower.
On the one hand, given that electromagnetic showers are independently sampled in the Cherenkov and scintillation fibers with a corresponding resolution σ_S and σ_C, the combination of the two signals on event-by-event basis allows to improve the electromagnetic resolution:

$$E = \frac{E_S/\sigma_S^2 + E_C/\sigma_C^2}{1/\sigma_S^2 + 1/\sigma_C^2}. \tag{2.12}$$

On the other hand, the response to single hadrons is studied with the *dual-readout method* [90], where the particle energy is expressed as a function of scintillating and Cherenkov signals:

$$\begin{aligned} S &= [f_{em} + (h/e)_S \times (1 - f_{em})] \times E, \\ C &= [f_{em} + (h/e)_C \times (1 - f_{em})] \times E. \end{aligned} \tag{2.13}$$

The electromagnetic fraction within the hadron shower is denoted as f_{em} and the ratios of non-electromagnetic to electromagnetic response for S and C signals are referred to as $s = (h/e)_S$ and $c = (h/e)_C$, respectively. The electromagnetic (e) and non-electromagnetic (h) response are shown in Fig. 2.18 as the average values of the signal distributions of the deposited energy for the e and h components of hadron showers. The f_{em} and the energy E can be derived on event-to-event basis as:

$$\begin{aligned} f_{em} &= \frac{c - s(C/S)}{(C/S)(1-s) - (1-c)}, \\ E &= \frac{S - \chi C}{1 - \chi} \quad \text{with} \quad \chi = \frac{1-s}{1-c}, \end{aligned} \tag{2.14}$$

where the factor χ is independent from both energy and particle type. The proposal of the IDEA calorimeter is a 2 m long detector, corresponding to about 9 λ, that surrounds the preshower covering the full volume down to 100 mrad of the z-axis without inactive regions. It features a unique longitudinally unsegmented detector equipped with scintillation and Cherenkov fibers, so that it is sensitive to the independent signals from the production of scintillation and Cherenkov light. An event-by-event measurement of the electromagnetic fraction can be performed, applying a correction to the reconstructed primary hadron

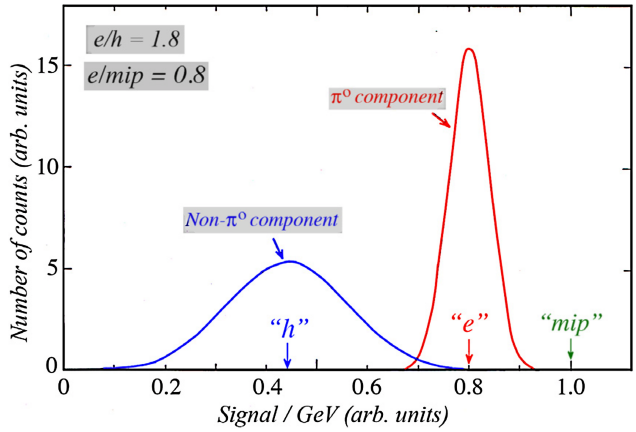


FIGURE 2.18: Distributions of the signal per unit deposited energy for the e and h components of hadron showers normalized to the response for MIPs [90].

energy for its fluctuations. By the combination of the two signals, an excellent energy resolution for both electromagnetic and hadron showers is obtained. See the discussion presented in Chapter 7 for all details. In addition, when considering isolated particles, the DR calorimeter provides very good particle identification capabilities to discriminate between muons, electrons/photons, and hadrons thanks to the separation given by the C/S ratio.

To achieve the energy resolutions needed in experiments at future e^+e^- colliders, the most challenging obstacle is represented by the fluctuations among the electromagnetic and non-electromagnetic component of hadronic induced showers. The calorimeter can be calibrated using electrons, which have an electromagnetic fraction equal to 1. The S and C signals can be independently calibrated for each *tower*, which indicates a group of fibers, by steering electrons of known energy. Finally, the fine transverse granularity allows to distinguish close showers guaranteeing good matching to tracks in the preshower signals and also to muon tracks. For this reason, it represents a good candidate for efficient PF reconstruction. The main issue is related to the possibility of disentangling signals produced by overlapping electromagnetic and hadron showers, which requires longitudinal segmentation as well. Currently, several ways to implement such a segmentation are under study.

Muon system

The muon system consists of three layers of chambers interleaved within the return yoke of the magnetic field and is based on the μ-RWELL technology (see Chapter 7) like the preshower detector. A cost effective chamber technology has been adopted because of the substantial area to be covered to ensure high efficiency in the detection of muons and to provide standalone measurement of the muon tracks, useful also for long lived particles searches. The choice of the μ-RWELL detectors is mainly due to their position resolution. The readout is segmented with a larger pitch strips of 1.5 mm with respect to the preshower.

2.4 Conclusion

In this chapter, a presentation of the LHC machine and the main features of the CMS detector has been provided. In particular, the emphasis was put on the overview of the experimental apparatus needed to perform the $H \to ZZ^* \to 4\ell$ analysis, which represents the main topic of this thesis. A careful examination of all steps of the analysis is going to be the content of Chapters 3, 4, 5 and will flow into Chapter 6 with the discussion of the results.

In view of the completion of the LHC physics research programme around the second half of the 2030s, the discussion about the future of high energy physics and the type of future accelerators needed to explore the broad range of fundamental questions in particle physics is very intense within the scientific community. After the very recent outcome of the ESPP update, the highest priority has been set for an electroweak and H boson factory. The CERN proposal is an electron-positron circular collider in a 100 km tunnel operating at a very high luminosity at different center of mass energies, ranging from the Z pole production to above the top pair production threshold (from 90 to 365 GeV). In particular, the operation of a similar machine at the Higgs-strahlung cross section peak will allow to investigate the unique nature of the H boson in a very clean environment and to study the couplings of the H boson with other particles with unprecedented precision. Two alternative proposals of e^+e^- circular colliders have been developed in the last years: FCC-ee at CERN and CEPC in China. Currently, the highest priority is the study and optimization of the performances of the various detector alternatives. The content of Chapter 7 will focus on this aspect, which includes my original contribution in the context of future accelerator studies.

Chapter 3

The golden channel and its main physics objects

The work presented in this thesis is focused on the study of the decay of the H boson into a pair of Z bosons, subsequently decaying into two leptons (muons or electrons). This particular decay channel played an extremely important role in the discovery of this new scalar particle.
The $H \to ZZ^* \to 4\ell$ decay channel, briefly denoted as $H \to 4\ell$, is often referred to as the "golden channel" since it is one of the most important final states to study the H boson profile. Across the next chapters, the highlights of the evolving strategy of the $H \to 4\ell$ analysis in the context of the H boson properties measurements are underlined. The reader will be accompanied on a journey showing how a particle physics analysis is not a static framework, but it can evolve in every aspect according to the data-taking conditions, the amount of data available and, in general, the goal of the study. The Run 1 analysis of the $H \to ZZ^* \to 4\ell$ process was located in the context of a discovery era, representing one of the key channels for the first observation of the H boson. At that time, the primary goal was to preserve as much signal as possible and maximize the sensitivity for a H signal within a large mass range. Thus, it was important to achieve a very high lepton selection efficiency in a wide range of momenta. Then, during the LHC Run 2 running period the statistical significance of the observation of the new boson in the four-lepton final state was already achieved. Consequently, the discovery of the H in all main production and decay modes became the new target of the CMS physics program. In this context, the focus of the $H \to ZZ^* \to 4\ell$ analysis switched to precision measurements of the H boson properties. This point is examined in more detail in the next chapters.

To introduce the first step of the the long way which leads to the measurements of the H boson properties exploiting the four-lepton final state, this chapter will provide a characterization of the process and will describe in detail the physics objects that constitute the inputs to build our analysis. Each section will briefly focus on the kinematic distributions of the objects under study after the full selection showing the good agreement between the data and their description in the simulated samples. These checks represented my regular task and my personal contribution to the analysis.

3.1 The four-lepton final state

Despite the low branching fraction of the process, which is 0.0124% considering a 125 GeV SM H boson, the four-lepton decay mode represents a key contribution to the measurements of many H properties. A plethora of results based on the Run 1 and Run 2 datasets comes from the study of this decay channel performed by both the ATLAS and CMS experiments, such as the determination of the mass of the H boson and of its spin-parity quantum numbers [28, 48, 92–95], the measurement of its decay width exploting both direct and indirect

approaches [96–99], of signal strengths and couplings to fermions and bosons [48, 95, 100] and of its anomalous couplings with a pair of neutral gauge bosons in both on-shell and off-shell regions [94, 95, 97, 101, 102], and the study of fiducial cross sections both integrated and differential as a function of various parameters [103–106]. Here we want to focus on the specific features of this decay channel:

* The complete reconstruction of its final state decay products allows to exploit the four-lepton invariant mass ($m_{4\ell}$) together with the kinematic information of the leptons as very discriminating variables to identify the presence of a H boson with respect to the other identical signatures coming from background processes. Having a way to select signal events killing, at the same time, the majority of the background processes was the main challenge of the Run 1 strategy, aiming to achieve a p-value as high as possible and declare the observation of the H boson in the four-lepton final state.

* The excellent momentum resolution of electrons and muons in the LHC detectors leads to a narrow resonance peak in the $m_{4\ell}$ distribution, allowing to measure the H boson mass with an extreme precision. In fact, the signal sensitivity strongly depends on the 4ℓ invariant mass resolution.

* The signal yield is not very high because of the $\mathcal{B}$ of the process but also the background yield is very low. The typical signal to background ratio is of the order of 2 to 1 in the mass window around the H peak.

A good knowledge of the different background sources is crucial in order to isolate the H boson signal. There are two different sources of background in a particle physics analysis: the irreducible background, which consists of a final state identical to the signal but is produced via other SM processes that are not of interest, and the reducible background, which presents a topologically different final state eventually identified as a signal candidate because of erroneous object identification. The $\mathrm{H} \to 4\ell$ decay channel copes with the following background processes.

Irreducible background The irreducibile nonresonant background around 125 GeV is due to the SM ZZ or Zγ^* production and comes from direct production of neutral vector bosons via $\mathrm{q\bar{q}}$ annihilation or gluon fusion, denoted as $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ and $\mathrm{gg} \to \mathrm{ZZ}$, respectively, and is presented in Fig. 3.1.

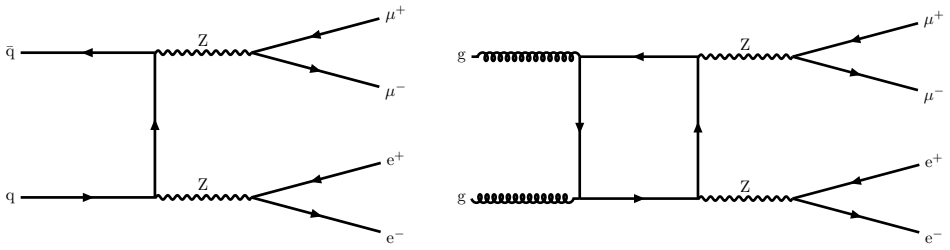


FIGURE 3.1: Feynman diagrams representing the irreducible $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ and $\mathrm{gg} \to \mathrm{ZZ}$ background processes.

Reducible background The reducible background mainly consists of events including bosons or pairs of top quarks and additional jets (Z/Zγ+jets, WZ/WW+jets, $t\bar{t}$+jets), leading to final states with two or three isolated leptons and jets. Extra leptons can either come from heavy flavor jets producing secondary leptons or from misidentification of heavy flavor hadron decays, in-flight decays of light mesons within jets, and (in the electron case) decays of charged hadrons overlapping with π^0, due to imperfections in reconstruction. Given that the dominant contribution comes from Z+jets events, the reducible background is commonly denoted as Z+X and some illustrative processes are shown in Fig. 3.2.

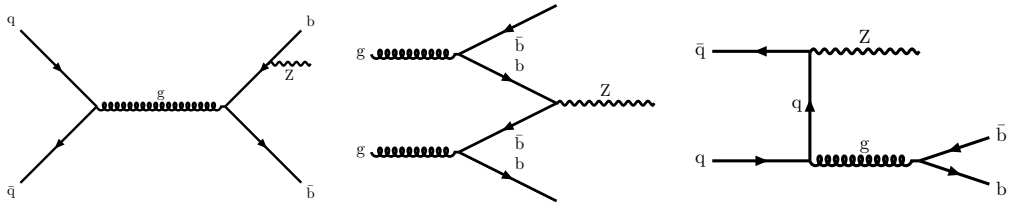


FIGURE 3.2: Feynman diagrams representing an example of reducible Z+X background processes.

Three mutually exclusive final states are considered in the analysis: 4μ, $4e$, and $2e2\mu$. In fact, each of them is characterized by different mass resolution and reducible background rates. Moreover, given the clean signature of the four leptons in the detector, different production modes of the H boson can be identified exploting the information given by the presence of other objects in the event like jets, missing transverse momentum, and extra leptons. In the following sections, we are going to see in which way the raw detector information is used to reconstruct these elementary objects needed to classify the events.

3.2 Muons

The CMS physics program largey relies on muons, involved in many of the signatures studied by the experiment concerning both precision measurements of the SM and searches for new physics up to the TeV scale. Given that muons are the only charged particles able to reach the outest layers of the detector, they provide a clear signature, easy to trigger on and to reconstruct with high efficiency and purity.

3.2.1 Muon reconstruction: standalone, global and tracker muons

As a first step to reconstruct muons with the CMS detector, a local reconstruction exploiting the information from the muon spectrometer alone is performed to identify hits in the individual DT and CSC chambers (see Section 2.2.2). Then they are used to form *segments*, that can be approximated by straight lines because the magnetic field is almost totally confined in the steel return yoke of the magnet.

On the one hand, DT track segments are reconstructed by the hits originating from the local DT electronics. Two different views are exploited to provide a measurement of the radial position and the bending angle. The hit reconstruction in a DT cell specifies the transverse distance between the wire and the intersection of the muon trajectory with the plane containing the wires in the layer. On the other hand, the CSC provides local charged-track segments by combining the cathode and anode hits (one per each layer of each station) in the transverse plane. Afterwards, an iterative approach based on the Kalman Filter technique [107] is adopted to build the tracks, taking into account the effect of the magnetic

field, as well as the energy loss and multiple scattering in the steel. Starting from the segments, the position and direction vectors is used to provide an initial estimate of the muon transverse momentum p_{T}^{μ}. This information is used to generate *seeds* that are propagated to each station layer looking for the most compatible segment based on the χ^2 of the fit. At this stage, hit clusters reconstructed in the RPC system (see Section 2.2.2) are also included in the tracks.

The result of this initial step is a collection of reconstructed track objects, which are referred to as *standalone muons*. In order to improve the momentum resolution, the beam spot position in the transverse plane can be used in the fit. The full muon tracks given by the local reconstruction in the muon spectrometer get further combined with tracks reconstructed in the inner detector. A *global muon* (GLB) is obtained by matching standalone muon tracks and inner tracks using the Kalman Filter technique. In the global reconstruction, as well as in the standalone one, the minimum requirement for a track is to have hits or segments in two station layers with at least one from DTs or CSCs. On the contrary, a *tracker muons* (TRK) is reconstructed by extrapolating an inner track inside-out to all the DT and CSC station layers and looking for compatible DT and CSC segments. The lower p_{T} reach of tracker muons is improved with respect to global muons reaching. Every track with p > 2.5 GeV and p_{T} > 0.5 GeV matched with at least one segment in the muon spectrometer is labeled as tracker muon. Thereby, the $\mathrm{H} \rightarrow 4\ell$ analysis can benefit significantly from this approach. Moreover, this algorithm also compensates for inefficiencies in the reconstruction of close dimuon pairs given by the ambiguity resolution in the standalone tracks.

The large majority of muons produced in p-p collisions (around 99%) with a sufficient high momentum is reconstructed by both algorithms. The information from all muon track types is finally combined into a single object to provide a coherent view of a muon candidate and to ensure the most efficient reconstruction. Energy deposits in the calorimeters can get also associated. More details about the muon reconstruction procedure can be found in Ref. [68, 108].

3.2.2 Muon identification

Given that the original collection of reconstructed muons contains misidentified charged hadrons or muons coming from in-flight decays, additional quality requirements must be applied to have as pure as possible samples of muon candidates. A set of variables and selection criteria, such as requirements on the type and number of missing hits, the track χ^2, and the compatibility between tracker tracks and standalone muon tracks when considering global muons, are defined to classify muons balancing efficiency and purity according to the analysis under study. The main identification (ID) criteria for muons adopted in the CMS physics analyses are summarized in the following with some performance plots reported in Fig. 3.3 comparing the three of them.

* A *loose muon* is a tracker or global muon selected by the PF algorithm. The loose ID is oriented to identify prompt muons originating at the primary vertex as well as muons from light and heavy flavor decays, avoiding the increase of the misidentification rate of charged hadrons as muons.

* The *medium muon* ID is optimized for prompt muons and for muons from heavy flavor decay. A medium muon is a loose muon reconstructed from a tracker track with many hits in the inner tracker layers. In case a muon is only reconstructed as a tracker muon, the *muon segment compatibility* is required to be larger than 0.451. It is a variable, ranging from 0 to 1, computed on the basis of the distance between the inner track, extrapolated to all the muon stations, and the closest DT and CSC segments from the same stations. If the muon is reconstructed as both a tracker muon and a

global muon, a less stringent cut on the muon segment compatibility is complemented by some quality requirements on the χ^2 of the track. See Ref. [68] for further details.

* A *tight muon* is a loose muon reconstructed from a tracker track with at least one pixel hit and hits from at least six layers of the inner tracker. It must be reconstructed as both a tracker muon and a global muon and must be compatible with the primary vertex. The tracker muon must have segment matching in at least two of the muon stations, consistently with the logic of the muon trigger. The global muon fit includes at least one hit from the muon system. The goal of such a tight criteria is to suppress muons from decay in-flight and from hadronic punch-through, hereafter referred to as *fake muons*.

* The *soft muon* ID is optimized for low p_{T} muons especially used in the context of b physics and quarkonia analyses. It is a tracker muon with a high purity tracker track which uses hits from at least six layers of the inner tracker including at least one pixel hit. The tracker track is matched with at least one muon segment in any station in both x and y coordinates and arbitrated.

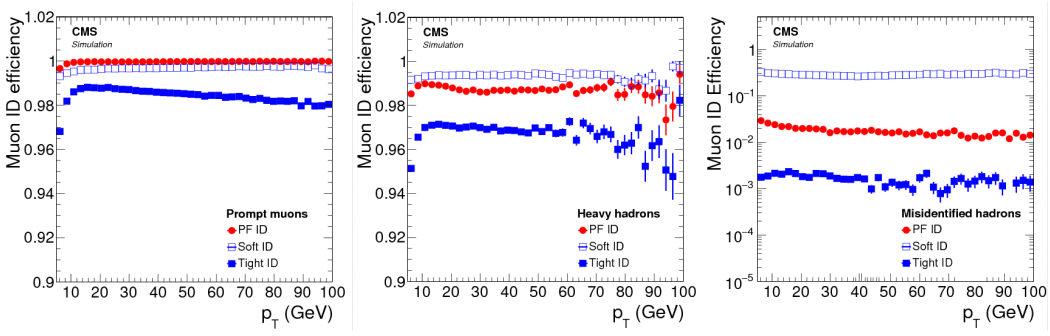


FIGURE 3.3: Efficiency of three ID algorithms (PF, soft, and tight) of a simulated muon track reconstructed as a tracker muon, as a function of the p_{T} of the reconstructed track, shown for prompt muons (left), for muons from heavy flavor decays (middle), and for misidentified hadrons (right) [71].

The $\mathrm{H} \rightarrow \mathrm{ZZ}^* \rightarrow 4\ell$ analysis aims to maximize its acceptance for muons with low transverse momentum down to 5 GeV preserving the kinematic acceptance for H boson candidates with a soft lepton in the final state. Therefore, a loose selection is chosen, further enhanced by means of the isolation and impact parameter (IP)[1] criteria to reject nonprompt muons. A loose ID is necessary to build the control regions needed for the analysis. Muons have to be reconstructed as global and tracker muons, thus standalone only muon tracks are discarded. Since the measurement of the momentum at low p_{T} achievable from the muon spectrometer information is not very precise because of the multiple scattering, the charge and momentum of this type of muons are extracted from the tracker track. An additional *ghost cleaning* step is performed to remove possible duplicates of the same muon track. Tracker muons that are not global muons are required to be arbitrated, for example if two muons are sharing 50% or more of their segments, the muon with lower quality is removed.

A **loose muon** is required to have a transverse momentum larger than 5 GeV in the acceptance region $|\eta^{\mu}| < 2.4$. The momentum acceptance is mostly limited by the increase of fake muons at very low p_{T}^{μ}, which makes the measurement of efficiencies and momentum

[1] Please note that IP in the previous chapter was used in place of *interaction point*, while hereafter always refers to the impact parameter.

scales more challenging. In addition, a set of cuts on the IP observables has to be satisfyed:

$$|d_{xy}| < 0.5 \text{ cm and } |d_z| < 1 \text{ cm}, \tag{3.1}$$

where d_{xy} and d_z are defined with respect to the primary vertex in the transverse plane and in the longitudinal direction, respectively, together with a cut on the muon significance of the impact parameter, $\text{SIP}_{3\text{D}}$. This quantity represents the ratio of the IP of the lepton track in three dimensions ($\text{IP}_{3\text{D}}$) with respect to the primary vertex position and its uncertainty:

$$\text{SIP}_{3\text{D}} \equiv \frac{|\text{IP}_{3\text{D}}|}{\sigma_{\text{IP}_{3\text{D}}}} < 4. \tag{3.2}$$

The goal of these requirements is to discard fake muons coming from in-flight decays of hadrons and cosmic rays and to identify only leptons coming from the primary vertex. The requirement on the significance of the impact parameter to the event vertex is particularly effective in the suppression of the $\text{Z}\text{b}\bar{\text{b}}$ and $\text{t}\bar{\text{t}}$ backgrounds. In fact, the high SIP tail is mainly populated by muons from displaced vertices tipically present in $\text{t}\bar{\text{t}}$+jets events. The effect of the cut on SIP is shown as an example in Fig. 3.4, where the distributions for signal muons coming from a Z boson and fake muons are compared. Signal and fakes muons are selected from a 2016 DY sample by using generator-level information in two different p_{T} regions: muons with a transverse momentum smaller (larger) than 10 GeV populate the low (high) p_{T} bin, respectively.

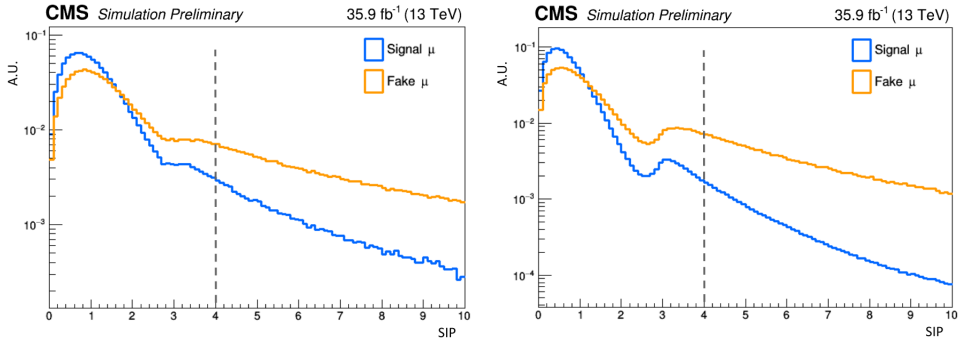


FIGURE 3.4: SIP distribution of signal and fake muons with a p_{T} smaller (left) and larger (right) than 10 GeV selected from a 2016 DY sample by using generator matching.

Then, to define a **tight muon** in the context of the $\text{H} \to 4\ell$ analysis, loose muons with p_{T}^{μ} smaller than 200 GeV have to pass the PF muon ID. The identification of a final PF muon (see Section 2.2.4) derives from three steps and aims to ensure high efficiency in selecting also genuine muons produced in hadronic jets.

1. Considering global muons, the activity in a cone of $\Delta R < 0.1$ around the muon candidate is evaluated, asking that the sum of the track p_{T} does not exceed 10% of the muon p_{T}^{μ}. On top of this preliminary isolation criterion based on detector-level information, another isolation specific of the $\text{H} \to 4\ell$ analysis defined in Section 3.2.3 will be applied, using the particle candidates reconstructed from the PF algorithm.

2. Between nonisolated candidates, PF muons are selected looking at the minimum number of hits in the muon stations and at the compatibility between the track and the segments as well as between track p_{T} and deposits in the calorimeters.

3. Finally, looser identification requirements based on the number of hits and the matching between track and hits are applied to accept PF muons between the remaining muon candidates featuring small energy deposition in the calorimeters with respect to the track momentum.

To identify a tight muon when its p_{T} is larger than 200 GeV, a tracker high p_{T} ID is required. All details about the definition are reported in Table 3.1. A relaxed selection of high p_{T} muons leads to increase signal efficiency for high mass searches, recovering selection efficiency on very collimated muons coming from very heavy resonance decays to two highly boosted Z bosons.

Plain text description	Technical description
Muon station matching	Inner track muon matched to segments in at least two muon stations
Good p_{T} measurement	$\frac{p_{\mathrm{T}}^{\mu}}{\sigma_{p_{\mathrm{T}}^{\mu}}} < 0.3$
Vertex compatibility (x-y)	$\lvert d_{xy}\rvert < 2$ mm
Vertex compatibility (z)	$\lvert d_z\rvert < 5$ mm
Pixel hits	At least one pixel hit
Tracker hits	Hits in at least six tracker layers

TABLE 3.1: Requirements to identify a tracker high p_{T} muon.

3.2.3 Muon isolation

In order to distinguish prompt muons coming from Z boson decays and muons arising from QCD processes where hadrons decay within a jet via electroweak interaction, the energy flow in their vicinity is required to be below a given threshold. The definition of the relative isolation $\mathcal{I}_{\mu}$ with respect to the muon p_{T} relies on the PF isolation. The relative isolation has to be smaller than 0.35. To subtract the pileup contribution to the isolation cone for the muons, the so-called $\Delta\beta$ correction is applied. During the Run 2, data was collected with an average of thirty interactions per bunch crossing, so that a large contribution can potentially affect the energy flow used for the muon isolation leading to a loss in signal efficiency. The factor $\Delta\beta$, defined as

$$\Delta\beta = 0.5 \sum_{\text{charged}} p_{\mathrm{T}}^{\mathrm{PU}}, \tag{3.3}$$

gives an estimate of the undesirable energy deposit of neutral particles (hadrons and photons) from pileup vertices and the factor of 0.5 corrects for the different fraction of charged and neutral particles in the neighborhood of the considered muon. Therefore, it is defined as

$$\mathcal{I}_{\mu} \equiv \frac{1}{p_{\mathrm{T}}^{\mu}} \Big(\sum_{\text{charged}} p_{\mathrm{T}} + \max\big[0, \sum_{\text{neutral}} p_{\mathrm{T}} + \sum_{\text{photons}} p_{\mathrm{T}} - \Delta\beta\big] \Big). \tag{3.4}$$

The quantities included in Eq. (3.4) correspond to the scalar sum of the transverse momenta of charged hadrons originating from the chosen primary vertex of the event, of neutral hadrons, and of photons, respectively. The isolation sums involved are all restricted to a volume bounded by a cone of angular radius $\Delta R = 0.3$ around the muon direction at the

primary vertex, where the angular distance between two particles i and j is

$$\Delta R(i,j) = \sqrt{(\eta^i - \eta^j)^2 + (\phi^i - \phi^j)^2}. \tag{3.5}$$

The effect of the $\Delta\beta$ correction is visible in Fig. 3.5 (right). After the application of the $\Delta\beta$ corrections, the stability of the isolation efficiency in data and simulated samples as a function of the number of vertices is significantly improved with respect to the 2011 scenario, shown in Fig. 3.5 (left).

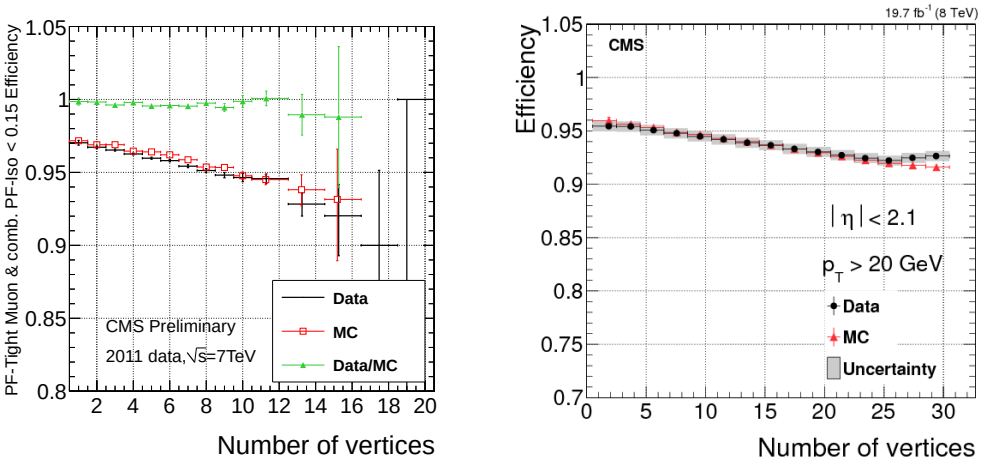


FIGURE 3.5: Isolation efficiency in data and simulated samples for muons from Z boson decays as a function of the number of vertices without (left) [109] and with the application of the $\Delta\beta$ correction [71].

3.2.4 Corrections and energy calibrations for muons

The muon resolution affects the invariant mass of a resonance and a possible muon scale bias also manifests as distortions or shifts in the invariant mass peak position of a resonance. For this reason, it is crucial to probe muon p_{T} scale and resolution and eventually extract appropriate corrections. This can be done exploiting the J/Ψ and Z invariant mass distributions. Several methods are used to measure the muon scale and resolution to cover the full muon p_{T} spectrum up to 1 TeV scale. For the low and intermediate p_{T} range, the *Rochester method* [110] is used, developed to correct biases in muon momentum reconstruction due to the mismodeling of detector alignment, the magnetic field, and energy losses. In fact, it is sensitive to the tracker alignment because the p_{T} measurement in the mentioned p_{T} range is fully driven by the tracker. It derives the calibration constants using Z and J/Ψ events with a dimuon decay and matching the data and simulated events to achieve a perfectly aligned scenario. The matching procedure is carried out in several steps. Concerning the muon scale, muons in data and Monte Carlo (MC) samples are corrected to match generated muons, while for the resolution the simulated events are smeared to match data.

The muon momentum scale is measured in data by fitting a Crystal Ball function to the dimuon mass spectrum around the Z peak in the $\mathrm{Z} \to \mu^+\mu^-$ control region. The energy scale and resolution for the 2016, 2017, and 2018 dataset is shown in Fig. 3.6 (top). The time evolution of the quantities related to the Z boson and leptons is also monitored across different runs of the same data period in order to check regularly that everything was working well, keep an eye on the lepton momentum corrections, and ensure an optimal reconstruction of the objects needed for the analysis. This task represented part of my

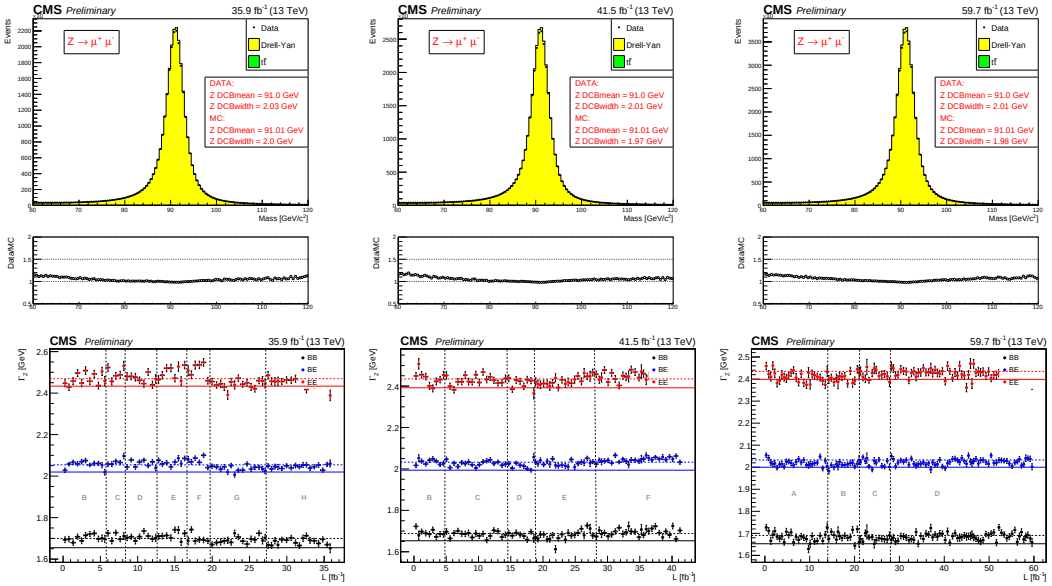


FIGURE 3.6: Muon energy scale and resolution (top) measured for all muons in the $Z \to \mu^+\mu^-$ control region in 2016 (left), 2017 (middle), and 2018 (right). The average measurement of the width (bottom) of the reconstructed Z bosons for each luminosity section (0.5 fb^{-1}) of the 2016, 2017, and 2018 running periods is also shown. Events presenting two muons in the barrel, two muons in the endcap, or one in the barrel and the other one in the endcap are denoted as BB, EE, and EB, respectively.

personal contribution to the analysis group. Fig. 3.6 (bottom) shows for example the average measurement of the width of the reconstructed Z bosons for each luminosity section (0.5 fb^{-1}) of the 2016, 2017, and 2018 LHC running periods.

3.2.5 Efficiency of muon reconstruction and identification

Once that reconstructed particle candidates are ready, an intermediate step is needed before using them in the search for a H boson signal. A careful study of the efficiencies in data and simulated samples for each object has to be performed exploiting dedicated control regions. In fact, the description of the detector response to signal muons may not be perfect in the simulation of the CMS subdetectors. Then, any observed difference between data and MC can be corrected by applying specific scale factors (SFs) to the simulated events.

The measurement of the muon efficiencies exploits the *tag-and-probe* (TnP) method [108] and is performed on $Z \to \mu^+\mu^-$ and $J/\Psi \to \mu^+\mu^-$ events in bins of p_T and η. The TnP is a generic method that uses narrow dilepton resonances to measure any user defined per-object efficiency from data (reconstruction, trigger, selection, ...). On the one hand, a Z sample is generally used to measure the muon reconstruction and identification efficiency at intermediate p_T. On the other hand, a J/Ψ sample is better to measure the reconstruction efficiency at low p_T, since it benefits from a better purity in that kinematic regime. A brief explanation of the approach is presented here:

* Events with dimuon decays from J/Ψ and Z resonances are collected triggering on a single muon. The triggering muon is used as *tag* and is normally a high quality muon, matched to a trigger object. It is required to pass tight identification and isolation criteria to improve the purity of the selected sample. Then, the *probe* is required to have same flavor and opposite charge and satisfy a loose identification. For example,

in the measurement of reconstruction and identification efficiencies, the probes are tracks reconstructed in the inner tracker.

* The *passing probes* are defined according to whatever is the efficiency under study. To measure the muon reconstruction and identification efficiency, the passing probes are the muons reconstructed as a global or tracker muon and satisfying the loose muon identification explained above.
* For each p_{T} and η bin, the invariant mass of the resonance decaying into "tag + passing probe" or "tag + failing probe" pairs is fit separately with a signal plus background model. The efficiency measurement is performed from the ratio of events related to the sum of passing and non-passing probes for the signal events.

The efficiency in data and simulation measured with J/Ψ events is shown in Fig. 3.7 for the 2018 dataset. Similar results are obtained for the 2016 and 2017 scenarios with plots reported in Appendix A, together with the product of all the data to simulation SFs for muon tracking, reconstruction, identification, impact parameter, and isolation requirements for all three years. The SFs have been found very close to one in all three years with deviations within 2%. The largest deviations are due to the SIP, which is very hard to model properly in the simulation, and happen for low p_{T} muons in the barrel region and in the very forward region $\eta > 2$.

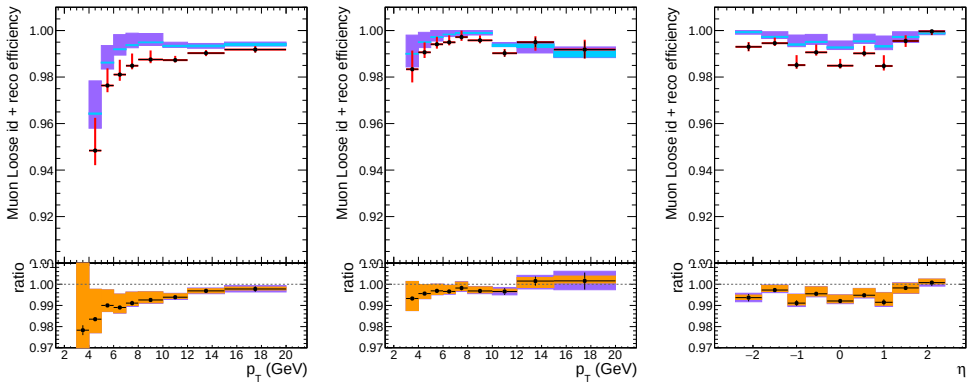


FIGURE 3.7: Muon reconstruction and identification efficiency at low p_{T} for the 2018 dataset. It is measured with the TnP method on J/Ψ events, as a function of p_{T} in the barrel (left) and endcap (center) regions, and as a function of η for $p_{\mathrm{T}} > 5$ GeV (right). In the upper panel, the black error bars on data are purely statistical while the red ones include also the systematical uncertainties. Similarly, the smaller azure rectangles represent the statistical uncertainty on MC while the larger orange rectangles include also the systematical contribution. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.

To conclude this part, some validation plots for the muon candidates selected after the full selection are shown as an example for the 2016 dataset in Fig. 3.8. Similar results have been produced extensively for all the muon variables in all three years.

3.3 Electrons

Electrons are reconstructed with high purity and efficiency in the CMS experiment, leaving a distinctive signal in the ECAL as an isolated energy deposit associated with hits in the

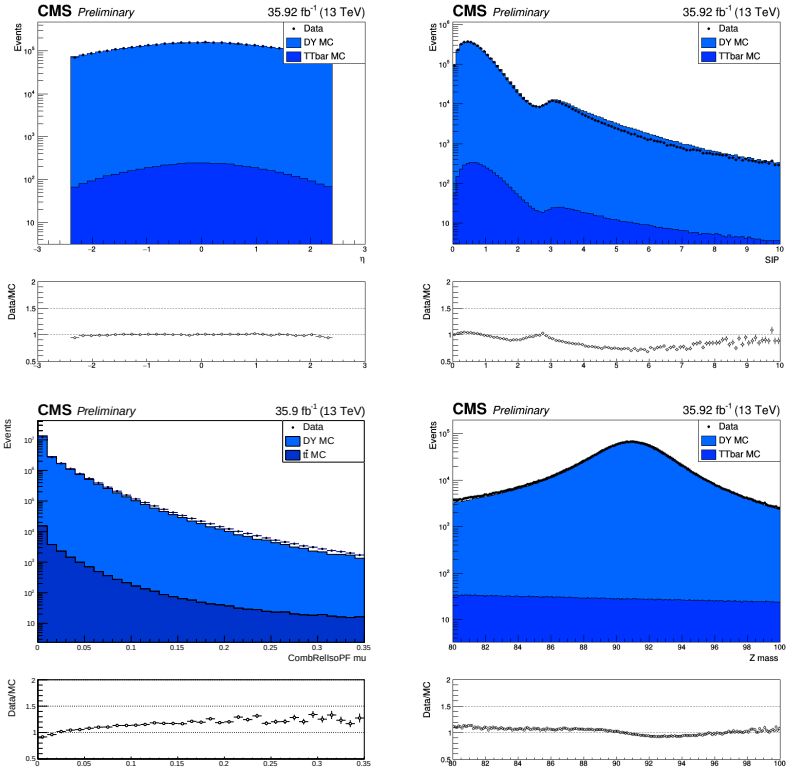


FIGURE 3.8: Distribution of the pseudorapidity (top left), SIP (top right), and isolation (bottom left) of the muon candidates with a p_{T} larger than 10 GeV selected after the full selection. Data/MC agreement is shown also for the invariant mass of the $\mathrm{Z} \to \mu^{+}\mu^{-}$ in the mass window $m_{\mu\mu} = [80, 100]$ GeV (bottom right).

tracker layers: their momentum is estimated by the combination of the energy measurement in the ECAL with the momentum measurement in the tracker. Given the excellent energy resolution of the ECAL, electrons represent an optimal ingredient to use both in precision measurements and in searches for BSM physics at the LHC. The energy resolution currently measured for electron is better than 2% (5%) in the barrel (endcap) region in p-p collisions for energies above 40 GeV and the energy scale in the same energy range is measured with an uncertainty lower than 0.1% (0.3%) in the barrel (endcap) acceptance.
The bremsstrahlung energy emitted by the electron as it propagates through the material in front of the ECAL is one of the main challenge in the electron reconstruction and measurement because of the amount of tracker material located between the collision point and the ECAL. The material budget in the tracker upstream to the ECAL before the upgrade of the silicon pixel detector (see Section 2.2.2) is shown in Fig. 3.9 (left). Indeed, by the time an electron reaches the electromagnetic calorimeter, it may easily interact with the material producing multiple electrons and photons. The electron emits bremsstrahlung photons and then the resulting photons possibly convert to electron pairs. Thus, a dedicated algorithm is needed to combine the energy deposits from the individual particles of the shower into a single object and recover the energy of the primary electron. Moreover, the emission of bremsstrahlung photons modifies the curvature of the electron trajectory in the tracker. Therefore, a dedicated tracking algorithm based on the Gaussian Sum Filter (GSF) [111] is required to estimate the tracker parameters. The strategy adopted is summarized in the next section. The complete reconstruction procedure used in Run 2 for p-p collisions at $\sqrt{\mathrm{s}} = 13$ TeV and all the details on reconstruction performances can be found in Ref. [112, 113].

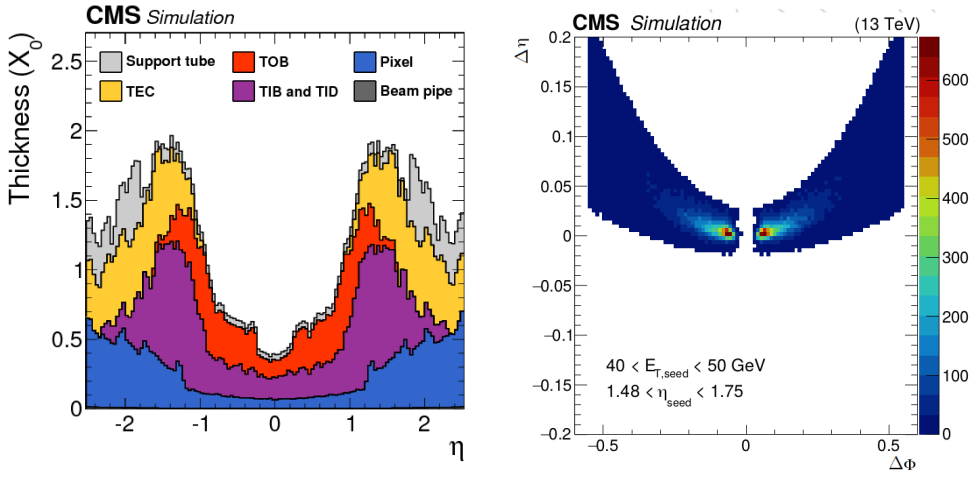


FIGURE 3.9: Material budget in the tracker upstream to the ECAL in units of radiation length X_0 as a function of particle pseudorapidity η in the $|\eta| < 2.5$ acceptance region (left), before the pixel detector upgrade [71]. Distribution of $\Delta\eta$ (seed-cluster, cluster) versus $\Delta\phi$ (seed-cluster, cluster) for simulated photons with a transverse energy $40 < E_{T,seed} < 50$ GeV; the z-axis represents the occupancy of the number of PF clusters (right) [113].

3.3.1 Electron reconstruction: superclustering and track association

Reconstruction of electron tracks, both for primary electrons and for photon conversions, is performed by searching for track seeds compatible with the ECAL superclusters (SC). The first step of the energy reconstruction algorithm consists in the formation of topological clusters by grouping together cells with energy above a given threshold starting from a seed cluster (see next paragraph). Then, the multiple ECAL clusters within a geometric area around the seed cluster are assembled into a single SC, the compatibility between SC position and the trajectory seeds in the pixel detector is checked, and the GSF tracking step is performed. Finally, the reconstruction algorithm based on ECAL superclusters is complemented by a PF-based one to differentiate between electrons and photons.

ECAL superclustering

Given that the electron energy spreads out over several crystals of the ECAL, as explained above, to clusterize energy deposits in several ECAL channels each local maximum above a certain energy threshold is assumed to correspond to a single particle reaching the detector. The seed cluster is identified as a local maximum of the energy deposited in any considered region and crystals with at least one side in common with a crystal already in the cluster are aggregated to it. A single ECAL energy deposit may be shared between overlapping clusters and a Gaussian profile of the shower is used to determine the fraction of the energy deposit to be assigned to each of the clusters. Therefore, the multiple ECAL clusters are combined into a single SC that collects the energy of the original electron so that bremsstrahlung losses and photon conversions are taken into account. The superclustering exploits the information on the η-ϕ correlation between the position of the electron and of bremsstrahlung clusters and the dependency on the transverse energy E_T. This approach allows to improve the energy determination of the supercluster and the rejection of clusters coming from pileup interactions close to the electromagnetic energy deposit, in particular for electrons with low p_T and large pseudorpidity, avoiding a bias in the energy collection and cluster shapes. Two different algorithms are used for the superclustering procedure:

* The **mustache algorithm** relies on the information only from the ECAL crystals and preshower detectors and it is used to properly measure low energy deposits. A seed cluster above a certain threshold is identified and additional clusters are merged in case they fall into a region in the $\Delta\eta$ (seed-cluster, cluster) versus $\Delta\phi$ (seed-cluster, cluster) space featuring a shape similar to a mustache, as shown in Fig. 3.9 (right). This particular slight bend in the $\Delta\eta$-$\Delta\phi$ plane is caused by the solenoidal structure of the CMS magnetic field, which tends to spread the radiated energy along the ϕ direction rather than along η. Consequently, the extension of this region depends on the transverse energy, because the curvature of particles with a higher $\mathrm{E_T}$ in the magnetic field is less significant. The superclusters reconstructed in this way are used to seed electrons, photons and conversion-finding algorithm.

* The **refined algorithm** exploits also the information from the tracking to extrapolate bremsstrahlung tangents and conversion tracks and determine when a cluster has to be merged in the SC. It can be applied as an independent method to recover additional conversion and bremsstrahlung clusters or dependently on the mustache SC, used as a starting point and refined using the information from other subdetectors. The risk of including spurious clusters is minimal. At the end of the procedure, a given ECAL cluster can belong to only one refined SC. The refined SC are used for the determination of all ECAL-based quantities of electron and photon objects.

Track reconstruction and association

To perform the tracking of electrons (both for primary electrons and for photon conversions), the GSF algorithm is used searching for a compatibility between track seeds in the pixel detector and the SC position during the pattern recognition stage. This allows to properly account for radiative losses from bremsstrahlung and large angle scattering in the material, providing also a measurement of the energy loss due to radiation by comparing the local curvatures of the trajectory at the two endpoints. First of all, a small hit patterns compatible with an electron trajectory is identified as a seed exploiting an "ECAL-driven" or "tracker-driven" approach. This seeding procedure is needed because the GSF tracking procedure can be very time consuming.

* **ECAL-driven approach** The better performance is on high p_T isolated electrons. Mustache SCs having a transverse energy $\mathrm{E_{SC,T}}$ > 4 GeV and a ratio between the energy deposits in HCAL and in ECAL (H/E) smaller than 0.15 within a cone of $\Delta R < 0.15$ centered on the SC are selected as starting point. Each of them is compared to a set of track seeds formed by the combination of pairs or triplets of hits in the inner tracker. The trajectory of the electron corresponding to a given SC is computed from the SC position, $\mathrm{E_{SC,T}}$ and the B field intensity, and is assumed to be helical.
* **Tracker-driven approach** It helps to recover efficiency for low p_T or nonisolated electrons located in the ECAL regions where the energy measurement is less precise. All general tracks are considered, referring to a collection of filtered tracks with p_T > 2 GeV. These tracks are reconstructed starting from hits in the pixel detector and applying an iterative algorithm known as the Kalman Filter (KF). If any of these KF tracks is associated to an ECAL cluster, its track seed is extrapolated toward the ECAL to match a SC and used to seed a GSF track. To identify the general tracks possibly originating from conversions, a dedicated algorithm is used.

By combining the ECAL-driven and tracker-driven seeds, a final collection of electron seeds is selected and used to initiate the electron track reconstruction. To build the track, an iterative procedure evaluates the track parameters at each tracker layer using the KF

algorithm and modelling the electron energy loss with a Bethe-Heitler function. Multiple candidate trajectories are created in case of multiple hits compatible with the predicted position in the successive layer and then the track χ^2 is evaluated. After having collected all the hits with the KF algorithm, a GSF fit is applied to estimate the tracks parameters at each layer; a mixture of Gaussian distributions is used to approximate the energy loss in each layer. Finally, in order to perform track-cluster associations, GSF tracks are extrapolated toward ECAL assuming a homogeneous magnetic field. The electron candidates are identified by associating the GSF tracks to the SC. The position of the SC is defined as the energy-weighted average of the constituent ECAL cluster positions. Then, a Boosted Decision Tree (BDT) classifier is used. It combines information about the track (kinematical variables and quality), the SC, and their association (geometric and kinematic observables) in order to take a decision about the association between the GSF track and an ECAL cluster. The final step is the estimation of the electron charge and momentum.

The electron reconstruction efficiency has been measured in 2017 data and in DY simulated samples (Fig. 3.10) together with the data to simulation SFs using the TnP technique.

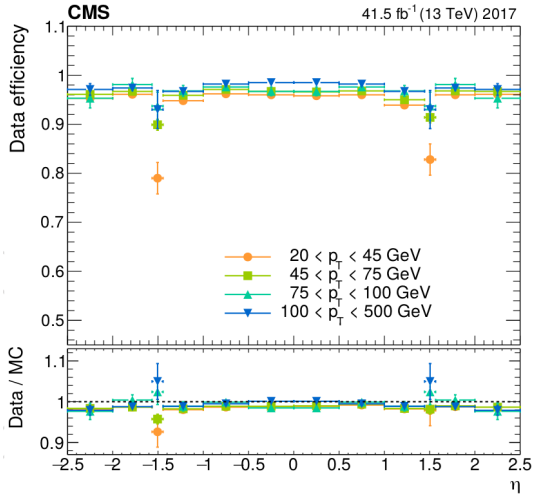


FIGURE 3.10: Electron reconstruction efficiency as a function of the pseudorapity in data (top) and the data-to-simulation ratios (bottom) using DY samples for the 2017 data-taking period, as provided by the CMS Electron and Gamma Physics Object Group (EGM POG) [113].

The ratios are very close to the unity almost in the entire range and are applied as a function of the p_{T} and η of the SC. The efficiency on the electron reconstruction is very high and it is comparable to the one obtained during Run 1, which is a good achievement considering the much harsher conditions related to the higher pileup of the Run 2 scenario. The average number of overlapping events during Run 1 was 9 (21) at 7 TeV (8 TeV) and it increased to 30 pileup interactions at 13 TeV, up to 40. The relative electron resolution evaluated in 2016 MC samples is shown as a function of the electron transverse momentum in Fig. 3.11.

3.3.2 Electron selection, identification and isolation

After the reconstruction step, some quality requirements are applied to identify genuine electrons coming from the H signal among all the objects that can be selected as electron candidates by mistake. The electron candidates preselection consists of loose cuts on the track-cluster matching observables in order to preserve high efficiency and, in parallel, to discard part of the QCD background. In the $\mathrm{H} \to 4\ell$ analysis, they are required to have a transverse momentum $p_T^e > 7$ GeV in the acceptance region $|\eta^e| < 2.5$: such electrons

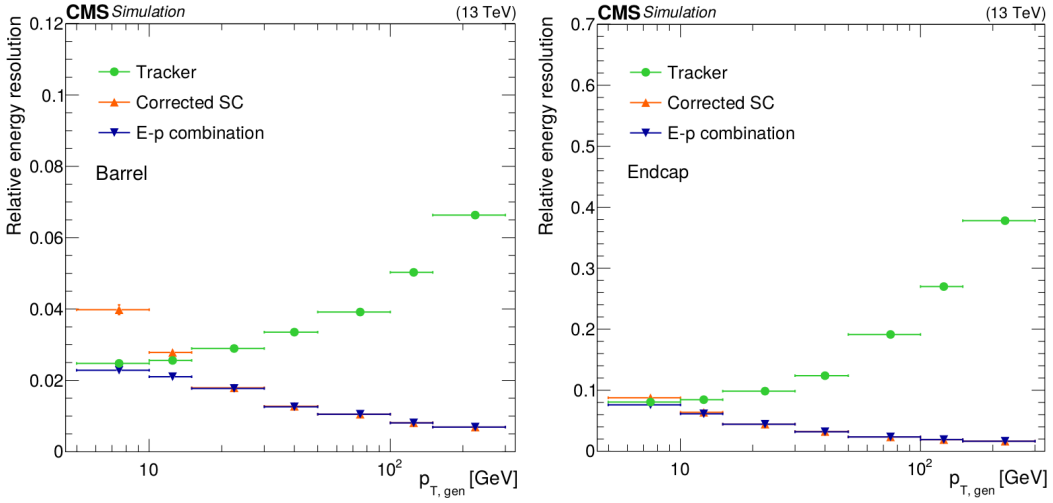


FIGURE 3.11: Relative electron resolution versus electron p_T found in 2016 MC samples for barrel (left) and endcaps (right) electrons. Curves measured by the ECAL (corrected SC) and by the tracker separately and given by the combination of ECAL and tracker information (E-p combination) are reported [113].

are called **loose electrons**. Similar to muons, they must satisfy a loose primary vertex constraint defined as d_{xy} < 0.5 cm and d_z < 1 cm and a 3D impact parameter significance with respect to the primary vertex smaller than 4 is required in order to suppress fake electrons originating from photon conversions.

Then, the offline-level electron identification relies on a broad range of observables condensed into a single multivariate discriminant. A multivariate classifier algorithm based on the XGBoost (e**X**treme **G**radient **Boost**ing) library [114], designed to be highly efficient, flexible and portable, and able to implement machine learning algorithms under the Gradient Boosting framework, is applied to identify isolated electrons, combining identification and isolation variables. The coherent approach used to select electrons in all data-taking periods represents one of the main improvements brought in the $H \to 4\ell$ analysis using the whole Run 2 dataset. The training exploits observables related to the electromagnetic cluster and the geometrical matching between the cluster and the electron track, but also observables based exclusively on tracking measurements as well as the PF isolation sums. Table 3.2 provides an overview of the main classes of observables and the specific variables used as inputs to the identification classifier.

The agreement between data and simulated samples in representing these quantities has been checked in all three years. Some distributions related to the isolation variables included in the electron MVA are shown as an example in Fig. 3.12 for the 2016 dataset. These plots are produced looking only at electrons from the Z_1 candidate in the control region Z+jets. Some disagreement is observed, especially considering the charged hadron component in the region at low isolation values. The most probable explanation is that we are not removing sufficiently the QCD background in data and the QCD contribution is missing in the simulated events.

The model is trained on 2016, 2017, and 2018 DY with jets MC samples for both signal and background, providing a separate training for each data-taking period in order to guarantee an optimal performance in the whole Run 2 dataset analyzed in this work. A matching to generated electrons from Z decays is required to select signal electrons, while reconstructed electrons which are not matched to any generated electron are treated as

Classes of oservables	Variable	Definition
Cluster shape	$\sigma_{i\eta i\eta}$, $\sigma_{i\varphi i\varphi}$	RMS of the energy-crystal number i spectrum along η and ϕ;
	η-ϕ width	SC width along η and ϕ
	H/E	Ratio of the hadronic energy behind the electron SC to the SC energy
	$(E_{5\times5} - E_{5\times1})/E_{5\times5}$	Circularity: $E_{5\times5}$ is the energy computed in the 5×5 block of crystals centered in the highest energy crystal of the cluster seed; $E_{5\times1}$ is the energy computed in the strip of crystals containing it
	R_9 = $E_{3\times3}/E_{\text{SC}}$	Sum of the seed and adjacent 3×3 crystals centered on the highest energy crystal divided by the SC energy
	E_{PS}/E_{raw}	Energy fraction in the preshower subdetector divided by the untransformed SC energy (for endcap training bins)
Track-cluster matching	E_{tot}/p_{in}, E_{ele}/p_{out}, $1/E_{tot} - 1/p_{in}$	Energy-momentum agreement
	$\Delta\eta_{in}$, $\Delta\varphi_{in}$, $\Delta\eta_{seed}$	Position matching
Tracking	$f_{brem} = 1 - p_{out}/p_{in}$	Fractional momentum loss
	χ^2_{KF}, χ^2_{GSF}	Reduced χ^2 of the KF and GSF track
	N_{KF}, N_{GSF}	Number of hits of the KF and GSF track
	$N_{miss.hits}$	Number of expected but missing inner hits
	P_{conv}	Probability transform of conversion vertex fit χ^2
Isolation	$\mathcal{I}_{\gamma}$	PF photon isolation sum
	$\mathcal{I}_{\text{ch. had.}}$	PF charged hadrons isolation sum
	$\mathcal{I}_{\text{neu. had.}}$	PF neutral hadrons isolation sum
For PU-resilience	ρ	Mean energy density in the event

TABLE 3.2: Overview of the input variables to the electron identification classifier.

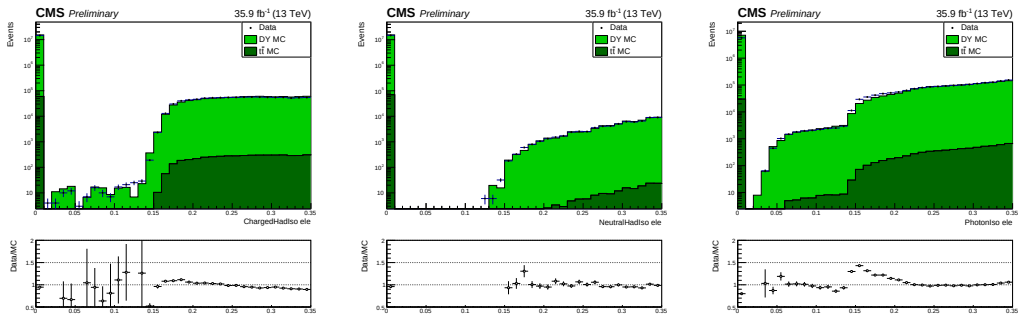


FIGURE 3.12: Data/MC agreement for the isolation variables in 2016 data included in the BDT for the electron identification.

background. To improve the performance on the low energy electron identification, the training is splitted in six bins of p_{T} and η considering separately electrons with a transverse momentum smaller/larger than 10 GeV in three different pseudorapidity regions: $|\eta^e| < 0.8$, $0.8 < |\eta^e| < 1.479$, and $|\eta^e| > 1.479$.

Fig. 3.13 shows the output of the multiclassifier discriminant, i.e. the MVA score for prompt electrons and misidentified electrons originating from jets in DY events, provided by the 2018 model. The performance of the model trained on 2018 MC using electron identification and isolation features is shown as an example in Fig. 3.14.

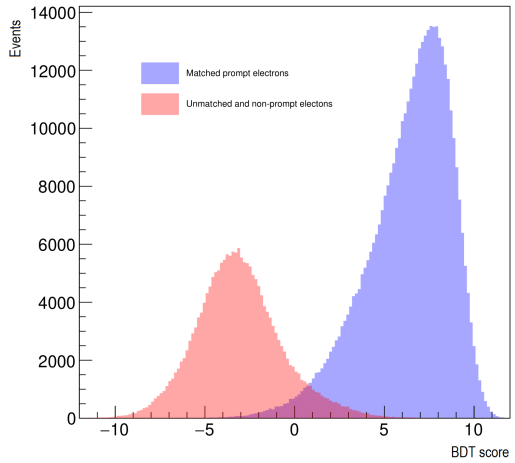


FIGURE 3.13: Output of the multiclassifier discriminant for prompt electrons matched to truth electrons from Z decay (blue) and for fakes (red) taken from a 2018 simulated sample of DY events.

The working points shown are chosen to achieve a similar level of signal efficiency as in Run 1. To be selected as final objects for the analysis, loose electrons have to pass MVA identification and isolation working points listed separately for each training (2016, 2017, and 2018) in Table 3.3.
The corresponding signal and background efficiencies are similar in all three years: larger than 95% for electrons with a $p_{\mathrm{T}} > 10$ GeV and around 80% (70%) for low p_{T} electrons in the barrel (endcap) region.

2016 Datasets			
minimum BDT score	$\lvert\eta^e\rvert < 0.8$	$0.8 < \lvert\eta^e\rvert < 1.479$	$\lvert\eta^e\rvert > 1.479$
$5 < p_T < 10$ GeV	0.9503	0.9461	0.9387
$p_T > 10$ GeV	0.3782	0.3587	-0.5745
2017 Datasets			
minimum BDT score	$\lvert\eta^e\rvert < 0.8$	$0.8 < \lvert\eta^e\rvert < 1.479$	$\lvert\eta^e\rvert > 1.479$
$5 < p_T < 10$ GeV	0.8521	0.8268	0.8694
$p_T > 10$ GeV	0.9825	0.9692	0.7935
2018 Datasets			
minimum BDT score	$\lvert\eta^e\rvert < 0.8$	$0.8 < \lvert\eta^e\rvert < 1.479$	$\lvert\eta^e\rvert > 1.479$
$5 < p_T < 10$ GeV	0.8956	0.9111	0.9401
$p_T > 10$ GeV	0.0424	0.0047	-0.6042

TABLE 3.3: Minimum BDT score required for passing the electron identification considering separately the 2016, 2017, and 2018 trainings.

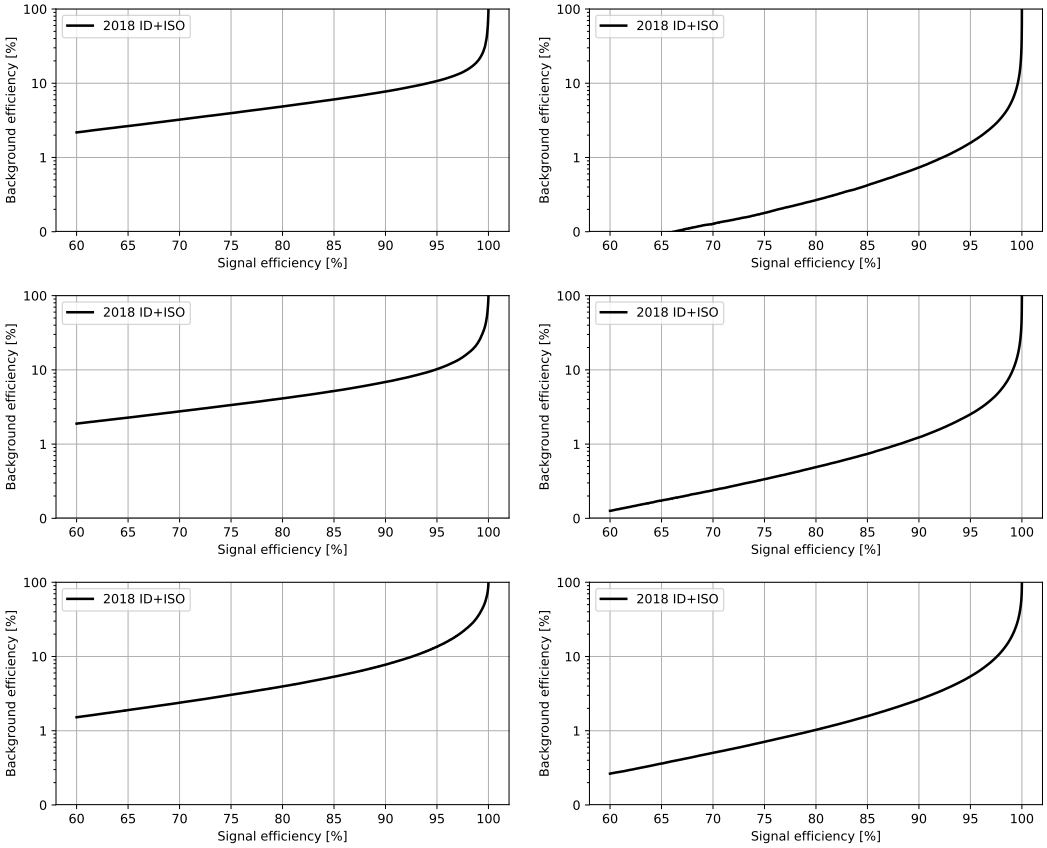


FIGURE 3.14: The receiver operating characteristic curves (background efficiency vs signal efficiency) of the MVA developed for the electron identification in 2018 data, trained on the 2018 simulated samples. Performance are shown for electrons with $5 < p_{\mathrm{T}}^{e} < 10$ GeV (left) and $p_{\mathrm{T}}^{e} > 10$ GeV (right) in three different η^{e} regions: $|\eta^{e}| < 0.8$ (top), $0.8 < |\eta^{e}| < 1.479$ (middle), and $|\eta^{e}| > 1.479$ (bottom).

3.3.3 Energy calibrations for electrons

The energy measurement provided by the electromagnetic calorimeter dominates the combination of the tracker and ECAL information used to measure electrons with a transverse momentum above 20 GeV, especially if they undergo large energy losses in the tracker material. Moreover, it is used to determine the energies of reconstructed photons. The energy calibration at the level of individual detector channels exploits the azimuthal symmetry of the energy flow, with corrections for the displacement of the beam spot in the transverse plane and photons from π^0 decays. The photon pairs from π^0 decays also provide a first absolute energy scale calibration; then, this first calibration is also validated using electrons from $W \to e\nu$ and $\mathrm{Z} \to e^{+}e^{-}$.

Electrons in data are corrected in bins of p_{T} and η according to the ECAL energy scale. These corrections, calculated using a $\mathrm{Z} \to e^{+}e^{-}$ sample, aim to align the dielectron mass spectrum observed in data to the one obtained in the MC sample and to minimize its width. The $\mathrm{Z} \to e^{+}e^{-}$ mass resolution in MC is made to match data by applying a pseudorandom Gaussian smearing to electron energies, with Gaussian parameters varying in bins of p_{T} and η, convoluting the electron energy spectrum with a Gaussian. Finally, the electron energy scale is measured in data by fitting a Crystal Ball function to the dielectron mass spectrum

around the Z peak in the Z $\rightarrow e^+e^-$ control region. The energy scale and resolution for the 2016, 2017, and 2018 dataset is shown in Fig. 3.15 (top) together with the time evolution of the Z boson mass measured across different runs of the same data period (bottom). A clear shift in the scale is visible in the 2018 Z mass distribution, where the right tale raises for mass values larger than 95 GeV. Those effects can come from crystal transparency loss in ECAL, but also from the noise and pedestal in electronics that is not constant in time. This effect can be corrected in the "Ultra Legacy (UL) ReReco" re-reconstructed dataset that is expected to arrive by the first half of 2021.

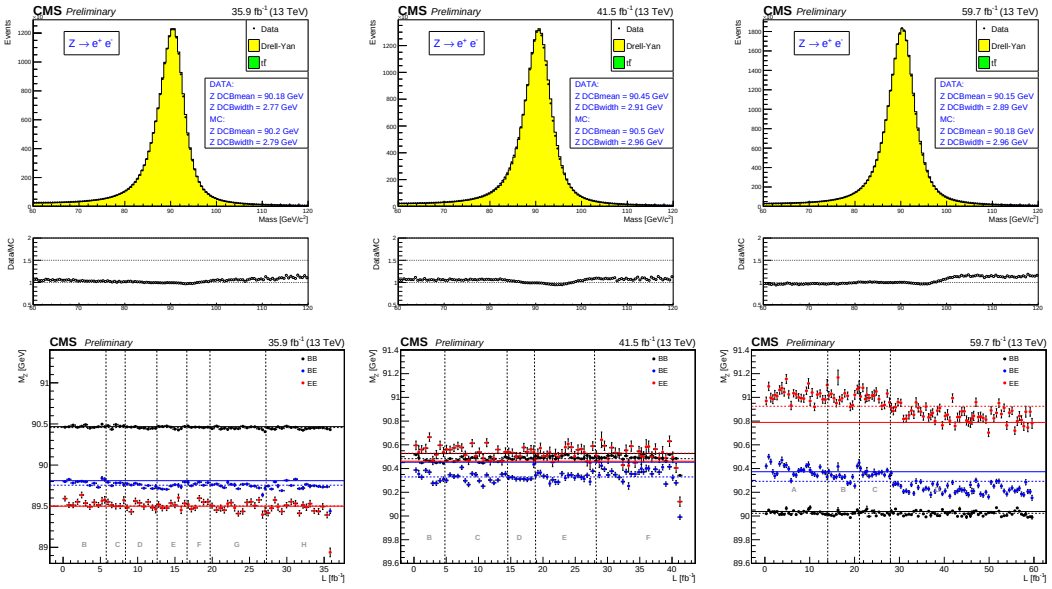


FIGURE 3.15: Electron energy scale and resolution (top) measured for all electrons in the Z $\rightarrow e^+e^-$ control region in 2016 (left), 2017 (middle), and 2018 (right). The average measurement of the mass of the reconstructed Z bosons for each luminosity section (0.5 fb^{-1}) of the 2016, 2017, and 2018 running period is also shown (bottom). Events presenting two electrons in the barrel, two electrons in the endcap, or one in the barrel and the other one in the endcap are denoted as BB, EE, and EB, respectively.

3.3.4 Electron efficiency measurements

Electrons suffer from intrinsically higher background rates in the low p_{T} range compared to muons. The main reason is related to the fact that J/Ψ samples were not available. Consequently, the number of signal events in the failing probes were not sufficient for the measurement and the background rates were very high. Therefore, a more complex selection procedure is needed to extract the final SFs with respect to the muon case.
Some quality requirements on the tag electrons are needed: a single electron trigger is required, the transverse momentum has to be larger than 30 GeV and $\eta_{\mathrm{SC}} < 2.17$. In addition, the tag and the probe need to have opposite charge. Considering the first electron bin in the range 7-20 GeV, more stringent criteria are required to clean the background and make the fits more reliable. The tag p_{T} cut is increased to 50 GeV and the charge is determined with the so-called *selection method*, asking all three estimates of the electron charge to agree. Finally, the tag has to pass a cut on the MVA ID > 0.92 and $\sqrt{2 \cdot \mathrm{MET}_{\mathrm{PF}} \cdot p_{\mathrm{T}}(tag) \cdot (1 - cos(\phi_{\mathrm{MET}} - \phi_{tag}))} < 45$ GeV. Probe electrons only need to be reconstructed as GSF electrons.

For the efficiency measurements a template fit is used. The m_{ee} signal shape of the passing and failing probes is taken from MC and convoluted with a Gaussian. The data is then fitted with the convoluted MC template and an error-function with a one-sided exponential tail. For the low p_{T} bins, a Gaussian is added to the signal model for the failing probes. The electron selection efficiency is measured as a function of the probe electron p_{T} and η_{SC}, considering separately electrons falling in the ECAL gaps (see Section 2.2.2). The p_{T} and η efficiencies measured in 2017 data are shown in Fig. 3.16, while similar results for 2016 and 2018 data are presented in Appendix A.

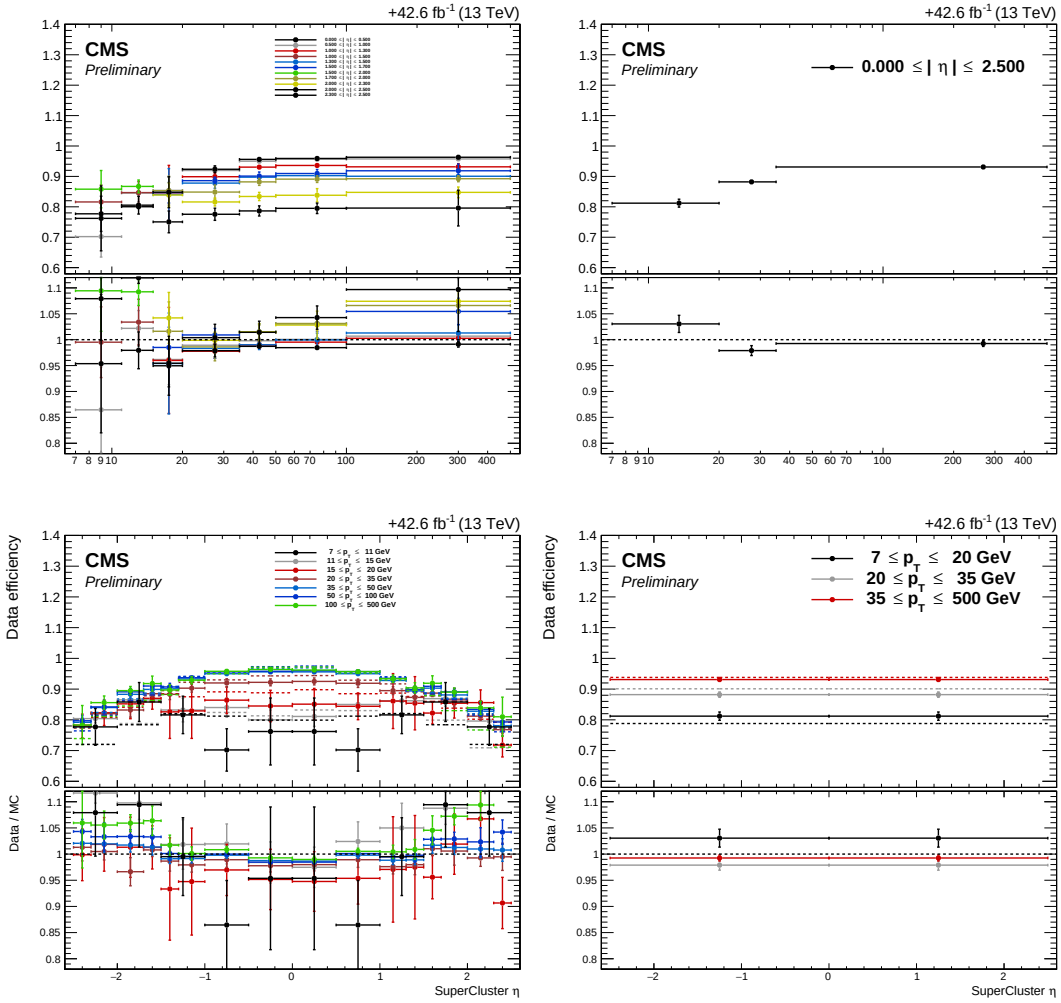


FIGURE 3.16: Electron selection efficiencies as a function of p_{T} (top) and η (bottom) measured in 2017 data using the TnP technique for non-gap electrons (left) and gap electrons (right), together with the corresponding data/MC ratio at the bottom of each plot.

Electron efficiencies are found to be around 95% for electrons with $p_{\mathrm{T}} > 30$ GeV and are similar in all three years. The uncertainties on the measurement of low p_{T} electrons are very large because of the poor statistics available. The total uncertainty for the measurement of the SFs is the quadratic sum of the statistical uncertainties returned from the fit and the previously mentioned systematic uncertainties. The large uncertainty on the SF measurement is the largest uncertainty source in our measurement. As mentioned above, it is mostly statistical and it comes from the low p_{T} bins where we have not enough events

to perform the fit in the failing probes. This aspect has to be fixed in view of Run 3. One possible solution is to verify if J/Ψ events can be found in our minimum bias samples or $Z \to 4\ell$ events are enough for the low p_T measurement. Otherwise, the development of dedicated trigger to collect those can be considered.

As a conclusion of this section, some validation plots for the electron candidates selected after the full selection are shown as an example for the 2016 dataset in Fig. 3.17. Similar results have been produced extensively for all the electron variables in all three years.

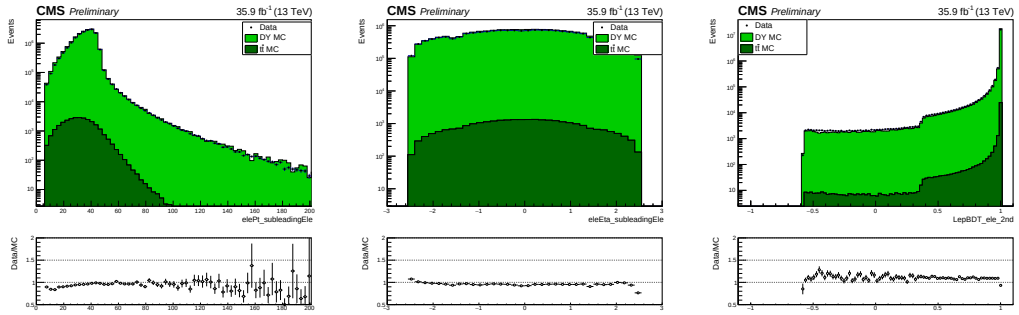


FIGURE 3.17: Distribution of the transverse momentum (left), pseudorapidity (middle), and BDT score (right) of the subleading electrons selected after the full selection in the 2016 dataset.

3.4 Photons: the FSR photon recovery

After that electrons and muons have been presented, a brief explanation of the treatment of photons in the $H \to ZZ^* \to 4\ell$ is needed because of a phenomenon named final-state radiation (FSR). In fact, the probability that a high energy photon is radiated in a $Z \to \ell^+\ell^-$ decay is around 8% (15%) considering dimuon (dielectron) decays and a p_T^γ threshold of 2 GeV. Consequently, the four-momentum of the photons have to be included in the computation of the H boson candidate invariant mass to avoid a degradation of the accuracy in the reconstruction of the four-lepton kinematics. Moreover, the removal of FSR photons from the isolation cone of the muons allows to improve the selection efficiency preserving them from failing the isolation requirement.

In order to identify genuine FSR photons with respect to other photons coming from initial-state radiation (ISR), pileup interactions, and π^0 decays, a FSR recovery algorithm is applied exploting the kinematics of the FSR photons, expected to be isolated from other particles and collinear with the emitting lepton. The selection of FSR photons is performed per-lepton looking at the angular distance between the lepton and the photon over the transverse energy of the photon squared ($\Delta R(\gamma, \ell)/E^2_{T,\gamma}$) and does not depend on any Z mass criterion. Starting from the collection of PF photons, they are preselected requiring $p_T^\gamma > 2$ GeV, $|\eta^\gamma| < 2.4$, and a relative PF isolation smaller than 1.8, defined as

$$\mathcal{I}^\gamma \equiv \frac{1}{p_T^\gamma}\left(\sum_{\text{photons}} p_T + \sum_{\text{neu. had.}} p_T + \sum_{\text{ch. had.}} p_T\right). \tag{3.6}$$

To compute the isolation, a cone of radius $R = 0.3$ is used, applying a threshold of 0.2 GeV on charged hadrons with a veto cone of 0.0001 and 0.5 GeV on neutral hadrons and photons with a veto cone of 0.01. Moreover, the contribution from pileup vertices (with the same radius and threshold as per charged isolation) is taken into account.
After that, all photon candidates associated with any electron passing kinematic and impact

parameter cuts are removed in order to avoid double counting with the electron supercluster (supercluster veto); the matching is performed by directly associating the two PF candidates. Then, considering all the leptons passing both the loose ID and SIP cuts, photons are associated to the closest lepton in the event and are discarded when $\Delta R(\gamma, \ell)/E^2_{T,\gamma} < 0.012$ and $\Delta R(\gamma, \ell) < 0.5$ cuts are not satisfied. In case more than one photon is associated to the same lepton, the one with the lowest $\Delta R(\gamma, \ell)/E^2_{T,\gamma}$ is selected. Finally, each selected FSR photon is excluded from the calculation of the isolation sum of all the good muons in the event, considering photons which are in the isolation cone ($\Delta R < 0.4$) and outside the isolation veto of muons leptons ($\Delta R > 0.01$).

From a study performed using signal gluon fusion MC samples, it is observed that the FSR algorithm affects about 4.5% of all signal events in our analysis. The percentages for each final state in each year are presented in Table 3.4.

% ev. with FSR	ggH ($m_{4\ell}$ = [105,140] GeV)		
Final state	2016	2017	2018
4μ	6.3%	3.9%	4.9%
$4e$	5.3%	3.3%	4.2%
$2e2\mu$	5.2%	3.1%	4.2%

TABLE 3.4: Percentage of events with FSR in the $4e$, 4μ, and $2e2\mu$ signal events using ggH MC samples and considering the mass range used in the statistical analysis $m_{4\ell}$ = [105,140] GeV.

Given that usually the energy of the photon is already included in the ECAL energy measurement in the case of electrons, the effect is more relevant for muons than for electrons. The comparison of the four-lepton invariant mass for events affected by FSR before and after the FSR correction is shown for all three years in Fig. 3.18.

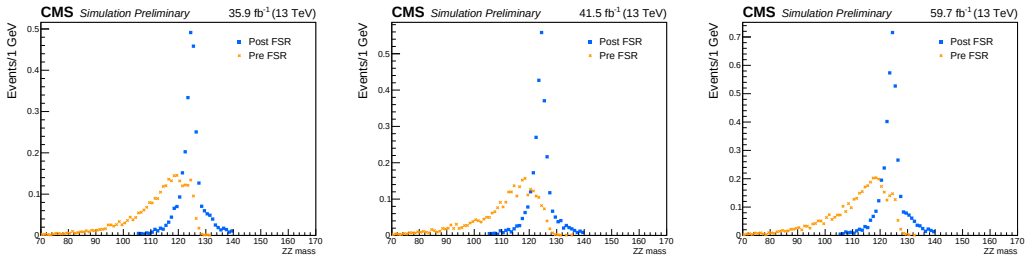


FIGURE 3.18: Comparison of the invariant mass of the four-lepton system for events affected by FSR before and after the FSR correction considering the reduced mass range $m_{4\ell}$ = [105,140] GeV in 2016 (left), 2017 (middle), and 2018 (right) ggH simulated samples.

3.5 Jets

Jet reconstruction, identification and classification is of prime interest in a hadronic environment such as the LHC. Although jets are not part of the decay of the H boson in the golden channel, they play a crucial role in the correct categorization of different production processes allowing to study each of them in detail. Indeed, the main difference between VBF and other production mechanisms is related to the jet kinematics. Moreover, jets are

very important also in the context of differential cross section studies, which represent one of the key measurements in the search for BSM hints in the Higgs sector.
Given that the scope of the analysis presented in this thesis is focused on the study of the H boson properties in its different production modes and different phase spaces, a good knowledge and modelling of these objects was particularly important and I spent a large effort to define the final selection of jets and to investigate the effect of some well-known data quality issues related to them. For this reason, in the current section the approach to reconstruct and identify jets is presented, together with the validation of the agreement between jets in data and in MC samples.

3.5.1 Jet reconstruction and corrections

The large amount of quarks and gluons produced in high energy processes is not directly detectable because of the fragmentation and hadronization of the particles which produce collimated showers of objects. In order to measure the kinematics of the initial parton, an algorithm to recollect and clusterize the hadronization products is needed. Particle flow candidates are combined in a jet according to the anti-k_t algorithm [115, 116], a sequential recombination algorithm which hierarchically clusters objects starting from the pair with the smallest distance in the event according to a metric:

$$d_{ij} = min\left(p_T^{-2}(i), p_T^{-2}(j)\right)\frac{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}{R^2}. \tag{3.7}$$

The final jet is obtained when the p_T^{-2} of the cluster is smaller than any remaining distance d_{ij}. The size of the jet conic shape is determined by a distance parameter R, fixed at 0.4.

The jet momentum is determined as the vectorial sum of the clustered PF candidates four momenta in the jet and is found from simulation to be within 5-10% of the true momentum over the whole p_{T} spectrum and detector acceptance. Additional p-p interactions within the same or nearby bunch crossings can contribute with additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, tracks identified as coming from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions. A good understanding of the jet energy scale and jet energy resolution is crucial because they are usually an important component of the systematic uncertainties. The calibration of the jet response is achieved applying a set of corrections which use the information of generated particles in a simulation. Jet energy corrections (JECs) are derived from simulation to bring the measured response of jets to that of particle-level jets on average and ensure the proper mapping of the measured jet energy deposition to the particle-level jet energy, taking into account the contribution from pileup in the event and nonlinearities in the detector response to hadrons. In situ measurements of the momentum balance in dijet, γ+jets, Z+jets, and multijet events are used to account for any residual differences in jet energy scale in data and simulation [117]. The jet energy scale (JES) is calibrated sequentially with pileup offset subtraction, detector response correction from simulation, residual corrections for differences between data and detector simulation, and optional corrections for jet flavor composition. In this analysis, the standard CMS JECs are applied to the reconstructed jets for both simulated samples and data.

* **L1 Offset/FastJet corrections** are responsible for the removal of energy coming from pileup events and of electronic noise. The pileup offset corrections are p_{T}-dependent and are determined from a simulated sample of QCD dijet events processed with and without pileup overlay.

* **L2 Relative and L3 Absolute corrections** are derived as a function of jet pseudorapidity and transverse momentum and make the response uniform over these two variables.
* **L2 and L3 residuals** are meant to correct for possible remaining differences between data and detector simulation (of the percent order) within the jet response in data and MC.

Then, jet energies in simulation are smeared to match the resolution in data. The jet energy resolution (JER) amounts typically to 16% at 30 GeV, 8% at 100 GeV, and 4% at 1 TeV [118]. As an example, the jet energy resolution as a function of the particle-level jet p_{T} at CMS is shown in Fig. 3.19. The most recent versions of JECs and JER have been tested in detail in our control regions and are currently applied in the analysis.

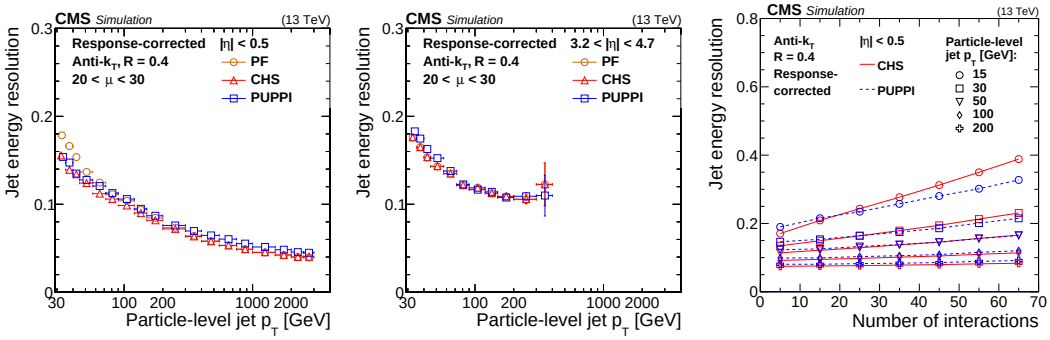


FIGURE 3.19: Jet energy resolution as a function of the particle-level jet p_{T} for different kind of jets in the barrel (left) and endcap (middle) region, measured in QCD multijet simulation at a center-of-mass energy of 13 TeV. JER is also shown as a function of the number of vertices in the event (right) [119].

3.5.2 Jet identification and pileup jet identification

The corrected jets considered in the analysis for categorization purpose are required to be within $|\eta^{\mathrm{jet}}| < 4.7$ area and to have a transverse momentum above 30 GeV. To reduce the possible instrumental background, a **tight identification** criteria is imposed. The efficiency is found to be larger than 98-99% for all pseudorapidity regions, whereas the background rejection is more than 98% for $|\eta| < 3$. Moreover, a pileup jet identification (PU jet ID) [118–120] is applied to jets with a transverse momentum below 50 GeV, to deal with a high number of pileup jets. It relies on three kinds of properties of the jets:

* **the trajectories of tracks** associated to the jets inside the tracker acceptance are used to check the compatibility of the jet with the primary interaction vertex;
* **the jet shape topology** helps to disentangle jets arising from the overlap of multiple interactions from truly hard jets;
* **the object multiplicity** is used as an additional handle.

A BDT is developed to integrate the information of all the discriminating variables able to quantify such properties and it is trained with 2016 samples. Then, three working points are defined: loose, medium, and tight. Efficiencies of the MVA PU jet ID are shown in Fig. 3.20 for a loose working point. A tight working point is chosen in the $\mathrm{H} \rightarrow 4\ell$ analysis to improve the purity of the event categorization. Finally, the jets are cleaned from any of the tight leptons which pass the SIP and isolation cut computed after FSR correction and from

FSR photons where a minimal angular distance between all selected lepton candidates or any selected FSR photons and the considered jet is required: $\Delta R(\ell/\gamma, \text{jet}) > 0.4$.

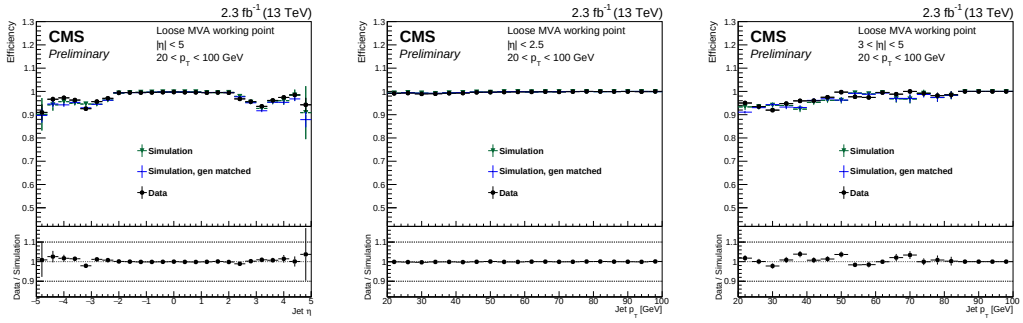


FIGURE 3.20: Comparison between data and MC of the MVA PU jet ID efficiency on the Z($\rightarrow \mu\mu$)+jets sample for PF jets with p_{T} >20 GeV using the loose working point. The efficiency is shown as a function of the jet pseudorapidity (left) and as a function of the transverse momentum for jets with $|\eta|$<2.5 (middle) and 3<$|\eta|$<5 (right) [118].

3.5.3 Additional criteria on jets

Additional selection criteria are applied to remove jets potentially dominated by anomalous contributions from various subdetector components or reconstruction failures related to some issues during the data-taking. In principle, a complete re-reconstruction of the data is needed in order to fully fix these well-known problems with jets during Run 2. This is expected to be solved in the UL dataset. However, to preserve the quality of the jet reconstruction and ensure high quality results with the currently available dataset, the possibilty of adding some criteria on the jets to cope with these issues was studied in detail.

L1 Prefiring In 2016 and 2017, the gradual timing shift of ECAL was not properly propagated to L1 trigger primitives (TP) resulting in a significant fraction of high η TP being mistakenly associated to the previous bunch crossing. Given that L1 rules forbid two consecutive bunch crossings to fire, an unpleasant consequence is that events can self veto if a significant amount of ECAL energy is found in the region between $2 < |\eta^{\text{jet}}| < 3$. This effect present in data is not described by the simulation. Thereby, a weight to take into account this effect is computed for each event and applied to the simulation in 2016 and 2017 samples. Fig. 3.21 shows the impact of the L1 prefiring weights on the 2017 signal samples considering ggH and VBF samples. As expected, the effect is more significant in the VBF production mode, in particular in the endcap region (at the level of 2-3%), while it is minor for the ggH signal.

HEM 15/16 failures In 2018, a part of the HCAL detector was not working for a significant part of the data-taking. Following a CMS-wide power interlock on 30$^{\text{th}}$ June, the power-on of CAEN A3100HBP modules that provide low voltage power to the on-detector HE front-end electronics led to irreversible damage of two sectors on the HE minus side, HEM15 and HEM16. As a consequence, extra electrons, jets, and photons in the problematic HEM 15/16 region between -2.4 < η < -1.4 and -1.6 < ϕ < -0.6 can be present in the events. The overall effect on our signal and control regions in the mentioned η-ϕ space, shown in Fig. 3.22, is found not statistically significant. It is well covered by the selection efficiency

uncertainties in our signal region, so that a specific strategy was not developed to cope with this.

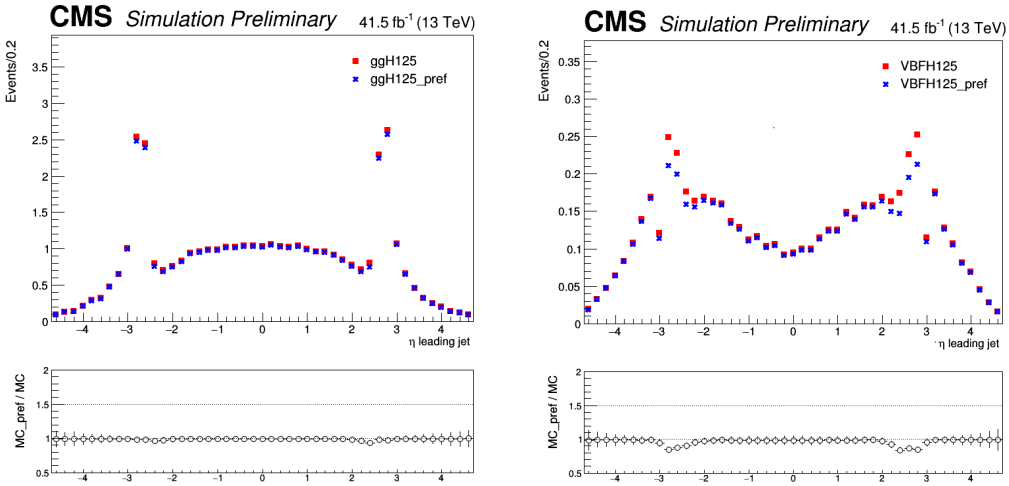


FIGURE 3.21: Comparison between 2017 MC samples with (blue) and without (red) L1 prefiring weights for ggH (left) and VBF (right) signals. The ratio is shown at the bottom of each plot.

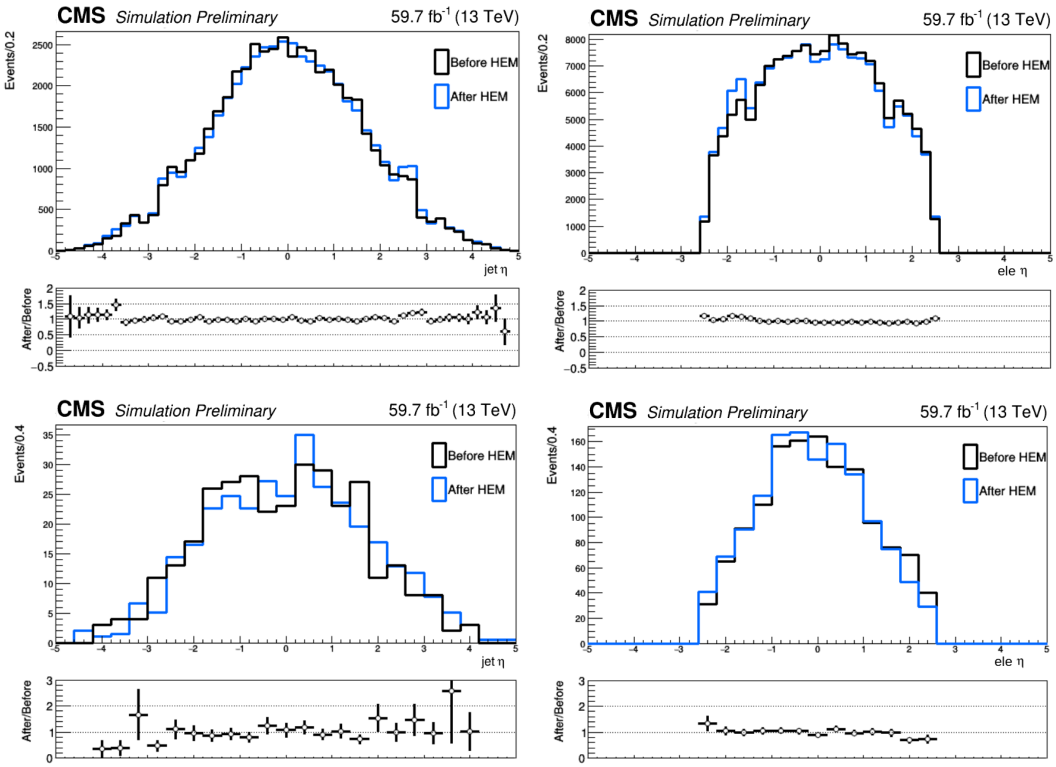


FIGURE 3.22: Comparison between the η distribution of jets (left) and electrons (right) in the control region Z+2ℓ (top) and in our signal region (bottom) before (black) and after (blue) HEM problem. The ratio between the events before and after the HEM issue is shown: no significant increase of jets and electrons in the region between -2.4 < η < -1.4 and -1.6 < ϕ < -0.6 is observed.

Removal of noisy jets A larger jet multiplicity was reported for 2017 data, related to an increase of the ECAL noise (PU and bunch crossing dependent), getting worse as luminosity increases. This issue causes a weird structure similar to "horns" in the η^{jet} distribution in the forward region $2.5 < \eta^{\text{jet}} < 3$. Given that the problem can be fixed only in the next reprocessing of data in the UL ReReco, the impact of rejecting jets with a transverse momentum before any correction $p_{\text{T}}^{\text{jet}}(\text{raw}) < 50$ GeV in the region $2.65 < \eta^{\text{jet}} < 3.139$ was checked. As preliminary studies showed no significant improvement in the data/MC agreement, it was decided not to apply any additional selection criteria to jets in order to avoid the rejection of events possibly targeting the VBF topology. The result of the study is shown in Fig. 3.23.

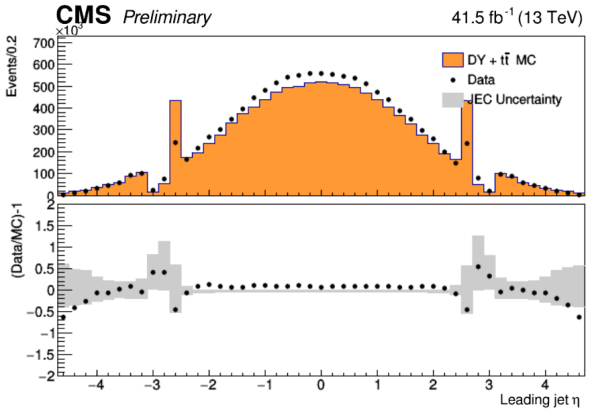


FIGURE 3.23: Comparison between data and MC of the leading jet η distribution discarding jets with a raw p_{T} smaller than 50 GeV in the region $2.65 < \eta^{\text{jet}} < 3.139$.

The effect of varying the jet ID and jet pileup ID criteria has been also checked. The study shown in Fig. 3.24 demonstrates that the joint application of both jet ID and jet pileup ID with a tight working point is particularly effective to reduce the horn structure in the forward η region both in data and MC.

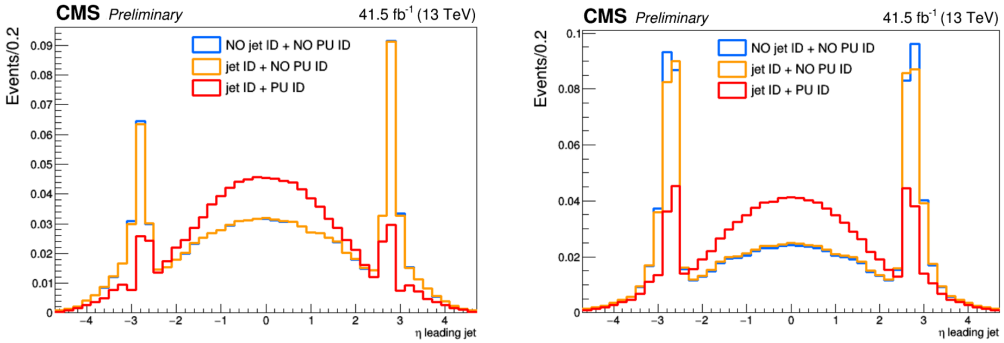


FIGURE 3.24: Leading jet η distribution in data (left) and MC (right) considering three differen scenarios: without the application of the tight jet ID and the tight PU jet ID (blue), applying only the tight jet ID (orange), an applying both of them (red).

3.5.4 B tagging

Given that the event categorization relies on the knowledge about the jet flavor, jets are tagged as b jets using the DeepCSV algorithm [121], which, in particular, combines information about impact parameter significance, the secondary vertex, and jet kinematics using information of tracks. The b tag output discriminator is computed with a Deep Neural Network (DNN). Data to simulation SFs for the b tagging efficiency are shown in Fig. 3.25 for all three years. They are applied as a function of $p_{\mathrm{T}}^{\mathrm{jet}}$, η^{jet}, and flavor to simulated jets by downgrading (upgrading) the b tagging status of a fraction of the b-tagged (untagged) jets that have a SF smaller (larger) than one.

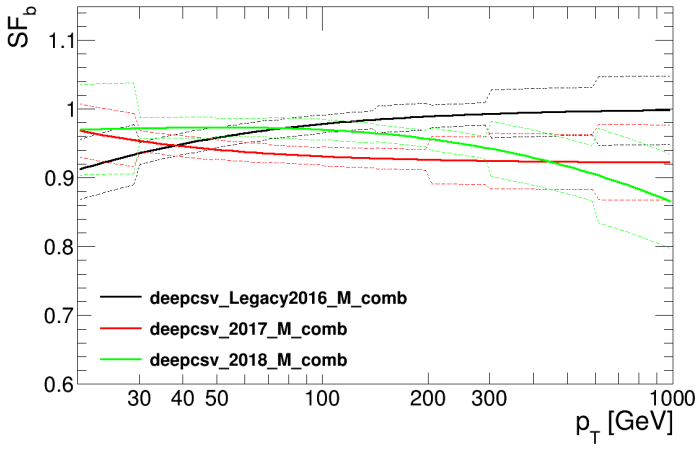


FIGURE 3.25: Scale factors for the DeepCSV b tagging algorithm using a medium working point evaluated in all three years (left): 2016 (black), 2017 (red), and 2018 (green).

In the $\mathrm{H} \rightarrow \mathrm{ZZ}^{*} \rightarrow 4\ell$ analysis, a jet is identified as b-tagged jet if it passes a medium working point. The distribution of the b tagging score in the Z+jets CR is shown as an example in Fig. 3.26 for data and MC samples.

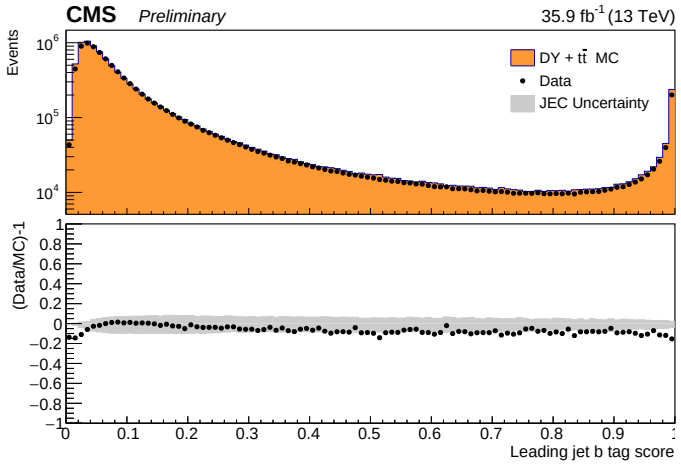


FIGURE 3.26: The distribution of the b tagging score of the leading jet in the Z+jets CR for data and MC samples (DY and $\mathrm{t\bar{t}}$).

3.5.5 Validation studies on jets

Extensive validation studies to probe the agreement between data and simulated samples for jets with $p_{\mathrm{T}}^{\mathrm{jet}} > 30$ GeV and $|\eta^{\mathrm{jet}}| < 4.7$ after the full selection have been performed. Events containing $\mathrm{Z} \to e^+e^-$ or $\mathrm{Z} \to \mu^+\mu^-$ plus additional jets are considered in order to compare basic properties of jets in data events and simulated events of DY and $\mathrm{t\bar{t}}$ processes. Fig. 3.27 and Fig. 3.28 show the pseudorapidity and transverse momentum of the leading jet and the jet multiplicity in data and MC for the three data-taking periods. The uncertainties coming from the combination of all contributions to the energy corrections are also displayed on the data/MC ratio plots. In the end, the agreement between data and MC samples is found to be good and covered by the large uncertainties in the forward region, allowing us to determine H boson production properties.

3.6 Conclusion

In this chapter all the ingredients needed to identify the experimental signature of the signal processes under study have been presented. A deep knowledge of the objects reconstruction and selection and of the corrections needed to improve their description is crucial to guarantee high quality performances of the H boson reconstruction. Since the beginning of my work in the $\mathrm{H} \to 4\ell$ analysis, a part of my effort has been dedicated to a detailed study of the objects in order to ensure a good description both in data and MC, focusing in particular on jets. I also performed several studies on the muon description in data and MC trying to finalize the development of a multivariate-based identification similar to the approach reported for electrons. Moreover, I studied the description of the missing transverse energy in p-p events at 13 TeV, because it is suitable to include an additional bin in the event categorization. These two aspects are not detailed here because they are not part of the current strategy of the analysis, but they will be resumed in view of the Run 3 in order to further improve the analysis for the increasing statistics that will become available.

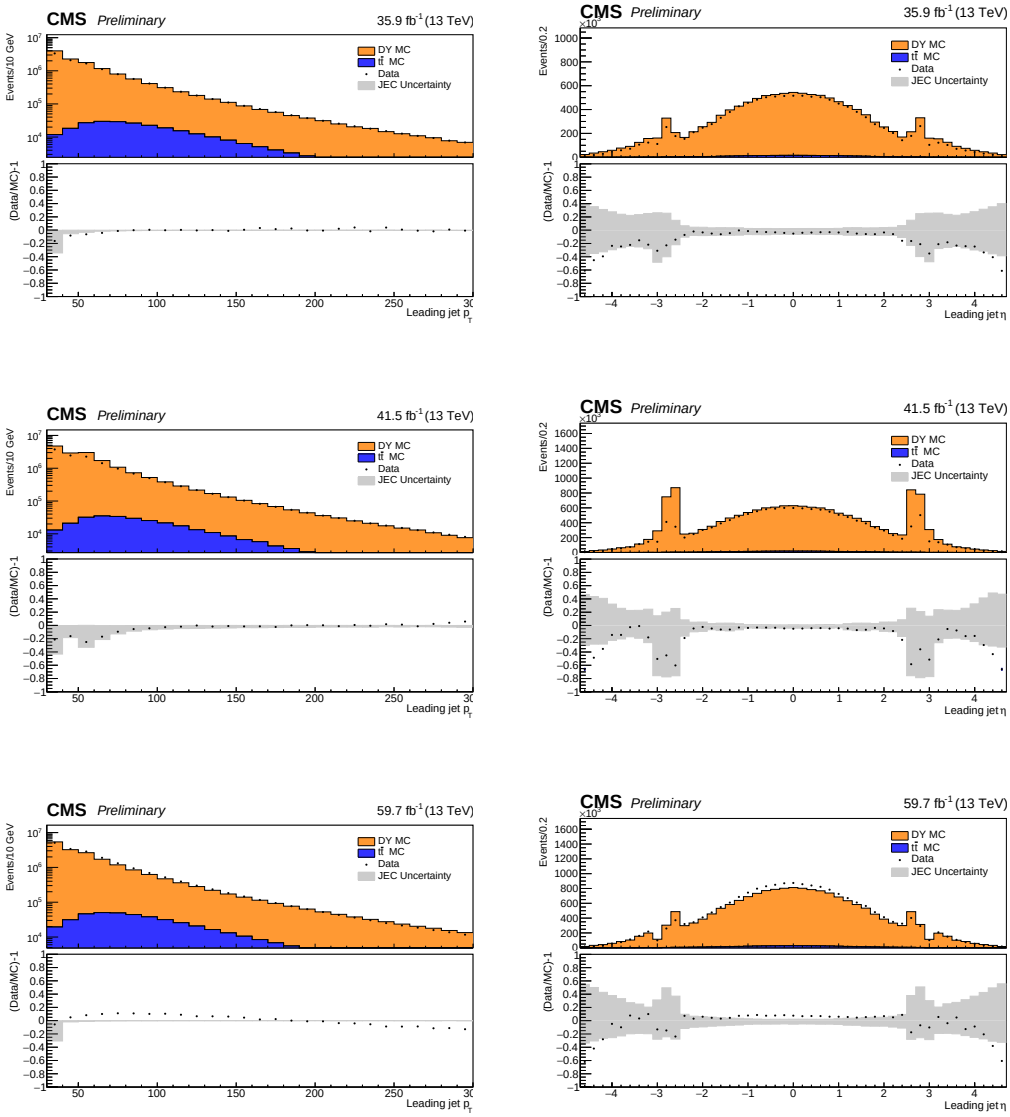


FIGURE 3.27: Comparison between the transverse momentum and pseudo-rapidity (right) of the leading jet (left) after the full selection in data and MC samples considering each year: 2016 (top), 2017 (middle), and 2018 (bottom). DY (orange) and $t\bar{t}$ (blue) MC samples are considered. Data/MC ratio plots are shown in the bottom of each plot, together with the uncertainties (gray histograms) coming from the combination of the jet energy corrections.

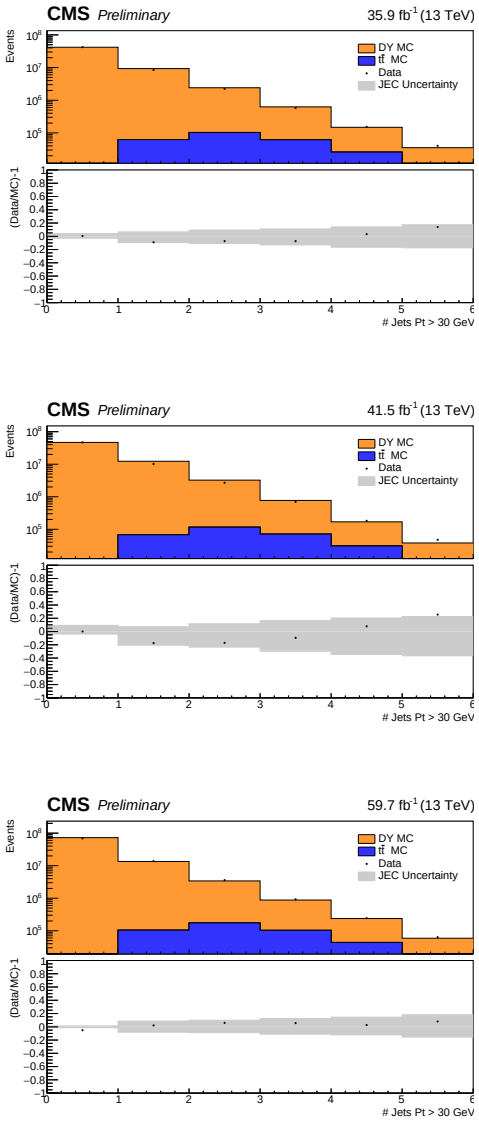


FIGURE 3.28: Comparison between the jet multiplicity after the full selection in data and MC samples considering each year: 2016 (top), 2017 (middle), and 2018 (bottom). DY (orange) and $t\bar{t}$ (blue) MC samples are considered. Data/MC ratio plots are shown in the bottom of each plot, together with the uncertainties (gray histograms) coming from the combination of the jet energy corrections.

Chapter 4

Event selection and categorization

In the previous chapter an overview of the basic ingredients needed to perform the analysis has been provided: the physics objects. Now we can proceed to present in which way the objects satisfying all selection requirements and quality criteria are combined into a $\mathrm{H} \to 4\ell$ candidate.
The definition and optimization of the signal region selection will be presented, followed by the properties exploited to discriminate the H boson signal against backgrounds and to classify the events depending on the specific topology of each H boson production mode. In order to study the H boson production mechanisms and improve the sensitivity of the cross section measurement for each of them, the selected events are classified into mutually exclusive categories based on the kinematical features of the reconstructed objects associated with the $\mathrm{H} \to 4\ell$ candidates. The categorization scheme that targets all main H production mechanisms is going to be described in detail, because it represents an essential point to understand the results presented in Chapter 6. The strategy adopted relies on the Simplified Template Cross Section (STXS) framework and benefits largely from the significant increase in the number of golden H boson events that have been recorded during the whole Run 2 data-taking period, representing a common basis for the latest H boson measurements performed by ATLAS and CMS. Concerning the observables and the categorization, a significant focus on the evolution of the approach since the Run 1 analysis will be discussed.

The topics mentioned here represent the first part of the core workflow of the $\mathrm{H} \to 4\ell$ analysis. The second part of the chain to close the circle relates to the modelling of the signal and background processes. Chapters 4 and 5 are meant to give to the reader a complete overview of the analysis strategy.

4.1 Event selection

In order to study the properties of the SM H boson it is crucial to define an optimal selection procedure on all possible events recorded by CMS. The selection approach of this analysis targets high purity on the signal H boson production while discarding background events as much as possible.
The first step of the event selection is represented by the application of the trigger requirements, which are necessary to store events offline for subsequent analysis, followed by the selection of the final state objects according to the discussion in Chapter 3. The four-lepton candidates are built starting from the selected leptons, defined as loose if they pass the reconstruction, kinematics, and primary vertex cuts, and then declared tight if they pass in addition the identification and isolation requirements, as described in Chapter 3. Most of the object selection is shared between the three analyzed data-taking periods: a summary of the physics object selection for the Run 2 $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis is presented in Table 4.1.

Electrons
$\mathrm{p_T^e} > 7$ GeV $\quad |\eta^e| < 2.5$
$\lvert d_{xy}\rvert < 0.5$ cm $\quad \lvert d_z\rvert < 1$ cm
$\mathrm{SIP_{3D}} < 4$
BDT ID with isolation with cuts from Table 3.3

Muons
Global or tracker Muon
Discard standalone muon tracks if reconstructed in muon system only
$\mathrm{p_T^\mu} > 5$ GeV $\quad |\eta^\mu| < 2.4$
$\lvert d_{xy}\rvert < 0.5$ cm $\quad \lvert d_z\rvert < 1$ cm
$\mathrm{SIP_{3D}} < 4$
PF muon ID if p_T < 200 GeV, PF muon ID or high p_T muon ID (Table 3.1) if p_T > 200 GeV
$\mathcal{I}^\mu < 0.35$
Ghost cleaning procedure (as explained in Section 3.2.2)

FSR photons
$p_T^\gamma > 2$ GeV $\quad |\eta^\gamma| < 2.4$
$\mathcal{I}^\gamma < 1.8$
$\Delta R(\ell,\gamma) < 0.5$ and $\frac{\Delta R(\ell,\gamma)}{(p_T^\gamma)^2} < 0.012$ GeV^{-2}

Jets
$p_T^\mathrm{jet} > 30$GeV $\quad |\eta^\mathrm{jet}| < 4.7$
$\Delta R(\ell/\gamma, \mathrm{jet}) > 0.4$
Cut based jet ID (tight WP)
Jet pileup ID (tight WP) for jets with a p_T < 50 GeV
DeepCSV b tagging algorithm (medium WP)

TABLE 4.1: Overview of the physics object selection in the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis using the full Run 2 dataset.

The requirement of four isolated leptons leads to the suppression of any background process with the exception of the irreducible continuum electroweak ZZ production. Moreover, the excellent resolution on the four-lepton invariant mass allows to separate the signal from this background. The small branching ratio of the decay process and the experimental challenges related to the reconstruction of very low momentum leptons, needed not to loose a ZZ candidate with a very soft lepton in the final state, represent the main limiting factors to the sensitivity of the decay channel under study.

4.1.1 Trigger requirements

The trigger selection represents the first step of the event selection. It is necessary to save data events offline that potentially have H boson decays for subsequent analysis among billions of collisions detected by the CMS detector every second. The trigger, described in more detail in Section 2.2.3, is designed and optimized to reach an efficiency close to 100% on the signal. The events recorded offline for this analysis are collected with a dedicated collection of HLT triggers which are designed to have muon and electron candidates in the event passing loose identification and isolation requirements. The minimal p_{T} of the leading and subleading leptons is not the same in all three analyzed years to take into account the evolution of the data-taking conditions. The main triggers select either a pair of muons or electrons; moreover, mixed electron-muon triggers are taken into account. Triggers requiring three leptons with relaxed transverse momentum thresholds and without isolation requirements are also used, as are isolated single-electron and single-muon triggers, aiming to maximize the coverage of the $\mathrm{H} \to 4\ell$ phase space. A collision event is stored for the analysis in case at least one of the considered HLT paths is fired without dependendence on the selected final state. This approach is due to the fact that we are targeting different production modes which can come with additional leptons in the event. For example, the HLT path selecting events with three electrons can be associated to the $4e$ final state but also to a $2e2\mu$ VH event with additional true electrons originating from the decay of the associated W or Z boson. Therefore, an OR logic of all trigger paths allows to slightly improve the trigger efficiency.

This analysis uses a dataset recorded by the CMS experiment during 2016, 2017, and 2018 corresponding to $35.9\ \mathrm{fb}^{-1}$, $41.5\ \mathrm{fb}^{-1}$, and $59.7\ \mathrm{fb}^{-1}$, respectively [122–124], resulting in a total integrated luminosity of $137\ \mathrm{fb}^{-1}$. The datasets used for the 2016, 2017, and 2018 analysis are listed in Appendix B along with the integrated luminosities. The analysis relies on different primary datasets (PDs), grouping the following types of events:

* **DoubleMuon**: two or three muons are present in the event;
* **SingleMuon**: only one muon is present in the event;
* **MuEG**: one muon and one electron are present in the event;
* **DoubleEG/DoubleEle**: two or three electrons are present in the event;
* **SingleElectron**: only one electron is present in the event;
* **EGamma**: this PD used in 2018 merges the previous two classes of events with one or two electrons in the event.

Each PD combines certain collections of HLT paths, where additional requirements on the objects are applied, e.g. on the p_{T} threshold, the isolation, or the pseudorapidity. The considered triggers separate events according to the number and type of final state objects: only one electron or muon (SingleEle/SingleMuon triggers), two electrons or two muons (DiEle/DiMuon triggers), three electrons or three muons (TriEle/TriMuon triggers), one electron and two muons (DiMuEle trigger) and viceversa (MuDiEle trigger), one electron and one muon (MuEle trigger). Later, individual triggers are logically combined to define

PDs. The complete list of HLT paths used to cover the phase space of the 4ℓ signal is reported in Appendix B for all three considered data-taking periods. The majority of our signal events is taken from the triggers requiring a pair of electrons or muons.
The minimal transverse momentum for the leading and subleading lepton required in these triggers is reported in Table 4.2. To avoid duplicate events from different primary datasets, events are taken:

* from EGamma if they pass the DiEle or TriEle or SingleEle triggers,
* from DoubleMuon if they pass the DiMuon or TriMuon triggers and fail the DiEle and TriEle triggers,
* from MuEG if they pass the MuEle or MuDiEle or DiMuEle triggers and fail the DiEle, TriEle, SingleEle, DiMuon and TriMuon triggers,
* from SingleMuon if they pass the SingleMuon trigger and fail all the above triggers.

DiEle/DiMuon triggers	2016	2017	2018
Leading electron	17	23	23
Subleading electron	12	12	12
Leading muon	17	17	17
Subleading muon	8	8	8

TABLE 4.2: Minimal p_{T} thresholds (GeV) used for the double muon and double electron triggers in 2016, 2017, and 2018 data.

The efficiency of the trigger selection has to be determined in collision data and to be reproduced in simulation in order to correct for possible mismatches. The efficiency in data of the combination of triggers used in the analysis with respect to the offline reconstruction and selection is measured by considering 4ℓ events triggered by single lepton triggers and using a method based on the TnP technique. A geometrical matching is performed between one of the four reconstructed leptons (tag) and a trigger object passing the final filter of one of the single muon or single electron triggers. The other three leptons are used as probe objects. Four possible TnP combinations can be present in each 4ℓ event; all possible combinations are counted in the denominator to compute the efficiency. For each of the three probe leptons all matching trigger filter objects are collected. Then, the matched trigger filter objects of the three probe leptons are combined trying to reconstruct any of the triggers used in the analysis. If any of the analysis triggers can be formed using the probe leptons, the set of probes is also counted in the numerator of the efficiency.

This method does not have a perfect closure in MC events because the presence of a fourth lepton increases the trigger efficiency and this effect is not taken into account. In the measurement, in fact, only three leptons are considered in data, because the fourth one is used as a tag. In addition, in the $2e2\mu$ final state the three probe leptons cannot be combined to form all possible triggers which can collect events with two electrons and two muons (e.g. if the tag lepton is an electron, the three remaining leptons cannot pass a double electron trigger). Therefore, the same trigger requirements are applied onto MC simulated events and the difference between the resulting trigger efficiencies in data and MC is used to determine the reliability of the simulation and associate a systematic uncertainty. The trigger efficiency plotted as a function of the minimum p_{T} of the third probe lepton is measured in 2016, 2017, and 2018 data and MC and is shown in Fig. 4.1 for the 4ℓ final state. Similar results are produced for each considered final state and are reported in Appendix B. The MC efficiency describes well the data within the statistical uncertainties. The trigger efficiency in simulation is found to be similar in all three years: it reaches a plateau at 99% for a lepton transverse momentum larger than 20 GeV in each final state and it is around 95% (98%) for smaller values of the electron (muon) p_{T}.

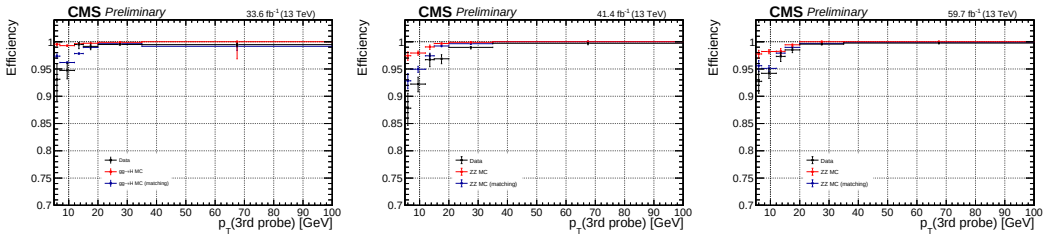


FIGURE 4.1: Trigger efficiency measured in 2016, 2017, and 2018 data and MC samples using 4ℓ events collected by single lepton triggers for the 4ℓ final state.

4.1.2 ZZ candidate selection

The design of the event selection aims to extract signal candidates from the combinatorial reconstruction of $\ell^+\ell^-$pairs from events containing at least four well-identified and isolated leptons, each originating from the primary vertex and possibly accompanied by a FSR photon candidate. A lepton cross cleaning is applied by discarding electrons which are within $\Delta R < 0.05$ of selected muons. The signal region definition is mostly determined by the selection of one ZZ candidate per event, on which any subsequent physics result will rely. The event selection of the $\mathrm{H} \to 4\ell$ analysis is identical for all three analyzed periods of data-taking and proceeds according to the following sequence:

1. **Primary vertex**
   The event is required to contain at least one reconstructed primary vertex (PV) passing the following quality requirements: a small radius of the PV ($R_{PV} < 2$ cm), high number of degrees of freedom ($N_{PV} > 4$), and collisions restricted along the z-axis ($z_{PV} < 24$ cm).

2. **Z candidates**
   To build Z candidates in the event, all possible pairs of selected leptons of opposite charge and matching flavor, either e^+e^- or $\mu^+\mu^-$, are considered. The reconstructed mass must satisfy $m_{\ell\ell(\gamma)}$ = [12, 120] GeV, where the computation of the invariant mass takes into account the FSR photons four-momentum in case of radiative decays.

3. **ZZ candidates**
   Among all selected lepton pairs, ZZ candidates are formed as all possible pairs of Z candidates that do not share a same lepton. The Z candidate with dilepton invariant mass $m_{\ell\ell(\gamma)}$ closest to the nominal Z boson mass given by the PDG [125] is selected and denoted as Z_1. The second, nonoverlapping, dilepton pair is referred to as Z_2. Depending on the flavor of the involved leptons, three mutually exclusive subchannels are defined: 4μ, $4e$, and $2e2\mu$. The following list of requirements on the ZZ candidates is meant to improve the quality of the signal candidate and, as a consequence, the sensitivity to H boson decays.

   * Ghost removal: all six pairs that can be built with the selected four leptons must be separated by $\Delta R(\eta, \phi) > 0.02$

   * Lepton p_T: in order to ensure that the selected events are on the plateau of the trigger efficiency, at least two leptons are required to have p_T > 10 GeV and at least one is required to have p_T > 20 GeV.

- Z_1 invariant mass: m_{Z_1} > 40 GeV.

- Alternative pairing check: in the 4μ and $4e$ final states, an alternative ZZ candidate can be formed out of the same four leptons. It is denoted as Z_aZ_b, where Z_a is the one closest to the nominal Z boson mass. In this scenario, the Z_aZ_b candidate is excluded if m_{Z_b} < 12 GeV. This requirement allows to discard 4μ and $4e$ events containing an on-shell Z and a low mass $\ell^+\ell^-$resonance.

- QCD suppression: among all four opposite charge pairs of leptons in the ZZ candidate, the requirement on the invariant mass of each pair $m_{\ell^+\ell^-}$ > 4 GeV is applied, regardless of lepton flavor. The goal is to suppress reducible background events with leptons originating from the decay of a heavy flavor hadron in jet fragmentation or from the decay of low mass resonances. Here, selected FSR photons are not used to compute $m_{\ell^+\ell^-}$, because a QCD-induced low mass dilepton resonance (e.g from J/Ψ) may have photons nearby (e.g from π_0).

- Four-lepton invariant mass: the mass range $m_{4\ell}$ > 70 GeV represents the phase space of interest for the subsequent steps of the analysis.

4. **Signal events**
   The events that contain at least one selected ZZ candidate form the signal region.

5. **Best ZZ candidate**
   If more than one ZZ candidate is present in the event after the full selection procedure, the one with the highest value of $\mathcal{D}^{\text{kin}}_{\text{bkg}}$, a kinematic discriminant which will be defined in Section 4.2, is selected as the best ZZ candidate. However, when different ZZ candidates involve the same four leptons, they have identical values of $\mathcal{D}^{\text{kin}}_{\text{bkg}}$: thus, the candidate with m_{Z_1} mass closest to the nominal Z boson mass is retained. The approach adopted differs from the Run 1 analysis, where the choice of the best ZZ candidate was performed looking at the Z candidate closest in mass to the nominal Z boson. The $\mathcal{D}^{\text{kin}}_{\text{bkg}}$-based choice naturally favours ZZ candidates which have a kinematic configuration more consistent with a H boson decay.

4.2 Observables: decay and production discriminants

The exploration of the H boson production in specific decay channels requires the experimental capability to discriminate signal-like events from SM backgrounds and categorize events. Moreover, the measurement of any physics parameter requires to understand which observables are most sensitive to it. For this reason, a dedicated set of observables is defined in the $H \to ZZ^* \to 4\ell$ analysis and is used for the classification of the events and as input to likelihood fits for the extraction of final results.

Since the analysis under study involves a complete reconstruction of the final state, the four-lepton invariant mass is the most emblematic observable to resolve a $H \to 4\ell$ signal over the continuum ZZ background. However, the sensitivity of the analysis can be increased by exploiting also the information given by kinematics properties such as the different angular distribution of the four leptons, the number of jets in the event, and the four-momentum of the final state objects. This information is combined into more complex observables. Kinematic discriminants related to either the H boson decay products or associated particles in its production are built using matrix element computations based on the Matrix Element Likelihood Approach MELA package [126–128], shortly referred to as *MEM-based* calculations, which provide the full set of processes studied in this analysis. The

JHUGen and MCFM matrix elements are used for the signal and the background, respectively. These discriminants rely on a complete set of mass and angular input observables $\vec{\Omega}$ to describe the kinematics at LO in QCD, as shown in Fig. 4.2. The p_{T} of either the combined H boson and the dijet system for the production discriminant (e.g. $\mathcal{D}_{\mathrm{2jet/1jet/VH}}$) [97, 129] or the H boson itself for the decay discriminants (e.g. $\mathcal{D}^{\mathrm{kin}}$) [14, 94] is not included in the input observables in order to reduce associated QCD uncertainties.

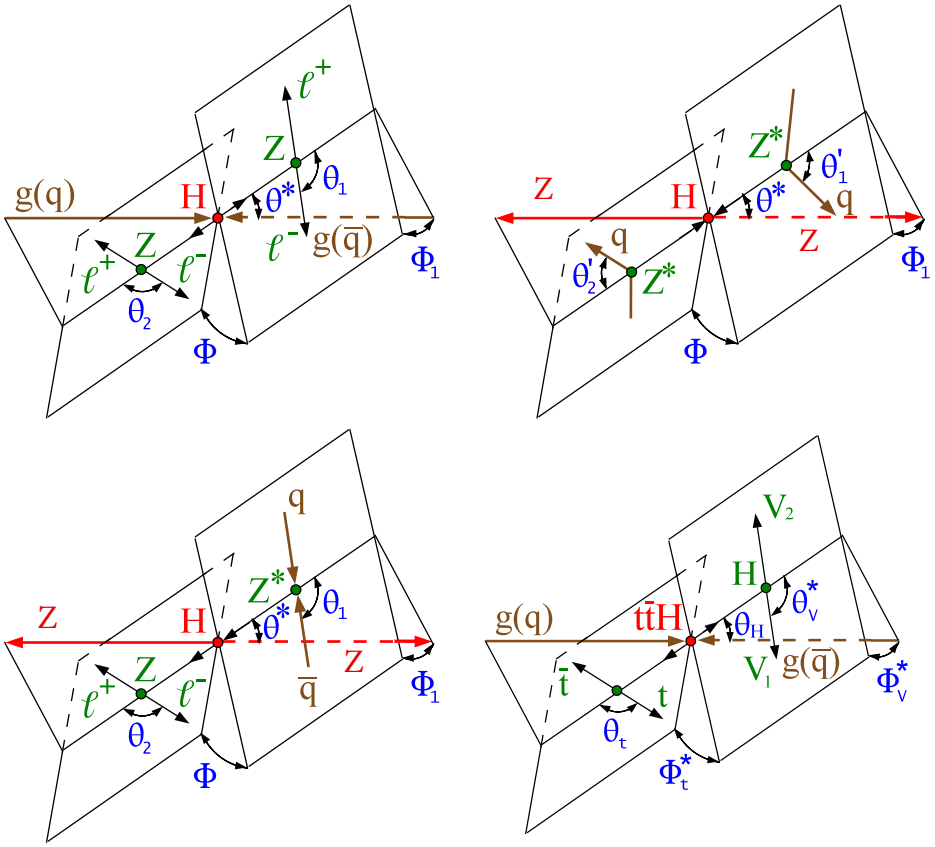


FIGURE 4.2: Illustration of different H boson production and decay processes: $\mathrm{gg/q\bar{q}} \to \mathrm{H} \to \mathrm{ZZ} \to 4\ell$ (top left), VBF $\mathrm{qq'} \to \mathrm{qq'H}$ (top right), $\mathrm{q\bar{q}} \to \mathrm{V^*} \to \mathrm{VH}$ (bottom left), and $\mathrm{gg/q\bar{q}} \to \mathrm{t\bar{t}H}$ (bottom right). Angles and Z boson invariant masses fully characterize the orientation of the production and decay chain and are defined in the suitable rest frames [126–128]. The θ^*, Φ, and Φ_1 angles are defined in the 4ℓ rest frame rest frame, while θ_1 and θ_2 are defined in the Z_1 and Z_2 rest frames, respectively.

Both the H boson decay kinematics and the kinematics of the associated production of H+jet, H+2jets, VBF, ZH, and WH are explored. The signal includes both the four-lepton decay kinematics in the processes $\mathrm{gg/q\bar{q}} \to \mathrm{X} \to \mathrm{ZZ/Z\gamma^*/\gamma^*\gamma^*} \to 4\ell$, and kinematics of associated particles in production H+jet, H+2 jets, VBF, ZH, WH, $\mathrm{t\bar{t}H}$, tHq, or $\mathrm{b\bar{b}H}$. The background includes $\mathrm{q\bar{q}} \to \mathrm{ZZ/Z\gamma^*/\gamma^*\gamma^*} \to 4\ell$ processes. In the following discussion, decay- and production-related discriminants describing the full event kinematics will be presented separately. It is interesting here to highlight that the development of production-related observables reflects the goal to investigate deeply the H production mechanisms, while in the context of the Run 1 analysis the decay-only observables represented a sufficiently powerful approach for the H search.

4.2.1 Decay discriminants

The decay part of the kinematic information can help in various aspects of the analysis: it allows to separate the H boson signal from the SM ZZ backgrounds and it can be used in the measurements of its mass, width, and spin-parity quantum numbers. Different decay-related observables are defined starting from the best ZZ candidate: the four-lepton invariant mass and kinematic discriminants $\mathcal{D}_{\text{bkg}}$ oriented to the signal-background separation. In the statistical analysis, a 2D fit which exploits the information from the invariant mass of the four leptons and $\mathcal{D}_{\text{bkg}}$ is used to extract final results, as explained in Section 6.1.2.

The four-lepton invariant mass

The most natural discriminating variable is the invariant mass of the four leptons selected to build the ZZ candidate, $m_{4\ell}$, which makes the H boson signal appear as a narrow resonance peak in the mass window around 125 GeV on top of a background which is almost flat. The discriminating power of this variable is directly related to the optimization of the lepton momentum resolution of the CMS detector. Unprecedented precision on the H boson mass measurement has been achieved thanks to the great performance of the CMS experiment together with additional fine tuning and offline corrections to data.

The $\mathcal{D}^{\text{kin}}_{\text{bkg}}$ kinematic discriminant

The kinematic configuration of the four-lepton decay of a scalar boson originally differs from the topology of the irreducible ZZ background and it is fully characterized by eight variables: $m_{4\ell}$, the Z boson invariant masses, and five angles, as illustrated in Fig. 4.2. These variables, except for $m_{4\ell}$, are collectively referred to as $\Omega^{\text{H}\to 4\ell}$. The $\mathcal{D}^{\text{kin}}_{\text{bkg}}$ variable sensitive to the $\text{gg}/\text{q}\bar{\text{q}} \to 4\ell$ kinematics is calculated as [14, 48, 94]:

$$\mathcal{D}^{\text{kin}}_{\text{bkg}} = \left[1 + \frac{\mathcal{P}^{q\bar{q}}_{\text{bkg}}(\vec{\Omega}^{\text{H}\to 4\ell}|m_{4\ell})}{\mathcal{P}^{gg}_{\text{sig}}(\vec{\Omega}^{\text{H}\to 4\ell}|m_{4\ell})}\right]^{-1}, \tag{4.1}$$

where the denominator contains the probability for the signal and the numerator includes the probability for the dominant $q\bar{q} \to 4\ell$ background process. It provides powerful discrimination between the signal and the ZZ background, characterizing only the decay of the H without providing information on its production mechanism.

However, the contamination from ggH process is significant in VBF and VH-hadronic categories. For this reason, two dedicated production-dependent $\mathcal{D}_{\text{bkg}}$ discriminants are used in the VBF category with two jets and in the hadronic VH category, defined as:

$$\mathcal{D}^{\text{VBF+dec}}_{\text{bkg}} = \left[1 + \frac{c^{\text{VBF+dec}}(m_{4\ell}) \times [\mathcal{P}^{\text{VBS+VVV}}_{\text{bkg}}(\vec{\Omega}^{\text{H+JJ}}|m_{4\ell}) + \mathcal{P}^{\text{QCD}}_{\text{bkg}}(\vec{\Omega}^{\text{H+JJ}}|m_{4\ell})]}{\mathcal{P}^{\text{VBF+VH}}_{\text{sig}}(\vec{\Omega}^{\text{H+JJ}}|m_{4\ell})}\right]^{-1} \tag{4.2}$$

$$\mathcal{D}^{\text{VH+dec}}_{\text{bkg}} = \left[1 + \frac{c^{\text{VH+dec}}(m_{4\ell}) \times [\mathcal{P}^{\text{VBS+VVV}}_{\text{bkg}}(\vec{\Omega}^{\text{H+JJ}}|m_{4\ell}) + \mathcal{P}^{\text{QCD}}_{\text{bkg}}(\vec{\Omega}^{\text{H+JJ}}|m_{4\ell})]}{\mathcal{P}^{\text{VBF+VH}}_{\text{sig}}(\vec{\Omega}^{\text{H+JJ}}|m_{4\ell})}\right]^{-1}, \tag{4.3}$$

where $\mathcal{P}^{\text{VBF+VH}}_{\text{sig}}$ is the probability for VBF and VH signal, $\mathcal{P}^{\text{VBS+VVV}}_{\text{bkg}}$ is the probability for vector boson scattering and the triboson background process, $\mathcal{P}^{\text{QCD}}_{\text{bkg}}$ is the probability for QCD production, and $c^{p}(m_{4\ell})$ is the $m_{4\ell}$-dependent constant to calibrate the distribution for the considered category p.

The performance of these new background discriminants in the VBF and VH-hadronic categories are compared with the default $\mathcal{D}^{\text{kin}}_{\text{bkg}}$ in Fig. 4.3. The capability of separating the

VBF and VH signals from the qqZZ background and the ggH process have been studied: the receiver operating characteristic (ROC) curves obtained using dedicated production-dependent discriminants outperforms the model based on the $\mathcal{D}^{\text{kin}}_{\text{bkg}}$ discriminant only. Consequently, the sensitivity of the analysis on the VBF and VH-hadronic categories is significantly enhanced by this approach. To quantify the gain, a test running the full statistical analysis was performed indicating an improvement of about 10-15% in the precision of the signal strength measurement in these categories.

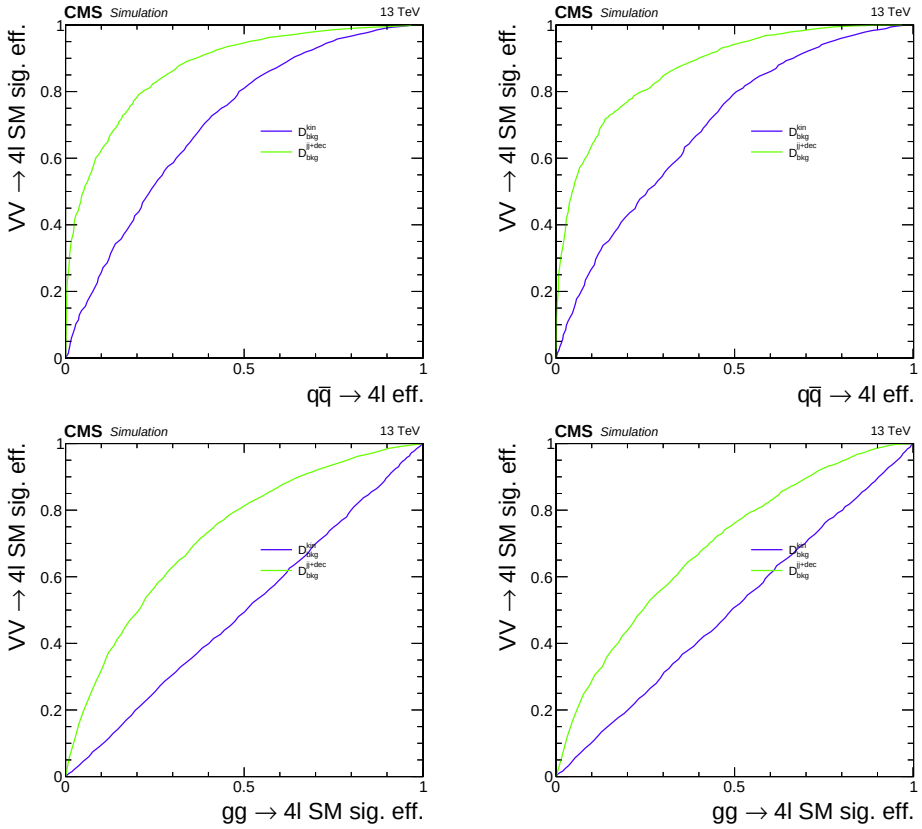


FIGURE 4.3: Comparison of the $\mathcal{D}^{\text{VBF+dec}}_{\text{bkg}}$ / $\mathcal{D}^{\text{VH+dec}}_{\text{bkg}}$ (green) and $\mathcal{D}^{\text{kin}}_{\text{bkg}}$ (blue) performances in the VBF category with two jets (left) and in the VH-hadronic category (right), respectively, in the discrimination between the VBF or VH process and the $q\bar{q}$ background (top left and top right, respectively) and between the VBF or VH process and the gluon fusion signal process (bottom left and bottom right, respectively).

4.2.2 Production discriminants

The usage of production sensitive observables enhances the discriminating power of the event categorization, which will be described in more detail in Section 4.3.3. For this reason, the jet information is exploited together with the already present information on the decay to better separate ggH from the other production modes and from the backgrounds. The number of jets in the event and the number of extra leptons represent crucial observables for the categorization of the different production mechanisms. In order to increase the purity of the target production mechanism in the event categorization (see Section 4.3.3), four additional matrix element production discriminants are calculated following the prescription in Refs. [48, 97, 129]. Two of them are sensitive to the VBF signal topology with one or two associated jets and the other two are sensitive to the VH signal topology, either

ZH or WH, with two associated jets. Note that during the Run 1, the event categorization relied only on two observables. In the 0/1-jet category, the transverse momentum of the four-lepton system was used to distinguish VBF production and associated production with a weak boson, VH, from gluon fusion. In the dijet category, the linear discriminant $\mathcal{D}_{\rm jet}$ was formed combining two VBF-sensitive variables in order to maximize the separation between vector boson fusion and gluon fusion processes: the absolute difference in pseudorapidity between the jet pair ($|\Delta\eta_{\rm jj}|$) and the invariant mass of the two leading jets (m_{jj}).

In the Run 2 observables evolution, the discriminant sensitive to the VBF signal topology with one or two associated jets are calculated as:

$$\begin{aligned}\mathcal{D}_{\rm 1jet} &= \left[1+\frac{\mathcal{P}_{\rm HJ}(\vec{\Omega}^{\rm H+J}|m_{4\ell})}{\int d\eta_{\rm J}\mathcal{P}_{\rm VBF}(\vec{\Omega}^{\rm H+JJ}|m_{4\ell})}\right]^{-1},\\ \mathcal{D}_{\rm 2jet} &= \left[1+\frac{\mathcal{P}_{\rm HJJ}(\vec{\Omega}^{\rm H+JJ}|m_{4\ell})}{\mathcal{P}_{\rm VBF}(\vec{\Omega}^{\rm H+JJ}|m_{4\ell})}\right]^{-1},\end{aligned} \tag{4.4}$$

where $\mathcal{P}_{\rm VBF}$ and $\mathcal{P}_{\rm HJ}/\mathcal{P}_{\rm HJJ}$ are probabilities obtained from the JHUGen matrix elements for the VBF process and gluon fusion (technically a combination of gg/qg/qq′ parton collisions) in association with one or two jets within the MELA framework. The $\int d\eta_{\rm J}\mathcal{P}_{\rm VBF}$ is the integral of the 2-jets VBF matrix element probability over the η_J values of the unobserved jet with the constraint that the total transverse momentum of the HJJ system is zero. The discriminant in Eq. 4.4 is equally efficient in separating VBF from either $\rm gg \to H + 2\,jets$ signal or $\rm gg/q\bar{q} \to 4\ell + 2\,jets$ background because jet correlations in these processes are distinct from the VBF process.

The discriminant sensitive to the VH signal topology with two associated jets is calculated as:

$$\begin{aligned}\mathcal{D}_{\rm WH} &= \left[1+\frac{\mathcal{P}_{\rm HJJ}(\vec{\Omega}^{\rm H+JJ}|m_{4\ell})}{\mathcal{P}_{\rm ZH}(\vec{\Omega}^{\rm H+JJ}|m_{4\ell})}\right]^{-1},\\ \mathcal{D}_{\rm ZH} &= \left[1+\frac{\mathcal{P}_{\rm HJJ}(\vec{\Omega}^{\rm H+JJ}|m_{4\ell})}{\mathcal{P}_{\rm WH}(\vec{\Omega}^{\rm H+JJ}|m_{4\ell})}\right]^{-1},\end{aligned} \tag{4.5}$$

where $\mathcal{P}_{\rm VH}$ and $\mathcal{P}_{\rm HJJ}$ are probabilities obtained from the JHUGen matrix elements for the VH (either ZH or WH) process and gluon fusion in association with two jets ($\rm H + 2\,jets$) within the MELA framework.

Fig. 4.4 shows distributions of the four production discriminants discussed above. Moreover, dedicated discriminants to isolate $\rm t\bar{t}H$, $\rm b\bar{b}H$, and $\rm tHq$ production mechanisms can be calculated using kinematics shown in the last graph of Fig. 4.2. However, the presently expected statistics in these categories is very small, so that the full reconstruction needed for the matrix element calculation would be highly inefficient. Details of such matrix element calculations and possible future applications can be found in Refs. [130, 131].

Now we have in hands the concepts needed to understand the approch used for the event categorization. To enhance the purity of the categories used in the analysis, all discriminants presented here are exploited. Working points are then defined to obtain a good compromise between expected category yields and purities.

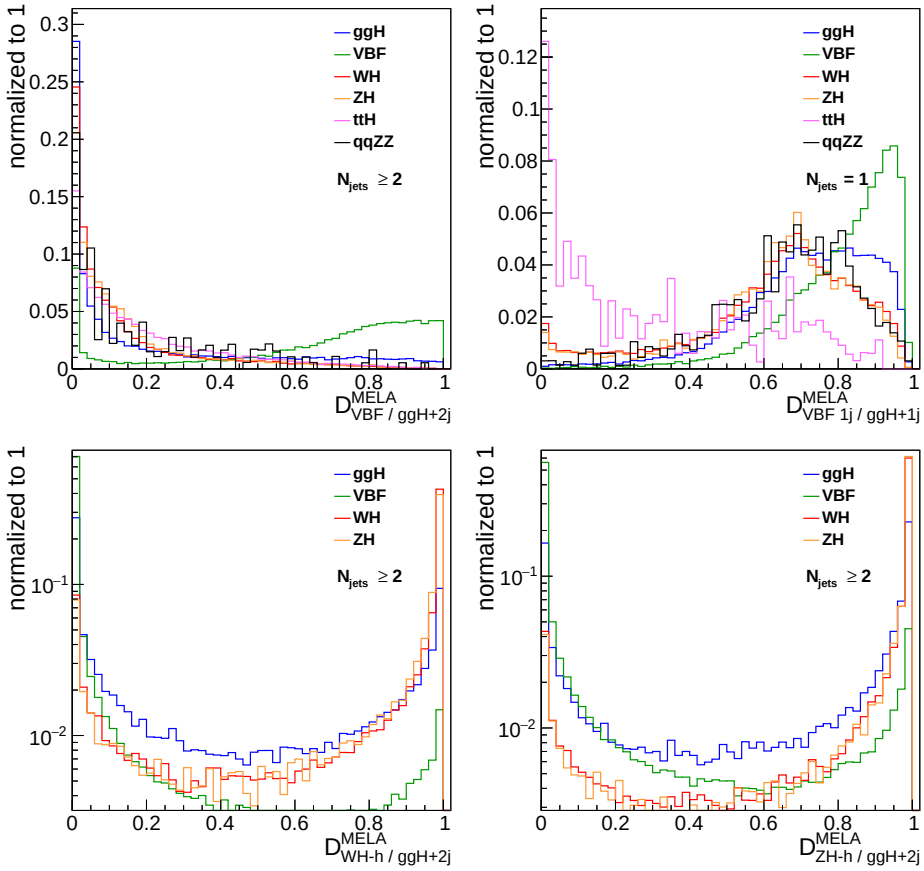


FIGURE 4.4: Distribution of the $\mathcal{D}_{2\text{jet}}$ (top left), $\mathcal{D}_{1\text{jet}}$ (top right), $\mathcal{D}_{\text{WH}}$ (bottom left), $\mathcal{D}_{\text{ZH}}$ (bottom right) discriminant for various production mechanisms of the H(125) signal and background in the $\text{H} \rightarrow 4\ell$ analysis. MC simulation at 13 TeV is shown.

4.3 Event categorization in the Simplified Template Cross Section framework

One of the goals of the work presented in this thesis is the measurement of the signal strength modifiers and simplified template cross sections of the H boson in the four-lepton final state. Therefore, it is crucial to achieve high sensitivity to the H boson production mechanisms. The event categorization allows to improve the sensitivity to specific experimental signatures and it is primarily designed to separate ggH, VBF, VH, and $\text{t}\bar{\text{t}}\text{H}$ processes. For this reason, the selected events are classified into seven mutually exclusive categories based on the features of the reconstructed objects associated with the $\text{H} \rightarrow 4\ell$ candidates. Therefore, a finer subdivision in twenty-two reconstructed event categories is performed (see Section 4.3.3) in order to study in a deeper way the structure within each production mechanism. Rare production modes like $\text{b}\bar{\text{b}}\text{H}$ and tH are considered explicitely in the analysis despite the little sensitivity. The splitting in mutually exclusive kinematic regions of phase space, named *bins*, is carried out by matching the recommended binning of the Simplified Template Cross Sections (STXS) framework [19, 132, 133], adopted by the LHC experiments as a common framework for H boson measurements. Its purpose and adoption to this analysis is described in Sections 4.3.1 and 4.3.2.

It is worth to underline the significantly evolving approach of this stage of the analysis with respect to the Run 1 strategy. In fact, originally the analysis targeted only two categories based on the jet multiplicity in order to improve the sensitivity to the H production, aiming to distinguish in particular the gluon fusion process from the vector boson fusion production: the 0/1-jet category, accounting for events with fewer than two jets, and the dijet category, containing events with at least two jets. Later, when the analysis of the 2016 data collected at 13 TeV started, a finer splitting in seven categories targeting the main production modes was introduced: this approach still represents the initial step of our categorization. At the end, increasing the Run 2 available statistics, the STXS approach started to become more important from a theoretical point of view. Thereby, new categories able to provide STXS measurements were designed, as presented in the following.

4.3.1 Overview of the STXS scheme

The kinematics of processes involving a H boson can be significantly modified by BSM physics. For example, the kinematic region of the H boson production at high transverse momentum can be particularly affected by BSM processes. This fact explains the primary interest in the measurement of the cross sections in specific regions of the phase space in addition to the coupling measurement. Considering the increasing precision of the H boson measurements, the κ-framework described in Section 1.2.2 does not allow to fully explore the information provided by experimental results. Moreover, given the complexity of the experimental analyses it is often difficult to unfold the final results in terms of the underlying H boson properties.
According to the STXS approach, fiducial regions in the phase space of the H boson production are defined providing more fine-grained measurements for the individual H boson production modes. On the one hand, the main purpose is to the reduce model dependence on theory predictions and theoretical uncertainties that are directly folded into the measurements as much as possible, maximizing the experimental sensitivity and isolating possible BSM effects. On the other hand, a similar approch makes easier the global combination of the measurements between different decay channels as well as between experiments, allowing to develop advanced and optimized analysis techniques and ensuring an easier and theoretically cleaner interpretability of the results.

Given that more granular STXS regions can be defined depending on the statistical power of individual analyses, a staged approach has been adopted by experiments to account for the evolving experimental sensitivity. The **stage 0** STXS framework corresponds to the individual production modes (`ggH`, `qqH`, `VH`, `ttH`) in the H boson rapidity range of $|y_{\mathrm{H}}| < 2.5$, since current measurements have no sensitivity beyond this rapidity range. However, it could be possible to use electrons at very forward rapidity regions (up to $|\eta| \sim 5$) in the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis and extend the accessible rapidity range. Therefore, an inclusive bin for $|y_{\mathrm{H}}| > 2.5$ can be added for each production process.
The **stage 1** predefined template bins correspond to a further fragmentation of the phase space into kinematic regions depending on the analysis sensitivity and is defined for `ggH`, `VBF`, and `VH` using information such as the transverse momentum of the H boson, the number of jets, and the invariant mass of jet pairs in the event. The bin definition is implemented at the truth level and is achieved by applying abstracted kinematic cuts which are explicitely kept simpler with respect to the exact fiducial volumes of the individual analyses in different H boson decay channels. The STXS measurements are complementary to differential measurements. The truth final state objects (leptons, jets, and the H boson itself) need to be defined unambiguously:

* **Higgs boson**: a global cut on the H boson rapidity at $|y_{\mathrm{H}}| < 2.5$ is included in all bins.

* **Jets**: generator-level jets are defined as anti-κ_T jets with a radius of 0.4 and a common p_{T} threshold of 30 GeV without restrictions on their rapidity. They are built from all stable particles (referring to all particles with a lifetime greater than 10 ps), including neutrinos, photons, and leptons from hadron decays or produced in the parton shower.

* **Leptons and leptonically decaying vector bosons**: no restriction on the transverse momentum and rapidity is placed on generator-level leptons. Vector boson decays are reconstructed taking into account the sum of all the involved leptonic decay products: electrons and muons from such leptonically decaying vector bosons are defined as dressed, i.e., all FSR photons should be added back to the electron or muon, and neutrinos are included.

The definition of the STXS bins is shared among all decay modes of the H boson. Several bins can be merged in each individual analysis and, consequently, only their sum can be measured according to the sensitivity of each analysis and decay channel. At the end, both individual and combined STXS measurements can be used as inputs for subsequent interpretations in and beyond the SM. For example, it is possible to determine overall signal strengths, coupling scale factors, or EFT coefficients (see Section 1.2.2) and to test specific BSM scenarios. A summary of the STXS framework is reported in Fig. 4.5, described in Ref. [16]. The main focus is on the three dominant H production processes: ggH, VBF, and VH.

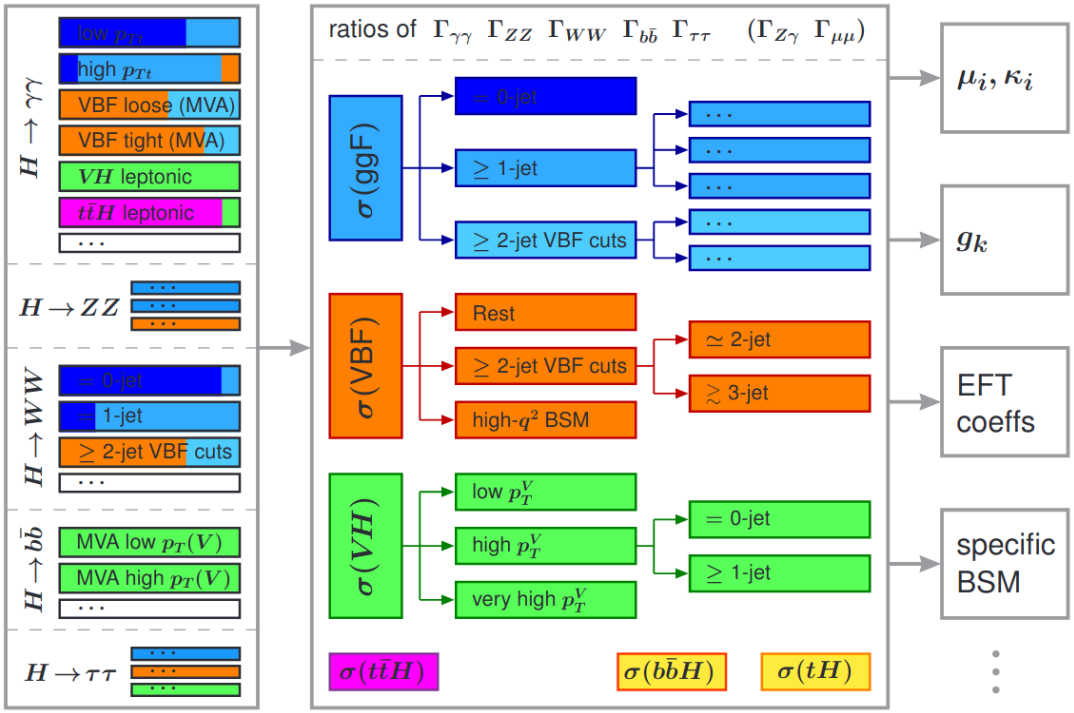


FIGURE 4.5: An overall view of the STXS scheme. In the left part, the experimental analyses targeting STXS measurements are shown. The main production modes included in the framework are reported in the middle: ggH (blue), VBF (orange), VH (green), and ttH (purple). Rare production modes are shown in yellow. Finally, the results can be used to constrain coupling modifiers, EFT coefficients, and specific BSM scenarios as summarized in the right part.

Currently, a revised **stage 1.2** STXS binning has been defined for the ongoing measurements by the ATLAS and CMS experiments using the full Run 2 dataset, aiming to better exploit the potential improvements in the full Run 2 measurements and enhance sensitivity on the VBF kinematics and low p_{T} ggH processes. Although the sensitivity of the analysis under study is not enough to fully investigate a so granular phase space, the whole implementation of the stage 1.2 is relevant as a preparation for the LHC Run 3 and can be interesting in the context of a CMS combination of the Run 2 results provided by different analyses.

4.3.2 STXS kinematic bins

In this section, we are going to define explicitely the stage 1.2 bins for individual H boson production based on the truth information in the MC simulation. The most recent scheme proposed by the LHC Higgs Cross Section Working Group as the result of many discussions and dedicated studies performed by the ATLAS and CMS Collaborations and the theory community is presented. Some final remarks concerning the adaptation of the scheme in the context of the $\mathrm{H} \to \mathrm{ZZ}^{*} \to 4\ell$ analysis are also included.

Gluon fusion production

The $\mathrm{gg} \to \mathrm{H}$ process together with higher order QCD and electroweak corrections is used to define the gluon fusion template process. Virtual electroweak corrections to the Born $\mathrm{gg} \to \mathrm{H}$ process and real electroweak radiation, referring in particular to the $\mathrm{gg} \to \mathrm{Z}(q\bar{q})\mathrm{H}$ process, are included in the calculations. Fig. 4.6 shows the treatment of the $\mathrm{gg} \to \mathrm{H}$ production process in the stage 1.2 of the STXS framework.

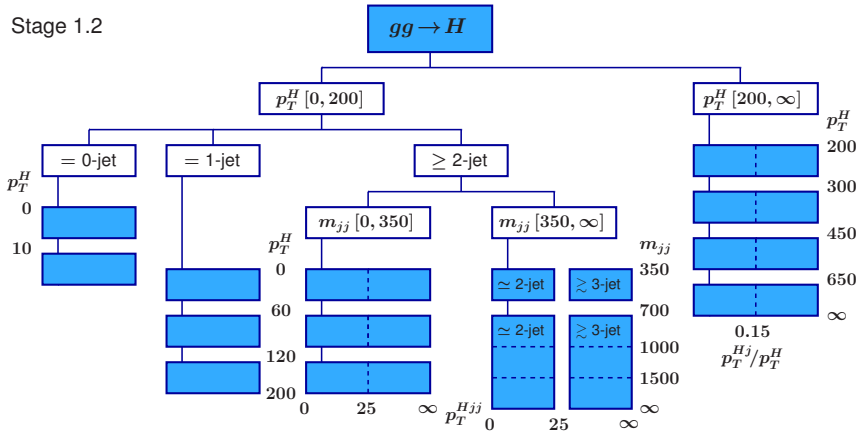


FIGURE 4.6: Binning of the ggH production process in the stage 1.2 of the STXS framework.

The $\mathrm{gg} \to \mathrm{H}$ production process is split into events with $p_{\mathrm{T}}^{\mathrm{H}} < 200$ GeV and $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV. On the one hand, the high $p_{\mathrm{T}}^{\mathrm{H}}$ region, primarily sensitive to BSM effects, is divided into four different bins. Moreover, a finer split into more $p_{\mathrm{T}}^{\mathrm{H}}$ bins is foreseen in a future iteration to allow a better consideration of the very high p_{T} region, which can be probed by dedicated boosted analyses. On the other hand, most of the cross section is given by events with $p_{\mathrm{T}}^{\mathrm{H}} < 200$ GeV and this kinematic region is split into events with zero, one, and two or more jets. Events with zero jets are divided into two $p_{\mathrm{T}}^{\mathrm{H}}$ bins with a boundary at 10 GeV to probe the very low p_{T} region of the H boson production, while events with one jet are divided into three $p_{\mathrm{T}}^{\mathrm{H}}$ bins with boundaries at 60 and 120 GeV. Events with two or

more jets are further split according to the dijet invariant mass using a cut at 350 GeV. The events with $m_{jj} < 350$ GeV are split into three production bins according to the H boson p_{T} analogously to the 1-jet bin. The $m_{jj} > 350$ GeV bin includes a small fraction of the total $\mathrm{gg} \to \mathrm{H}$ cross section, representing the main background to the VBF production. For this reason, the same splitting used in the corresponding high m_{jj} VBF bin is assumed with boundaries defined at m_{jj} equal to 700, 1000, and 1500 GeV. Moreover, a bin boundary is defined at a $p_{\mathrm{T}}^{\mathrm{H}jj}$ value of 25 GeV, providing a separation into phase space regions with two jets and three or more than three jets similarly to the approach adopted in the qqH scheme presented below. This allows for an easier uncertainty treatment for $\mathrm{gg} \to \mathrm{H}$ as background to the VBF process.

Electroweak qqH production

The electroweak production bin (qqH) includes the production of the H boson in association with two quarks from either VBF or VH events with hadronic decays of the vector boson V. The final state of the two topologies is identical: in fact, they represent the t-channel and s-channel contributions to the same physical process. The cross section is first split into events with zero, one, or two jets. Then, the 2-jet bin is split into two classes to distinguish low m_{jj} (divided into three bins with boundaries at 60 and 120 GeV) and high m_{jj} events using a cut at 350 GeV. A further splitting based on the $p_{\mathrm{T}}^{\mathrm{H}}$ is performed in the high dijet mass region, distinguishing events with two jets and three or more jets in the region $p_{\mathrm{T}}^{\mathrm{H}} <$ 200 GeV. Fig. 4.7 shows the treatment of the electroweak production process in the stage 1.2 of the STXS framework.

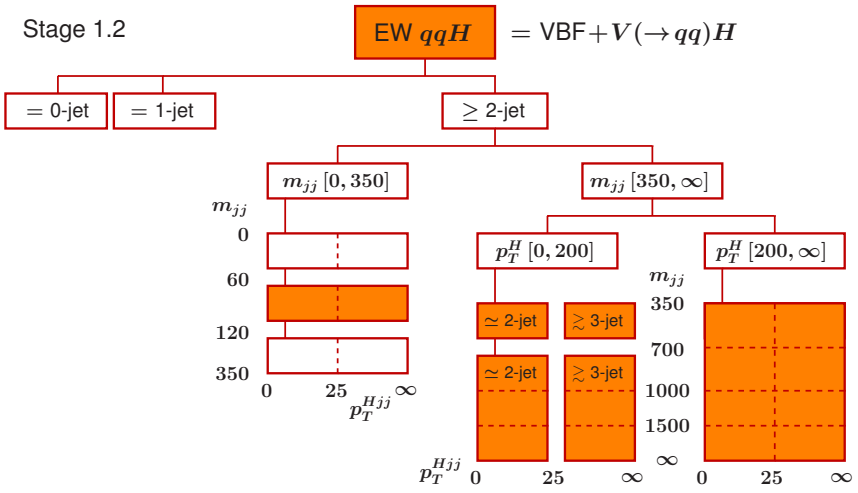


FIGURE 4.7: Binning of the qqH production process (combination of VBF and VH with hadronic V decay) in the stage 1.2 of the STXS framework.

Associated VH production

The associated VH production bin (VH) includes $\mathrm{pp} \to \mathrm{V}(\to \text{leptons})\mathrm{H}$ events with a leptonically decaying vector boson, separated into the three subprocesses $\mathrm{q\bar{q}}' \to \mathrm{W}(\to \ell\bar{\nu})\mathrm{H}$, $\mathrm{q\bar{q}} \to (\mathrm{Z} \to \ell\bar{\ell})\mathrm{H}$, and $\mathrm{gg} \to (\mathrm{Z} \to \ell\bar{\ell})\mathrm{H}$. The $\mathrm{q\bar{q}}' \to \mathrm{WH}$ and $\mathrm{q\bar{q}} \to \mathrm{ZH}$ underlying processes are split into $p_{\mathrm{T}}^{\mathrm{V}}$ bins with boundaries at 75, 150, 250, and 400 GeV, and exactly the same binning is used for $\mathrm{q\bar{q}} \to \mathrm{ZH}$ and $\mathrm{gg} \to \mathrm{ZH}$. The $150 < p_{\mathrm{T}}^{V} < 250$ GeV bin differentiates explicitly between events with zero and one or more jets. The subdivision of the VH production bin in the stage 1.2 of the STXS framework is reported in Fig. 4.8.

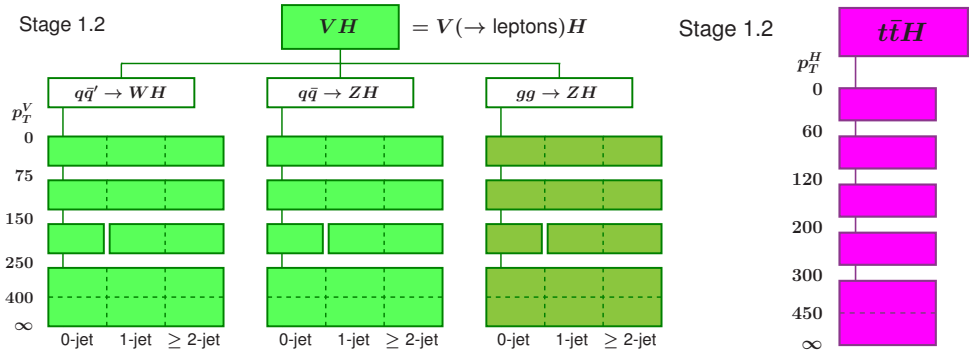


FIGURE 4.8: Binning of the VH production process with leptonic V decay and of the and $\mathrm{t\bar{t}H}$ process in the stage 1.2 of the STXS framework.

Other production modes

In stage 1.2 of the STXS framework the ttH stage 0 production bin is split in five different bins according to the p_{T} of the H boson, as shown in Fig. 4.8. The goal is the identification of bins which allow to constrain possible BSM effects, especially CP-odd contributions, or provide sensitivity to the Higgs self-coupling. Finally, the small $\mathrm{b\bar{b}H}$ and tH production processes are merged with the ggH and ttH production bins, respectively.

The merged stage 1.2 production bins in the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis

To conclude this section, we focus on the specific merging scheme adopted in the $\mathrm{H} \to 4\ell$ analysis starting from the organization presented above and shown in Fig. 4.9, considering that the full set of bins cannot be measured separately with the current statistics.
Concerning the ggH bin, the 2-jet events with $m_{jj} > 350$ GeV are all placed into one production bin merging the four bins originally suggested in stage 1.2. Moreover, given the limited sensitivity in the high $p_{\mathrm{T}}^{\mathrm{H}}$ region to the SM Higgs boson, this phase-space region is treated inclusively in jets, so that the events with $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV are placed into a single bin.
For the qqH bin, the production bins with zero jet, one jet, or with two or more jets with $m_{jj} < 60$ GeV or $120 < m_{jj} < 350$ GeV are of secondary importance in the analysis because they are very hard to access experimentally. Thereby, they remain merged into one single bin, while the events with two or more jets and $60 < m_{jj} < 120$ GeV are considered independently. The events with two or more jets and $m_{jj} > 350$ GeV are split into events with $p_{\mathrm{T}}^{\mathrm{H}} < 200$ GeV and $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV. Then, on the one hand events with $p_{\mathrm{T}}^{\mathrm{H}} > 200$ GeV are placed into one single production bin and on the other hand events with $p_{\mathrm{T}}^{\mathrm{H}} < 200$ GeV are split into two production bins in case $p_{\mathrm{T}}^{\mathrm{H}jj} < 25$ GeV and, otherwise, they are merged into a single bin.
Finally, the three production processes WH, ZH, and gluon fusion ZH are combined together and the proposed production bins are merged together into just two bins according to the p_{T} of the vector boson with a boundary at 150 GeV; the splitting of the ttH bin is condensed in a single stage 0 bin that includes the tH production process as well because of the very low expected yields.

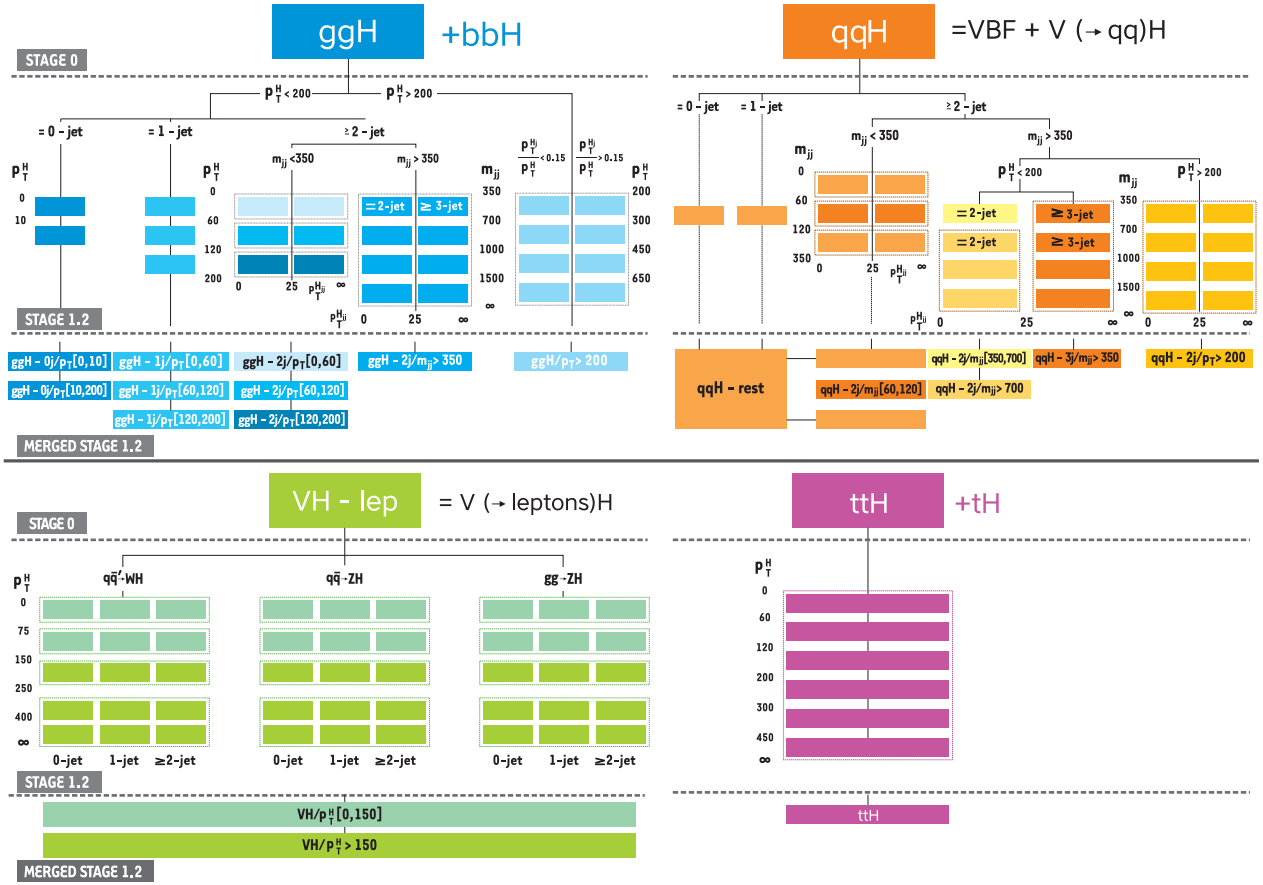


FIGURE 4.9: Binning of the gluon fusion production process, the electroweak production process (combination of VBF and VH with hadronic V decay), the VH production process with leptonic V decay (combining WH, ZH, and gluon fusion ZH production), and the ttH production process in the merged stage 1.2 of the STXS framework used in the analysis.

4.3.3 Reconstructed event categories - Stage 1.2

In order to increase the sensitivity to different production bins in the H boson production phase space, the ZZ candidates identified after the full event selection presented in Section 4.1 are classified into several dedicated reconstructed event categories. With the full Run 2 statistics, analyses gain sensitivity to the STXS stage 1.2 binning. Therefore, the finer splitting of the STXS stage 0 bins based on the event kinematic is also implemented in the $H \rightarrow ZZ^* \rightarrow 4\ell$ analysis resulting in an event categorization scheme with twenty-two reconstructed categories. These categories try to closely match the true STXS stage 1.2 predefined template bins (see Section 4.3.2). In addition, when all the reconstructed categories are merged together, signal strengths of different H boson production mechanisms can also be extracted. The category definition proceeds in two steps and relies on the kinematic information of the event to target the main production modes, taking into account:

- the multiplicity of jets;
- the number of b-tagged jets;
- the number of additional leptons which are not selected to form the ZZ candidates but satisfy the identification, vertex compatibility, and isolation requirements;
- MEM-based discriminants described in Section 4.2 (Eqs. 4.4 and 4.5);
- the invariant mass of the two leading jets;
- the transverse momentum of the ZZ or ZZ+j candidates.

At the beginning, the ZZ candidates are split into seven initial categories to target the main H boson production mechanisms corresponding to the stage 0 production bins. An event is considered for the subsequent category only if it does not satisfy the requirements of the previous category, as shown in Fig. 4.10, and the following order is adopted to apply the categorization criteria:

1. The **VBF-2jet-tagged** category requires exactly four leptons. In addition there must be either two or three jets of which at most one is b-tagged, or at least four jets and no b-tagged jets. Finally, $\mathcal{D}_{2\mathrm{jet}} > 0.5$ is required.

2. The **VH-hadronic-tagged** category requires exactly four leptons. In addition there must be two or three jets, or at least four jets and no b-tagged jets. Finally, $\mathcal{D}_{\mathrm{VH}} > 0.5$ is required.

3. The **VH-leptonic-tagged** category requires no more than three jets and no b-tagged jets in the event, and exactly one additional lepton or one additional pair of opposite sign same flavor leptons. Events with no jets and at least one additional lepton are also included in this category.

4. The $\mathrm{t\bar{t}H}$**-hadronic-tagged** category requires at least four jets of which at least one is b-tagged and no additional leptons.

5. The $\mathrm{t\bar{t}H}$**-leptonic-tagged** category requires at least one additional lepton.

6. The **VBF-1jet-tagged** category requires exactly four leptons and one jet; in addition, $\mathcal{D}_{1\mathrm{jet}} > 0.7$ is required.

7. The **untagged** category consists of the remaining events.

In principle, an additional VH category can be used at this level of the categorization to differentiate events with a considerable amount of missing transverse energy (MET), denoted as **V**H-**MET-tagged**. However, the modelling of the MET after the application of all corrections was not satisfying enough. Therefore, we decided to not include this category in the

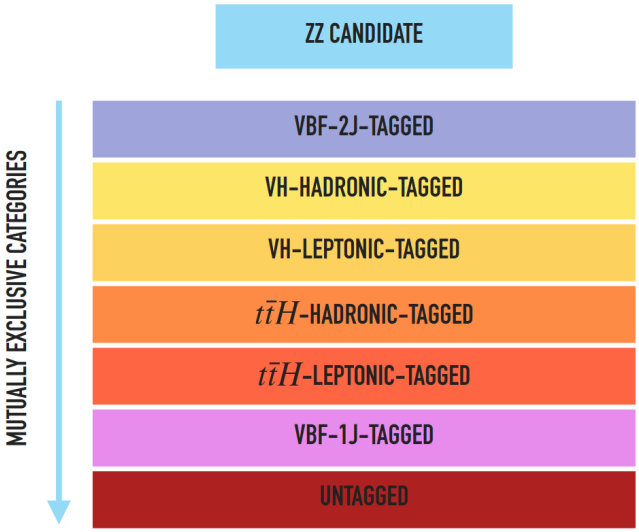


FIGURE 4.10: Stage 0 STXS scheme used in the $H \to ZZ^* \to 4\ell$ analysis: seven mutually exclusive categories are defined and used as input to the second step of the event categorization process.

current analysis, considering the very low sensitivity to the process and, as a consequence, the marginal impact on the whole analysis. This aspect will be resumed in view of the Run 3 analysis. After that, twenty-two reconstructed event categories are determined to match the merged stage 1.2 production bins presented in the Section 4.3.2. A further splitting of the untagged, VBF-2jet-tagged, VH-hadronic-tagged, and VH-leptonic-tagged categories is performed. Each reconstructed category mainly corresponds to a production bin (with the label reported in parentheses).

Untagged categories:

1. The **untagged-0j-$p_{\mathrm{T}}^{4\ell}$[0,10]** category requires the event to be in the untagged category without reconstructed jets and $p_{\mathrm{T}}^{4\ell}$ = [0,10] GeV (`ggH-0j/pT[0-10]`);
2. The **untagged-0j-$p_{\mathrm{T}}^{4\ell}$[10,200]** category requires the event to be in the untagged category without reconstructed jets and $p_{\mathrm{T}}^{4\ell}$ = [10,200] GeV (`ggH-0j/pT[10-200]`);
3. The **untagged-1j-$p_{\mathrm{T}}^{4\ell}$[0,60]** category requires the event to be in the untagged category with one reconstructed jet and $p_{\mathrm{T}}^{4\ell}$ = [0,60] GeV (`ggH-1j/pT[0-60]`);
4. The **untagged-1j-$p_{\mathrm{T}}^{4\ell}$[60,120]** category requires the event to be in the untagged category with one reconstructed jet and $p_{\mathrm{T}}^{4\ell}$ = [60,120] GeV (`ggH-1j/pT[60-120]`);
5. The **untagged-1j-$p_{\mathrm{T}}^{4\ell}$[120,200]** category requires the event to be in the untagged category with one reconstructed jet and $p_{\mathrm{T}}^{4\ell}$ = [120,200] GeV (`ggH-1j/pT[120-200]`);
6. The **untagged-2j-$p_{\mathrm{T}}^{4\ell}$[0,60]** category requires the event to be in the untagged category with two reconstructed jets, $p_{\mathrm{T}}^{4\ell}$ = [0,60] GeV, and the invariant mass of the jet pair $m_{jj} < 350$ GeV (`ggH-2j/pT[0-60]`);
7. The **untagged-2j-$p_{\mathrm{T}}^{4\ell}$[60,120]** category requires the event to be in the untagged category with two reconstructed jets, $p_{\mathrm{T}}^{4\ell}$ = [60,120] GeV, and the invariant mass of the jet pair $m_{jj} < 350$ GeV (`ggH-2j/pT[60-120]`);

8. The **untagged-2j-$p_{\mathrm{T}}^{4\ell}$[120,200]** category requires the event to be in the untagged category with two reconstructed jets, $p_{\mathrm{T}}^{4\ell}$ = [120,200] GeV, and the invariant mass of the jet pair m_{jj} < 350 GeV (ggH-2j/pT[120-200]);

9. The **untagged-$p_{\mathrm{T}}^{4\ell}$>200** category requires the event to be in the untagged category and $p_{\mathrm{T}}^{4\ell}$ > 200 GeV (ggH/pT>200);

10. The **untagged-2j-m_{jj}>350** category requires the event to be in the untagged category with two reconstructed jets and m_{jj} > 350 GeV (ggH-2j/mJJ>350);

VBF-tagged categories:

11. The **VBF-1jet-tagged** category requires the event to be in the VBF-1jet-tagged category (qqH-rest);

12. The **VBF-2jet-tagged-$m_{\mathrm{jj}}[350, 700]$** category requires the event to be in the VBF-2jet-tagged category with $p_{\mathrm{T}}^{4\ell}$ < 200 GeV, $p_{\mathrm{T}}^{4\ell jj}$ < 25 GeV, and m_{jj} = [350,700] GeV (qqH-2j/mJJ[350-700]);

13. The **VBF-2jet-tagged-m_{jj}>700** category requires the event to be in the VBF-2jet-tagged category with $p_{\mathrm{T}}^{4\ell} < 200$ GeV, $p_{\mathrm{T}}^{4\ell jj}$ < 25 GeV, and m_{jj} > 700 GeV (qqH-2j/mJJ>700);

14. The **VBF-3jet-tagged-m_{jj}>350** category requires the event to be in the VBF-2jet-tagged category with $p_{\mathrm{T}}^{4\ell}$ < 200 GeV, $p_{\mathrm{T}}^{4\ell jj}$ > 25 GeV, and m_{jj} > 350 GeV (qqH-3j/mJJ>350);

15. The **VBF-2jet-tagged-$p_{\mathrm{T}}^{4\ell}$>200** category requires the event to be in the VBF-2jet-tagged category with $p_{\mathrm{T}}^{4\ell}$ > 200 GeV and $m_{jj} >$ 350 GeV (qqH-2j/pT>200);

16. The **VBF-rest** category consists of the remaining events to be in the VBF-2jet-tagged category (qqH-rest);

VH-hadronic-tagged categories:

17. The **VH-hadronic-tagged-m_{jj}[60,120]** category requires the event to be in the VH-hadronic-tagged category with m_{jj} = [60,120] GeV (qqH/mJJ[60-120]);

18. The **VH-rest** category consists of the remaining events to be in the VH-hadronic-tagged category (qqH-rest);

VH-leptonic-tagged categories:

19. The **VH-leptonic-tagged-$p_{\mathrm{T}}^{4\ell}$[0,150]** category requires the event to be in the VH-leptonic-tagged category with $p_{\mathrm{T}}^{4\ell}$ < 150 GeV (VH-lep/pT[0-150]);

20. The **VH-leptonic-tagged-$p_{\mathrm{T}}^{4\ell}$>150** category requires the event to be in the VH-leptonic-tagged category with $p_{\mathrm{T}}^{4\ell}$ > 150 GeV (VH-lep/pT>150);

$t\bar{t}$H-tagged categories:

21. The **$t\bar{t}$H-leptonic-tagged** category (ttH);

22. The **$t\bar{t}$H category** (ttH).

Fig. 4.11 shows the signal relative purity of the merged stage 1.2 production bins in these twenty-two reconstructed event categories togehter with the expected number of signal events in each reconstructed event category. The fraction of events from the merged stage 1.2 production bins that populate each reconstructed event category is reported in the twenty-two horizontal lines. The purity of the untagged categories, targeting ggH events, is increasingly affected by the contribution of qqH events with the increase of the number of jets. Looking at the VBF-1jet-tagged, it can be noticed that the $gg \rightarrow H$ process represents a large background to the VBF production mode. This is also visible in the VBF-rest category as well as in the VH-rest category.

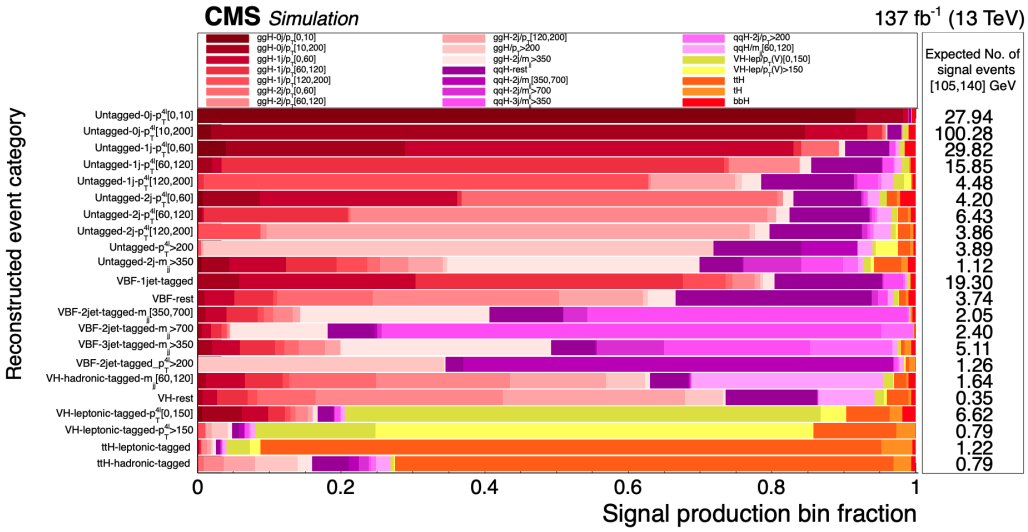


FIGURE 4.11: Signal relative purity of the twenty-two reconstructed event categories in terms of merged stage 1.2 production bins. In the right column, the expected number of signal events in each reconstructed event category is reported, considering the mass window $m_{4\ell}$ = [105,140] GeV used in the final statistical analysis.

Chapter 5

Modelling of the physics processes

In order to achieve a reliable exploration of the data events selected as output of the analysis workflow, an accurate estimate of the signal and background rates and their kinematics is essential. Therefore, the simulation of signal and relevant background processes performed with the Monte Carlo technique is crucial to perform various steps of the analysis. They are used for the optimization of the analysis strategy, for the comparison between the observed data and the predictions, for the evaluation of signal shape and acceptance, for the estimate of background events, and for the detection of the presence of a signal. Without a solid description of all the possible processes in a particle physics analysis, we could be insensitive to new physics. In case we are not able to achieve a good description of a physics process with the MC simulation, a dedicated strategy to estimate it using data is developed. The data-driven estimate of the reducible background in this analysis, due to the poor statistics of simulated events, is presented in Section 5.3.
At the end, any possible discrepancy or ignorance in the knowledge of the modelling of signal and background processes has to be considered in the analysis as a systematic uncertainty, essential part of the statistical interpretation of the data. In fact, as long as the amount of data and simulated events is large enough, the evaluation of the systematics is becoming even more important. All the components are finally combined together in a statistical analysis using a likelihood approach to measure physics parameters of interest, presented in Chapter 6 as the conclusive part of the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis story. In this chapter, we are going to discuss the modelling of the signal and then the MC simulation and data-driven estimation of the main background sources that affect this decay channel. A special focus is put on the challenging task of the modelling of the reducible background. This is due to the fact that large part of my contribution in the context of the $\mathrm{H} \to 4\ell$ analysis has been oriented on this topic since the beginning. Finally, the contribution and treatment of various systematic uncertainties is also detailed here, while their effect on the sensitivity of the measurements will be summarized in the next chapter.

A relevant aspect of this chapter consists in the fact that during the Run 1 analysis, which aimed for the observation of the H boson, an inclusive approach was adopted for the production categorization to verify that all signal and background predictions were on spot. The novelty of the Run 2 analysis presented here consists in a dedicated treatment of the modelling of signals and backgrounds according to the production mode, considering the twenty-two reconstructed categories derived from the STXS framework. A detailed study of the rare signal and background processes represents a significant challenge. In this context, my work was oriented to a full understanding of the reducible background estimate and modelling for each stage 1.2 production bin, representing an original contribution to the full Run 2 $\mathrm{H} \to 4\ell$ analysis. At the end of this chapter, all the concepts needed to understand the procedure to extract the $\mathrm{H} \to 4\ell$ signal and all subsequent steps that lead to the measurements of physics parameters will be in place.

5.1 Signal modelling of the SM Higgs

The Higgs production mechanisms are modelled using a MC simulation, which is described in detail in Section 5.1.1. Then, we will discuss the corrections applied to simulated samples in order to generate an even more accurate signal model (Section 5.1.2). However, the final measurements are not directly based on the $m_{4\ell}$ histogram templates provided from the simulation. In fact, the signal shape for $m_{4\ell}$ is described by an analytical function which helps to smooth out the irregularities due to the finite number of simulated events. The probability density functions (PDFs) used to describe the signal shapes for each mass point in different event categories will be discussed in Section 5.1.3.

5.1.1 Signal samples

The production of simulated samples for a CMS analysis is a very complex procedure which accounts for both the physics and experimental effects by interfacing different event generators with each other. In general, it consists of five main steps. First of all, the hard physics process with decays of the involved particles is generated providing a full list of particles with the information on the four-momenta calculated from the SM Lagrangian. A specific event generator is chosen depending on the physics process under study. Then, the simulation of the hadronization of quarks and gluons, jet fragmentation and showering, and the inclusion of underlying events is needed. In addition, pileup interactions are included in the simulated sample. In the final step, a detailed simulation of the CMS detector and its effect on the event reconstruction is applied, providing the raw detector information as the output. To conclude, the emulation of the high level trigger and reconstruction of events with the same algorithms used for data is performed. Each level of the simulation needs a different type of generator, as we are going to detail below.

To describe the SM H boson production, the POWHEG 2.0 [134–136] generator is used for the five main production processes, generating signal events at next-to-leading order (NLO) in perturbative QCD (pQCD): ggH [137] including quark mass effects [138], VBF [139], VH (either WH or ZH) [140], and $\mathrm{t\bar{t}H}$) [141]. In the case of WH and ZH the MiNLO HVJ extension of POWHEG is used. The ZH production mode occurs in two ways: $\mathrm{q\bar{q}} \to \mathrm{ZH}$ and the much smaller contribution from $\mathrm{gg} \to \mathrm{ZH}$ which is generated at leading order (LO) using JHUGen 7.0.2 [126–128, 130]. In addition to the main production processes, the contributions due to the rare H boson production mechanism are also taken into account. This is a novelty introduced in Run 2 because some of them may significantly contribute to some of the twenty-two categories discussed in Section 4.3.3. The H boson production in association with a single top quark and either a quark (tHq) or a W boson (tHW) are generated at LO using JHUGen and MadGraph5_aMC@NLO 2.2.2 [142, 143], respectively. The associated production with a pair of bottom quarks ($\mathrm{b\bar{b}H}$) is generated at LO with JHUGen. The list of the signal processes considered in the analysis and the corresponding $\sigma\mathcal{B}$ are reported in Table 5.1. Moreover, signal samples are generated for different values of the H boson mass to allow a parametrization of the expectations as a function of m_{H}, except for rare production modes: tHW, tHq, $\mathrm{b\bar{b}H}$. For the low mass studies presented here, the H signal is simulated using other four mass points in addition to H(125): 120, 124, 126, and 130 GeV. Finally, additional signal samples are generated to evaluate part of the systematic uncertainties, for example using different event generators for a comparison and varying the parameters for the UE (see Section 2.1.2) tune or the hadronization scale. This topic will be mentioned at the end of this chapter when discussing systematic uncertainties (Section 5.4).

In all cases, the modelling of the decay of the H boson to four leptons is obtained using the JHUGen generator, that properly takes into account the interference effects associated

Process	$\sigma\mathcal{B}(\times\epsilon_{\text{filter}})$
$gg \to H \to ZZ \to 4\ell$	13.34 fb
$q\bar{q} \to Hqq \to ZZqq \to 4\ell qq$	1.038 fb
$q\bar{q} \to W^+H \to W^+ZZ \to 4\ell + X$	0.231 fb
$q\bar{q} \to W^-H \to W^-ZZ \to 4\ell + X$	0.146 fb
$q\bar{q} \to ZH \to ZZZ \to 4\ell + X$	0.618 fb
$gg \to t\bar{t}H \to t\bar{t}ZZ \to 4\ell + X$	0.139 fb
$gg \to b\bar{b}H \to b\bar{b}ZZ \to 4\ell + X$	0.134 fb
$q\bar{q}/qg \to tHq \to tqZZ \to 4\ell + X$	0.021 fb
$qg \to tHW \to tWZZ \to 4\ell + X$	0.015 fb

TABLE 5.1: Signal processes considered in the analysis with their $\sigma\mathcal{B}$ including possible efficiency filters at the generation level. The rare H production mechanisms included in the presented full Run 2 analysis are highlighted in blue.

with permutations of identical leptons in the $4e$, 4μ, and 4τ final states. Considering also other combinations of decay products, $2e2\mu$, $2e2\tau$, and $2e2\mu2\tau$, six final states are added at the generator level in total, although reconstructed τ leptons are not used in the analysis. Therefore, a small amount of events involving τ pairs can be misreconstructed as $4e$, 4μ, or $2e2\mu$ events owing to the existence of leptonic decays of the τ. In the case of associated production with a vector boson or top quark, the H boson is allowed to decay to H $\to$ ZZ $\to$ 2ℓ2X so that four-lepton events where two leptons come from the decay of the associated particle are also taken into account in the simulation. The theoretical predictions used for the various production and decay modes can be found in Refs. [144–166] and are summarized in Ref. [16]. The NNPDF3.0 set of PDFs is used [167]. To simulate the showering of parton-level events with multiparton interactions and hadronization effects, all signal and background event generators are interfaced with PYTHIA 8 [168]. The PYTHIA event generator has been extensively compared to LHC data, resulting in different sets of tunings for the parameters needed at the generator level to simulate the parton shower. Tuning all parameters at the same time leads to a high dimensional parameter space. This is of particular relevance since the tune of a parameter in the parton shower can potentially bias particular new physics searches. The PYTHIA tune used to produce the 2016 MC samples is called CUETP8M1 [169], while for the 2017 and 2018 data-taking periods the CP5 [170] tune is used. Then, the generated events are processed through a detailed simulation of the CMS detector based on GEANT4 [171, 172] and are reconstructed with the same algorithms that are used for data.

Additional pileup interactions are added in simulated events. Given that an accurate prediction of the profile of the pileup is impossible, a dedicated pileup reweighting of simulated samples has to be taken into account in order to match the distribution of the number of interactions per LHC bunch crossing observed in data, as shown in Fig. 5.1. The average number of overlapping p-p interactions is either computed from the number of reconstructed primary vertices or from the measured instantaneous luminosity per bunch crossing. Then, the distributions obtained in simulation and data are used to compute pileup weights for a particular data-taking period. The minimum bias cross section used for each year is 69.2 mb. This procedure of pileup reweighting is found to have only a subpercent impact on expected yields in the signal region.

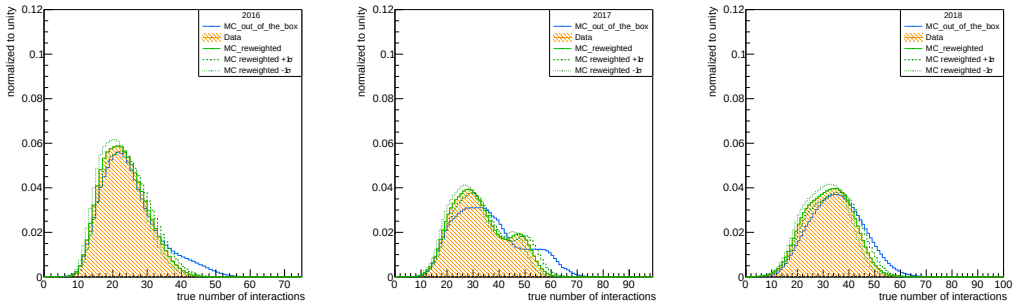


FIGURE 5.1: Distribution of pileup events in simulated samples and data is shown before and after the application of pileup weights, together with the up and down variations of 5% in the minimum bias cross section when calculating the weights. An illustration of the magnitude of pileup weights in 2016 (left), 2017 (middle), and 2018 (right) scenario is presented.

5.1.2 Signal corrections

To achieve a signal modelling as accurate as possible, the latest greatest N3LO QCD + NLO EW computations have been used for the ggH inclusive cross section, as recommended by the LHC Higgs Cross Section Working Group [50, 145]. Moreover, the p_{T} spectrum of the H boson ($p_{\mathrm{T}}^{\mathrm{H}}$) was tuned in the simulation of the dominant gluon fusion production mechanism with an event reweighting procedure to match closely the predictions from full phase space calculations at 13 TeV [138, 173, 174]. Weights are defined as function of generated $p_{\mathrm{T}}^{\mathrm{H}}$ and number of generated jets and then are used to reweight the POWHEG signal sample to the next-to-next-to-leading order parton shower (NNLOPS) event generator [175]. In practice, events originating from the ggH process are split into classes with 0, 1, 2, and $\geq$3 jets. The jets are clustered starting from all stable particles with p_{T} > 30 GeV except for the decay products of the H boson or associated vector bosons. The weights are obtained as the ratios of the $p_{\mathrm{T}}^{\mathrm{H}}$ distributions from the NNLOPS and the POWHEG generators for each event class. The sum of the weights in each sample is normalized to the inclusive cross section.

The effect of the NNLOPS reweighting on the spectrum of the H boson transverse momentum and on the distribution of the number of jets is shown in Fig. 5.2 considering the 2018 gluon fusion MC sample.

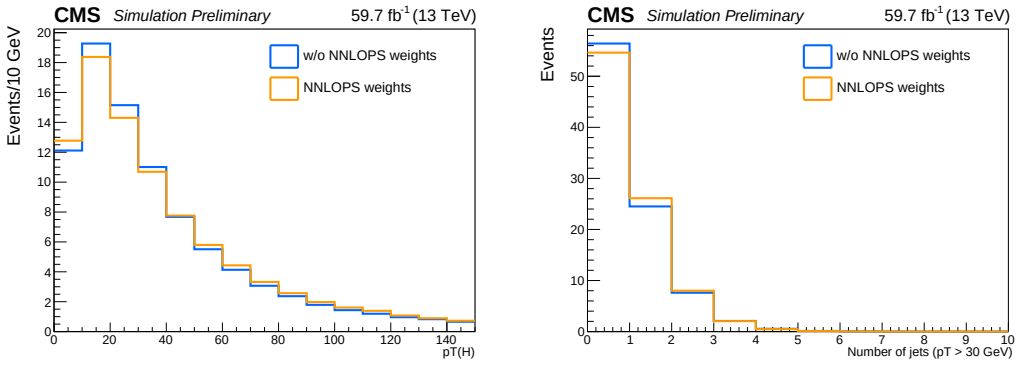


FIGURE 5.2: Comparison of the H boson p_{T} spectrum (left) and the number of jet distribution (right) with (orange) and without (blue) the NNLOPS reweighting using the 2018 ggH(125) MC sample.

5.1.3 Signal lineshape modelling

To measure any physics parameter for a certain value of the H boson mass, a continuous parametrization of signal predictions as a function of m_{H} is necessary. For measurements related to the on-shell H(125) boson presented in this thesis, the parametrization is built for the [105,140] GeV mass window starting from the few simulated samples produced with different m_{H} points (see Section 5.1.1).

The distribution of the signal invariant mass for the H boson depends only on the experimental resolution of lepton four-momentum, as the nominal width of the H boson is extremely narrow. The signal mass shape is parametrized by means of a double-sided Crystal Ball function [93] around $m_{\mathrm{H}} \sim 125$ GeV. A Landau function is added in the total PDF for the nonresonant part of the signal for the case of WH, ZH, and $\mathrm{t\bar{t}H}$ production modes. The double Crystal Ball function of six independent parameters is defined as:

$$dCB(\xi) = N \cdot \begin{cases} A \cdot (B + |\xi|)^{-n_L}, & \text{for } \xi < \alpha_L \\ A \cdot (B + |\xi|)^{-n_R}, & \text{for } \xi > \alpha_R \\ \exp\left(-\xi^2/2\right), & \text{for } \alpha_L \leq \xi \leq \alpha_R \end{cases} \tag{5.1}$$

where N is the normalizing constant and $\xi = (m_{4\ell} - m_{\mathrm{H}} - \Delta m_{\mathrm{H}})/\sigma_m$. It is intended to capture the Gaussian core of the four-lepton mass resolution function (σ_m) and the systematic mass shift of the peak (Δm_{H}), describing the two asymmetric non-Gaussian tails of the distribution with two parameters for each part. The tails originate from leptons emitting bremsstrahlung in the tracker material, present for both electrons and muons, from final state radiation, and energy leakage in the ECAL. The prominence of the left-hand/right-hand tail is defined as the power n_L/n_R, respectively. The parameters α_L/α_R define the position of the boundary between the core and the tails in units of σ_m. The two additional parameters in the formula, A and B, are not independent: they are defined by requiring the continuity of the function itself and its first derivatives.
The dependency of the six parameters p_i ($\sigma_m, \Delta m_{\mathrm{H}}, n_L, n_R, \alpha_L, \alpha_R$) on m_{H} has to be determined for each stage 1.2 STXS bin and each final state. Here, a linear approximation is used:

$$p_i(m_{\mathrm{H}}) = C_i + D_i \cdot (m_{\mathrm{H}} - 125\ \mathrm{GeV}). \tag{5.2}$$

Each parameter of the double-sided Crystal Ball function has a linear dependence on m_{H}, for a total of twelve free parameters. The initial value for the parameters is obtained by fitting the H(125) sample alone. The parameters of the PDF are extracted from a simultaneous fit to the simulated signal events for different H boson mass hypotheses in the [105,140] GeV mass range. The parameterization has been found to model very well the simulated data. The parametric signal model for each year of data-taking separately and for the sum of all years together is shown for the untagged event category as an example in Fig. 5.3.
To conclude, the normalization of the H signal in the peak region is directly taken from simulation. Using simulated samples for five mass points (120, 124, 125, 126, and 130 GeV), polynomial fits of the expected signal yields in the $m_{4\ell}$ = [105,140] GeV window around the H boson peak are performed to model separately each bin of the twenty-two reconstructed categories in three final states. Some examples of the fits are shown in Fig. 5.4 for different stage 1.2 STXS bins.

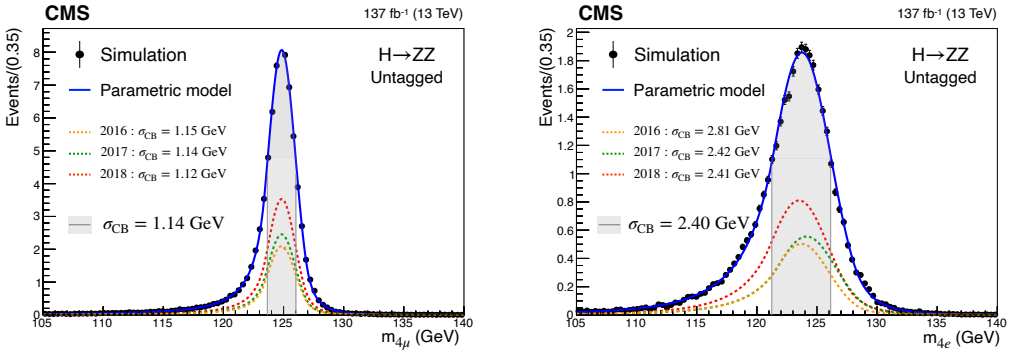


FIGURE 5.3: Parametric signal shape for each year of simulated data and for their sum considering the 4μ (left) and $4e$ (right) final states in the untagged event category. Weighted simulation events of the ggH(125) production mode and the corresponding model are represented by black points and blue line, respectively. The contribution of the signal model from each year of data-taking is illustrated with the dotted lines. The σ_{CB} value (half the width of the narrowest interval containing 68% of the invariant mass distribution) is shown in the grey shaded area.

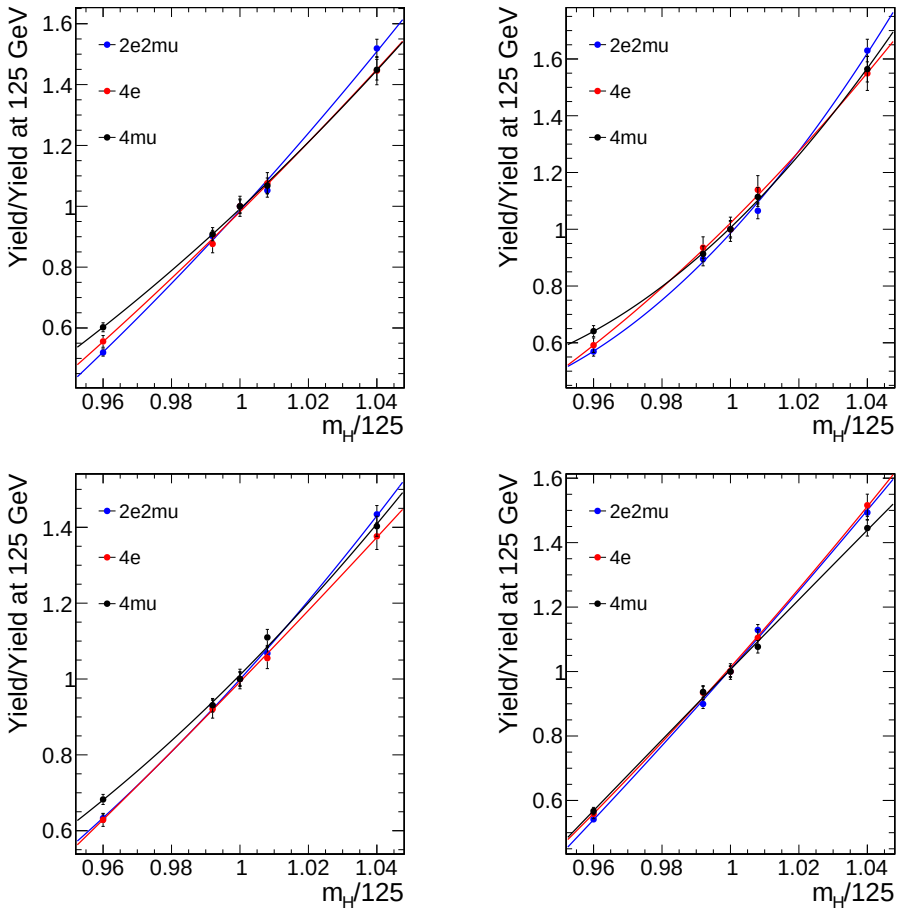


FIGURE 5.4: Fits of the expected signal yields after the full event selection in the $m_{4\ell}$ = [105,140] GeV mass window for some of the stage 1.2 bins taken as an example: the ggH bin with one jet in the $p_{\mathrm{T}}^{\mathrm{jet}}$ range [60,200] GeV (top left), the VBF bin with 1 jet (top right), the VH-hadronic (bottom left) and VH-leptonic (bottom right) bins.

5.2 Irreducible background modelling

To have a reliable signal extraction in the analysis, a proper understanding of all possible sources of background is essential. For this purpose, we want to achieve a good prediction of the background contribution and of its associated uncertainty. Having in mind the general procedure to produce simulated samples as described in Section 5.1.1, we can discuss the simulated samples produced to describe irreducible background processes. They are summarized in Table 5.2 and their contributions and shapes are determined directly from the simulation.

5.2.1 Background samples

The contribution from the quark-antiquark annihilation to the SM ZZ background is generated at NLO pQCD with POWHEG 2.0 [134–136] and PYTHIA 8, using the same settings as for the H boson signal. Given that this simulation covers a large range of ZZ invariant masses, dynamical QCD factorization and renormalization scales have been chosen, equal to m_{ZZ}. Event simulation for the $gg \to ZZ$ process is done at LO with the generator MCFM 7.0 [176, 177]. In order to match the distribution of the $gg \to ZZ$ transverse momentum predicted by POWHEG at NLO, different PYTHIA 8 settings are used for the showering in MCFM samples, allowing only emissions up to the parton-level scale (*wimpy shower*). Both simulated backgrounds require the application of dedicated k-factors to consider missing higher order corrections, as explained below.

In the analysis presented here, the triboson backgrounds ZZZ, WZZ, and WWZ, as well as the smaller $t\bar{t}Z$, $t\bar{t}WW$, and $t\bar{t}ZZ$ processes, are included for the first time in the background estimation. They are modelled using MadGraph5_aMC@NLO [142, 143]. Finally, the modelling of events containing Z bosons with associated jets (Z+jets) or WZ bosons and the $t\bar{t}$ background are also modelled using MadGraph5_aMC@NLO. They are not directly used to model data observations, because their contribution is actually estimated using a data-driven technique and dedicated control regions defined in data. Nevertheless, they are necessary for the optimization and validation of the methods and for testing the data/MC agreement in distributions of the object quantities and to ensure a good description of data in simulation. Moreover, DY samples represent the source of signal and background leptons in lepton-level optimization studies and efficiency measurements. They are generated using MadGraph5_aMC@NLO.

5.2.2 Irreducible background estimate

In principle, it is impossible to separate signal events from irreducible background processes on a single event basis because they have exactly the same final state and the same topology. Nonetheless, the study of a larger number of events exploiting some observables which are able to show the different behaviour of signal and background processes allows to identify a signal event occurring among the expected background processes.

The irreducible background to the H boson signal in the 4ℓ channel, originating from the production of ZZ via $q\bar{q}$ annihilation or gluon fusion, is estimated using simulation. Also some rare backgrounds such as triboson processes (ZZZ, WZZ, and WZZ) and events with $t\bar{t}Z$, $t\bar{t}WW$, and $t\bar{t}ZZ$, jointly referred to as the electroweak (EWK) backgrounds, are included in the evaluation of the irreducible background and are all estimated from the simulation.

Process	$\sigma\mathcal{B}(\times\epsilon_{\text{filter}})$
$qq \to ZZ \to 4\ell$	1.2560 pb
$gg \to ZZ \to 4e/4\mu/4\tau$	0.0016 pb
$gg \to ZZ \to 2e2\mu/2e2\tau/2\mu2\tau$	0.0032 pb
$Z \to \ell\ell$(M50) + jets	6225.4 pb
$WZ \to 3\ell\nu$	4.6700 pb
$t\bar{t} \to 2\ell2\nu2b$	87.310 pb
ZZZ	0.0140 pb
WZZ	0.0557 pb
WWZ	0.1651 pb
$t\bar{t}$ + ZZ	0.0016 pb
$t\bar{t}$ + WW	0.0079 pb
$t\bar{t}$ + Z(jets) (M10)	0.2590 pb
$t\bar{t}$ + Z($\to \ell\ell\nu\nu$) (M10)	0.2527 pb
$t\bar{t}$ + Z($\to \ell\ell$) (M1-M10)	0.0470 pb

TABLE 5.2: Background processes considered in the analysis with their $\sigma\mathcal{B}$ including possible efficiency filters at the generation level. The new samples included for the first time in the full Run 2 analysis are highlighted in blue.

qq → ZZ corrections

The $q\bar{q} \to ZZ$ process is simulated at NLO, but the fully differential cross section is computed at NNLO [178], which is not available yet in a partonic level event generator. For this reason, NNLO/NLO k-factors are applied to the POWHEG sample differentially as a function of m_{ZZ}, varying from 1.0 to 1.2 and being around 1.1 at $m_{ZZ} = 125$ GeV. According to the prescription in Ref. [179], additional NLO electroweak corrections depending on the initial state quark flavor and kinematics are also applied to the $q\bar{q} \to ZZ$ background process in the region $m_{ZZ} > 2m_Z$ where the corrections have been computed. The differential QCD and electroweak k-factors can be seen in Fig. 5.5.

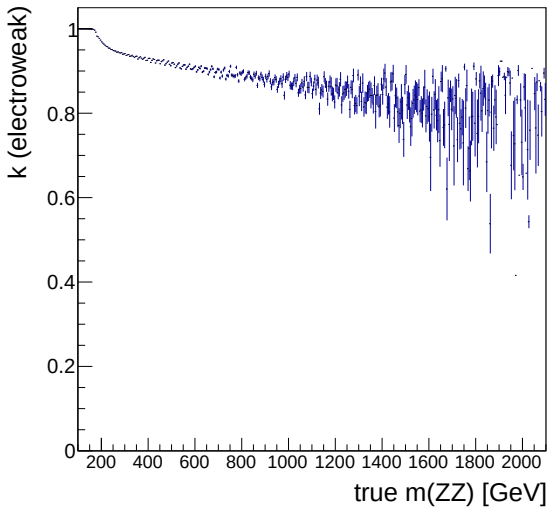


FIGURE 5.5: NNLO/NLO electroweak k-factor for the $q\bar{q} \to ZZ$ background as a function of m_{ZZ} for the 4ℓ final state.

Table 5.3 shows the inclusive cross sections obtained using the same PDF and renormalization and factorization scales as the POWHEG sample at LO, NLO, and NNLO.

QCD Order	$\sigma_{2\ell 2\ell'}$(fb)	$\sigma_{4\ell}$(fb)
LO	$218.5^{+16\%}_{-15\%}$	$98.4^{+13\%}_{-13\%}$
NLO	$290.7^{+5\%}_{-8\%}$	$129.5^{+4\%}_{-6\%}$
NNLO	$324.0^{+2\%}_{-3\%}$	$141.2^{+2\%}_{-2\%}$

TABLE 5.3: Inclusive cross sections for $q\bar{q} \to ZZ$ production at 13 TeV.

$gg \to ZZ$ corrections

The ZZ production via gluon fusion contributes at NNLO in pQCD, but currently no exact calculation exists beyond the LO. Even so, some studies [180] demonstrate that the soft collinear approximation describes well the background cross section and the interference term at NNLO. Additional calculations also show that the k-factors are very similar at NLO for the signal and background [181] and at NNLO for the signal and interference terms [182]. Thereby, the same k-factor computed for the signal is used also for the background [183]. The ratio between the NNLO and LO $gg \to H \to 2\ell 2\ell'$ cross sections computed at the small H boson decay width of 4.07 MeV with the HNNLO v2 MC program [174, 184, 185] is used to obtain the NNLO k-factor for the signal as a function of $m_{4\ell}$. The k-factor for the $gg \to ZZ$ background varies from 2.0 to 2.6 and is 2.27 at $m_{ZZ} = 125$ GeV. A systematic uncertainty of 10% on its determination is applied to the background process in the analysis. The illustration of the NNLO as well as the NLO k-factors and the cross sections from which they are derived is shown in Fig. 5.6, along with the NNLO, NLO, and LO cross sections at the SM H boson decay width [186].

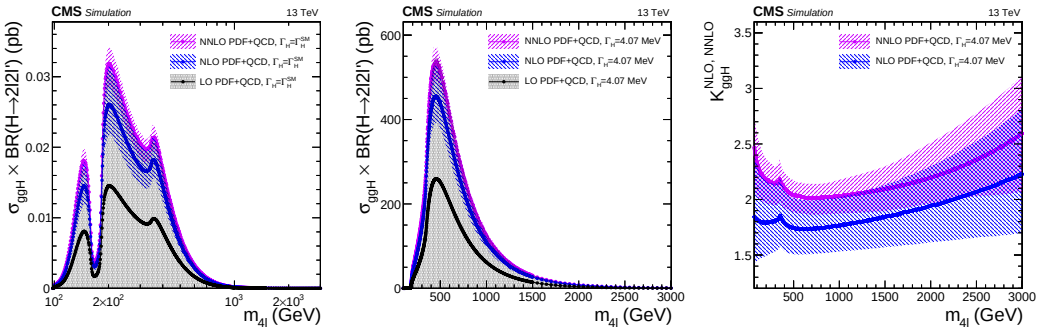


FIGURE 5.6: $gg \to H \to 2\ell 2\ell'$ cross sections at NNLO, NLO, and LO at each H pole mass using the SM H boson decay width (left) or at the fixed and small decay width of 4.07 MeV (middle). The cross sections at the fixed value are used to obtain the k-factor for both the signal and the continuum background contributions as a function of $m_{4\ell}$ (right).

5.3 Reducible background modelling

Due to the intrinsic differences between the irreducible and reducible background, they are estimated separately using different techniques. While the irreducible background is modelled from the simulation, a more challenging task is the estimate of the Z+X reducible background. It derives from processes that contain final state objects misidentified as leptons coming from a Z boson. This is due to imperfections of the detector performance. Ideally, no events coming from the reducible background would ever be selected as a final ZZ candidate with a perfect detector. In order to mitigate this effect, a dedicated set of selection cuts is designed to reject as much of the Z+X background as possible without affecting

the signal efficiency, as described in Chapter 4. This step reduces the amount of reducible background that enters final event selection, but there is still a nonnegligible contribution which needs to be considered.
The final estimate of the Z+X yields and the study of the kinematics is performed in dedicated control regions (CR) defined in data. This approach is common to many particle physics analyses. A CR is a region of the phase space designed to be orthogonal to the signal region (SR) and more populated than it. I took care of this technical aspect of the analysis during my activity in collaboration with the CMS $H \to 4\ell$ analysis group, focusing on the optimization of the method developed for the Run 1 analysis at the time of the discovery of the H boson in order to provide the Z+X estimate in each reconstructed category. To model the reducible background, dedicated studies per category have been performed regarding both the yields and the shapes of the process in each final state.

As discussed in Section 3.1, the dominant sources of nonprompt leptons in the four-lepton final state are heavy flavor jets producing secondary leptons, jets originating from light flavor quarks misreconstructed as leptons, and electrons from a photon conversion. In this discussion, each of these objects will be considered a *fake lepton* if it passes the loose selection criteria accounting for kinematic and impact parameter cuts, as defined in Section 3.3.2 and 3.2.2 for electrons and muons, respectively. The workflow which provides the final Z+X estimate is composed of two main parts. First of all, the rate of these background processes is estimated by measuring the f_e and f_μ probabilities, hereafter referred to as fake ratios or fake rates (FRs), for fake electrons and fake muons to also pass the final selection criteria (defined in Section 4.1.2). Then, these probabilities are applied in dedicated control samples in order to extract the expected Z+X background yields in the SR.

In the following sections, two independent methods are presented to measure both the yields and shapes of the reducible background, called *opposite sign* (OS) method and *same sign* (SS) method, and differences between both methods will be described. The final result combines the outcome of the two approaches.

5.3.1 Reducible background estimate with the opposite sign method

The Z+X estimate relies on two kinds of control regions, called Z+ℓ and Z+$\ell\ell$, where Z and ℓ denotes a Z candidate and a loose lepton, respectively. The former is used to calculate muon and electron FRs, and the latter is composed of events where the FRs are applied to get the final estimate of the Z+X background yields. All loose leptons can be divided into two orthogonal categories: passing leptons (P), defined as leptons that pass the full lepton selection, and failing leptons (F), defined as loose leptons that fail the isolation or identification cut, or both.

The Z+ℓ CR consists of events with a Z candidate and exactly one loose lepton in the event. The p_{T} cuts on leptons coming from the Z candidate and the QCD suppression cut specified in Section 4.1.2 are applied to define the phase space of interest for the FR calculation. Then, two different Z+$\ell\ell$ control samples are identified as subsets of the four-lepton events, denoted as 3P+1F and 2P+2F. Two opposite sign leptons where either one lepton is failing lepton and the other one is passing lepton or both leptons are failing leptons, respectively, are required to have a Z+$\ell\ell$ candidate. Two Z candidates are present in this CR where all kinematic cuts for the H boson phase space selection (see Section 4.1.2) are applied: a normal Z candidate (Z_1) and a Z candidate built with relaxed selection criteria (Z_2). The FSR recovery is treated in the same way as in the signal region, while the jet cleaning procedure is extended to also involve the loose and failing leptons of selected Z+$\ell\ell$ candidates and their associated photons. The distribution of the invariant mass of the 2P+2F and 3P+1F events stacked on the distribution of the WZ and the ZZ/Z$\gamma^* \to 4\ell$ irreducible background taken from the simulation is shown in Fig. 5.7 and 5.8 for the 2017 dataset. Distributions

related to the 2016 and 2018 datasets are reported in Appendix C. Given the discrepancy observed between data and simulated samples and the poor statistics of simulated events, it is not possible to rely on the MC simulation to estimate the reducible background. This is why a dedicated data-driven method has been developed.

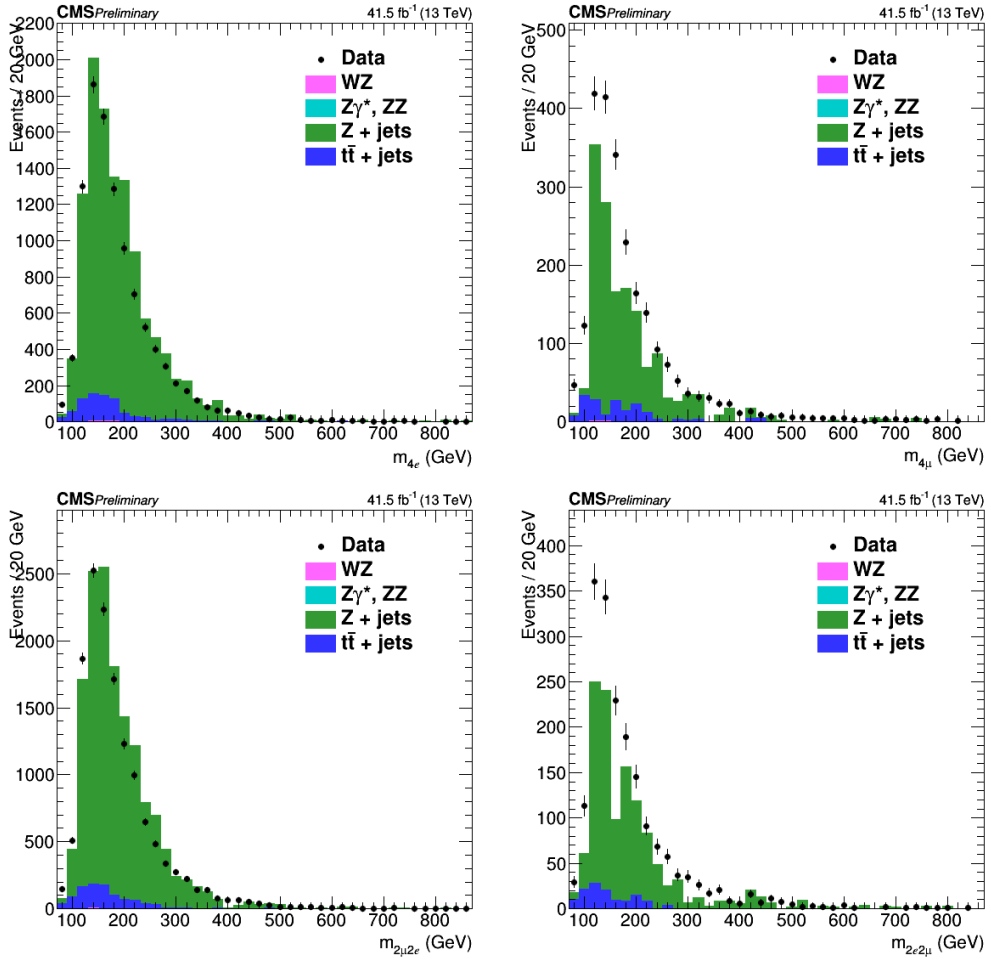


FIGURE 5.7: Invariant mass distribution of the events selected in the 2P+2F control sample in the 2017 dataset for all the considered channels: $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right).

In order to avoid overlaps, events selected in the SR are not used to define CR candidates, and no more than one candidate is chosen in each event for each method. However, one event can still provide candidates for both the 2P+2F and 3P+1F regions: thus, this possibility is taken into account in the calculation, as presented below.

Fake rate determination (OS method) To measure the lepton fake ratios f_e and f_μ, samples of $\mathrm{Z}(\ell\ell) + e$ and $\mathrm{Z}(\ell\ell) + \mu$ events are selected. They are expected to be completely dominated by final states which include a Z boson and a fake lepton. These events are required to have two same flavor, opposite charge leptons with p_T larger than 20 GeV and 10 GeV, respectively. They have to pass the tight selection criteria and thus they form the Z candidate. In addition, there is exactly one lepton passing the loose selection criteria as defined above, which is used as the probe lepton for the FR measurement. The invariant mass of this lepton and the opposite sign lepton from the reconstructed Z candidate should satisfy $m_{\ell\ell} > 4$ GeV.

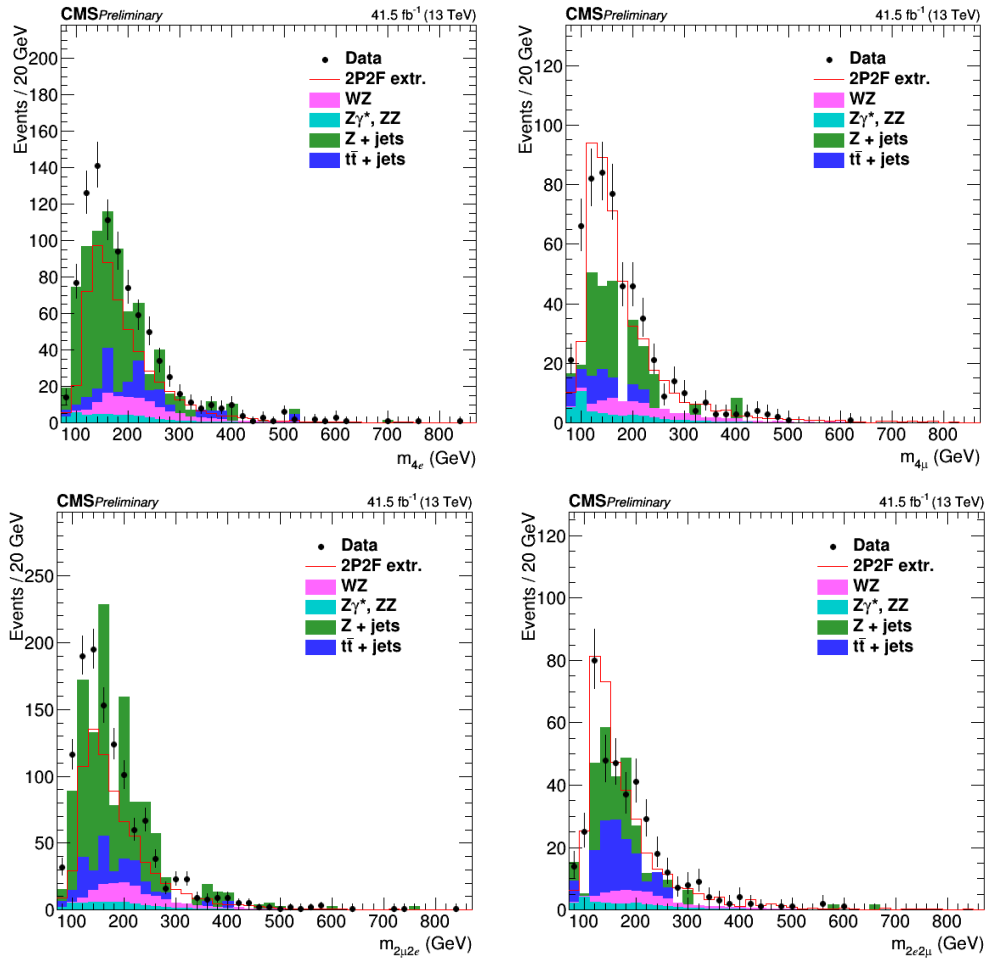


FIGURE 5.8: Invariant mass distribution of the events selected in the 3P+1F control sample in the 2017 dataset for all considered channels: $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right).

The FRs are evaluated using the tight requirement $|m_{\ell_1\ell_2} - m_Z| < 7$ GeV to reduce the contribution from the asymmetric photon conversions populating the low mass region and asking for $p_T^{miss} < 25$ GeV. The muon and electron FRs are measured in bins of the transverse momentum of the loose lepton in the barrel and endcap regions, separately for the 2016, 2017 and 2018 data-taking periods, and they are shown in Fig. 5.9.

Fake rate application (OS method) The application of FRs to estimate the reducible background in the SR is performed in the two dedicated control samples introduced before, which are enriched with fake leptons and are orthogonal by construction to the SR:

* the 2P+2F sample is expected to be populated with events that intrinsically have only two prompt leptons (mostly DY, with a small fraction of $t\bar{t}$ and Z+γ events);

* the 3P+1F sample is expected to include the same type of events as in the 2P+2F region, albeit with different relative proportions, as well as WZ events which intrinsically have three prompt leptons.

To compute the expected number of reducible background events in the 3P+1F region, N_{3P+1F}^{bkg}, it is necessary to account for the number of events observed in the 2P+2F control

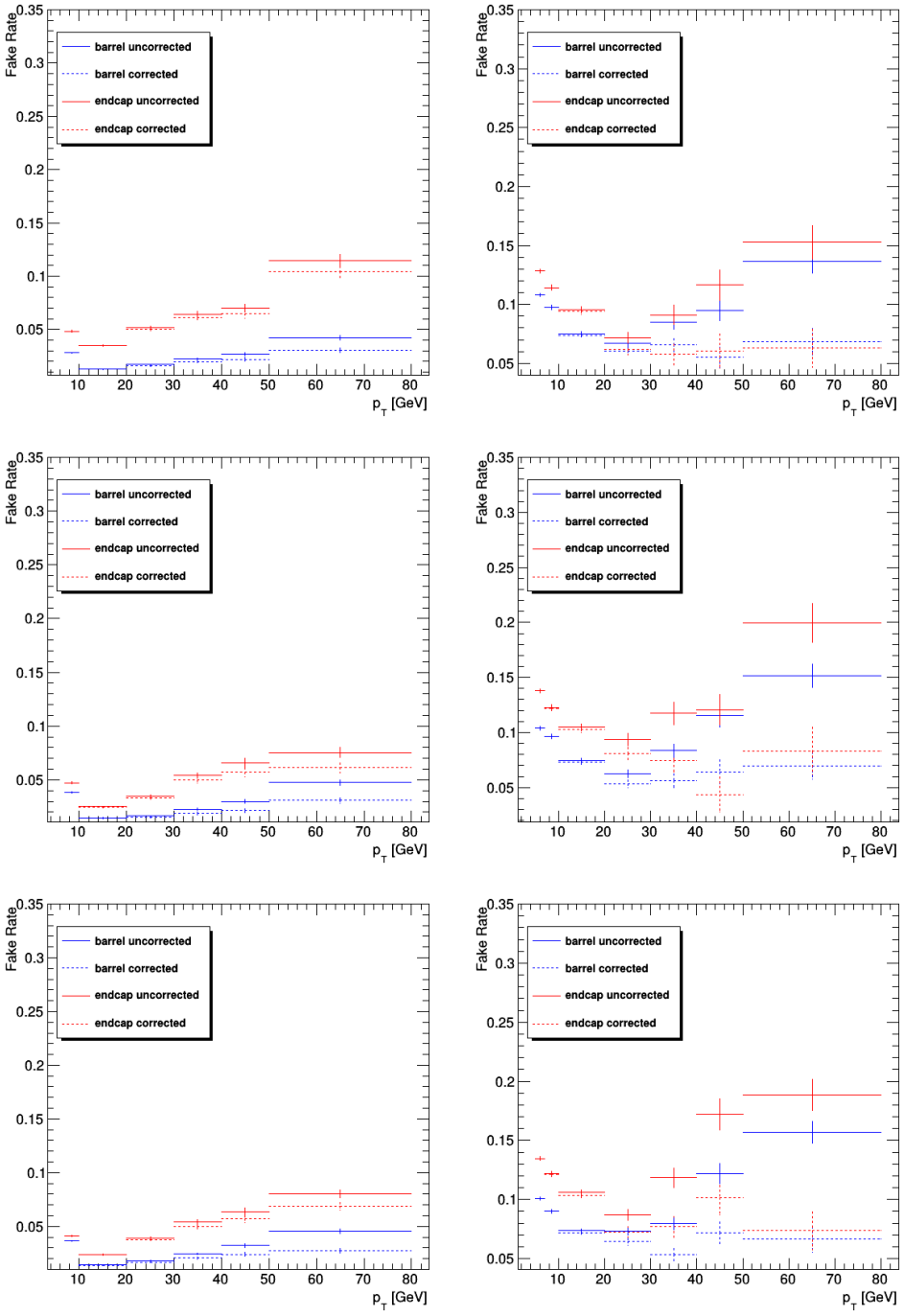


FIGURE 5.9: Fake rates as a function of the probe p_{T} for electrons (left) and muons (right) which satisfy the loose selection criteria, measured in a $\mathrm{Z}(\ell\ell) + \ell$ sample in the 2016 (top), 2017 (middle), and 2018 (bottom) data at 13 TeV. The barrel selection includes electrons (muons) up to $|\eta|$ = 1.479 (1.2). The FRs are shown before (dotted lines) and after (plain lines) removal of the WZ contribution from the simulation.

region, $N^{\text{bkg}}_{\text{2P+2F}}$, by weighting each event in the region with a factor $(\frac{f_i}{1-f_i} + \frac{f_j}{1-f_j})$, where f_i and f_j correspond to the fake ratios of the two loose leptons:

$$N^{\text{bkg}}_{\text{3P+1F}} = \sum (\frac{f_i}{1-f_i} + \frac{f_j}{1-f_j}) N^{\text{bkg}}_{\text{2P+2F}}. \tag{5.3}$$

In case the FR is measured in a sample that has exactly the same background composition as the 2P+2F sample, the difference between the observed number of events in the 3P+1F sample and the expected background predicted from the 2P+2F sample would solely amount to the small WZ and Z+γ contribution. Instead, large differences arise because the FRs used in Eq. 5.3 do not properly account for the background composition of the 2P+2F control sample. In particular, the difference observed in Fig. 5.8 between the 3P+1F distribution and the expectation from 2P+2F in the channels with loose electrons ($4e$ and $2\mu 2e$), especially in the low mass region, is due to the missing contribution of photon conversions (confirmed explicitly by the simulation). This means that the FR calculation does not properly account for the background composition of the 2P+2F sample. More precisely, the expected reducible background in the signal region is given by the sum of two terms:

* a 2P+2F component, obtained from the number of events observed in the 2P+2F control region, $N_{\text{2P+2F}}$, by weighting each event in that region with the factor $\frac{f_i}{1-f_i}\frac{f_j}{1-f_j}$, where f_i and f_j correspond to the fake ratios of the two loose leptons;

* a 3P1F component, obtained from the difference between the number of observed events in the 3P+1F control region, N_{3P1F}, and the expected contribution from the 2P+2F region and ZZ processes in the signal region, $N^{\text{ZZ}}_{\text{3P1F}} + N^{\text{bkg}}_{\text{3P1F}}$. The $N^{\text{bkg}}_{\text{3P1F}}$ is given by Eq. 5.3 and $N^{\text{ZZ}}_{\text{3P1F}}$ is the contribution from ZZ which is taken from the simulation. The difference $N_{\text{3P1F}} - N^{\text{bkg}}_{\text{3P1F}} - N^{\text{ZZ}}_{\text{3P1F}}$, which may be negative, is obtained for each (p_T, η) bin of the failing lepton, and is weighted by $\frac{f_i}{1-f_i}$, where f_i denotes the fake rate of this lepton. This 3P+1F component accounts for the contribution of reducible background processes with only one fake lepton (like WZ events) and for the contribution of other processes (e.g. photon conversions) that are not properly estimated by the 2P+2F component.

Therefore, the full expression for the prediction can be symbolically written as:

$$N^{\text{bkg}}_{\text{SR}} = \sum \frac{f_i}{(1-f_i)}(N_{\text{3P+1F}} - N^{\text{bkg}}_{\text{3P+1F}} - N^{\text{ZZ}}_{\text{3P+1F}}) + \sum \frac{f_i}{(1-f_i)}\frac{f_j}{(1-f_j)} N_{\text{2P+2F}}. \tag{5.4}$$

Previous equation can be condensed in:

$$N^{\text{bkg}}_{\text{SR}} = (1 - \frac{N^{\text{ZZ}}_{\text{3P+1F}}}{N_{\text{3P+1F}}}) \sum_{j}^{N_{\text{3P+1F}}} \frac{f^j_a}{1-f^j_a} - \sum_{i}^{N_{\text{2P+2F}}} \frac{f^i_3}{1-f^i_3}\frac{f^i_4}{1-f^i_4}. \tag{5.5}$$

Different contributions mentioned in Eqs. 5.4 and 5.5 are shown in Fig. 5.10 for the 2017 samples, while similar plots for the 2016 and 2018 datasets are reported in Appendix C. The contribution of the 3P+1F component is positive for final states where the Z_2 candidate is made of two electrons, and amounts to typically 30% of the total predicted background. Considering channels with loose muons (4μ and $2e2\mu$) coming from the Z_2, the 3P+1F sample is rather well described by the prediction from 2P+2F, as it is visible in Fig. 5.8, and the 3P+1F component is mainly driven by statistical fluctuations in the 3P+1F sample, which are larger than the expectation from WZ production.

Table 5.4 shows the expected number of events in the signal regions from the reducible background processes at 13 TeV for each considered final state and for all three years using

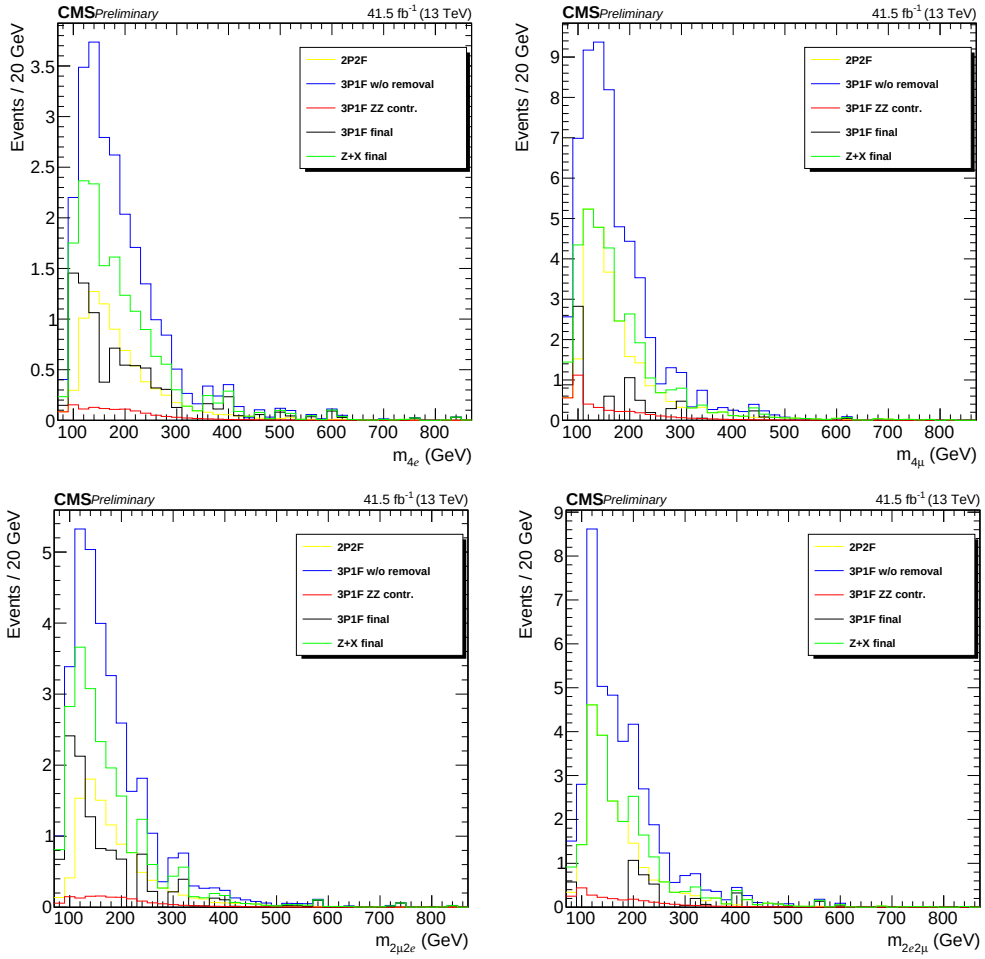


FIGURE 5.10: Four-lepton invariant mass of the different processes which contribute to the 3P+1F estimate of the reducible background in the 2017 dataset for all the considered channels: $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right).

the OS method. The invariant mass distribution of the Z+X events obtained from the combination of the results in the 2P+2F and 3P+1F control samples is shown in Fig. 5.11 for all three years.

Channel	4e	4μ	2e2μ	2μ2e
2016	20.1 ± 6.1	26.9 ± 8.6	25.2 ± 8.0	22.3 ± 6.8
2017	16.2 ± 5.0	32.7 ± 10.3	24.0 ± 7.7	21.3 ± 6.5
2018	25.4 ± 7.7	49.4 ± 15.4	34.2 ± 10.7	33.0 ± 10.0

TABLE 5.4: Contributions of reducible background processes in the signal region predicted from measurements in 2016, 2017 and 2018 data using the OS method. The predictions correspond to 35.9 fb^{-1}, 41.5 fb^{-1}, and 59.7 fb^{-1} of data at 13 TeV, respectively.

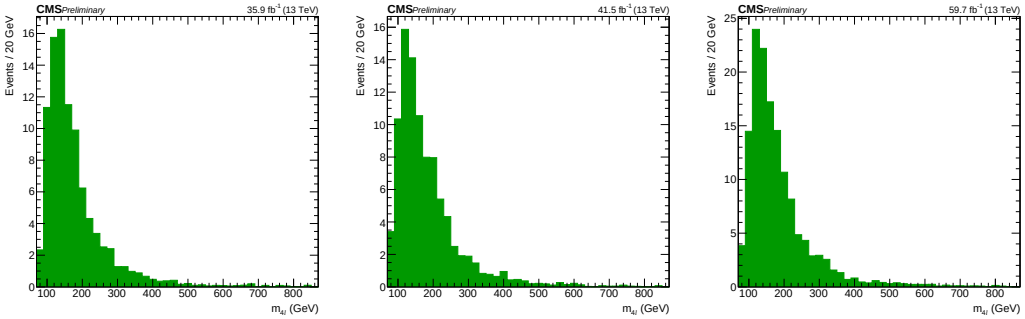


FIGURE 5.11: Invariant mass distribution of the 4ℓ events selected by the combination of the 2P+2F and 3P+1F methods for the 2016 (left), 2017 (middle), and 2018 (right) dataset.

5.3.2 Reducible background estimate with the same sign method

The main difference between the two approaches used to predict the reducible background is that the SS method allows to have an inclusive measurement of the all main reducible backgrounds at the same time.

Both methods exploit the same Z+ℓ CR to estimate the FRs. Then, they rely on a different definition of the Z+$\ell\ell$ CR. Two loose leptons with same sign (to avoid signal contamination) and same flavor passing the cuts on the IP are required to identify a Z+$\ell\ell$ candidate, while no identification requirements are imposed. It is referred to as 2P+2L_{SS} CR. The reconstructed invariant mass of the SS pair of leptons is requested to be in the range $m_{\ell\ell} = [12, 120]$ GeV, the reconstructed four-lepton invariant mass is required to satisfy $m_{4\ell} > 70$ GeV, and the QCD suppression cut is applied. Some adaptations to the kinematic cuts mentioned before are needed in the 2P+2L_{SS} CR: the QCD suppression cut is applied only to the three possible opposite sign pairs, while the alternative pairing check only involves the pairs of opposite sign leptons coming from a Z boson decay. The invariant mass distribution of the 2P+2L_{SS} events is shown in Fig. 5.12 together with the prediction from the MC simulation for channels with loose electrons ($4e$ and $2\mu 2e$) and channels with loose muons (4μ and $2e2\mu$) using the 2017 dataset at 13 TeV. Same results obtained with the 2016 and 2018 datasets are reported in Appendix C. A decent agreement between data and MC distributions for electrons is observed. Considering muons, we have a significant discrepancy but at the end it has no impact on the final estimate which mainly relies on data.

This subset of events is used to derive the inclusive number of reducible background events in the SR, evaluating the probability for the two additional leptons to pass the isolation and identification criteria of the H $\rightarrow 4\ell$ analysis. Starting from the control sample previously described, the final reducible background prediction in the signal region is given by the following expression:

$$N_{SR}^{Z+X} = \left(\frac{OS}{SS}\right) \sum_{i}^{N_{2P2L_{SS}}} f_3^i \times f_4^i \tag{5.6}$$

where

* $N_{2P2L_{SS}}$ is the number of events in the CR;
* f_3 and f_4 are the FR of each additional loose lepton, parameterized as a function of p_T and η;

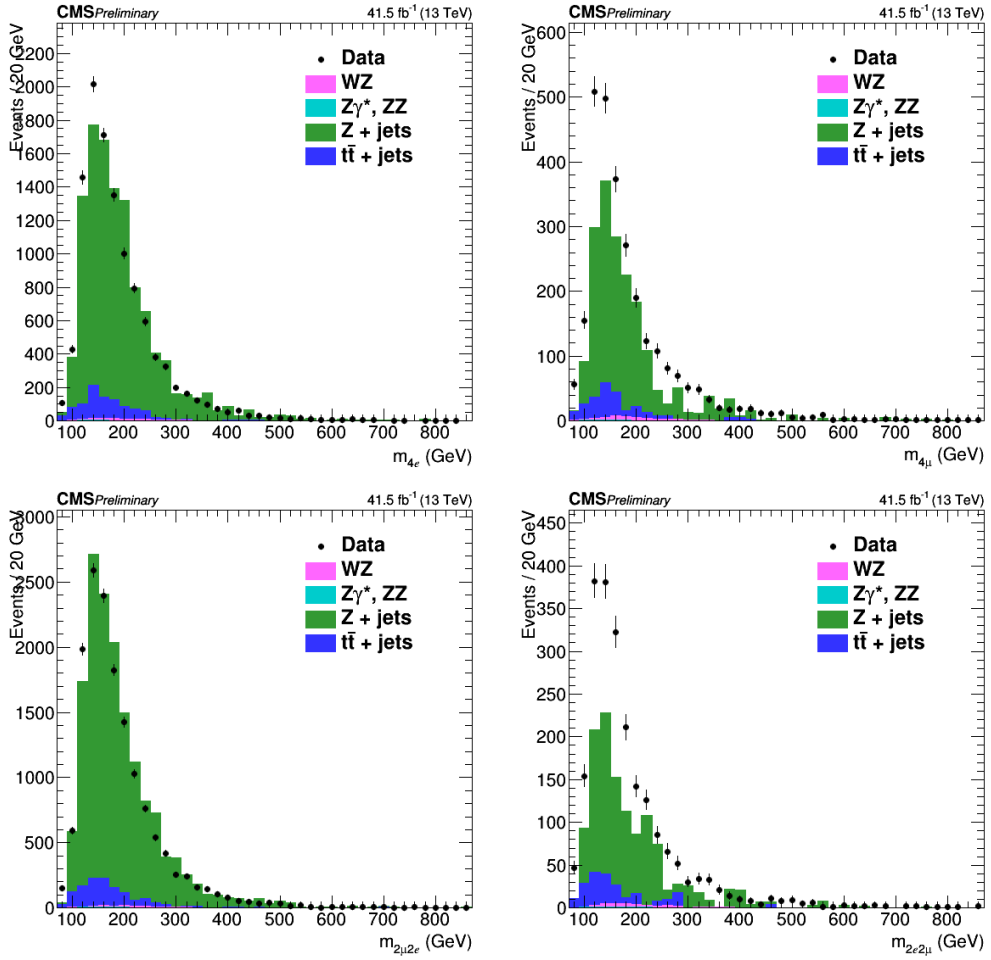


FIGURE 5.12: Invariant mass distribution of the events selected in the 2P+2L$_{SS}$ control sample for all considered final states in 2017 data: $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right).

* $\frac{OS}{SS}$ is a correction factor for the final data-driven estimate which represents the ratio between the number of events in the OS and SS control samples. The differences in rates are estimated using data and are reported for each year in Table 5.5.

Channel	4e	2μ2e	4μ	2e2μ
2016	1.00 ± 0.01	1.00 ± 0.01	1.00 ± 0.03	1.00 ± 0.03
2017	1.01 ± 0.01	1.00 ± 0.01	1.04 ± 0.03	1.01 ± 0.03
2018	1.01 ± 0.01	1.00 ± 0.01	1.03 ± 0.02	1.03 ± 0.02

TABLE 5.5: The $\frac{OS}{SS}$ ratios used to estimate the number of Z+X events with the SS method for each final state in all three years.

Fake Rate determination (SS method) The approach used to determine the lepton FRs is very similar to the one described for the OS method. Samples of Z$(\ell\ell) + e$ and Z$(\ell\ell) + \mu$ events are selected in the same way except that the mass window around the nominal Z

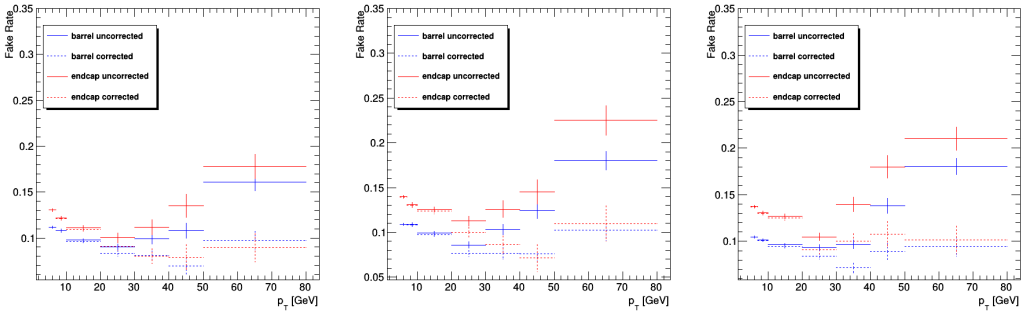


FIGURE 5.13: Fake rates as a function of the probe p_{T} for muons which satisfy the loose selection criteria, measured in a $\mathrm{Z}(\ell\ell)+\mu$ sample in the 2016 (left), 2017 (middle), and 2018 (right) data at 13 TeV. The barrel selection includes muons up to $|\eta|$ = 1.2. The FRs are shown before (dotted lines) and after (plain lines) the removal of the WZ contribution from the simulation.

mass is set to $m_{\ell_1\ell_2} = [40, 120]$ GeV as in the SR, identifying the SS phase space. Given that the cut on the missing transverse energy at 25 GeV leaves a nonnegligible contamination of real leptons from WZ events at high p_{T}, this contribution is subtracted from the final FR values using the estimation given by the simulation. The muon FRs are shown in Fig. 5.13 for both muons in the barrel ($|\eta| < 1.2$) and in the endcap ($|\eta| > 1.2$) region.

Regarding electrons, events containing an asymmetric conversion of a photon, where one low p_{T} leg is not identified, contribute to the $\mathrm{Z}(\ell\ell)+e$ sample that is used to measure the electron FR. While the requirement $|m_{\ell_1\ell_2} - m_{\mathrm{Z}}| < 7$ GeV in the OS phase space largely suppresses FSR of photons radiated off the lepton legs, the rate of this process is large in the SS phase space. The challenging point is that the relative fraction of FSR conversions is not the same in the SS phase space and in the CR where the FRs are applied. A correction accounting for this difference must be applied to the FR measured within the SS phase space, in order to obtain average FR that are appropriate for the control sample. To determine this correction, different FR measurements are performed by applying different cuts on $m_{\ell_1\ell_2}$ or on $m_{\ell_1\ell_2 e}$ in the subset of $\mathrm{Z}(\ell\ell)+e$ events. The requirement $|m_{\ell_1\ell_2} - m_{\mathrm{Z}}| < 7$ GeV, applied also in the computation of the FR with the OS method, identify a subset of events where a minimal amount of conversions from FSR photons is ensured. In addition, a sample which is maximally enriched in FSR conversions by the cut $|m_{\ell_1\ell_2 e} - m_{\mathrm{Z}}| < 5$ GeV is defined, as well as samples with an intermediate contamination from FSR conversions ($m_{\ell_1\ell_2} = [40, 120]$ GeV and $m_{\ell_1\ell_2} = [60, 120]$ GeV). In each sample, the FR and the average value of the expected inner missing hits, referred to as $N_{miss.hits}$), are determined in several bins of p_{T} and η for the loose electron. The $N_{miss.hits}$ variable is very useful to tag conversions. In fact, both the measured FRs and the average $< N_{miss.hits} >$ are expected to grow linearly with the fraction of conversions in a given bin. Hence, a linear dependence of the FR with respect to $< N_{miss.hits} >$ is expected. This is demonstrated in results produced at 13 TeV with 35.9 fb^{-1}data and is shown in Fig. 5.14. Linear fits are made in each bin of p_{T} and η, which relate the FR to $< N_{miss.hits} >$. Finally, considering the loose electrons in the Z+$\ell\ell$ CR with their p_{T} and η and the $< N_{miss.hits} >$ measured in each bin, the information from this fit can be used to determine the average FR in any other CR by simply determining the average number of missing hits. Examples of the corrected FRs obtained from this procedure are shown in Fig. 5.15 together with the FRs measured in the SS phase space.

The correction for the actual fraction of conversions which is present in the control sample lowers the FRs with respect to what is measured in the SS phase space. The determination of these corrected FRs mostly suffers from the limited statistics of the control

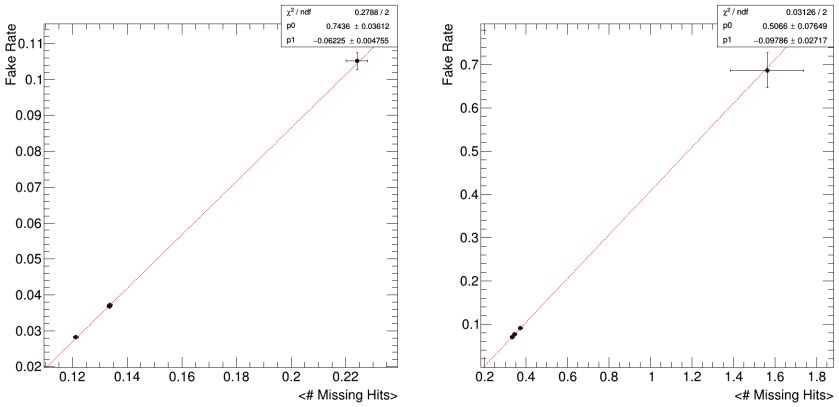


FIGURE 5.14: Examples of the correlation between the FRs and the number of missing hits in the pixel detector for electrons with a p_{T} of [7,10] GeV in the barrel (left) and for electrons a p_{T} of [30,40] GeV in the endcaps (right). Each dot shows the measurements made in a given Z+e sample using the 2016 dataset.

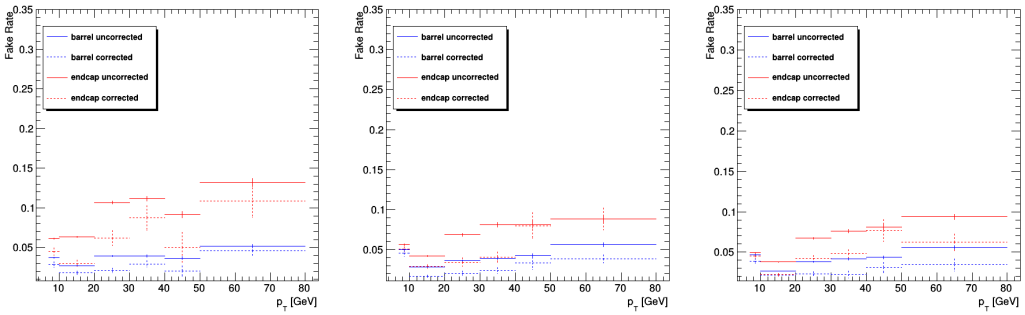


FIGURE 5.15: Average FRs to be applied to the 2P2L SS control sample (plain line), compared to the FRs measured in the SS phase space (dotted line), for electrons in the barrel (blue) and in the endcaps (red). Results are produced at 13 TeV using the 2016 (left), 2017 (middle), and 2018 (right) datasets.

sample, which translates into a large uncertainty on $< N_{miss.hits} >$.

Fake Rate application (SS method) The application of the FRs to estimate the reducible background in the SR is performed in the 2P+2L$_{\mathrm{SS}}$ control sample presented above. The invariant mass distribution of the Z+X events obtained from 2P+2L$_{\mathrm{SS}}$ control sample is shown in Fig. 5.16 for all three years. Table 5.6 shows the expected number of events in the SR from the reducible background processes at 13 TeV for each considered final state and for all three years using the SS method. Considering very large uncertainties, results from the two methods are found to be in agreement in all three years, except for the $4e$ final state. This is due to the different treatment of electrons coming from photon conversions, as already mentioned.

5.3.3 Reducible background estimate per category

In the context of STXS measurements, a dedicated estimate of the reducible background in each stage 1.2 production bin is needed and it has been developed for the full Run 2 analysis for the first time. At the time of the discovery, only an inclusive measurement was provided.

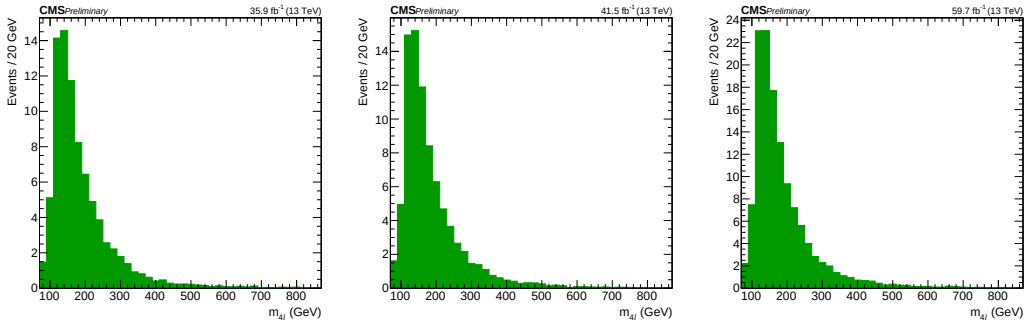


FIGURE 5.16: Invariant mass distribution of the 4ℓ events selected in the 2P+2L$_{\mathrm{SS}}$ control sample for the 2016 (left), 2017 (middle), and 2018 (right) dataset.

Channel	$4e$	4μ	$2e2\mu$	$2\mu2e$
2016	13.0 ± 5.4	29.7 ± 9.1	24.8 ± 7.6	16.7 ± 7.0
2017	10.9 ± 4.0	33.6 ± 10.3	26.3 ± 8.1	14.7 ± 5.5
2018	16.0 ± 5.9	52.2 ± 15.8	37.4 ± 11.4	23.3 ± 8.5

TABLE 5.6: The contribution of reducible background processes in the SR predicted from measurements performed using the SS method in the 2016, 2017, and 2018 datasets. The predictions correspond to $35.9\ \mathrm{fb}^{-1}$, $41.5\ \mathrm{fb}^{-1}$, and $59.7\ \mathrm{fb}^{-1}$ of data at 13 TeV, respectively.

In the following, the approach adopted for this purpose and the possible improvements in view of the LHC Run 3 are going to be described.

Fake Rate determination per category The FRs are currently evaluated inclusively in the analysis. Average FRs are used to evaluate Z+X yields in each STXS category instead of dedicated ones. Given that the VBF category can be particularly affected by this approach, a detailed study of the FR variation in different Z+ℓ phase spaces designed to mimic the VBF category has been performed using the 2018 data. However, a realistic design of the VBF phase space is not completely reproducible in the three-lepton CR without exploiting the information given by the kinematic discriminants. In order to identify events targeting the VBF phase space, hereafter referred to as VBF-like categories, the following requirements on jets are applied:

- an angular separation between the additional lepton and each jet larger than 0.4, otherwise the jet is discarded;
- the presence of two or three jets and at most one b-tagged jet OR at least four jets without b-tagged jets;
- an angular separation between the pair of leading jets larger than 0.5 and a dijet invariant mass greater than 450 GeV.

Futhermore, FRs in the Z+ℓ CR with zero, one, or two jets have been studied as well as the FRs in the phase space complementary to the VBF-like one. The FR curves obtained in each category for both electrons and muons in the barrel and endcap region using SS and OS methods are shown in Fig. 5.17. While the curve in the region complementary to the VBF-like space is perfectly in agreement with the inclusive one, as a result of the

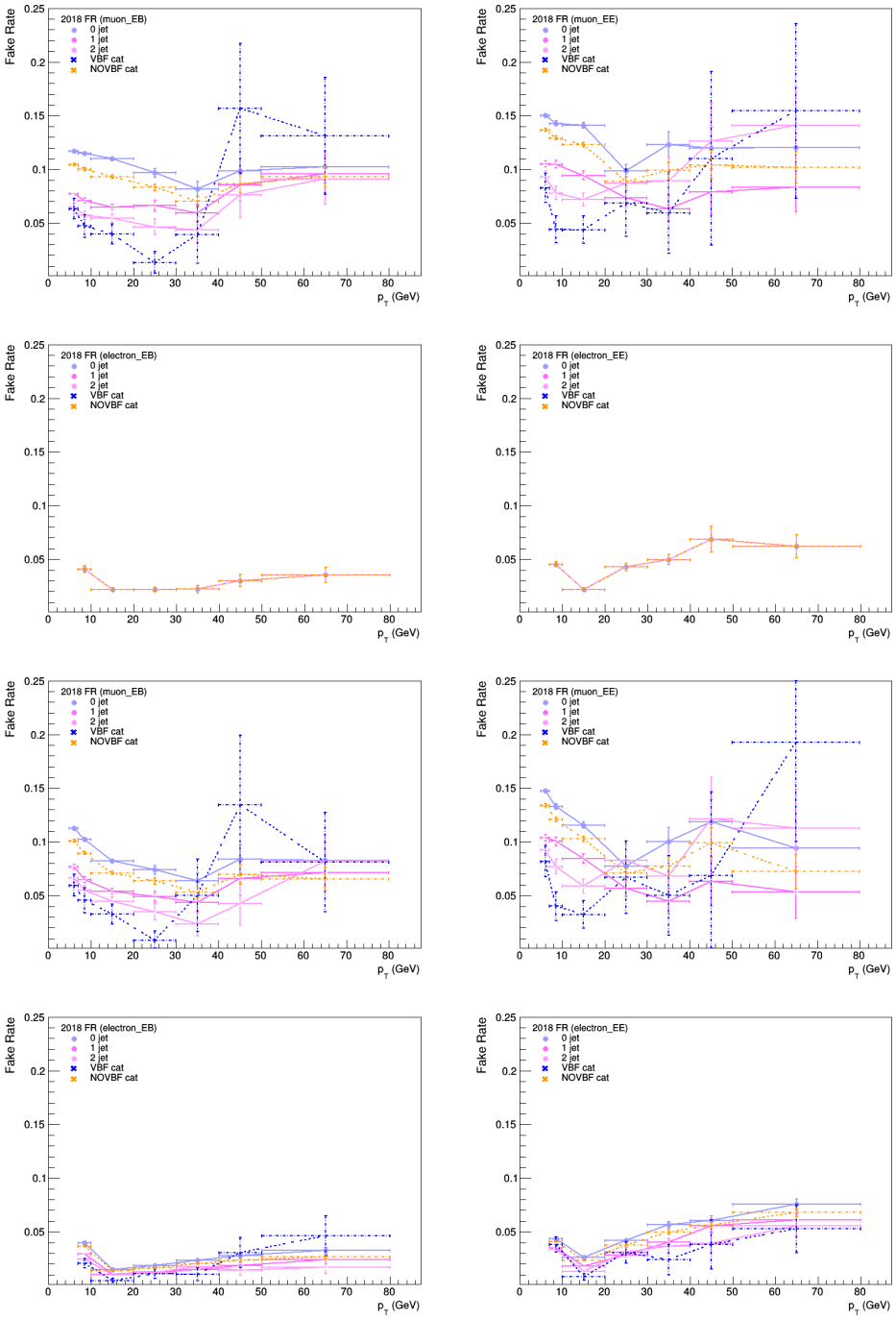


FIGURE 5.17: Fake rate curves in the 0/1/2-jet, VBF-like, and non VBF-like categories for muons (rows 1 and 3) and electrons (rows 2 and 4) in the barrel (left) and endcap (right) region obtained using SS method (rows 1 and 2) and OS method (rows 3 and 4).

combined contribution of the 0/1/2-jet categories, the FR variation in the 0/1/2-jet and VBF-like categories is significant, especially for muons. As a consequence, large discrepancies are observed in the final estimated yields in the VBF categories. The Z+X yield estimate is performed by using these newly defined in the Z+2ℓ CR, where we can use the real event categorization differently from the Z+1ℓ CR. Results are reported in Table 5.7.

	Inclusive FR				Dedicated FR			
Category	$4l$	4μ	$4e$	$2e2\mu$	$4l$	4μ	$4e$	$2e2\mu$
VBF_1j	1.05	0.39	0.15	0.51	0.66	0.19	0.11	0.37
VBF_2j	0.88	0.30	0.10	0.49	0.05	0.01	0.02	0.03
VBF_2j_mjj_350_700_2j	0.04	0.03	0.00	0.01	0.01	0.00	0.00	0.00
VBF_2j_mjj_GT700_2j	0.02	0.01	0.00	0.00	0.09	0.02	0.04	0.03
VBF_2j_mjj_GT350_3j	1.00	0.44	0.10	0.45	0.00	0.00	0.00	0.00
VBF_GT200_2J	0.02	0.02	0.00	0.00	0.02	0.00	0.00	0.01
Inclusive VBF	3.01	1.19	0.35	1.46	0.83	0.22	0.17	0.44

TABLE 5.7: Comparison between combined Z+X yields in the VBF-like categories using inclusive and dedicated FRs.

In principle, the uncertainty on the background composition of the CR was designed to cover for this effect, but considering that the effect seems very large, the possible impact on the analysis has been checked explicitly. An extreme situation has been studied by assuming that we have no knowledge at all about the behaviour of the Z+X background in the VBF-like categories. This assumption is reflected in the choice of inflating the Z+X uncertainty for VBF-like categories by 100% in the statistical analyis, which will be better explained in Chapter 6. Then, the signal strengths, both expected and observed, obtained using the standard Z+X uncertainty and inflating it in the VBF categories are compared. Results are presented in Table 5.8, showing that the expected uncertainty does not change and the observed signal strengths change only very slightly well within the uncertainties. In conclusion, the current approach based on inclusive FRs assumed identical for all the production categories is not correct because FRs are indeed different for VBF-like categories. Nevertheless, the final impact on the analysis is found to be not significant due to the kinematic discriminant used in the fit and the fact that Z+X yields are rather insignificant. The current strategy is kept for this analysis and more effort to have a dedicated Z+X estimate for each category will be invested in view of the future Run 3 analysis.

	Inflated Z+X uncertainty		Nominal Z+X uncertainty	
	Expected	Observed	Expected	Observed
$\mu_{\mathrm{t\bar{t}H,tH}}$	$1.00\ ^{+1.36}_{-0.78}$	$0.23^{+0.95}_{-0.23}$	$1.00\ ^{+1.36}_{-0.78}$	$0.22^{+0.95}_{-0.22}$
μ_{WH}	$1.00\ ^{+2.01}_{-1.00}$	$1.68^{+1.76}_{-1.44}$	$1.00\ ^{+2.01}_{-1.00}$	$1.71^{+1.79}_{-1.71}$
μ_{ZH}	$1.00\ ^{+8.33}_{-1.00}$	$0.00^{+5.24}_{-0.00}$	$1.00\ ^{+8.33}_{-1.00}$	$0.00^{+5.44}_{-0.00}$
μ_{VBF}	$1.00\ ^{+0.56}_{-0.46}$	$0.54^{+0.51}_{-0.41}$	$1.00\ ^{+0.56}_{-0.46}$	$0.56^{+0.50}_{-0.41}$
$\mu_{\mathrm{ggH,b\bar{b}H}}$	$1.00\ ^{+0.16}_{-0.14}$	$1.02^{+0.15}_{-0.13}$	$1.00\ ^{+0.16}_{-0.14}$	$1.03^{+0.15}_{-0.13}$

TABLE 5.8: Comparison between the best fit values and $\pm 1\sigma$ uncertainties for the expected and observed signal strength modifiers, inflating Z+X uncertainties in VBF categories by 100% and keeping nominal Z+X uncertainties.

Fake Rate application per category In order to provide a dedicated Z+X estimate per category, the inclusive FRs are applied separately for each of the stage 1.2 bins defined in Section 4.3.3 in the mass range of the statistical analysis $m_{4\ell}$ = [105,140] GeV and the event yields expected from the Z+X background in the SR are calculated for each final state. As an example, the Z+X yield in the $4e$ final state for the full mass range $m_{4\ell}$ > 70 GeV in 2018 data is reported in Table 5.9. Details about the different contributions to the total error reported here, obtained with a quadrature sum of each source of uncertainty, are provided in Section 5.3.5.

Category	exp. $N^{\mathrm{SR}}_{\mathrm{Z+X}}$	(stat.)	(syst. - FR)	(syst - Bkg comp)	Tot. unc.
ggH_0J_PTH_0_10	0.35	$\pm$0.02	$\pm$0.07	$\pm$0.11	$\pm$0.13
ggH_0J_PTH_10_200	6.75	$\pm$0.08	$\pm$1.40	$\pm$2.03	$\pm$2.48
ggH_1J_PTH_0_60	2.18	$\pm$0.04	$\pm$0.44	$\pm$0.65	$\pm$0.79
ggH_1J_PTH_60_120	1.51	$\pm$0.04	$\pm$0.32	$\pm$0.45	$\pm$0.56
ggH_1J_PTH_120_200	0.53	$\pm$0.02	$\pm$0.12	$\pm$0.16	$\pm$0.20
ggH_2J_PTH_0_60	0.65	$\pm$0.02	$\pm$0.13	$\pm$0.20	$\pm$0.23
ggH_2J_PTH_60_120	0.74	$\pm$0.03	$\pm$0.16	$\pm$0.22	$\pm$0.27
ggH_2J_PTH_120_200	0.31	$\pm$0.02	$\pm$0.07	$\pm$0.09	$\pm$0.12
ggH_PTH_200	0.45	$\pm$0.02	$\pm$0.11	$\pm$0.14	$\pm$0.17
ggH_VBF	0.28	$\pm$0.02	$\pm$0.06	$\pm$0.08	$\pm$0.10
VBF_1j	0.47	$\pm$0.02	$\pm$0.09	$\pm$0.14	$\pm$0.17
VBF_2j	0.30	$\pm$0.02	$\pm$0.06	$\pm$0.09	$\pm$0.11
VBF_2j_mjj_350_700_2j	0.02	$\pm$0.00	$\pm$0.00	$\pm$0.01	$\pm$0.01
VBF_2j_mjj_GT700_2j	0.02	$\pm$0.002	$\pm$0.003	$\pm$0.01	$\pm$0.01
VBF_2j_mjj_GT350_3j	0.40	$\pm$0.02	$\pm$0.09	$\pm$0.12	$\pm$0.15
VBF_GT200_2J	0.03	$\pm$0.01	$\pm$0.01	$\pm$0.01	$\pm$0.01
VH_Had	0.57	$\pm$0.02	$\pm$0.12	$\pm$0.17	$\pm$0.21
VBF_rest_VH	0.10	$\pm$0.01	$\pm$0.02	$\pm$0.03	$\pm$0.04
VH_lep_0_150	0.13	$\pm$0.01	$\pm$0.03	$\pm$0.04	$\pm$0.05
VH_Lep_GT150	0.01	$\pm$0.00	$\pm$0.00	$\pm$0.00	$\pm$0.00
ttH_Lep	0.04	$\pm$0.01	$\pm$0.01	$\pm$0.01	$\pm$0.02
ttH_Had	0.19	$\pm$0.01	$\pm$0.04	$\pm$0.06	$\pm$0.07
Inclusive	16.0	$\pm$0.12	$\pm$3.34	$\pm$4.8	$\pm$5.85

TABLE 5.9: Number of Z+X events expected for the stage 1.2 categories in the SR in the $4e$ final state considering a mass window $m_{4\ell}$ > 70 GeV as predicted from the SS measurement in the 2P+2L$_{\mathrm{SS}}$ control sample in 2018 data corresponding to an integrated luminosity of 59.7 fb^{-1}. The total uncertainty associated to the SS measurement is shown in the table together with the splitting in its three contributions: statistical error and both systematic errors due to fake rate variation and background composition.

5.3.4 Modelling of the Z+X background

Having presented both methods used to get the final estimate of the Z+X contribution to the total background in the H $\rightarrow 4\ell$ analysis in terms of yields, we can now discuss the final modelling of the Z+X events in view of the statistical analysis also in terms of shape. At first, we will focus on the combination of the SS and OS methods showing inclusive results for each considered final state. Then, a discussion oriented to the estimate per-category will be presented with a description of the $m_{4\ell}$ shape in each category.

Combination of the OS and SS methods

Two independent methods for the reducible background estimate have been presented in Sections 5.3.1 and 5.3.2. Table 5.10 shows the summary of the yields obtained with each method and associated uncertainties for all three years together with results given by the combination of the two approaches. Predictions of the two methods are combined assuming no correlation between the uncertainties, because the measurements are performed in independent CRs and take into account partially different sources of systematic uncertainty. Combined mean values are obtained by weighting the individual means according to the corresponding variances. The $m_{4\ell}$ distribution is described by using a Landau function in each considered final state normalized to the total yield given by the combination of the two methods: $4e$, 4μ, and $2e2\mu$.

2016	$4e$	4μ	$2e2\mu$
Method OS	$20.2 \pm 6.2_{\text{tot.}}$	$27.0 \pm 8.6_{\text{tot.}}$	$47.7 \pm 10.5_{\text{tot.}}$
Method SS	$13.1 \pm 5.5_{\text{tot.}}$	$29.6 \pm 9.0_{\text{tot.}}$	$41.5 \pm 10.3_{\text{tot.}}$
Combined	16.1	28.2	44.4
2017	$4e$	4μ	$2e2\mu$
Method OS	$16.2 \pm 5.0_{\text{tot.}}$	$32.7 \pm 10.3_{\text{tot.}}$	$45.7 \pm 10.2_{\text{tot.}}$
Method SS	$10.9 \pm 4.1_{\text{tot.}}$	$33.4 \pm 10.3_{\text{tot.}}$	$40.8 \pm 9.7_{\text{tot.}}$
Combined	13.0	33.1	43.1
2018	$4e$	4μ	$2e2\mu$
Method OS	$25.4 \pm 7.7_{\text{tot.}}$	$50.1 \pm 15.5_{\text{tot.}}$	$67.7 \pm 14.8_{\text{tot.}}$
Method SS	$16.1 \pm 5.9_{\text{tot.}}$	$51.9 \pm 15.8_{\text{tot.}}$	$60.7 \pm 14.2_{\text{tot.}}$
Combined	19.4	50.7	63.9

TABLE 5.10: Summary of the results given by the two methods for the prediction of the contribution of reducible background processes in the signal region for all three years. Symmetric individual uncertainties for the two methods are shown, while the treatment of the error on the combined value is explained in the next section. Results are given for an integrated luminosity of 35.9, 41.5, and 59.7 fb^{-1} in 2016, 2017, and 2018 data, respectively.

Shapes in each STXS stage 1.2 bin

In order to extract the shape of the $m_{4\ell}$ distribution for the reducible background used in the final analysis, shapes for each category and each final state have been studied in the mass range [70, 300] GeV in both SS and OS methods using the 2016 dataset. Then the standard [105,140] GeV window is used in the final analysis.

The study is focused on the SS method because of the better statistics available. Contrary to the SS method, the OS method combines several contributions to get the final Z+X estimate, leading to higher statistical fluctuations especially in the low statistics categories. A fit to Landau function is performed to provide the $m_{4\ell}$ shape in each category separately for the $4e$, 4μ, $2\mu2e$, and $2e2\mu$ final states, as shown in Fig. 5.18 for the 2016 dataset. The consistency of the approach has been checked also in the 2017 and 2018 datasets. Additional plots are reported in Appendix C. The study have been performed using a range wider than the one used in the statistical analysis to simplify the extraction of the Landau shape. Since

not all categories are populated enough, the minimum number of events needed to extract the shape for a single category has been evaluated. In the very low populated categories with less than fifty events selected in the mass window, shapes obtained from the inclusive distributions in each final state are used.
The comparison between different categories shows that for most of the categories the bulk of the distribution is very close to the H peak, but for some of them it is well above the mass range for the H analysis. Consequently, only a small tail of the distribution lies is in the range around 125 GeV. For this reason, the inclusion of an additional shape uncertainty associated to the Z+X modelling can be considered in the future. In the current analysis, given the very low number of events in such categories, the possible effect of a shape uncertainty corresponding to a shift of the distribution to the left which can cause a variation of the background yield in the H mass window is not taken into account.

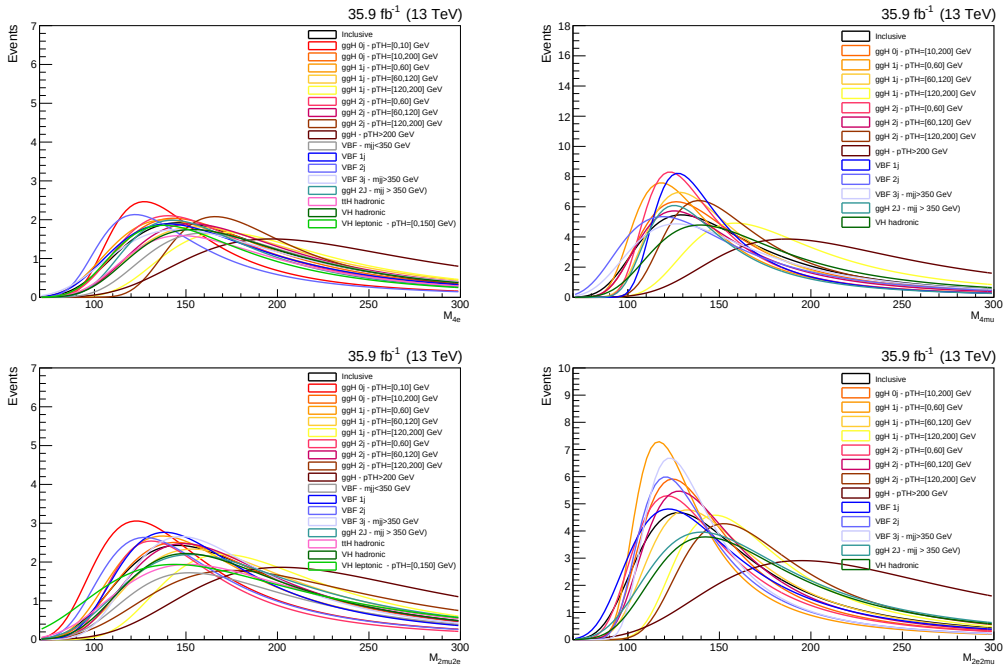


FIGURE 5.18: Shapes of the $m_{4\ell}$ distribution for the reducible background in all the considered categories in the $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right) final states using the SS method in the 2016 dataset. Shapes are normalized to the same number of events.

On the one hand, the results from the two methods are found to be more or less identical in the 4μ and $2e2\mu$ final states, as shown in Fig. 5.19. On the other hand, some differences are visible in the $4e$ and $2\mu 2e$ distributions but they are mainly due to the normalization. As discussed in Section 5.3.2, a discrepancy between yields obtained from the two methods is expected in the final states with electrons. Taking into account the merged $2e2\mu$ and $2\mu 2e$ final states, a fit to Landau function is still performed to obtain the $m_{4\ell}$ shape. In fact, since the muon final state has a larger contribution, the addition of the $2\mu 2e$ component does not distort the single Landau shape.

Looking at the comparison between the $m_{4\ell}$ distribution for the Z+X background obtained from the SS and OS methods in each category separately for all the final states, shape differences between the two methods are found to be not statistically significant. Moreover, it is checked if the shape obtained from the SS method can reasonably describe the Z+X distribution given by the OS method considering each category individually. As an example,

results for one category in each considered final states are shown in Fig. 5.20.

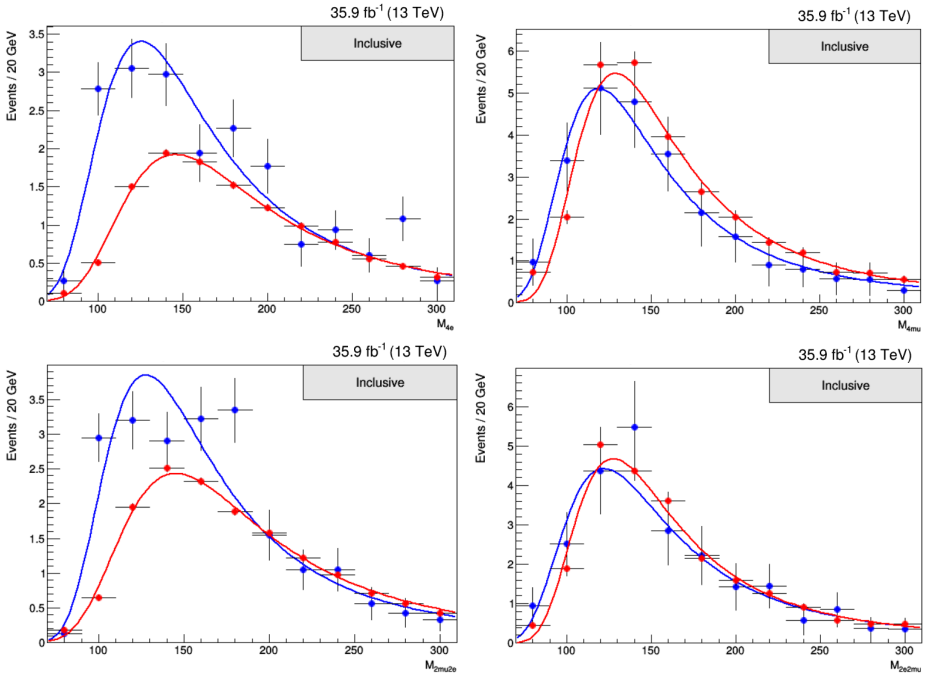


FIGURE 5.19: Comparison between the $m_{4\ell}$ distribution for the Z+X background obtained from the SS (red line) and OS (blue line) methods using the 2016 dataset in the inclusive category for each considered final state: $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right).

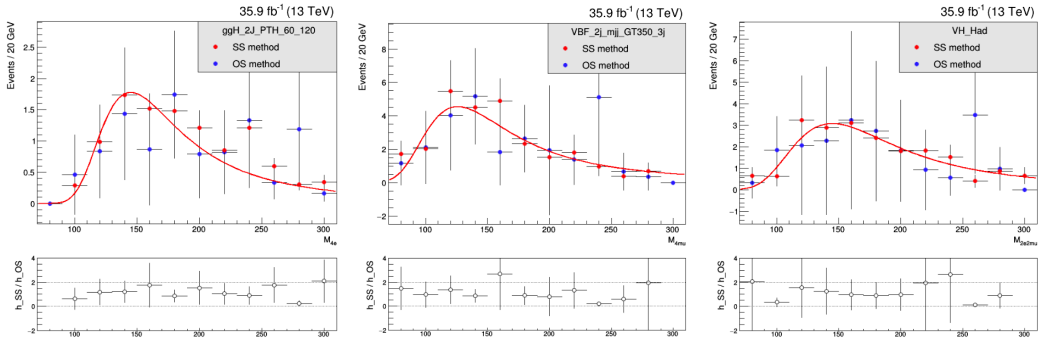


FIGURE 5.20: Comparison between the $m_{4\ell}$ distribution for the Z+X background obtained from the SS (red points) and OS (blue points) methods using the 2016 dataset fitted by the SS shape (red line) for three of the stage 1.2 categories taken as an example: the ggH bin with two jets in the $p_{\mathrm{T}}^{\mathrm{jet}}$ range [60,120] GeV in the $4e$ final state (left), the VBF bin with three jets in the 4μ final state (middle), and the VH hadronic bin in the $2e2\mu$ final state (right), respectively. The bottom panel show the ratio between the four-lepton invariant mass obtained from the SS and OS methods.

Given that a Landau shape is found to model in a good way the $m_{4\ell}$ invariant mass obtained from the SS method in different STXS stage 1.2 bins, it is then used in the mass range [105,140] GeV for the final modelling of the background in each category. If very

low populated categories are considered, shapes obtained from the inclusive distributions in the given final state are used. The result of the fit is shown as an example in Fig. 5.21 for one of the STXS stage 1.2 categories. This shape is finally normalized to the yield obtained from the combination of the SS and OS method results. The discrepancies between the two methods are covered by the yield uncertainty which is around 40%. Indeed, a dedicated shape uncertainty is not included in the analysis.

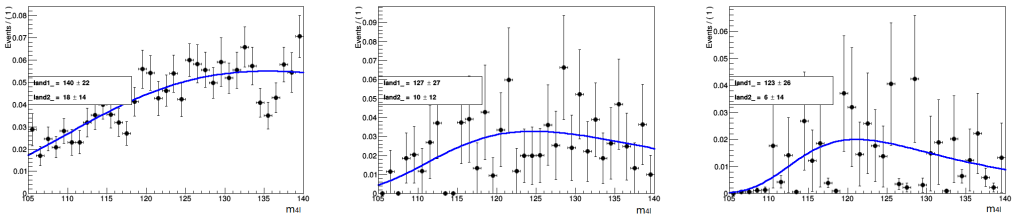


FIGURE 5.21: Shape of the Z+X background events in the mass range [105,140] GeV for three of the stage 1.2 categories taken as an example: the ggH bin with zero jets in the $p_{\mathrm{T}}^{\mathrm{jet}}$ range [10,200] GeV in the $4e$ final state (left), the ggH bin with one jet in the $p_{\mathrm{T}}^{\mathrm{jet}}$ range [60,120] GeV in the 4μ final state (middle), and the VBF bin with three jets in the $2e2\mu$ final state (right).

5.3.5 Uncertainties on the Z+X estimate

One of the highest sources of uncertainties in the analysis comes from the data-driven estimate of the reducible background. Since this uncertainty has a very low impact on all of the measurements, no additional effort has been invested to reduce it for now. The small impact is mainly due to the kinematic discriminants $\mathcal{D}^{\mathrm{kin}}$ described in Section 4.2.1 that allow to achieve a successful discrimination between the signal and the Z+X background. There are three different sources of the uncertainty to be taken into account:

Statistical uncertainty The statistical uncertainty of the methods is driven by the limited size of the samples in the control regions where the FRs are measured (Z+ℓ) and later applied to get the final estimate of Z+X yields (Z+$\ell\ell$). It is typically in the range of 1-10%. Concerning the SS measurement, an additional uncertainty derives from the statistical error bars of the measurement of the average number of missing hits $< N_{miss.hits} >$ in the 2P+2L$_{\mathrm{SS}}$ control sample, which propagate to the electron FRs and, consequently, finally to the estimated number of Z+X events in the SR.

Systematic uncertainty due to fake rate variation A systematic uncertainty given by the variation of the expected yield considering up and down variations in the FR measurement has to be taken into account.

Systematic uncertainty due to the background composition The main source of the total uncertainty associated to the Z+X estimate is due to the different composition of the reducible background processes (DY, $\mathrm{t\bar{t}}$, WZ, Z$\gamma^{(*)}$) in the regions where we measure and where we apply the fake ratios. On the one hand, the OS method corrects for the resulting bias via the 3P+1F component of its prediction. On the other hand, the SS method corrects explicitely the electron FR by using the fraction of photon conversions. To evaluate the sensitivity of the estimate to background composition, the residual bias in the two methods can be estimated by measuring the FRs for individual background processes in the Z+ℓ region in simulated samples. The weighted average of these individual FR values represents the FRs

that we measure in simulation. The exact composition of the background processes in the 2P+2F region where we plan to apply the FRs can be determined from simulation. Thereby, the individual FRs can be reweighted according to the 2P+2F composition. The difference between the reweighed FRs and the average value can be used as an estimate of the uncertainty on the FR measurement. The effect of this systematic uncertainty is propagated to the final estimates and it amounts to about 38% for $4e$, 33% for $2e2\mu$, and 30% for 4μ final state.

The shape systematic uncertainty is not associated to the Z+X estimate. It is absorbed in the statistical and systematic uncertainties on the estimated yields. In order to evaluate the uncertainty on the m_{4l} shape, the difference between the shapes of predicted background distributions for all three channels and between the shapes given by SS and OS methods was checked and was estimated to be roughly in the range of 5-15%. Given that the difference of the shapes slowly varies with m_{4l}, it is taken as a constant versus m_{4l} and is absorbed into the much larger uncertainty on the predicted yield of background events. The final uncertainty associated to the combined Z+X yield is given by the sum in quadrature of these three contributions considering the full mass range of the m_{4l} distribution in the SS method. Table 5.11 shows the Z+X uncertainty for each final state in each year.

Channel	$4e$	4μ	$2e2\mu$
2016	41%	30%	35%
2017	38%	30%	33%
2018	37%	30%	33%

TABLE 5.11: Total uncertainty associated to the Z+X estimate for each final state in all three years.

5.4 Systematic uncertainties

In the context of a scientific analysis, a central concept is that of uncertainty. It reflects the precision of the measurement and is composed of two different contributions. We have a statistical uncertainty related to the amount of data collected by the experiment. More data enhances the precision of a measurement or the reach of a search, so that during each year the goal of the ATLAS and CMS experiments is to record as much data as possible. For this purpose, the increased instantaneous luminosity as well as detector and machine developments are exploited. Differently, we also deal with many other effects arising from differences between the simulation and observation in data, imperfections in the knowledge of the detector response, and uncertainties in the theoretical prediction. All these sources of uncertainty can affect the modelling of signal and background processes that was presented up to now, worsen the prediction of signal and background yields and shapes, and cause migrations between the event categories. They are considered as systematic uncertainties in the final analysis and are modelled as nuisance parameters that are profiled in the maximum likelihood fit described in Section 6.1.

A good estimate of the systematic uncertainties is crucial for a valuable comparison of two experimental results of a measurement and the corresponding theory predictions, especially when the statistical component is not dominant in the overall uncertainty and the two contributions are comparable. In fact, when the amount of data available for a measurement is too small, the sensitivity on the systematic uncertainty is not a big deal so that a too much optimistic or conservative estimate is not dramatic.

The $H \to ZZ^* \to 4\ell$ analysis is located exactly at this border line where the magnitude of the systematic and statistical components are becoming comparable, being an analysis dedicated to precision properties measurement and no more a discovery analysis. Therefore, the content of this part is particularly relevant for the comprehension of final results and the effort oriented to the understanding and evaluation of all possible systematics sources is essential. All significant sources of systematic uncertainties on the measurements of H boson properties are divided into two groups. On the one hand, the experimental uncertainties are related to the imperfect knowledge of the detector, for example the luminosity, the calibrations, efficiencies, and resolutions, or to the data control regions. On the other hand, the theoretical uncertainties are mainly related to the cross sections and generators used for the modelling of the signal and background processes. In the following discussion, we are going to provide a complete overview of the different sources of systematic uncertainty and their effect on the analysis. Instead, the impact of the dominant systematic uncertainties on some of the measurements will be discussed with the final results in Section 6.2.

5.4.1 Experimental uncertainties

The dominant sources of experimental uncertainties originate from the luminosity measurement, the lepton reconstruction and selection efficiency, the lepton and jet scale and resolution, the b tagging efficiency, and the reducible background estimate. In the combination of the three data-taking periods, the ones related to reconstruction and selection efficiency, the lepton scale and resolution, and the b tagging efficiency are considered correlated, while the others are treated as uncorrelated.

* The **measurement of the integrated luminosity** of the dataset is performed with Van der Meer scans [187]. It is affected by a normalization uncertainty which ranges from 2.3 to 2.6% depending on the data-taking period and affects all final states, both signal and background.

* The **lepton reconstruction and selection efficiency** represents another experimental uncertainty common to all final states. All the steps from trigger to IP significance and finally identification and isolation requirements are included in the selection efficiency. The uncertainty ranges from 1 to 2.3% in the 4μ channel and from 11 to 15.5% in the $4e$ channel. This large discrepancy between muons and electrons is mainly due to the low p_{T} measurement. In fact, low mass dimuon resonances are used to measure the efficiency in the low p_{T}^{μ} regions, while in the electron case the efficiency measurement relies solely on the Z boson resonance, resulting in a higher uncertainty in the low p_{T}^{e} region.

* The **lepton momentum scale and resolution** uncertainties modify the signal shape by allowing the corresponding parameters of the double-sided Crystal Ball function to vary. They are estimated from the study of the $Z \to \ell\ell$ invariant mass distribution in data and simulation. Events are classified according to the p_{T} and η of one of the two leptons, determined randomly, and integrated over the other. Then, the distribution of the dilepton mass is fit by a Breit-Wigner parameterization convoluted with a double-sided Crystal Ball function, as described in Section 5.1.3. The scale uncertainty is around 0.04% in the 4μ channel and 0.3% in the $4e$ channel, while the resolution uncertainty is found to be 20% for both channels. The impact on the measurements presented here is found to be negligible compared to the other uncertainty sources. The importance of this uncertainty is related to the measurement of the H boson mass and width.

* The **jet energy corrections** do not impact the signal selection efficiency, as jets are not part of the final state under study, but they play a significant role in the migration of events between the categories. They can also modify the shape of the discriminants, but a dedicated study showed that the effect on the shape is negligible. The uncertainty on the jet energy scale ranges from 1% in the high jet p_{T} range and increases up to 5% in the low jet p_{T} range, while the uncertainty on the jet energy resolution ranges from 1 to 2%. A detailed description of the determination of the jet energy scale and smearing uncertainties can be found in [188]. The impact on the analysis is studied by propagating the uncertainties and estimating the effect on event migration in each of the twenty-two reconstructed categories. Their impact on the inclusive measurements is found to be negligible. However, they represent one of the leading sources of uncertainty in the measurements targeting the VBF and VH production modes and differential cross section measurements, strongly related to the jet kinematics.

* The **b tagging efficiency** introduces a systematic uncertainty around 3% (1%) in the low (high) $p_{\mathrm{T}}^{\mathrm{jet}}$ range. The impact from the category migration on the final measurements is found to be negligible in all categories.

* The **reducible background estimate** represents one of the main source of systematic uncertainty in the analysis. As described in Section 5.3.5, it is mainly due to the background composition and misidentification rate uncertainties and varies between 30% and 45% depending on the final state and category. In the final numbers a component for the statistical limitation of the considered CR is also included. These uncertainties are treated as uncorrelated between different final states. Each of them can be viewed as a separate measurement. Although their magnitude is significant, the impact of this uncertainty on the final measurements is found to be negligible.

5.4.2 Theoretical uncertainties

In the combination of the three data-taking periods, all theoretical uncertainties are treated as correlated. They affect both signal and background estimation and come from the theoretical computations, in particular the renormalization and factorization scale and the choice of the PDF set. The theoretical uncertainties on the overall signal yield are not included in the measurement when cross sections, rather than signal strengths, are extracted. A list of the main contributions to the overall theoretical uncertainty is reported below.

* A **QCD uncertainty** is applied to every signal and irreducible background sample. The uncertainty from the renormalization and factorization scale is determined by varying these scales between 0.5 and 2 times their nominal value while keeping their fraction between 0.5 and 2. It can affect the signal and background yields and can also cause some anti-correlated migrations of events between the categories. An additional uncertainty of 10% given by the use of the NNLO/LO k-factor described in Section 5.2.2 is attached to the ggZZ process. The electroweak corrections for the $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ background prediction are also considered in all measurements.

* The choice of a **set of PDFs** represents a source of uncertainty which is determined independently for different sets of processes grouped by initial state. It is defined, following the PDF4LHC recommendations [189], by taking the root mean square of the variation when using different replicas of the default NNPDF set [190, 191].

* A systematic uncertainty of 2% [16] in the **branching fraction** of $H \to 4\ell$ is applied to the yields of all signal processes.

* The impact of the **modelling of hadronization** and fragmentation processes and the UE is determined using dedicated generator-level samples simulated with the variation of the nominal PYTHIA 8 parameters for the underlying event tune and hadronization scale (see Section 5.1.1). The **UE modelling uncertainty** is determined by varying initial and final state radiation renormalization scales between 0.25 and 4 times their nominal value. The only effect of this additional theoretical uncertainty is related to the migration of signal and background events between categories.

* A dedicated scheme of uncertainties for the ggH and qqH processes is included, according to Ref. [16]. They are considered as uncertainties in the SM predictions in the cross section measurements.
  Nine NPs are introduced in the ggH uncertainty scheme to take into account different sources of uncertainty: the cross section prediction for exclusive jet bins (including the migration between the 0- and 1-jet, as well as between the 1- and $\geq$2-jet bins), the 2-jet and $\geq$3 jet VBF phase space, different $p_{\mathrm{T}}^{\mathrm{H}}$ regions, and the uncertainty in the $p_{\mathrm{T}}^{\mathrm{H}}$ distribution due to missing higher order finite top quark mass corrections.
  In the qqH uncertainty scheme, the uncertainty in the modelling of the $p_{\mathrm{T}}^{\mathrm{H}}$, m_{jj}, and $p_{\mathrm{T}}^{\mathrm{Hjj}}$ distributions, and the number of jets in the event is taken into account. Six NPs account for the migrations of events across m_{jj} boundaries at 60, 120, 350, 700, 1000, and 1500 GeV. Two additional NPs consider the migrations across the $p_{\mathrm{T}}^{\mathrm{H}} = 200$ GeV and $p_{\mathrm{T}}^{\mathrm{Hjj}} = 25$ GeV bin boundaries. Finally, a single source is defined to account for a migration between the zero and one, and the two or more jet bins. In each case, the uncertainty is evaluated by varying the renormalization and factorization scales and recalculating the fractional breakdown of qqH STXS stage 1.2 cross sections.

Chapter 6

Measurements of the Higgs boson properties

The exciting Run 1 led us to the landmark discovery of the H boson with an extremely rich physics program. Then, the successful and productive Run 2 has allowed to gain a deep knowledge about H boson properties, its main production and decay modes, and its couplings to the heaviest quarks and leptons of the third generation. Thanks to the collection of much more data, the reach of the LHC experiments is impressive.
The $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis continuously played a central role in the context of the ATLAS and CMS research, providing quickly interesting results as soon as a new set of data was available. For example, the analysis of the first Run 2 dataset of 2.8 fb^{-1} recorded in 2015 found an evidence for the H boson signal for the first time at $\sqrt{s} = 13$ TeV [192]. The observation of the H boson at 13 TeV happened in the middle of the 2016 data-taking, analyzing a dataset of 12.9 fb^{-1} collected between April and June [193]. A significant goal in the scope of the $\mathrm{H} \to 4\ell$ analysis was represented by the publication of the impressive results related to the full 2016 dataset, corresponding to 35.9 fb^{-1} [48, 95].

At the end of the Run 2 of the LHC, the contribution of the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis in the activity of the CMS experiment was still in the forefront. The first results using the whole Run 2 integrated luminosity, referred to as legacy analysis, were released as preliminary results last year during spring 2019. They provided measurements of the signal strength modifiers and simplified template cross sections [194]. To analyze the total amount of Run 2 data, corresponding to an integrated luminosity of 137 fb^{-1}, significant improvements in many aspects of the analysis have been introduced with respect to the previously reported measurements by the CMS Collaboration.

I joined the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ group exactly at that time in the context of the legacy analysis effort with the goal of finalizing and optimizing the results and finally publishing them as a journal paper. I soon became more involved in the analysis, contributing in many key areas, from the object selection (see Chapter 3), running the analysis selection chain to provide the inputs to all results (see Chapter 4), to the estimate and modelling of the reducible background (see Section 5.3). The final steps of the review of the analysis are ongoing and I had a chance to follow them in detail, contributing to the preparation of the conclusive checks and results required before the internal approval and publication of the work.

In this chapter, a description of the statistical procedure used to extract the H boson properties will be provided. This topic represents the only missing part to understand the final results of this work, which are the outcome of the interplay of all components discussed in the previous chapters. Then, we will finally move to the discussion of the most recent results produced in order to probe the H boson properties in the four-lepton final state to a new precision and assess its compatibility with the predictions for a SM H boson.

6.1 Statistical analysis for the H boson measurements

During the searches for the SM H boson in Run 1, the joint work of the ATLAS and CMS Collaborations and of the LHC Higgs Combination Group put in place a common statistical framework in the context of the Higgs analyses combination [195] in order to coordinate the effort of all groups involved. At that time, the main focus of the search was to look for an excess of data with respect to the expectation given by the SM and to quantify it. Then, having established the observation of the H boson, the main focus shifted from an inclusive search to precision measurements of its properties.

In the following section we are going to describe the statistical methods exploited to glue together all the inputs previously discussed, such as selected events, estimate of backgrounds, and uncertainties, and evaluate the presence or absence of a signal in the observed data to derive the final results. The definition of the method includes the choice of a test statistic for the parameter(s) of interest (POIs) and specifies the treatment of nuisance parameters (NPs) in the construction of the test statistic and in generating pseudo-data. The NPs represent all other parameters in the statistical model except for the parameter(s) of interest.

6.1.1 Likelihood function and statistical interpretation

The process of drawing conclusions about evidence or exclusion of a signal starting from collected data is called *statistical inference*.
First of all, the methodology requires that the information about the modelling of signal, backgrounds, and systematic uncertainties is encoded in mathematical form into a likelihood function. If one assumes that the outcome of a measurement is identified as x, referring to, for example, a number of events or a histogram, the distribution of x is a probability density function which may depend on other parameters $\boldsymbol{\alpha}$ and is denoted as $P(x|\boldsymbol{\alpha})$. If $P(x|\boldsymbol{\alpha})$ is evaluated from the measured data depending on the parameters $\boldsymbol{\alpha}$, then it represents the likelihood function, indicated as $\mathcal{L}(\boldsymbol{\alpha}) \equiv P(x|\boldsymbol{\alpha})$. The term *model* can be used to refer to the full function $P(x|\boldsymbol{\alpha})$ that contains the dependence both on x and $\boldsymbol{\alpha}$. Note that the notation $\boldsymbol{\alpha}$ adopted here generally refers to all parameters of the model for now, both POIs and NPs.
Then, statistical methods are used to translate this information into statistical significances or exclusion limits. We can distinguish between two kinds of measurements: the measurement of continuous physics parameters of a model, such as the production cross sections, the couplings, or the mass, and the comparison of two different models. In the context of the $\mathrm{H} \rightarrow \mathrm{ZZ}^* \rightarrow 4\ell$ analysis in Run 2, we are dealing with a parameter estimation and an approach based on the *maximum likelihood estimation* (MLE) is adopted. This procedure only needs the likelihood function $\mathcal{L}(x|\boldsymbol{\alpha})$, which represents the probability for the observed data in the signal region, x, as a function of the parameter(s) of the model, $\boldsymbol{\alpha}$. The maximum likelihood approach selects the value of the model parameter(s), $\hat{\boldsymbol{\alpha}}$, which maximizes $\mathcal{L}(x|\boldsymbol{\alpha})$ identifying the configuration that makes the probability of observing a given data distribution as large as possible.

The likelihood function of a H boson analysis for a fixed value of the H boson mass can be defined in terms of:

1. the observed data;
2. a global signal strength $\mu = \sigma/\sigma_{\mathrm{SM}}$, where σ_{SM} is the expected SM H boson cross section, as the POI of the model;
3. the signal and background models.

Hereafter, the expected SM H boson signal event yields and the total background event yields will be referred to as s and b, respectively.
Given that the yield expectation is affected by the systematic uncertainties that have been discussed in Section 5.4, every independent source of systematic uncertainty is represented in the statistical model as a *nuisance parameter* θ_i and the entire set of θ_i is collectively denoted as $\boldsymbol{\theta}$. Consequently, the expected signal and background yields will be denoted as $s \equiv s(\boldsymbol{\theta})$ and $b \equiv b(\boldsymbol{\theta})$. Considering that we extract our results within the general scenario of a SM-like H boson (see Section 1.2.2), where the cross sections for all production channels are scaled by the same factor μ and $\mathcal{B}$ are assumed to be preserved, the expected signal yield can be expressed as $\mu \cdot s(\boldsymbol{\theta})$.
Each θ_i is profiled during the minimization procedure, affecting either one or multiple processes and depending on the considered final state or the event category. Two types of nuisances can be identified according to the effect of the considered uncertainty: a *normalization uncertainty* affects only the number of expected events and it is taken into account by scaling the yield in a multiplicative way; a *shape uncertainty* can modify the shape of the discriminating variables described in Section 4.2 and its modelling relies on a set of alternative templates, which is usually governed by one single NP. The two classes of NPs are described by a lognormal distribution and a Gaussian PDF, respectively. Finally, a good understanding of the correlation effects related to the considered NPs is crucial. The same NP is associated to correlated sources of uncertainty, while uncorrelated uncertaintes have a dedicated NP in the likelihood fit.

The possible deviation of a quantity from the input value derived from a priori considerations or separate measurements is expressed as a systematic uncertainty and is represented through a variable $\hat{\theta}_i$ and a PDF $p_i(\hat{\theta}_i|\theta_i)$. The value of $p_i(\hat{\theta}_i|\theta_i)$ describes probability to measure a value $\hat{\theta}_i$, which denotes the estimation of the NP, given the true value θ_i. When constructing the general form of our likelihood function for a given set of data, either actual data or a pseudo-experiment, and the measurements of $\hat{\boldsymbol{\theta}}$, we can write it as the product of the likelihood of the data and the individual PDF p_i for the associated measurements of the NPs:

$$\mathcal{L}(x, \hat{\boldsymbol{\theta}}|\mu, \boldsymbol{\theta}) = \mathcal{L}(x|\mu, \boldsymbol{\theta}) \times \prod_{i=1}^{N_\theta} p_i(\hat{\theta}_i|\theta_i) \tag{6.1}$$

The $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis presented in this work exploits simulation to build the PDFs for the POIs. Therefore, unbinned distributions are used aiming to take the advantage of all the information about the individual values of a quantity x, such as the four-lepton invariant mass. Consequently, the likelihood function is expressed as the product of a Poisson term and the overall probability density function of all observed events e,

$$\mathcal{L}(x|\mu, \boldsymbol{\theta}) = P(x|\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})) \times \prod_{e=1}^{N_E} f(x|\mu, \boldsymbol{\theta}), \tag{6.2}$$

with $f(x|\mu, \boldsymbol{\theta})$ defined as the normalized distributions of the signal and background events:

$$f(x|\mu, \boldsymbol{\theta}) = \frac{\mu \cdot s(\boldsymbol{\theta})}{\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})} f_S(x|\boldsymbol{\theta}) + \frac{b(\boldsymbol{\theta})}{\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})} f_B(x|\boldsymbol{\theta}), \tag{6.3}$$

where $f_S(x|\boldsymbol{\theta})$ and $f_B(x|\boldsymbol{\theta})$ represent the PDF of the variable x for signal and background events. Combining the three equations above, the likelihood function for a given observation x can be written as

$$\mathcal{L}(x, \hat{\boldsymbol{\theta}}|\mu, \boldsymbol{\theta}) = P(x|\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})) \times p(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}). \tag{6.4}$$

In conclusion, the statistical significance of an excess over the background-only hypothesis ($\mu = 0$), is quantified by a test statistic q_0, defined as

$$q_0 = -2\,ln\,\frac{\mathcal{L}(x, \hat{\boldsymbol{\theta}}|0, \hat{\boldsymbol{\theta}}_0)}{\mathcal{L}(x, \hat{\boldsymbol{\theta}}|\hat{\mu}, \hat{\boldsymbol{\theta}})}, \tag{6.5}$$

where the numerator and denominator of the ratio gives information about the background-only expectation and an unconstrained signal-plus-background hypothesis, respectively. The new notation which appears here $(\hat{\boldsymbol{\theta}}, \hat{\mu})$ indicates the values of the parameters $\boldsymbol{\theta}$ and μ that maximize the likelihood in the denominator. Similarly, $\hat{\boldsymbol{\theta}}_0$ refers to the value which maximizes the background-only hypothesis. This test statistic q_0 is often mentioned as the *negative log-likelihood*. Therefore, Eq. 6.5 corresponds to

$$q_0 = -2\,ln\,\Delta\mathcal{L}. \tag{6.6}$$

This strategy allows to measure central values and uncertainty intervals. In order to determine confidence intervals, we know from the Wilks theorem [196] that, considering a model with n parameters of interest, the test statistic q_0 can be approximated by a χ^2 distribution with n degrees of freedom in the limit of a large number of events (asymptotic regime). An approximate 68% and 95% CL interval for the measurement of a single parameter μ is obtained by requiring $q_0 < 1$ and $q_0 < 3.84$, respectively. In general, the results are both quoted as central values with 68% CL intervals, and displayed graphically as scans of $-2\,ln\,\Delta\mathcal{L}$. To estimate the sensitivity of a measurement, we may want to provide expected results for some nominal values of the parameters: in this case, a large number of pseudo-experiments have to be generated together with the determination of their median outcome. However, the Asimov dataset [26] provides a very good approximation in this context, representing one single dataset where the observed rates and distributions are identical to the predictions under the nominal set of NPs.

6.1.2 Measurement methodology

Up to now, we talked about likelihood fits assuming them as one-dimensional fits, which represent the simplest way of measuring, for example, the global signal strength modifier of the H boson signal. In this case, the likelihood function exploits only the information given by the four-lepton invariant mass distribution. Nevertheless, it is possible to build a multi-dimensional likelihood function which contains simultaneously the information of different variables and improves the sensitivity of the model. In the context of this work, the choice of the strategy depends on the kind of measurement, as explained in the following.

A multi-dimensional likelihood fit for the signal strength measurement

The signal extraction for measurements presented in Section 6.2 is performed by using a multi-dimensional fit that relies on two variables: the invariant mass of the four-lepton system ($m_{4\ell}$) and one of the matrix element kinematic discriminants (K_D), that have been described in Section 4.2. The total PDF can be written in the following way:

$$\mathcal{L}_{2D}(m_{4\ell}, K_D) = \mathcal{L}(m_{4\ell})\mathcal{L}(K_D|m_{4\ell}). \tag{6.7}$$

The mass dimension is unbinned because of the small number of expected events in the mass peak and the resolution model used to describe $m_{4\ell}$ has been presented in Section 5.1.3. The inclusion of a second dimension which accounts for the K_D shape and its strong correlation with $m_{4\ell}$ greatly improves the model. A specific kinematic discriminant enters the likelihood depending on the considered production category:

- the decay-only kinematic discriminant $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ is used to separate the H boson signal from the background in most of the subcategories, as described in Eq. (4.1);

- the $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$ discriminant (see Eq. 4.2), sensitive to the VBF production mechanism, is included for the VBF-2jet-tagged subcategories;

- the $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ discriminant (see Eq. 4.3), sensitive to the VH production mechanism, is used for the VH-hadronic-tagged subcategories.

The distribution of the decay-only kinematic discriminant $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ and of the production-dedicated $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$ and $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ discriminants in the mass range $m_{4\ell} = [118, 130]$ GeV are shown in Fig. 6.1 together with their correlation with the four-lepton invariant mass in the mass range $m_{4\ell} = [105, 140]$ GeV.

To properly consider the correlation between the kinematic discriminant and the mass in the fit, two-dimensional histogram templates of $m_{4\ell}$ versus K_D are normalized to one in slices of $m_{4\ell}$. As a consequence, the information provided by the $\mathcal{L}(K_D|m_{4\ell})$ term decouples the strong correlation between the invariant mass and the kinematic discriminant. The 2D templates are created separately for signal and background and for each final state. A different template for every main production mode is built using the simulation of the signal (ggH, VBF, WH, and ZH samples), while ggH templates are used for all other production modes. Then, the templates for the irreducible background are obtained from the simulation of the $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ and $\mathrm{gg} \to \mathrm{ZZ}$ processes, while the ones for the reducible background are derived from the control region using the SS method. The POIs are estimated with their corresponding confidence intervals using a profile likelihood ratio test statistic [26]. The choice of the POIs depends on the specific measurement under consideration. Then, the remaining parameters are treated as NPs. Both experimental and theoretical uncertainties are incorporated via NPs.

At the beginning of this section we mentioned the measurement of the global signal strength as the target of this strategy. However, other similar measurements can be performed using the same framework, as it will be presented in Section 6.2, defining different signal strength modifiers or grouping some of them.

The framework for the fiducial cross section measurement

The kinematic distributions of the SM Higgs boson with m_{H} =125 GeV in different decay channels and production modes can be sensitive to the effects coming from BSM processes, so that H boson properties are not determined by a simple scaling of couplings anymore (see Section 1.2.2). Therefore, the measurement of the cross sections can be performed in a restricted part of the phase space, referred to as *fiducial phase space*, in order to test for similar deviations in the H boson kinematics. Fiducial measurements can be inclusive or differential. This approach allows to minimize the model dependence on the specific model assumed to study the H boson production and properties. In fact, the acceptance and selection efficiency for the $\text{H} \to 4\ell$ decays can be significantly different between various production mechanisms of the H boson and different exotic models. Therefore, the region for the measurement is only defined by using the same experimental selection and acceptance cuts applied at the reconstruction level, so that the results are largely independent from the assumptions on the relative fractions and kinematic distributions of the individual production modes.

For all these reasons, measurements within a fiducial phase space represent a powerful way to test SM predictions on the Lagrangian structure of the H boson interactions in the full spectra of the observables under study and, eventually, to probe BSM effects, non-standard production modes, or new tensorial couplings. The procedure accounts for the

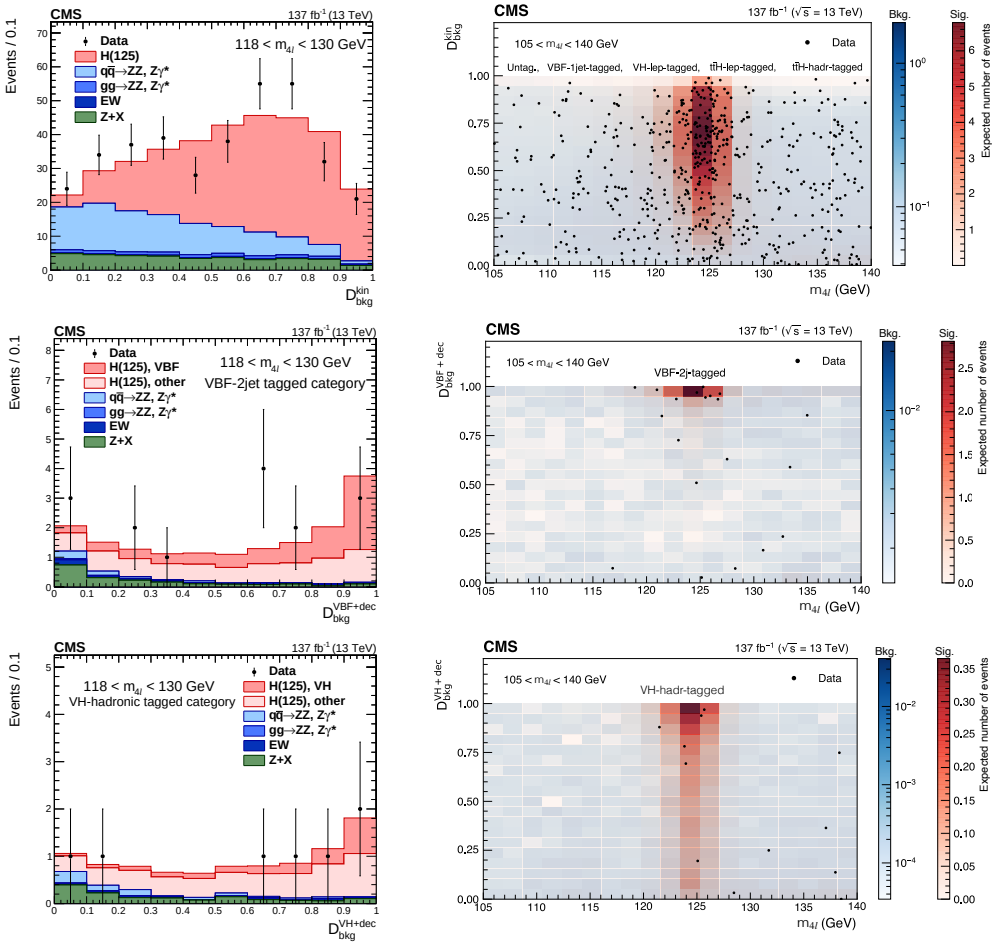


FIGURE 6.1: Distribution of the $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ (top left), $\mathcal{D}_{\text{bkg}}^{\text{VBF+dec}}$ (middle left), and $\mathcal{D}_{\text{bkg}}^{\text{VH+dec}}$ (right left) kinematic discriminants in the mass region $m_{4\ell} = [118, 130]$ GeV. Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes. The H(125) signal, the ZZ background, and the rare electroweak backgrounds are normalized to the SM expectations, while the Z+X background to the estimation from data. In the case of discriminants dedicated to the VBF and VH categories, the SM H boson is separated into two components: the production mode which is targeted by the dedicated discriminant and other production modes, where the gluon fusion process is dominant. On the right, the correlation between each of the discriminants and the four-lepton reconstructed mass $m_{4\ell}$ in the mass region $m_{4\ell} = [105, 140]$ GeV is shown. The red scale represents the number of expected SM H(125) boson signal events, while the blue scale represents the expected total number of ZZ, rare electroweak, and Z+X background events. The points show the data from the categories listed in the legend.

unfolding of detector effects from the observed distributions. In this context, the precision of the simulated processes, the theoretical knowledge about the expected BSM effects as well as their possible effect on the unfolding procedure represent a delicate problem. In the following part, the fiducial volume definition and the specific strategy adopted for the cross section measurement of the pp $\rightarrow$ H $\rightarrow 4\ell$ process within a fiducial volume will be discussed.

Requirements for the $H \to 4\ell$ fiducial phase space	
Lepton kinematics and isolation	
Leading lepton p_T	$p_T > 20$ GeV
Next-to-leading lepton p_T	$p_T > 10$ GeV
Additional electrons (muons) p_T	$p_T > 7(5)$ GeV
Pseudorapidity of electrons (muons)	$\lvert\eta\rvert < 2.5(2.4)$
Sum of scalar p_T of all stable particles p within $\Delta R_{p,\ell} < 0.3$	$< 0.35 p_T$
Event topology	
Existence of at least two same flavor OS lepton pairs, where leptons satisfy the criteria above	
Invariant mass of the Z_1 candidate	$m_{Z_1} = [40, 120]$ GeV
Invariant mass of the Z_2 candidate	$m_{Z_2} = [12, 120]$ GeV
Distance between selected four leptons	$\Delta R(\ell_i, \ell_j) > 0.02$ for any $i \neq j$
Invariant mass of any opposite sign lepton pair	$m_{\ell^+\ell'^-} > 4$ GeV
Invariant mass of the selected four leptons	$m_{4\ell}$ = [105,140] GeV

TABLE 6.1: Summary of requirements used in the definition of the fiducial phase space for the $H \to 4\ell$ cross section measurements.

The fiducial volume definition at the generator-level is designed to match the experimental acceptance in terms of the lepton kinematics and topological event selection (see Ref. [104]). All details about the construction of the fiducial region are reported in Table 6.1. In the selection of the phase space we consider only *dressed* leptons, obtained by adding the four-momentum of photons within $\Delta R < 0.3$ to the bare leptons. The isolation criteria is also required, matching the Run 2 reconstruction level isolation. Leptons are considered isolated at generator level if the sum of the transerse momentum of all particles (except for electrons, muons, and neutrinos) within a cone $\Delta R < 0.3$ is smaller than $0.35 p_T$. An increase of the model dependence compared to Ref. [104] is observed when using the ZZ candidate selection at reconstruction level where the candidate with the best $\mathcal{D}_{bkg}^{kin}$ discriminant value is chosen. For this reason, the ZZ pair selection in case of multiple candidates is simplified. The Z_1 candidate is chosen to be the one with $m(Z_1)$ closest to the nominal Z boson mass and the pair of leptons with the largest sum of the p_T magnitudes is chosen when there are multiple Z_2 candidates which satisfy all criteria. The candidate selection applied at the reconstruction level for the fiducial cross section measurements is based on the same requirements used for the fiducial-level selections.

The extraction of the fiducial cross section for the $pp \to H \to 4\ell$ process (σ_{fid}) is performed by using a maximum likelihood fit of the signal and background parametrizations to the observed four-lepton invariant mass, $N_{obs}(m_{4\ell})$. Differently from the approach presented for the signal strength measurements, here no event categorization is applied and the $\mathcal{D}_{bkg}^{kin}$ observable is not included in the measurement to minimize the model dependence. On the one hand, the inclusive fit is performed simultaneously in all final states. On the other hand, the branching ratio of the H boson to different final states ($4e, 4\mu, 2e2\mu$) is a free parameter in the fit. For all measurements presented in this chapter, the most precise measurement of the H mass currently available from the combination of the 2016 results of the $H \to 4\ell$ and $H \to \gamma\gamma$ analyses [30, 48, 49] is used: 125.38 ± 0.14 GeV.
The systematic uncertainties described in Section 5.4 enter the likelihood in the form of nuisance parameters and the results are provided by using an asymptotic approach [197] with a test statistic based on the profile likelihood ratio [25]. The unfolding of detector effects is the same as in Refs. [104] and [198]. Considering a certain observable, the number of expected events in each final state f and in each bin i can be expressed as a function of

$m_{4\ell}$:

$$\begin{aligned} N^{f,i}_{\text{exp}}(m_{4\ell}) &= N^{f,i}_{\text{fid}}(m_{4\ell}) + N^{f,i}_{\text{nonfid}}(m_{4\ell}) + N^{f,i}_{\text{nonres}}(m_{4\ell}) + N^{f,i}_{\text{bkg}}(m_{4\ell}) \\ &= \epsilon^{f}_{i,j}\left(1 + f^{f,i}_{\text{nonfid}}\right)\sigma^{f,j}_{\text{fid}}\mathcal{L}\mathcal{P}_{\text{res}}(m_{4\ell}) \\ &\quad + N^{f,i}_{\text{nonres}}\mathcal{P}_{\text{nonres}}(m_{4\ell}) + N^{f,i}_{\text{bkg}}\mathcal{P}_{\text{bkg}}(m_{4\ell}). \end{aligned} \tag{6.8}$$

It is worth presenting in more detail the content of this formula. A double Crystal Ball (see Section 5.1.3) is used to describe the shape of the resonant signal contribution, $\mathcal{P}_{\text{res}}(m_{4\ell})$; the normalization is proportional to the fiducial cross section.
Reconstructed events not originating from the fiducial phase space represent an additional resonant contribution. They arise from detector effects which cause differences between the quantities used for the fiducial region definition and the corresponding ones at the reconstruction level. This contribution is treated as background and is referred to as the *non-fiducial signal* contribution. Given that the shape of these events is identical to the shape of the fiducial signal, the normalization is fixed to be a fraction of the fiducial signal component. The value of this fraction, denoted as f_{nonfid}, has been determined from the simulation for each of the considered signal models. The detector response matrix that maps the number of expected events in a given observable bin j at the fiducial level to the number of expected events in the bin i at the reconstruction level is represented by the $\epsilon^{f}_{i,j}$. The f^{i}_{nonfid} fraction describes the ratio of the non-fiducial and fiducial signal contribution in bin i at the reconstruction level. The efficiency is measured using simulated samples of the signal and corrected for residual differences between data and MC, resulting in a single efficiency for the integrated fiducial cross section measurement. The variation between different models within the experimental constraint is directly related to the model dependence of the measurement and it is expressed by the factor $(1 + f_{\text{nonfid}})\epsilon$.
The nonresonant signal contribution is expressed by the function $\mathcal{P}_{\text{nonres}}(m_{4\ell})$ multiplied by the number of nonresonant events, $N^{f,i}_{\text{nonres}}(m_{4\ell})$. It is determined by the WH, ZH, and $\text{t}\bar{\text{t}}\text{H}$ contributions where one of the leptons from the H boson decay is lost or not selected and it is modeled by a Landau distribution with shape parameters constrained in the fit to be within a range determined from simulation. This contribution is referred to as the *combinatorial signal* and is treated as a background in this measurement.
The contribution of the other background sources in bin i as a function of $m_{4\ell}$ is given by the term $N^{f,i}_{\text{bkg}}(m_{4\ell})\mathcal{P}_{\text{bkg}}(m_{4\ell})$, where $\mathcal{P}_{\text{bkg}}(m_{4\ell})$ represents the corresponding background PDF.

6.2 Results: signal strength and STXS measurements

After a discussion of the strategy adopted for the measurements, we proceed to show and quantify the inputs to the final results provided by the analysis after the full selection. The event yields are reported, together with the distribution of the signal and background events with associated uncertainties. Different regions of the four-lepton invariant mass distribution are considered: the statistical analysis is performed considering the restricted mass window $m_{4\ell}$ =[105,140] GeV, but the full signal region is presented in the $m_{4\ell}$ distributions. Although it is not directly part of the scope of this work, it is relevant for analyses thoroughly connected with the study of the SM H(125) in the four-lepton final state, such as the off-shell analysis and the study of the anomalous couplings or of the H boson mass and width.

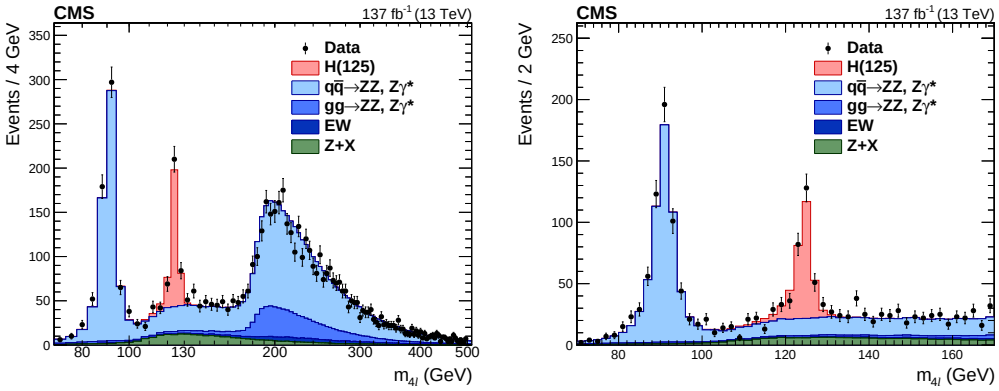


FIGURE 6.2: Four-lepton mass distribution, $m_{4\ell}$, up to 500 GeV with 4 GeV bin size (left) and in the low mass range with 4 GeV bin size (right). Points with error bars represent the data and stacked histograms represent the expected distributions for the signal and background processes. The SM Higgs boson signal with m_{H} =125 GeV, denoted as H(125), the ZZ and rare electroweak backgrounds are normalized to the SM expectation, the Z+X background to the estimation from data.

The observed four-lepton mass distribution is presented in Fig. 6.2 for the full Run 2 dataset, considering inclusively all final states ($4e$, 4μ, and $2e2\mu$) and all production modes. A comparison with the expectations from signal and all background processes is also shown. On the one hand, the H(125) signal, the ZZ background, and the rare electroweak backgrounds are normalized to the SM expectation and the information about the shapes is provided by the MC samples, and on the other hand, the normalization of the Z+X background comes from the estimation from data. The EWK contribution has been studied and included in these kind of plots for the first time in the story of the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis. The data points are plotted with associated error bars which correspond to the so-called Garwood confidence intervals at 68% confidence level (CL) [199]. The agreement between the observed distribution and the SM expectation is found to be within the statistical uncertainties over the whole spectrum. The reconstructed dilepton invariant masses selected as Z_1 and Z_2 are shown in Fig. 6.3 together with their correlation, both in the full range of $m_{4\ell}$ and focusing on a $m_{4\ell} = [118, 130]$ GeV narrow mass window around the H boson peak. The observed four-lepton invariant mass is reported in Fig. 6.4 considering each final state separately in the two mass ranges. The distribution of the kinematic discriminants used for the categorization of various production mechanisms ($\mathcal{D}_{2\mathrm{jet}}$, $\mathcal{D}_{1\mathrm{jet}}$, $\mathcal{D}_{\mathrm{VH}}$) along with the corresponding working point values are shown in Fig. 6.5 (right) in the $m_{4\ell} = [118, 130]$ GeV mass window. Events with at least two selected jets are considered, except for $\mathcal{D}_{1\mathrm{jet}}$ that is related to events with exactly one selected jet. The correlation of each production discriminant with the four-lepton invariant mass is shown in Fig. 6.5 (right).

To conclude, the number of the H boson candidates observed in the data corresponding to an integrated luminosity of 137 fb^{-1} and the expected yields for the H(125) signal and all background processes after the full event selection are provided in Table 6.2. The mass window around the H boson peak $m_{4\ell}$ =[105,140] GeV is considered and results are reported for each of the twenty-two reconstructed event categories (see Section 4.3.3). The number of expected and observed events in every stage 1.2 reconstructed category is presented graphically in Fig.6.6 for the full Run 2 dataset.

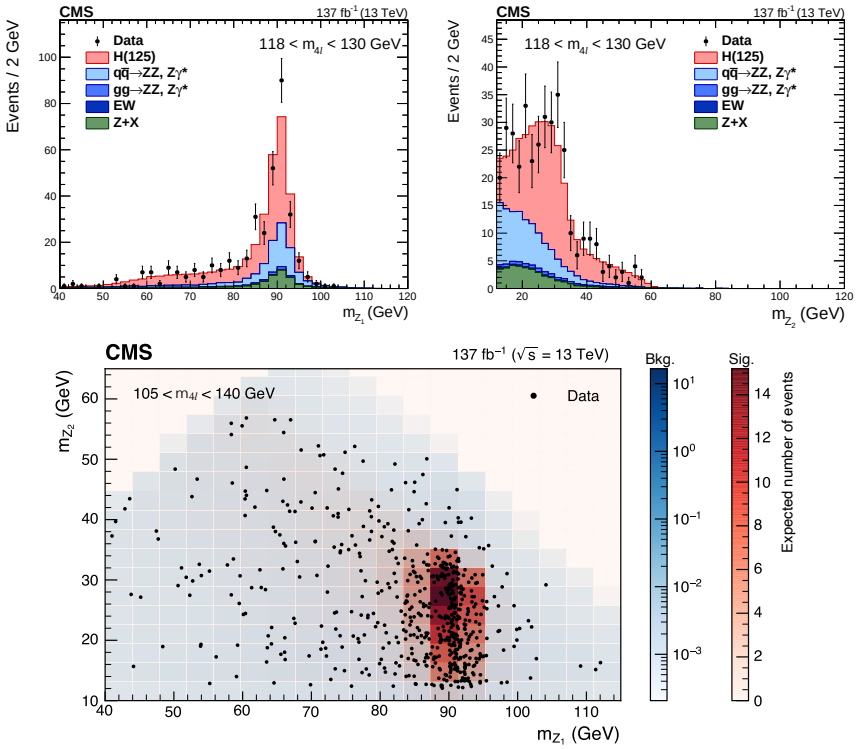


FIGURE 6.3: Distribution of the Z_1 (top left) and Z_2 (top right) invariant mass in the low mass range $m_{4\ell} = [118, 130]$ GeV and their two-dimensional Z_2 versus Z_1 distribution (bottom) in the mass range $m_{4\ell} = [105, 140]$ GeV. The stacked histograms and the red and blue scales represent expected distributions of the signal and background processes. The points represent the data. The H(125) signal as well as the ZZ and rare electroweak backgrounds are normalized to the SM expectation, while the Z+X background yield is normalized to the estimation from data.

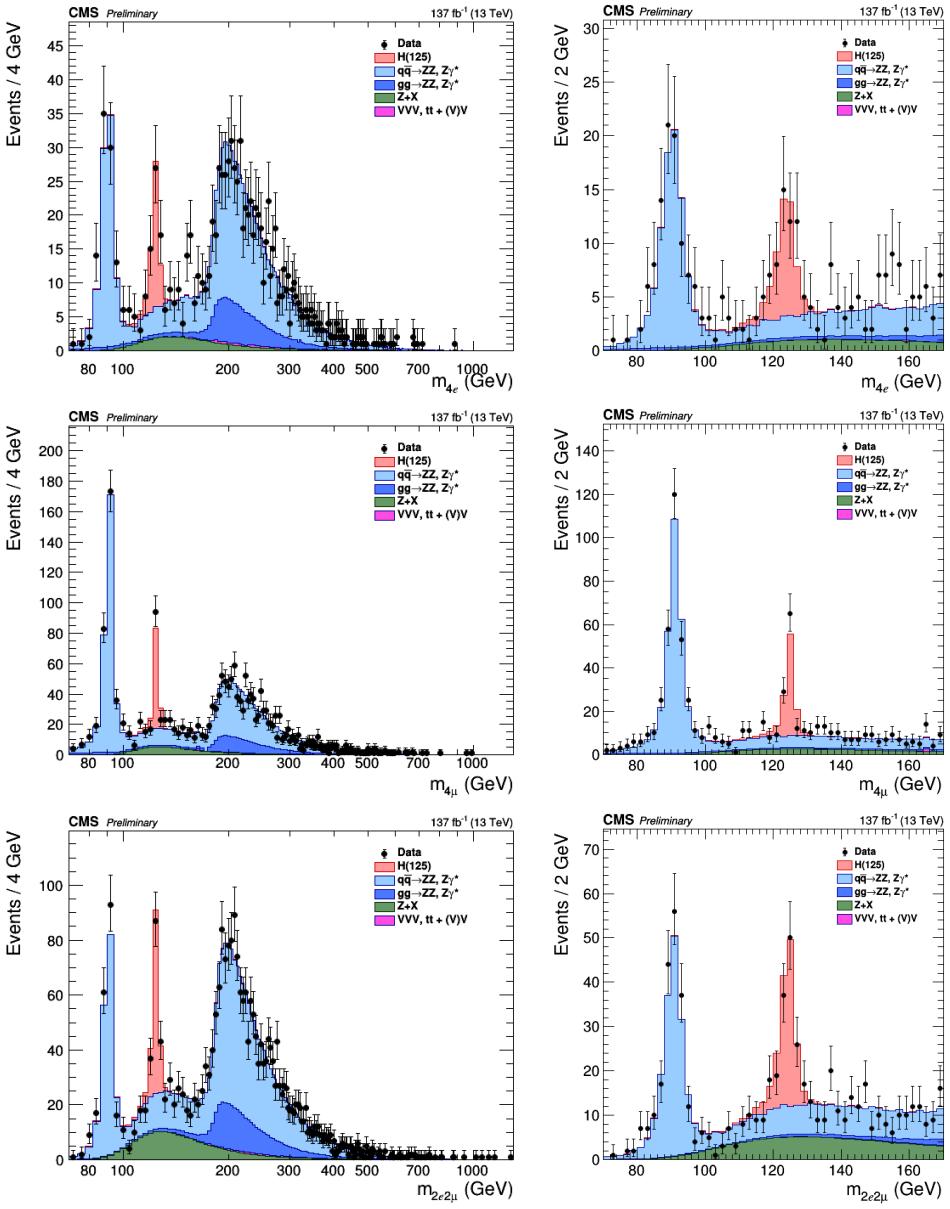


FIGURE 6.4: Four-lepton reconstructed mass for the full mass spectrum (left) and the low mass range (right) considering three different subchannels: $4e$ (top), 4μ (middle), and $2e2\mu$ (bottom).

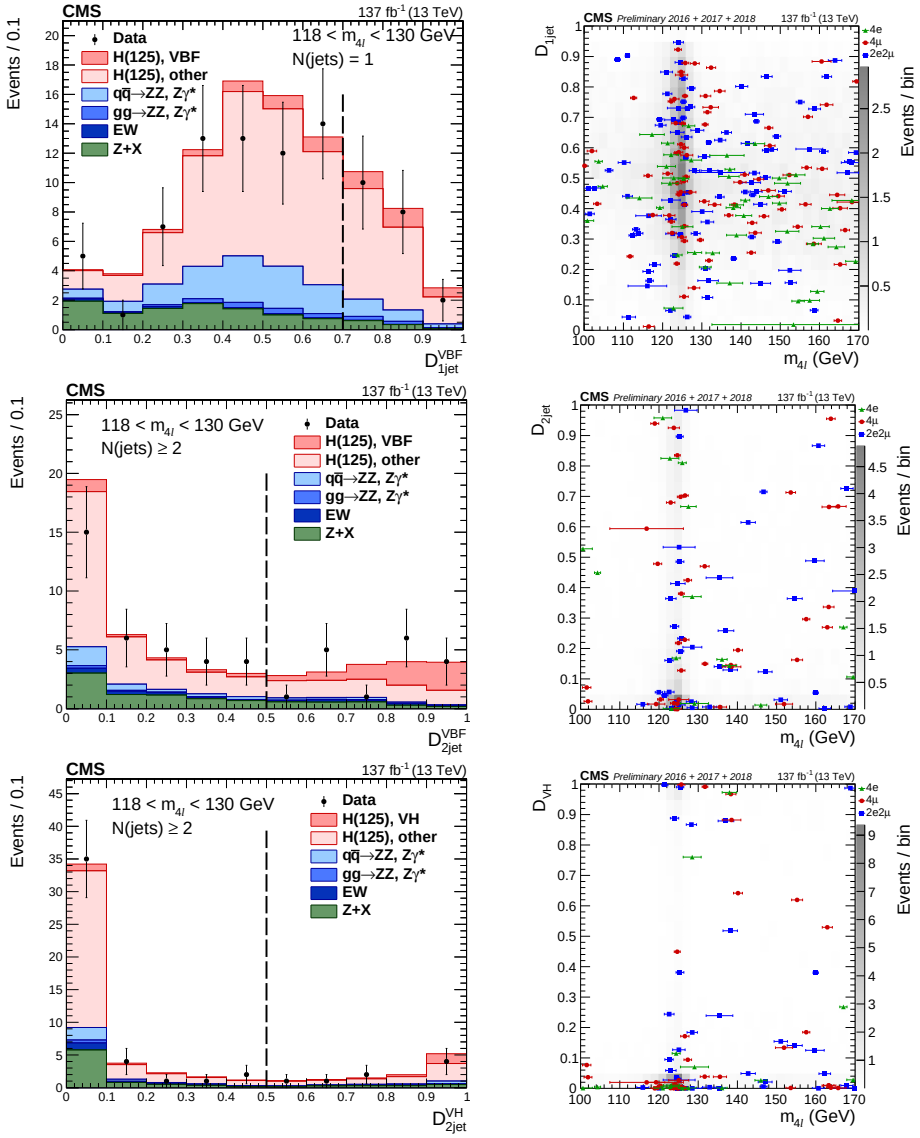


FIGURE 6.5: Distribution of the $\mathcal{D}_{2\text{jet}}$ (left), $\mathcal{D}_{1\text{jet}}$ (middle), and $\mathcal{D}_{\text{VH}} = \max(\mathcal{D}_{\text{WH}}, \mathcal{D}_{\text{ZH}})$ (right) production discriminants used for the event categorization in the mass region $m_{4\ell} = [118, 130]$ GeV. Points with error bars represent the data and stacked histograms represent expected distributions. The H(125) signal and the ZZ backgrounds are normalized to the SM expectation, while the Z+X background to the estimation from data. The vertical grey dashed lines denote the working points used in the event categorization. On the right, the correlation between each of the production discriminants and the four-lepton reconstructed mass $m_{4\ell}$ in the mass region $m_{4\ell} = [118, 130]$ GeV is shown. The grey scale represents the expected relative density of ZZ background plus H(125) signal, while the points show the data and the horizontal bars represent the measured mass uncertainties.

Reconstructed event	Signal							Background				Expected		Observed
category	ggH	VBF	WH	ZH	ttH	$b\bar{b}H$	tHq	qqZZ	ggZZ	EW	Z+ X	signal	total	
Untagged-0j-$p_{\mathrm{T}}^{4\ell}[0,10]$	27.7	0.09	0.03	0.03	0.00	0.15	0.00	71.5	3.06	0.01	3.21	27.9±0.1	106±0	114
Untagged-0j-$p_{\mathrm{T}}^{4\ell}[10,200]$	96.2	1.69	0.60	0.77	0.01	1.01	0.00	98.1	11.6	0.35	37.8	100±0	248±1	278
Untagged-1j-$p_{\mathrm{T}}^{4\ell}[0,60]$	26.8	1.51	0.56	0.48	0.01	0.45	0.01	25.3	3.02	0.64	14.2	29.8±0.1	72.9±0.4	74
Untagged-1j-$p_{\mathrm{T}}^{4\ell}[60,120]$	13.5	1.31	0.51	0.41	0.02	0.11	0.01	7.81	0.82	0.62	7.95	15.9±0.1	33.1±0.3	20
Untagged-1j-$p_{\mathrm{T}}^{4\ell}[120,200]$	3.51	0.60	0.17	0.17	0.01	0.02	0.00	1.15	0.19	0.25	1.63	4.48±0.05	7.69±0.16	11
Untagged-2j-$p_{\mathrm{T}}^{4\ell}[0,60]$	3.45	0.29	0.15	0.14	0.08	0.09	0.02	2.14	0.32	0.63	4.75	4.20±0.06	12.1±0.2	14
Untagged-2j-$p_{\mathrm{T}}^{4\ell}[60,120]$	5.26	0.56	0.24	0.19	0.12	0.04	0.03	2.19	0.30	0.72	4.14	6.43±0.06	13.8±0.2	15
Untagged-2j-$p_{\mathrm{T}}^{4\ell}[120,200]$	3.07	0.40	0.16	0.13	0.07	0.01	0.02	0.75	0.14	0.34	1.19	3.86±0.05	6.28±0.14	7
Untagged-$p_{\mathrm{T}}^{4\ell}$>200	2.79	0.62	0.21	0.17	0.07	0.01	0.02	0.43	0.21	0.21	0.73	3.89±0.04	5.47±0.11	3
Untagged-2j-m_{jj}>350	0.77	0.16	0.06	0.04	0.05	0.01	0.01	0.34	0.06	0.31	1.71	1.12±0.02	3.54±0.14	3
VBF-1jet-tagged	15.5	3.29	0.22	0.16	0.00	0.13	0.01	6.85	1.53	0.20	2.44	19.3±0.1	30.3±0.2	27
VBF-2jet-tagged-$m_{jj}[350,700]$	0.83	1.19	0.01	0.01	0.00	0.01	0.00	0.19	0.07	0.11	0.14	2.05±0.03	2.55±0.05	2
VBF-2jet-tagged-m_{jj}>700	0.43	1.96	0.00	0.00	0.00	0.00	0.00	0.07	0.05	0.12	0.03	2.40±0.02	2.67±0.03	1
VBF-3jet-tagged-m_{jj}>350	2.52	2.35	0.06	0.06	0.03	0.03	0.05	0.62	0.21	0.64	2.43	5.11±0.05	9.01±0.17	12
VBF-2jet-tagged-$p_{\mathrm{T}}^{4\ell}$>200	0.44	0.79	0.01	0.01	0.01	0.00	0.01	0.03	0.03	0.04	0.06	1.26±0.02	1.42±0.03	0
VBF-rest	2.48	0.94	0.13	0.09	0.04	0.04	0.01	0.98	0.20	0.39	2.18	3.74±0.05	7.49±0.17	5
VH-hadronic-tagged-$m_{jj}[60,120]$	4.11	0.25	1.09	0.96	0.13	0.06	0.02	1.69	0.22	0.52	2.93	6.62±0.06	12.0±0.2	12
VH-rest	0.57	0.03	0.09	0.06	0.03	0.01	0.00	0.16	0.02	0.06	0.33	0.79±0.02	1.36±0.06	0
VH-leptonic-tagged-$p_{\mathrm{T}}^{4\ell}[0,150]$	0.33	0.04	0.85	0.26	0.10	0.03	0.03	2.16	0.36	0.19	1.11	1.64±0.02	5.47±0.13	10
VH-leptonic-tagged-$p_{\mathrm{T}}^{4\ell}$>150	0.02	0.01	0.21	0.06	0.04	0.00	0.01	0.05	0.01	0.03	0.08	0.35±0.01	0.52±0.03	0
ttH-leptonic-tagged	0.02	0.01	0.02	0.02	0.68	0.00	0.03	0.08	0.01	0.23	0.21	0.79±0.01	1.32±0.07	0
ttH-hadronic-tagged	0.18	0.05	0.03	0.05	0.86	0.01	0.03	0.03	0.01	0.82	1.06	1.22±0.01	3.15±0.14	2

TABLE 6.2: Number of the expected background and signal events and number of observed candidates after full analysis selection in the mass range $m_{4\ell}$ =[105,140] GeV for an integrated luminosity of 137 fb^{-1}. The yields are given for the different production modes considering the twenty-two reconstructed event categories. Signal, ZZ background from $\mathrm{q\bar{q}} \to \mathrm{ZZ}$ and $\mathrm{gg} \to \mathrm{ZZ}$ processes, and rare electroweak backgrounds are estimated from MC simulation, while Z+X is estimated from data.

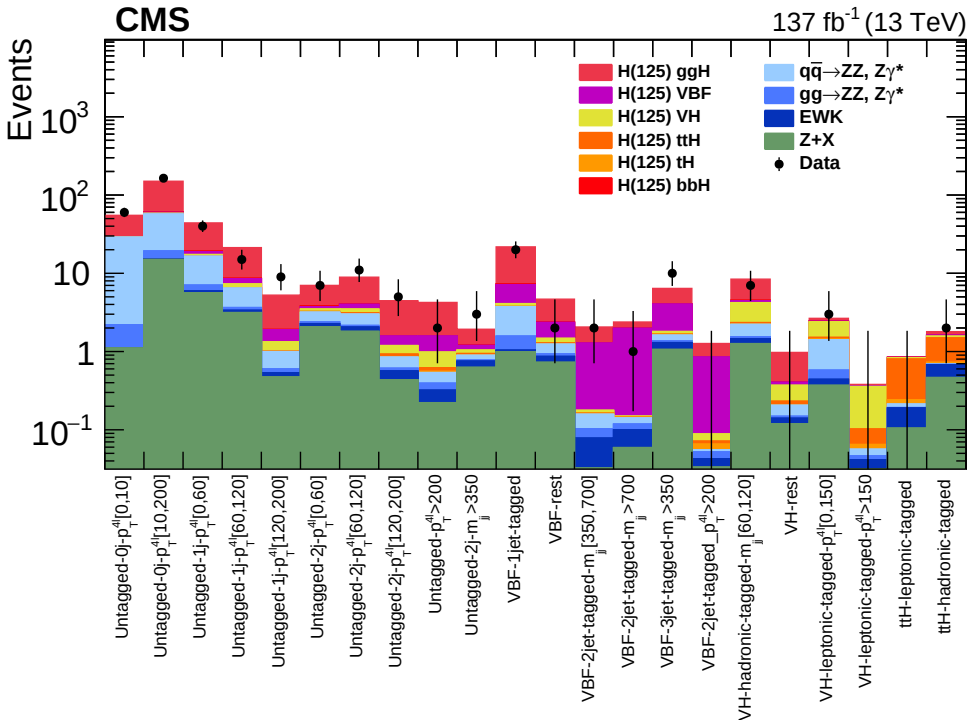


FIGURE 6.6: Distributions of the expected and observed number of events for the reconstructed event categories in the mass region $m_{4\ell}$ =[105,140] GeV. Points with error bars represent the data and stacked histograms represent the expected numbers of the signal and background events. The different H boson production mechanisms with m_{H} =125 GeV, denoted as H(125), and the ZZ background, and the rare electroweak backgrounds are normalized to the SM expectation, while the Z+X background to the estimation from data.

6.2.1 Signal strength measurement

The measurement of the inclusive signal strength of the H boson in the four-lepton final state relies on the methodology presented in the first part of Section 6.1.2. However, the same strategy can be easily adapted to produce other interesting results, that we are going to discuss in this section.

The combined measurement of the inclusive signal strength modifier with a simultaneous fit to all categories is:

$$\mu = 0.94\,^{+0.12}_{-0.11} = 0.94 \pm 0.07\ (\mathrm{stat})\,^{+0.07}_{-0.06}\ (\mathrm{th})\,^{+0.06}_{-0.05}\ (\mathrm{exp}),$$

with the mass of the H boson fixed to 125.38 ± 0.14 GeV. The SM cross sections are extrapolated at m_{H} = 125.38 GeV using numbers reported in the most recent Yellow Report (YR4) as a reference [16]. Later, results are also provided by removing the assumpion on the H boson mass and considering it as a freely floating parameter in the fit, so we can say that it is *profiled* in the fit. The observed and expected profile likelihood scans of the inclusive signal-strength modifier is shown in Fig. 6.7. The total uncertainty is affected in a similar way by the experimental and theoretical contributions. The fact that these sources of uncertainty in the measurement are found to be similar in magnitude remarks significantly a new way of exploring the H boson properties, moving from a discovery approach to precision measurements. The experimental uncertainties on the lepton identification efficiencies and

luminosity measurement represent the dominant source of the systematic uncertainty. The main contribution from the theoretical side comes from the uncertainty related to the total gluon fusion cross section. The impact of the dominant sources of systematic uncertainty on the measurement is summarized later in Table 6.7 and illustrated in Fig. 6.16.

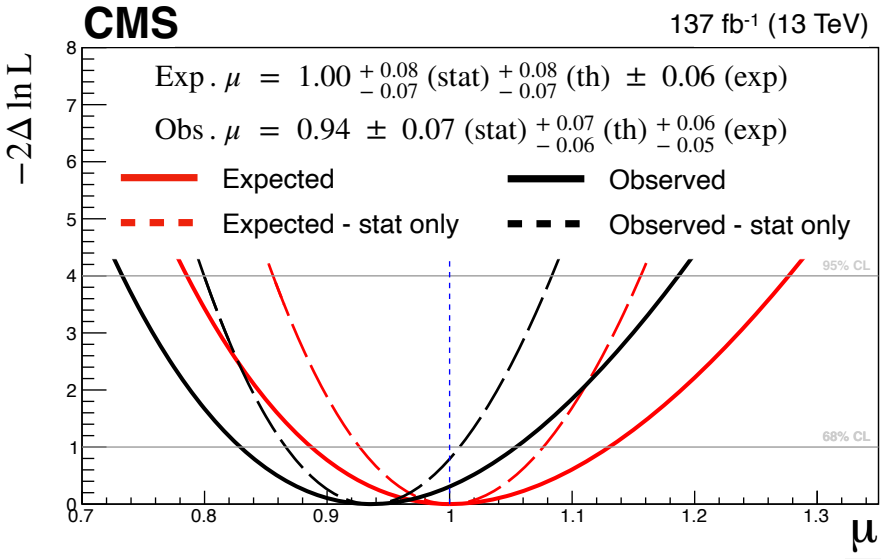


FIGURE 6.7: The observed and expected profile likelihood scans of the inclusive signal-strength modifier, derived both with (solid line) and without (dashed line) systematic uncertainties.

In addition to the inclusive μ measurement, signal strengths per production mode can be measured. A definition of a signal strength modifier for all five main SM H boson production mechanisms is needed. Then, the basic model previously discussed, $\mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})$, is substituted by:

$$\mu_{\mathrm{ggH,b\bar{b}H}} \cdot (s(\boldsymbol{\theta})_{\mathrm{ggH}} + s(\boldsymbol{\theta})_{\mathrm{b\bar{b}H}}) + \mu_{\mathrm{VBF}} \cdot s(\boldsymbol{\theta})_{\mathrm{VBF}} + \mu_{\mathrm{WH}} \cdot s(\boldsymbol{\theta})_{\mathrm{WH}} + \mu_{\mathrm{ZH}} \cdot s(\boldsymbol{\theta})_{\mathrm{ZH}} + \mu_{\mathrm{t\bar{t}H,tH}} \cdot (s(\boldsymbol{\theta})_{\mathrm{t\bar{t}H}} + s(\boldsymbol{\theta})_{\mathrm{tH}}) + b(\boldsymbol{\theta}). \tag{6.9}$$

The contributions of the $\mathrm{b\bar{b}H}$ and tH production modes are also taken into account, merging them with the ggH and $\mathrm{t\bar{t}H}$ category, respectively. The relative normalizations of the $\mathrm{b\bar{b}H}$ (tH) and the gluon fusion ($\mathrm{t\bar{t}H}$) contributions are kept fixed in the fit. Therefore, the results of likelihood scans for the considered scenarios are presented in Fig. 6.8 (left): $\mu_{\mathrm{ggH,b\bar{b}H}}$, μ_{VBF}, μ_{WH}, μ_{ZH}, $\mu_{\mathrm{t\bar{t}H,tH}}$. The corresponding numerical values are given in Table 6.3, together with the decomposition of the uncertainties into statistical and systematic components and the corresponding expected uncertainties.

It is worth to note that the WH and ZH processes are treated separately as this approach represents a novelty of the legacy analysis presented here. In fact, they were merged in the past and later split according to the decay of the associated vector boson into either hadronic or leptonic decays because of the lack of sensitivity to the WH and ZH production modes separately. The splitting of VH into two subcategories leads to a minor amount of events in each bin, resulting in an increase of the associated statistical uncertainty, especially for the ZH category where the cross section is significantly smaller with respect to WH. The results coming from the merging of the WH and ZH contributions in a single VH bin are shown in Fig. 6.8 (right).

Profiling the mass of the H boson in the fit, the best value obtained is $\mu = 0.97^{+0.12}_{-0.11}$, which is perfectly in agreement with the observed value quoted above. The best fit mass is found to be $\hat{m}_{\mathrm{H}} = 125.09^{+0.15}_{-0.14}$ GeV, as shown in Fig. 6.10, compatible with the CMS mass

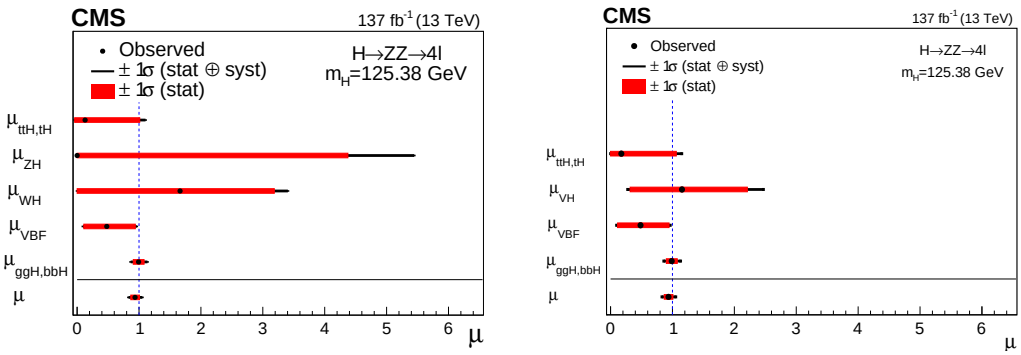


FIGURE 6.8: Results of likelihood scans for the signal strength modifiers corresponding to the five main SM Higgs boson production mechanisms, compared to the inclusive signal strength modifier μ shown as a vertical line (left). The thick black lines indicate the 1σ confidence intervals including both statistical and systematic sources. The thick red lines indicate the statistical uncertainties corresponding to the 1σ confidence intervals. On the right, the same results are shown considering the merging of the WH and ZH production mode.

measurement performed by combining the $\mathrm{H} \to 4\ell$ and $\mathrm{H} \to \gamma\gamma$ analyses. A limiting factor on the combined H boson mass measurement was the amount of data available. Now that the analysis is no longer statistically limited, it is possible to measure the H boson mass with the same precision exploiting the four-lepton final state alone. However, the precise determination of the H boson mass and its uncertainties is beyond the scope of this work. To conclude on this topic, even considering the five main production mechanisms separately, no drastic changes in the results are found, as shown in Fig. 6.9.

Another interesting study in the context of the signal strength modifiers measurements can be performed by defining $\mu_{\mathrm{ggH,t\bar{t}H,\,b\bar{b}H,\,tH}}$ and $\mu_{\mathrm{VBF,VH}}$ as scale factors for the fermion and vector boson induced contributions to the expected SM cross section. The former scales

	Expected	Observed
$\mu_{\mathrm{ggH,b\bar{b}H}}$	$1.00^{+0.10}_{-0.10}(\mathrm{stat.})\,^{+0.12}_{-0.10}(\mathrm{syst.})$	$0.99^{+0.09}_{-0.09}(\mathrm{stat.})\,^{+0.11}_{-0.09}(\mathrm{syst.})$
μ_{VBF}	$1.00^{+0.53}_{-0.44}(\mathrm{stat.})\,^{+0.18}_{-0.12}(\mathrm{syst.})$	$0.48^{+0.46}_{-0.37}(\mathrm{stat.})\,^{+0.14}_{-0.10}(\mathrm{syst.})$
μ_{ZH}	$1.00^{+4.79}_{-1.00}(\mathrm{stat.})\,^{+6.76}_{-0.00}(\mathrm{syst.})$	$0.00^{+4.38}_{-0.00}(\mathrm{stat.})\,^{+3.24}_{-0.00}(\mathrm{syst.})$
μ_{WH}	$1.00^{+1.83}_{-1.00}(\mathrm{stat.})\,^{+0.75}_{-0.00}(\mathrm{syst.})$	$1.66^{+1.52}_{-1.66}(\mathrm{stat.})\,^{+0.85}_{-0.00}(\mathrm{syst.})$
$\mu_{\mathrm{t\bar{t}H,tH}}$	$1.00^{+1.23}_{-0.77}(\mathrm{stat.})\,^{+0.51}_{-0.06}(\mathrm{syst.})$	$0.17^{+0.88}_{-0.17}(\mathrm{stat.})\,^{+0.42}_{-0.00}(\mathrm{syst.})$
μ	$1.00^{+0.08}_{-0.07}(\mathrm{stat.})\,^{+0.10}_{-0.08}(\mathrm{syst.})$	$0.94^{+0.07}_{-0.07}(\mathrm{stat.})\,^{+0.09}_{-0.08}(\mathrm{syst.})$

TABLE 6.3: Best fit values and $\pm\,1\sigma$ uncertainties for the expected and observed signal strength modifiers. The value of m_{H} is fixed to the best CMS value $m_{\mathrm{H}} = 125.38$ GeV for all the results. The statistical and systematic uncertainties are given separately.

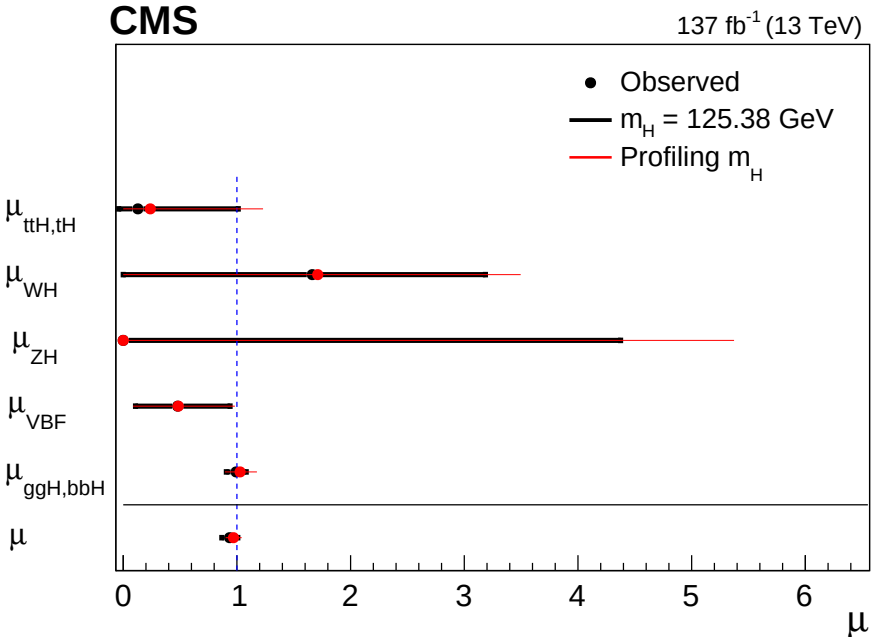


FIGURE 6.9: Comparison between the results of likelihood scans for the signal strength modifiers corresponding to the five main SM H boson production mechanisms obtained in two different scenario: fixing the mass of the H at 125.38 GeV (black points) and profiling it in the fit (red points).

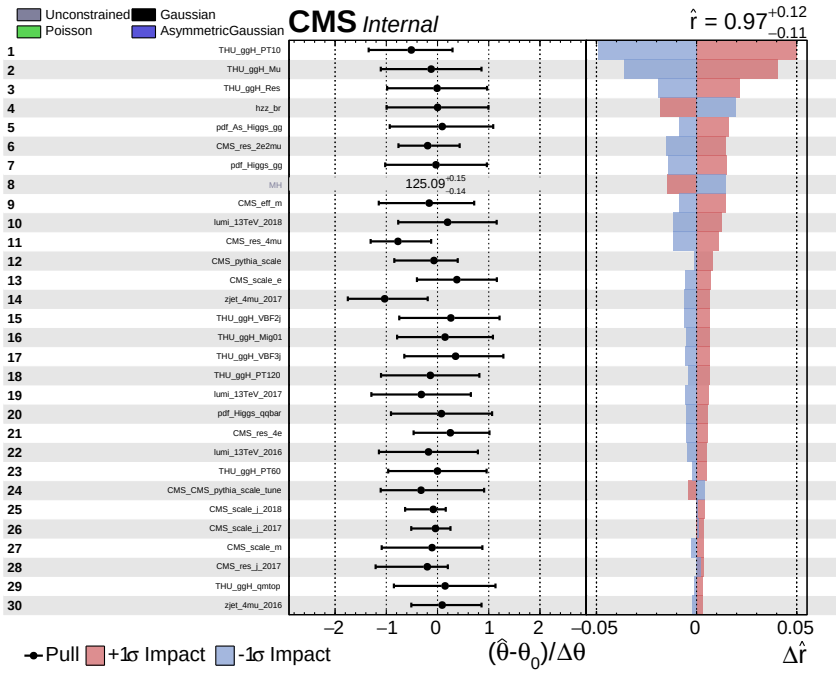


FIGURE 6.10: Systematic uncertainties with the largest impact on the signal strength μ measurement with m_{H} profiled in the fit: value and error of the corresponding nuisance parameter after the maximum likelihood fit to the observed data are shown.

the H boson production cross sections in modes related to couplings to fermions, the latter scales the cross sections in modes related to couplings to electroweak gauge bosons. They will enter the model in the form:

$$\mu_{\mathrm{ggH,t\bar{t}H,\,b\bar{b}H,\,tH}} \cdot (s(\boldsymbol{\theta})_{\mathrm{ggH}} + s(\boldsymbol{\theta})_{\mathrm{t\bar{t}H}} + s(\boldsymbol{\theta})_{\mathrm{b\bar{b}H}} + s(\boldsymbol{\theta})_{\mathrm{tH}}) + \mu_{\mathrm{VBF,VH}} \cdot (\boldsymbol{s(\theta)}_{\mathrm{VBF}} + s(\boldsymbol{\theta})_{\mathrm{VH}}) + b(\boldsymbol{\theta}). \quad (6.10)$$

A two-parameter fit is performed simultaneously to the events reconstructed in all categories with m_{H} fixed to 125.38 GeV, leading to the expected and observed measurements reported in Table 6.4. The 68% and 95% CL contours in the $(\mu_{\mathrm{ggH,t\bar{t}H,\,b\bar{b}H,\,tH}}, \mu_{\mathrm{VBF,VH}})$ plane are shown in Fig. 6.11 and the SM predictions lie within the 68% CL regions of this measurement.

	Expected	Observed
$\mu_{\mathrm{ggH,t\bar{t}H,\,b\bar{b}H,\,tH}}$	$1.00\ ^{+0.15}_{-0.13}$	$0.96\ ^{+0.14}_{-0.12}$
$\mu_{\mathrm{VBF,VH}}$	$1.00\ ^{+0.39}_{-0.33}$	$0.82\ ^{+0.36}_{-0.31}$

TABLE 6.4: Best fit values and $\pm\,1\sigma$ uncertainties for the expected and observed signal strength modifiers $\mu_{\mathrm{ggH,t\bar{t}H,\,b\bar{b}H,\,tH}}$ and $\mu_{\mathrm{VBF,VH}}$, with the value of m_{H} fixed to $mH = 125.38$ GeV.

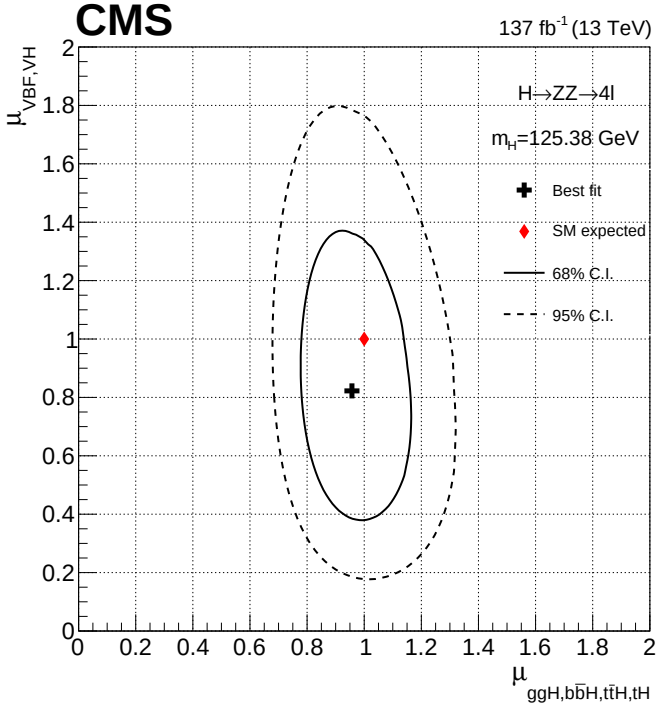


FIGURE 6.11: The 2D likelihood scan for the $\mu_{\mathrm{ggH,t\bar{t}H,\,b\bar{b}H,\,tH}}$ and $\mu_{\mathrm{VBF,VH}}$ signal strength modifiers, where the solid and dashed contours show the 68% and 95% CL regions. The best fit value and the expected value for the SM H boson are represented with a cross and a diamond, respectively.

6.2.2 STXS measurements

In the following part, the first extensive measurements of the H boson production cross section in the context of the STXS framework accounting for both stage 0 and stage 1.2 scenarios are presented.
Considering the production bins defined in Section 4.3.2, the measurement of the H boson production cross section, $(\sigma\mathcal{B})_{\text{obs}}$, and the branching fraction normalized by the SM expectation, $(\sigma\mathcal{B})_{\text{SM}}$, are reported in Tables 6.5 and 6.6 for the stage 0 and for the reduced stage 1.2 categorization. Results are obtained with the H boson mass fixed to 125.38 GeV. The corresponding plots for the stage 0 and for the reduced stage 1.2 production bins are shown in Fig. 6.12 and in Fig. 6.13. The uncertainties in the SM expectation are not taken into account in the ratio calculation, while the theoretical uncertainties are considered because they can induce migration of events between the different categories. The correlation matrices are shown in Fig. 6.14 and Fig. 6.15.

	$(\sigma\mathcal{B})_{\text{obs}}$ [fb]	$(\sigma\mathcal{B})_{\text{SM}}$ [fb]	$(\sigma\mathcal{B})_{\text{obs}}/(\sigma\mathcal{B})_{\text{SM}}$
ttH	$3\ ^{+16}_{-3}$	15.9 ± 1.4	$0.16\ ^{+0.98}_{-0.16}$
VH-lep	$41\ ^{+52}_{-35}$	25.9 ± 0.8	$1.56\ ^{+1.99}_{-1.34}$
qqH	$61\ ^{+53}_{-44}$	122 ± 6	$0.50\ ^{+0.44}_{-0.36}$
ggH	$1214\ ^{+135}_{-125}$	1192 ± 95	$1.02\ ^{+0.11}_{-0.11}$

TABLE 6.5: Best fit values and $\pm\ 1\sigma$ uncertainties for the measured cross sections $\sigma\mathcal{B}$ and the SM predictions $(\sigma\mathcal{B})_{\text{SM}}$ for the stage 0 production bins. The results are obtained with m_{H} fixed to 125.38 GeV.

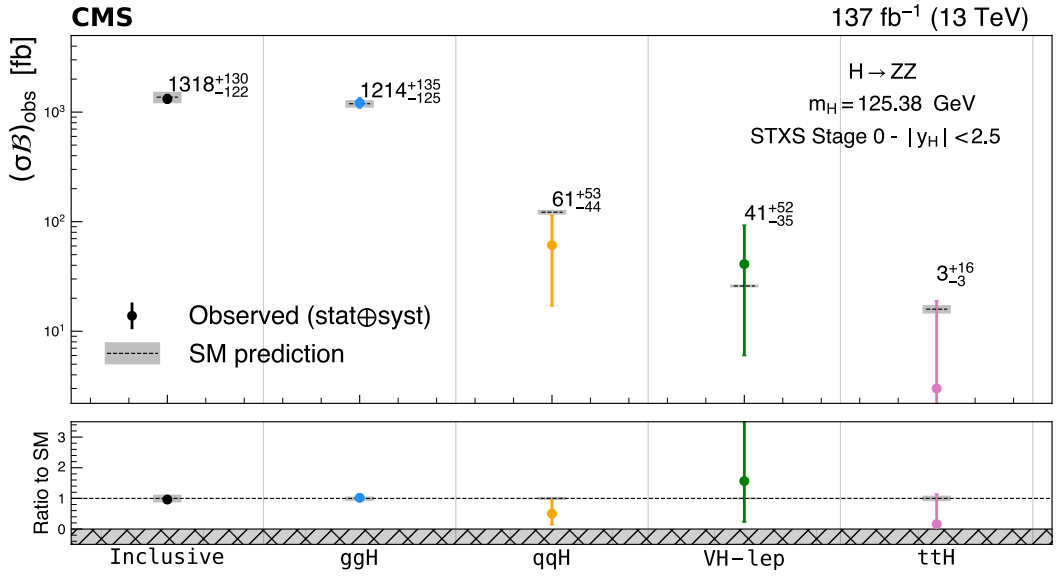


FIGURE 6.12: The measured cross sections $(\sigma\mathcal{B})_{\text{obs}}$ and the SM predictions $(\sigma\mathcal{B})_{\text{SM}}$ for the stage 0 production bins with m_{H} fixed to 125.38 GeV. The measured values are represented by points with error bars while the SM predictions are represented by black dashed lines with gray uncertainty bands. Ratios of the measured cross sections and the SM predictions are shown in the bottom panel with corresponding uncertainties for each of the bins and the inclusive measurement.

	$(\sigma\mathcal{B})_{\text{obs}}$ [fb]	$(\sigma\mathcal{B})_{\text{SM}}$ [fb]	$(\sigma\mathcal{B})_{\text{obs}}/(\sigma\mathcal{B})_{\text{SM}}$
$\mathtt{ggH}-\mathtt{0j}/\mathtt{p_T}[0,10]$	$145\ ^{+45}_{-40}$	164 ± 11	$0.89\ ^{+0.28}_{-0.24}$
$\mathtt{ggH}-\mathtt{0j}/\mathtt{p_T}[10,200]$	$611\ ^{+98}_{-90}$	561 ± 87	$1.09\ ^{+0.17}_{-0.16}$
$\mathtt{ggH}-\mathtt{1j}/\mathtt{p_T}[0,60]$	$214\ ^{+78}_{-87}$	177 ± 18	$1.21\ ^{+0.44}_{-0.49}$
$\mathtt{ggH}-\mathtt{1j}/\mathtt{p_T}[60,120]$	$59\ ^{+44}_{-53}$	121 ± 14	$0.48\ ^{+0.37}_{-0.44}$
$\mathtt{ggH}-\mathtt{1j}/\mathtt{p_T}[120,200]$	$53\ ^{+25}_{-22}$	20 ± 4	$2.62\ ^{+1.24}_{-1.08}$
$\mathtt{ggH}-\mathtt{2j}/\mathtt{p_T}[0,60]$	$0\ ^{+27}_{-0}$	35 ± 6	$0.00\ ^{+0.76}_{-0.00}$
$\mathtt{ggH}-\mathtt{2j}/\mathtt{p_T}[60,120]$	$78\ ^{+41}_{-37}$	51 ± 9	$1.53\ ^{+0.81}_{-0.73}$
$\mathtt{ggH}-\mathtt{2j}/\mathtt{p_T}[120,200]$	$27\ ^{+22}_{-19}$	26 ± 6	$1.06\ ^{+0.87}_{-0.72}$
$\mathtt{ggH}-\mathtt{2j}/\mathtt{m_{jj}} > 350$	$4\ ^{+72}_{-4}$	23 ± 3	$0.17\ ^{+3.2}_{-0.17}$
$\mathtt{ggH}/\mathtt{p_T} > 200$	$7\ ^{+8}_{-7}$	15 ± 6	$0.47\ ^{+0.56}_{-0.47}$
$\mathtt{qqH}-\mathtt{rest}$	$11\ ^{+161}_{-11}$	71 ± 5	$0.15\ ^{+2.27}_{-0.15}$
$\mathtt{qqH}-\mathtt{2j}/\mathtt{m_{jj}}[60,120]$	$12\ ^{+30}_{-12}$	12.1 ± 1.2	$1.01\ ^{+2.45}_{-1.01}$
$\mathtt{qqH}-\mathtt{2j}/\mathtt{m_{jj}}[350,700]$	$15\ ^{+23}_{-15}$	10.5 ± 0.7	$1.41\ ^{+2.21}_{-1.41}$
$\mathtt{qqH}-\mathtt{2j}/\mathtt{m_{jj}} > 700$	$0\ ^{+12}_{-0}$	15 ± 1	$0.00\ ^{+0.77}_{-0.00}$
$\mathtt{qqH}-\mathtt{3j}/\mathtt{m_{jj}} > 350$	$43\ ^{+30}_{-43}$	8.9 ± 0.5	$4.84\ ^{+3.38}_{-4.84}$
$\mathtt{qqH}-\mathtt{2j}/\mathtt{p_T} > 200$	$0\ ^{+3}_{-0}$	4.2 ± 0.2	$0.00\ ^{+0.72}_{-0.00}$
$\mathtt{VH}-\mathtt{lep}/\mathtt{p_T}(\mathtt{V})[0,150]$	$56\ ^{+58}_{-40}$	22.3 ± 1.1	$2.49\ ^{+2.60}_{-1.79}$
$\mathtt{VH}-\mathtt{lep}/\mathtt{p_T}(\mathtt{V}) > 150$	$0\ ^{+10}_{-0}$	3.6 ± 0.1	$0.00\ ^{+2.79}_{-0.00}$
$\mathtt{ttH}$	$0\ ^{+15}_{-0}$	15.9 ± 1.4	$0.00\ ^{+0.91}_{-0.00}$

TABLE 6.6: Best fit values and $\pm\ 1\sigma$ uncertainties for the measured cross sections $\sigma\mathcal{B}$ and the SM predictions $(\sigma\mathcal{B})_{\text{SM}}$ for the stage 1.2 production bins. The results are obtained with m_{H} profiled in the fit.

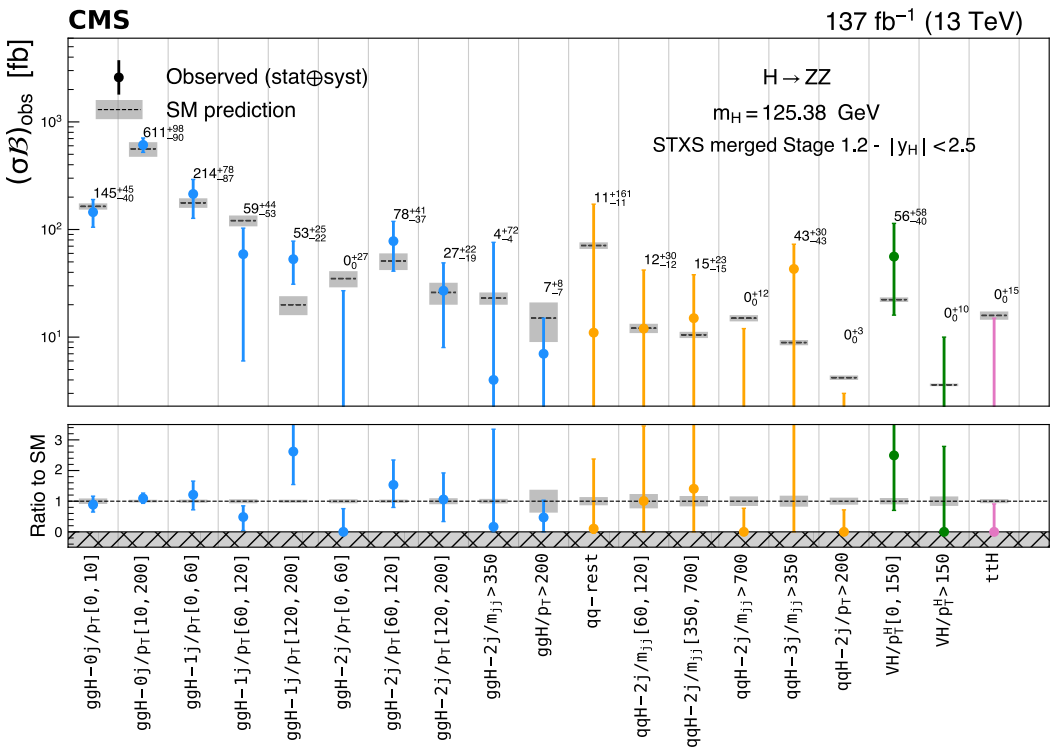


FIGURE 6.13: The measured cross sections $(\sigma\mathcal{B})_{\text{obs}}$ and the SM predictions $(\sigma\mathcal{B})_{\text{SM}}$ for the merged stage 1.2 production bins with m_{H} fixed to 125.38 GeV. The measured values are represented by points with error bars while the SM predictions are represented by black dashed lines with gray uncertainty bands. Ratios of the measured cross sections and the SM predictions are shown in the bottom panel with corresponding uncertainties for each of the bins and the inclusive measurement.

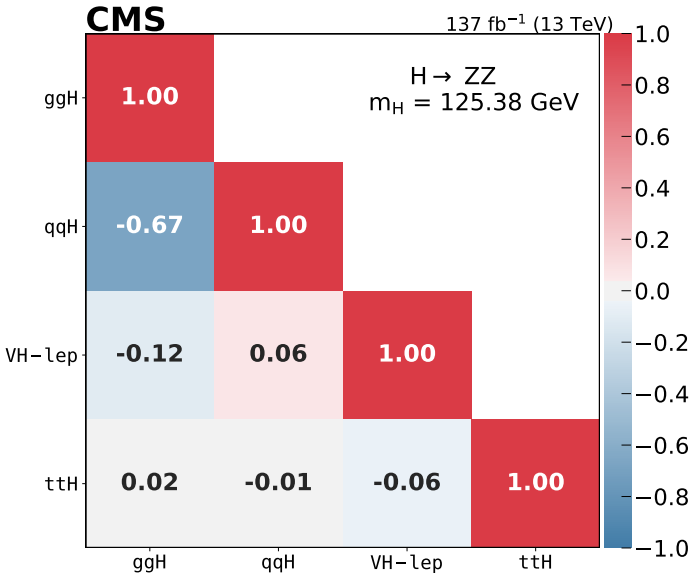


FIGURE 6.14: The correlation matrices between the measured cross sections for the stage 0.

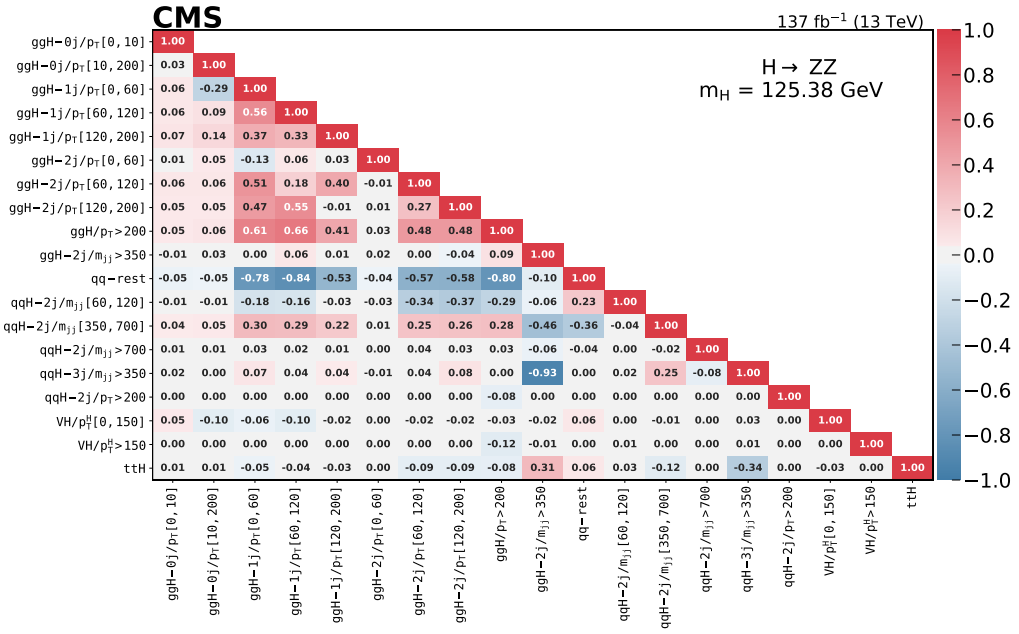


FIGURE 6.15: The correlation matrices between the measured cross sections for the merged stage 1.2.

The dominant experimental sources of systematic uncertainty are the same as in the measurement of the signal strength modifiers, while the dominant theoretical source is the uncertainty in the category migration for the ggH process. The impact of the dominant systematic uncertainties on the STXS measurements is summarized in Table 6.7 and shown in Fig. 6.17. Only dominant experimental and theoretical sources are considered: in the first case, the contribution given by the integrated luminosity uncertainty (Lumi.), the lepton reconstruction and selection efficiency (Leptons), the jet energy scale and resolution (Jet), the b tagging efficiency (B tag), and the reducible background estimation uncertainty (Red. bkg.) is included; in the second case, the ggH and qqH cross section uncertainty scheme (THU), the effect of the renormalization and factorization scale (QCD) uncertainty, the choice of the PDF set (PDF), the branching fraction of $H \to 4\ell$ ($\mathcal{B}$), the modelling of hadronization and the underlying event (Hadr.), and the background modelling (Bkg mod.) are reported. The impact plots with more details on the single contributions to the overall uncertainty for each stage 0 production bin can be found in Appendix D.

	Experimental uncertainties					Theoretical uncertainties					
	Lumi.	Lepton eff.	Jet	B tag	Red. bkg	THU	QCD	PDF	$\mathcal{B}$	Hadr.	Bkg mod.
μ	2.0	5.0	1.0	<0.5	1.0	7.0	0.5	2.0	2.0	1.0	1.0
ggH	2.0	6.0	<0.5	<0.5	1.0	-	<0.5	<0.5	2.0	1.0	<0.5
qqH	2.0	4.0	9.5	<0.5	8.5	-	2.5	<0.5	2.0	15.0	3.0
VH-lep	2.0	10.0	3.0	<0.5	7.5	-	10.0	1.0	2.0	3.0	2.5
ttH	2.0	4.0	2.0	<0.5	10.0	-	6.0	1.0	1.0	20.0	1.0

TABLE 6.7: Impact of the dominant systematic uncertainties (in percent) on the inclusive signal strength and stage 0 production mode cross section measurements. The uncertainties are rounded to the nearest 0.5%.

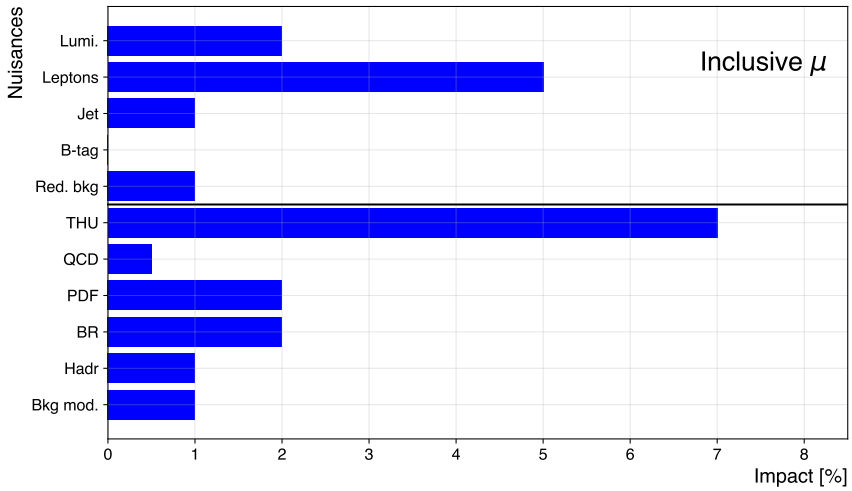


FIGURE 6.16: Impact of the dominant systematic uncertainties (in percent) on the measurements of the inclusive signal strength. The uncertainties are rounded to the nearest 0.5%.

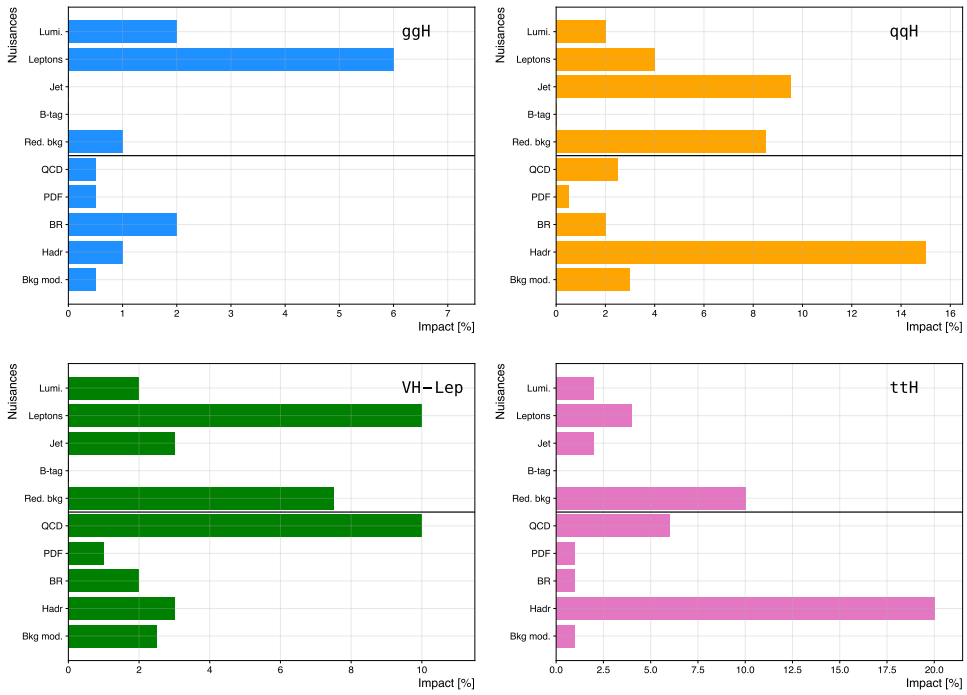


FIGURE 6.17: Impact of the dominant systematic uncertainties (in percent) on the measurements of the stage 0 production mode cross sections. The uncertainties are rounded to the nearest 0.5%.

6.3 Results: fiducial cross section measurements

Based on the measurement strategy presented in the second part of Section 6.1.2, an overview of the results obtained for the integrated fiducial cross section is reported in Table 6.8. The acceptance $\mathcal{A}_{\text{fid}}$ has a strong model dependence and it affects individual H(125) production modes by up to 60%. On the one hand, the reconstruction efficiency (ϵ) represents the fraction of signal events originating from individual SM production modes that come from the fiducial phase space but are not reconstructed. On the other hand, the fraction of events reconstructed outside the fiducial phase space is referred to as f_{nonfid}. Information about the model dependence is provided by the variation of the factor $(1 + f_{\text{nonfid}})\epsilon$ reported in the final column of Table 6.8 when the relative fraction of each production mode is varied within these experimental constraints.
The result of the simultaneous fit to to the $m_{4\ell}$ spectrum is shown for each final state in Fig. 6.19.

The integrated fiducial cross section at $m_{\text{H}} = 125.38$ GeV is measured to be

$$\sigma_{\text{fid.}} = 2.84^{+0.35}_{-0.31} = 2.84^{+0.23}_{-0.22}(\text{stat.})^{+0.26}_{-0.21}(\text{syst.})\ \text{fb},$$

where the corresponding SM expectation is:

$$\sigma^{\text{SM}}_{\text{fid.}} = 2.85 \pm 0.15\ \text{fb}.$$

The measured inclusive fiducial cross section in different final states and integrated as a function of the energy in the center of mass are shown in Fig. 6.18.
The impact of the dominant systematic uncertainties on the measurement is summarized in Table 6.9.

Signal process	$\mathcal{A}_{\text{fid}}$	ϵ	f_{nonfid}	$(1+f_{\text{nonfid}})\epsilon$
gg→H (POWHEG) 125 GeV	0.402 ± 0.001	0.598 ± 0.002	0.054 ± 0.001	0.631 ± 0.002
VBF 125 GeV	0.445 ± 0.002	0.615 ± 0.002	0.043 ± 0.001	0.641 ± 0.003
WH 125 GeV	0.329 ± 0.002	0.604 ± 0.003	0.078 ± 0.002	0.651 ± 0.004
ZH 125 GeV	0.340 ± 0.003	0.613 ± 0.005	0.082 ± 0.004	0.663 ± 0.006
$t\bar{t}H$ 125 GeV	0.315 ± 0.004	0.588 ± 0.007	0.181 ± 0.009	0.694 ± 0.010

TABLE 6.8: Overview of the fraction of signal events for various SM H signal production modes within the fiducial phase space (acceptance $\mathcal{A}_{\text{fid}}$), the reconstruction efficiency (ϵ) for signal events from within the fiducial phase space, and the ratio between events reconstructed outside the fiducial phase space and events from within the fiducial phase space (f_{nonfid}); the factor $(1+f_{\text{nonfid}})\epsilon$ which regulates the signal yield for a given fiducial cross section is also reported. For all production modes the values are given for $m_{\text{H}} = 125\,\text{GeV}$. The uncertainties listed are statistical only; the theoretical uncertainty is smaller than 1% in $\mathcal{A}_{\text{fid}}$ for the SM.

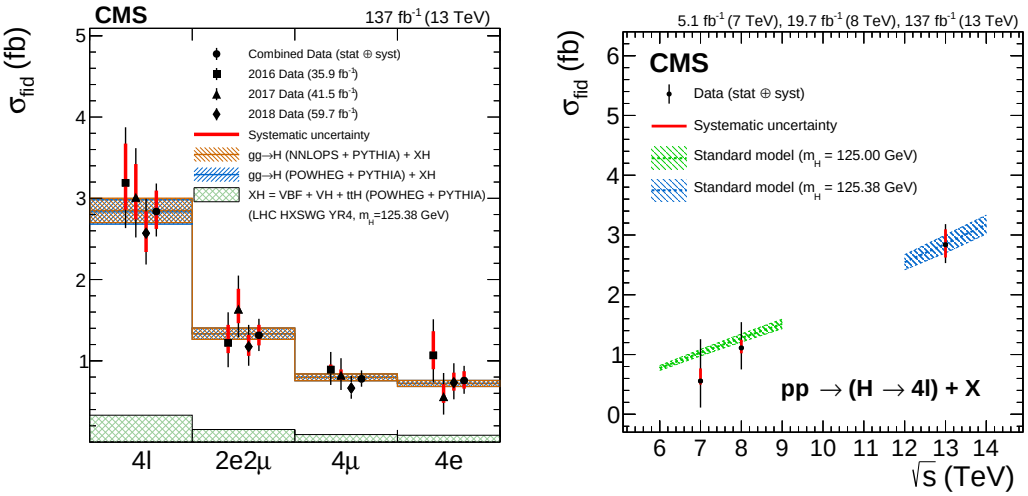


FIGURE 6.18: The measured inclusive fiducial cross section in different final states (left) and the measured fiducial cross section as a function of $\sqrt{s}$ (right). The acceptance is calculated using POWHEG at $\sqrt{s}$=13 TeV and HRES [173, 174] at $\sqrt{s}$=7 and 8 TeV, and the total gluon fusion cross section and uncertainty are taken from Ref. [145]. The fiducial volume for $\sqrt{s}$=6-9 TeV uses the lepton isolation definition from Ref. [104], while for $\sqrt{s}$=12-14 TeV the definition described in the text is used.

	Experimental uncertainties					Theoretical uncertainties					
	Lumi.	Lepton eff.	Jet	B tag	Red. bkg	THU	QCD	PDF	$\mathcal{B}$	Hadr.	Bkg mod.
μ	2.0	9.0	-	-	1.0	-	0.5	0.5	-	-	

TABLE 6.9: Impact of the dominant systematic uncertainties (in percent) on the fiducial cross section measurement. The uncertainties are rounded to the nearest 0.5%.

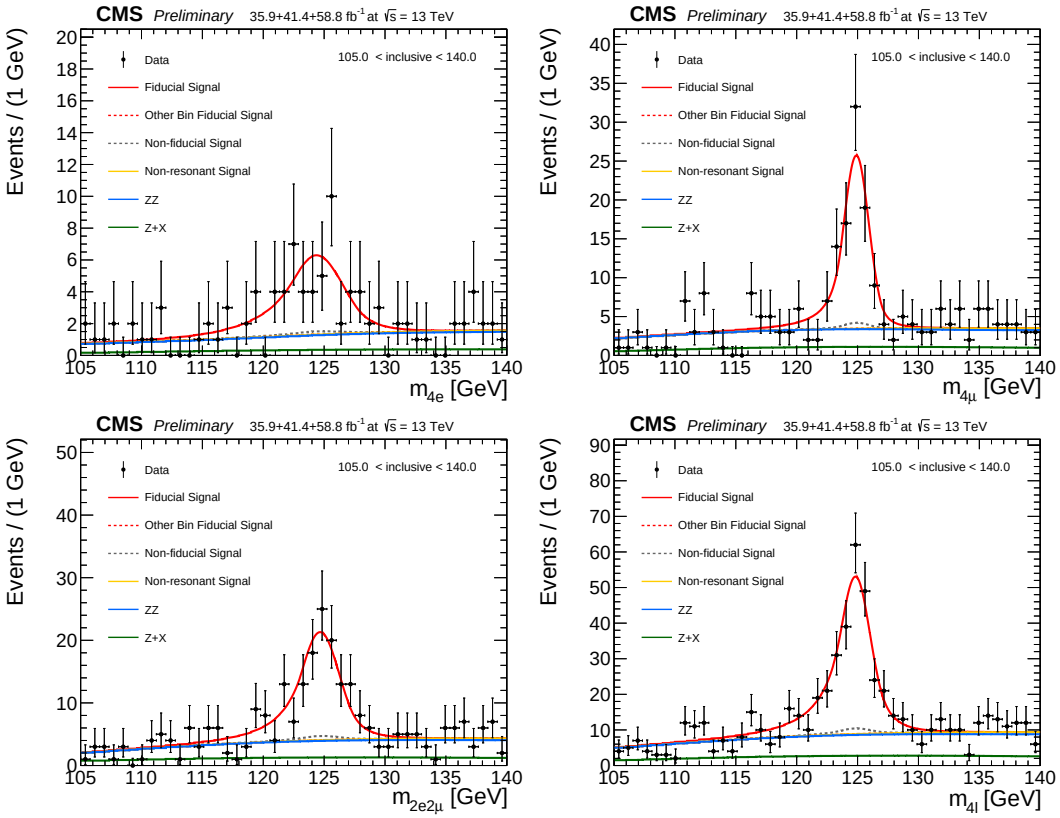


FIGURE 6.19: Result of the simultaneous fit on the 2016, 2017, and 2018 samples for the integrated fiducial cross section measurement in each final state.

The differential fiducial cross sections are measured as a function of several kinematic observables: the transverse momentum of the four-lepton system ($p_{\mathrm{T}}(\mathrm{H})$), the rapidity of the H boson ($|y(\mathrm{H})|$), the transverse momentum of the leading jet in the event ($p_{\mathrm{T}}(\mathrm{jet})$), and the number of associated jets (N(jets)). They are also shown in Fig. 6.20. Examples of the efficiency matrices for the $p_{\mathrm{T}}(\mathrm{H})$, $|y(\mathrm{H})|$, and N(jets) observables in the gluon fusion production mode are reported in Appendix D, considering separately the 2016, 2017, and 2018 scenarios. The lepton identification efficiencies and the luminosity measurement represent the dominant systematical uncertainty in the measurement, while the theoretical contribution to the overall uncertaity is smaller. The unfolding procedure is repeated using different response matrices created by varying the relative fraction of each SM production mode within its experimental constraints in order to assess the model dependence of the measurement. The uncertainty is determined to be negligible with respect to the experimental systematic uncertainties.

6.4 Conclusion and prospects

Several measurements of the H boson properties in the four-lepton final state at a center-of-mass energy of 13 TeV have been presented, probing the consistency with the expectations for the SM H boson within their uncertainties. A dataset corresponding to an integrated luminosity of 137 fb^{-1} has been used to measure the inclusive signal strength modifier μ, the signal strength modifiers for the main H boson production modes, and the integrated fiducial cross section. Results within the framework of the STXS have also been produced, in

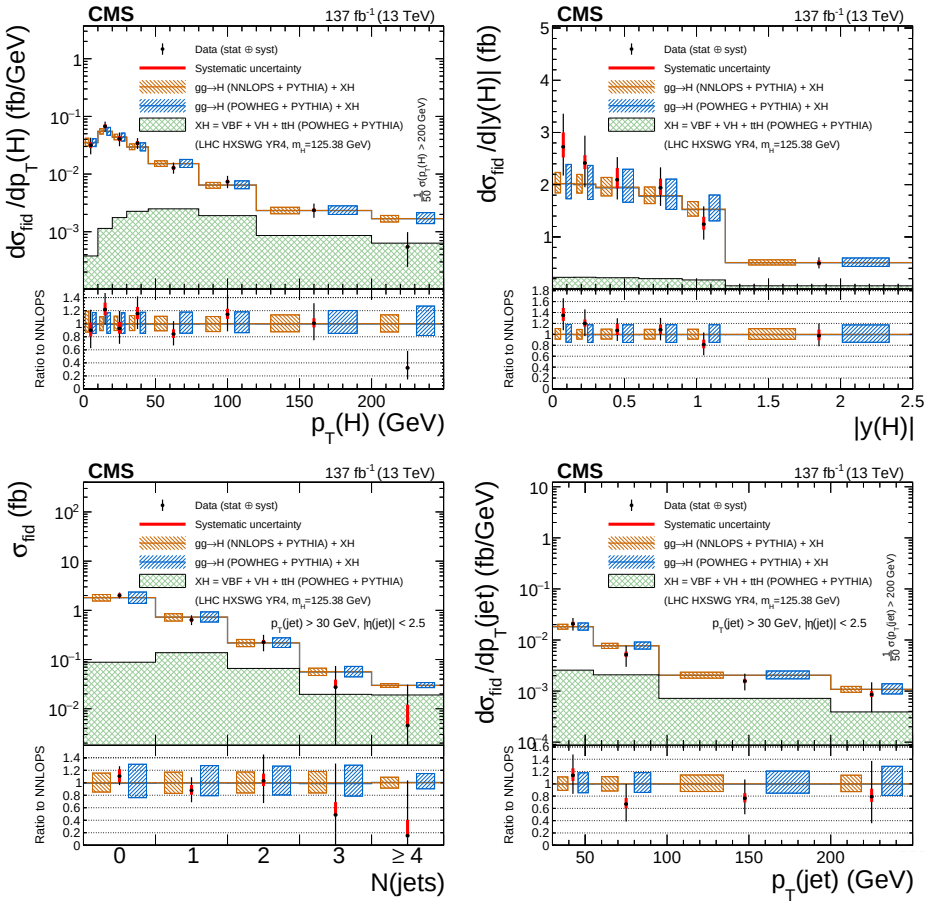


FIGURE 6.20: Differential cross sections as a function of p_T^{H} (top left), $|y_{\mathrm{H}}|$ (top right), N(jets) (bottom left), and p_{T} of the leading jet (bottom right). The acceptance and theoretical uncertainties in the differential bins are calculated using POWHEG. The subdominant component of the the signal ($\mathrm{VBF} + \mathrm{VH} + \mathrm{t\bar{t}H}$) is denoted as XH.

order to quantify the different H boson production processes in specific regions of the phase space. Finally, the differential cross sections as a function of the transverse momentum and rapidity of the four-lepton system, the transverse momentum of the leading jet, and the jet multiplicity are determined.

The $\mathrm{H} \to 4\ell$ analysis has been improved in many aspects during the preparation of the full Run 2 legacy paper. In order to further improve the analysis in view of the LHC Run 3 for the increasing statistics that will become available, some studies started during this work have to be finalized:

* reduction of the electron uncertainty, which is the largest uncertainty source in our measurements. It is due to the efficiency measurement in the low p_{T} bins, for which a J/Ψ sample was not available, and is mainly statistical. One possible solution is to verify if J/Ψ events can be found in our minimum bias samples or $\mathrm{Z} \to 4\ell$ events are enough for the low p_{T} measurement. Otherwise, the development of dedicated trigger to collect those events can be considered;
* the development of a muon identification based on a multivariate analysis, following the approach presented for the electrons;
* the inclusion of an additional VH bin in the event categorization considering the decay of the V boson to invisible. For this purpose, a detailed study of the MET is needed;

- dedicated fake rate measurements for all the STXS stage 1.2 categories instead of an inclusive FR.

In parallel to the work presented in this thesis, studies of CP violation and anomalous couplings of the H boson to vector bosons and fermions have been performed [47] exploiting the four-lepton final state and using the same criteria to select the objects and the events. This work focused on on-shell H boson measurements. The possible extension of the analysis to the off-shell region allows to set joint constraints on Γ_{H} and its couplings. More details can be found in Ref. [102] which includes the results obtained with a partial Run 2 dataset. A full detector simulation of all kinematic effects in the H boson decay and associated particle production is performed, exploiting matrix element techniques to identify the production mechanism and to increase sensitivity to the H couplings. The kinematic information on the H decay into four-lepton and the H production in association with a vector boson, hadronic jets, or a top quark pair is used. Finally, simultaneous measurement of up to five HVV, two Hgg, and two Htt couplings is performed. The results have been produced in the framework of the anomalous coupling measurements, providing also an interpretation in the EFT framework with SU(2)×U(1) symmetry for the HVV couplings. Although I had the opportunity to work closely with the main analyzers of this work, these interesting results are out of the scope of my thesis.

The next step in the context of the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis will be the study of the H boson mass and total width (Γ_{H}), exploiting the full Run 2 dataset. As already mentioned, the current best measurement of the H boson mass comes from the combination of the $\mathrm{H} \to 4\ell$ and $\mathrm{H} \to \gamma\gamma$ analyses [30, 48, 49]. Concerning the width, the precision on Γ_{H} from the latest on-shell measurements of the width of the resonance peak alone is limited by the four-lepton invariant mass resolution, so that it is about 1 GeV. However, the limit on Γ_{H} can be significantly improved relying on the indirect measurement of the H width. This indirect measurement exploits the relative measurement of the off-shell and on-shell production and profits from the fact that the SM H boson width is proportional to the ratio of the event yields in the off-shell and on-shell regions. The Run 1 results by the CMS and ATLAS Collaborations [96, 98] set an upper bound on Γ_{H} of 13 MeV at the 95% CL using the off-shell production method with a predicted value for the SM H boson of Γ_{H} = 4.07 MeV. From the latest CMS result based on the off-shell method and the Run 1 and Run 2 combination (5.1 fb^{-1} + 19.7 fb^{-1} + 80.2 fb^{-1}) [102], a constraint on Γ_{H} was set at $3.2^{+2.8}_{-2.2}$ MeV, while the expected constraint based on simulation was $4.1^{+5.0}_{-4.0}$ MeV. The optimization of the analysis methods presented in this thesis and the increase of the analyzed integrated luminosity given by the the full Run 2 dataset will lead to a substantial improvement in the precision of the H boson total width measurement using the off-shell technique, either under the assumption of SM couplings or with BSM effects.

At the end of this overview of the H boson measurements in the four-lepton final state in the context of the LHC Run 2, one can be sure that the road has to be continued. The $\mathrm{H} \to 4\ell$ analysis will still represent an important benchmark during the Run 3 and the high luminosity phase of the LHC (HL-LHC) given the increasing interest in the precision studies of the Higgs boson properties (see Section 1.3). This item will be discussed more in detail in the next chapter in the context of a specific proposal for the future circular colliders post the LHC era: an electron-positron circular collider.

Chapter 7

Future exploration of the Higgs sector: the IDEA detector

This chapter is oriented to discuss the future exploration of the H sector and the status of a specific detector project. The ongoing development of the design and simulation tools of the IDEA detector (see Section 2.3.3) for the future H boson factory colliders, like FCC-ee or CEPC (see Section 2.3.1), will be exposed.

The process of preparing a proposal for a new detector, including the realization of a Conceptual Design Report of the project, the development of the appropriate technologies and of the software tools, the study of the performances, and the design optimization, is very long, as attested by the story of the LEP and LHC experiments. The possibility to partecipate in this process following step by step the evolution of the project in every aspect is unusual and exciting.
I was involved in this work when I started my Ph.D. working in the international FCC study group and in the Italian INFN RD-FA group, which stands for R&D for Future Accelerators. I had the chance to contribute to different aspects of the development of the IDEA detector concept, both from the hardware and the software point of view. The subdivision in two parts of the chapter follows this approach, with a quick overview of the discovery potential at the FCC-ee at the beginning.
In the first section, the first test beam (TB) on a full slice of the IDEA detector will be described. It was performed in September 2018 in order to test the feasibility of the project, setting a milestone for the development of the detector. During the TB, I focused on the micropattern gaseous detectors (MPGDs) which constituted the preshower and muon detector of IDEA. In the context of the subsequent data analysis, I acted as the contact person for the analysis of the data collected during the TB with the preshower and muon detector.
In the second section, the current status of the simulation tools available to describe the IDEA detector concept will be presented, focusing on both the full (GEANT4-based) and fast (DELPHES-based) simulations. They represent essential tools in order to study the performance of the proposed detector and to optimize its geometry providing a final design.
In January 2020, a meeting between many of the people involved in the process of defining future colliders and detectors took place in Hong Kong. In this occasion, I had the chance to present the status of the software simulation of the IDEA detector on behalf of the IDEA Proto-Collaboration.

7.1 Discovery potential at FCC-ee

Among the different proposals of future colliders, the FCC-ee represents the most ambitious lepton collider at the electroweak (EW) scale. The project design aims to maximize the opportunity of major fundamental discoveries, proposing a broad program of exploration of the Higgs and EW sectors. The FCC-ee enables the measurement of the Z, W, and H boson

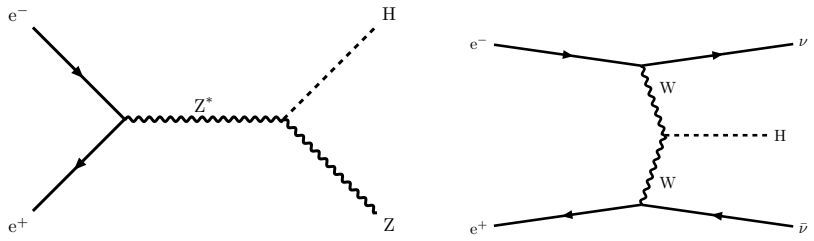


FIGURE 7.1: Feynman diagrams of the Higgs-strahlung production of a H boson accompanied by a Z boson (left), that is the dominant process at a center-of-mass energy of 240 GeV, and of the VBF production mechanism (right), that becomes dominant at high $\sqrt{s}$ values.

properties and the exploration of the top quark physics with high precision on a comprehensive set of EW and H observables. The EW quantum corrections are sensitive to particles with EW couplings and much higher masses, not directly accessible with the center-of-mass energy available. Subsequently, the FCC-ee allows to tightly constrain a large number of the SM parameters, such as the electromagnetic and strong coupling constants at the Z mass scale, thus providing indirect sensitivity to new particles. The FCC-ee also permits to observe possible small deviations or rare new processes with respect to the SM expectations. For this reason, a large interest and animated effort was devoted by many people in the last years to study of the machine and the detector proposals and the collaboration around the FCC study group grew up significantly.

Given that the main focus of the $\mathrm{H} \to 4\ell$ analysis presented in this thesis is the precise measurement of the H couplings to the other SM particles, this introduction will show the expected precision at FCC-ee in order to offer an interesting comparison. At lepton colliders, the dominant production mode at low values of $\sqrt{s}$ is the Higgs-strahlung (HZ), where the H boson is produced in association with a real Z boson, while at high $\sqrt{s}$ values the VBF process becomes dominant. The Feynman diagrams of the two production mechanisms are reported in Fig. 7.1. The H production will be measured inclusively from its presence as a recoil to the Z in the $e^+e^- \to \mathrm{HZ}^* \to \mathrm{HZ}$ process.
The measurement through the recoil mass spectrum is unique and provides a decay-mode independent measurement of the HZ coupling. Under the assumption that the coupling structure presents the same form as in the SM, the cross section is proportional to the square of the H coupling to the Z ($\mathrm{g_{HZZ}}$):

$$\begin{aligned} m^2_{\mathrm{recoil}} &= (\sqrt{s} - E_{\ell\ell})^2 - |\vec{p_{\ell\ell}}|^2 \\ \sigma(e^+e^- \to \mathrm{ZH}) &\propto \mathrm{g^2_{HZZ}} \end{aligned} \tag{7.1}$$

The cross section of the HZ process is maximal around 240-250 GeV, as shown in Fig. 7.2. The FCC-ee will produce one million HZ events with 5 ab^{-1} collected at 240 GeV, adding another 180000 HZ events with 1.5 ab^{-1} of data at 365 GeV FCC-ee. Thanks to the high statistics and the extremely clean experimental environment, a precision of 0.7% on the measurement of the HZ cross section can be achieved. The events can be tagged using the Z decay products without depending on the H decay and thus allowing the absolute measurement of the H coupling to the Z. Consequently, a model-independent determination of its total width and

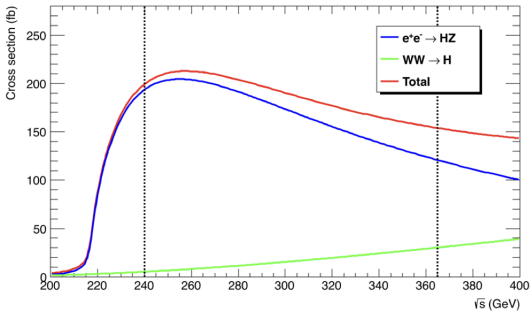


FIGURE 7.2: The H boson production cross section at lepton colliders as a function of the center-of-mass energy. The vertical dotted lines indicate the center-of-mass energies of choice at the FCC-ee for the measurement of the H properties [75].

thus of its other couplings (g_{HXX}) through branching ratio measurements is accessible:

$$\begin{aligned}\sigma(e^+e^- \to \mathrm{ZH}, \mathrm{H} \to \mathrm{ZZ}) &= \sigma(e^+e^- \to \mathrm{ZH})\frac{\Gamma_{\mathrm{H}\to\mathrm{ZZ}}}{\Gamma_{\mathrm{H}}} = \frac{\mathrm{g}^4_{\mathrm{HZZ}}}{\Gamma_{\mathrm{H}}},\\ \sigma(e^+e^- \to \mathrm{ZH}, \mathrm{H} \to \mathrm{XX}) &\propto \frac{\mathrm{g}^2_{\mathrm{HXX}} \cdot \mathrm{g}^2_{\mathrm{HZZ}}}{\Gamma_{\mathrm{H}}}.\end{aligned} \tag{7.2}$$

From the combination of the measurements performed at 240 GeV and at 365 GeV, a precision on the width of the H boson of 1.6% can be obtained. Therefore, the measured HZ cross section and the total width can be exploited to perform absolute measurements of the H couplings to all the SM particles.
The leading H couplings to SM particles, denoted $\mathrm{g}_{\mathrm{HXX}}$ for particle X, will be measured by FCC-ee with a subpercent precision. The relative precision on the H couplings, $\delta g_{\mathrm{HXX}}/g_{\mathrm{HXX}}$, achievable at FCC-ee $\sqrt{s}$ = 240 GeV is reported in Table 7.1, providing about an order of magnitude improvement with respect to the HL-LHC potential. Further improvements on the precision reported here can be obtained by using also the data collected at $\sqrt{s}$ = 365 GeV. Moreover, the measurements performed at HL-LHC can help to constrain a few of the couplings, in particular $g_{\mathrm{H}\mu\mu}$ and $g_{\mathrm{H}\tau\tau}$.

Coupling (%)	FCC-ee at 240 GeV	+FCC-ee at 365 GeV	+HL-LHC
$\delta g_{\mathrm{HZZ}}/g_{\mathrm{HZZ}}$	0.20	0.17	0.16
$\delta g_{\mathrm{HWW}}/g_{\mathrm{HWW}}$	1.30	0.43	0.40
$\delta g_{\mathrm{Hbb}}/g_{\mathrm{Hbb}}$	1.30	0.61	0.56
$\delta g_{\mathrm{Hcc}}/g_{\mathrm{Hcc}}$	1.70	1.21	1.18
$\delta g_{\mathrm{Hgg}}/g_{\mathrm{Hgg}}$	1.60	1.01	0.90
$\delta g_{\mathrm{H}\tau\tau}/g_{\mathrm{H}\tau\tau}$	1.4	0.74	0.67
$\delta g_{\mathrm{H}\mu\mu}/g_{\mathrm{H}\mu\mu}$	10.1	9.0	3.8
$\delta g_{\mathrm{H}\gamma\gamma}/g_{\mathrm{H}\gamma\gamma}$	4.8	3.9	1.3
$\delta g_{\mathrm{Htt}}/g_{\mathrm{Htt}}$	-	-	3.1
$\delta\Gamma_H/\Gamma_H$	2.7	1.3	1.1

TABLE 7.1: Relative precision (in %) determined in the κ-framework for the H couplings to SM particles and total decay width achievable at FCC-ee [75]. In the third column the values obtained combining also the data taken with FCC-ee at $\sqrt{s}$ = 365 GeV are reported, while on the last column the values obtained combining also the results of HL-LHC are listed. All numbers indicate 68% CL sensitivity.

7.2 Test beam of a full slice of the IDEA detector

The concept of the IDEA detector (see Section 2.3.3) is based on innovative detector technologies developed over several years of R&D. However, additional development and optimization is needed for the final design. For this reason, prototypes of the proposed subdetectors have been tested together for the first time exploiting the SPS beam at CERN. I was specifically involved in the work on the micropattern gaseous detectors components of the preshower and muon detector subsystems.

In this section, the overall setup of the TB is presented and results from standalone subdetector measurements and their combined performance are shown. A specific focus is reserved to the MPGDs which were part of the test: the Gas Electron Multiplier (GEM) [200] and the micro-Resistive WELL (μ-RWELL) [86–88] detectors. During the one-week TB period, the commissioning and the calibration of the subdetectors lasted two days, then two and a half days were used to reproduce a slice of the IDEA detector and test all subdetectors together. Finally, two days were dedicated for the calorimeter standalone program, which is not treated in this thesis.

7.2.1 The test beam layout: a combined program

A slice of IDEA was reproduced at the H8 beam line in the experimental area of the SPS north area of CERN (Prévessin site). The origin of the coordinate system was defined as the beam exit point. The x-axis corresponded to the nominal beam line and the y-axis was perpendicular to the x-axis, horizontal with respect to the ground, while the z-axis was perpendicular to the x-y plane, positive to the top vertical direction.

Some prototypes of each kind of subdetector involved in the baseline proposal (see Section 2.3.3) were used: a drift chamber, a preshower detector, a dual-readout calorimeter, and muon chambers. The instrumentation provided by the beam facility, composed of two delay wire chambers (DWCs) and scintillators, was also exploited. The test beam design, reported in Fig. 7.3, accounted for:

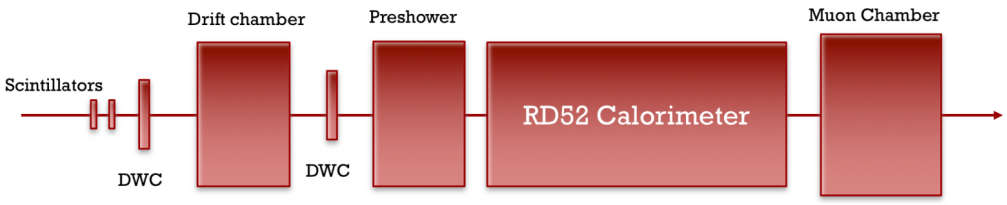


FIGURE 7.3: Schematic view of a slice of the IDEA detector reproduced at the H8 line at CERN.

Drift chamber The drift chamber (DCH) prototype, represented in Fig. 7.4 (left), is a high transparency, single-volume, full-stereo, cylindrical chamber. It consists of twelve layers composed of twelve 1×1 cm^2 drift cells each, for a total of 144 channels. The design of the DCH is inspired by the KLOE experiment prototype [201], later developed as the MEGII drift chamber [202, 203]. The resolution, measured for a similar MEGII prototype, is of the order of 100 μm in the x-y plane, and 1 mm in the z plane. The use of the cluster counting technique, based on the information on the number of clusters (N_{cl}), allows to improve particle identification with respect to the classical dE/dx (stopping power) method [204].

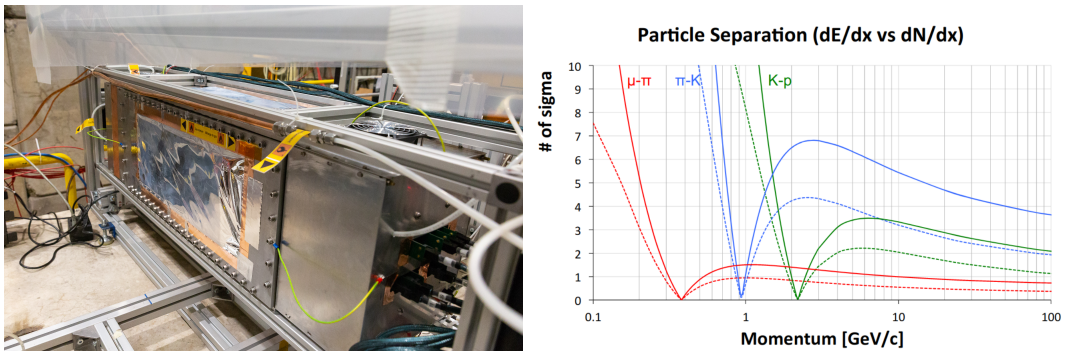


FIGURE 7.4: Drift chamber used during the 2018 IDEA test beam (left) and particle separation as a function of the particle momentum [205].

The dN_{cl}/dx information allows to reach a resolution smaller than 3%, defined as:

$$\frac{\sigma_{dN_{cl}/dx}}{dN_{cl}/dx} = (\delta_{cl} \cdot L_{\text{trk}})^{-\frac{1}{2}}, \tag{7.3}$$

where $\delta_{cl} = 12.5$ cm for He/iC_4H_{10} (90/10) and L_{trk} indicates the length of the track. An analytic evaluation is shown in Fig. 7.4 (right). Moreover, the inclusion of timing techniques can further improve the DCH performance. From the ordered sequence of the electrons arrival times, it is possible to reconstruct the most probable sequence of clusters drift times by considering the average time separation between clusters and their time spread due to diffusion.
All these aspects are currently under study and are not part of the discussion.

Dual-readout calorimeter The dual-readout (DR) calorimeter [89–91, 206], shown in Fig. 7.5 (left), was designed by the CERN RD52 group. It alternates clear and scintillating fibers in a metal matrix, where scintillating fibers are sensitive to all charged particles and clear fibers sense only the Cherenkov light, mostly due to electrons and positrons. It consists of nine lead modules ($9.3 \times 9.3 \times 250$ cm^3 each), with a sampling fraction of 5%. The hadronic energy resolution in a 4π geometry is expected to be around $30\%/\sqrt{E}$ from simulation studies in GEANT4. The readout is made with photomultipliers (PMTs).

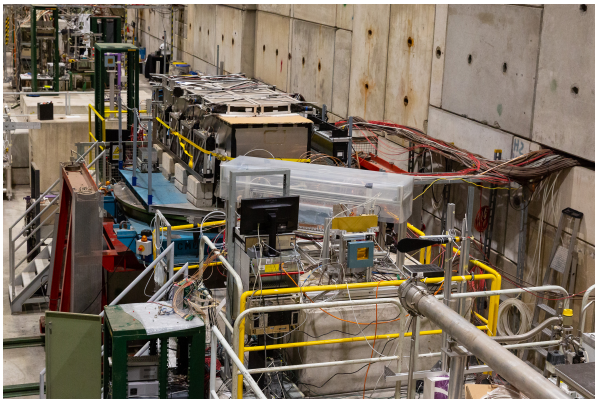


FIGURE 7.5: RD52 dual-readout calorimeter used during the 2018 IDEA test beam. The full slice at the H8 line is also visible.

Preshower and muon chambers The preshower is composed of two triple-GEM detectors (see Section 7.2.2 for details) with a surface of 10×10 cm^2 and strip readout with 650 μm pitch, as shown in Fig. 7.6 (left). The efficiency of this prototype from various past tests is expected to be around 97%, with a spatial resolution of the order of 100 μm.
A lead absorber with variable thickness was placed behind the preshower system in front of the second DWC to verify the effect of varying the material budget, roughly reproducing the presence of the magnet. It was composed of a fixed 5 mm target plus an additional changeable layer of 3, 6, or 10 mm, for an overall radiation length ranging from 1 to 2.5 X_0.

The muon detector, represented in Fig. 7.6 (right), is composed of one triple-GEM detector, similar to the ones of the preshower, followed by two μ-RWELL 10×10 cm^2 prototypes with a pitch of 400 μm. The operating principle of these detectors is the same and all chambers shared the same DAQ system. In addition to the GEM detectors specifications, the μ-RWELL detector has an excellent spatial resolution of around 60-70 μm. More details about these innovative gaseous detectors and working conditions at the TB are given in the next section.

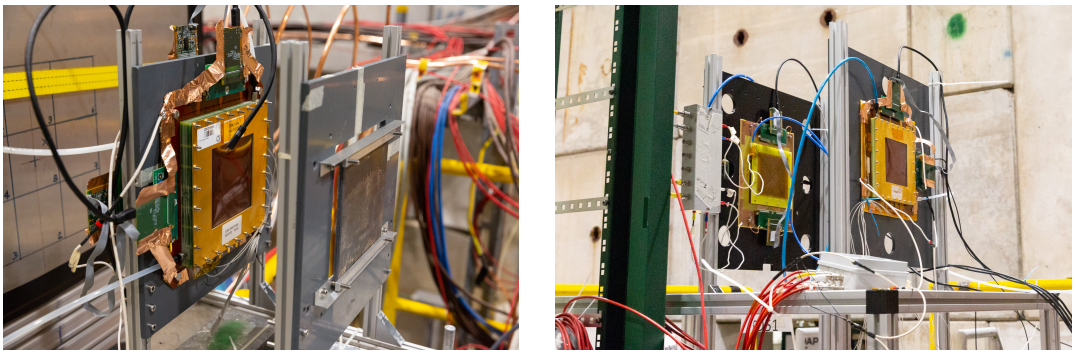


FIGURE 7.6: MPGD components of the preshower (left) and muon detector (right) subsystems used during the 2018 IDEA test beam.

Ancillary detectors Two DWCs were placed before and after the DCH to provide information about the beam direction. Three small scintillation counters provided the signal that were used to trigger the data acquisition (DAQ) systems. Moreover, two additional scintillators were exploited to define a global signal selection able to distinguish between electrons, muons, and hadrons. Finally, leakage counters were installed around the calorimeter to measure the energy leakage.

Trigger and data acquisition system Different beam conditions were available and exploited: muon (40 GeV), electron (from 10 to 60 GeV) and hadron (from 50 to 60 GeV) beams. The trigger was realized using three scintillator counters with the third one featuring a 10 mm radius hole: the first and second scintillators (T_1 and T_2) are represented as a single component in Fig. 7.3, followed by the third scintillators (T_H). The anti-coincidence between these three counters logic signals provided the trigger ($T_1 \cdot T_2 \cdot \bar{T_H}$). Three different data acquisition systems (DAQs) were running in parallel: the central one comprised the trigger, the calorimeter, and the leakage detectors around the calorimeter; the DAQ for the preshower and muon systems made of MPGD prototypes; the drift chamber DAQ. Data from the three different systems were acquired simultaneously during the data-taking and later merged and synchronized offline.

7.2.2 The MPGD-based subsystems

In this section, a description of the experimental setup of the preshower and the muon system during the TB, the GEM and μ-RWELL technologies, and the data-taking procedure is provided. Then, the data analysis performed in the subsequent months is also presented. The detection efficiency of the MPGD detectors used for the TB has been evaluated and the detector response to electron, muon, and hadron beams has been studied varying the lead thickness in front of the preshower. The information from the preshower chambers is exploited to estimate the number of particles produced in the lead by the incoming beam. The main goal of this subdetector in the design of IDEA is to provide an information about the number of particles expected at the calorimeter surface, their position and their energy.

GEM and μ-RWELL in the preshower and muon detector

Gaseous detectors are suitable to cover large surfaces mantaining excellent performances: good efficiency together with spatial and time resolution. In the last years, a new generation of detectors, collectively called Micro Pattern Gas Detectors (MPGDs), is extensively under study and some prototypes have been already adopted in many experimental setups, such as LHCb, CMS, ATLAS, and BES III. The technology at the basis of these new structures makes them very flexible and able to overcome some limitations affecting the traditional Multiwire Proportional Chamber (MWPC) [207] and Micro Strip Gas Counter (MSGC) [208]. The main limiting factors are the stability of the mechanical anode wires structure (especially in large area detectors), the low rate capability, and the fast deterioration due to sustained irradiation. The MPGD detectors represent a robust technology, featuring a clear separation between the anode and the cathode, a localized amplification region, and the capability to fastly evacuate the ions.

In the following, the main features and the operating principle of two innovative devices are summarized: the GEM and μ-RWELL detectors. The former has been used during the TB as a component of the preshower substituting μ-RWELL prototypes which are the proposed technology for the IDEA detector. The latter have been tested with a GEM chamber in the muon system.

The GEM technology The basic unit of a GEM chamber is a thin Kapton foil of 50 μm, metal-coated on both sides with a 5 μm copper layer. The foils are chemically etched forming holes of 70 (50) μm of external (internal) diameter, which occur with high density (tipically 50-100 mm^2) and with a pitch of 140 μm, as shown in Fig. 7.7 (left). The chamber is filled with an appropriate gas mixture, as detailed below, and is designed to work in the avalanche mode [204]. The GEM foil is placed between a drift and a charge collection electrode and develops equipotential field lines near the holes under the application of adequate potentials. When applying a large difference of potential between the two foil faces, a high field is created in the holes (Fig. 7.7). Thus, the charged particles released in the upper region and drifting toward the holes acquire sufficient energy to ionize the molecules of the gas. A multiplication process happens in the holes and particles are then collected by an electrode. A mixture of argon (Ar) and carbon dioxide (CO_2) is particularly convenient thanks to non-flammability and chemical stability. The presence of CO_2 is exploited to absorb electron and photon energy during the avalanche reducing the probability of the propagation of a discharge. The addition of carbon tetrafluoride (CF_4) allows to improve the drift velocity of the gas mixture and, subsequently, the time resolution of the detector. However, it requires special precautions because of its bad environmental effects.

From the combination of various GEM foils, it is possible to build a multi-GEM structure, such as the triple-GEM configuration. The subsequent advantage derives from the

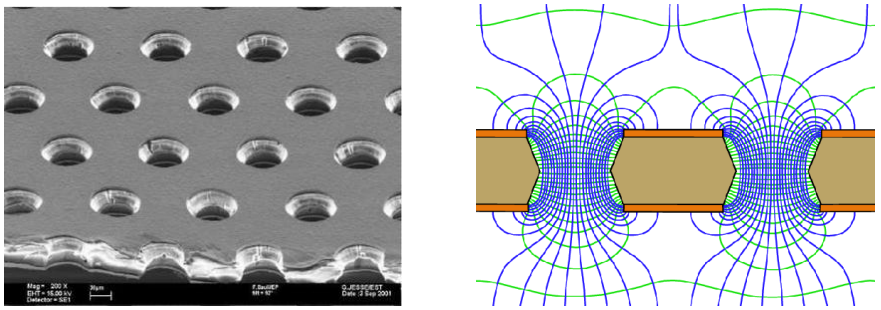


FIGURE 7.7: Structure of a GEM foil observed with an electron microscope (left) and electric field lines in a GEM foil (right) [200].

lower amplification field values applied to the GEM foil, that implies a decrease of the gain but also of the discharge probability. The cascade replication of three GEM foils permits to obtain high proportional gains up to 10^4 with a single amplification factor of at least twenty. Moreover, thanks to the lower field gap between the two faces of the GEM foil, the probability of discharges is highly reduced allowing to preserve the fragile front-end readout electronics. In this case, after the first multiplication process particles are injected into a second multiplying region. The preamplification and transfer process is repeated three times. At the third step, the electrical signal is then readout by a printed circuit board (PCB) equipped with strip electrodes.

The μ-RWELL technology The starting point of the μ-RWELL detector concept is represented by the operating principle of a single GEM foil. The addition of a resistive layer with edge grounding is then oriented to suppress sparks in order to reach higher gains in a single amplification stage. This novel MPGD has been recently developed by the LNF - Detector Development Group (DDG) of Frascati (Italy). It is a compact, simple and robust detector for very large area HEP applications able to operate in very harsh environments. It consists of two parts, the cathode PCB and the μ-RWELL PCB, with the drift region of 3-6 mm corresponding to the gap between the two, as represented in Fig. 7.8. The former is a simple foil of FR_4, a composite material of fiberglass and epoxy resin, with a thin copper layer on one side. The latter represents the core of the detector forming both the amplification stage and the readout. It is composed of three elements:

* a 50 μm thick kapton structure micro-patterned with a top copper layer of 5 μm, which features conical wells with opening (base) diameter of 70 (50) μm. It works as the amplification stage and is manufactured, as for GEMs, with standard photolitographic technique;
* a resistive layer for discharge suppression realized with a 100÷200 nm thin Diamond-Like-Carbon (DLC) which has a surface resistivity of 10-100 MΩ/□;
* a standard PCB readout glued to the DLC.

By applying a proper voltage (see later for detailed values considering a specific case) between the two sides of the kapton foil, the avalanche of the electrons originated in the drift gas gap occurs. The subsequent resistive layer collects the charge and the transition from streamer to spark [204], allowing to safely prevents from sparks and obtain high gains up to 10^4 with a single amplification stage. The current induced on the resistive layer causes a local drop of the amplifying voltage. In fact, the charge collected on the resistive layer flows toward the ground with a characteristic time which depends on the surface resistivity

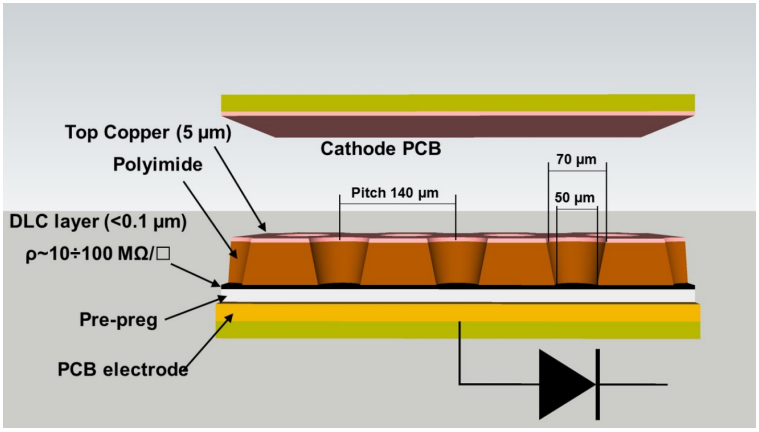


FIGURE 7.8: Structure of a μ-RWELL chamber [87].

(ρ_s) and on the capacitive coupling (c) with the readout:

$$\tau = \rho_s \, c = \rho_s \, \frac{\epsilon_0 \epsilon_r}{d} \tag{7.4}$$

where d is the distance between the resistive and readout layers, ϵ_0 is the vacuum permittivity, and ϵ_r indicates the relative permittivity of the material. As a consequence, the charge depletion time limits the capability of the detector to stand very high radiation fluxes and dedicated charge evacuation schemes are needed to work in a very high rate (HR) environment at the level of 1 MHz/cm^2 or higher. More sophisticated configurations are under study for HR purposes as an alternative to the consolidated single-resistive layer scheme to speed up the charge evacuation. The single-resistive layer scheme has been extensively tested and validated. It was also the prototype used during the TB. However, the IDEA design is based on HR μ-RWELL detectors, which are currently implemented in the IDEA preshower simulation (see Section 7.3.1).

Two different HR prototypes can be realized to improve these features [87]. The first approach relies on the introduction of a double resistive layer. An additional resistive sheet can be connected to the first through high density of metalized vias. Then, a second matrix of vias is used as a connection between the second DLC layer and the readout electrodes to provide the final grounding of the resistive stage. Otherwise, a single resistive layout with a grounding silver grid network, later referred to as *Silver-Grid* layout (SG), can be adopted (Fig. 7.9). A 2D current evacuation scheme is implemented by depositing a conductive grid on the DLC layer to reduce the charge paths to the ground. The conductive grid can be screen-printed or etched by photolitography. In this case, the introduction of small dead zones on the amplification stage above the grid lines is needed to avoid discharges on the DLC over the grids.

The choice of the μ-RWELL technology for the IDEA project is mainly due to the excellent spatial resolution, which can reach 40 μm, combined with high rate capabilities up to 10 MHz/cm^2 in alternative configurations under study. Given the readout scheme presented before, the current μ-RWELL prototypes are able to provide information about one single spatial coordinate. One of the main goal of the ongoing R&D is to design μ-RWELL detectors able to give a 2D information about the particle position.

Details on setup and hit digitization During the IDEA TB, the preshower detector was reproduced using two triple-GEM detectors with a x-y readout of 128 channels per view and a gas mixture of Ar/CO_2/CF_4 (45/15/40), hereafter referred to as GEM 0 and GEM 1. In

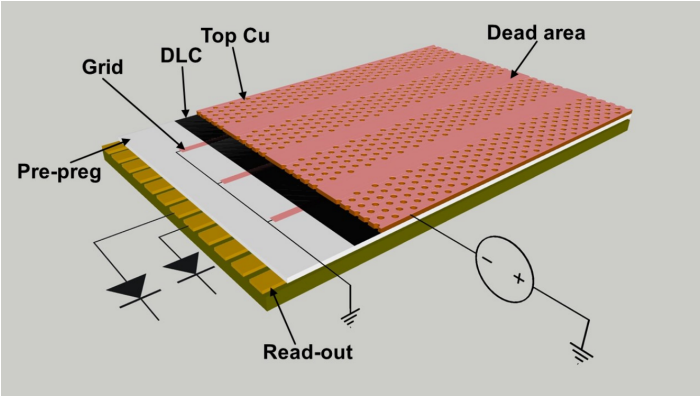


FIGURE 7.9: Sketch of a μ-RWELL PCB in the Silver-Grid layout [87].

the muon detector, a similar GEM chamber (GEM 2) was used with a x-v readout, where v indicates a coordinate in the diagonal plane. The high voltage (HV) setup for each GEM foil is presented in Table 7.2. A coupled HV was used for GEM 0 and GEM 1. During the HV configuration of the detectors, the GEM 0 revealed very low efficiency so that it was impossible to identify a working point to assure the expected high efficiency for both detectors. The chosen working point granted to have high efficiency in GEM 1 without risk to damage the detectors. For this reason, the majority of the results presented below have been produced with the GEM 1.

HV (V)	GEM 0 - GEM 1	GEM 2
Drift	000	000
Foil 1	425	420
Transfer 1	600	600
Foil 2	420	415
Transfer 2	600	600
Foil 3	415	410
Transfer 3	600	600

TABLE 7.2: High voltage scheme of the GEM detectors: GEM 0 and GEM 1 in the preshower and GEM 2 in the muon system.

The muon system was reproduced using the mentioned GEM 2 and two μ-RWELL detectors with 256 strips per view. The first one was placed with the readout in the x direction and the second one with the readout in the y direction in order to provide a bidimensional information on the particle position. The same gas mixture of $Ar/CO_2/CF_4$ (45/15/40) was used and the amplification voltage was set to 570 V and 585 V, respectively.

For the event reconstruction and the analysis of raw data, a customized software framework developed in the last three years by the INFN Ferrara group, called GEM Reconstruction And Analysis Library (GRAAL), was used. A schematic view of the event reconstruction performed within the GRAAL framework is reported in Fig. 7.10. The goal of the first reconstruction is the transformation of the data acquired with the APV25 into collections of hits and clusters, which are then used in the analysis. Contiguous strips on the same detector and view, representing single hits, are collected together to create a cluster. Then, the track reconstruction is realized exploiting the cluster information and the alignment is performed, providing the final reconstructed data for the analysis.

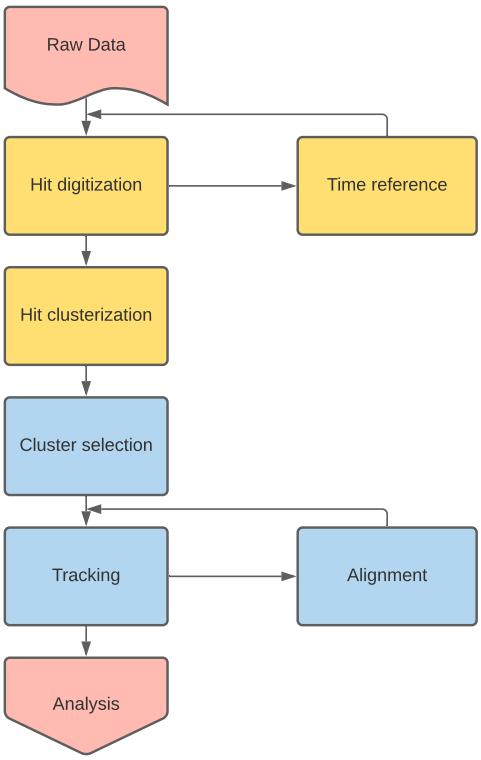


FIGURE 7.10: Flow of the event reconstruction performed within the GRAAL framework.

Alignment and efficiency measurement

The ancillaries detectors have been exploited to align the preshower and muon detectors. To profit of the muon system information, a run with 40 GeV muons has been used used. Only events with one cluster in each chamber have been taken into account. The residuals distributions have been evaluated as $(x_{\text{chamber}} - x_{\text{track}})$ and then used for the alignment applying a shift correction to the chamber position, as shown in Fig. 7.11. The angle between the track and the detectors has to be orthogonal. Thus, the rotation of the entire system is used to constrain the angle to reconstruct an orthogonal track. The first GEM is kept fixed and every other chamber is referred to it. An iterative procedure is used for ten times to obtain the most refined result.

The efficiency measurement have been performed using a muon run. All events that are tagged as muon events and have a track from the DWCs are considered in the denominator. Then, the numerator is filled if a cluster is found within 3σ from the extrapolation of the DWC track to each chamber. Results are shown in Table 7.3. Note that for the GEM in the muon system only the efficiency of the x readout plane is computed because it presents a x-v configuration.

GEM 0 (x)	GEM 0 (y)	GEM 1 (x)	GEM 1 (y)	GEM 2 (x)	μ-RWELL 1 (x)	μ-RWELL 2 (y)
32%	28%	98%	97%	96%	96 %	94%

TABLE 7.3: Efficiency measurement of each readout plane of the GEM and μ-RWELL detectors used in the preshower and muon system.

Cluster studies with muon and electron beams

A thorough study of the clusters created by the incident beam in the GEM and μ-RWELL detectors has been performed in order to understand the best strategy to count the number of particles. For this purpose, data from five runs collected during the TB has been used, corresponding to five different configurations: muon beam without additional lead and electron beam with different lead thicknesses, starting from no additional lead to a maximum of 10 mm of additional lead. The list of runs analyzed with the preshower and muon system MPGD chambers is reported in Table 7.4.

RUN_{MPGD}	RUN_{calo}	Particle	Energy	Lead (mm)	Lead (X_0)
51	12688	Muon	40 GeV	0+5	$\sim$1.0 X_0
71	12709	Electron	20 GeV	0+5	$\sim$1.0 X_0
66	12705	Electron	20 GeV	3+5	$\sim$1.5 X_0
65	12704	Electron	20 GeV	6+5	$\sim$2.0 X_0
64	12703	Electron	20 GeV	10+5	$\sim$2.5 X_0

TABLE 7.4: List of runs analyzed with the preshower and muon system MPGD detectors.

At first, the dimension of the clusters produced in the preshower GEM and the μ-RWELLs of the muon system by an incident muon and electron beam has been compared, showing similar shapes. The comparison of the number of strips fired per cluster by a beam of muons and electron in the configuration without additional lead is shown in Fig. 7.12, together with the distribution of the cluster size obtained in the first μ-RWELL detector. The cluster size peaks at two strips both for muons and electrons looking at the GEM 1. Considering the μ-RWELL chamber, the distribution is slightly shifted on the right and peaks at three, as expected. This is due to the different strip width of the two chambers, which is 650 μm and 400 μm for the GEM and μ-RWELL, respectively. Therefore, the cluster size of the two different subdetectors is compatible.

A muon and electron selection based on the information of the preshower and muon system has been defined to evaluate the purity of the local signal selection with respect to the global one:

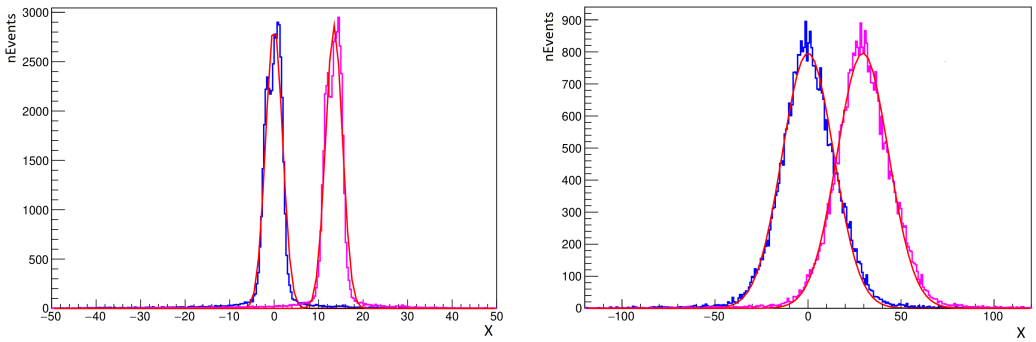


FIGURE 7.11: Residual distribution before (pink) and after (blue) the alignment in the x-view of the second GEM of the preshower (left) and of the first μ-RWELL of the muon system (right). The Gaussian fit of the distributions is also shown in red.

* **Muon selection**: $\geq$ 2 fired layers in the preshower and $\geq$ 3 layers in the muon system in order to assure hits in more than one chamber. "Layer" refers to a single dimension of the chamber (x separated from y).

* **Electron selection** $\geq$ 2 fired layers in the preshower and no particles in the muon system.

The efficiency of the local selection with respect to the global one is found to be around 90% for both electrons and muons in the scenario without additional lead, with a fake rate of 54% and 37%, respectively. Considering the configuration with 10 mm of additional lead, the efficiency of the electron selection increases up to around 100% but the fake rate rises up to 85%.
Given these results, the global selection previously presented has been adopted also for the local analysis because of its better purity. This choice has permitted also an easy comparison of the results with the other subsystems.

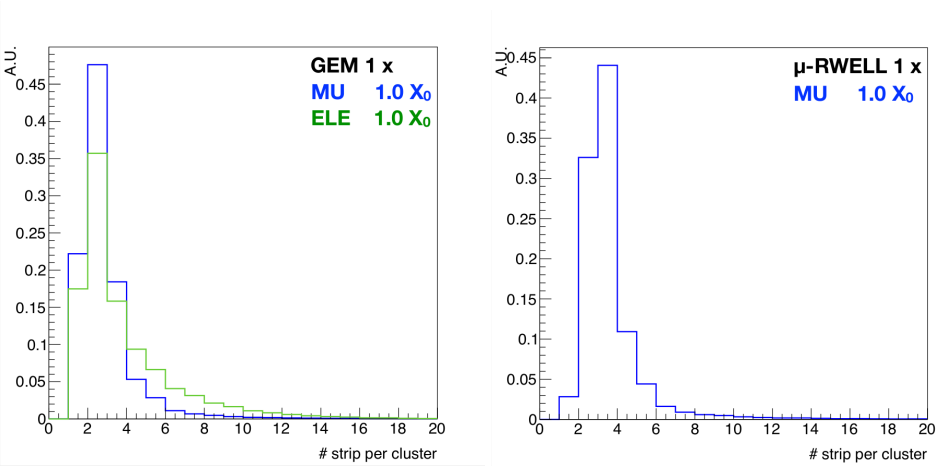


FIGURE 7.12: Comparison of the cluster size for muon and electron beams for the x-direction of the GEM 1 detector (left) and for the first μ-RWELL chamber of the muon system (right) in the configuration without additional lead. Note that the two subdetectors have a different strip width, which is 650 μm and 400 μm for the GEM and μ-RWELL, respectively.

In addition, the variation of the number of clusters and their size as a function of the different lead thickness placed in front of the preshower has been studied. The number of clusters reconstructed in the GEM detectors increases with the increase of the material, while the average cluster size does not change significantly with the increase of the radion length. The variation is found to be smaller than 300 μm. The distribution of the number of clusters and of the cluster size in the x-direction of the GEM 1 are shown in Fig. 7.13. The same study has been done for the y-direction and for the GEM 0 leading to similar results. However, we will focus on the results obtained from the analysis of data collected with the GEM 1 chamber because GEM 0 was inefficient.

Finally, the distribution of the charge released from the particles in the strip-segmented readout has been studied using data collected by the two GEM detectors. In particular, the charge of each cluster has been computed for all lead configurations. It is shown in Fig. 7.14 (left) in decreasing order for the first eight clusters. The average cluster charge for each scenario is reported in Fig. 7.14 (right).
From a comparison between muon and electron runs, no significant effect is visible on the cluster size, while an evident decrease of the charge is visible. Increasing the lead thickness, the charge of all clusters shifts toward greater values.

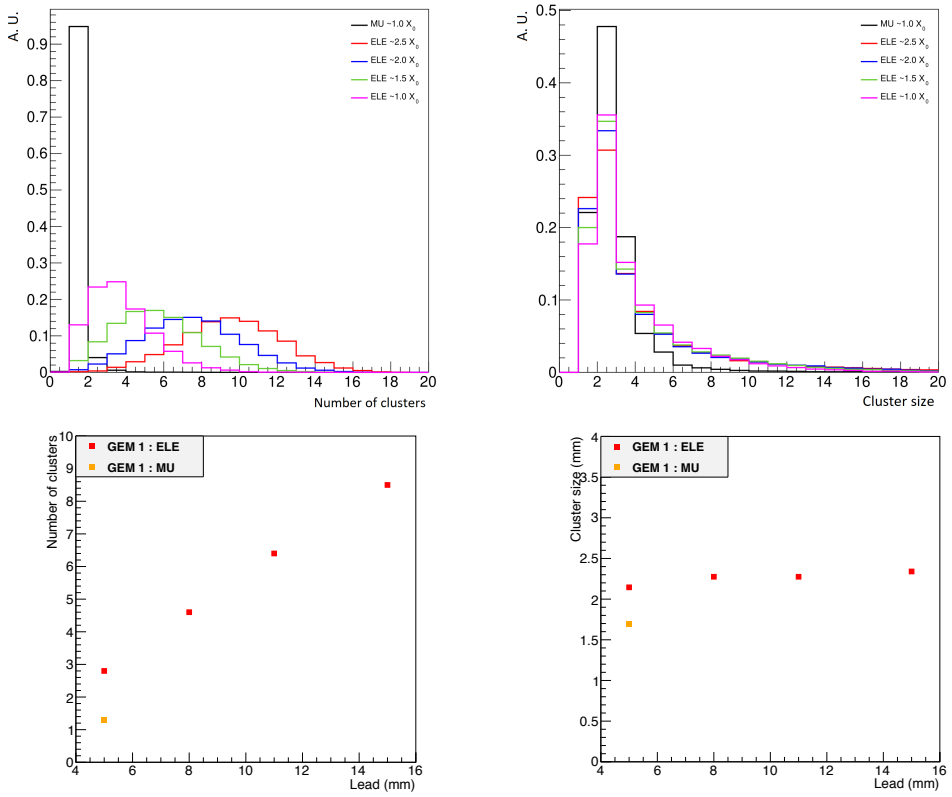


FIGURE 7.13: Number of clusters (top left) and cluster size (top right) distributions in the x coordinate of the second GEM detector of the preshower, considering the electron beams with different lead thickesses and the muon beam as a reference. The corresponding average number of clusters and average dimension (in mm) of clusters is also shown (bottom), considering both electrons (red) and muons (orange).

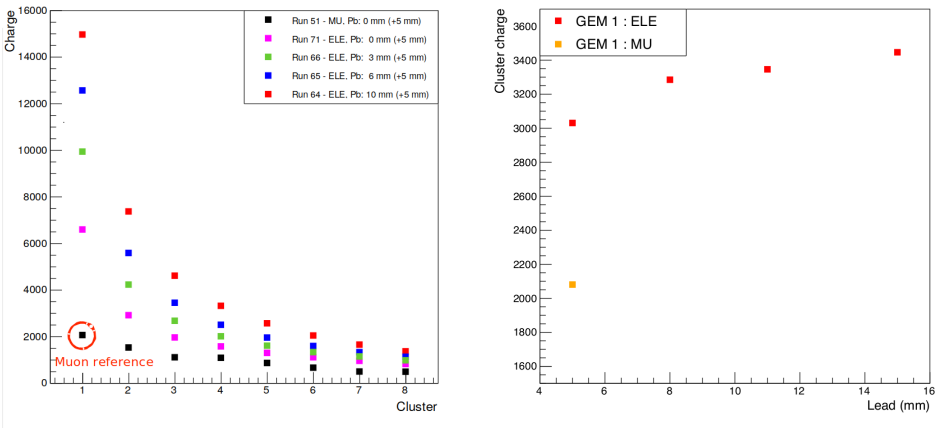


FIGURE 7.14: Comparison of the cluster charge in decreasing order for the first eight clusters considering electrons with and without lead and muons without lead (left). Muon charge is taken as a reference in the particle counting. The average cluster charge for each scenario is also shown (right): muon beam without additional lead and lead thickness scan with electron beam.

Particle counting

From the studies presented above, different methods oriented to count the number of particles produced in the lead by the incoming beam have been tested. The integrated charge is proportional to the energy released by the electromagnetic shower and thus to the number of particles. Moreover, it has been verified that the larger clusters correspond to clusters with highest charge. The choice of the method of counting has been guided by this correleration between the number of particles and the charge and the size of the clusters.

The final strategy adopted to provide the number of particles (n) expected at the calorimeter surface is:

1. Define the muon average cluster charge of the highest charge cluster as a reference (Q_{ref}) to identify one single particle, as shown in Fig. 7.14 (left).
2. Divide the cluster charge of electron clusters (Q_{ref}) by the muon reference cluster charge.
3. Apply a correction to take into account the number of saturated strips n_{sat}. If the ratio Q_{ref}/Q_{ref} is smaller than n_{sat}, n=n_{sat}, otherwise n is rounded to the lower integer (assuming n=1 in case n<1).

This approach is referred to as "charge method". Given the similar behaviour observed for the cluster size looking at the first eight clusters in each event (Fig. 7.15), the same procedure can be used exploiting the size of electron clusters divided by the reference muon cluster size. In this case, the approach is called "size method" and the correction for the number of saturated strips is not applied. From a simulation of the energy release of electrons and muons in the 5 mm gas volume of a GEM or μ-RWELL chamber with a gas mixture of $Ar/CO_2/CF_4$ performed with Garfield[1], it has been demonstrated that electrons and muons produce the same amount of primary electrons. Therefore, it is reasonable to assume that the greater cluster size is due to electrons produced in the lead.

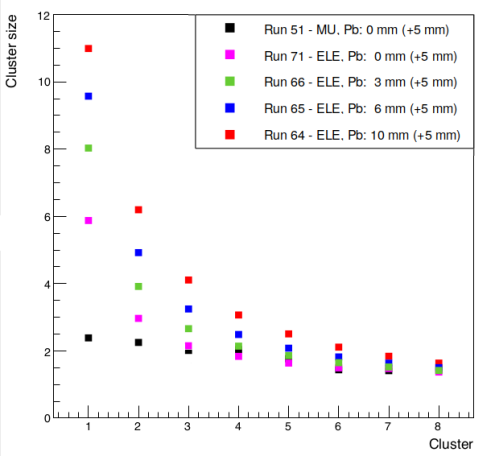


FIGURE 7.15: Comparison of the cluster size in decreasing order for the first eight clusters.

To confirm the effect of different lead thicknesses on the electrons, a GEANT4 simulation of the two lead volumes and the GEM surface has been realized. Then, the passage of 1000 electrons with an energy of 20 GeV has been simulated to verify the expected number of particles on the GEM 0 surface. The comparison between results obtained from the two methods and from the GEANT4 simulation is presented in Fig. 7.16.

[1]Software generally used for the simulation of micropattern gaseous detectors based on C++.

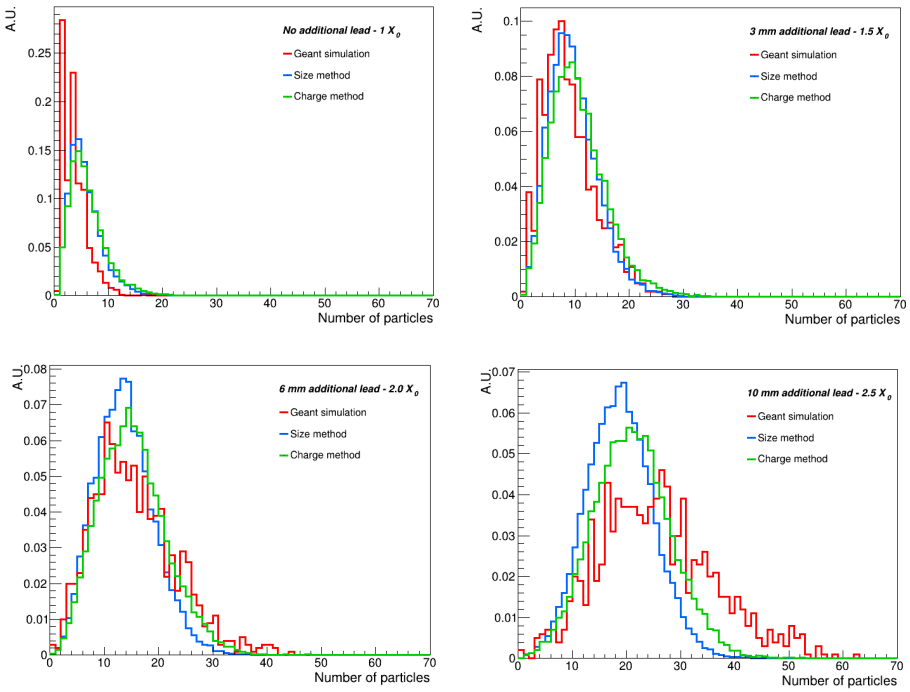


FIGURE 7.16: Comparison of the number of particles obtained from the GEANT4 simulation, the size method, and the charge method for different lead thicknesses: $\sim$1.0 X_0 (top left), $\sim$1.5 X_0 (top right), $\sim$2.0 X_0 (bottom left), and $\sim$2.5 X_0 (bottom right).

The reliability of the simulation has been checked by applying the same clustering algorithm used in data to the simulated events applying a minimum energy cut of 5 MeV to electrons. An average cluster size of 1.7 mm is used for the clustering as obtained from data in the muon run. The average number of clusters between simulation and data is in agreement, as reported in Fig. 7.17. Thus, an alternative promising method to count particles is to perform a clustering from simulation data and then to use an inverse function to compute the number of particles from data. This procedure can be further investigated in the future. In fact, it would require a lot of work to simulate the preshower detector reproducing the same TB conditions, while the goal of this TB was a feasibility study of a MPGD-based preshower as a counter.

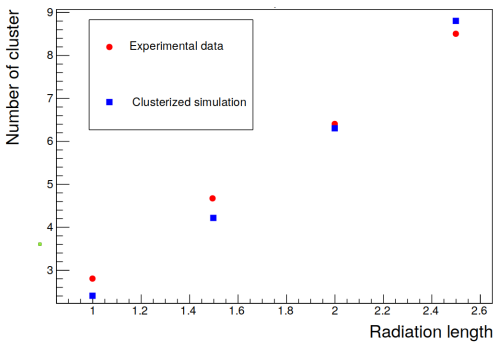


FIGURE 7.17: Average number of clusters in experimental data and as a result of the clustering in simulated events.

7.2.3 Results of the combined program

The primary goal of the TB was to test all subdetectors together and exploit the whole detector to obtain complete information about the tracks. For this reason, correlation studies between quantities registered by different DAQs have been performed to be sure that the DAQ systems of all subdetectors were properly synchronized. As an example, the correlation between the y position of events measured by one GEM detector of the preshower and the y coordinate acquired by the first DWC is shown in Fig. 7.18 (left). The correlation is clearly visible and is confirmed in the x coordinate, too. Then, data from all subsystems have been merged offline in order to perform combined studies.

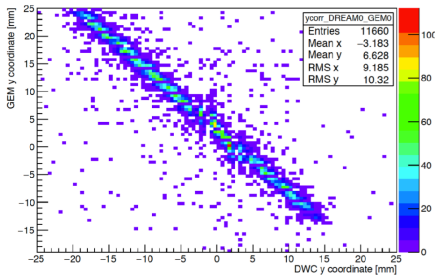


FIGURE 7.18: Correlation plot of the y coordinate of events measured by the GEM detector of the preshower versus the y coordinate acquired by the DWC.

To evaluate the energy response of the calorimeter with different lead thicknesses in front, scintillation channels were studied with an electron beam of 20 GeV. The energy response without additional lead (fixed 5 mm only) and with maximum lead thickness (5+10 mm) is represented in Fig. 7.19. There is no noticeable impact of the different lead thicknesses on the mean energy and resolution of the Gaussian fit performed in the range between -1.5σ and +1.5σ. However, a small impact of the different lead absorbers is visible in the shower shape. The energy-weighted shower width, shown in Fig. 7.20 (left), is evaluated as:

$$\mathrm{R_W} = \frac{\sum(E_{ch} \cdot \sqrt{x_{ch}^2 + y_{ch}^2})}{\sum E_{ch}} \tag{7.5}$$

according to the energy and the position of each channel (ch). Each channel corresponds to a calorimetric tower, which is a group of fibers.

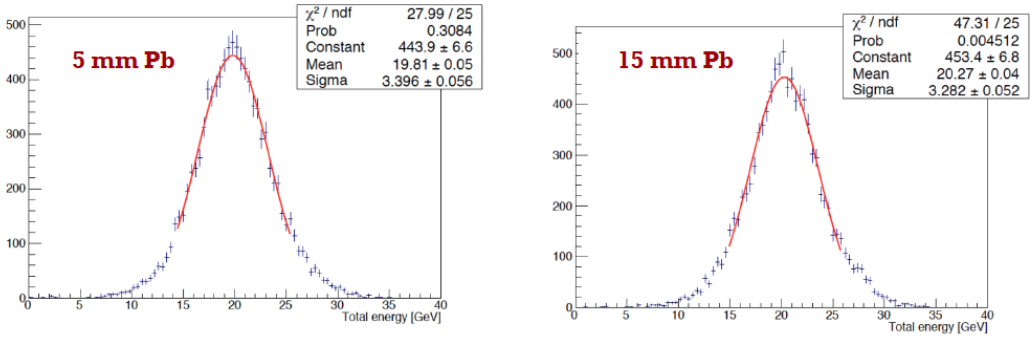


FIGURE 7.19: Calorimeter energy response without additional lead (left) and with maximum lead thickness (right).

As explained in the previous section, the main goal of the preshower is to provide information about the position and the number of particles generated in the material before the calorimeter and reaching the calorimeter surface. According to the studies performed with the preshower data, the number of clusters increases with the material while the dimension of the clusters remains almost constant. The correlation between the number of clusters in the preshower and the shape measured in the calorimeter has been verified, as shown in Fig. 7.20 (right).

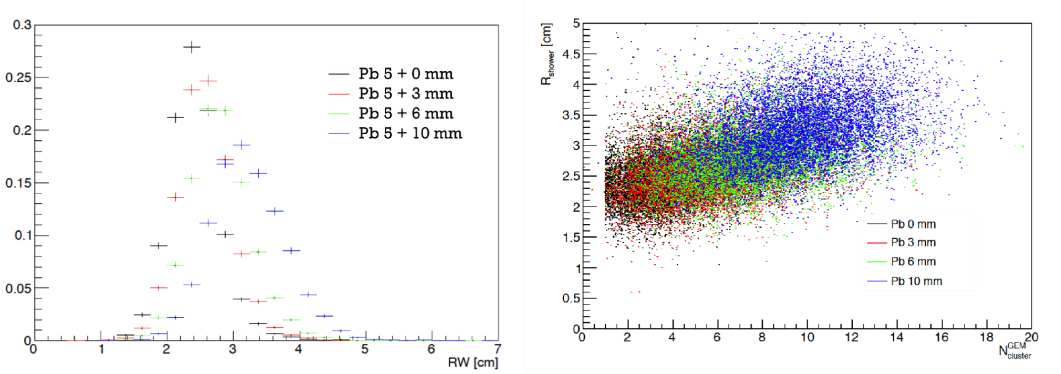


FIGURE 7.20: Energy-weighted shower width (R_W) measured in the calorimeter for different lead thicknesses (left) and its correlation with the number of clusters present in the second GEM of the preshower (right).

The charge and the size of the clusters can be used to provide a reliable estimate of the number of particles reaching the calorimeter. Moreover, an extrapolation procedure from the GEM 1 to the calorimeter surface has been performed for the scenario without additional lead exploiting the information of the DWCs to reconstruct the track direction (Fig. 7.21). The position is provided starting from the center of the cluster. As a closure test of the reliability of the information provided by the preshower, the expected position on the calorimeter surface should be compared with the calorimeter data.

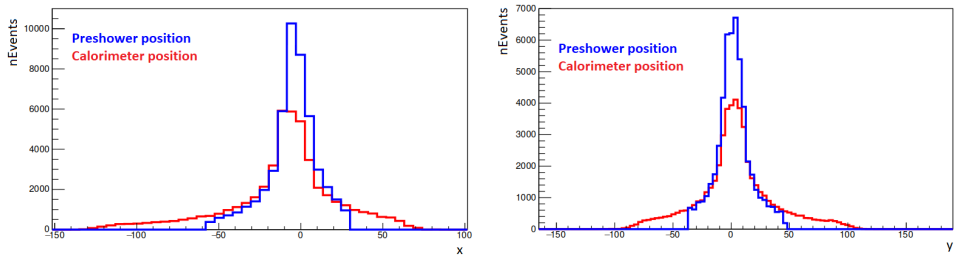


FIGURE 7.21: Expected position of particles detected in the preshower on the calorimeter surface in the x (left) and y (right) directions.

The feasibility of a preshower made of MPGD chambers has been successfully proved. The comparison between the number of particles expected by the simulation and found in data indicates that it can be exploited as a counter in the IDEA detector. Additional studies are needed to investigate the possible contribution of the preshower to the particle identification.

7.3 The simulated description of the IDEA detector

From the software point of view, most of the ongoing work is focused on the simulation of IDEA. It is crucial to study the optimization of the IDEA design and to improve the reconstruction of each subdetector. Considering the significant interplay between full and fast simulation in the process of defining a detector design, two aspects are presented in the current section: the implementation of a standalone full simulation of the IDEA detector in GEANT4 [171] and the description of the detector characteristics in the DELPHES [209] fast simulation.

The study of the detector performances is typically performed using some benchmark physics channels. For this purpose, the combined performance of the IDEA subdetectors have to be investigated with a complete full simulation. To guarantee the quick development of the IDEA geometry, it was decided to merge all IDEA subdetectors in a standalone GEANT4 geometry. This approach aims to facilitate the importing of the detector description in a common framework for the FCC and CEPC projects. In parallel, the implementation of IDEA in the DELPHES fast simulation framework has been realized, based on the output of the dedicated GEANT4 simulation of the IDEA tracker and DR calorimeter. Consequently, the two parts of this section are intimately connected. The DELPHES software demonstrated to be flexible enough to provide a fast simulation of the IDEA detector. In addition, significant improvements can be obtained by the inclusion of the full covariance matrix to provide tracks smearing and by the development of new algorithms dedicated to e^+e^- physics.

7.3.1 Status of the IDEA standalone full simulation in GEANT4

A complete description of the tracker system and of the DR calorimeter proposed for the IDEA detector concept is already available and fully exploited for performance studies. In addition, during the last year an effort focused on the implementation of the MPGD-based IDEA subdetectors, the preshower and the muon system, started in order to collect all the pieces needed for a complete standalone GEANT4 simulation of IDEA.
In the meanwhile, the project to move toward a common software framework for future experiments, named *key4hep*, has started.

The tracker system

The IDEA central tracker simulation includes the inner-outer-forward components of the silicon vertex detector, the 400 cm long drift chamber, composed of 112 layers, and the silicon wrappers. A representation of the simulated DCH is reported in Fig. 7.22 together with a schematic view of the various tracker components. While the drift chamber is well described in its geometry details, the silicon detectors are simulated as a simple layer and overall equivalent material.

The signal hits creation is simulated for the silicon detectors and the wire chamber. On the one hand, all the coherent hits in a cell of the drift chamber are grouped together to create a hit with the proper distance of closest approch (DCA) smeared with a resolution of 100 μm. On the other hand, the hits in the silicon detectors are translated in pixel/strip information. Considering the full tracking performance, a transverse momentum resolution of 2.2×10^{-5} can be reached (Fig. 7.23), improved by a factor of 2.5 with respect to not including the silicon wrapper information. Being a very light and transparent detector, the contribution of the multiple scattering is low.

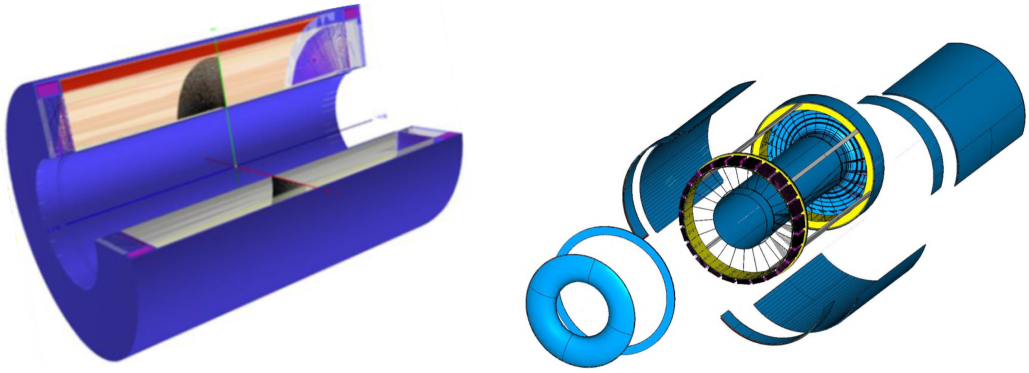


FIGURE 7.22: Simulation of the IDEA drift chamber geometry in GEANT4 (left) and the full tracking system (right) with its different components: vertex detector, drift chamber, and silicon wrappers.

The dual-readout calorimeter

The detailed simulation of a fully projective tower-based DR calorimeter, shown in Fig. 7.24, has been developed during 2019. The simulation work is largely based and validated using TB data coming from the RD52 Collaboration prototypes [91, 210]. Its geometry is highly modular. It is composed of 5400 copper-based towers 2 m long, corresponding to about 9 λ, and it is unsegmented in the longitudinal direction. Towers are trapezoids with slightly different shapes changing with θ. The whole calorimeter consists of about 130 million fibers featuring 1 mm diameter and 0.5 mm of absorber material (copper) between two adjacent fibers.

Equalization constants are extracted per each tower by sending electrons of known energy at their center and collecting signals (photoelectrons) at the end of the fibers. Then, calibration constants are rescaled to take into account the different response while sampling electromagnetic shower tails. From the combination of the scintillation and the Cherenkov signals, which sample the electromagnetic shower independently, the electromagnetic energy resolution (Fig. 7.25) is obtained:

$$\frac{\sigma}{E} = \frac{11\%}{\sqrt{E}} + 0.8. \tag{7.6}$$

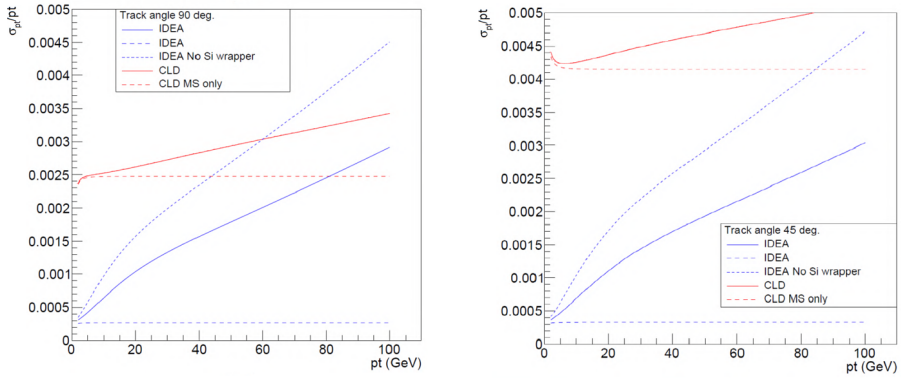


FIGURE 7.23: Transverse momentum resolution of the IDEA tracker system at 90 degree (asymptotic behaviour) on the left and at 45 degree on the right (solid blue line) compared with the CLD tracker performance (solid red line). The contribution of the multiple scattering only is also shown (dotted lines).

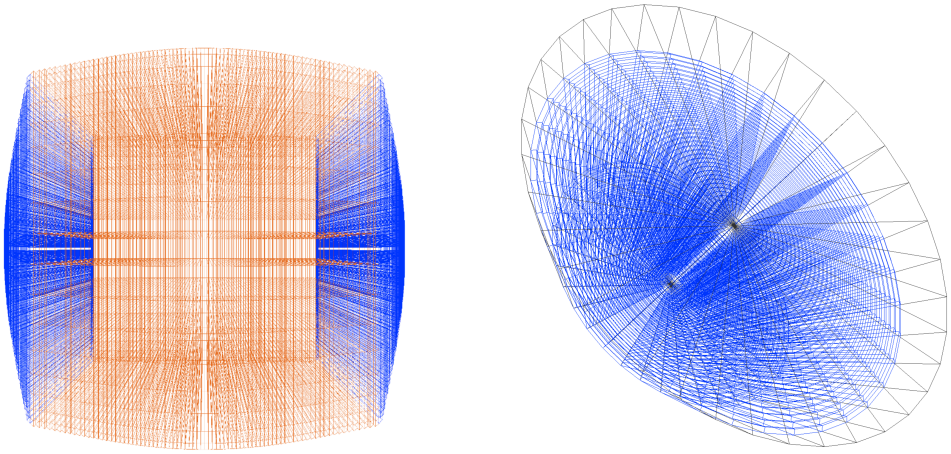


FIGURE 7.24: Simulation of the DR calorimeter in GEANT4: barrel (left) and endcap (right) [210].

In addition, the DR calorimeter relies on a very good angular resolution. Thanks to the extremely high granularity coming from the fibers, the impact point in the calorimeter can be reconstructed very precisely. The reconstruction of the shower barycenter exploits both S and C signals, assuming that the readout of each fiber is individually performed with a SiPM. Firstly, the barycenter is evaluated separately for S and C; thus, the shower position is given by the average of the S and C coordinates. The expected θ and ϕ angular resolutions are respectively:

$$\begin{aligned} \sigma_\theta &= \frac{1.4}{\sqrt{E}} + 0.018\,(\text{mrad}), \\ \sigma_\phi &= \frac{1.8}{\sqrt{E}} + 0.088\,(\text{mrad}). \end{aligned} \tag{7.7}$$

Finally, single hadron performance has been studied and the DR compensation applies well to hadrons fully simulated with GEANT4. The combined resolution on 20-80 GeV charged pions is shown in Fig. 7.25 and is found to be close to $\frac{\sigma}{E} \simeq \frac{30\%}{\sqrt{E}}$. Further studies on how to optimize this resolution are ongoing both for single hadrons and jets. However, all these items are not treated here. Some details can be found in Ref. [210].

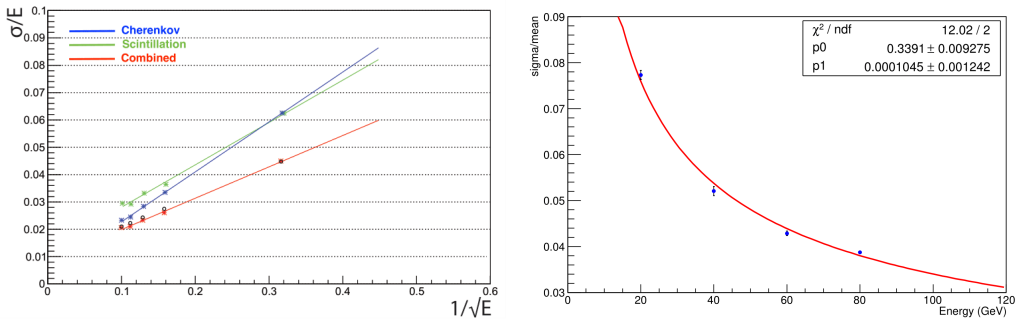


FIGURE 7.25: The electromagnetic (left) and hadronic (right) resolution of the DR calorimeter in GEANT4 [210].

The preshower

The full barrel geometry of the preshower has been built exploiting the basic module of the MPGD-based IDEA subdetectors, a μ-RWELL chamber. The 2 T thin solenoid made of absorber material (lead) in front of the preshower is also included in the simulation. Given that I was responsible for the realization of this part of the IDEA detector and I started from scratch, a specific focus is reserved to the preshower implementation in GEANT4.

The geometry of the barrel account for two layers of μ-RWELL detectors, as shown in Fig. 7.26, in order to provide a x-y information on the particle position. This is due to the fact that current prototypes of μ-RWELL provide information on one single spatial coordinate, as explained in Section 7.2.2. However, one of the main goals of the ongoing R&D is to design μ-RWELL detectors with a bidimensional readout.

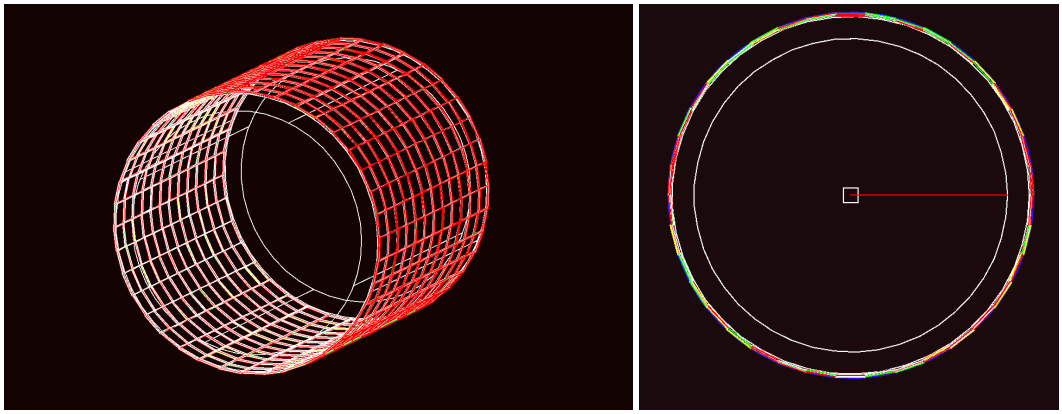


FIGURE 7.26: The full barrel geometry of the preshower including also the 2 T thin solenoid which surrounds the tracker in the longitudinal (left) and transverse (right) view.

The implementation of the basic module of the MPGD-based IDEA subdetectors, a μ-RWELL chamber with the high rate layout and the SG2++ configuration (see Section 7.2.2), has been realized. The size of each chamber is 50×50 cm^2. An occupancy study has been performed using a 2D CAD software to understand the real active area of the 50×50 cm^2 chamber taking into account cables, connectors, and APVs. The active area has been evaluated to be around 41.3×41.3 cm^2, which is the size adopted in the current implementation. Considering smaller chambers with the exclusion of the service area, an uniform geometry can be assumed to cover the whole barrel.
Chambers are placed in a cylindrical and staggered geometry with the cathode pointing toward the interaction point. The structure of each chamber accounts for the cathode, the drift gap, and the μ-RWELL with the readout PCB (see Section 7.2.2). Thicknesses and materials of each layer of the μ-RWELL chamber, schematically presented in Fig. 7.27, are reported in Table 7.5. The overall thickness of the chamber amounts to 9.46 mm.

The description of the copper and kapton materials is contained in the default GEANT4 libraries. The Diamond-like-Carbon (DLC) is defined as a new material with carbon density of 2.00 g/cm^3); the same density is assumed to describe the pre-preg layer, which is a film glue. The FR4 is a composite material of fiber glass (60%) and epoxy (40%) with a density of 1.99 g/cm^3 and 1.25 g/cm^3, respectively. It is simulated as permaglass with FR4 density (1.85 g/ g/cm^3). The gas mixture $ArCO_2CF_4$ is realized using argon and CO_2 from GEANT4 libraries, with a density of 1.661 kg/m^3 and 1.842 kg/m^3, respectively, together with CF_4, which is implemented as a new material with density: 3.78 kg/m^3. The density of each component is weighted according to their volume percentage (45/15/40) by defining fraction mass values: f_{Ar} = 0.295, f_{CO_2} = 0.109, and f_{CF_4} = 0.596. Finally, in order to take into

μ-RWELL component	Thickness of each layer	Material
Cathode	1.6 mm	FR_4
	35 μm	Copper
Gas gap	6 mm	$ArCO_2C_4$
μ-RWELL + readout PCB	5 μm	Copper
	50 μm	Kapton
	0.1 μm	DLC
	35 μm	Copper
	100 μm	Film glue (same DLC density)
	35 μm	Copper
	1.6 μm	FR_4

TABLE 7.5: Thickness and material of each layer of the μ-RWELL chamber implemented in GEANT4. Copper and kapton (in blue) description takes into account holes and dead zones on the amplification stage, while the copper in the readout PCB (red) accounts for strips.

account holes and dead zones in the amplification stage, copper and kapton density have been redefined by using a weight to distinguish active and dead areas on the amplification stage. Shape, diameter, thickness, and pitch of the holes and size and pitch of the strips are considered to define this weight. The final result of the implementation of a μ-RWELL chamber in GEANT4 is shown in Fig. 7.27.

Considering the prototype described here, the barrel geometry is made of twelve detectors for each sector along the beam axis and of thirty-eight sectors to cover the cylinder. A dedicated effort to design and implement the description of the endcaps will follow up this preliminary version of the preshower. Then, having the minimal unit of the μ-RWELL subdetectors implemented in GEANT4, the next step will be the construction of the muon system using the preshower implementation as a basis.

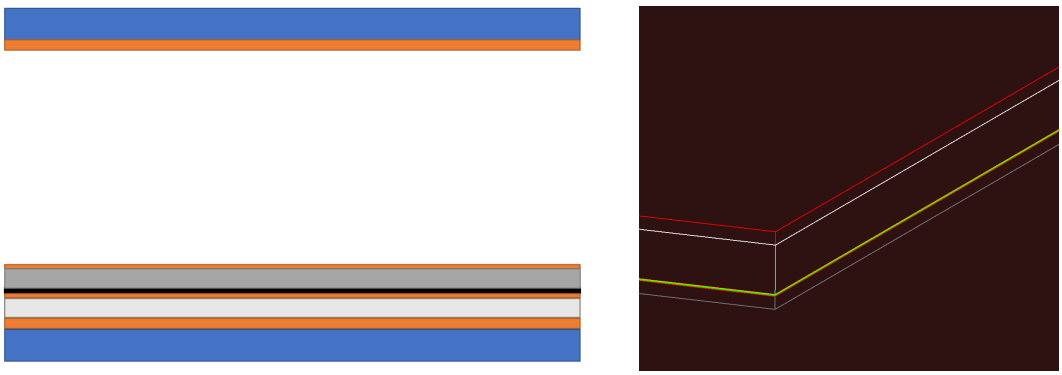


FIGURE 7.27: Schematic view of the various layers accounted for the description of the μ-RWELL detector in GEANT4 (left). From the top to the bottom: the cathode, the drift gap filled of gas (white), and the μ-RWELL structure glued to the readout PCB, composed of top copper and the kapton layer (amplification stage), DLC layer, copper grid to evacuate the charge, pre-preg, copper strips for the readout, and FR_4 layer. An angle of the μ-RWELL chamber implemented in GEANT4 is also shown (right).

7.3.2 The fast simulation of the IDEA detector

The DELPHES fast simulation describes a detector in a parametric way considering each subdetector type with its extension, segmentation, and resolution. The software only needs general volumes for acceptance, a resolution driven segmentation, efficiency formulas, and resolution parametrizations. Thanks to recent studies, the full covariance matrix calculation for track parameter smearing can be used to describe the track resolution in DELPHES. Then the particle trajectory is followed in the detector and the particle is affected by a smearing in its transverse momentum and energy passing different volumes. No local details of detector response are needed. The output is a ROOT file containing reconstructed physics objects, such as leptons, jets, b jets, τ jets, and transverse missing energy.

Currently, a baseline description of the IDEA detector is implemented in DELPHES and included in the official release. This topic represented my first contribution to the IDEA software tools developments. A dedicated card to run the software and reconstruct generated events with the IDEA detector has been provided as a result of the work. The main features of the IDEA card in DELPHES and the most outstanding aspects concerning its validation will be presented in the subsequent discussion.

The current description of the detector in DELPHES is based on the output of dedicated GEANT4 simulation of the IDEA tracker and DR calorimeter.

* The **magnetic field** with an intensity of 2 T is included, affecting the region of the detector within a length of 5 m and up to a radius of 2.25 m.

* The **tracker system** is described through an overall efficiency formula depending on the particle energy and the pseudorapidity region (Table 7.6) and a transverse momentum resolution as a function of η and p_{T}, equal for electrons, muons, and charged hadrons.

$\lvert\eta\rvert \geq 3.0$	0.0%
E $\geq$ 500 MeV ($\lvert\eta\rvert \geq 3.0$)	99.7%
E = [300, 500] MeV ($\lvert\eta\rvert \geq 3.0$)	65.0%
E $\leq$ 300 MeV ($\lvert\eta\rvert \geq 3.0$)	6.0%

Table 7.6: Tracking efficiency formula used in DELPHES for electrons, muons, and charged hadrons.

* The **DR calorimeter** is implemented as a monolithic calorimeter using a single longitudinal segmentation and applying two different energy resolutions (equal for barrel and endcap) to describe its different response to electromagnetic and hadronic showers, as presented in the previous section (see Fig. 7.25).

Given that the DR calorimeter has never been implemented and studied in a fast simulation, the main focus of my work has been directed to its description. In particular, some crucial checks and considerations guided the work. In DELPHES the calorimeter geometry is described as a segmentation of the associated cylinder in cells in the η-ϕ directions, as shown in Fig. 7.28. Each tower reconstructed in the calorimeter corresponds to a single cell in DELPHES workflow. If two particles with energy E_1 and E_2 deposit their energy in the same cell, the reconstructed tower corresponds to a cell. For this reason, the granularity has to be defined accordingly to the physics dimensions of showers, because a clustering approach is not performed in the DELPHES. It is impossible to reproduce the real granularity (less than millimetric) of the DR calorimeter fibers. Then, according to the original

DELPHES approach, a different energy resolution formula affects particles which deposit all their energy in ECAL ($e^{\pm}$, γ, π^0) or in HCAL, partially or totally.

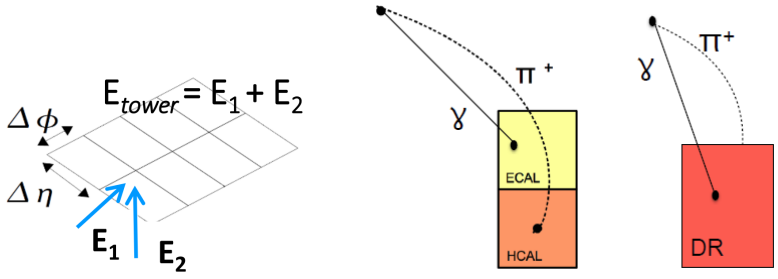


FIGURE 7.28: Traditional energy flow in DELPHES (left) and scheme of the modified energy flow to take into account the presence of the DR calorimeter (right).

However, this distinction does not exist in the DR calorimetry philosophy. Since we have a unique volume to describe the calorimeter, we need to take into account the possibility of having overlaps between particles. For this purpose, the traditional DELPHES energy flow has been modified assuming that, in case of an electromagnetic and hadronic deposit in the same cell, a pessimistic hadronic resolution is applied (Fig. 7.28).
The validation of the energy resolution for reconstructed objects considering both hadronic and electromagnetic particles in different regions of pseudorapidity is performed using particle gun events and is shown in Fig. 7.29.
Then, different cell size configurations and various physics processes involving jets have

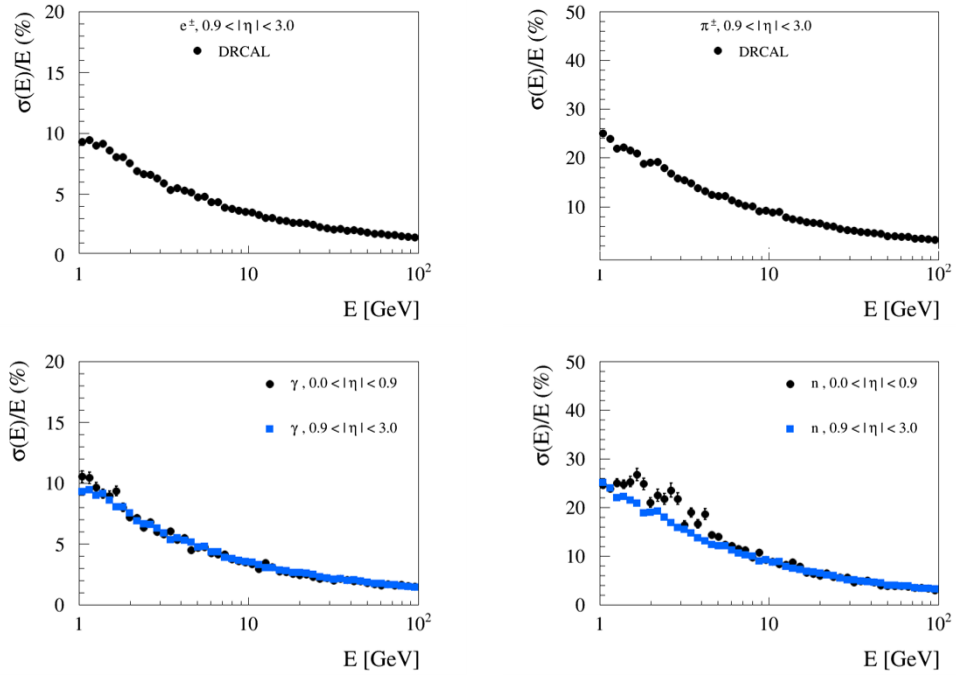


FIGURE 7.29: Energy resolution for electrons (top left), pions (top right), photons (bottom left), and neutrons (bottom right) in DELPHES. The factors $11\%/\sqrt{E}$ and $30\%/\sqrt{E}$ reported in the previous section are clearly visible in electromagnetic and hadronic particles, respectively.

been studied in order to check the effect in DELPHES of changing the DR calorimeter granularity. Jets are reconstructed using the anti-k_t algorithm and requiring a radius of 0.4. The jet energy resolution and the angular resolution have shown not to be affected by the variation of the cell size, both when considering jets reconstructed with only the calorimeter information and when looking at particle flow jets with the additional information of the tracker.

In order to identify the optimal granularity for the DELPHES dual-readout calorimeter, we studied processes involving τ leptons. Tau leptons are particularly suitable to understand the capability of separating very close decay products and its dependence on the granularity. The study of ZH events with the Z and H bosons decaying into a τ pair and light quarks, respectively, shows clearly that the number of calorimetric towers with both an electromagnetic and a hadronic deposit increases when decreasing the granularity (Fig. 7.30).

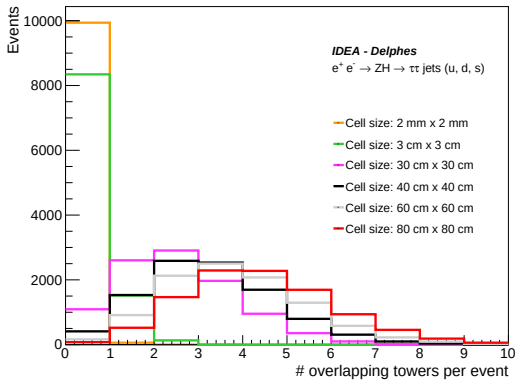


FIGURE 7.30: Number of overlapping particles in the DR calorimeter cells when increasing the cell size.

Then, ZH events have been simulated with the PYTHIA8 generator where the H boson decays in light jets and the Z boson in a pair of τ, forced to decay through a ρ resonance in a π_0 and a charged π associated to a neutrino. This final state is categorized in three different categories according to the number of reconstructed photons in the event (zero, one, or two) together with a charged pion. The percentage of selected events in each category is reported in Table 7.7. The invariant mass of the ρ resonance obtained for indvidual categories and for the combination of the three is shown in Fig. 7.31. The degradation of the photon reconstruction resulting from the increased cell size is evident. As a consequence of the study, a conservative cell size of 6 cm $\times$ 6 cm has been chosen for the current version of the IDEA card. However, investigation will continue in parallel with the development of the full simulation.

To conclude this section, it is worth to briefly mention a new study that was recently performed. A significant improvement of the track description derives from the inclusion of the full covariance matrix of the tracking parameters as an input to the DELPHES parametrization of the tracks. Therefore, the track parameter smearing can be performed according to this matrix. This new feature has been recently changed and validated in DELPHES. In this way, multiple scattering effects can be included in the fast simulation. Moreover, this additional information can be exploited to study the impact of the material, which is not accounted for in the default DELPHES version, and a realistic heavy flavor (HF) tagging simulation.

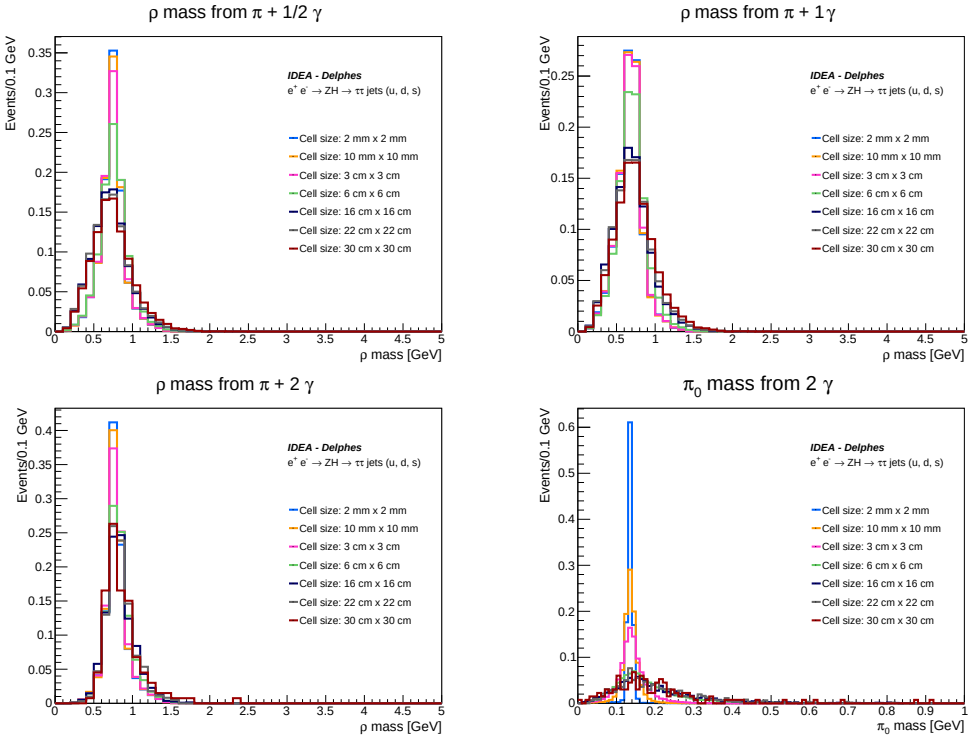


FIGURE 7.31: Invariant mass of the ρ resonance obtained from the combination (top left) of three different categories of selected events: a charged pion plus two, one, or zero reconstructed photons. The invariant mass of the $\pi^{\pm} + 1\gamma$ (top right) and $\pi^{\pm} + 2\gamma$ (bottom left) are also shown, together with the invariant mass of the π^0 decaying into two photons.

Granularity	$\pi^{\pm} + 2\gamma$	$\pi^{\pm} + 1\gamma$	$\pi^{\pm} + 0\gamma$
2 mm×2 mm	51.0%	35.0%	14.0%
10 mm×10 mm	51.0%	35.0%	14.0%
3 cm×3 cm	51.0%	35.0%	14.0%
6 cm×6 cm	43.0%	43.0%	14.0%
16 cm×16 cm	6.0%	47.0%	14.0%
22 cm×22 cm	2.0%	32.0%	14.0%
30 cm×30 cm	0.4%	19.3%	80.3%

TABLE 7.7: Percentage of selected events in each of the considered categories: a charged pion plus two, one, or zero reconstructed photons.

7.4 Conclusion

An extensive effort oriented to the development of innovative detector technologies is ongoing in view of the construction of future colliders post the LHC era. A specific detector proposal for a large e^+e^- circular collider, IDEA, has been presented in this thesis. Recently, the interest for this detector proposal has significantly increased given the outcome of the European Strategy in June 2020, which defined an e^+e^- collider as a priority for the future of high energy physics. This chapter presented an overview of different activities related to the development of the IDEA detector concept. I have been personally involved in many aspects of this work, both from the hardware and the software point of view.

Summary

After the announcement of the discovery of the H boson in 2012, the precise measurement of the various H boson interactions with other SM particles has become one of the main goals in the physics program of the ATLAS and CMS experiments at the LHC. In this context, the current period is very exciting because many analyses which exploit the total amount of data collected during the successful LHC Run 2 are going to be finilized. They will provide the most precise measurements of the last discovered elementary particle. Results of the $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ analysis presented in this thesis are part of this story, covering an extensive set of measurements of the H boson properties. The signal strength modifiers for the main H boson production mechanisms and the integrated fiducial cross section have been measured. Moreover, thanks to the large amount of data available, highly granular STXS measurements and precise fiducial differential measurements have been performed targeting specific regions of the phase space. In parallel, limits on CP violation and anomalous couplings of the H boson to vector bosons and fermions have been produced.
The SM predictions are confirmed by all measurements within their uncertainties, that are continuously decreasing thanks to the great efforts from the theoretical and experimental side even if we are coping with more challenging data-taking conditions. The full Run 2 dataset is definitely introducing us in the H boson precision realm, offering the chance to access small deviations from the SM predictions that could provide strong hints about possible BSM effects. Inclusive H boson measurements start to be limited by systematic uncertainties in several channels. For this reason, differential measurements of the H boson properties are becoming much more relevant, both defining fiducial phase spaces and studying STXS.

In parallel, a large effort to improve analysis strategies in view of Run 3 and to prepare the upgrades of the experiments in view of the HL-LHC ten-year phase of operation is ongoing. In the ambitious landscape of the HL-LHC, only a few percent of the expected data has been collected up to now. The great expectation for HL-LHC is due to the fact that a huge amount of H bosons per year will be produced, extending the sensitivity for new physics searches. Furthermore, it will give access to the study of the H H pair production and the H boson self-coupling, which represents one of the main goals of the HL-LHC program.

As it has been discussed, the investigation of the Higgs sector is in the front line of the high energy physics research. It is also guiding the very animated discussion about the future of the high energy physics and the kind of future accelerators needed, that will shape the activities of the particle physics research for decades. An extensive effort focused on the development of innovative detector technologies is ongoing in view of the construction of future colliders post the LHC era. Recently, the outcome of the European Strategy in June 2020 oriented the primary interest of the HEP community to the realization of a lepton electroweak and H factory collider.

To conclude, during my activity as a Ph.D. student I was lucky enough to be a part of the H boson precision measurements, starting from the $\mathrm{H} \to 4\ell$ analysis with the CMS experiment at the LHC and also performing studies for the future of the Higgs sector at large lepton colliders.

Appendix A

Objects in the H $\rightarrow 4\ell$ analysis: additional plots

Muon scale factors

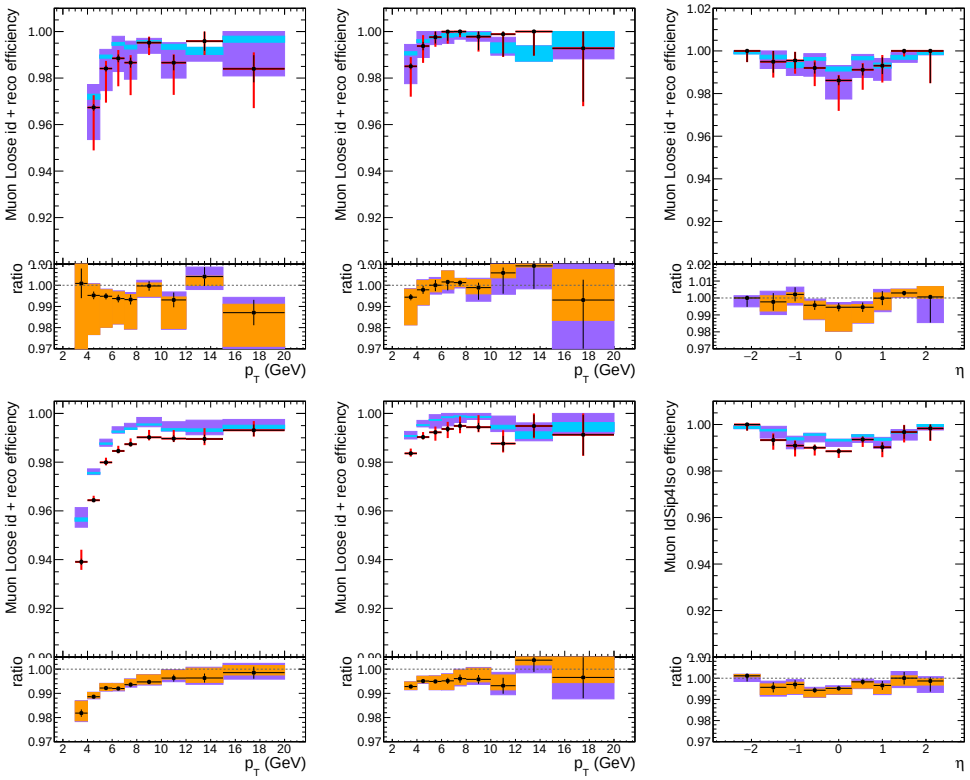


FIGURE A.1: Muon reconstruction and identification efficiency at low p_{T}, measured with the TnP method on J/Ψ events, as function of p_{T} in the barrel (left) and endcaps (center), and as function of η for $p_{\mathrm{T}} > 5$ GeV (right), for the 2016 (top) and 2017 (bottom) data set. In the upper panel, the black error bars on data are purely statistical while the red ones include also the systematical uncertainties. Similarly, the smaller azure rectangles represent the statistical uncertainty on MC while the larger orange rectangles include also the systematical contribution. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.

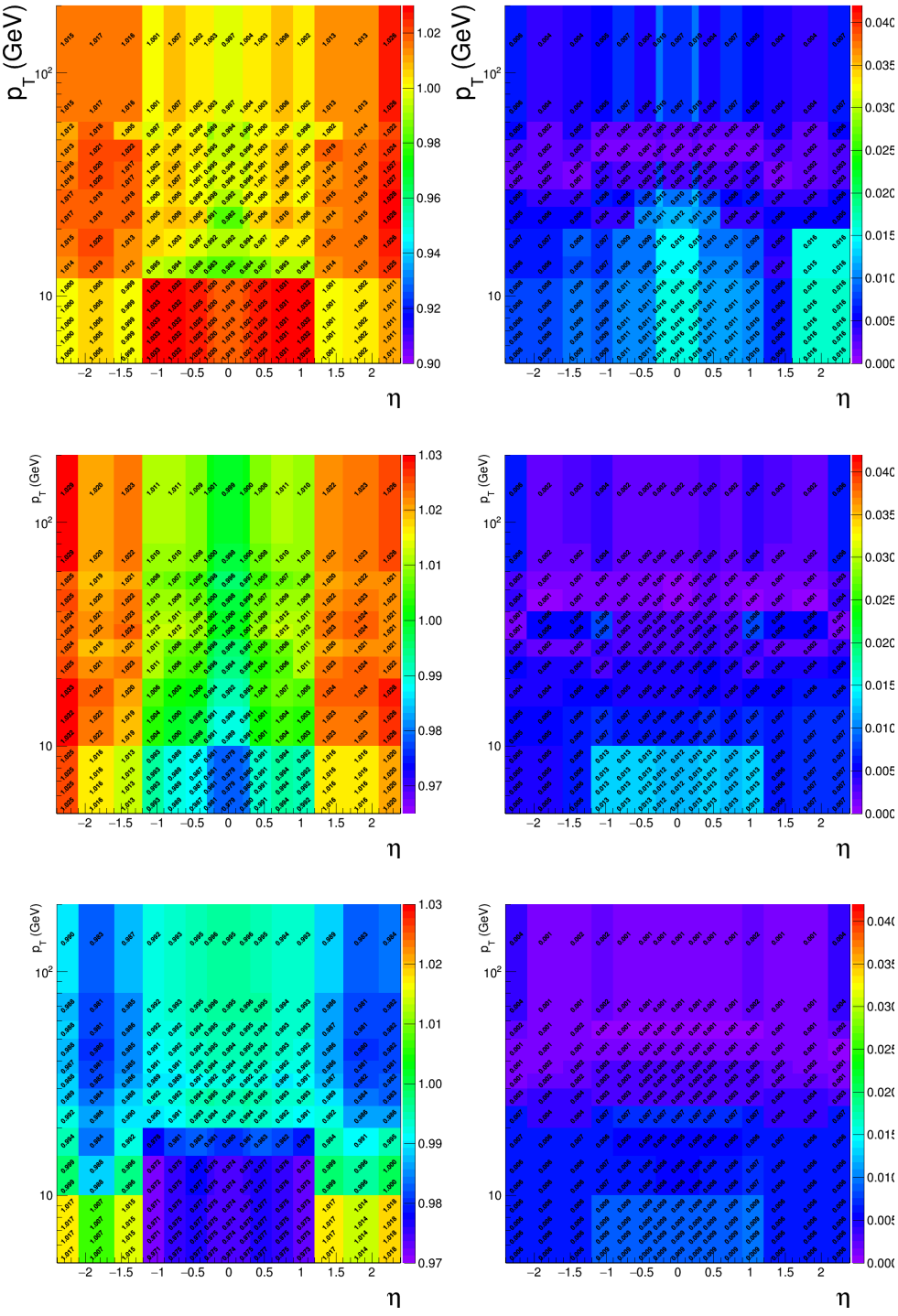


FIGURE A.2: Overall data to simulation scale factors for muons, as a function of p_{T} and η (left). Uncertainties on data to simulation scale factors for muons, as a function of p_{T} and η (right). Results are shown for 2016 (top), 2017 (middle), and 2018 (bottom).

Electron scale factors (2016 and 2018)

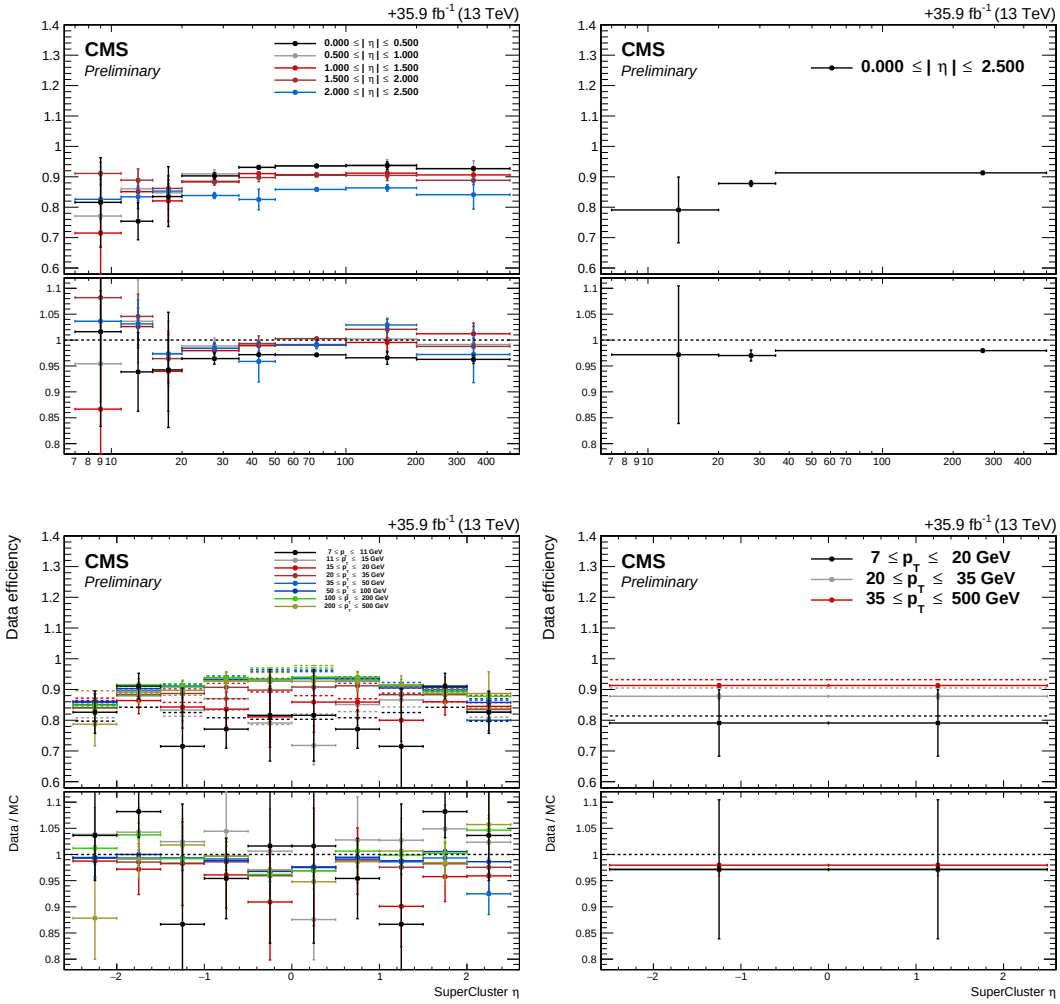


FIGURE A.3: Electron selection efficiencies as a function of p_{T} (top) and η (bottom) measured in 2016 data using the TnP technique for non-gap electrons (left) and gap electrons (right), together with the corresponding data/MC ratio at the bottom of each plot.

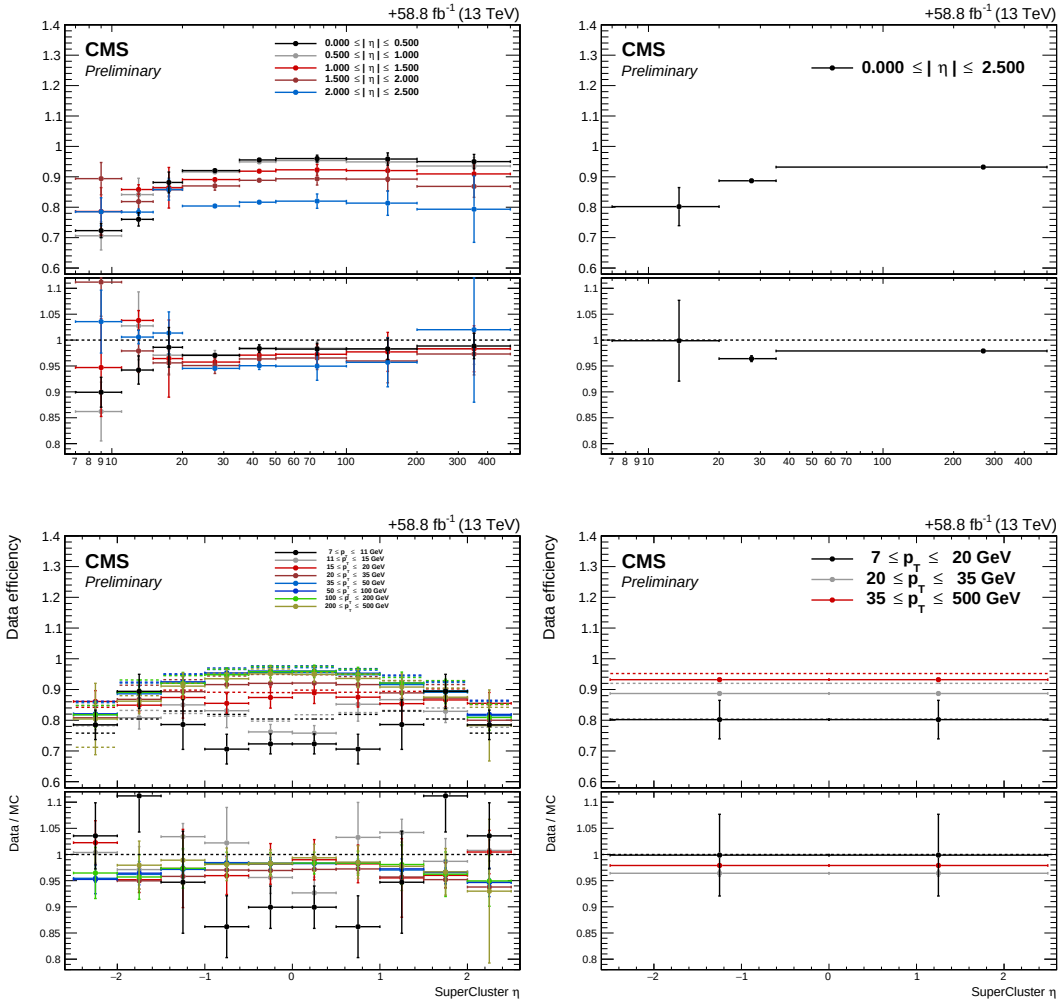


FIGURE A.4: Electron selection efficiencies as a function of p_{T} (top) and η (bottom) measured in 2018 data using the TnP technique for non-gap electrons (left) and gap electrons (right), together with the corresponding data/MC ratio at the bottom of each plot.

Appendix B

Datasets and HLT paths used in the $H \rightarrow 4\ell$ analysis

2018 datasets and HLT paths

Run-range	Dataset	Integrated luminosity
315252-316995	/DoubleMuon/Run2018A-17Sep2018-v2/MINIAOD /MuonEG/Run2018A-17Sep2018-v1/MINIAOD /EGamma/Run2018A-17Sep2018-v2/MINIAOD /SingleMuon/Run2018A-17Sep2018-v2/MINIAOD	$14.02\ \text{fb}^{-1}$
317080-319310	/DoubleMuon/Run2018B-17Sep2018-v1/MINIAOD /MuonEG/Run2018B-17Sep2018-v1/MINIAOD /EGamma/Run2018B-17Sep2018-v1/MINIAOD /SingleMuon/Run2018B-17Sep2018-v1/MINIAOD	$7.06\ \text{fb}^{-1}$
319337-320065	/DoubleMuon/Run2018C-17Sep2018-v1/MINIAOD /MuonEG/Run2018C-17Sep2018-v1/MINIAOD /EGamma/Run2018C-17Sep2018-v1/MINIAOD /SingleMuon/Run2018C-17Sep2018-v1/MINIAOD	$6.90\ \text{fb}^{-1}$
320673-325175	/DoubleMuon/Run2018D-PromptReco-v2/MINIAOD /MuonEG/Run2018D-PromptReco-v2/MINIAOD /EGamma/Run2018D-PromptReco-v2/MINIAOD /SingleMuon/Run2018D-PromptReco-v2/MINIAOD	$31.75\ \text{fb}^{-1}$

TABLE B.1: Datasets used in the 2018 analysis corresponding to an integrated luminosity of $59.7\ \text{fb}^{-1}$.

HLT path	prescale	primary dataset
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*	1	DoubleEG
HLT_DoubleEle25_CaloIdL_MW_v*	1	DoubleEG
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*	1	DoubleMuon
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*	1	MuonEG
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*	1	MuonEG
HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*	1	MuonEG
HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ_v*	1	MuonEG
HLT_Ele32_WPTight_Gsf_v*	1	SingleElectron
HLT_IsoMu24_v*	1	SingleMuon

TABLE B.2: Trigger paths used in 2018 collision data.

2017 datasets and HLT paths

Run-range	Dataset	Integrated luminosity
297046-299329	/Run2017B-31Mar2018-v1/MINIAOD /DoubleEG/Run2017B-31Mar2018-v1/MINIAOD /MuonEG/Run2017B-31Mar2018-v1/MINIAOD /SingleElectron/Run2017B-31Mar2018-v1/MINIAOD /SingleMuon/Run2017B-31Mar2018-v1/MINIAOD	$4.79\ \text{fb}^{-1}$
299368-300676	/DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD /DoubleEG/Run2017C-31Mar2018-v1/MINIAOD /MuonEG/Run2017C-31Mar2018-v1/MINIAOD /SingleElectron/Run2017C-31Mar2018-v1/MINIAOD /SingleMuon/Run2017C-31Mar2018-v1/MINIAOD	$9.64\ \text{fb}^{-1}$
302030-303434	/DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD /DoubleEG/Run2017D-31Mar2018-v1/MINIAOD /MuonEG/Run2017D-31Mar2018-v1/MINIAOD /SingleElectron/Run2017D-31Mar2018-v1/MINIAOD /SingleMuon/Run2017D-31Mar2018-v1/MINIAOD	$4.24\ \text{fb}^{-1}$
303824-304797	/DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD /DoubleEG/Run2017E-31Mar2018-v1/MINIAOD /MuonEG/Run2017E-31Mar2018-v1/MINIAOD /SingleElectron/Run2017E-31Mar2018-v1/MINIAOD /SingleMuon/Run2017E-31Mar2018-v1/MINIAOD	$9.31\ \text{fb}^{-1}$
305040-306462	/DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD /DoubleEG/Run2017F-31Mar2018-v1/MINIAOD /MuonEG/Run2017F-31Mar2018-v1/MINIAOD /SingleElectron/Run2017F-31Mar2018-v1/MINIAOD /SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	$13.55\ \text{fb}^{-1}$

Table B.3: Datasets used in the 2017 analysis corresponding to an integrated luminosity of $41.5\ \text{fb}^{-1}$.

HLT path	prescale	primary dataset
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_*	1	DoubleEG
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL	1	DoubleEG
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL	1	DoubleEG
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8	1	DoubleMuon
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8	1	DoubleMuon
HLT_TripleMu_12_10_5	1	DoubleMuon
HLT_TripleMu_10_5_5_D2	1	DoubleMuon
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	1	MuonEG
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	1	MuonEG
HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	1	MuonEG
HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ	1	MuonEG
HLT_Mu8_DiEle12_CaloIdL_TrackIdL	1	MuonEG
HLT_Mu8_DiEle12_CaloIdL_TrackIdL_DZ	1	MuonEG
HLT_Ele35_WPTight_Gsf_v*	1	SingleElectron
HLT_Ele38_WPTight_Gsf_v*	1	SingleElectron
HLT_Ele40_WPTight_Gsf_v*	1	SingleElectron
HLT_IsoMu27	1	SingleMuon

Table B.4: Trigger paths used in 2017 collision data.

2016 datasets and HLT paths

Run-range	Dataset	Integrated luminosity
272007-275376	/DoubleMuon/Run2016B-17Jul2018-v1/MINIAOD /DoubleEG/Run2016B-17Jul2018-v1/MINIAOD /MuonEG/Run2016B-17Jul2018-v1/MINIAOD /SingleElectron/Run2016B-17Jul2018-v1/MINIAOD /SingleMuon/Run2016B-17Jul2018-v1/MINIAOD	$5.75\ \mathrm{fb}^{-1}$
275657-276283	/DoubleMuon/Run2016C-17Jul2018-v1/MINIAOD /DoubleEG/Run2016C-17Jul2018-v1/MINIAOD /MuonEG/Run2016C-17Jul2018-v1/MINIAOD /SingleElectron/Run2016C-17Jul2018-v1/MINIAOD /SingleMuon/Run2016C-17Jul2018-v1/MINIAOD	$2.57\ \mathrm{fb}^{-1}$
276315-276811	/DoubleMuon/Run2016D-17Jul2018-v1/MINIAOD /DoubleEG/Run2016D-17Jul2018-v1/MINIAOD /MuonEG/Run2016D-17Jul2018-v1/MINIAOD /SingleElectron/Run2016D-17Jul2018-v1/MINIAOD /SingleMuon/Run2016D-17Jul2018-v1/MINIAOD	$4.25\ \mathrm{fb}^{-1}$
276831-277420	/DoubleMuon/Run2016E-17Jul2018-v1/MINIAOD /DoubleEG/Run2016E-17Jul2018-v1/MINIAOD /MuonEG/Run2016E-17Jul2018-v1/MINIAOD /SingleElectron/Run2016E-17Jul2018-v1/MINIAOD /SingleMuon/Run2016E-17Jul2018-v1/MINIAOD	$4.02\ \mathrm{fb}^{-1}$
277772-278808	/DoubleMuon/Run2016F-17Jul2018-v1/MINIAOD /DoubleEG/Run2016F-17Jul2018-v1/MINIAOD /MuonEG/Run2016F-17Jul2018-v1/MINIAOD /SingleElectron/Run2016F-17Jul2018-v1/MINIAOD /SingleMuon/Run2016F-17Jul2018-v1/MINIAOD	$3.10\ \mathrm{fb}^{-1}$
278820-280385	/DoubleMuon/Run2016G-17Jul2018-v1/MINIAOD /DoubleEG/Run2016G-17Jul2018-v1/MINIAOD /MuonEG/Run2016G-17Jul2018-v1/MINIAOD /SingleElectron/Run2016G-17Jul2018-v1/MINIAOD /SingleMuon/Run2016G-17Jul2018-v1/MINIAOD	$7.48\ \mathrm{fb}^{-1}$
280919-284044	/DoubleMuon/Run2016H-17Jul2018-v1/MINIAOD /DoubleEG/Run2016H-17Jul2018-v1/MINIAOD /MuonEG/Run2016H-17Jul2018-v1/MINIAOD /SingleElectron/Run2016H-17Jul2018-v1/MINIAOD /SingleMuon/Run2016H-17Jul2018-v1/MINIAOD /DoubleMuon/Run2016H-17Jul2018-v1/MINIAOD /DoubleEG/Run2016H-17Jul2018-v1/MINIAOD /MuonEG/Run2016H-17Jul2018-v1/MINIAOD /SingleElectron/Run2016H-17Jul2018-v1/MINIAOD /SingleMuon/Run2016H-17Jul2018-v1/MINIAOD /DoubleMuon/Run2016H-17Jul2018-v1/MINIAOD /DoubleEG/Run2016H-17Jul2018-v1/MINIAOD /MuonEG/Run2016H-17Jul2018-v1/MINIAOD /SingleElectron/Run2016H-17Jul2018-v1/MINIAOD /SingleMuon/Run2016H-17Jul2018-v1/MINIAOD	$8.75\ \mathrm{fb}^{-1}$

Table B.5: Datasets used in the 2016 analysis corresponding to an integrated luminosity of $35.9\ \mathrm{fb}^{-1}$.

HLT path	prescale	primary dataset
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	1	DoubleEG
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	1	DoubleEG
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL	1	DoubleEG
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL	1	DoubleEG
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	1	DoubleMuon
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	1	DoubleMuon
HLT_TripleMu_12_10_5	1	DoubleMuon
HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL	1	MuonEG
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	1	MuonEG
HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL	1	MuonEG
HLT_Mu8_DiEle12_CaloIdL_TrackIdL	1	MuonEG
HLT_DiMu9_Ele9_CaloIdL_TrackIdL	1	MuonEG
HLT_Ele25_eta2p1_WPTight	1	SingleElectron
HLT_Ele27_WPTight	1	SingleElectron
HLT_Ele27_eta2p1_WPLoose_Gsf	1	SingleElectron
HLT_IsoMu20 OR HLT_IsoTkMu20	1	SingleMuon
HLT_IsoMu22 OR HLT_IsoTkMu22	1	SingleMuon

Table B.6: Trigger paths used in 2016 collision data.

Trigger efficiencies per final state

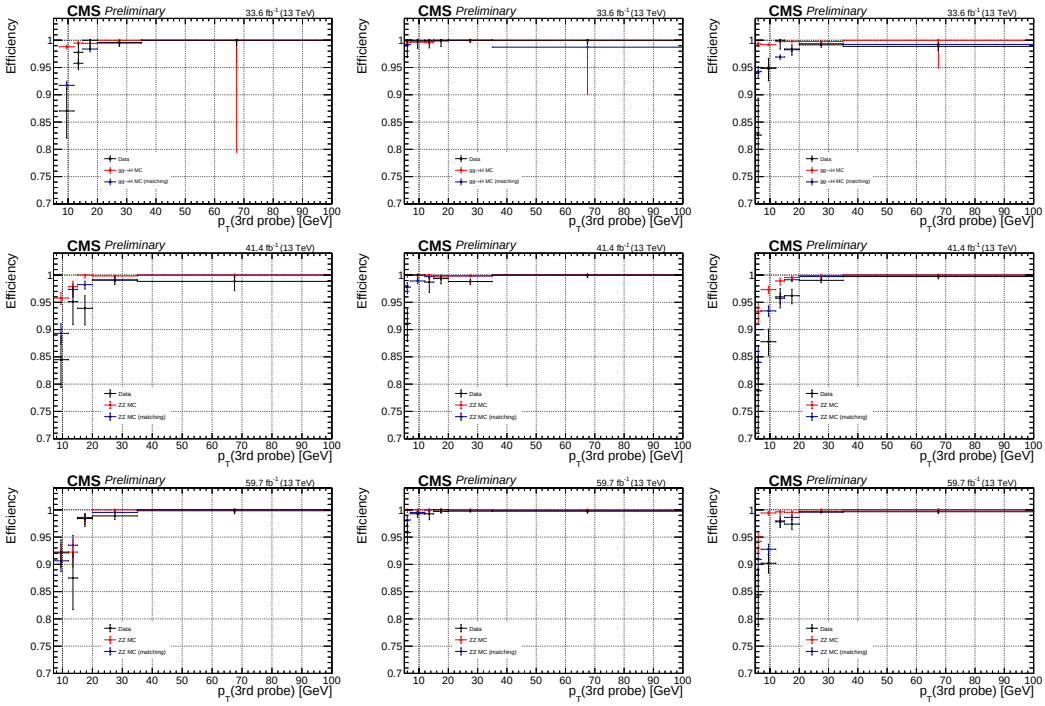
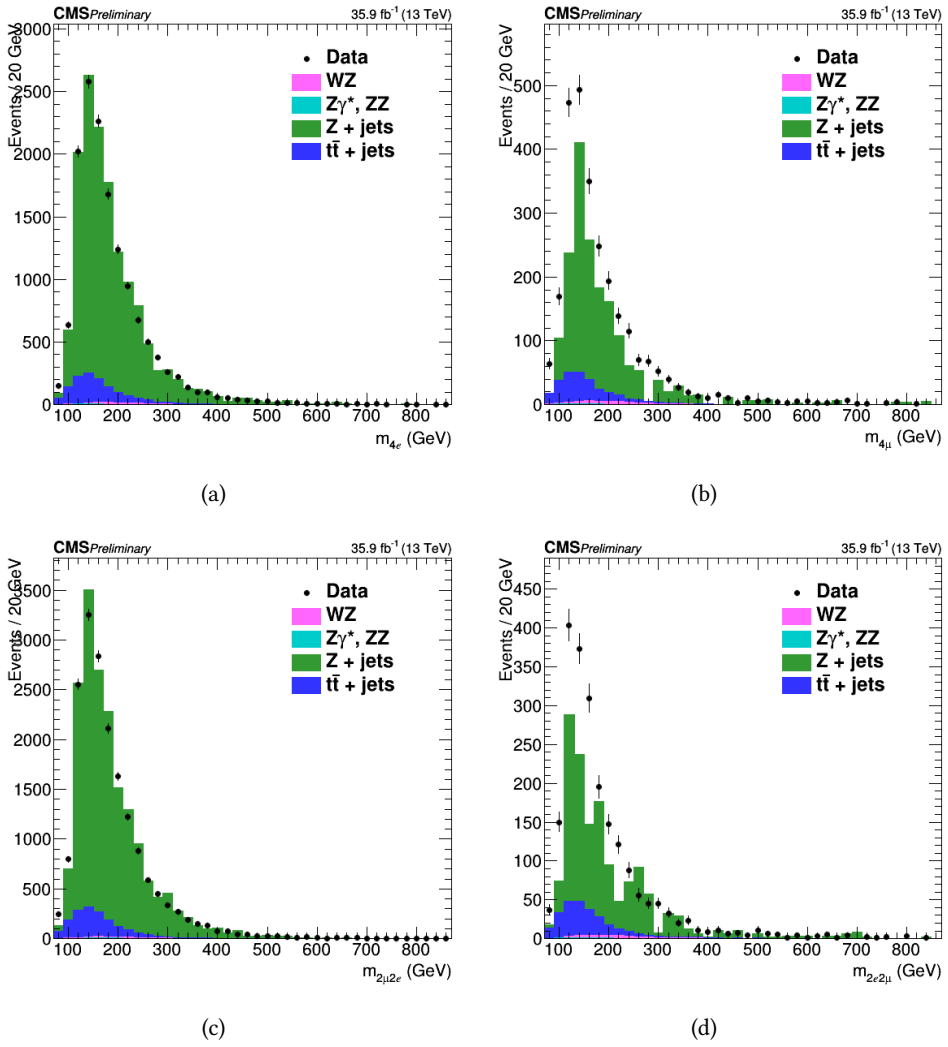


FIGURE B.1: Trigger efficiency measured in 2016 (top line), 2017 (middle line), and 2018 (bottom line) data using 4ℓ events collected by single lepton triggers for the $4e$ (left), 4μ (middle), and $2e2\mu$ (right) final states.

Appendix C

Estimate of the Z+X background in the H $\rightarrow 4\ell$ analysis

Additional distributions for the 2016 dataset



(a) (b)

(c) (d)

FIGURE C.1: Invariant mass distribution of the events selected in the 2P+2F SS control samples for all the considered final states in 2017 data: $4e$ (a), 4μ (b), $2\mu 2e$ (c), and $2e2\mu$ (d).

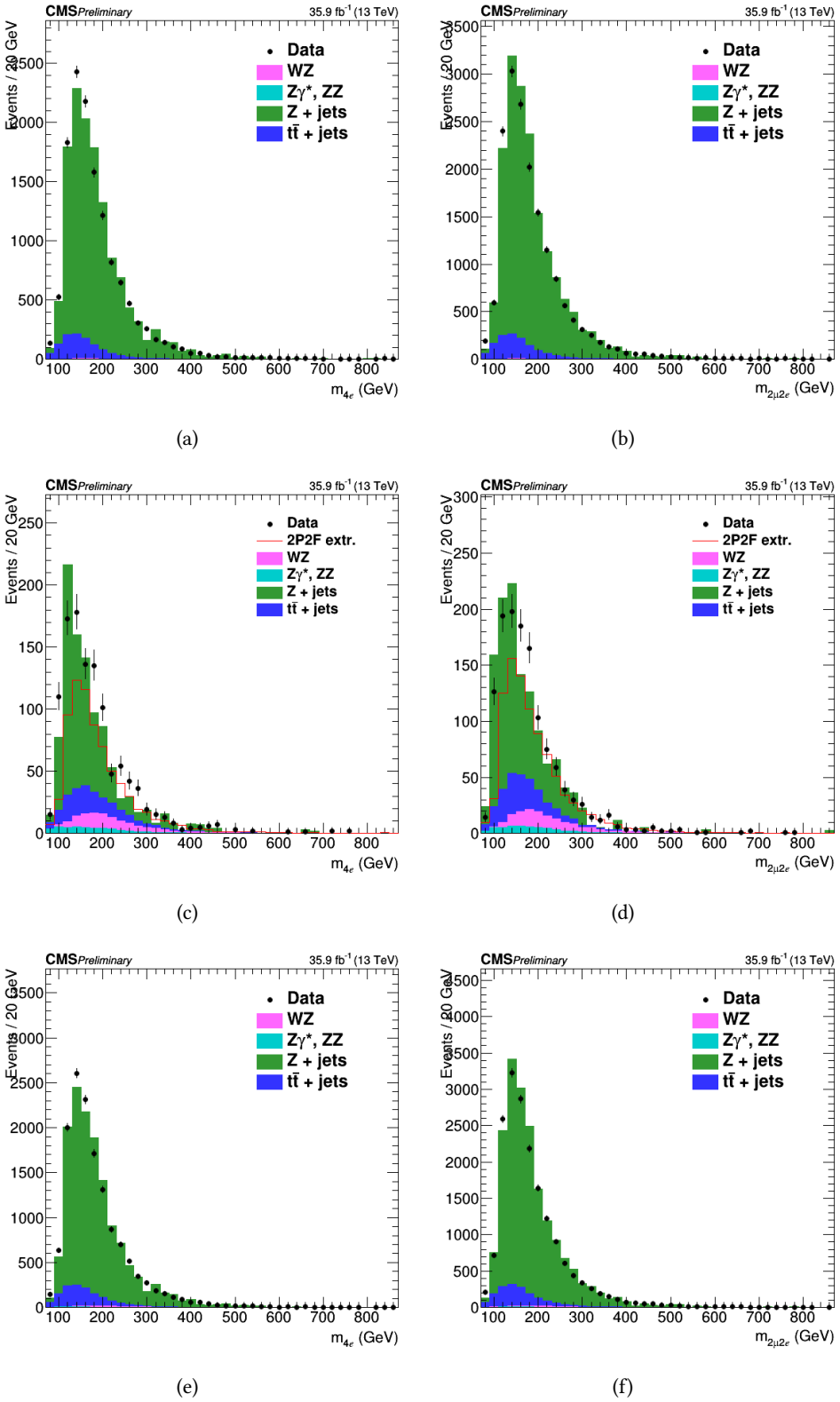


FIGURE C.2: Invariant mass distribution of the events selected in the 2P+2F (top) and 3P+1F (middle) control sample and the combination of the two CR using the OS method in the 2016 dataset for $4e$ (left) and $2\mu 2e$ (right) final state.

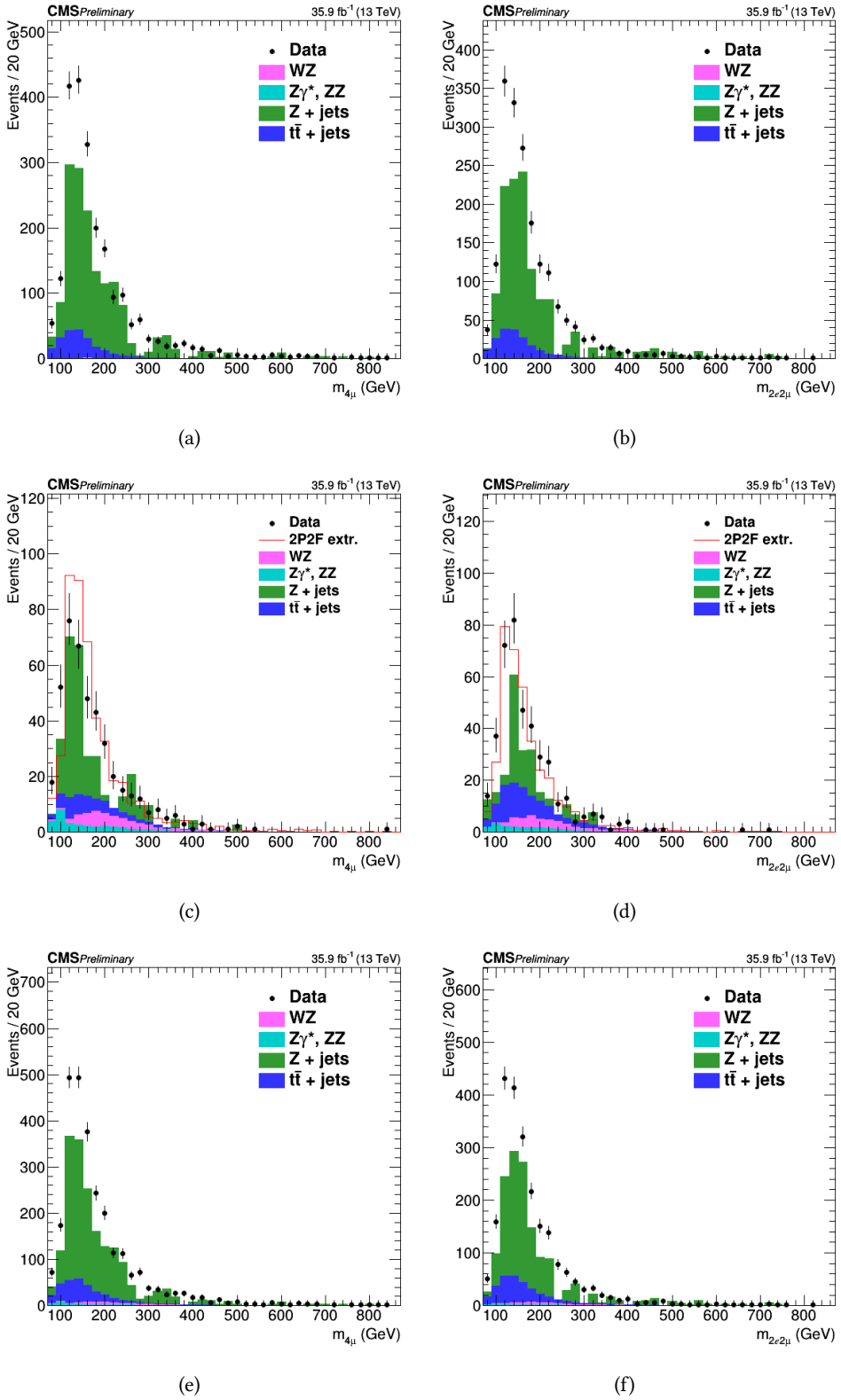


FIGURE C.3: Invariant mass distribution of the events selected in the 2P+2F (top) and 3P+1F (middle) control sample and the combination of the two CR using the OS method in the 2016 dataset for 4μ (left) and $2e2\mu$ (right) final state.

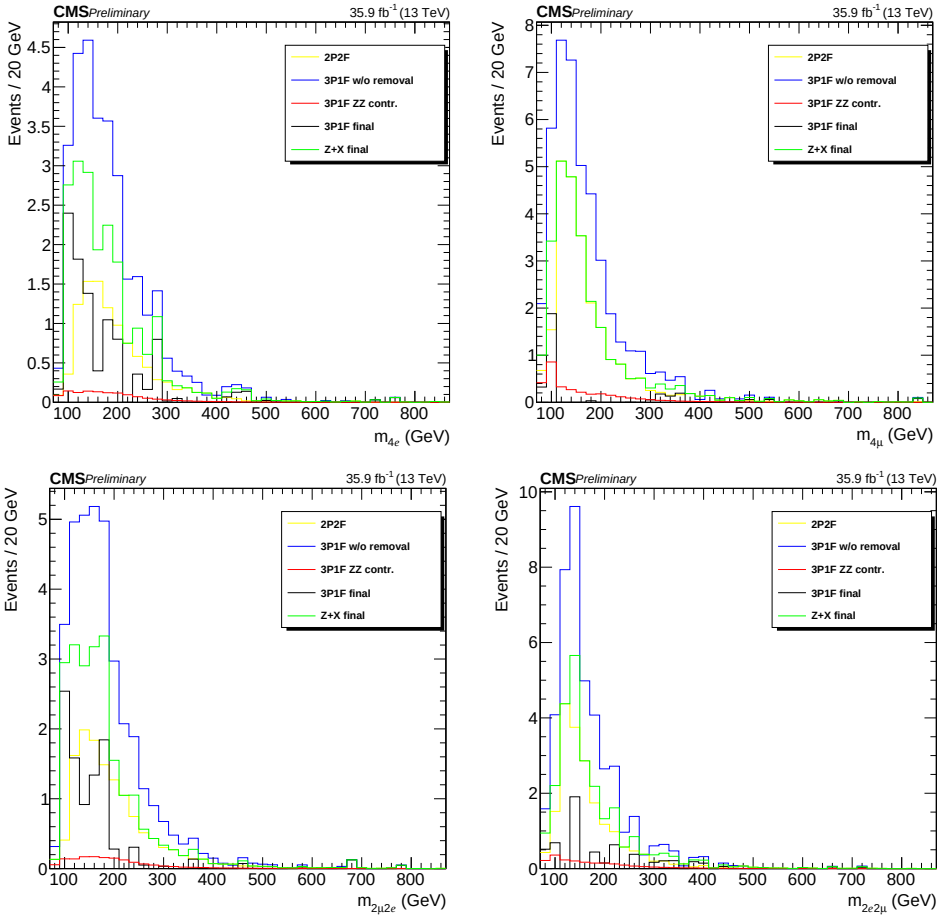
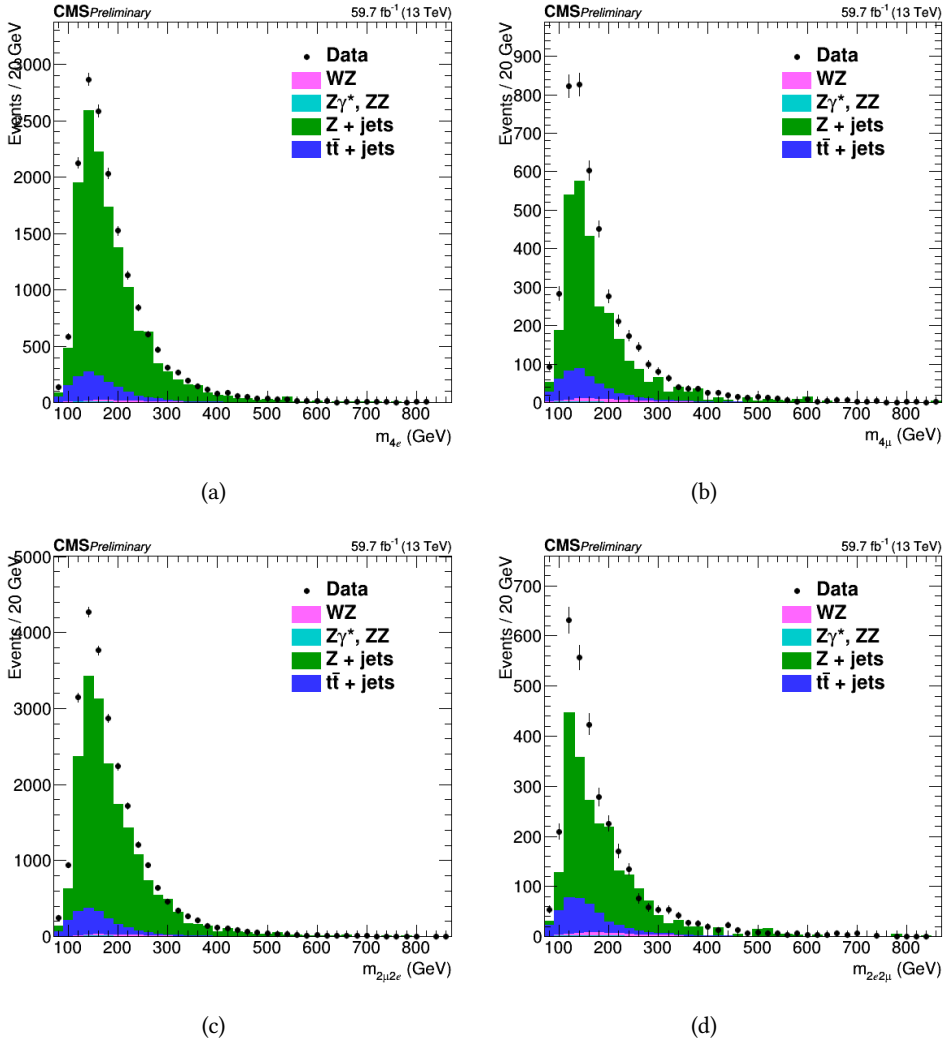


FIGURE C.4: Four-lepton invariant mass of the different processes which contribute to the 3P+1F estimate of the reducible background in the 2016 dataset for all the considered channels: $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right).

Additional distributions for the 2018 dataset



(a) (b)

(c) (d)

FIGURE C.5: Invariant mass distribution of the events selected in the 2P+2F SS control samples for all the considered final states in 2018 data: $4e$ (a), 4μ (b), $2\mu 2e$ (c), and $2e2\mu$ (d).

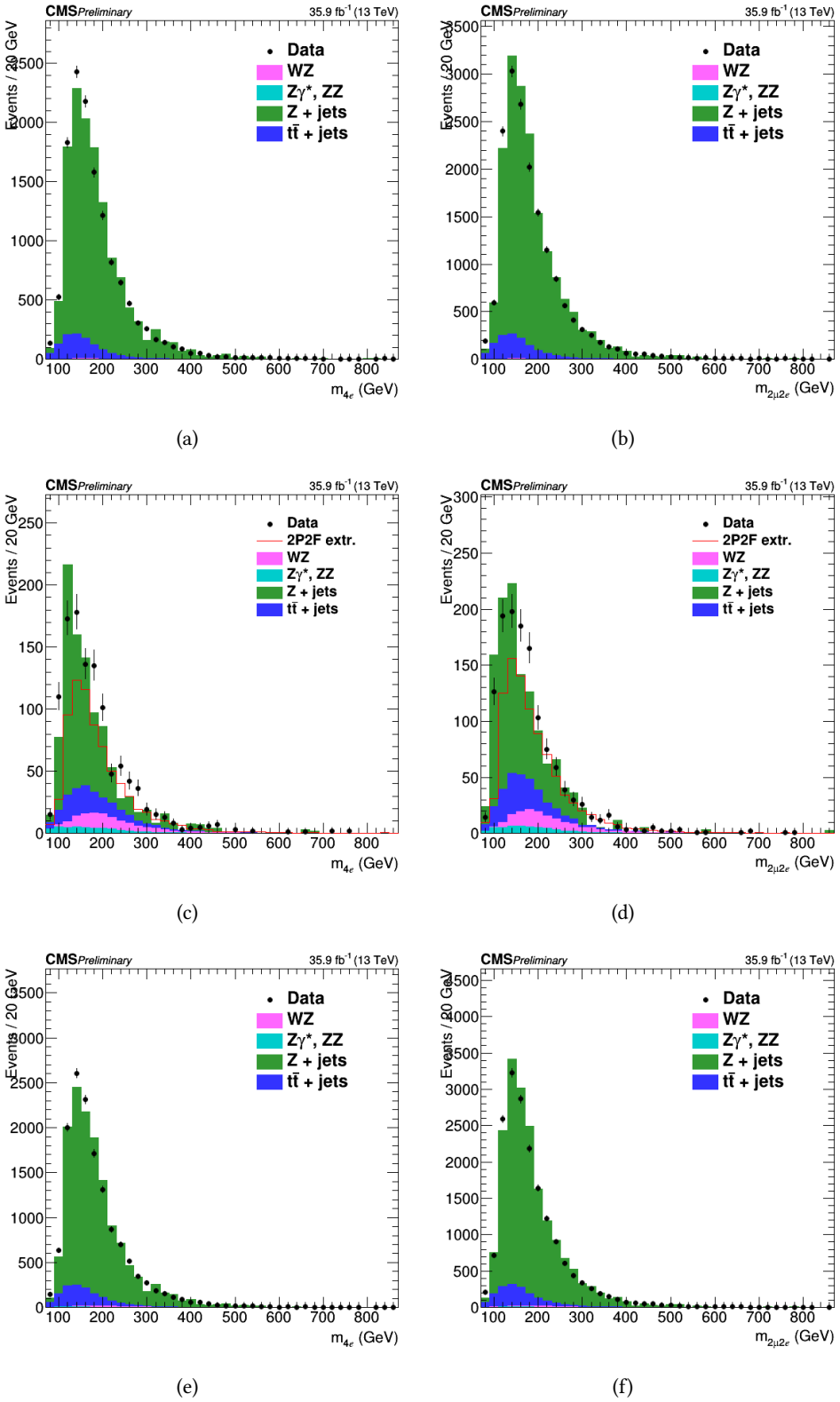


FIGURE C.6: Invariant mass distribution of the events selected in the 2P+2F (top) and 3P+1F (middle) control sample and the combination of the two CR using the OS method in the 2016 dataset for $4e$ (left) and $2\mu 2e$ (right) final state.

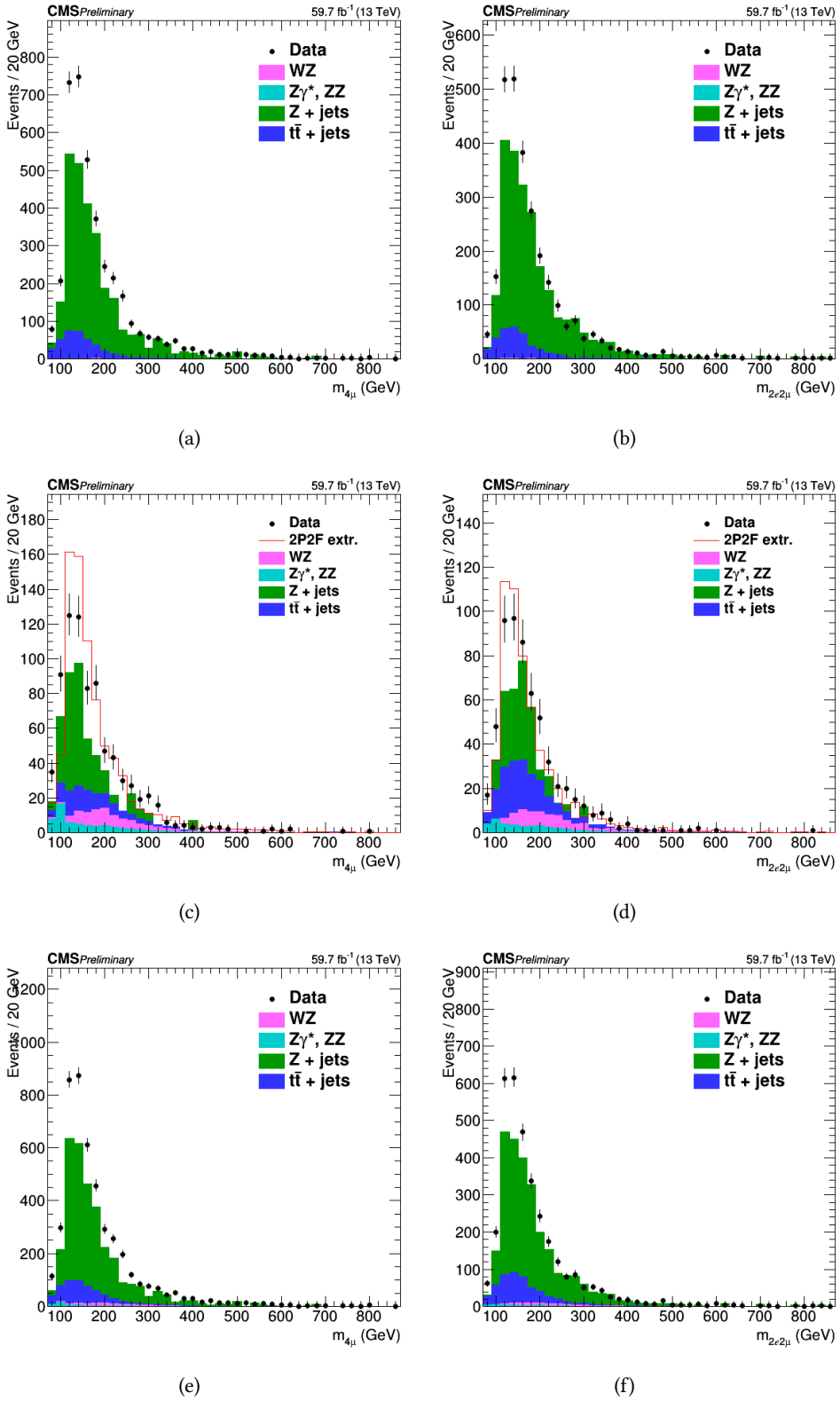


FIGURE C.7: Invariant mass distribution of the events selected in the 2P+2F (top) and 3P+1F (middle) control sample and the combination of the two CR using the OS method in the 2018 dataset for 4μ (left) and $2e2\mu$ (right) final state.

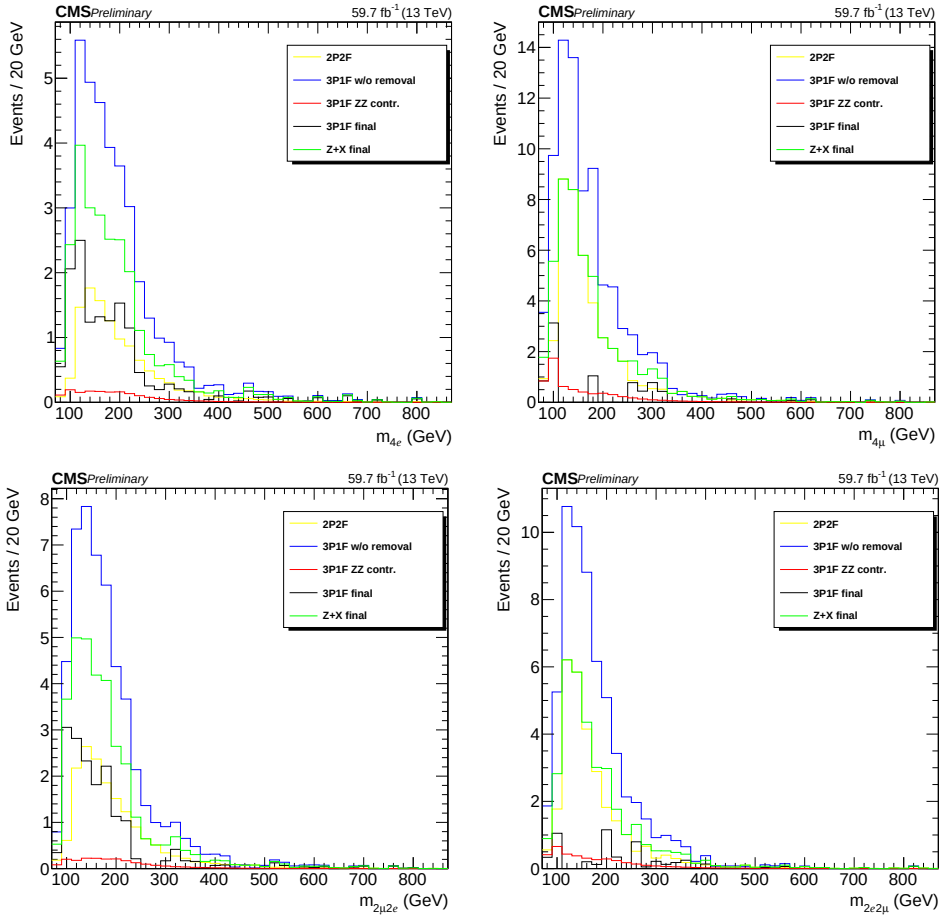


FIGURE C.8: Four-lepton invariant mass of the different processes which contribute to the 3P+1F estimate of the reducible background in the 2018 dataset for all the considered channels: $4e$ (top left), 4μ (top right), $2\mu 2e$ (bottom left), and $2e2\mu$ (bottom right).

Shapes in different STXS stage 1.2 bin (2017-2018)

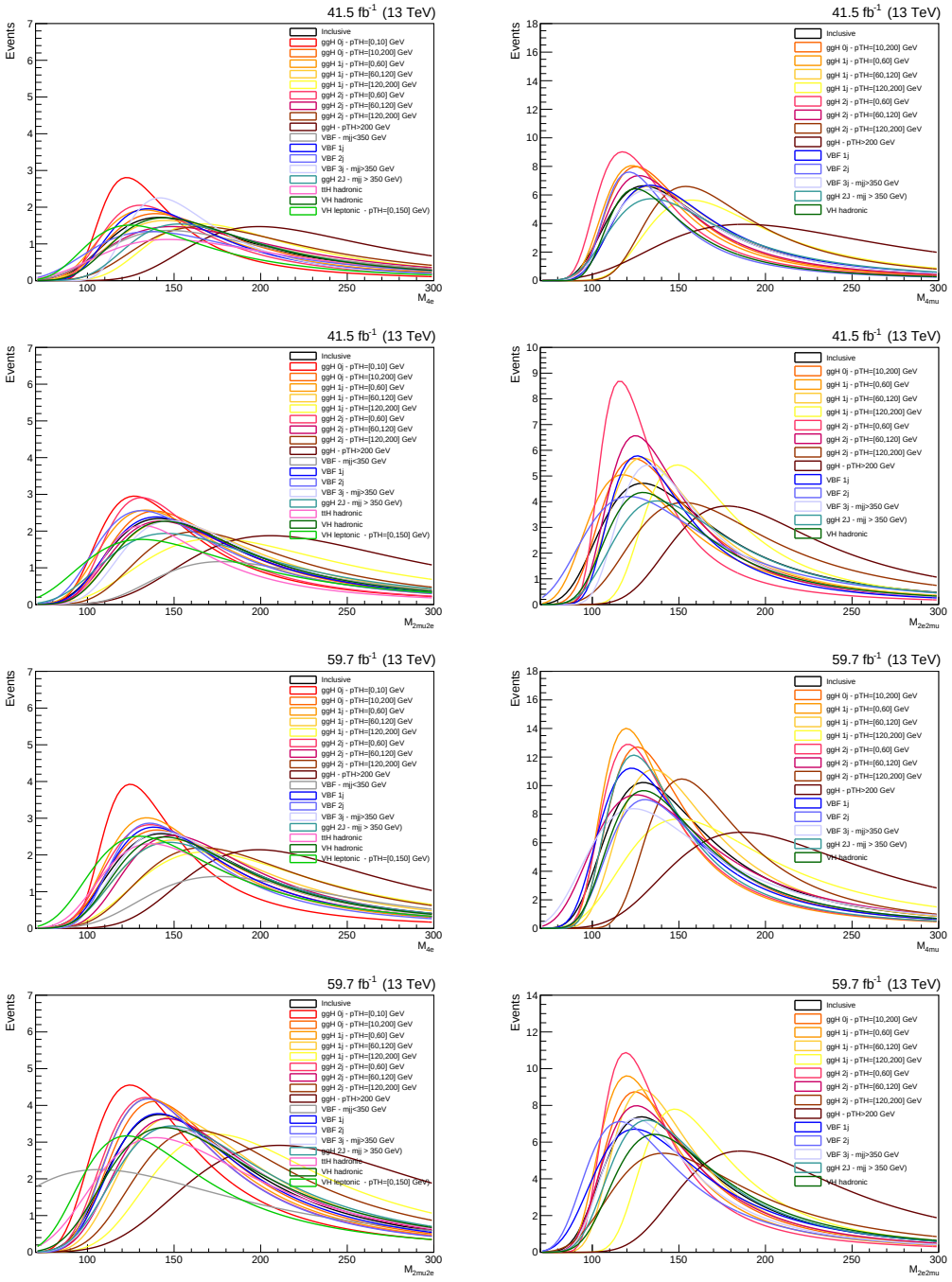


FIGURE C.9: Shapes of the $m_{4\ell}$ distribution for the reducible background in the 2017 (rows 1 and 2) and 2018 (rows 3 and 4) in all the considered categories using the SS method. Shapes are normalized to the same number of events.

Appendix D

Results of the H $\rightarrow 4\ell$ analysis: additional plots

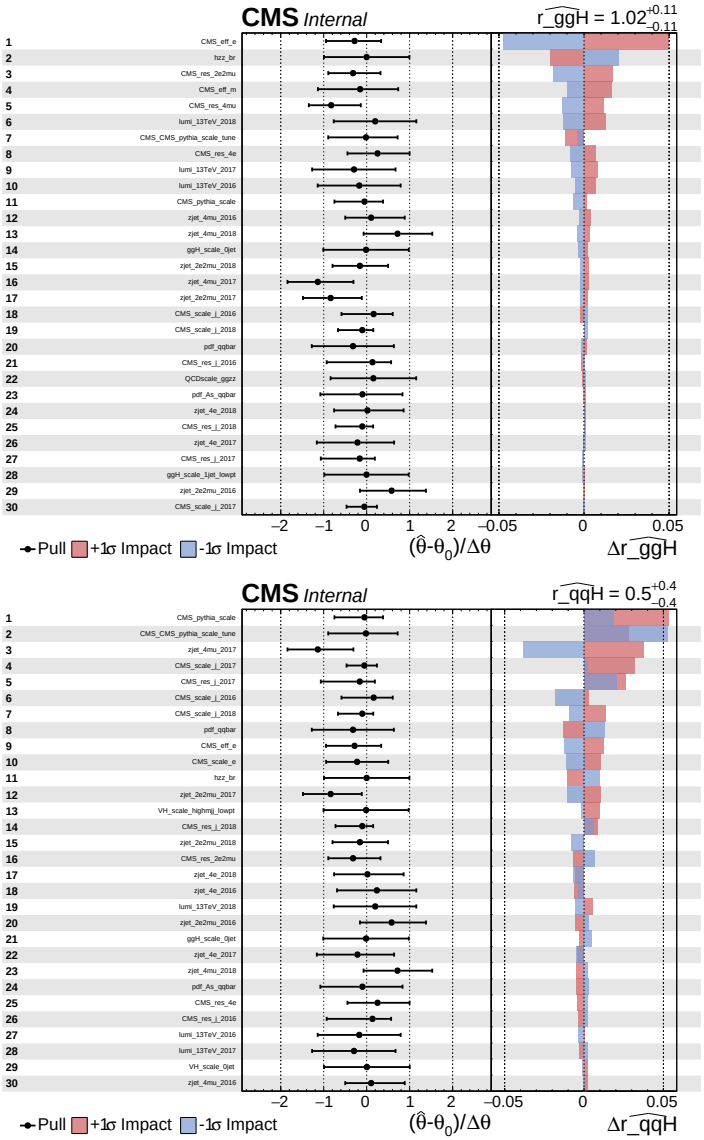


FIGURE D.1: Systematic uncertainties with the largest impact on the ggH (top) and qqH (bottom) cross section measurement with $m_{\mathrm{H}} = 125.38$ GeV. Value and error of the corresponding nuisance parameter after the maximum likelihood fit to the observed data are shown.

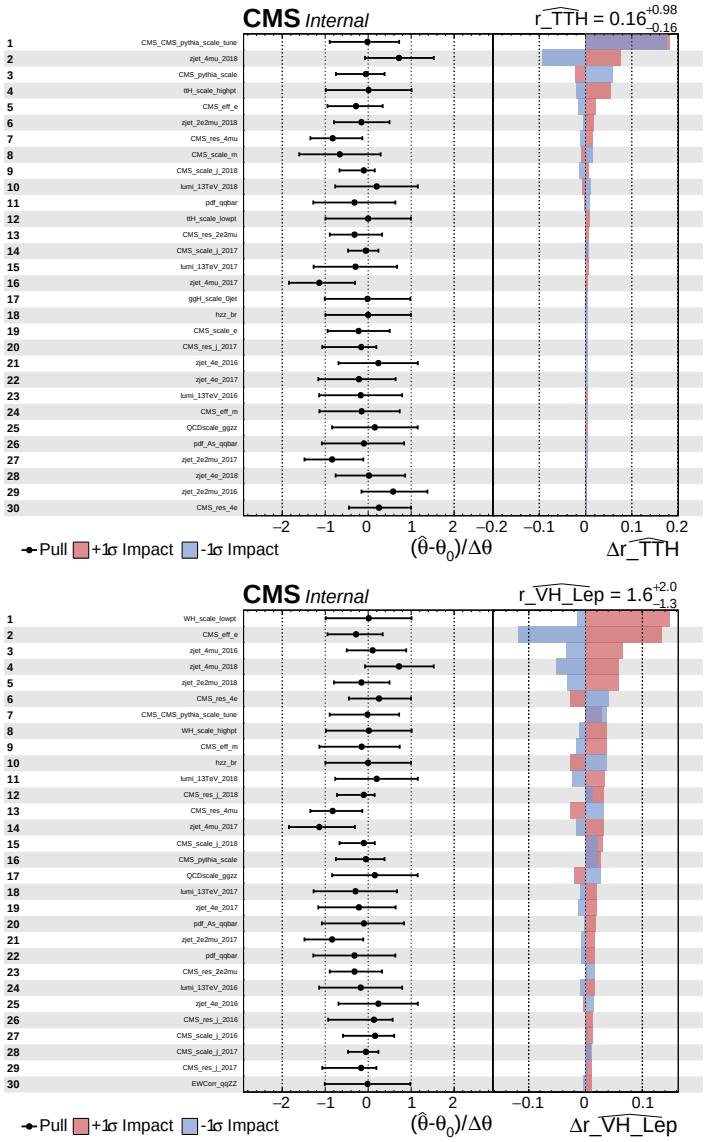


FIGURE D.2: Systematic uncertainties with the largest impact on the `ttH` (top) and `VH-lep` (bottom) cross section measurement with $m_{\mathrm{H}} = 125.38$ GeV. Value and error of the corresponding nuisance parameter after the maximum likelihood fit to the observed data are shown.

Efficiency matrices of the fiducial differential cross section measurements

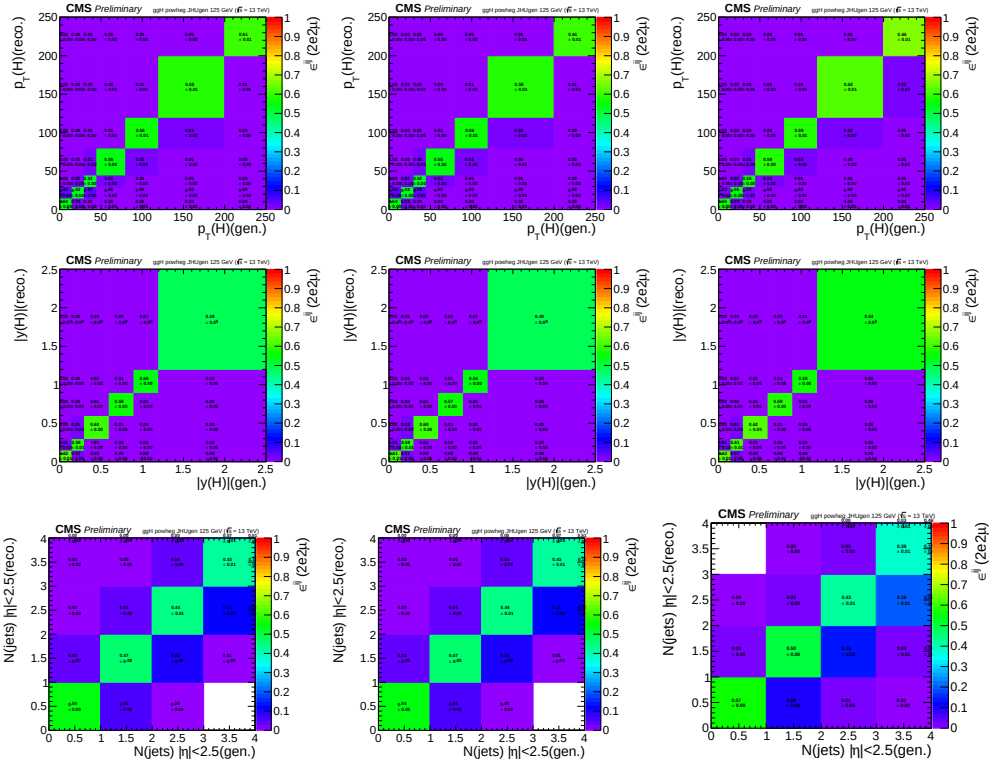


FIGURE D.3: Efficiency matrices for the the transverse momentum (top) and rapidity (middle) of the H boson and the jet multiplicity (bottom) for the gluon fusion production mode in the $2e2\mu$ final state in 2016 (left), 2017 (middle), and 2018 (right).

Bibliography

[1] A. Zee, *Group Theory in a Nutshell for Physicists*, Princeton University Press, USA, 2016.

[2] I. Aitchison and A. Hey, *Gauge theories in particle physics: A practical introduction. Vol. 2: Non-Abelian gauge theories: QCD and the electroweak theory*, CRC Press, 2012.

[3] C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions: Second Edition*, Princeton University Press, USA, 9 2013.

[4] D. Griffiths, *Introduction to elementary particles*, Wiley, 2008.

[5] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction.* Prog. Theor. Phys. **49** (1973) 652.

[6] P. W. Higgs, *Broken symmetries, massless particles and gauge fields.* Phys. Lett. **12** (1964) 132.

[7] P. W. Higgs, *Broken symmetries and the masses of gauge bosons.* Phys. Rev. Lett. **13** (1964) 508.

[8] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons.* Phys. Rev. **145** (1966) 1156.

[9] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons.* Phys. Rev. Lett. **13** (1964) 321.

[10] G. Guralnik, C. Hagen, and T. Kibble, *Global conservation laws and massless particles.* Phys. Rev. Lett. **13** (1964) 585.

[11] T. Kibble, *Symmetry breaking in nonabelian gauge theories.* Phys. Rev. **155** (1967) 1554.

[12] G. Guralnik, C. Hagen, and T. Kibble, *Broken symmetries and the Goldstone theorem.* Adv. Part. Phys. **2** (1968) 567.

[13] 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. B **716** (2012) 1.

[14] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.* Phys. Lett. B **716** (2012) 30.

[15] CMS Collaboration, *Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ = 7 and 8 TeV.* JHEP **06** (2013) 081.

[16] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector.* CERN-2017-002-M (CERN, Geneva, 2016).

[17] CMS Collaboration, *https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.*

[18] LHC Higgs Cross Section Working Group, *LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle*, Tech. Rep. arXiv:1209.0040. LHCHXSWG-2012-001. CERN-PH-TH-2012-284, CERN, Sep 2012.

[19] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties.* CERN-2013-004 (2013).

[20] M. Cepeda, S. Gori, et al., *Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC.* CERN Yellow Rep. Monogr. **7** (2018) 221.

[21] J. De Blas et al., *Higgs Boson Studies at Future Particle Colliders.* JHEP **01** (2020) 139.

[22] P. Teixeira-Dias, *Higgs boson searches at LEP.* J. Phys. Conf. Ser. **110** (2008) 042030.

[23] R. Barate et al., *Search for the standard model Higgs boson at LEP.* Phys. Lett. B **565** (2003) 61.

[24] G. Cowan, *Statistical data analysis*, Oxford University Press, 1998.

[25] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics.* Eur. Phys. J. C **71** (2011) 1554.

[26] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics.* The European Physical Journal C **71** (2011).

[27] CDF and D0 Collaborations, *Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.2 fb^{-1} of Data.* arXiv:1109.0576 (2011).

[28] CMS Collaboration, *Study of the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs.* Phys. Rev. Lett. **110** (2013) 081803.

[29] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment.* Eur. Phys. J. C **76** (2016).

[30] CMS Collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV.* Eur. Phys. J. C **75** (2015) 212.

[31] ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV.* JHEP **08** (2016) 45.

[32] ATLAS Collaboration, *Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment.* Phys. Rev. D **101** (2020) 012002.

[33] CMS Collaboration, *Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV.* Eur. Phys. J. C **79** (2019) 421.

[34] CMS Collaboration, *Evidence for Higgs boson decay to a pair of muons* (2020).

[35] CMS Collaboration, *https://cds.cern.ch/record/2725728.*

[36] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector.* Phys. Lett. B **784** (2018) 173.

[37] CMS Collaboration, *Observation of $t\bar{t}H$ production.* Phys. Rev. Lett. **120** (2018) 231801.

[38] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector.* JHEP **04** (2015) 117.

[39] CMS Collaboration, *Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector.* Phys. Lett. B **779** (2018) 283.

[40] ATLAS Collaboration, *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector.* Phys. Lett. B **786** (2018) 59.

[41] CMS Collaboration, *Observation of Higgs boson decay to bottom quarks.* Phys. Rev. Lett. **121** (2018) 121801.

[42] E. Fontanesi and S. Braibant, *Measurement of the Higgs Self-Coupling in the $HH \to VVbb$ Channel at the FCC-hh Collider.* PoS **LHCP2018** (2018) 090.

[43] E. Fontanesi and L. Borgonovi, *Projections for HH measurements in the bbZZ(4l) final state with the CMS experiment at the HL-LHC.* PoS **LHCP2019** (2019) 038.

[44] L. Borgonovi, *Higgs self-coupling measurements with the CMS detector at the LHC and at future colliders in the $HH \to ZZ(4l)bb$ decay channel*, PhD thesis, alma, Marzo 2020.

[45] ATLAS Collaboration, *Projections for measurements of Higgs boson cross sections, branching ratios, coupling parameters and mass with the ATLAS detector at the HL-LHC*, Tech. Rep. ATL-PHYS-PUB-2018-054, CERN, Geneva, Dec 2018.

[46] CMS Collaboration, *Sensitivity projections for Higgs boson properties measurements at the HL-LHC*, Tech. Rep. CMS-PAS-FTR-18-011, CERN, Geneva, 2018.

[47] CMS Collaboration, *Constraints on anomalous Higgs boson couplings to vector bosons and fermions in production and decay in the $H \to 4\ell$ channel* (2020, in publication).

[48] CMS Collaboration, *Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV.* JHEP **11** (2017) 047.

[49] CMS Collaboration, *A measurement of the higgs boson mass in the diphoton decay channel.* Physics Letters B **805** (2020) 135425.

[50] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, CERN Yellow Reports: Monographs, CERN, Oct 2016.

[51] A. Azatov, R. Contino, G. Panico, and M. Son, *Effective field theory analysis of double Higgs boson production via gluon fusion.* Phys. Rev. D **92** (2015) 035001.

[52] M. L. Mangano, G. Ortona, and M. Selvaggi, *Measuring the Higgs self-coupling via Higgs-pair production at a 100 TeV p-p collider*, Tech. Rep. arXiv:2004.03505, CERN, Geneva, Apr 2020.

[53] O. Brüning, H. Burkhardt, and S. Myers, *The Large Hadron Collider.* Prog. Part. Nucl. Phys. **67** (2012) 705.

[54] O. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, Geneva, 2004.

[55] T. S. Pettersson and P. Lefèvre, *The Large Hadron Collider: conceptual design*, Tech. Rep. CERN-AC-95-05-LHC, CERN, Oct 1995.

[56] E. Mobs, *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018* (2018). General Photo.

[57] CMS Collaboration, *https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults*.

[58] CMS Collaboration, *CMS: The Compact Muon Solenoid: Letter of intent for a general purpose detector at the LHC*. CERN-LHCC-92-03 (1992).

[59] CMS Collaboration, *The CMS experiment at the CERN LHC*. JINST **3** (2008) S08004.

[60] CMS Collaboration, *The CMS magnet project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

[61] CMS Collaboration, *The CMS tracker system project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

[62] CMS Collaboration, *CMS Technical Design Report for the Pixel Detector Upgrade*, Tech. Rep. CERN-LHCC-2012-016. CMS-TDR-11, CERN, Sep 2012.

[63] CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

[64] P. Adzic et al., *Energy resolution of the barrel of the CMS electromagnetic calorimeter*. JINST **2** (2007) P04004.

[65] CMS Collaboration, *The CMS hadron calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

[66] S. Abdullin et al., *The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c*. Eur. Phys. J. C **60** (2009) 359. [Erratum: Eur.Phys.J.C 61, 353–356 (2009)].

[67] CMS Collaboration, *The CMS muon project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

[68] CMS Collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at* $\sqrt{s} = 13$ *TeV*. JINST **13** (2018) P06015.

[69] CMS Collaboration, *The CMS trigger system*. JINST **12** (2017) P01020.

[70] CMS Collaboration, *CMS Technical Design Report for the Level-1 Trigger Upgrade*, Tech. Rep. CERN-LHCC-2013-011. CMS-TDR-12, CERN, Jun 2013.

[71] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*. Journal of Instrumentation **12** (2017) 10003.

[72] I. Bejar, O. Brüning, P. Fessia, L. Rossi, R. Tomas, and M. Zerlauth, *The HL-LHC Machine*. Adv. Ser. Direct. High Energy Phys. **24** (2015) 31.

[73] European Strategy Group, *2020 Update of the European Strategy for Particle Physics*, CERN Council, Geneva, 2020.

[74] A. Abada et al., *Fcc physics opportunities: Future circular collider conceptual design report volume 1.* European Physical Journal C **79** (2019).

[75] M. Benedikt, A. Blondel, et al., *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2. Future Circular Collider*, Tech. Rep. CERN-ACC-2018-0057. 2, CERN, Geneva, Dec 2018.

[76] CEPC Study Group, *CEPC Conceptual Design Report: Volume 1 - Accelerator.* arXiv:1809.00285 (2018).

[77] CEPC Study Group, *CEPC Conceptual Design Report: Volume 2 - Physics \& Detector.* arXiv:1811.10545 (2018).

[78] T. Behnke, J. E. Brau, B. Foster, J. Fuster, M. Harrison, J. M. Paterson, M. Peskin, M. Stanitzki, N. Walker, and H. Yamamoto, *The International Linear Collider Technical Design Report - Volume 1: Executive Summary.* arXiv:1306.6327 (2013).

[79] K. Fujii et al., *Physics Case for the 250 GeV Stage of the International Linear Collider.* arXiv:1710.07621 (2017).

[80] M. Aicheler, P. Burrows, M. Draper, T. Garvey, P. Lebrun, K. Peach, N. Phinney, H. Schmickler, D. Schulte, and N. Toge, *A Multi-TeV Linear Collider Based on CLIC Technology: CLIC Conceptual Design Report*, CERN Yellow Reports: Monographs, CERN, Geneva, 2012.

[81] M. J. Boland et al., *Updated baseline for a staged Compact Linear Collider*, CERN Yellow Reports: Monographs, CERN, Geneva, Aug 2016. Comments: 57 pages, 27 figures, 12 tables.

[82] A. Abada et al., *FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3.* Eur. Phys. J. ST **228** (2019) 755.

[83] M. Ahmad et al., *CEPC-SPPC Preliminary Conceptual Design Report. 1. Physics and Detector.* IHEP-CEPC-DR-2015-01 (2015).

[84] M. Ahmad et al., *CEPC-SPPC Preliminary Conceptual Design Report. 2. Accelerator.* IHEP-CEPC-DR-2015-01 (2015).

[85] M. Krammer, *The update of the european strategy for particle physics.* Physica Scripta **T158** (2013) 014019.

[86] G. Bencivenni, R. De Oliveira, M. G., and M. Poli Lener, *The micro-resistive well detector: a compact spark-protected single amplification-stage mpgd.* Journal of Instrumentation **10** (2015) P02008.

[87] G. Bencivenni, R. De Oliveira, G. Felici, M. Gatta, M. Giovannetti, G. Morello, A. Ochi, M. Poli Lener, and E. Tskhadadze, *The μ-RWELL layouts for high particle rate.* JINST **14** (2019) P05014.

[88] G. Bencivenni et al., *On the space resolution of the μ-RWELL.* arXiv:2007.03223 (2020).

[89] S. Lee, M. Livan, and R. Wigmans, *Dual-readout calorimetry.* Reviews of Modern Physics **90** (2017).

[90] S. Lee, M. Livan, and R. Wigmans, *Dual-Readout Calorimetry.* Rev. Mod. Phys. **90** (2018) 025002.

[91] R. Ferrari, *Dual-readout calorimetry: recent results from RD52 and plans for experiments at future* $\mathrm{e^+e^-}$ *colliders.* JINST **13** (2018) C02050.

[92] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector.* Phys. Rev. D **91** (2015) 012006.

[93] CMS Collaboration, *Measurement of the properties of a Higgs boson in the four-lepton final state.* Phys. Rev. D **89** (2014) 092007.

[94] CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV.* Phys. Rev. D **92** (2015) 012004.

[95] ATLAS Collaboration, *Measurement of the Higgs boson coupling properties in the* $H \to ZZ^* \to 4\ell$ *decay channel at* $\sqrt{s}$ *= 13 TeV with the ATLAS detector.* JHEP **03** (2018) 095.

[96] CMS Collaboration, *Constraints on the Higgs boson width from off-shell production and decay to z-boson pairs.* Phys. Lett. B **736** (2014) 64.

[97] CMS Collaboration, *Limits on the Higgs boson lifetime and width from its decay to four charged leptons.* Phys. Rev. D **92** (2015) 072010.

[98] ATLAS Collaboration, *Constraints on the off-shell Higgs boson signal strength in the high-mass* ZZ *and* WW *final states with the ATLAS detector.* Eur. Phys. J. C **75** (2015) 335.

[99] ATLAS Collaboration, *Constraints on off-shell Higgs boson production and the Higgs boson total width in* $ZZ \to 4\ell$ *and* $ZZ \to 2\ell 2\nu$ *final states with the ATLAS detector.* Phys. Lett. B **786** (2018) 223.

[100] ATLAS Collaboration, *Higgs boson production cross-section measurements and their EFT interpretation in the* 4ℓ *decay channel at* $\sqrt{s}$ *=13 TeV with the ATLAS detector.* Eur. Phys. J. C **80** (2020) 957.

[101] CMS Collaboration, *Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state.* Phys. Lett. B **775** (2017) 1.

[102] CMS Collaboration, *Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state.* Phys. Rev. D **99** (2019) 112003.

[103] ATLAS Collaboration, *Fiducial and differential cross sections of Higgs boson production measured in the four-lepton decay channel in* pp *collisions at* $\sqrt{s}$*=8 TeV with the ATLAS detector.* Phys. Lett. B **738** (2014) 234.

[104] CMS Collaboration, *Measurement of differential and integrated fiducial cross sections for Higgs boson production in the four-lepton decay channel in pp collisions at* $\sqrt{s} = 7$ *and 8 TeV.* JHEP **04** (2016) 005.

[105] ATLAS Collaboration, *Measurement of inclusive and differential cross sections in the* $H \to ZZ^* \to 4\ell$ *decay channel in* pp *collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector.* JHEP **10** (2017) 132.

[106] ATLAS Collaboration, *Measurements of the Higgs boson inclusive and differential fiducial cross sections in the* 4ℓ *decay channel at* $\sqrt{s}$ *= 13 TeV.* Eur. Phys. J. C **80** (2020) 942.

[107] W. Adam, B. Mangano, T. Speer, and T. Todorov, *Track Reconstruction in the CMS tracker*, Tech. Rep. CMS-NOTE-2006-041, CERN, Geneva, 2006.

[108] CMS Collaboration, *Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV*. JINST **7** (2012) P10002.

[109] CMS Collaboration, *https://twiki.cern.ch/twiki/pub/CMSPublic/PhysicsResultsMUO*.

[110] A. Bodek, A. van Dyne, J. Han, W. Sakumoto, and A. Strelnikov, *Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments*. Eur. Phys. J. C **72** (2012) 2194.

[111] 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** (2005) N9.

[112] CMS Collaboration, *Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV*. JINST **10** (2015) P06005.

[113] CMS Collaboration, *Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC* (2020).

[114] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*. arXiv:1603.02754 (2016).

[115] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*. JHEP **04** (2008) 063.

[116] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual*. Eur. Phys. J. C **72** (2012) 1896.

[117] CMS Collaboration, *Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV*. JINST **12** (2017) P02014.

[118] CMS Collaboration, *Jet algorithms performance in 13 TeV data* (2017).

[119] CMS Collaboration, *Pileup mitigation at CMS in 13 TeV data*. JINST **15** (2020) P09018.

[120] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*. Phys. Lett. B **659** (2008) 119.

[121] CMS Collaboration, *Identification of heavy-flavour jets with the cms detector in pp collisions at 13 TeV*. JINST **13** (2018) P05011.

[122] CMS Collaboration, *CMS luminosity measurements for the 2016 data-taking period*, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017.

[123] CMS Collaboration, *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*, CMS Physics Analysis Summary CMS-PAS-LUM-17-004, CERN, 2017.

[124] CMS Collaboration, *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13TeV$*, CMS Physics Analysis Summary CMS-PAS-LUM-18-002, CERN, 2019.

[125] M. Tanabashi et al., *Review of particle physics*. Phys. Rev. D **98** (2018) 030001.

[126] Y. Gao, A. V. Gritsan, Z. Guo, K. Melnikov, M. Schulze, and N. V. Tran, *Spin determination of single-produced resonances at hadron colliders.* Phys. Rev. D **81** (2010) 075022.

[127] S. Bolognesi, Y. Gao, A. V. Gritsan, K. Melnikov, M. Schulze, N. V. Tran, and W. A., *On the spin and parity of a single-produced resonance at the LHC.* Phys. Rev. D **86** (2012) 095031.

[128] I. Anderson, S. Bolognesi, F. Caola, Y. Gao, A. V. Gritsan, C. B. Martin, K. Melnikov, M. Schulze, N. V. Tran, A. Whitbeck, and Y. Zhou, *Constraining anomalous HVV interactions at proton and lepton colliders.* Phys. Rev. D **89** (2014) 035007.

[129] CMS Collaboration, *Search for a Higgs boson in the mass range from 145 to 1000 GeV decaying to a pair of W or Z bosons.* JHEP **10** (2015) 144.

[130] A. V. Gritsan, R. Röntsch, M. Schulze, and M. Xiao, *Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques.* arXiv:1606.03107 (2016).

[131] CMS Collaboration, *Feasibility of CP parity measurement of the Higgs boson in ttH, bbH, and tqH production*, CMS Physics Analysis Note CMS-AN-16-185, CERN, 2016.

[132] N. Berger et al., *Simplified Template Cross Sections - Stage 1.1.* arXiv:1906.02754 (2019).

[133] *Les Houches 2017: Physics at TeV Colliders Standard Model Working Group Report*, 2018.

[134] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms.* JHEP **11** (2004) 040.

[135] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method.* JHEP **11** (2007) 070.

[136] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX.* JHEP **06** (2010) 043.

[137] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG.* JHEP **04** (2009) 002.

[138] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM.* JHEP **02** (2012) 088.

[139] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG.* JHEP **02** (2010) 037.

[140] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, *$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO.* JHEP **10** (2013) 1.

[141] H. B. Hartanto, B. Jager, L. Reina, and D. Wackeroth, *Higgs boson production in association with top quarks in the POWHEG BOX.* Phys. Rev. D **91** (2015) 094003.

[142] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and Z. M., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations.* JHEP **07** (2014) 079.

[143] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations.* JHEP **07** (2014) 079.

[144] C. Anastasiou, C. Duhr, F. Dulat, F. Herzog, and B. Mistlberger, *Higgs boson gluon-fusion production in QCD at three loops.* Phys. Rev. Lett. **114** (2015) 212001.

[145] C. Anastasiou, C. Duhr, F. Dulat, E. Furlan, T. Gehrmann, F. Herzog, A. Lazopoulos, and B. Mistlberger, *High precision determination of the gluon fusion Higgs boson cross-section at the LHC.* JHEP **05** (2016) 58.

[146] M. Ciccolini, A. Denner, and S. Dittmaier, *Strong and electroweak corrections to the production of a Higgs boson+2 jets via weak interactions at the Large Hadron Collider.* Phys. Rev. Lett. **99** (2007) 161803.

[147] M. Ciccolini, A. Denner, and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC.* Phys. Rev. D **77** (2008) 013002.

[148] P. Bolzoni, F. Maltoni, S. Moch, and M. Zaro, *Higgs production via vector-boson fusion at NNLO in QCD.* Phys. Rev. Lett. **105** (2010) 011801.

[149] P. Bolzoni, F. Maltoni, S. Moch, and M. Zaro, *Vector boson fusion at next-to-next-to-leading order in QCD: Standard model Higgs boson and beyond.* Phys. Rev. D **85** (2012) 035002.

[150] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders.* Phys. Lett. B **579** (2004) 149.

[151] M. L. Ciccolini, S. Dittmaier, and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders.* Phys. Rev. D **68** (2003) 073003.

[152] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, and P. M. Zerwas, *Higgs radiation off top quarks at the Tevatron and the LHC.* Phys. Rev. Lett. **87** (2001) 201805.

[153] W. Beenakker, S. Dittmaier, M. Kramer, B. Plumper, M. Spira, and P. M. Zerwas, *NLO QCD corrections to $\mathrm{t\bar{t}H}$ production in hadron collisions.* Nucl. Phys. B **653** (2003) 151.

[154] S. Dawson, L. H. Orr, L. Reina, and D. Wackeroth, *Associated top quark Higgs boson production at the LHC.* Phys. Rev. D **67** (2003) 071503.

[155] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackeroth, *Associated Higgs production with top quarks at the Large Hadron Collider: NLO QCD corrections.* Phys. Rev. D **68** (2003) 034022.

[156] Z. Yu, M. Wen-Gan, Z. Ren-You, C. Chong, and G. Lei, *QCD NLO and EW NLO corrections to $t\bar{t}H$ production with top quark decays at hadron collider.* Phys. Lett. B **738** (2014) 1.

[157] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, *Weak corrections to Higgs hadroproduction in association with a top-quark pair.* JHEP **09** (2014) 65.

[158] F. Demartin, F. Maltoni, K. Mawatari, and M. Zaro, *Higgs production in association with a single top quark at the LHC.* Eur. Phys. J. C **75** (2015) 267.

[159] F. Demartin, B. Maier, F. Maltoni, K. Mawatari, and M. Zaro, *tWH associated production at the LHC.* Eur. Phys. J. C **77** (2017) 34.

[160] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira, *Standard model Higgs-boson branching ratios with uncertainties.* Eur. Phys. J. C **71** (2011) 1753.

[161] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension.* Comput. Phys. Commun. **108** (1998) 56.

[162] A. Djouadi, J. Kalinowski, M. Muhlleitner, and M. Spira, *An update of the program HDECAY*, in The Les Houches 2009 workshop on TeV colliders: The tools and Monte Carlo working group summary report, 2010.

[163] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, *Precise predictions for the Higgs-boson decay $H \to WW/ZZ \to 4$ leptons.* Phys. Rev. D **74** (2006) 013004.

[164] A. Bredenstein, A. Denner, S. Dittmaier, and M. M. Weber, *Radiative corrections to the semileptonic and hadronic Higgs-boson decays $H \to WW/ZZ \to 4$ fermions.* JHEP **02** (2007) 80.

[165] S. Boselli, C. M. Carloni Calame, G. Montagna, O. Nicrosini, and F. Piccinini, *Higgs boson decay into four leptons at NLOPS electroweak accuracy.* JHEP **06** (2015) 23.

[166] S. Actis, G. Passarino, C. Sturm, and S. Uccirati, *NNLO computational techniques: the cases $H \to \gamma\gamma$ and $H \to gg$.* Nucl. Phys. B **811** (2009) 182.

[167] R. D. Ball et al., *Parton distributions for the LHC Run II.* JHEP **04** (2015) 040.

[168] T. Sjöstrand et al., *An introduction to PYTHIA 8.2.* Computer Physics Communications **191** (2015) 159.

[169] V. Khachatryan et al., *Event generator tunes obtained from underlying event and multiparton scattering measurements.* Eur. Phys. J. C **76** (2016) 155.

[170] CMS Collaboration, *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements.* Eur. Phys. J. C **80** (2020) 4.

[171] S. Agostinelli et al., *GEANT4: a simulation toolkit.* Nucl. Instrum. Meth. A **506** (2003) 250.

[172] J. Allison et al., *Geant4 developments and applications.* IEEE Trans. Nucl. Sci. **53** (2006) 270.

[173] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini, *Higgs boson production at the LHC: transverse momentum resummation effects in the $H \to \gamma\gamma$, $H \to WW \to \ell\nu\ell\nu$ and $H \to ZZ \to 4\ell$ decay modes.* JHEP **06** (2012) 132.

[174] M. Grazzini and H. Sargsyan, *Heavy-quark mass effects in Higgs boson production at the LHC.* JHEP **09** (2013) 129.

[175] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, *NNLOPS simulation of Higgs boson production.* JHEP **10** (2013) 222.

[176] J. M. Campbell and R. K. Ellis, *MCFM for the Tevatron and the LHC.* Nucl. Phys. Proc. Suppl. **205** (2010) 10.

[177] J. M. Campbell, R. K. Ellis, and C. Williams, *Bounding the Higgs width at the LHC using full analytic results for* $\mathrm{gg} \to \mathrm{e}^-\mathrm{e}^+\mu^-\mu^+$. JHEP **04** (2014) 060.

[178] M. Grazzini, S. Kallweit, and D. Rathlev, *ZZ production at the LHC: Fiducial cross sections and distributions in NNLO QCD.* Phys. Lett. B **750** (2015) 407.

[179] A. Bierweiler, T. Kasprzik, and J. H. Kuhn, *Vector-boson pair production at the LHC to* $\mathcal{O}(\alpha^3)$ *accuracy.* JHEP **12** (2013) 071.

[180] M. Bonvini, F. Caola, S. Forte, K. Melnikov, and G. Ridolfi, *Signal-background interference effects in* $gg \to H \to WW$ *beyond leading order.* Phys. Rev. D **88** (2013) 034032.

[181] K. Melnikov and M. Dowling, *Production of two Z-bosons in gluon fusion in the heavy top quark approximation.* Phys. Lett. B **744** (2015) 43.

[182] C. S. Li, H. T. Li, D. Y. Shao, and J. Wang, *Soft gluon resummation in the signal-background interference process of* $gg(\to h^*) \to ZZ$. JHEP **08** (2015) 065.

[183] G. Passarino, *Higgs CAT.* Eur. Phys. J. C **74** (2014) 2866.

[184] S. Catani and M. Grazzini, *An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC.* Phys. Rev. Lett. **98** (2007) 222002.

[185] M. Grazzini, *NNLO predictions for the Higgs boson signal in the* $H \to WW \to \ell\nu\ell\nu$ *and* $H \to ZZ \to 4\ell$ *decay channels.* JHEP **02** (2008) 043.

[186] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, CERN Report CERN-2013-004, CERN, 2013.

[187] S. van der Meer, *Calibration of the effective beam height in the ISR*, Tech. Rep. CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968.

[188] CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS.* JINST **6** (2011) P11002.

[189] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II.* J. Phys. G **43** (2016) 023001.

[190] M. Botje et al., *The PDF4LHC working group interim recommendations.* arXiv:1101.0538 (2011).

[191] S. Alekhin et al., *The PDF4LHC working group interim report.* arXiv:1101.0536 (2011).

[192] CMS Collaboration, *Studies of Higgs boson production in the four-lepton final state at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. CMS-PAS-HIG-15-004, CERN, Geneva, 2016.

[193] CMS Collaboration, *Measurements of properties of the Higgs boson and search for an additional resonance in the four-lepton final state at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. CMS-PAS-HIG-16-033, CERN, Geneva, 2016.

[194] CMS Collaboration, *Measurements of properties of the Higgs boson in the four-lepton final state in proton-proton collisions at* $\sqrt{s} = 13$ TeV, Tech. Rep. CMS-PAS-HIG-19-001, CERN, Geneva, 2019, in publication.

[195] ATLAS and CMS Collaborations and LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in Summer 2011*, Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

[196] S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses.* Annals Math. Statist. **9** (1938) 60.

[197] ATLAS and CMS Collaborations and LHC Higgs Combination Group, *Procedure for the LHC Higgs boson search combination in summer 2011*, ATL-PHYS-PUB/CMS NOTE 2011-11, 2011/005, CERN, 2011.

[198] CMS Collaboration, *Measurement of differential cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s}$=8 TeV.* Eur. Phys. J. C **76** (2015) 13.

[199] F. Garwood, *Fiducial limits for the Poisson distribution.* Biometrika **28** (1936) 437.

[200] F. Sauli, *The gas electron multiplier (GEM): Operating principles and applications.* Nucl. Instrum. Meth. A **805** (2016) 2.

[201] M. Adinolfi et al., *The tracking detector of the KLOE experiment.* Nucl. Instrum. Meth. A **488** (2002) 51.

[202] A. Baldini et al., *MEG Upgrade Proposal.* arXiv:1301.7225 (2013).

[203] G. Chiarello, *The full stereo drift chamber for the MEG II experiment.* JINST **12** (2017) C03062.

[204] W. Leo, *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach*, Springer, 1994.

[205] G. Chiarello, A. Corvaglia, F. Grancagnolo, A. Miccoli, A. Panareo, and G. Tassielli, *The tracking system for the idea detector at future lepton colliders.* Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **936** (2019) 503. Frontier Detectors for Frontier Physics: 14th Pisa Meeting on Advanced Detectors.

[206] S. Lee et al., *Hadron detection with a dual-readout fiber calorimeter.* Nucl. Instrum. Meth. A **866** (2017) 76.

[207] G. Charpak, R. Bouclier, T. Bressani, J. Favier, and C. Zupancic, *The Use of Multiwire Proportional Counters to Select and Localize Charged Particles.* Nucl. Instrum. Meth. **62** (1968) 262.

[208] A. Oed, *Position Sensitive Detector with Microstrip Anode for electron Multiplication with Gases.* Nucl. Instrum. Meth. A **263** (1988) 351.

[209] J. De Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment.* JHEP **02** (2014) 057.

[210] M. Antonello, M. Caccia, R. Ferrari, G. Gaudio, L. Pezzotti, G. Polesello, E. Proserpio, and R. Santoro, *Expected performance of the IDEA dual-readout fully projective fiber calorimeter.* JINST **15** (2020) C06015.