UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Search for double Higgs events produced via a vector boson fusion mechanism in the decay channel bb̄4l with the CMS experiment at the LHC

Candidate:
Brunella D'Anzi

Supervisor:
Prof. Nicola De Filippis

A thesis submitted to the
Physics Department, University of Bari Aldo Moro (UniBa)
in partial fulfillment of the requirements for the degree of
Master of Science (M.Sc.)

in

Nuclear, Subnuclear and Astroparticle Physics

"*Ohana* means family.

Family means nobody gets left behind, or forgotten".

— Lilo & Stitch

To my beloved sister Antonella,
my mummy Maria, and my daddy Piero.

Search for double Higgs events produced via a vector boson fusion mechanism in the decay channel bb̄4l with the CMS experiment at the LHC

by Brunella D'Anzi

Abstract

The subject of this thesis is the search for Higgs boson pair production events (HH) produced via a vector boson fusion (VBF) mechanism using proton-proton collision data collected at $\sqrt{s} = 13$ TeV with the **C**ompact **M**uon **S**olenoid (CMS) experiment at the **L**arge **H**adron **C**ollider (LHC), **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire (CERN) . Events in which the two Higgs particles decay in two b-quarks and four charged leptons ($ZZ^* \to 4l$) respectively, also accompanied by two energetic forward jets usually coming from light quarks ($HHqq' \to b\bar{b}4lqq'$), are explored to investigate the **V**ector **B**oson **F**usion (VBF) non-resonant production mechanism. The VBF HH production provides a unique means to infer information about the quartic VVHH Higgs coupling and the already studied HHH trilinear Higgs self-coupling in the Higgs production via the gluon-gluon fusion process, which are fundamental points to shed light on the reality of **B**rout-**E**nglert-**H**iggs (BEH) mechanism to give the particles mass and consequently to probe **B**eyond **S**tandard **M**odel (BSM) theories.

The difficulty of the search, uninvestigated up to now, is mainly due to the small value of the signal production cross section weighted with the branching ratios BRs (for the HH production via VBF, with the Higgs mass set to its best fit value of 125.09 GeV, the cross section at $\sqrt{s} =13$ TeV is approximately 1.723 fb and the corresponding **B**ranching **R**atio (BR)s are 2.79×10^{-4} for $H \to ZZ^* \to 4l$, with $l = e, \mu, \tau$, and 5.75×10^{-1} for $H \to b\bar{b}$), thus requiring an exclusive event selection in order to efficiently perform a background rejection. Indeed, multivariate analysis techniques of machine learning are applied to improve the signal discrimination from background, the latter being dominated by the associated production of a Higgs boson and a top quark-antiquark pair. The investigation of the $HHqq' \to b\bar{b}4lqq'$ process explores three decay modes of

the system $4l$ that is $4e, 4\mu, 2e2\mu$ and has required a detailed study of leptons (electrons and muons), light quarks and b-jets reconstruction. Results in terms of upper limits on the signal strength are derived using an integrated luminosity of $59.74~\text{fb}^{-1}$ from the 2018 Run II campaign.

" We should perhaps finish our paper with an apology and a caution. We apologize to experimentalists for having no idea what is the mass of the Higgs boson , [. . .] and for not being sure of its couplings to other particles, except that they are probably all very small. For these reasons, we do not want to encourage big experimental searches for the Higgs boson , but we do feel that people doing experiments vulnerable to the Higgs boson should know how it may turn up . "

— Ellis, Gaillard, and Nanopoulos [1]

Acknowledgments

Throughout the writing of this dissertation I have received a great deal of support and assistance. I would first like to thank my supervisor, Prof. Nicola De Filippis [1], for accepting me into his CMS working group, and for the constant support I have been granted. His expertise was invaluable in formulating the right research questions and methodology. His insightful feedback pushed me to sharpen my thinking and brought my work to a professional level. During these two years, I have been given the opportunity to participate in many schools,workshops and conferences, the rare freedom to pursue some of my own research interests and ideas. I would like to acknowledge colleagues met in these amazing experiences, especially Master and Ph.D. students from the online DESY school and the three-month CERN internship for indirectly motivating me in improving my professional skills even during a worldwide pandemic. I am much indebted to my tutors, Dr. Giorgia Miniello [2] , Walaa Elmetenawee [2] , Ilirjan Margjeka [2] for their precious guidance throughout my studies. They provided me with the tools from lines of code to papers that I needed to choose the right direction and successfully complete my dissertation. A special thank goes to Professor Domenico Diacono [2] for having involved me in INFN projects that goes beyond physics but has always fascinated me. In addition, I wish to thank my parents for their wise counsel and sympathetic ear. They have been next to me all the time, sustaining me in the difficult moments and bearing all the difficulties with me. Without their strength and constant encouragement this thesis would not have been possible. And finally, I could not have completed this dissertation without the support of my friend, Alessandro, my beloved sister Antonella and my dear Raffaele, who have provided stimulating discussions as well as enjoyable distractions to rest my mind outside of my research.

---

1 Politecnico of Bari and INFN Section of Bari
2 Department of Physics – University of Bari *Aldo Moro* and INFN Section of Bari

Bari, Italy   The author
October 28, 2021   *Brunella D'Anzi*

Contents

List of Figures

List of Tables

Acronyms

ALICE   **A** Large **I**on **C**ollider **E**xperiment

ANN   **A**rtificial **N**eural **N**etwork

AOD   **A**nalysis **O**bject **D**ata

API   **A**pplication **P**rogramming **I**nterface

ATLAS   **A** **T**oroidal **L**HC **A**pparatu**S**

APDs   **A**valanche **P**hoto-**D**iodes

BEH   **B**rout-**E**nglert-**H**iggs

BPIX   **B**arrel **PIX**el detector

BR   **B**ranching **R**atio

BSM   **B**eyond **S**tandard **M**odel

BX   **B**unch **C**rossing

CERN   **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire

CKM   **C**abibbo **K**obayashi **M**askawa

C.L.   **C**onfidence **L**evel

CMS   **C**ompact **M**uon **S**olenoid

CP   **C**harge conjugation and **P**arity symmetry

CR   **C**ontrol **R**egion

CSC   **C**athode **S**trip **C**hamber

CTF   **C**ombinatorial **T**rack **F**inder

DAQ   **D**ata **A**c**Q**uision

DIELE   **diEl**ectron

DIMUELE   **DiMu**on**El**ectron

DQM   **D**ata **Q**uality **M**onitoring

DNN	**D**eep **N**eural **N**etwork
DT	**D**rift **T**ube
EB	**E**CAL **B**arrel
ECAL	**E**lectromagnetic **CAL**orimeter
EE	**E**CAL **E**ndcap
EFT	**E**ffective **F**ield **T**heory
EGI	**E**uropean **G**rid **I**nfrastructure
EM	**E**lectro**M**agnetic
EPIX	**E**ndcap **PIX**el detector
ES	**E**M Pre**S**hower
EW	**E**lectro**W**eak
EWSB	**E**lectro**W**eak **S**ymmetry **B**reaking
FASER	**F**orw**A**rd **S**earch **E**xpe**R**iment
FNAL	**F**ermi **N**ational **A**ccelerator **L**aboratory
FR	**F**ake **R**ate
FSR	**F**inal **S**tate **R**adiation
GEANT4	**GE**ometry **AN**d Tracking
HB	**H**CAL **B**arrel detector
HCAL	**H**adronic **CAL**orimeter
HE	**H**CAL **E**nd-cap detector
HEP	**H**igh **E**nergy **P**hysics
HF	**H**CAL **F**orward detector
HL-LHC	**H**igh **L**uminosity **L**arge **H**adron **C**ollider
HLT	**H**igh **L**evel **T**rigger
HO	**H**CAL **O**uter calorimeter
HPDs	**H**ybrid **P**hoto**D**iodes

IP   Impact Parameter

INFN   Istituto Nazionale di Fisica Nucleare

ISR   Initial State Radiation

JEC   Jet Energy Correction

L1   Level-1

L1 GT   L1 Global Trigger

LEP   Large Electron-Positron collider

LHC   Large Hadron Collider

LHCOPN   LHC Optical Private Network

LHCʙ   Large Hadron Collider beauty experiment

LHCꜰ   Large Hadron Collider forward

LHE   Les Houches Event

LINAC2   Linear Accelerator 2

LO   Leading Order

LS   Long Shutdown

LS2   Long Shutdown 2

LS3   Long Shutdown 3

MELA   Matrix Element Likelihood Analysis

ML   Machine Learning

MɪɴɪAOD   MiniAnalysis Object Data

MC   Monte Carlo

MoEDAL   Monopole and Exotics Detector At the LHC

MB   Muon Barrel

ME   Muon Endcap

MSM   Minimal Standard Model

MuEʟᴇ   MuonElectron

MUDIELE MuonDiElectron

MVA MultiVAriate analysis

NEIC Nordic e-Infrastructure Collaboration

NLO Next to Leading Order

NN Neural Network

NNLO Next to Next to Leading Order

OSG Open Science Grid

PDF Parton Distribution Function

PF Particle Flow

PMNS Pontecorvo Maki Nakagawa Sakata

PMTs Photomultiplier Tubes

PS Parton Shower

PSB Proton Synchrotron Booster

p_T Transverse Momentum

PU Pile-Up

RECAS Rete di Calcolo per SuperB

SINGLEELE singleElectron

RECO RECOnstructed

SND Scattering and Neutrino Detector

QCD Quantum Chromo Dynamics

QED Quantum ElectroDynamics

QFT Quantum Field Theory

PV Primary Vertex

RF RadioFrequency cavities

ROC Receiver Operating Characteristic curve

RPC Resistive Plate Chamber

SIP   **S**ignificance of the **I**mpact **P**arameter

SM   **S**tandard **M**odel

SPS   **S**uper **P**roton **S**ynchrotron

SR   **S**ignal **R**egion

SSV   **S**imple **S**econdary **V**ertex

SUSY   **SU**per **SY**mmetry

SV   **S**econdary **V**ertex

TEC   **T**racker **E**ndcaps

TIB   **T**racker **I**nner **B**arrel

TID   **T**racker **I**nner **D**isks

TC   **T**rack **C**ounting

TOB   **T**racker **O**uter **B**arrel

TMVA   **T**oolkit for **M**ultivariate **D**ata **A**nalysis

TOTEM   **T**otal Cross **S**ection, **E**lastic **S**cattering and **D**iffraction **D**issociation

TRIELE   **triEle**ctron

VEV   **V**acuum **E**xpectation **V**alue

VPTs   **V**acuum **P**hoto**T**riodes

VBF   **V**ector **B**oson **F**usion

WLCG   **W**orld LHC Computing Grid

WLFs   **W**avelength **S**hifting **F**ibers

JBP   **J**et **B** **P**robability

JP   **J**et **P**robability

2HDM   **T**wo-**H**iggs-**d**oublet **M**odel

3D   **T**hree **D**imensional

Introduction

D URING the 20th century, aiming to give a detailed account in terms of
fundamental laws on how our Universe works and what it is made of,
Elementary Particle Physics has stimulated many physicists around the world
to persevere in this research field by providing ideas, theories and experiments.
Remarkable results were achieved with no equal in any other experimental science
that had, and is continuing to produce, a profound impact on our modern society
that the scientific community appreciated by awarding more than 50 Nobel prizes
to **H**igh **E**nergy **P**hysics (HEP) experts.

After the first exploration of the sub-atomic scales (a timeline guide is reported
in Figure 0.1) happened at the end of the 19th century (1897) by Sir Joseph
John Thomson, while he was working with cathode rays and discovered that they
were made up of negatively charged light corpuscles with unique charge/mass
ratio called electrons [3], our understanding of the constituents of the matter and
their relation to the space and time coordinates with which their interactions are
described is now well supported by the two main theoretical pillars of particle
physics, quantum mechanics and special relativity. Together with the general
relativity, they provide a mathematical description that is capable of representing
the behaviour of our Universe at spatial scales that extend over more than 40
orders of magnitude, and they can elucidate its past history, and tell more about
its future.

The mathematical theory that supports the description of the subnuclear world
is the Standard Model (SM) of particle physics, a **Q**uantum **F**ield **T**heory (QFT)
formalized between the 1960s and 1970s. It relies on the mathematically concept
of local gauge invariance under symmetry groups of its Lagrangian $\mathcal{L}$, the function
from which the equations of motion of a physical system can be derived. The SM
is at the origin of the unification of **E**lectro**M**agnetic (EM), from the **Q**uantum
Electro**D**ynamics (QED) model, and weak forces, by predicting the existence of
the vector (massless) photon γ mediator, the intermediate vector (massless at this
level) bosons $W^{\pm}$ and Z^0, and determining their mutual interactions and those

with the matter. Moreover, it explains the existence of the eight massless vector particles called gluons as the mediator of the strong force and its interactions with quarks, by incorporating in this way the **Q**uantum **C**hromo **D**ynamics (QCD) into a more complete model.
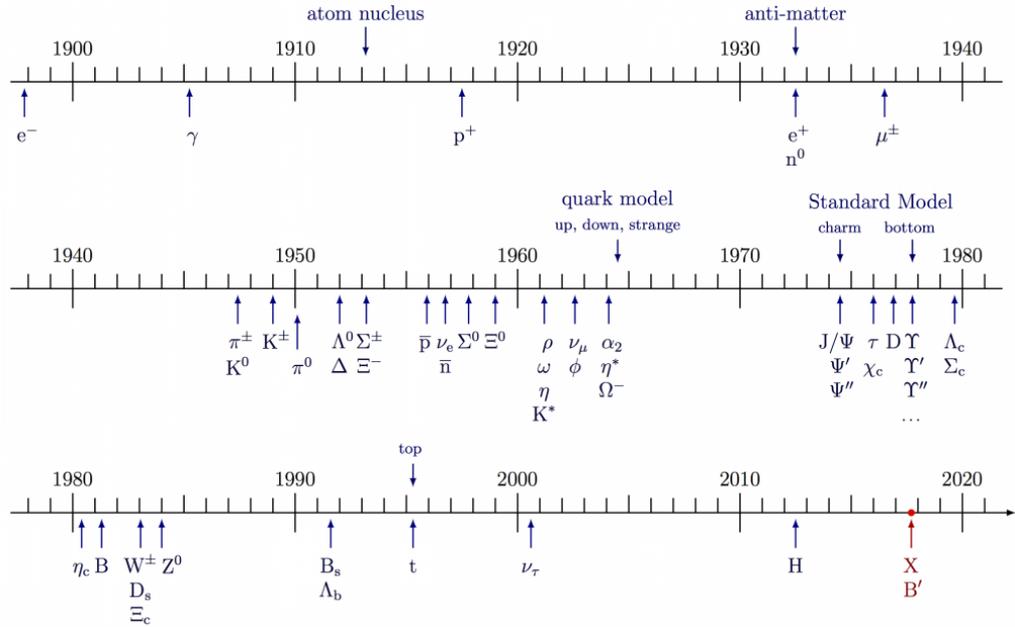
Figure 0.1: Timeline of particle physics discoveries [4].

It has precisely predicted a wide variety of phenomena and so far successfully explained almost all experimental results in particle physics. Indeed, after the forecasted discovery of the $W^{\pm}$ and Z^0 bosons at the CERN **S**uper **P**roton **S**ynchrotron (SPS) in 1983 by UA1 and UA2 Collaborations [5–9] and the observation of the top quark at the **F**ermi **N**ational **A**ccelerator **L**aboratory (FNAL) Tevatron in 1995 by CDF and DØ Collaborations [10–12], further verifications of its predictions have been performed, most notably at the CERN **L**arge **E**lectron-**P**ositron collider (LEP). The SM has been capable of correctly describing all the measurements performed until now at the GeV and TeV [3] energy scales.

However, this theoretical picture alone cannot explain the experimental observation of massive fermions and $W^{\pm}$ and Z^0 bosons, and predicts unphysical

---

3 1 GeV $= 10^9$ eV, 1 TeV $= 10^{12}$ eV . An electronvolt (symbol eV, also written electron-volt and electron volt) is the measure of an amount of kinetic energy gained by a single electron accelerating from rest through an electric potential difference of one volt in vacuum. When used as a unit of energy, the numerical value of 1 eV in Joules (symbol J) is equivalent to the numerical value of the charge of an electron in Coulombs (symbol C). Under the 2019 redefinition of the International System (SI) base units, this sets 1 eV equal to the exact value 1.60217663410^{-19} J.

properties for vector boson scattering processes [13]. The great innovation of the SM particle physics theory relies on the introduction of the BEH mechanism, as called from the name of Belgian (Robert Brout, Francois Englert) and Scotsman (Peter Higgs) physicists who made a proposal in order to solve this problem in 1964 [14–18]. The BEH mechanism postulates the existence of a doublet of complex scalar fields that causes an **E**lectro**W**eak **S**ymmetry **B**reaking (EWSB), thus giving masses to the vector bosons, providing naturally mass terms for the fermion masses via a Yukawa interaction, and ensuring the unitarity of the theory. The BEH mechanism introduces in the SM model the presence a scalar boson field, called the Higgs field, the last missing piece with a mass m_H that is a free parameter of the model. Particles that interact intensely with the Higgs field are heavy, while those that have feeble interactions are light.

Following the incorporation of the BEH mechanism in the electroweak model by Weinberg and Salam in 1964 [19] and the proof by 't Hooft and Veltman in 1972 that the theory was renormalizable [20], the search for the Higgs boson, the particle associated with the Higgs field, became the main goal of the researches performed at high energy particle colliders in the late 1980s. However, even if some constrains on the Higgs boson mass with theoretical arguments and direct searches at the LEP and Tevatron colliders [21, 22] were conducted, its existence could not be directly proved for almost half a century.

In October 1995, the LHC technical design report was published for constructing a new particle hunter of the previously undetected Higgs boson running underneath the frontier between Switzerland and France and exploring the physics at the TeV energy scale. Its main purpose is colliding protons at a centre-of-mass energy up to 14 TeV, in four interactions points where the four main experiments of the LHC operate. They consist in two general-purpose detectors, ATLAS and the CMS, with two independently designed detectors (this is vital for cross-confirmation of any new discoveries made) and the B physics **L**arge **H**adron **C**ollider **b**eauty experiment (LHCв) and the heavy-ion physics **A** **L**arge **I**on **C**ollider **E**xperiment (ALICE) detectors, focusing on specific phenomena. These four detectors sit underground in huge caverns on the LHC ring. Moreover, five smaller special-purpose experiments are present at LHC : the **L**arge **H**adron **C**ollider **f**orward (LHCF), **M**onopole and **E**xotics **D**etector **A**t the **L**HC (MoEDAL), **T**otal **C**ross **S**ection, **E**lastic **S**cattering and **D**iffraction **D**issociation (TOTEM), **F**orw**A**rd **S**earch **E**xpe**R**iment (FASER) and **S**cattering and **N**eutrino **D**etector (SND). By reporting the words of the LHC Physics Center Distinguished Researcher Luca Cadamuro [23]:

"The realization of the LHC has thousands of physicists and engineers from the entire world to design, build, commission, and run what is probably the most complex machine ever realized .

On July 4, 2012, the ATLAS and CMS collaborations announced that decades of searching and generations of experiments had finally culminated in the discovery of a new boson with a mass close to 125 GeV[4] [24, 25], which confirmed the BEH mechanism. However, finding it is not the end of the story, and researchers have to study the Higgs boson in detail to measure its properties and pin down its rarer decays. For this purpose, the data collected between 2010 and 2012, in the machine operations denoted as Run I, were crucial. It turned out that the Higgs boson is the only known elementary boson with a 0 spin, which means a scalar field, and whose existence does not arise from a local gauge invariance mechanism. It breaks the degeneracy between the three families of fermions by giving them their mass with couplings of different strengths, and its observation proves that the origin of particles' mass is purely a quantum-relativistic effect. The measurements performed with the LHC Run I data showed a good agreement with the predictions of the SM. It is now extremely important to precisely characterize the properties and couplings of the Higgs boson. In this context, the Higgs boson self-interactions are of particular interest because they provide invaluable information to reconstruct the shape of the scalar potential itself. As little is currently known experimentally about these interactions, their measurement represents one of the main elements for a complete characterization of the scalar sector of the SM.

The SM does not provide a mechanism that determines its 18 input parameters (if neutrino are massless we refer to it as the **M**inimal **S**tandard **M**odel (MSM)). In particular, it does not explain why three families of fermions exist, what is at the origin of their couplings to the Higgs boson, and why their values range over several order of magnitude. Being responsible for the breaking of the degeneracy between the fermion families, the scalar sector of the SM has a direct role in this context. Moreover, the mass of the Higgs boson is not protected by any fundamental symmetry of the theory, making it largely sensitive to divergent radiative corrections. These corrections need to be finely tuned to account for the observed Higgs boson mass of about 125 GeV, raising an issue about the naturalness of the theory. Finally, this specific mass value implies an instability of the vacuum, as the evolution of the Higgs self-coupling (and consequently of the shape of the scalar potential) with the energy implies that at higher scales this can become negative. This results in a metastable conditions of our Universe, that could collapse into a different vacuum state. These theoretical considerations should be regarded together with the limitations of the SM in explaining the experimental observations at cosmological scales.

---

4 Researchers in particle physics, nuclear physics and astrophysics adopt "natural units", where $\hbar$ = 1 and c = 1 and the unit of energy is the GeV. All basic quantities (length, area, time, rate, momentum, mass) can be expressed in terms of powers of the eV.

The SM does not provide a mechanism that is responsible for the observed matter-antimatter asymmetry in the Universe, nor it accounts for the existence of a stable particle that is compatible with the dark matter, whose are actually observed gravitational effects. Although being extremely successful at describing the phenomenology of collider experiments, the SM seems incomplete. When considered together, these open questions seem to indicate that the SM is only a part of a more extended physics model. The existence of such physics BSM is a puzzle which currently focuses the efforts of the theoretical and experimental communities. Many different models have been devised to solve these problems and provide an explanation to the arbitrariness and unnaturalness of many aspects of the SM.

Being something intimately different from all the other particles of the SM, the Higgs boson provides a preferential way to search for these SM extensions. No clear signs of its presence have been found in Run I data at the LHC. With the restart of the LHC in 2015 for its Run II these questions might find an answer. The increase in the centre-of-mass energy of pp collisions from 8 to 13 TeV and the higher instantaneous luminosity allow for a more precise test of the Higgs boson properties and the exploration of its rare decay modes, where deviations from the SM might be found. At the same time, it opens the way to direct searches for experimental signatures of BSM physics at the TeV scale.

The work presented in this thesis is situated in this context and aims at the exploration of the EWSB properties and of extensions of the SM via the study of Higgs boson pair (HH, di-Higgs) production via vector boson mechanism. This process allows for probing the very fundamental nature of the Higgs scalar field, since it directly involves the trilinear self-coupling and the quartic coupling of the Higgs boson which, as mentioned above, depends on the shape of the scalar potential itself. The observation of this process thus represents a crucial test of the validity of the SM. Its measurement at the LHC is particularly challenging because of its small cross section [5]. However, the direct relation to the scalar potential makes HH production very sensitive to the presence of BSM contributions, that could manifest either directly as new states decaying to a HH pair (resonant production), or as contributions in the quantum loops that would modify its cross section and kinematic properties (non-resonant production).

The search for HH production discussed in this thesis explores the decay channel where one Higgs boson decays to a b-quark pair and the other to four charged lepton from a ZZ* vector boson pair by using data collected by the CMS experiment during Run II. This is particularly difficult in the dense environment of the LHC, where proton bunches collide at the centre of the CMS detector

---

5 In particle physics, the cross section is used to express the normalized rate or probability of a given particle interaction. It has the dimension of a surface and is usually expressed in barns (b): 1 b = 10^{-2} m^2.

every 25 ns and up to 40 simultaneous interactions on average take place at each crossing. The sensitivity thus crucially depends on the efficiency of the identification of these decays and on the rejection of the abundant jet background. In this exciting exploration, occasionally marked by fluctuations subsequently disproved by the analysis of larger data-sets, no evidence for BSM physics has been found thus far, with no exception for the HH search in the $b\bar{b}qq'4l$ final state. These results however have tried to constrain the parameter space of BSM physics models, as well as anomalous couplings of the Higgs boson.

Thesis outline

The material presented in this thesis is based on work performed within the CMS collaboration and the **I**stituto **N**azionale di **F**isica **N**ucleare (INFN) Bari section, thus relying on the contribution of many people, in particular the Double-Higgs research group. The author's personal contribution to this Di-Higgs search is here briefly summarised.

Chapter 1 and Chapter 2 introduce the reader to the theoretical and experimental frameworks, more specifically discuss the reasons of a search for double Higgs HH production via vector boson fusion mechanism, and describe the experimental apparatus that is used to perform the search, the CMS detector. The subsequent chapters report my personal contribution to this subject and summarize the activity done during the last seven months of my Master's studies in Experimental Particle Physics.

Chapter 3 describes how the informations collected from CMS sub-detectors is translated to physical objects: insights into the algorithms used for the identification of different particles, for the reconstruction of their tracks and for the measurements of their kinematic quantities are given. Among all the detectable particles, a particular attention will be focused on the description of electrons, muons and jets, which are the main characters of the search carried out in this thesis. Besides that, an overview of how physics events are generated and simulated in the detector is provided.

In Chapter 4, the search for VBF $HH \to b\bar{b}4l$ decay is explained in detail. The description of the search strategy is provided along with the analysis steps, from the optimization of the selections applied, to the training of a multivariate discriminator used to better separate the signal-like events from the background. Finally, the interpretation of the resultant expected upper limit is presented.

Part I

THEORETICAL MODELS

Double Higgs boson production

THE Standard Model of particle physics (SM) is a renormalizable QFT that describes the phenomena at the subnuclear scales. It provides a unified description of the strong, weak, and electromagnetic forces, and incorporates a minimal scalar sector that is at the origin of the spontaneous breaking of the electroweak symmetry and of the masses of the fermions. The SM is well corroborated by experimental observations at collider experiments, and received further confirmation with the recent discovery of the Higgs boson in 2012.

Despite the excellent agreement with direct experimental tests performed so far, some observations from the subnuclear to the astrophysical scales, as well as theoretical considerations, suggest that it is incomplete and that a broader theory exists beyond its current formulation. This physics BSM is possibly connected to the scalar sector, and the Higgs boson discovery opens new ways to its exploration. Being intimately related to the nature of the scalar sector, the production of Higgs boson pairs (HH) at the LHC can give invaluable information in this context. It allows for determining the Higgs boson self-interaction, and provides a fertile ground to search for the signs of BSM physics. This chapter discusses the importance of the study of HH production in the context of both SM and BSM physics.

After introducing the SM gauge structure and its scalar sector, with a focus on the Higgs boson properties and couplings, HH production in the SM is reviewed. This process is subsequently discussed in the context of BSM models with non-resonant HH signatures. Finally, the phenomenology at collider experiments and results previously obtained at the LHC are discussed [26, 27].

1.1 The Standard Model of particle physics

The gauge sector and the scalar sector are the foundations of the SM. They are closely interconnected but rely on different theoretical concepts of gauge invariance and spontaneous symmetry breaking. They are discussed separately in what follows, with a focus on the properties of the Higgs boson and a summary of experimental measurements.

1.1.1 Fundamental particles and gauge symmetries

The mathematical formulation of the SM is based on the local gauge invariance of its Lagrangian under the gauge group $SU(3)_C \otimes SU(2)_L \otimes SU(1)_Y$ to explain the strong, weak, and electromagnetic interactions. In particular, the $SU(3)_C$ invariance results in the existence of "gluons" (g) as the mediators of the strong force, which is described by the quantum chromodynamics QCD. The $SU(2)_L \otimes SU(1)_Y$ symmetry explains jointly the weak and electromagnetic forces, mediated by the $W^\pm$ and Z bosons and the photon γ, respectively. The SM formulation does not encompass the gravitational interaction,which is negligible at the subnuclear scales. The main constituents of the SM are summarised in Figure 1.1.

Matter is described in the SM by fermion fields of spin 1/2, whose interactions are mediated by spin-1 boson fields. Experimental observations show that twelve physical fermion fields, six "quark" fields and six "lepton" fields exist. They are organized in three families, made up of two quarks of electric charge $+\frac{2}{3}$ and $-\frac{1}{3}$ and two leptons of electric charge -1 and 0. Fermions in one family and their counterparts in the others have identical properties and only differ in their mass. This is related to their coupling strength to the scalar field, that is described in the next section. To each fermion corresponds an antiparticle with identical properties but opposite quantum numbers.

Quarks

Quarks are subject to all the three forces and, in particular, are the only fermions to possess a "colour" charge to which QCD owes its name. The first family of quarks is composed of the up (u) and down (d) quarks, with a mass of a few MeV/c^2. The former has a positive electric charge of +2/3 while the latter has a negative electric charge of -1/3. Being the lightest quarks (see Figure 1.1), they are stable and compose the ordinary matter. Their counterparts in the second family are the charm (c) and the strange (s) quarks, of masses of about 1.28 GeV/c^2 and 95 MeV/c^2 respectively. Finally, the third family is composed of the top (t) and bottom (b) quarks, which masses are about 173 and 4.2

Standard Model of Elementary Particles

Figure 1.1: A summary table with all the SM particles: the fermions (on the left), which are the matter particles, are divided into quarks (in violet) and leptons (in green), while the bosons (on the right), are classified into gauge bosons, i.e. the carriers of the interactions (in red) and the scalar Higgs boson (in yellow). For each particle, the mass, electric charge and spin values are reported.

GeV/c^2 respectively. Because of the QCD colour confinement properties, quarks do not exist as free states but can be experimentally observed only as bound states. Collider experiments thus detect "mesons", that are composed by a quark-antiquark pair, or "baryons", composed by three quarks. The proton and neutron composing the ordinary matter are stable examples of the latter. Mesons and baryons are collectively denoted as hadrons. The creation of hadrons from a single quark produced in a collider experiment is a complex process that takes the name of "hadronization". As its timescales, related to the QCD energy scales, are of the order of 10^{-24} s, hard scatter and hadronization phenomena can be treated separately thanks to a factorization of their effects. The top quark represents an exception in this sense, as its lifetime is so short ($\sim 0.5 \times 10^{-24}$ s) that it decays before bound states can be formed. Quark flavour is conserved in electromagnetic and strong interactions but not in weak ones, as quark mass eigenstates do not correspond to the weak interaction eigenstates. Their mixing is described by the **C**abibbo **K**obayashi **M**askawa (CKM) matrix.

Leptons

Leptons have no colour charge and are subject only to the electromagnetic and weak forces. The charged leptons of the three families are respectively denoted as the electron (e), muon (μ) and tau lepton (τ). The electron is stable, being the lightest lepton with a mass of 511 keV/c^2. The muon has a mass of 105.7 MeV/c^2 and a lifetime of 2.2 μs, that is sufficiently long to consider it as a stable particle at an LHC experiment given the detector size and the typical muon momentum.

Finally, the tau lepton has a mass of 1.8 GeV/c^2 and a lifetime of 2.9 $\times 10^{-13}$, that is instead short enough for observing it only through its decay products. Most notably, the tau is the only lepton that has a sufficient mass to decay semi-leptonically. To each lepton corresponds a neutrino, respectively denoted as ν_e, ν_μ, and ν_τ. Being electrically neutral, neutrinos interact with the matter only via the weak force and consequently they are not directly detectable at collider experiments. Little is known about their masses, but the observation of their flavour oscillations prove that they are not zero. The mixing of weak and mass eigenstates is represented by the **P**ontecorvo **M**aki **N**akagawa **S**akata (PMNS) matrix.

1.1.2   Quantum Chromodynamics (QCD)

QCD is built on the local gauge invariance under the $SU(3)_C$ group, the subscript denoting the relation to the colour charge arising from this symmetry. The free-field Dirac Lagrangian density of a massless spin-1/2 fermion (the quark fields in this case) is:

$$\mathcal{L} = \overline{\psi}(x)(i\gamma^\mu \partial_\mu)\psi(x) \tag{1.1}$$

where ψ is the fermion field at the space-time coordinate x and γ^μ are the Dirac matrices. The notation $\partial\!\!\!/_\mu \equiv \partial_\mu \gamma^\mu \equiv \partial^\mu \gamma_\mu$ is also used in some case. The following discussion on QCD is valid in presence of a $m\overline{\psi}\psi$ mass term. The reason to consider massless fermions is explained in the context of the electroweak interactions described in the next section. The fermion field transforms in the following way under the SU(3)$_C$ group:

$$\psi(x) \rightarrow e^{ig\frac{\lambda^a}{2}\theta_a(x)}\psi(x) \tag{1.2}$$

where $\lambda^a/2$ are the eight Gell-Mann matrices that generate the group. An important remark is that the derivatives $\partial_\mu \psi(x)$ do not transform in the same way.

Hence, the only way for the Lagrangian density 1.1 to be invariant under the transformation 1.2 is to redefine the derivative ∂_μ as covariant derivatives:

$$D_\mu = \partial_\mu - igA_\mu^a(x)\frac{\lambda^a}{2} \tag{1.3}$$

where the gauge vector fields $A_\mu^a(x)$ correspond to the eight gluons that mediate the strong force. To satisfy the local gauge invariance of the Lagrangian, the gluon fields must transform as:

$$A_\mu^a \rightarrow A_\mu^a + \partial_\mu\theta^a + gf^{abc}A_\mu^c\theta^c \tag{1.4}$$

The f^{abc} symbols denote the structure constants of the group that are defined from the commutation rules $[\frac{\lambda^a}{2}, \frac{\lambda^b}{2}] = if^{abc}\frac{\lambda^c}{2}$. The introduction of the vector fields ensure that the covariant derivative transforms in the same way as the fermion fields and that, consequently, the Lagrangian density is invariant under the local gauge transformation. The Lagrangian density can be completed with a kinetic term for the gluon fields in the form:

$$-\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c \tag{1.5}$$

The complete QCD Lagrangian density is given by:

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu\partial_\mu)\psi(x) - g\overline{\psi}(x)\gamma^\mu\frac{\lambda_a}{2}\psi(x)A_\mu^a - \frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a \tag{1.6}$$

where a summation over all quark fields is implied. The first term is the same as in the original Lagrangian density and represents the free-field propagation of the quark. The second one stems from the introduction of the covariant derivative and represents the interaction of the quark with the vector field A_μ. The strength of the interaction is parametrized by the constant g, usually redefined as the strong coupling constant $\alpha_S = g^2/4\pi$. The third term has been introduced as the kinetic term of the vector field. The generators of the $SU(3)_C$ group do not commute and therefore the structure constants f^{abc} are not all zero. As a consequence, the $gf^{abc}A_\mu^b A_\nu^c$ terms in Equation 1.5, when inserted in the kinetic term of Equation 1.6 , result in cubic and quartic self-interactions of the gluon fields. Such interactions between the force mediators are a general property of non-abelian gauge theories.

Requiring the local gauge invariance led to the introduction of gauge bosons (the gluons) and to the description of their interactions with the fermion fields (the quarks). Choosing the $SU(3)_C$ group implies the presence of eight generators,

the eight gluons, that are mathematically described by the adjoint representation of the group (8), and differ by the colour charge that they carry. Quarks and anti-quarks are instead described in the simplest non-trivial representations of $SU(3)_C$, 3 and $\bar{3}$, which explains the three colour quantum numbers of quarks. It is important to remark that explicit mass terms $A^a_\mu A^\mu_a$ for the gauge bosons in Equation 1.6 would break the gauge invariance of the Lagrangian.

1.1.3 Electroweak interaction

Electroweak interactions are explained in the SM with the same local gauge invariance mechanism as strong interactions by imposing a symmetry under the $SU(2)_L \times U(1)_Y$ group. Experimental observations show that parity is violated by weak interactions, which is accounted for in the theoretical description by assigning different interactions to fermions of opposite chiralities. The left and right chiral components of a field are defined from the $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$ matrix. This is used to define left and right chirality projection operators as $\frac{1-\gamma^5}{2}$ and $\frac{1+\gamma^5}{2}$, respectively. In the limit of a massless particle, the chirality corresponds to the helicity, that is defined as the normalized projection of the spin vector onto the spatial momentum vector. The $SU(2)_L$ gauge group is a non-abelian group to which the weak isospin quantum number (I_3) is associated. The gauge invariance under this group results in the presence of three gauge fields W^i_μ (i = 1, 2, 3). Fermion fields of left chirality are represented by $SU(2)_L$ doublets while fermions of right chirality are $SU(2)_L$ singlets and do not interact with the W^i_μ fields. The $U(1)_Y$ gauge group is abelian and is associated to the weak hypercharge Y. A single gauge field, denoted as B_μ, results from the $U(1)_Y$ local gauge invariance, and interacts with both ψ_L and ψ_R. The $U(1)_Y$ group can be seen in close analogy with the electromagnetic $U(1)_{em}$ group associated to the electric charge Q, but is distinct from it. The Gell-Mann–Nishijima formula determines the relation with the electric charge:

$$Q = I_3 + \frac{Y}{2} \tag{1.7}$$

Fields can therefore be represented as one doublet Ψ_L and two singlets ψ_R, ψ'_R:

$$\Psi_L \equiv \frac{1-\gamma^5}{2}\begin{pmatrix}\psi\\\psi'\end{pmatrix} = \begin{pmatrix}\psi_L\\\psi'_L\end{pmatrix}, \Psi_R \equiv \frac{1+\gamma^5}{2}\begin{pmatrix}\psi\\\psi'\end{pmatrix} = \begin{pmatrix}\psi_R\\\psi'_R\end{pmatrix} \tag{1.8}$$

The fields ψ and ψ' represent either the neutrino and charged lepton fields or the up and down–type quark fields. The two sectors are however separate and neither the strong nor the electroweak interactions can transform quark fields into lepton fields or vice-versa: a direct lepton-quark coupling is not predicted in the SM.

Under this notation, the Lagrangian density can be written as composed of three parts

$$\mathcal{L} = \mathcal{L}_{kin} + \mathcal{L}_{CC} + \mathcal{L}_{NC} \tag{1.9}$$

that explicitly are:

$$\mathcal{L} = i\overline{\Psi}_L \slashed{D} \Psi_L + i\overline{\psi}_R \slashed{D} \psi_R + i\overline{\psi}'_R \slashed{D} \psi'_R \tag{1.10}$$

where the covariant derivative, that is introduced to ensure the gauge invariance, is defined as:

$$D_\mu = \partial_\mu - ig W_\mu^i T_i - ig' \frac{Y_\psi}{2} B^\mu \tag{1.11}$$

where $T_i = \frac{\sigma_i}{2}$ (the Pauli matrices, generators of the $SU(2)_L$ group) for the Ψ_L field and 0 for ψ_R, ψ'_R. The coupling constants g and g' define the strength of the interactions. More interestingly, the total Lagrangian density can be rewritten as:

$$\mathcal{L}_{kin} = i\overline{\Psi}_L \slashed{\partial} \Psi_L + i\overline{\psi}_R \slashed{\partial} \psi_R + i\overline{\psi}'_R \slashed{\partial} \psi'_R$$

$$\mathcal{L}_{CC} = g W_\mu^1 \overline{\Psi}_L \gamma^\mu \frac{\sigma_1}{2} \Psi_L + g W_\mu^2 \overline{\Psi}_L \gamma^\mu \frac{\sigma_2}{2} \Psi_L$$
$$= \frac{g}{\sqrt{2}} W_\mu^+ \overline{\Psi}_L \gamma^\mu \sigma^+ \Psi_L + \frac{g}{\sqrt{2}} W_\mu^- \overline{\Psi}_L \gamma^\mu \sigma^- \Psi_L$$

$$\mathcal{L}_{NC} = \frac{g}{\sqrt{2}} W_\mu^3 (\overline{\psi_L} \gamma_\mu \psi_L - \overline{\psi'_L} \gamma_\mu \psi'_L) + \frac{g'}{2} B_\mu [Y_{\Psi_L} (\overline{\psi_L} \gamma^\mu \psi_L + \overline{\psi'_L} \gamma^\mu \psi'_L)$$
$$+ Y_{\psi_R} (\overline{\psi_R} \gamma^\mu \psi_R) Y_{\psi'_R} (\overline{\psi'_R} \gamma^\mu \psi'_R)]$$

where

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp i W_\mu^2) \sigma_\mu^\pm = \frac{1}{2} (\sigma^1 \pm i\sigma^2) \tag{1.12}$$

There is therefore a charged current interaction that couples the fields ψ_L and ψ'_L and is mediated by the $W_\mu^\pm$ fields, that correspond to the charged weak bosons interacting with fermions. A neutral current interaction also exists, although neither the W_μ^3 nor the B_μ fields can be interpreted as the photon field since they couple to neutral fields. However, we can express them in terms of the physical Z_μ field (the neutral Z boson field) and the A_μ field (the photon field) through a linear superposition parametrized with the Weinberg angle θ_W:

$$B_\mu = A_\mu cos\theta_W - Z_\mu sin\theta_W \tag{1.13}$$

$$W_\mu^3 = A_\mu sin\theta_W + Z_\mu cos\theta_W \tag{1.14}$$

Upon substituting this expression in the expression of $\mathcal{L}_{NC}$, two neutral current interactions appear, the one with the A_μ field being determined by a coupling of strength $gsin\theta_W I_3 + g'cos\theta_W \frac{Y}{2}$. The unification of the weak and electromagnetic forces is completed if I require this expression to be equal to the coupling constant of the photon field eQ. We can arbitrarily set $Y_{\Psi_L} = $ -1 as the hypercharge only appears multiplied by g', and solve the equation by using Q = 0 for the neutrino fields and Q = -1 for the lepton fields. Upon substitution, the following relation between the coupling constants is obtained:

$$gsin\theta_W = g'cos\theta_W = e \tag{1.15}$$

The full electroweak Lagrangian can be expressed in a compact form as:

$$\mathcal{L}_{EWK} = i\overline{\Psi}_L \slashed{D} \Psi_L + i\overline{\psi}_R \slashed{D} \psi_R + \overline{\psi'}_R \slashed{D} \psi'_R - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W^{\mu\nu}_i \tag{1.16}$$

where the field strength tensors are:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.17}$$

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{abc} W^b_\mu W^c_\nu \tag{1.18}$$

The Equation 1.16 contains the free fermion Dirac Lagrangian and the charged and neutral current interactions of the fermions previously discussed.

The kinetic terms of the W^i_μ field, once developed, show a rich structure of self-interactions of the gauge bosons. Trilinear and quadrilinear interactions are predicted by the theory. As already observed for the strong interactions, explicit mass terms of the gauge fields would break the gauge invariance. Direct fermion mass terms are also not allowed, because they are not invariant under the gauge transformation: the left and right chiralities of the fields transform differently under $SU(2)_L \times U(1)_Y$ while the mass can be decomposed as $m\overline{\psi}\psi = m(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R)$. A summary of the fermion fields are summarized under their $SU(2)_L$ representation in Table 1.1. The fermion fields in the SM, quarks and leptons have the same structure under the $SU(2)_L$ group. Left and right chirality fields are respectively a double and a singlet of the $SU(2)_L$ group and, consequently, only the former have a charged weak interaction, that is mediated by the $W_\pm$ bosons. Neutral weak interactions are mediated by the Z bosons, that interact with both chiral components, albeit with a different strength, thanks to the mixing of the gauge fields via the Weinberg angle θ_W. The electromagnetic force, mediated by the photon, is not sensitive to the chirality of the fermion fields and its interaction depends on the charge Q, that is related to the hypercharge Y and the weak isospin I_3. The table shows the different lepton fields according to their flavours and to their chirality. Quark fields exist in three additional types according to

Type	1st gen.	2nd gen.	3st gen.	I_3	Y	Q	SU(3)$_C$
Quarks	$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$ u_R d_R	$\begin{pmatrix} c_L \\ s_L \end{pmatrix}$ c_R s_R	$\begin{pmatrix} t_L \\ b_L \end{pmatrix}$ t_R b_R	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$ 0 0	1/3 4/3 $-2/3$	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$ 2/3 $-1/3$	triplet
Leptons	$\begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix}$ e_R $\nu_{e,R}$	$\begin{pmatrix} \nu_{\mu,L} \\ \mu_L \end{pmatrix}$ μ_R $\nu_{\mu,R}$	$\begin{pmatrix} \nu_{\tau,L} \\ \tau_L \end{pmatrix}$ τ_R $\nu_{\tau,R}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$ 0 0	-1 -2 0	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ 1 0	singlet

Table 1.1: Fermion fields under their SU(2)$_L$ representation. The L and R subscripts denote respectively the left and right chiralities. Quarks and lepton fields are separately listed. They differ for the weak hypercharge Y (and consequently by the electric charge Q = (I_3 + Y/2) quantum numbers, as well as for their colour charge under the SU(3)$_C$ group.

their colour charge, as they represent a triplet of the SU(3)$_C$ group. In contrast, leptons have no colour charge and thus do not interact with the strong force. Interactions can change the quantum numbers of the fields through the charge carried by the mediators. Charged weak interactions change the weak isospin (and consequently the electric charge) and strong interactions change the colour charge of quarks. This formulation of the SM is of extraordinary beauty and elegance. Matter fields are completely described in terms of quantum numbers, and their interactions follow from the application of a symmetry principle to the Lagrangian density. Differences in the representation of the fields under a specific group completely determine the phenomenology that we observe experimentally. The electromagnetic and weak force are jointly described, and all the relevant forces at the subnuclear scale are explained from a common symmetry principle.

However, the theory requires both the fermions and the gauge bosons to be massless, as any explicit mass term would violate the gauge invariance itself. This is in clear contrast with the experimental observation of massive weak bosons and fermions. A simple addition "a posteriori" of the mass terms produces a theory that is not renormalizable, and results in unhphysical predictions for scattering of longitudinally polarized vector bosons. The solution needed to ensure the unitarity of the theory and to explain bosons and fermions masses is provided by the Brout-Englert-Higgs mechanism, a natural way of breaking the SU(2)$_L$ × U(1)$_Y$ symmetry to U(1)$_{em}$ without explicitly violating the local gauge invariance.

1.1.4 The Brout-Englert-Higgs mechanism

The BEH mechanism was proposed in 1964 independently by physicists Englert and Brout, Higgs, and also by Guralnik, Hagen, and Kibble [14–17] as a solution to generate the gauge boson masses and explain the fermion masses. The mechanism is based on the concept of spontaneous symmetry breaking, a phenomenon that is often observed in Nature whenever individual ground states of a system do not satisfy the symmetries of the system itself. A familiar example is a pencil vertically placed on a table, for which the gravitational force is symmetric for rotations around the vertical axis, but the ground state corresponds to a pencil laying horizontally on the table and pointing to a specific direction: the ground state does not satisfy the symmetry of the force acting on the system. The symmetry is said to be "spontaneously broken" because the ground state is not invariant under it, but the ground state chosen is only one out of the continuum of the possible ground states. As these are interrelated by the original symmetry of the system, the latter is "hidden" among these multiple choices. In the BEH mechanism, spontaneous symmetry breaking is realized through the introduction of a complex scalar doublet (under $SU(2)_L$) of fields:

$$\Phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix} \tag{1.19}$$

An important remark is that the field must be scalar to satisfy space isotropy, otherwise the expectation value on the vacuum would be frame-dependent. Moreover, the expectation value on the vacuum must be constant to satisfy space homogeneity. The field has an hypercharge $Y = 1$ and thus its covariant derivative is:

$$D_\mu = \partial_\mu - igW_\mu^i \frac{\sigma_i}{2} - \frac{1}{2}ig'B_\mu \tag{1.20}$$

The BEH lagrangian can consequently be written as:

$$\mathcal{L}_{BEH} = (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi^\dagger\Phi) \tag{1.21}$$

where the "mexican hat" potential $V(\Phi^\dagger\Phi)$ is defined as:

$$V(\Phi^\dagger\Phi) = -\mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2 \tag{1.22}$$

with $\mu^2, \lambda > 0$. All the doublets that satisfy the condition:

$$|\Phi^2| = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.23}$$

are minima of these potential, and are connected through gauge transformations that change the phase of the field Φ but not its modulus. The quantity v is called the **V**acuum **E**xpectation **V**alue (VEV) of the scalar potential. Once a specific ground state is chosen, the symmetry is explicitly broken but the Lagrangian is still gauge invariant with all the important consequences for the existence of gauge interactions. If the symmetry is spontaneously broken to the ground state that is parallel to the ϕ^0 component of the doublet, it can be shown that this specific ground state is still invariant under the $U(1)_{em}$ symmetry group. As a consequence, the field expansion around this minimum is written as:

$$\Phi(x) = \frac{1}{\sqrt{2}} e^{\frac{i\sigma_i \theta^i(x)}{v}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \qquad (1.24)$$

This corresponds to the presence of a scalar real massive field H and of three massless fields θ^i. The latter are expected as consequence of the Goldstone theorem [28], that states that the spontaneous breaking of a continuous symmetry generates as many massless bosons (the Goldstone boson) as broken generators of the symmetry. However, such massless bosons are not observed in Nature. They can be removed with an $SU(2)_L$ transformation that consists in the choice of a specific gauge called "unitary gauge":

$$\Phi(x) \to \Phi'(x) = e^{-\frac{i\sigma_i \theta^i(x)}{v}} \Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \qquad (1.25)$$

After this transformation, only the real scalar field H(x) remains and its quanta correspond to a new physical massive particle, the Higgs boson (H). Upon substitution of the covariant derivative expression and that of $\Phi(x)$ in the unitary gauge, the BEH Lagrangian reads:

$$\mathcal{L}_{BEH} = \frac{1}{2}\partial^\mu H \partial_\mu H - \frac{1}{2}(2\lambda v^2)H^2 \qquad (1.26)$$

$$+ [(\frac{gv}{2})^2 W^{\mu+}W^-_\mu + \frac{1}{2}\frac{(g^2 + g'^2)v^2}{4} Z^\mu Z_\mu](1 + \frac{H}{v})^2 \qquad (1.27)$$

$$- \lambda v H^3 - \frac{\lambda}{4}H^4 + \frac{\lambda}{4}v^4 \qquad (1.28)$$

The first line represents the evolution of the scalar Higgs field, that has a mass $m_H^2 = 2\lambda v^2 = 2\mu^2$. It is a free parameter of the theory, directly related to the parameter μ of the scalar potential. The second line represents the mass terms of the weak bosons (those that multiply the constant term), of mass:

$$m_W^2 = \frac{g^2 v^2}{4} \qquad (1.29)$$

$$m_Z^2 = \frac{(g^2 + g'^2)v^2}{4} = \frac{m_W^2}{cos^2\theta_W} \qquad (1.30)$$

It can be observed that the Goldstone bosons, removed with the unitary transformation, are absorbed as additional degrees of freedom of the $W^\pm$ and Z bosons, corresponding to their longitudinal polarizations: the mechanism gives mass to the weak bosons. The second line of Equation 1.26 also describes the interactions of the weak bosons with the Higgs field. There are a HWW and a HZZ interactions from the $2H/v$ term and a HHWW and a HHZZ interaction from the H^2/v^2 term. The third line shows that cubic and quartic self-interactions of the Higgs boson are predicted. The BEH potential can be rewritten in terms of a trilinear and a quadrilinear coupling as:

$$V(H) = \frac{1}{2}m_H^2 H^2 + \lambda_{HHH} v H^3 + \frac{1}{4}\lambda_{HHHH} H^4 - \frac{\lambda}{4}v^4 \qquad (1.31)$$

with the self-coupling constants defined as:

$$\lambda_{HHH} = \lambda_{HHHH} = \lambda = \frac{m_H^2}{2v^2} \qquad (1.32)$$

An important remark is that both Higgs boson self-couplings are directly related to the parameters of the scalar potential and are entirely determined from the Higgs boson mass and the VEV. Their measurement thus represents a test of the validity and coherence of the SM. In a wider perspective, the Higgs boson self-couplings have no equal in the SM: in contrast to the weak boson self-interactions, that have a gauge nature, the Higgs boson self-interactions are purely related to the scalar sector of the theory and they are responsible for the mass of the Higgs boson itself.

Their experimental determination is thus crucial to reconstruct the Higgs boson potential and explore the nature of the EWSB. Finally, there is a constant term in the Lagrangian density of BEH . While this is irrelevant in the SM, it contributes to the vacuum energy, which is related to the cosmological constant that determines the curvature of the Universe. The value of this constant predicted in the SM is not compatible with astronomical observations. This is a puzzle that requires either a proper quantum theory of gravity with additional interactions or a mechanism to reduce the Higgs field vacuum energy density. There are at this point two free parameters of the BEH mechanism: the VEV v and the Higgs boson mass m_H. The first corresponds to the energy scale of the electroweak symmetry breaking and can be computed from the Fermi constant G_F that is precisely determined from the muon lifetime:

$$\frac{G_F}{\sqrt{2}} = (\frac{g}{2\sqrt{2}})^2 \frac{1}{m_W^2} \Rightarrow v = \sqrt{\frac{1}{\sqrt{2}G_F}} \sim 246 GeV \qquad (1.33)$$

Fermions have been assumed to massless until now. Mass terms are generated by the Higgs field itself through a Yukawa interaction that couples the left and right chiral fields. Denoting as ψ and ψ' the up ($I_3 = +1/2$) and down-type ($I_3 = -1/2$) fermions, the Yukawa Lagrangian density is:

$$\mathcal{L}_{Yukawa} = -y_{f'}(\overline{\Psi}_L \Phi \psi'_R + \overline{\psi'}_R \Phi^\dagger \Psi_L) - y_f(\overline{\Psi_L}\tilde{\Phi}\psi_R + \overline{\psi}_R \tilde{\Phi}^\dagger \Psi_L) \qquad (1.34)$$

with

$$\tilde{\Phi} = i\sigma_2 \Phi^* = \begin{pmatrix} \phi_0^* \\ -\phi_+^* \end{pmatrix} \xrightarrow{EWSB} \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} \qquad (1.35)$$

The Yukawa Lagrangian density can be generalized to include fermion mixing through combinations of the mass eigenstates of the fields Ψ_L, ψ_R and ψ'_R. It is Lorentz and gauge invariant and renormalizable, so that it can be included in the SM Lagrangian density. After EWSB, the Yukawa lagrangian density term becomes:

$$\mathcal{L}_{Yukawa} = -\sum_f m_f(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L)(1 + \frac{H}{v}) \qquad (1.36)$$

where the sum runs on both up– and down–type fermions and the mass terms

$$m_{f'} = y_{f'}\frac{v}{\sqrt{2}} \qquad (1.37)$$

Fermion masses are thus explained in the SM as the interaction of the fermion fields with the Higgs field, which changes the chirality of the fermions. The strengths of the interactions are directly related to the fermion masses, and are free parameters of the theory. An important remark is that the SM does not explain the origin of these couplings and, consequently, the hierarchy of the three fermion families. In conclusion, the BEH mechanism solves the aforementioned problems of the electroweak theory of the SM. Upon breaking the electroweak symmetry, the scalar field generates Goldstone bosons that are absorbed as degree of freedoms of the vector boson fields, which become massive. The Higgs boson contributions to the quantum loops in the scattering of longitudinally polarized vector bosons regularizes the process and ensure its unitarity at the TeV scale and beyond. Finally, the Higgs boson couples the left and right chiral components of the fermion fields in a Yukawa interaction, determining the fermion masses with a purely quantum-relativistic mechanism.

After having provided an overview of the two fundamental theories of the SM (QCD and **E**lectro**W**eak (EW) theory) and the spontaneous symmetry breaking

mechanism that gives mass to the particles, it is presented the Standard Model Lagrangian:

$$\mathcal{L}_{SM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}W_{\mu\nu}W^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}$$
$$+\overline{\psi}i\gamma^{\lambda}D_{\lambda}\psi + (D_{\mu}\Phi)^{\dagger}(D^{\mu}\Phi) - V(\Phi^{\dagger}\Phi)$$
$$+\mathcal{L}_{Yukawa} + h.c.$$

where the covariant derivative D_{μ} can be retrieved from Equation 1.3, 1.11,1.20. Finally, the theory obtained by incorporating the BEH mechanism in the electroweak theory is renormalizable, as demonstrated by 't Hooft and Veltman [20].

1.2 The Higgs boson phenomenology and experimental status

Experimental confirmation of the BEH mechanism came in July 2012 with the discovery of a new scalar boson of a mass of approximately 125 GeV announced by the ATLAS and CMS Collaborations [24, 25].

The discovery was performed in the data collected at $\sqrt{s} = 8$ TeV (Run I) and lead by the high resolution H $\rightarrow$ ZZ* $\rightarrow l^{+}l^{-}l'^{+}l'^{-}(l, l' = e, \mu)$ and H $\rightarrow \gamma\gamma$ decay channels. The existence of this scalar particle is now firmly established and further confirmed with the data collected at $\sqrt{s} = 13$ TeV (Run II), as shown in Figure 1.2 and Figure 1.3. With the observation of the higgs boson and the measurement of its mass, the last important parameter of the SM has been determined. However, the observation of this new particle only represents the first step in the exploration of the EWSB. The efforts of the experimental community are thus focusing on the characterization of this boson. The measurements of its couplings via exclusive production modes and decay channels, of its spin-parity, and of its differential production cross section, need to be thoroughly investigated to verify that they correspond precisely to the SM predictions. Several mechanisms contribute to the production of a Higgs boson at a pp collider. The dominant one, with a cross section of about 49 pb, is the gluon fusion (ggF) production, that proceeds through a heavy quark loop. The second most frequent mechanisms, about 10 times rarer than ggF, is vector boson fusion VBF, where the Higgs boson is produced in association with a jet pair of large invariant mass. The third main mechanism is the production in association with a single vector boson (VH, V $= W^{\pm}$, Z). Studying the rare VH and VBF production mechanisms allows for probing the Higgs boson coupling to vector bosons. Finally, Higgs bosons can be produced in association with a pair of top quark ($t\bar{t}H$) or a
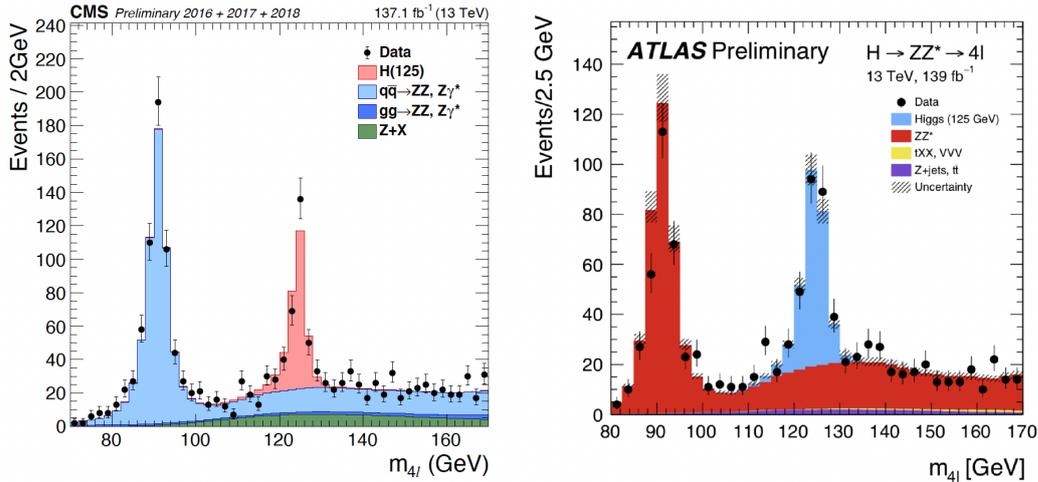
Figure 1.2: Distribution of the reconstructed four-lepton invariant mass m_{4l} in the low-mass range, with full Run II data, in CMS [29] (on the left) and ATLAS [30] (on the right) experiments. Points with error bars represent the data and stacked histograms represent expected distributions of the signal and background processes.

single top quark (tH). The $t\bar{t}H$ mechanism is of particular interest as it allows for the direct determination of the magnitude of the top Yukawa coupling y_t, in contrast with the indirect determination from ggF. The tH mechanism, about a factor of 10 rarer than the previous, allows for the determination of the sign of y_t. The cross sections of these production modes as a function of $\sqrt{s}$ are summarized in Figure 1.5a . The Run I Higgs boson discovery was performed inclusively for all the production mechanisms. The combination of the high-resolution channels between the ATLAS and CMS experiments resulted in a precise determination of m$_H$ of:

$$m_H = 125.09 \pm 0.21(stat.) \pm 0.11(syst.)GeV \tag{1.38}$$

Exclusive production modes are now being probed and, in particular,the $t\bar{t}H$ production has been observed [34]. Higgs boson decays branching fractions are shown in Figure 1.5b as a function of m_H, and are summarized in Table 1.2 for a Higgs boson of mass $m_H = 125.09$ GeV. The H $\to ZZ^*$ and $H \to \gamma\gamma$ decay modes are among the rarest but are experimentally advantageous because of the high signal-to-background (S/B) ratio and the excellent invariant mass resolution. The mass m$_H$ ~125 GeV turns out to be one of the most difficult values to reach experimentally, as all masses below about 120 GeV and above 130 GeV had been excluded with 2011 data [35]. However, after the Higgs boson discovery,
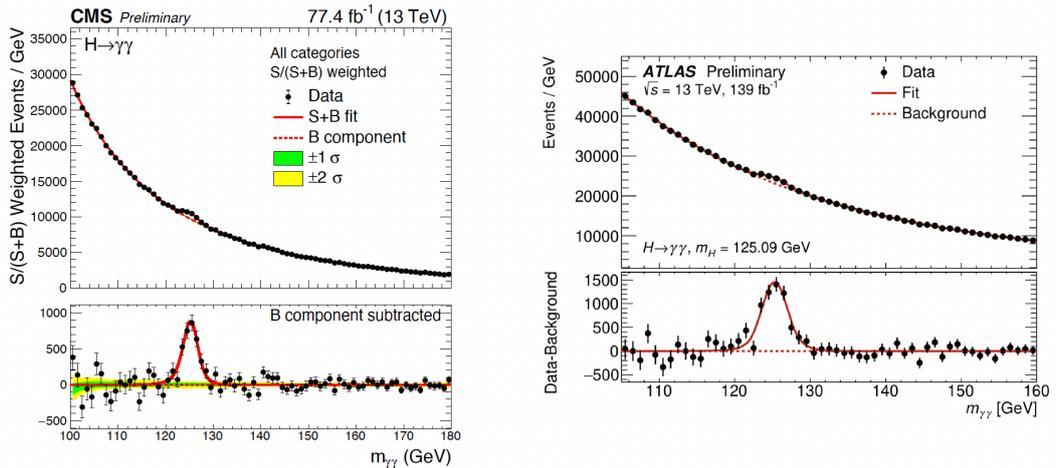
Figure 1.3: Distribution of the reconstructed diphoton invariant mass $m_{\gamma\gamma}$ in the low-mass range, with 2016 and 2017 data collected by CMS [31] (on the left) and the full Run II data collected by ATLAS [32] (on the right). The black dots represent the data.

this specific mass value provides several final states that can be simultaneously probed. Decays of the Higgs boson to $W^{\pm}W^{\mp*}$ or $Z\gamma$, as well as to fermions in the $H \to \tau^+\tau^-$, $H \to b\bar{b}$, $H \to \mu^+\mu^-$ and $H \to c\bar{c}$ decay channels, can thus be eventually probed at the LHC. The latter decay channel is challenging because of the low S/B ratio and the tiny branching fractions, but of particular interest as it represents the only direct way to probe Higgs boson decays to up-type fermions. The Higgs boson was also shown to have a spin-parity $J^P = 0^+$ [36], and the combination of several decay channels indicated that its decay rates and coupling strengths are compatible with the SM expectation [37]. The combined measurement performed by the ATLAS and CMS experiments [38] further confirmed the agreement with the SM predictions. The exploration of the Higgs boson properties with Run II data continues to reveal important information about this particle.

The existence of decays to fermions has been firstly established in the $\tau^-\tau^+$ decay mode by the CMS experiment [39], complementing the previous Run I observation from the combination of the two experiments [38]. Further incredible studies on the couplings to fermions are still going on and recently had brought to the evidence of the decay of the Higgs boson to muon pairs [40] as shown in Figure 1.6. A summary of the measurements of the Higgs mass, carried out by ATLAS and CMS experiments in the last years, is shown in Figure 1.4. The investigation of the Higgs boson properties is showing good agreement with the SM predictions. Although constraints on its couplings partially depend
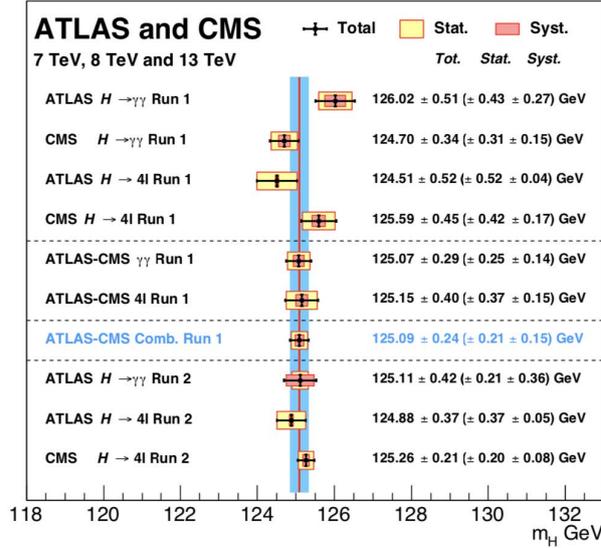
Figure 1.4: Summary of the CMS and ATLAS mass measurements in the $\gamma\gamma$ and ZZ channels in pp collisions during Run I and Run II [33].
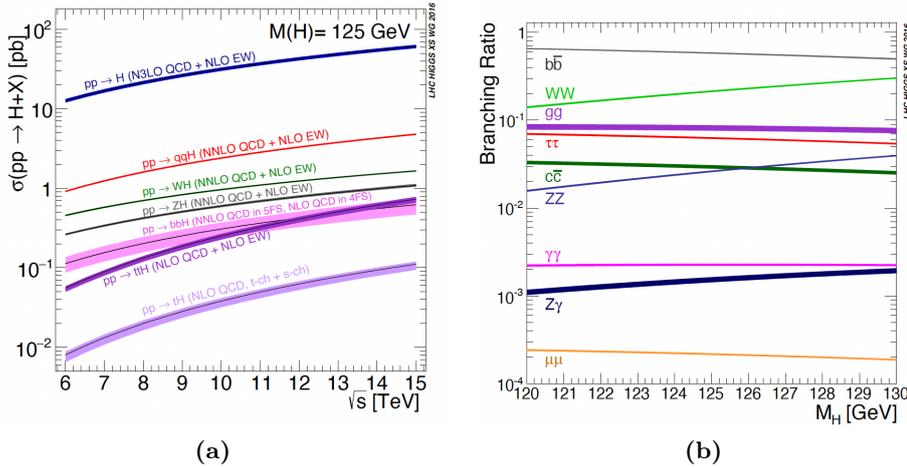
Figure 1.5: Figure 1.5a Higgs boson production cross section as a function of $\sqrt{s}$ for different production mechanisms. The theoretical uncertainties are indicated as bands. Figure 1.5b Branching fractions of the decay of a Higgs boson as a function of m_H. The theoretical uncertainties are indicated as bands.

on assumptions on BSM contributions and still leave some space for possible deviations, the scalar boson discovered almost ten years ago is today precisely known and is compatible within the uncertainties with a SM Higgs boson. Most importantly, it has been experimentally observed that this particle breaks the degeneracy between the three fermion families by coupling proportionally to their mass. As summarized in Figure 1.6, couplings of the Higgs boson are probed over about four orders of magnitude and the dependence of their strength on

Decay mode	$\mathcal{B}$ [%]
$H \to b\bar{b}$	$58.09^{+0.72}_{-0.73}$
$H \to W^{\pm}W^{\mp *}$	$21.52^{+0.33}_{-0.33}$
$H \to gg$	$8.18^{+0.42}_{-0.42}$
$H \to \tau^{+}\tau^{-}$	$6.27^{+0.10}_{-0.10}$
$H \to c\bar{c}$	$2.88^{+0.16}_{-0.06}$
$H \to ZZ^{*}$	$2.641^{+0.040}_{-0.040}$
$H \to \gamma\gamma$	$0.2270^{+0.0047}_{-0.0047}$
$H \to Z\gamma$	$0.1541^{+0.0090}_{-0.0090}$
$H \to \mu^{+}\mu^{-}$	$0.02171^{+0.00036}_{-0.00037}$

Table 1.2: Branching fractions of the main Higgs boson decay modes for a SM Higgs boson of mass $m_H = 125.09$ GeV. Theoretical uncertainties combine the uncertainties on the Higgs boson partial width, on the value of α_S, and on the quark masses.

the boson and fermion masses is established. However, one important element is missing in this figure: the Higgs boson itself. With its mass now known with precision, the value of its self-coupling can be computed from Equation 1.32 to be $\lambda_{HHH} \sim 0.13$, completely determined in the SM from m_H and v. Experimentally measuring λ_{HHH} would allow to verify if this coupling really fits in the global scheme illustrated in Figure 1.6 at an ordinate of $\sqrt{2\lambda_{HHH}} = m_H/v \sim 0.51$, providing a test of the validity of the SM. As this coupling is responsible for the Higgs boson mass itself, it is related to the very fundamental properties of the EWSB and of the BEH mechanism. The λ_{HHH} coupling can be directly probed in Higgs boson pair (HH) production. Similarly, the measurement of the quadrilinear coupling λ_{HHHH}, a further probe of the BEH potential, requires the study of triple Higgs final state. The production of the latter is however extremely rare in the SM, with a cross section of about 80 ab at $\sqrt{s} = 14$ TeV [41], out of the experimental reach of the LHC.

In contrast, HH production, although challenging, can be experimentally probed at the LHC. It has also been recently suggested that the trilinear coupling HHH could be determined from precision measurements at the LHC. Its effects could be observed either in electroweak precision observables [42], or from precision measurements of single Higgs boson production, where the radiative corrections due to the trilinear coupling can be sizeable [43]. In the latter case, significant information can be extracted from both the total cross section and the differential Higgs boson p_T distribution [44, 45]. However, as λ_{HHH} is completely determined from m_H and the vacuum expectation value v, these indirect constraints need to
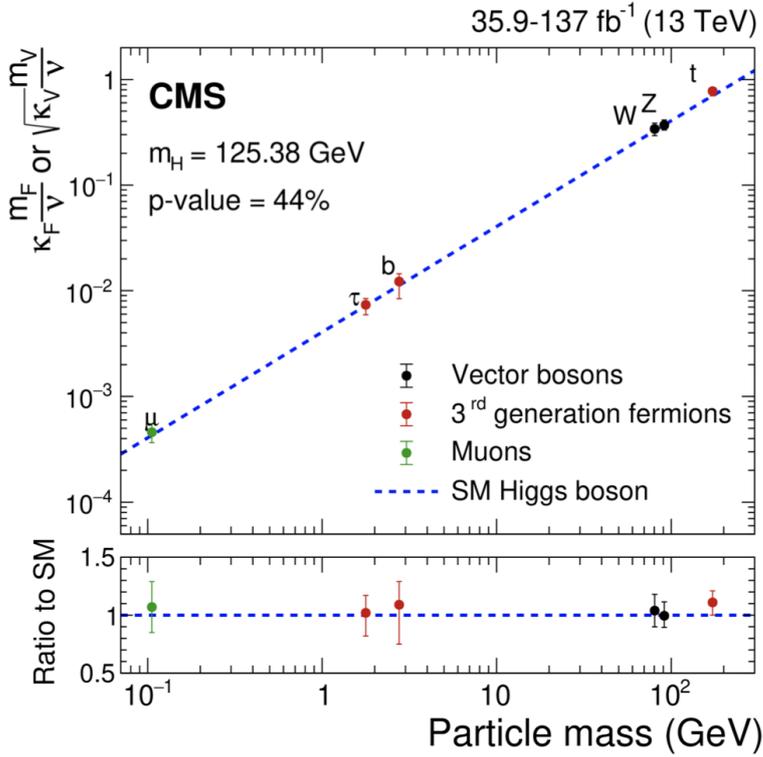
Figure 1.6: A summary of the CMS measurements of the Higgs boson couplings to the other fundamental particles, with the predictions by the standard model indicated by the dashed black line. In the lower panel, the ratio between the measured coupling and the standard model prediction is shown. This new CMS result presents the first measurement of the Higgs boson coupling to the muon, indicated by the left-most data point [40].

assume a variation of the trilinear coupling which often depends on the theoretical assumptions through which this is realized. The direct determination of λ_{HHH} from HH production is thus an essential step in the understanding of the BEH mechanism and, for this reasons, it represents one of the main goals of the LHC physics programme.

1.3 Higgs boson pair production

It is known since a long time that λ_{HHH} can be extracted from the measurement of the Higgs boson pair production cross section. The role of the trilinear coupling in this context was highlighted back in 1988, well before the Higgs boson discovery, when the first computation of the cross section was performed [46].

However, the λ_{HHH} coupling represents only one of the possible interactions that contribute to HH production. In general, a pair of on-shell Higgs bosons can be produced in the final state of a collision through any of the diagrams in Figure 1.9. The production mechanisms that are detailed in this section are
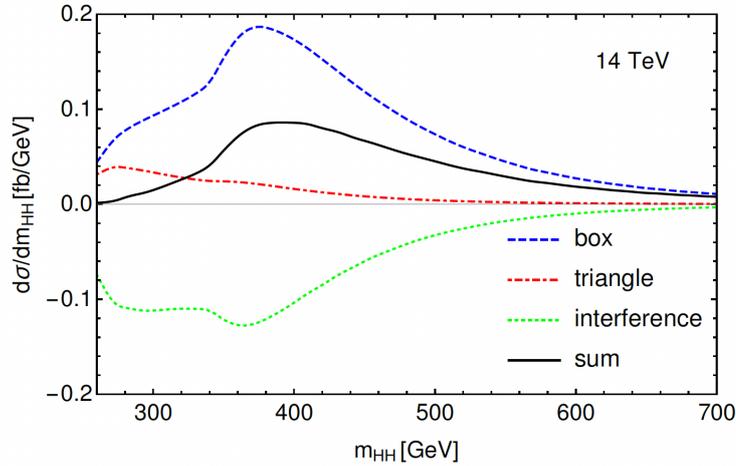
Figure 1.7: Higgs pair invariant mass distribution at leading order for the different contributions to the gluon fusion production mechanismand their interference [26].

characterized by different combinations of these interactions, in which the λ_{HHH} contribution must be disentangled from other effects.

1.3.1 The HH pair production mechanisms

At the LHC, Higgs boson pairs can be produced through the five main mechanisms that are listed below in decreasing order of their cross section. Some representative Feynman diagrams illustrate the Higgs boson couplings involved [47, 48]:

- **Gluon fusion** $gg \rightarrow HH$. It involves either the production of a Higgs boson pair through the trilinear Higgs boson self-coupling (triangle-type contributions), or the radiation of two on-shell Higgs bosons from a heavy quark loop (box-type contributions), as shown in the Feynman diagrams in Figure 1.9a [49]. The cross section consequently depends on λ_{HHH} and on the top quark Yukawa couplings y_t. The relative contribution of these two different pieces, as well as their interference, can be observed in the Higgs pair invariant mass distribution shown in Figure 1.7. The effect of the trilinear Higgs self-coupling in the **L**eading **O**rder (LO) total cross section amounts to a reduction of about 50% with respect to the box-only contribution, due to the large destructive interference. The QCD corrections are known up to **N**ext to **L**eading **O**rder (NLO), and at **N**ext to **N**ext to

Leading Order (NNLO) in the limit of heavy top quarks [50, 51] including partial finite top quark mass effects.

- **The vector-boson fusion (VBF)** $qq' \to HHjj$. It is the second-largest production mechanism at LHC, and it is dominated by t-channel W and Z exchange in analogy to single Higgs production. It involves continuum diagrams originating from two Higgs radiations off the virtual W or Z bosons, and diagrams in which a single Higgs boson (off-shell) splits into a Higgs pair as shown in Figure 1.9b (the key coupling dependency is highlighted for each diagram). In addition to the Higgs boson trilinear coupling, this process, also depends on the quadrilinear coupling of a Higgs boson pair to a vector boson pair as well as on the single Higgs boson coupling to vector boson. Despite its cross section being one order of magnitude smaller than the gluon fusion one, the two final state jets provide a clean signature that can be used to discriminate signal events from background. The QCD corrections are only known in the structure-function approach, i.e. where only the t-channel W and Z exchange is taken into account and interference effects for external quarks of the same flavour are neglected. This approximation is valid at the level of a percent similar to the single Higgs case.

- **Top quark pair associated production / Double Higgs bremsstrahlung off top quarks.** $qq'/gg \to t\bar{t}HH$. It is a HH variant of the single Higgs boson pair production in association to a top quark pair ($t\bar{t}H$), where either two Higgs bosons are radiated from the top quarks or are produced from the Higgs boson self-coupling as illustrated in Figure 1.9c. Its cross section exceeds the one from VBF HH production at high transverse momenta of the HH pair and for high centre-of-mass energies. As shown in Figure 1.8 it reaches a cross section value close to the vector-boson fusion cross section at a 100 TeV hadron collider.

- **Double Higgs-strahlung/Vector boson associated production** $qq' \to VHH$ with $V = W^{\pm}, Z$ as shown in Figure 1.9d. It involves the same Higgs boson couplings as VBF production, but an on-shell vector boson is present in the final state. Its production rate is significantly lower than VBF's one.

- **Single top quark associated production** $qq' \to tjHH$. It can proceed through either the t- or s-channel, that are respectively illustrated in the top and bottom row of the diagrams in Figure 1.9e. The t-channel diagrams are illustrated for simplicity in the so-called 5F scheme [52]. It is the only process that is sensitive at the same time to the HH couplings to vector bosons and to top quarks and to their relative phase. However, its cross

section is so small that it can hardly be investigated at the LHC, but could be studied in a future higher energy collider.

The cross sections of these production mechanisms at different centre-of-mass energies are summarized in Table 1.3. The cross section for gluon fusion is computed at the NNLO of the theoretical perturbative QCD calculation, including next-to-next-to-leading-logarithm (NNLL) corrections and finite top quark mass effects at NLO.

The theoretical uncertainties include uncertainties in the QCD factorization and renormalization scales, α_S , **P**arton **D**istribution **F**unction (PDF) [1], and unknown effects from the finite top quark mass at NNLO. The cross sections for VHH are computed at the NNLO and those of the other processes at the NLO of the perturbative QCD calculation. HH production is in general very rare at the LHC. As a consequence, experimental searches,including the one presented in this thesis, focus on the first two dominant production mechanisms, the gluon fusion and VBF HH production. The last one, subject of the studies of this work, provide additional handles for the measurement of the λ_{HHH} and give access to the VVHH interaction that only in the last two years is getting explored. The destructively interference of the three VBF production diagrams combined with the restricted phase space of production of two Higgs bosons, makes the HH production extremely sensitive to BSM physics. BSM physics contributions might alter the destructive interference and produce large modifications that can be probed with the current LHC data. In this sense, this research work looks at the HH production as a test of the SM and a probe of BSM physics which is the topic of the next section.

1.4 Beyond the Standard Model (BSM)

Theoretical considerations and experimental results indicate that the SM is incomplete. Once compared to astrophysical observations and combined with cosmological models, the SM cannot provide a suitable mechanism that is responsible for the matter-antimatter asymmetry in the Universe, nor it predicts the existence of a particle species that is compatible with the dark matter, which is observed from its gravitational effects. The SM is also unsatisfactory under three main theoretical aspects deeply related to Higgs boson physics.

First, it cannot currently provide an explanation for the existence of three families of fermions, identical under all aspects but for their couplings with the

---

1 Parton distribution functions give the probability to find partons (quarks and gluons) in a hadron as a function of the fraction x of the proton's momentum carried by the parton. They are conventionally defined in terms of matrix elements of certain operators. For more details, see Section 3.1.

Production mode	$\sigma\,[\text{fb}]$
	$\sqrt{s} = 8\,\text{TeV}$
Gluon fusion	$10.15^{+4.1\%}_{-5.7\%}\ (\text{scale}) \pm 3.1\%(\text{PDF}) \pm 2.6\%(\alpha_{\text{s}}) \pm 5.0\%(\text{top})$
VBF	$0.459^{+3.2\%}_{-3.6\%}\ (\text{scale}) \pm 2.6\%(\text{PDF} + \alpha_{\text{s}})$
$t\bar{t}$HH	$0.174^{+2.8\%}_{-10.6\%}(\text{scale}) \pm 3.9\%(\text{PDF} + \alpha_{\text{s}})$
W^+HH	$0.145^{+0.43\%}_{-0.52\%}(\text{scale}) \pm 2.8\%(\text{PDF} + \alpha_{\text{s}})$
W^-HH	$0.0677^{+1.0\%}_{-1.2\%}\ (\text{scale}) \pm 3.7\%(\text{PDF} + \alpha_{\text{s}})$
ZHH	$0.143^{+2.7\%}_{-2.3\%}\ (\text{scale}) \pm 2.6\%(\text{PDF} + \alpha_{\text{s}})$
tjHH	$0.00540^{+5.4\%}_{-3.1\%}\ (\text{scale}) \pm 5.6\%(\text{PDF} + \alpha_{\text{s}})$
	$\sqrt{s} = 13\,\text{TeV}$
Gluon fusion	$33.49^{+4.3\%}_{-6.0\%}\ (\text{scale}) \pm 2.1\%(\text{PDF}) \pm 2.3\%(\alpha_{\text{s}}) \pm 5.0\%(\text{top})$
VBF	$1.62^{+2.3\%}_{-2.7\%}\ (\text{scale}) \pm 2.3\%(\text{PDF} + \alpha_{\text{s}})$
$t\bar{t}$HH	$0.772^{+1.7\%}_{-4.5\%}\ (\text{scale}) \pm 3.2\%(\text{PDF} + \alpha_{\text{s}})$
W^+HH	$0.329^{+0.32\%}_{-0.41\%}(\text{scale}) \pm 2.2\%(\text{PDF} + \alpha_{\text{s}})$
W^-HH	$0.173^{+1.2\%}_{-1.3\%}\ (\text{scale}) \pm 2.8\%(\text{PDF} + \alpha_{\text{s}})$
ZHH	$0.362^{+3.4\%}_{-2.6\%}\ (\text{scale}) \pm 1.9\%(\text{PDF} + \alpha_{\text{s}})$
tjHH	$0.0281^{+5.2\%}_{-3.2\%}\ (\text{scale}) \pm 4.5\%(\text{PDF} + \alpha_{\text{s}})$
	$\sqrt{s} = 14\,\text{TeV}$
Gluon fusion	$39.59^{+4.4\%}_{-6.0\%}\ (\text{scale}) \pm 2.1\%(\text{PDF}) \pm 2.2\%(\alpha_{\text{s}}) \pm 5.0\%(\text{top})$
VBF	$1.95^{+1.8\%}_{-2.3\%}\ (\text{scale}) \pm 2.4\%(\text{PDF} + \alpha_{\text{s}})$
$t\bar{t}$HH	$0.949^{+1.8\%}_{-4.8\%}\ (\text{scale}) \pm 3.2\%(\text{PDF} + \alpha_{\text{s}})$
W^+HH	$0.368^{+0.33\%}_{-0.39\%}(\text{scale}) \pm 2.1\%(\text{PDF} + \alpha_{\text{s}})$
W^-HH	$0.197^{+1.2\%}_{-1.3\%}\ (\text{scale}) \pm 2.7\%(\text{PDF} + \alpha_{\text{s}})$
ZHH	$0.414^{+3.5\%}_{-2.7\%}\ (\text{scale}) \pm 1.8\%(\text{PDF} + \alpha_{\text{s}})$
tjHH	$0.0364^{+3.7\%}_{-1.3\%}\ (\text{scale}) \pm 4.7\%(\text{PDF} + \alpha_{\text{s}})$

Table 1.3: Cross section for different HH production modes assuming a Higgs boson mass of 125.09 GeV. The gluon fusion cross section is computed at NNLO of perturbative QCD calculation, with NNLO corrections and finite top quark mass effects at NLO. The cross section of the VHH,V = W$^{\pm}$, Z are computed at NNLO QCD and those of the other processes at NLO QCD. The values are taken from Ref. [41].

Higgs boson, that span over several orders of magnitude. Secondly, the mass of the Higgs boson is not protected by any fundamental symmetry of the SM and it is subject to quadratically divergent radiative corrections, that call for a mechanism to stabilize them to avoid an unnatural fine tuning. Finally, the validity of the theory up to the Planck scale can only be ensured if the scalar potential is bounded from below, which guarantees the stability of the vacuum. From the values of m_H and m_t presently measured, a meta-stability condition of the scalar potential appears to be favoured [53], challenging the long term existence of the electroweak vacuum. The stability of the Higgs potential at higher energy scales is also related to its possible role in the inflation of the primordial
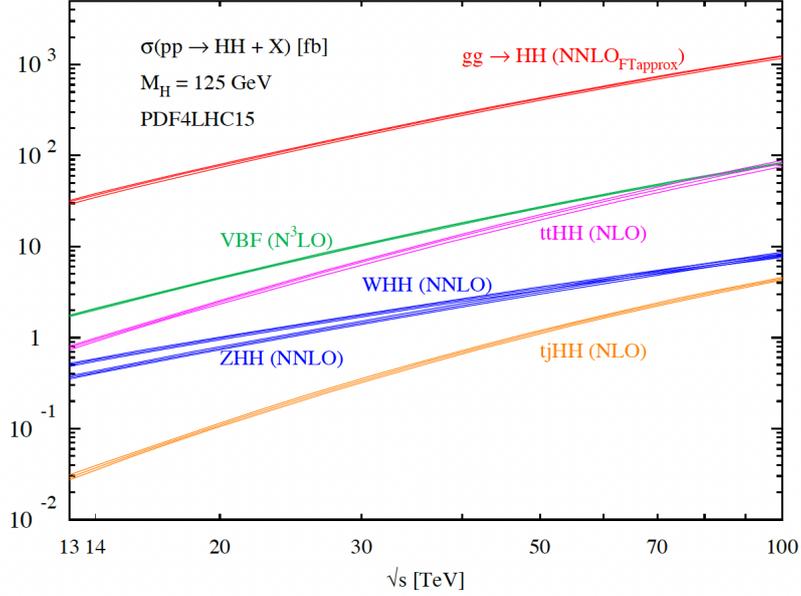
Figure 1.8: Total production cross sections for Higgs pairs within the SM via gluon fusion, vector-boson fusion, double Higgs-strahlung and double Higgs bremsstrahlung off top quarks. PDF4LHC15 parton distribution densities have been used and the size of the bands shows the total uncertainties originating from the renormalization and factorisation scale dependence and the PDF+α_S uncertainties [26].

Universe [54, 55]. Both the vacuum stability and the role of the Higgs as an inflaton depend on the shape of the scalar potential, that is determined from the running value of the BEH potential parameter λ. This parameter is deeply connected to the Higgs boson self-couplings which are one of the main topics investigated in this thesis. It is natural in this context to think that the SM is only the manifestation of a more extended theory beyond it, that exists below the Planck scale and regulates the problems of the SM. The presence of BSM physics could provide a solution to these problems by changing profoundly the structure of the SM while preserving its incredible success at describing the phenomenology of collider experiments until now. In this context, HH production is both a probe for BSM physics, and a balance to discriminate between possible alternatives. If the scale of BSM physics is at the LHC reach, new heavy states can be directly produced (in resolved - low-medium mass resonances 250 GeV $< m_X <$ 1500 GeV or boosted- high mass resonances 900 GeV $< m_X <$ 3000 GeV modes) and subsequently decay to a HH pair. The experimental signature of this **resonant production** mechanism is an enhancement of σ_{HH} at a specific value of m_{HH}, corresponding to mass of the resonance. Despite the nature and the mass of these hypothetical particles, in many of these theoretical models, their presence modifies the Higgs boson's couplings since they contribute to HH production through real and virtual processes. A sizeable effect is expected for several theoretical
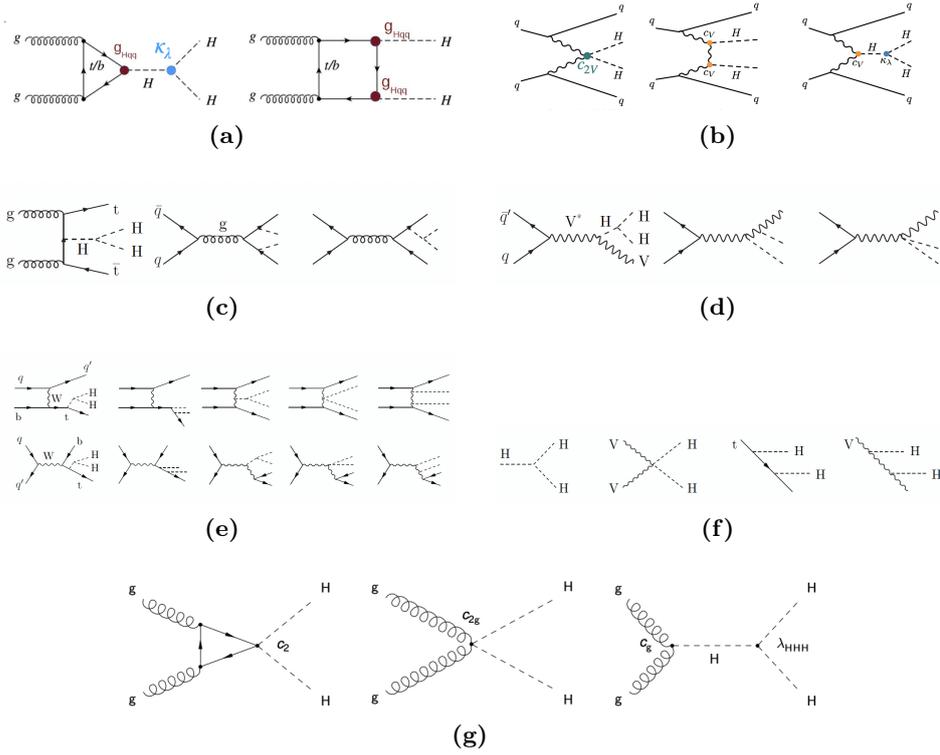
Figure 1.9: Diagrams contributing to Higgs pair production: Figure 1.9a gluon fusion diagrams, Figure 1.9b vector-boson fusion, Figure 1.9d double Higgs-strahlung and Figure 1.9c double Higgs bremsstrahlung off top quarks and Figure 1.9e single top quark associated production. Figure 1.9f Double Higgs boson interaction modes. Figure 1.9g BSM new Feynman diagrams predicted at the same perturbative order of SM diagrams in Figure 1.9a. The trilinear (quadrilinear) Higgs coupling λ_{HHH} (λ_{VVHH}) contribution is visible between three dashed (two dashed plus two wave) lines. In this work the κ_λ (C_{2V}) symbol will denote the ratio $\lambda_{HHH}/\lambda_{SM}$ ($\lambda_{VVHH}/\lambda_{SM}$).

models, like multiplet extensions of the scalar sector [56], additional Higgs doublets (**T**wo-**H**iggs-**d**oublet **M**odel (2HDM) like **SU**per **SY**mmetry (SUSY)) [57] and composite Higgs theories [58]. If instead the scale of BSM physics is significantly higher than the LHC centre-of-mass collision energy, its effects could still be observed as a **non-resonant enhancement** of the production cross section, due to either new particles in the quantum loops, modification of the Higgs boson pair kinematic properties, or to anomalous Higgs boson couplings. In this work, the latter case is examined.

1.4.1   Non-resonant BSM HH production

A general approach to non-resonant HH production, not focusing on a specific model, is discussed in [59]. If the scale of BSM physics is assumed to be beyond the direct reach of the LHC, we can approximate its effects through an addition

of higher order operators to the $d \leq 4$ SM Lagrangian. These additional operators are suppressed by powers of a scale Λ. From a bottom-up perspective,it can be interpreted as the scale up to which only SM fields propagate, while from a top-down perspective it is the energy scale of the BSM physics itself. Using an **E**ffective **F**ield **T**heory (EFT) model[2], BSM effects can be parameterized through a modification of existing couplings and by introducing additional operators. For instance, in the HH production via gluon-gluon fusion, physicists define κ_λ as λ/λ^{SM} and κ_t as y_t/y_t^{SM}. Three additional couplings not present in the SM Feynman diagrams are also considered in this parametrization: contact interactions between two Higgs bosons and two gluons (c_g), between one Higgs boson and two gluons (c_{2g}) and between two Higgs bosons. Since the HH and $t\bar{t}H$ production are sensitive to κ_λ and κ_t, these two categories are combined together to extract the constrains on aforementioned parameters. Moreover, since exploring all the possible combinations of the five couplings is clearly not feasible for an experimental search in terms of complexity of the combinations and computing time, an approach consists in defining **shape benchmarks**, combinations of the five EFT parameters which topologies are representative for large regions of the five-dimensional parameter space, few possible combinations of the five couplings are chosen in such a way that they are representative of the full phase-space. The value of λ_{HHH} is completely determined in the SM once the values of v and m_H are known. However, several BSM models predict a modification of the trilinear Higgs boson self-coupling, modifying the properties of HH production. This can provide the first hints at the LHC of the presence of BSM physics, and serve as an important criterion to discriminate between alternative models [61]. In this context, a parametric approach is adopted and consist in considering the λ_{HHH} value in Equation 1.1.4 as a free parameter. In the following, the HH phenomenology in presence of an anomalous trilinear (HHH) and quartic Higgs boson (VVHH) coupling is first discussed, and serves as in introduction to a more general approach based on an effective SM field theory.

---

2 The Effective Field Theory EFT is one of the profound ideas of quantum field theory that interactions of arbitrary complexity that act at short distances can be approximated systematically by a Lagrangian with an enumerable set of parameters. This Lagrangian provides an "effective" description of any underlying model in this class. The EFT Lagrangian might not be renormalisable in the strictest sense, but it is nevertheless possible to carry out precise calculations that relate the parameters of this Lagrangian to observables [60] and the non-renormalizability does not constitute a problem in this context as an EFT only represents the lower energy manifestation of a more extended (and renormalizable) theory at higher scales.Once the complete model is matched to the EFT, anomalous couplings arise and are directly related to the fundamental parameters of the original model. Indeed,the EFT Lagrangian is written as the SM Lagrangian with corrections described by addition of local operators $\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i$. In this context, BSM physics is fully parametrized in terms of the Wilson coefficients c_i. From an experimental point of view, the previous equation provides a generic parametrization to investigate several BSM signatures with a model-independent approach.The EFT formalism addresses the problem of calculating corrections to the predictions of the SM in a systematic way.

Anomalous Higgs boson couplings in VBF production mode

Vector boson fusion process for Higgs pair production, referred to as VBF HH, has the second highest cross section at the LHC, after the gluon gluon fusion (ggHH) mode. The VBF HH rate, 1.73 fb, is almost 15 times smaller than ggHH.

The leading order diagrams for Higgs pair production with VBF topology are depicted in Figure 1.9b. The key coupling dependency is highlighted for each diagram. Most generally, the actual values of these couplings can be different from those in SM. Hence experimental determination of these couplings is extremely important for Higgs characterization and their precision measurements would provide a possible insight into BSM physics. Various types of Higgs coupling modifiers with respect to SM are denoted as C_V (κ_V) for VVH with V = W, Z, C_3 (κ_λ) for HHH and C_{2V} (κ_{2V}) for VVHH. It is to be noted that C_V is already very well constrained from single Higgs production and decay measurements at the LHC. The di-Higgs production is of prime importance for establishing Higgs trilinear coupling HHH and VVHH. Though ggHH is the main mode for constraining κ_λ, VBF HH provides additional sensitivity due to the diagram in Figure 1.9b (right). As evident from the diagram on the left, VBF HH uniquely provides access to the coupling of Higgs pair to a pair of gauge boson. Hence we can constrain C_{2V} only by studying VBF HH. The diagramin in the middle does not involve κ_λ or C_{2V}, but plays interesting role in the overall analysis of VBF HH having 2 vertices involving C_V.

The main goal of the VBF HH analysis is to establish the presence of C_{2V} mediated process and consequently set a stringent limit on it as a probe to new physics beyond SM. Presence of any anomalous coupling due to new physics is likely to enhance the production rate. The general strategy is to study in detail the event characteristics of few points in the couplings parameters space, and then generalize the analyses to make it as sensitive as possible in the whole space.

For VBF HH study, **M**onte **C**arlo (MC) events for signal are generated at the LO accuracy using MadGraph5_aMC@NLO package [62], utilizing effective field theory calculations and by specifying the values of C_V, C_{2V} and κ_λ . Each event in the sample has a weight corresponding to a given set of values of these parameters. The cross section for VBF HH can be described as an analytic expression in terms of these parameters as presented in

$$
\begin{aligned}
\sigma(C_V, C_{2V}, \kappa_\lambda) &= |C_V \kappa_\lambda A + C_V^2 B + C_{2V} C|^2 \\
&= C_V^2 \kappa_\lambda^2 A^2 + C_V^4 B^2 + C_{2V}^2 C^2 + C_V^3 \kappa_\lambda (AB^* + A^*B) \\
&\quad + C_V \kappa_\lambda C_{2V} (BC^* + B^*C) + C_V^2 C_{2V} (CA^* + C^*A) \\
&= C_V^2 \kappa_\lambda^2 a + C_V^4 b + C_{2V}^2 c + C_V^3 \kappa_\lambda i_{ab} + C_V \kappa_\lambda C_{2V} i_{bc} + C_V^2 C_{2V} i_{ca}
\end{aligned}
$$
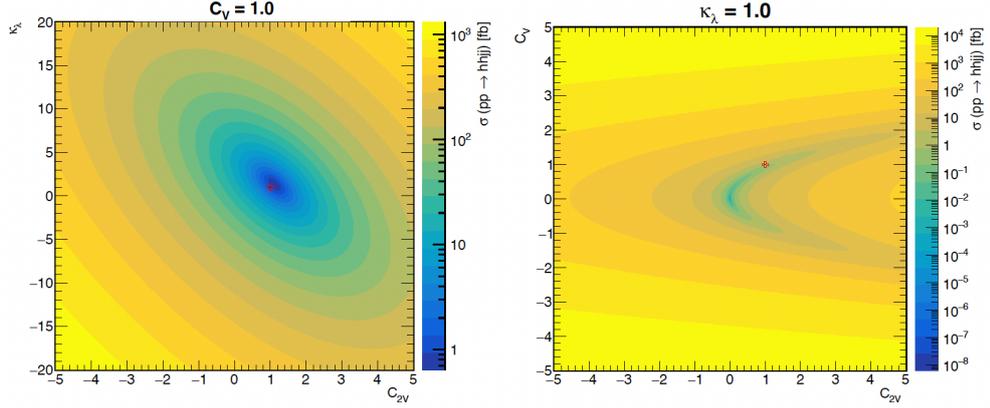
Figure 1.10: Ratio of production cross section with respect to the SM value in 2-d space of C_{2V}-κ_λ (left) and C_{2V}-C_V (right) parameters. The colours shades represent the ratio of generator level cross sections. [63]

Here a, b and c denote the individual square terms of the three diagrams, while i_{ab}, i_{bc} and i_{ca} are the interference terms among the three diagrams. Here to evaluate the cross section and extract the event kinematics for any given set of $(C_V, C_{2V}, \kappa_\lambda)$, seven samples with known parameter values and the corresponding cross sections are needed. Ultimately, the cross section for VBF HH can be written as:

$$\sigma(C_V, C_{2V}, \kappa_\lambda) = \sum_{i=1}^{7} f_i(C_V, C_{2V}, \kappa_\lambda)\sigma_i \qquad (1.39)$$

Here $f_i(C_V, C_{2V}, \kappa_\lambda)$ are the set of functions of $C_V, C_{2V}, \kappa_\lambda$ corresponding to seven different samples and they are orthogonal to each other. The explicit form of the functions f_i depends on the value of the Higgs coupling and σ_i. The previous expression is used to determine the theoretical value of cross section for any set of parameters.

The variation of the ratio of VBF HH cross section w.r.t. SM cross section versus the simultaneous variation of C_{2V} and κ_λ is presented in Figure 1.10.

1.4.2 Experimental searches for Higgs boson pair production at the LHC

The ATLAS and CMS collaborations have exploited a rich variety of signatures to search for HH pair production, exploiting the several Higgs boson decay modes shown in Figure 1.11. There is a large variety of BSM models that can manifest either in resonant or non-resonant HH production. Depending on the specific model, Higgs bosons can have a low transverse momentum or, inversely, be highly boosted. Searches at the LHC thus need to explore several HH decay channels

and to make use of complementary analysis techniques to be sensitive to this large variety of signals.

HH decay modes and results

Measuring the production of Higgs boson pairs at the LHC requires to reconstruct their decay products in the detector and to discriminate them from the large background. The choice of the decay channel of the HH system is crucial in this sense and determines a different trade-off between the branching fraction and the background contamination. Higgs boson pair production, at least in the context of the SM, is characterized by tiny cross sections, so that decay channels with a sizeable branching fraction are preferred. Referring to the single Higgs boson
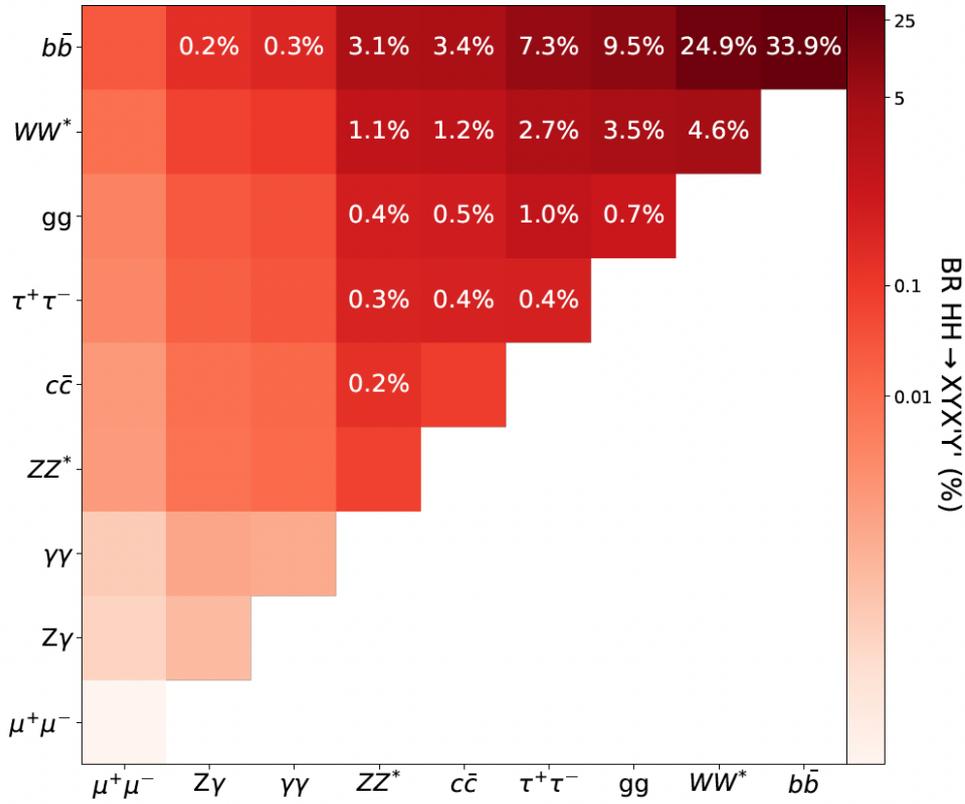
Figure 1.11: Branching fractions of the decay of an HH pair to a selected group of final states. The decay modes are shown on each axis by increasing probability. The numerical values are only shown if larger than 0.1%. The branching fractions of the Higgs boson are evaluated for $m_H = 125.0$ GeV/c^2 [26].

branching fractions of Table 1.2, this consists in requiring that least one Higgs boson decays to a bb or a $W^{\pm}W^{\mp*}$ pair. In the following, HH production and subsequent decays are considered as independent processes and, in particular, SM branching fractions are assumed. Although this might not be the case for some BSM scenarios, good agreement has been observed thus far between the

measured Higgs boson coupling strengths and the SM predictions. Anomalous Higgs boson couplings to fermions and gauge bosons can be probed in single Higgs measurements thanks to the larger cross section and are of limited interest for HH searches.

The sensitivity to HH production at the LHC is mainly due to four main channels:

- $HH \to b\bar{b}b\bar{b}$ is characterized by the highest branching fraction but is affected by a copious multi-jet background. It has consequently little sensitivity to low m_{HH} values but can profit from the large signal yields to probe regions up to $m_{HH} \sim 3$ TeV/c^2.

- $HH \to b\bar{b}VV$ (V = W$^{\pm}$, Z) profits from a sizeable branching fraction and a reduced background contamination. Searches typically focus on the $b\bar{b}$W$^+$W$^-$ decay where the two W (Z) bosons decay leptonically $ll\nu\nu$ (4l), which reduces the branching fraction by about a factor of 10. This channel suffers from a large contamination from $t\bar{t} \to b\bar{b}W^+W^-$ irreducible background.

- $HH \to b\bar{b}\tau^-\tau^+$ represents an optimal compromise between the branching fraction and the background contamination. Contributions from the irreducible $t\bar{t}$ background are suppressed with respect to the $b\bar{b}W^+W^-$ because of the branching fraction B($W \to \tau\nu_{\tau}$) $\sim$11%.

- $HH \to b\bar{b}\gamma\gamma$ is a very pure final state but suffers from a small branching fraction. The clean signature of the photon pairs results in a high signal selection efficiency and provides a powerful tool to separate signal events from the background through the use of the photon pair invariant mass. A similar behavior characterizes the decay channel into four charged leptons and two b-jets, which is analyzed in this thesis.

Many other final states can be studied at the LHC to improve the sensitivity of experiments to HH production. Those listed above represent nevertheless the decay channels whose combination can ensure the largest coverage of the possible HH topologies. Searches for Higgs boson pair production at the LHC in pp collisions at $\sqrt{s} = 8$ TeV (Run I) have been performed by both the ATLAS and CMS collaborations. ATLAS Collaboration explored the $bbbb$, $bb\tau\tau$, $bb\gamma\gamma$, and $WW^*\gamma\gamma$ final states, and evaluated their combined sensitivity [64]. The CMS Collaboration explored and combined the $b\bar{b}b\bar{b}$, $b\bar{b}\tau\tau$ and $b\bar{b}\gamma\gamma$ final states [65]. Results are found to be in agreement with the SM predictions and are used to set upper limits on both the resonant and non-resonant production mechanisms. Subsequently, all of the aforementioned decay channels have been explored by the ATLAS and CMS collaborations with full Run-II data-sets (126 - 139 fb^{-1}

13 TeV data) both in resonant (CMS [66–68], ATLAS [69–72]) and non-resonant condition (CMS [73–76], ATLAS [69, 70, 72, 77]) for what concerns the gluon fusion production mechanism and only recently for the VBF production mode as shown in the non-resonant results in Figure 1.12b and Figure 1.12a.

Experimental challenges are very different depending on the final state considered. The exploration of HH production in its $b\bar{b}b\bar{b}$ final state crucially relies on the capability to identify jets from b-quarks and to reject instrumental background from the mis-identification of gluon/light flavour quark jets. Inversely, final states such as $b\bar{b}VV$ and $b\bar{b}\gamma\gamma$ are mostly affected from irreducible backgrounds, that can be statistically suppressed only by exploiting the kinematic properties of the selected events.

The $b\bar{b}ZZ^* \rightarrow b\bar{b}4l$ final state in vector boson fusion, that is the topic of this thesis, represents a particularly interesting situation. As the the two forward and boosted jets can come from any light/heavy quark, they must be distinguished from instrumental backgrounds caused by the mis-identification of a quark/gluon jet and from the b-quarks jets from one Higgs boson. The properties of the $b\bar{b}ZZ^*$ in VBF decay channel and the major background contributions are further discussed in Section 4.3.1 throughout the rest of this thesis.
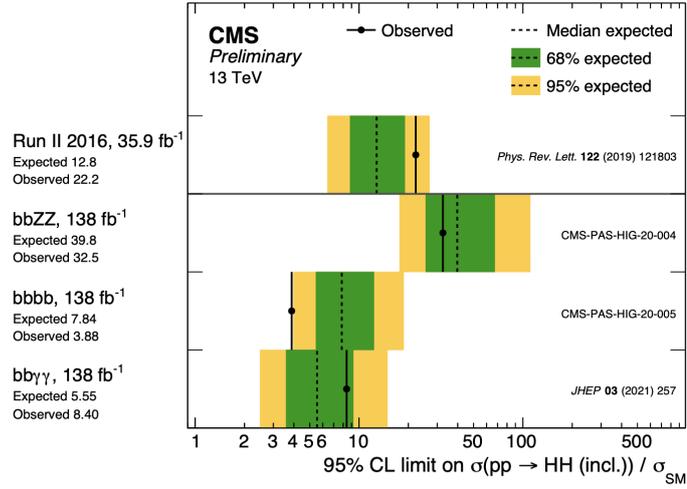
(a)
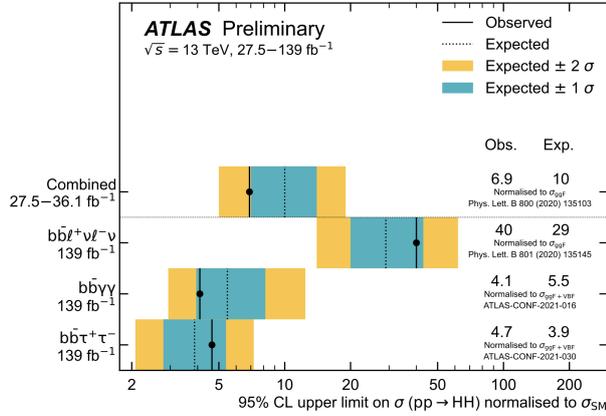
(b)

Figure 1.12: Figure 1.12a Upper limits at 95% C.L. on the ratio of non-resonant HH production cross-section to the Standard Model prediction, obtained over an expected hypothesis assuming the absence of the SM HH signal. Previous combination First Run II data, 36 fb^{-1} and new results exceeding from individual channels exceed the previous results. Figure 1.12b Upper limits at 95% C.L. on the ratio of non-resonant HH production cross-section to the Standard Model prediction, obtained over an expected hypothesis assuming the absence of the SM HH signal. Results are shown from the statistical combination of the $b\bar{b}\tau^+\tau^-$, $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, $W^+W^-W^+W^-$, $W^+W^-\gamma\gamma$ and $b\bar{b}W^+W^-$ searches with 36 fb^{-1} and from the searches using 139 fb^{-1} in the $b\bar{b}l^+\nu l^-\nu$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau^+\tau^-$. The 36 fb^{-1} combination limit is normalised to a cross-section of $\sigma_{ggF} = 33.5$ fb, calculated at NLO in α_S with the heavy top quark mass approximation, then numerically with full top quark mass dependence corrected to NNLO in α_S with NNLL resummation in the heavy top quark limit for $m_H = 125.00$ GeV. The 139 fb^{-1} limits are normalised to cross-sections computed at NNLO including an approximation of finite top quark mass effects. The $b\bar{b}l^+\nu l^-\nu$ limit is normalised to $\sigma_{ggF} = 31.05$ fb and the $b\bar{b}\tau^+\tau^-$ limit is normalised to $\sigma_{ggF+VBF} = 32.78$ fb, calculated for $m_H = 125.00$ GeV. The $b\bar{b}\tau^+\tau^-$ limit uses the most recent prescription for the top quark mass scheme theoretical uncertainties as detailed in the text. The $b\bar{b}\gamma\gamma$ limit is normalised to $\sigma_{ggF+VBF} = 32.74$ fb, calculated with m$_H = 125.09$ GeV [78].

Part II

EXPERIMENTAL SETUP

The Compact Muon Solenoid Detector at LHC

2

> *"Detectors... are really the way you express yourself. To say somehow what you have in your guts. In the case of painters, it's painting. In the case of sculptures, it's sculpture. In the case of experimental physicists, it's detectors. The detector is the image of the guy who designed it."*
>
> – Carlo Rubbia [79]

T HE CERN, founded on 29 September 1954, is one of the world's leading centres for particle physics producing cutting-edge science and technology. It inspires and trains generations of scientists, technicians and engineers worldwide in the pursuit of knowledge, counting a manpower of almost 10 000 people from more than 100 different countries in the world using its facilities.

The CERN laboratories host the LHC, consisting of a 27-kilometre ring of superconducting magnets with a number of accelerating structures to boost the energy of the particles (protons or lead ions) along the way. It first started up on the 10th of September 2008, and remains the latest addition to CERN's accelerator complex. It collides the particle beams in four interaction points, instrumented with an equal number of detectors at LHC.

In one of these four points is installed the CMS experiment, designed to explore the physics at the TeV scale and possibly reveal some violations that can be a hint of new physics, described by BSM theories. At the moment the CMS collaboration involves the active participation of more than 5500 scientists from 51 different countries.It is in this experimental team that the work presented in this thesis has been developed.

This chapter introduces the properties and operations of the LHC and the structure of the CMS detector used to collect the data analysed in this thesis. The algorithms to reconstruct the particles produced in the collisions from the raw

detector data are presented. Finally, the trigger system of the CMS experiment is described.

2.1 The Large Hadron Collider

The LHC is designed to collide protons at a centre-of-mass energy of $\sqrt{s} = 14$ TeV (7 TeV per beam) with a nominal peak luminosity $\mathcal{L} = 10^{34} cm^{-2}s^{-1}$[80, 81]. LHC do not only collide proton-proton (pp), but also lead-lead (Pb - Pb), proton-lead (p - Pb) and Xenon-Xenon (Xe - Xe) nuclei for the purpose of studying heavy-ion collisions at the center-of-mass energy $\sqrt{s} = 5.02, 8.16$ and 5.44 TeV, respectively[1]. The protons are accelerated to high energy by a series of accelerators before being injected into the LHC machine, each accelerator raises the energy of the protons plus transfers them to the next accelerator. Those series of accelerators are called the *LHC accelerator complex* which will be briefly described in this section.

2.1.1 Operations

The realization of the LHC constituted a two decade-long international effort for investigating the scalar sector of the SM and also the physics BSM. Scientists started thinking about the LHC in the early 1980s, when the previous accelerator, the LEP, was not yet running. In December 1994, CERN Council voted to approve the construction of the LHC and in October 1995, the LHC technical design report was published. LHC was finally inaugurated in 2008. On the 10th of September 2008, LHC succeeded to circulate the proton beam for the first time until the ninth day when a shutdown was forced due to an electrical connection faulty between the superconducting magnets, causing the release of helium into the tunnel and mechanical damage [82, 83]. On the 20th of November 2009, the proton beams were successfully circulated again and the first recorded proton-proton collisions occurred 3 days later at collision energy of 450 GeV per beam [84] and then later at $\sqrt{s} = 2.136$ TeV which corresponded to an higher center of mass energy w.r.t. the one of the Tevatron ring (1.96 TeV). This made the LHC the highest energy collider ever built. On 30th of March 2010 , LHC increased the energy of proton beams to 3.5 TeV leading to a new world record for the highest-energy man-made particle collider [85]. Data collection at $\sqrt{s}$= 8 TeV was done from 5th of April 2012 until the 6th of February 2013. This collision time-period (2010-2013) is known as *Run I* of the LHC. About $45pb^{-1}$ and 6 fb^{-1} were collected by the CMS experiment at $\sqrt{s}$= 7 TeV in 2010 and 2011, respectively, and a larger data-sets of 23 fb^{-1} was recorded at

---

1 The definition of luminosity and center of mass energy will be introduced in Section 2.1.3

$\sqrt{s} = 8$ TeV in 2012. These data allowed for the discovery of the Higgs boson and for a precise measurement of its properties. After Run I, LHC entered a period of **L**ong **S**hutdown (LS) for two years (known as LS1) for upgrades. Important consolidation works were performed to push the LHC performance towards the design parameters. LHC started its operation again in 2015 Run II with energy of 6.5 TeV per proton beam and the nominal spacing of 25 ns. On 3rd of December 2018 Run II ended after collecting a large amount of data ~ 150 fb^{-1} and LHC entered a **L**ong **S**hutdown 2 (LS2) period. After LS2, LHC is scheduled to operate again in 2022 at center-of-mass energy of 14 TeV (Run III). During the **L**ong **S**hutdown 3 (LS3) period, the accelerator infrastructure will be updated with new detectors and trigger technologies to cope with the increasing of energy and unprecedented instantaneous luminosity resulting between 5.0 and $7.5 \times 10^{34} cm^{-2} s^{-1}$ corresponding to a total of 3000 to 4000 fb^{-1} of collected data for the full accelerator lifetime. This project is called HL-LHC or Phase II. Figure 2.1 shows the detailed timeline of LHC and HL-LHC run operations with the evolution of the beam energy and integrated luminosity. The main purpose of the HL-LHC is to improve the search sensitivity for many physics processes such as rare processes which are statistically limited. The properties of the discovered Higgs boson will be measured with higher precision, which will allow searches for new physics associated with the Higgs sector. The searches will be extended also for new exotic and Supersymmetric models.

Figure 2.1: Timeline LHC and HL-LHC Plan vs the corresponding center-of mass energies and integrated luminosity of the collider experiments [86].

2.1.2 The LHC accelerator complex

The accelerator tunnel of the LHC was previously used between 1984 and 1989 to host the LEP collider [87], with an average depth of about 100 metres under the surface of the Earth (50 -175 m). It is situated beneath the France-Switzerland border near the city of Geneva [88].

In the LHC, two separate and parallel, counter-rotating particle beam lines are kept in orbit along two rings in two magnetic separated channels thanks to the field generated by superconducting niobium-titanium (NbTi) cables which allow the high currents to flow without losing any energy to electrical resistance. Indeed, the particles are steered by a magnetic field of ~ 8.33 Tesla – more than 100 000 times more powerful than the Earth's magnetic field - generated by a current of about 11 080 A in the 1232 dipole magnets, each measuring ~ 14.3 metres of length and 35 tonnes of weight. This high technology global magnet system allows to reaches the nominal LHC beam energy of 7 TeV. The system uses a total of about 9600 magnets. If normal magnets were used in the 27 km-long LHC instead of superconducting magnets, the accelerator would have to be 120 kilometres long to reach the same energy. The stability of the beam dynamics is ensured by 392 quadrupoles magnets measuring 5 to 7 metres of length, that focus the particles and keep them in a narrow beam. Special quadrupoles are installed in front of the collision points to squeeze the beams either vertically or horizontally and increase the proton density in the collisions. Dipoles are also equipped with sextupole, octupole and decapole magnets, which correct for small imperfections in the magnetic field at the extremities of the dipoles. Superconducting magnets are cooled with cryogenic system of 120 tonnes superfluid Helium-4 and kept to a working temperature of 1.9 K.

The LHC is the last element of an injection chain composed of several smaller particle accelerators [23, 89, 90], as reported in Figure 2.2 and Figure 2.3. To make protons, physicists inject hydrogen gas into the metal cylinder - *duoplasmatron* - then surround it with an electrical field to break down the gas into its constituent protons and electrons. This process yields about 70 percent protons that are accelerated to an energy of 50 MeV in **L**inear **A**ccelerator **2** (LINAC**2**) (LINAC3 is the injection line for lead ions) which feeds the **P**roton **S**ynchrotron **B**ooster (PSB) where protons are accelerated to ~ 1.4 GeV. The beam is then injected into the Proton Synchrotron (PS) for a further acceleration to ~ 25 GeV, and subsequently into the SPS where protons reach an energy of ~ 450 GeV. The proton beams are finally transferred to the two LHC beam pipes, where the beams are accelerated and synchronized temporally, grouping them into discrete packets, called *bunches* thanks to **R**adio**F**requency cavities (RF) operated at 400 MHz. Once the proton

reach the nominal energy for the Run and the beam dynamics is stabilized, protons are brought to collide in four points along the LHC ring.

The four collision points of the LHC are instrumented with particle detectors installed in underground caverns. ATLAS [91] and the CMS [92] experiments are installed in the diametrically opposite Points 1 and 5 of the LHC, where the highest instantaneous luminosity of collision is produced. They are designed as hermetic, multi-purpose detectors that surround the interaction point and measure the debris of proton and ion collisions. The LHCB [93], located at Point 8 is a forward, one-arm spectrometer devoted to the study of **C**harge conjugation and **P**arity symmetry (CP) violation in B hadrons. ALICE [94] is installed in Point 2 and is built to study heavy ion collisions and quark-gluon plasmas. The LHC also hosts three smaller size experiments. The LHCF and the TOTEM experiments, located a hundred of meters away on either side from the ATLAS and CMS interaction points, are dedicated to pp interaction cross section measurements and forward diffractive physics. LHCB shares its cavern with the MoEDAL experiment, dedicated to the search for magnetic monopoles. FASER, the newest LHC experiment, is situated 480 metres from the ATLAS collision point in order to search for light new particles and study neutrinos.
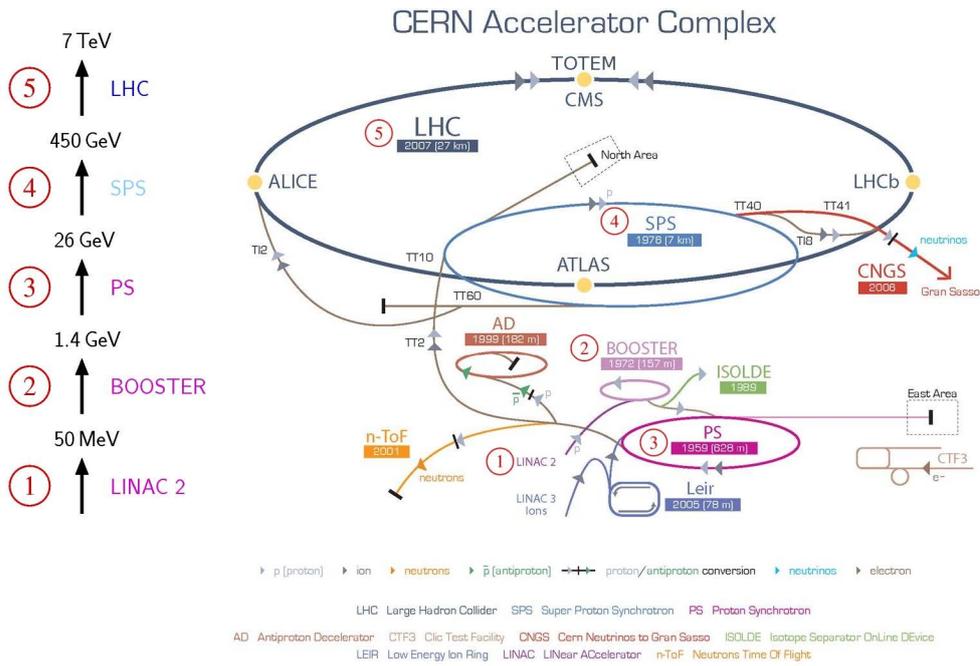
Figure 2.2: Schematic view of the CERN accelerator complex. The figure is taken from Ref. [95].
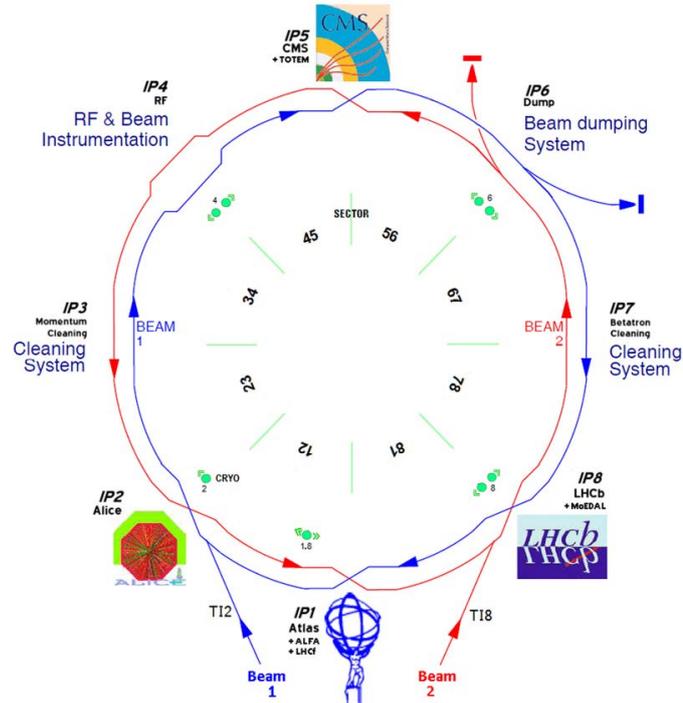
Figure 2.3: Layout of the LHC with the eight interaction points labelled IP1 to IP8.
The experiments ATLAS, ALICE, CMS and LHCB are installed in IP1,
IP2, IP5 and IP8, respectively. Beam 1 is injected close to IP2 and circulates
clockwise, Beam 2 is injected close to IP8 and circulates counter-clockwise.
The two beams exchange position between outside and inside of the ring
at every experiment to ensure that the path length is the same for both
beams. The two beam dumps are located around IP6 [92].

2.1.3   Design and specifications

One of the most important LHC parameters is the available energy at the collision
point to generate new physics processes. For the collider experiments, the total
energy available at the center-of-mass is the sum of the energy of the two incoming
beams paying attention of the type of physical interaction occurring. Indeed, when
two bunches of protons collide, several independent proton-proton interactions
can take place, from which particles can originate. Events produced in these
interactions can be of two types:

1. **Soft interactions**: characterized by a small amount of momentum trans-
   ferred. The particles produced have small transverse momentum, so most
   of them cannot be detected because they escape down the beam pipe. The
   final states of these interactions are called *Minimum Bias* events.

2. **Hard interactions**: characterized by high transverse momentum trans-
   ferred, due to the collisions between partons (quarks or gluons).

Not all the center-of mass energy contributes in the hard interaction but only a fraction of the protons energy. When two protons collide, two of its partons can participate in the interaction the effective center-of-mass $\sqrt{s}$ is smaller than the energy of the incoming proton beams, and it is proportional to the fractional energies x_a and x_b carried by the two interacting partons:

$$\sqrt{s} = \sqrt{x_a x_b s} \tag{2.1}$$

In the design of the LHC apparatus the instantaneous luminosity $\mathcal{L}$ is also a crucial specification. It determines the number of collisions that can be produced in a detector per unit time and unit cross-sectional area of the beams times. The higher the luminosity of the collider the higher the probability of particle interactions to happen. It depends on the geometrical and kinetic beam properties as [88, 96]:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta} F \tag{2.2}$$

where N_b is the number of particles in each of the n_b bunches per beam revolving in the accelerator ring with a frequency f_{rev}. The parameter γ denotes the relativistic Lorentz factor of the particles. The shape and focus of the beam are described by its renormalized transverse emittance ϵ_n [2] and its focal length (betatron function) β [3], at the collision point. The factor F accounts for the geometric reduction of the instantaneous luminosity, and depends on the angle between the two beams, the *beam crossing angle* θ and on the RMS bunch sizes in the plane orthogonal to the beam direction and along the beam line σ_{xy} and σ_z at the interaction point as:

$$F = (1 + \frac{\theta \sigma_z}{2\sigma_{xy}})^{-1/2} \tag{2.3}$$

It is easy to prove that the present technologies cannot permit to group a number of anti-protons per bunch large enough to reach the LHC nominal instantaneous luminosity, hence a simple particle-antiparticle accelerator collider configuration could not be used at LHC.

---

2 The renormalized transverse emittance ϵ_n is a beam quality concept reflecting the process of bunch preparation (the injector chain), extending all the way back to the source for hadrons. A low emittance particle beam is a beam where the particles are confined to a small distance and have nearly the same momentum. A beam transport system will only allow particles that are close to its design momentum, and of course they have to fit through the beam pipe and magnets that make up the system. In a colliding beam accelerator, keeping the emittance small means that the likelihood of particle interactions will be greater resulting in higher luminosity.

3 The amplitude function, β , is determined by the accelerator magnet configuration (basically, the quadrupole magnet arrangement) and powering. When expressed in terms of σ (cross-sectional size of the bunch) and the transverse emittance, the amplitude function β becomes: $\beta = \pi \sigma^2 / \epsilon_n$. So, β is roughly the width of the beam squared divided by the emittance. If β is low, the beam is narrower, "squeezed". If β is high, the beam is wide and straight. β has units of length. Sometimes β is referred as the distance from the focus point that the beam width is twice as wide as the focus point .

In addition to the previously mentioned parameters, the luminosity lifetime τ is an important parameter at LHC and colliders in general. Indeed, the luminosity decreases exponentially with the law [97]:

$$\mathcal{L} = \mathcal{L}_0 e^{-t/\tau} \tag{2.4}$$

where $\mathcal{L}_0$ is the peak instantaneous luminosity at time zero and τ the luminosity lifetime.

(a)

(b)
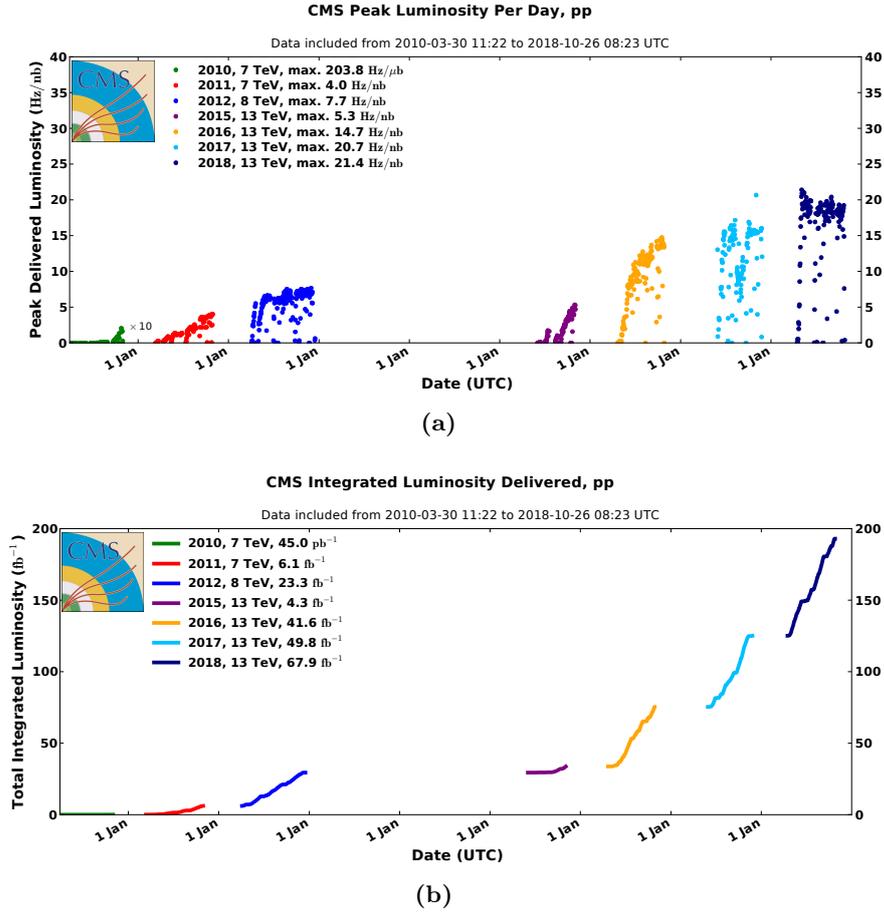
Figure 2.4: Figure 2.4a Peak luminosity versus day delivered to CMS during stable beams and for pp collisions. This is shown for data-taking in 2010 (green), 2011 (red), 2012 (blue), 2015 (purple), 2016 (orange), 2017 (light blue), and 2018 (navy blue). Figure 2.4b Cumulative delivered and recorded luminosity versus time for 2010-2012 and 2015-2018 during stable beams for pp collisions at nominal center-of-mass energy [98].

The integrated (cumulative) luminosity $L = \int \mathcal{L} \, dt$ is a measure of the total amount of collisions produced over the time experiment runs. During the last two years of Run II (2017 - 2018) operation, the LHC achieved a 50 percent up-time, meaning it was actively colliding protons for an average of 12 hours a day, seven days a week, for six months a year. During the last year of Run I,

proton - (anti)proton cross sections
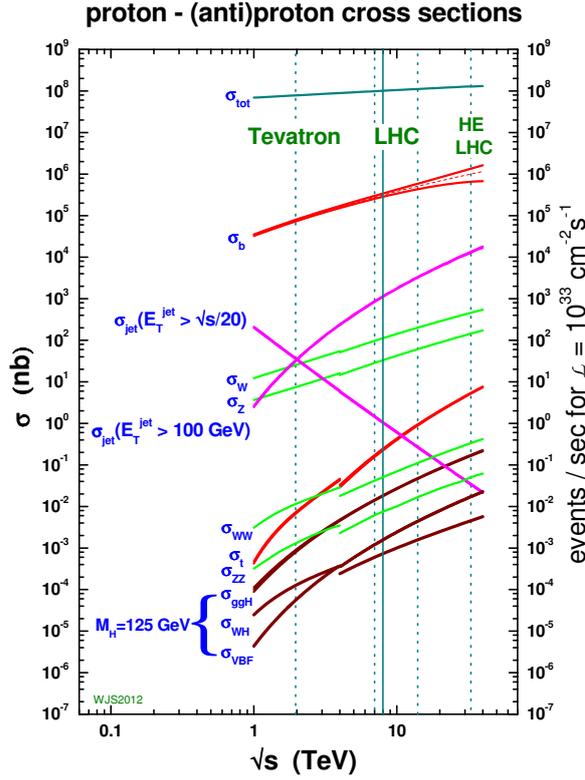
Figure 2.5: Standard Model cross sections as a function of collider energy, with 125 GeV/c^2 Higgs [99].

the accelerator's up-time was 36 percent. In 2016, the LHC achieved a record performance of colliding beam particles for about 70% of the operation time. The cumulative luminosity constitutes the coefficient of proportionality between the number of events N produced for a specific physics process and its cross section:

$$N = L \times \sigma \tag{2.5}$$

Therefore, in order to access rare processes (i.e. processes with very low cross sections), it is very important to maximize the luminosity of the accelerator. In Figure 2.5 the production cross sections at pp and $\overline{p}p$ colliders, predicted by the SM for different physics processes, are displayed as a function of the center-of mass energy of the accelerator. Significant physical processes such as the Higgs physics and the physics beyond the SM is (about 3-4) orders of magnitude lower than W/Z production cross section and much more w.r.t. inelastic pp cross section ($\sim 80mb$), but it increases with increasing the center-of-mass energy of the collision (see Figure 2.5). Therefore, increasing the luminosity of the collider and its center-of-mass energy permits to reach higher event rates. The total integrated luminosity delivered by LHC and recorded by CMS experiment for proton-proton

collisions since 2010 is shown in Figure 2.6a. The integrated luminosities for 2016, 2017 and 2018 data-takings are displayed in the plots in Figure 2.6b , Figure 2.6c and Figure 2.6d. In Figure 2.4a and Figure 2.4b, the peak luminosity and the integrated luminosity delivered per day by LHC to the CMS experiment are reported [98]. A summary of the nominal LHC machine parameters is given in Table 2.1.

Symbol	definition	LHC design
$\sqrt{s}$	center-of-mass energy	14 TeV
δt_b	bunch spacing	25 ns
n_b	bunches per beam	2808
N_b	particles per bunch	1.15×10^{11}
f_{rev}	revolution frequency	11.2 kHz
ϵ_n	transverse beam emittance	3.75 μm
β	beta function	0.55m
θ	crossing angle at i.p.	285 μrad
σ_{xy}	transverse r.m.s. bunch size	16.7 μm
σ_z	longitudinal r.m.s. bunch size	7.55 cm

Table 2.1: Nominal parameters of the LHC machine in pp collisions

During Run II, the LHC designed luminosity for the proton-proton collisions was reached in June 2016 and doubled in 2017. Data collected from CMS in 2018 are used to derive the results presented in this thesis. The average number of interactions for each bunch crossing depends on the beam parameters which I illustrate previously, e.g. the number of particles in a bunch, how much the bunch is focused, etc. In 2018 data-taking there were, on average, 37 interactions per bunch crossing, as shown in Figure 2.7. The presence of many primary vertices per bunch crossing is a challenge for the event reconstruction, since the particles originating from different primary vertices can be superimposed in the detector. Interactions besides the ones of interest, that one wants to study, are referred to as in-time PU. Moreover, collisions took place in the centre of the CMS experiment every 25 ns, implying that new waves of particles leave the interaction point before those produced in the previous bunch crossing have even escaped the external surface of the detector. This last effect is denoted as out-of-time pileup and overlap to the signal of interest represented by the hard-scatter interaction. These challenging conditions call for a detector design that is highly granular, fast in its response, and resistant to the radiation. At the same time, it must be capable of precisely measuring the energy and the momentum of the final state particles and to identify them. The design of the detector, described in Section 2.2, has been conceived to fulfil these requirements.
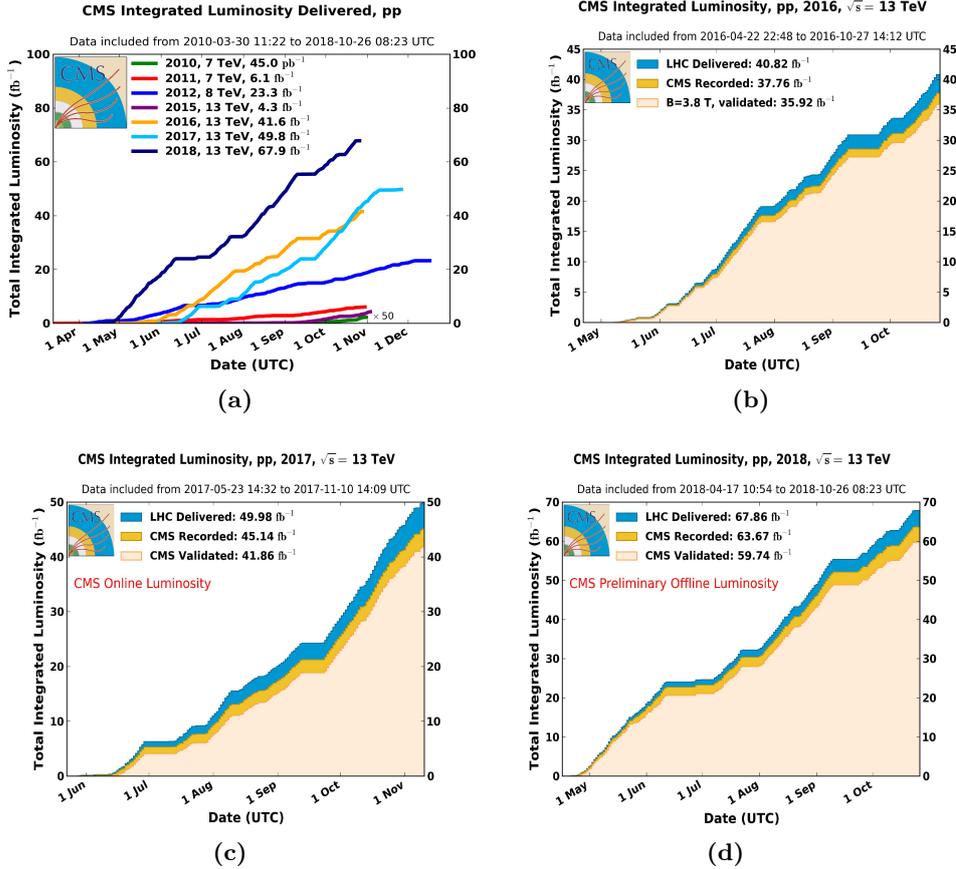
Figure 2.6: Figure 2.6a Total integrated luminosity delivered by the LHC to the CMS detector for proton-proton collisions during Run I and Run II, 2010-2018 [98, 100]. Total integrated luminosity delivered by the LHC to the CMS, and the total recorded versus validated luminosity (for a use in physics analyses, and include dead times of the experiment) by the CMS for different data-taking periods : 2016 Figure 2.6b, 2017 Figure 2.6c and 2.6d 2018 for pp collisions at a center-of-mass energy of 13 TeV.

2.2 The Compact Muon Solenoid experiment

The CMS Collaboration brings together members of the particle physics community from across the globe in a quest to advance humanity's knowledge of the very basic laws of our Universe. It includes over 5500 particle physicists, engineers, computer scientists, technicians and students from around 200 institutes and universities from more than 40 countries.

The collaboration operates and collects data from the CMS detector, one of the multi-purpose particle detectors at the Large Hadron Collider LHC facility that hermetically surrounds the interaction point in the underground caver of Point 5 in Cessy.

CMS Average Pileup

Legend:
- 2018 (13 TeV): $<\mu> = 37$
- 2017 (13 TeV): $<\mu> = 38$
- 2016 (13 TeV): $<\mu> = 27$
- 2015 (13 TeV): $<\mu> = 13$
- 2012 (8 TeV): $<\mu> = 21$
- 2011 (7 TeV): $<\mu> = 10$

$\sigma_{in}^{pp}\,(13\;TeV) = 80.0\;mb$
$\sigma_{in}^{pp}\,(8\;TeV) = 73.0\;mb$
$\sigma_{in}^{pp}\,(7\;TeV) = 71.5\;mb$
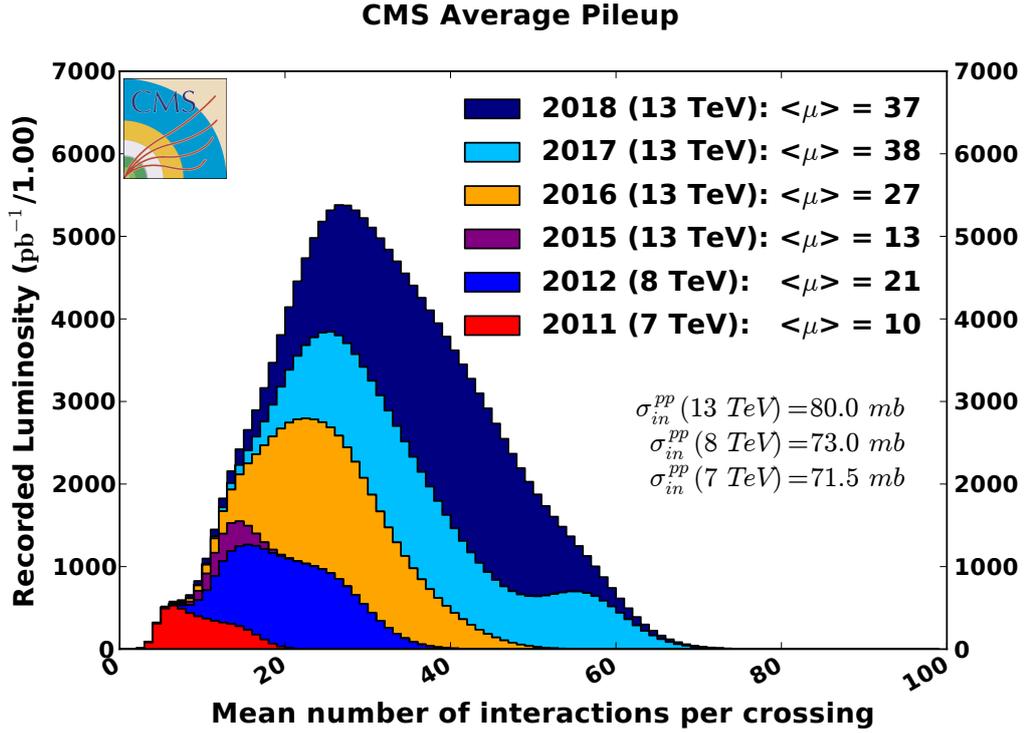
Figure 2.7: Distribution of the average number of interactions per crossing bunch (in-time pileup) for pp collisions in 2011 (red), 2012 (blue), 2015 (purple), 2016 (orange), 2017 (light blue), and 2018 (navy blue). The overall mean values and the minimum bias cross sections are also shown. These plots use only data that passed the "golden" certification (i.e., all CMS sub-detectors were flagged to be ok for any kind of usage in physics analysis), and the "LHC standard" values for the minimum bias cross sections, which are taken from the theoretical prediction from Pythia and should be used to compare to other LHC experiments [98]. The overall mean values and the minimum bias cross sections are also shown.

It has a broad physics program ranging from the investigation of the physics underlying the electro-weak symmetry breaking (including the scalar sector and the Higgs boson physics) to searching for extra dimensions and particles that could make up dark matter at TeV scale. Although it has the same scientific goals as the ATLAS experiment, it uses different technical solutions and a magnet-system design to explore the physics at the TeV scale in many different signatures and final states. The detector's denomination summarizes its main features :

1. The *«Compact»* first term is related to its small size compared to its mass and to the ATLAS's dimensions ;

2. the *«Muon»* adjective underlines the advanced muon system in the outer layer of the detector ;

3. the last noun *«Solenoid»* is due to a superconducting solenoid magnet.

CMS DETECTOR

Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm²) ~1 m² ~66M channels
Microstrips (80–180 μm) ~200 m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000 A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16 m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 2.8: The **T**hree **D**imensional (3D) model of the CMS detector. It was created in SketchUp via its Ruby API. The detector geometry were obtained from the CMS Detector Description stored in CMSSW packages. The figure is taken from Ref. [101].

2.2.1 The CMS detector structure and the Solenoidal Magnet

The CMS detector [92] has a cylindrical structure with a diameter of 15 m and a length of 21.5 m, and an overall weight of about 12 500 tonnes. It is constituted of a central section, or "*barrel*", and two forward regions, or "*endcaps*" that can be seen in Figure 2.8. It is composed by multiple, concentric layers of sub-detectors using several technologies for particle detection and reconstruction [102]. From the innermost part towards the outside, three layers of sub-detectors are situated inside a superconducting solenoid magnet: indeed, firstly the interaction point is surrounded by pixel tracker and strip tracker detectors to precisely measure the positions of the interaction points (or "*collision vertices*"), the trajectory and momentum of the charged particles. Then, the **E**lectromagnetic **CAL**orimeter (ECAL) and **H**adronic **CAL**orimeter (HCAL) are located around the tracking systems and designed to absorb electrons, photons, and hadrons within their volume to measure the energy deposited.

Finally, muons can traverse the calorimeters and are measured in muon tracking systems occupying the most external part of the CMS detector outside the superconducting solenoid interspersed within an iron magnetic flux return
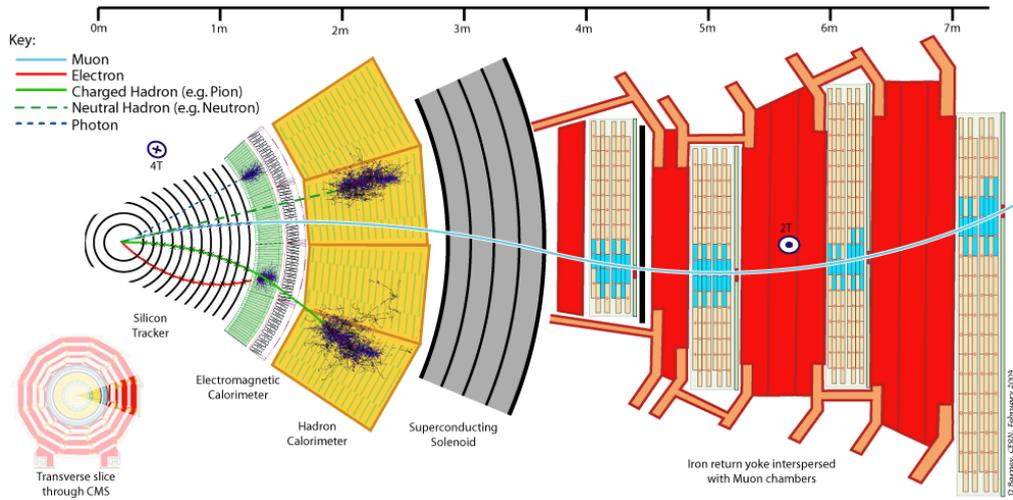
Figure 2.9: Schematic view of a transverse slice of the central part of the CMS detector. The figure is taken from Ref. [89].

yoke with three layers. The core of the experiment is a large niobium-titanium superconducting solenoid of 5.9 m inner diameter and 12.5 m long.

It is operated at a temperature of 4.5 K and generates, by circulating a current of 18 kA, a 3.8 T uniform magnetic field along the beam line axis, the z-axis, and therefore the particles trajectories are bended in the transverse (x,y) plane[4]. This strong magnetic field within a compact detector volume is used to bend the charged particles and measure their transverse momentum up to TeV/c with the tracking sub-detectors from their track curvatures [103]. The tracker and calorimeters systems are located inside the solenoid, which poses tight constraints of their size and, in the case of the calorimeters, requires high density materials to contain the incoming particles and their secondary interaction products. The return field of the magnet has an intensity of about 2 T and is used to measure the transverse momentum with the muon detectors located inside in the iron structure that surrounds the solenoid. This causes the muons trajectories to be bent in opposite directions in the inner tracker and muon systems, a characteristic feature to which the CMS experiment owes its logo. This latter part of the "*CMS onion*" is a redundant system based on three different gaseous detector technologies, namely **D**rift **T**ube (DT), **C**athode **S**trip **C**hamber (CSC), and **R**esistive **P**late **C**hamber (RPC). Figure 2.9 shows the signatures left when different types of the particles pass through the CMS sub-detector. Putting together the information from different sub-detectors helps in discriminating between distinctive species of elementary particles. Both photons and electrons leave showers in the electromagnetic calorimeters but photons don't leave tracks

---

4 The CMS reference system convention will be discussed in Section 2.2.2

in the tracking system permitting in this way to distinguish the two species. In the same way we can identify charged and neutral hadrons. Before a description of the sub-detectors, the coordinate system of the experiment must be mentioned.

2.2.2 The CMS coordinate system

A right-handed cartesian coordinate system, which is shown in Figure 2.10, is used to describe the detector and the collision products. It is defined with its centre in the nominal interaction point, the x-axis pointing radially inward to the centre of the LHC ring, the y-axis pointing vertically upwards, and the z-axis pointing in the anticlockwise proton beam direction.

Since the experiment has a cylindrical symmetry, it is very useful to define cylindrical coordinates to label the position of particles. In particular, a radial coordinate r and two angles are used. The r coordinate is measured in the (x, y) plane. This holds also to the azimuthal angle ϕ whose zero value corresponds to the x-axis directed toward the centre of the LHC ring and takes values of $[-\pi, +\pi]$ The polar angle called θ if defined such that null value lies along the beam-pipe in the positive z-axis and takes values of $[0, +\pi]$. Moreover, it is known that the particles produced by proton collisions are strongly boosted to the beam-line axis and it is more convenient that the particle production is constant as a function of the observable that is used for describing the reference-frame.

For these reasons, a different kinematic quantity is preferred, called pseudo-rapidity $\eta = -ln(tan(\theta/2)$, which is a good relativistic reference-frame independent kinetic observable and has the advantage to be independent on the Lorentz boost resulting from the initial momentum of the interacting partons along the z-axis, classified as one of the three Mandelstam variables[5]. Smaller (larger) values of the η represent the particles lying in a direction perpendicular (parallel) to the beam axis. Based on η-values, within the CMS detector the observer can distinguish different regions:

- two ***endcaps***, characterized by $1.2 < |\eta| < 2.4$;

- the ***barrel***, corresponding to the region with $|\eta| < 1.2$.

The spatial separation of two particles can be expressed in terms of their angular distance which is related to the cones around the particle direction with a radius of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The projections of the particle's momentum and energy onto the transverse plane ($\eta = 0$) are denoted as the **T**ransverse **M**omentum ($\boldsymbol{p_T}$) $= sin\theta$ and $E_T = E sin\theta$.

---

5 In theoretical particle physics, the Mandelstam variables are numerical quantities that encode the energy, momentum, and angles of particles in a scattering process in a Lorentz-invariant fashion.

(a)

(b)
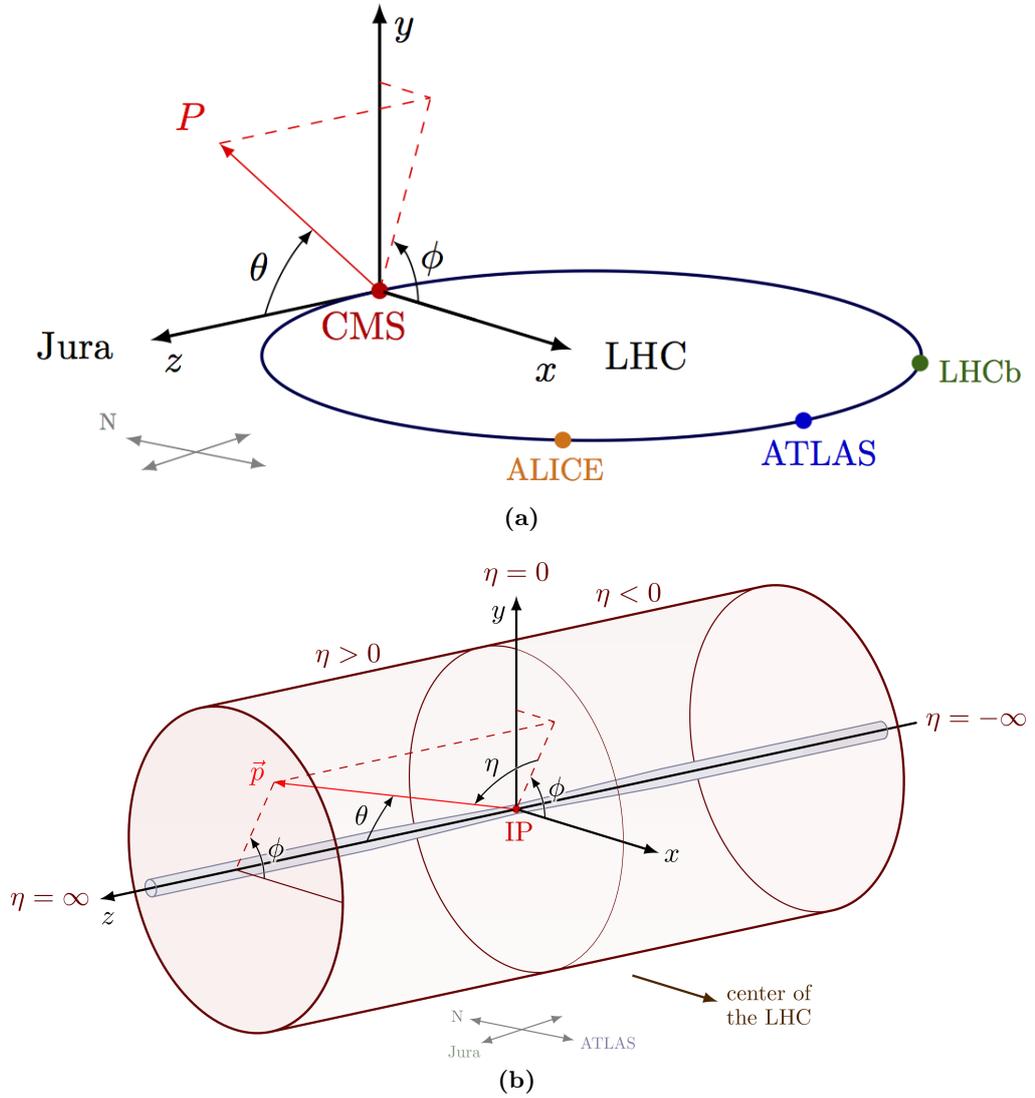
Figure 2.10: Figure 2.10a CMS conventional 3D coordinate system w.r.t. LHC accelerator complex. Figure 2.10b CMS 3D coordinate system within its cylinder structure [104].

2.2.3 Inner tracking system

The inner tracking system is the first CMS sub-detector which is located directly around the collision point within the pseudo-rapidity region $|\eta| < 2.5$. Thanks to the uniform magnetic field within the tracking detector volume, the informations on the position of charged particles within each detector, or "hits", are combined to measure with high resolution the momentum and charge of the particles as they propagate outward from the interaction point. The precision of its momentum measurement increases as p_T decreases, because of the larger curvature of the trajectory. The spatial measurement provided by the tracking system also allows for the determination of the hard scatter interaction point, i.e. the **P**rimary

Vertex (PV), and its discrimination against additional interactions from pileup in the event. It also allows for the reconstruction of in-flight decays such as those of B hadrons or τ leptons from the **S**econdary **V**ertex (SV).

Because of its spatial location near the interaction point, it is exposed to a very large flux of particles. For this reason, the CMS detector tracking system required a technology that is able to provide high segmentation information and to guarantee good radiation hardness, keeping to the minimum the amount of material, in order to limit the energy losses that the particles experience from passing through matter by multiple coulomb scattering, bremsstrahlung and nuclear interactions.
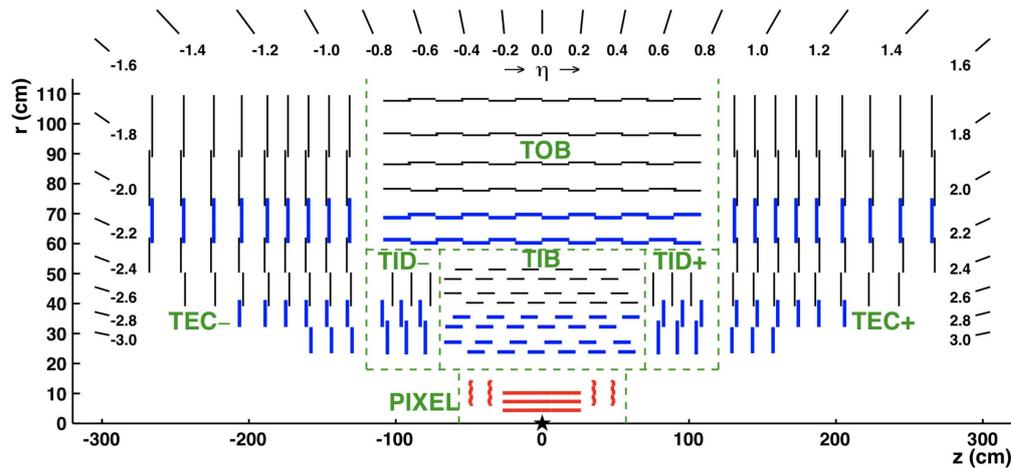
Figure 2.11: Schematic view of half of the inner tracking system. The centre of the tracker, corresponding to the approximate position of the pp collision point, is indicated by a star. The strip tracker detector is composed of the **T**racker **I**nner **B**arrel (TIB) and **T**racker **I**nner **D**isks (TID), surrounded by the **T**racker **O**uter **B**arrel (TOB) and the **T**racker **E**ndcaps (TEC) [105].

In order to fulfill all of these requirements, silicon technology with fast readout on-board electronics has been chosen for the whole tracker [106, 107], which is made up of two different silicon sensors that are sensitive to the passage of charged particles and occupy a volume of 5.6 m of length and 2.4 m of diameter as shown in Figure 2.11. Indeed, higher spatial precision is required close to the beam pipe for the identification of the interaction vertices.

The silicon pixel sub-detector

In the **innermost region** there is a silicon pixel sub-detector that consists of four **B**arrel **PIX**el detector (BPIX) layers lies around the beam-pipe, three **E**ndcap **PIX**el detector (EPIX) disks in each endcap region having pixel cells of ~ 100 μm $\times 150$ μm size (see Figure 2.11). It guarantees a spatial resolution of 10 μm in

the transverse plane (r, ϕ) and 20 μm in the z-direction, while the third coordinate is given by the sensor plane position.

In total, its 1440 modules cover an area of about 1 m^2 and have 66 million pixels [105]. Therefore it allows very precise measurements and provides a small **I**mpact **P**arameter (IP) resolution [6], that is crucial for good secondary vertex reconstruction [108]. Silicon pixels are connected in a reverse-bias during operation. When the charged particles pass through the pixels the electrons in the silicon are excited from the valence to the conduction band of the semiconductor creating small currents. These small currents are amplified by electronics. If these currents fire a trigger condition, the data will send to the **D**ata **A**c**Q**uision (DAQ) system or no trigger is received and the data is overwritten.
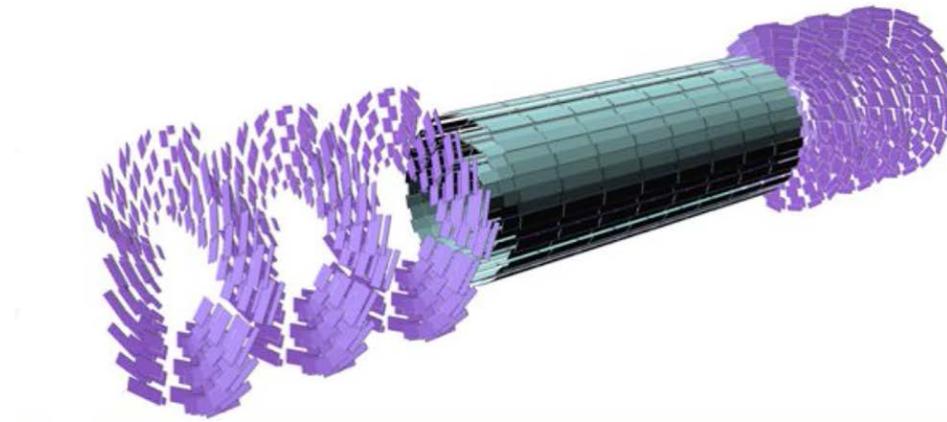
Figure 2.12: Longitudinal view of the CMS silicon pixel detector. The interaction point is surrounded by four cylindrical barrel layers at radii of 3, 4.4, 7.3 and 10.2 cm and three endcap disks on each side with modules between 6 and 15 cm from the beam axis [23, 109].

However, over the time, the pixel detector will suffer from degradation of the resolution due to radiation damage. To keep its high performance, a **Phase Pixel Upgrade** was performed during the technical stop at the end of 2016 [110, 111]. The fourth layer in barrel and the third disk in end-cap were added in the pixel upgrade phase during the technical stop providing an additional measurement point in both regions.

The strip tracker and the pixel detectors are operated at a temperature of about -15 °C and -20 °C, respectively. This is necessary to minimize the damage caused by ionizing radiation to silicon detectors, and requires the presence of an efficient cooling system to absorb the heath produced by the on-board electronics.

---

6 In colliding-beam experiments, the Impact Parameter IP is the closest distance from the primary vertex of the collision to a given track.

The silicon strip sub-detector

The outermost region uses sensors comprised of silicon micro-strips instead of pixels with two different strip pitches (i.e. the distances between neighbouring strips) depending on the distance from the interaction point. The silicon pixel sub-detector consists of a total of about 9.3 million strips, covering an active area of about 198 m^2 [105]. It extends the tracking measurements to $|\eta| < 1.6$ in the barrel and to $|\eta| < 2.5$ in the end-cap region and its modules work in a similar way of the pixel modules. The track transverse momentum resolution varies between 0.7 % to 1.5 % for tracks with 1 GeV to 100 GeV respectively. It consists of 10 cylinder layers in the barrel region around the pixel sub-detector:

- four-layer TIB cover r < 55 cm and |z| < 118 cm and provide a single-point resolution of 13-38 μm in the (r,ϕ) plane and 23 μm in the z-direction;

- six-layer TOB covers r > 55cm and |z| < 118 cm providing a resolution of 18-47 μm in the (r,ϕ) direction and 47 μm in z.

Moreover, 24 disks of end-cap modules containing concentric rings of silicon strip modules, 12 on each end of the barrel cylinders are installed:

- three TID providing the same resolution and coverage as TIB detectors;

- nine TEC disks with the same resolution as TOB detectors. They cover the region 124<|z|<282 cm.

Figure 2.13 shows the material budget[7] of the CMS tracker as a function of η, as estimated from simulation (with an accuracy better than 10 %).

2.2.4 The CMS calorimeters

The calorimeter sub-detector lies next to the tracking system and still within the solenoid magnet. These detectors are designed to measure the energy of both neutral and charged particles. It measures the energy lost by incident particles due to interactions with the detector material which help in the identification problem of the incident particle by studying the energy deposition profile in the detector. However, calorimeter cannot detect neutrinos which escape the detector, but their presence can be inferred as an apparent energy imbalance in the collision. The CMS detector includes three calorimetry sub-detectors: the ECAL,the HCAL and **H**CAL **F**orward detector (HF) calorimeter which is designed to detect the energies of particles in the very forward regions of the detector.

---

7 The radiation length X_0 is commonly used to determine the amount of material traversed by particles in a detector, i.e. its *thickness.*The detector thickness expressed in units of radiation length X_0 is called the material budget.
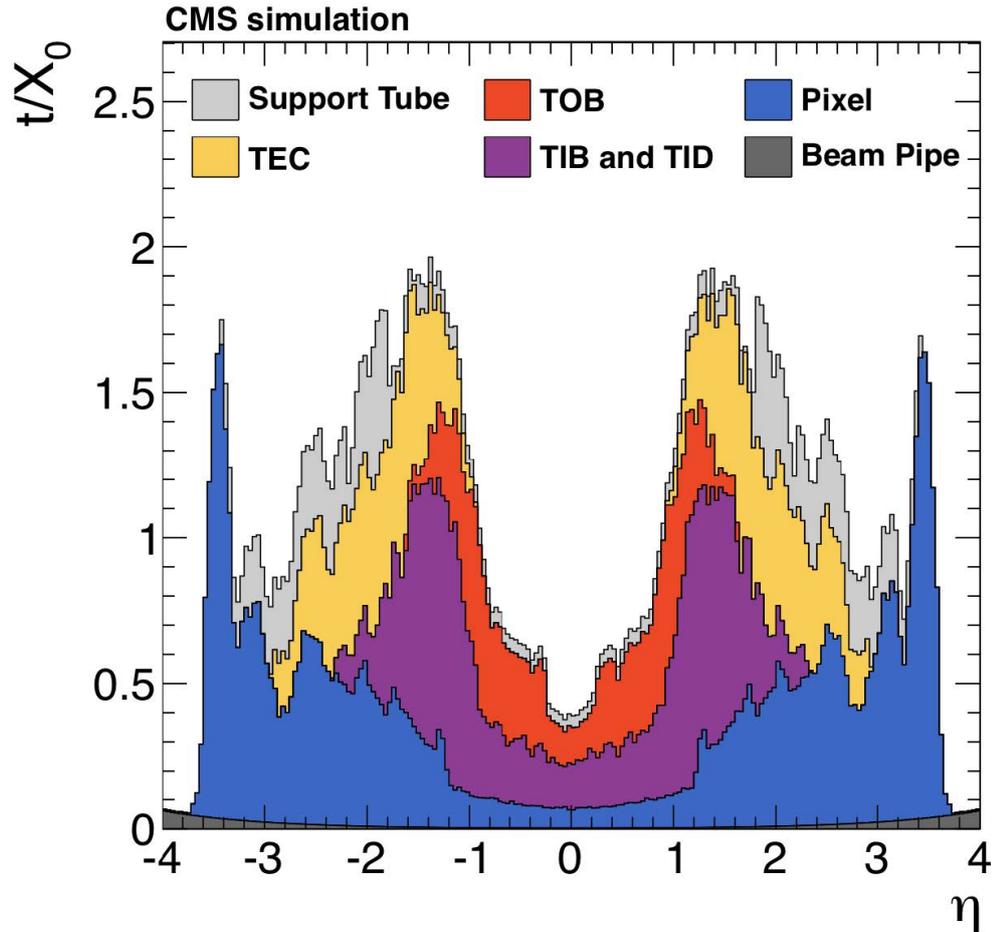
Figure 2.13: The contribution to the total material budget of each of the sub-systems that comprise the CMS tracker is shown, together with contributions from the beam pipe and from the support tube that surrounds the tracker [105].

Electromagnetic calorimeter

The electromagnetic calorimeter ECAL, is designed to measure the energy of incident electrons and photons. The CMS ECAL is a homogeneous and highly granular calorimeter constituted of lead tungstate inorganic crystals ($PbWO_4$). The energy measurement is based on the conversion of the incident electron or photon to an electromagnetic shower, that interacts with the crystal material producing scintillation light that is subsequently converted into photoelectrons by a photosensitive device. The crystals are at the same time the dense interacting material, which is able to contain the EM shower, and the active scintillating medium, resulting in an excellent energy resolution. The choice of $PbWO_4$ is motivated by its high density (8.28 g/cm^3), small radiation length ($X_0 = 0.89$ cm)

and short Molière radius[8] ($R_M = 2.2$ cm). These parameters ensure an excellent containment of the electromagnetic shower within the crystals, which have a length of approximately 25 X_0. The lead tungstate is radiation hard and about 80% of its scintillation light is produced within 25 ns (scintillation decay time), making it ideal for the high instantaneous luminosity collisions of the LHC and the proton bunch spacing. The disadvantage of this material is the relatively low light yield, corresponding to about 30 photons per MeV of deposited energy, which calls for the usage of photodetectors with internal amplification. The barrel part of the ECAL , abbreviated as **E**CAL **B**arrel (EB), is constituted by 61 200 crystals with cross sections from $22 \times 22 \ mm^2$ (front) to $26 \times 26 \ mm^2$ ((rear) and a length of 23 cm (25.8 X_0), and ensures the coverage of the region $|\eta| <$ 1.479. The two end-caps **E**CAL **E**ndcap (EE) are each made of 7324 crystals of a frontal transverse section from $28.62 \times 28.62 \ mm^2$ (front) to $30 \times 30 \ mm^2$ (rear) and a length of 22 cm (24.7 X_0), and extend the coverage up to $|\eta| < 3.0$.

The layout of the crystals in the ECAL is illustrated in Figure 2.17 with the arrangement of crystal grouped into modules and super-modules in barrel and two half-disks or dees in the end-caps, with pre-shower in front Figure 2.14. In both the barrel and the end-caps, crystals are mounted in a quasi-projective geometry, with their axes being tilted of with respect to the direction that points to the nominal interaction region. This ensures that no particle escapes the ECAL active volumes from the interstices between the crystals. The crystal scintillation light is read-out by detectors that are designed to work in the high magnetic field to which they are exposed and to be resistant to the radiation. The EB is instrumented with silicon **A**valanche **P**hoto-**D**iodes (APDs) while **V**acuum **P**hoto**T**riodes (VPTs) are deployed in the two EE. The signals from these photodetectors are amplified and shaped by the front-end electronics and sampled at a frequency of 40 MHz with a 12-bit analog-to-digital converter. The ECAL is operated at a temperature of 18 °C. To control the changes in the light yield of the crystals and of the photodetectors on the temperature, it is stabilized within 0.05 °C in EB and 0.1 °C in EE by a cooling system.

A sampling **E**M Pre**S**hower (ES) detector is installed in front of the two end-caps. It consists of a sampling calorimeter per end-cap, made up of two layers of lead absorber to initiate EM showers from incoming electrons and photons, followed by two 2-mm silicon strip detectors to measure the energy deposit and the transverse shower profile. The silicon strip detectors are placed orthogonal to each other to provide a two dimensional reconstruction of the electromagnetic shower initiated in the lead plates. The pre-shower system is fundamental to identify and reject the π_0 mesons decaying into two photons and to improve the

---

8 The Molière radius approximates the radius of a cylinder coaxial with the shower axis which contains 90% of the EM shower energy.
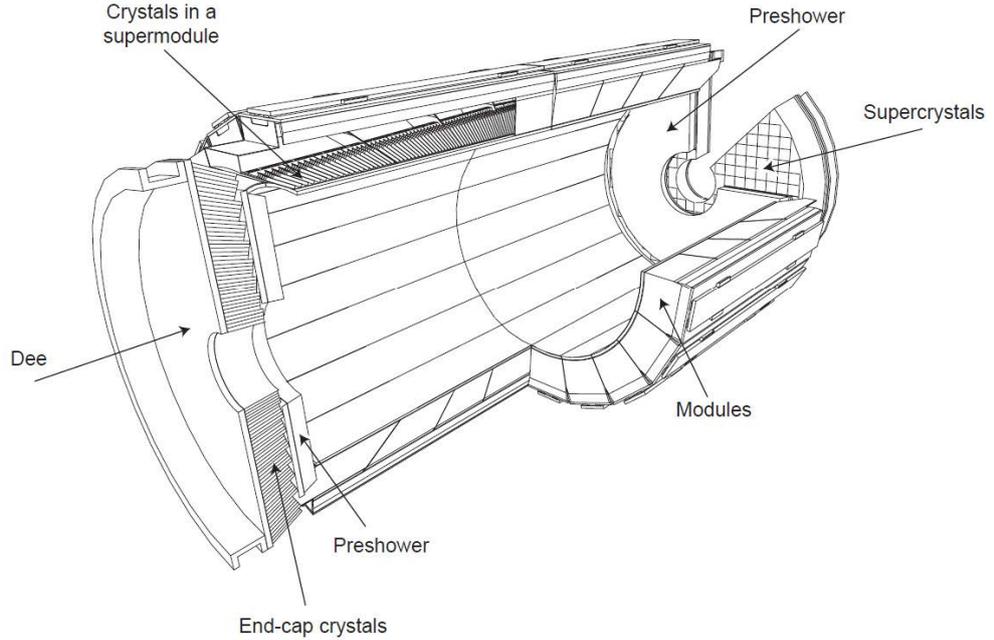
Figure 2.14: The CMS Electromagnetic Calorimeter (ECAL). The barrel section comprises 36 supermodule, each containing 4 modules while each endcap includes two half-disks or dees in each endcap [112].

measurement of the position of electrons and photons, because it has a higher granularity than the EE. The energy resolution of the CMS ECAL increases with the particle energy itself, because of the reduced impact of two of the three main effects that determine resolution of a generic calorimeter. The first one is a **stochastic term** that depends on the number of scintillation photons produced in the interaction as that is in turn proportional to the incident particle energy E. A second term accounts for the **noise** coming from the electronic of the readout chain and depends on the detector technique and on the features of the readout circuit (detector capacitance, cables, etc.). Finally, a third term is related to detector inhomogeneities (due to detector geometry,temperature gradients,radiation damage etc.), instrumental effects and calibration uncertainties resulting in an error that amounts to a constant term. The combined effect of these three factors results in a dependence of the ECAL energy resolution on the particle of energy E as:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E(GeV)}} \oplus \frac{12\%}{E(GeV)} \oplus 0.3\% \tag{2.6}$$
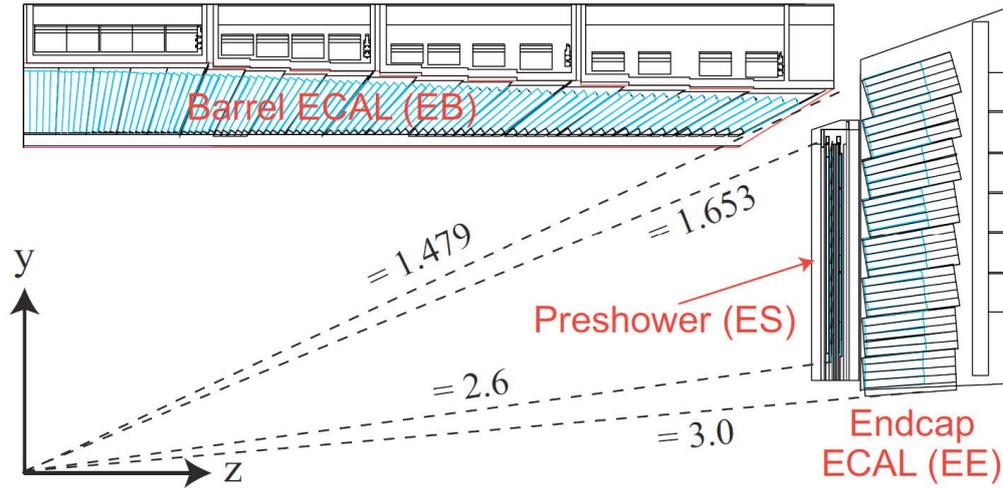
Figure 2.15: Longitudinal view of part of the CMS electromagnetic calorimeter showing the ECAL barrel and an ECAL endcap, with the preshower in front. [112].

where the symbol $\oplus$ indicates the sum in quadrature of the three terms. The ECAL energy resolution has been measured on one barrel super module during a test beam performed with incident electrons [113].

Hadronic calorimeter

Hadrons typically traverse the ECAL volume without being stopped. The HCAL, is designed to absorb them within its volume and measure their energy. Compared to electron and photon interactions in the ECAL, hadron energies are intrinsically more difficult to measure from hadron showers induced in the HCAL. Indeed, nuclear and hadronic interaction result in non-Poissonian (i.e. rare) effects in the shower development, where many undetectable particles can also be produced. The presence of π_0 decaying to photon pairs also results in an EM component of the shower with a different response from the HCAL itself. All these effects limit the intrinsic resolution on hadron energies, which can be improved offline with the usage of the **P**article **F**low (PF) reconstruction techniques detailed in Section 3.4. Despite these limitations, the HCAL is an essential element in the reconstruction of final states containing hadron jets or non-interacting particles such as neutrinos. It consists of 70,000 tiles in total. The HB and HE sections of the HCAL instrument respectively the regions $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$. Both the HB and HE are sampling calorimeters composed of a Brass absorber (70% Cu, 30% Zn, density of 8.53 g/cm^3) and of active plastic scintillating tiles sandwiched between the absorbers. Brass was chosen for its short interaction length $\lambda_I = 16.42$ cm and because it is a non-magnetic material. The scintillation
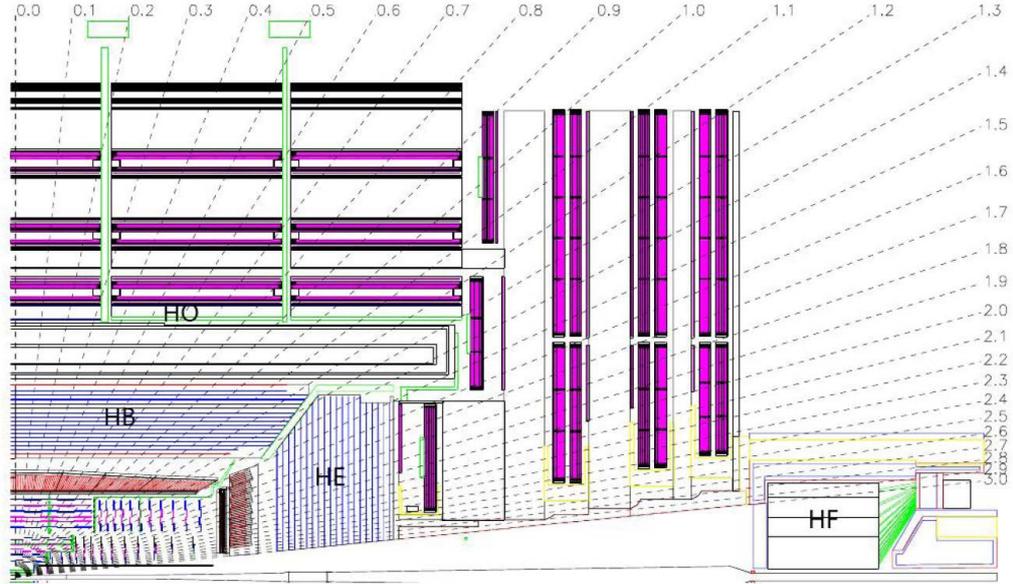
Figure 2.16: Longitudinal view of the HCAL layout, where are visible the barrel (**H**CAL **B**arrel detector (HB)) and endcaps (**H**CAL **E**nd-cap detector (HE)) detectors located inside the solenoid, the outer detector (**H**CAL **O**uter calorimeter (HO)) outside the solenoid, and the forward calorimeter (HF) [114].

light is collected by **W**avelength **S**hifting **F**ibers (WLFs) embedded in the tiles to modify the frequency of the scintillation light and optical fibers to transfer the light to hybrid photodiodes **H**ybrid **P**hotoDiodes (HPDs). Each readout cell is formed by the addition of a "tower" of scintillating tiles with a transverse dimension $\Delta\eta \times \Delta\phi$ of about 0.087×0.087 in the HB and of about 0.17×0.17 in the HE. As the HCAL is located between the ECAL and the internal surface of the solenoid, the limited space does not allow for a full containment of the secondary interaction shower. The detector is completed by a HO located outside the solenoid as a tail catcher, external with respect to the cryostat and within the return yoke, using the iron as absorber, which extends the total interaction depth [9] to about 11 λ_I, where the constant indicates the average interaction length of hadrons in the calorimeter. The energy measurement in the forward region is complemented by the HF, that is located 11.2 m away from the interaction point along the z-axis and measures hadron interactions from $|\eta| < 3.0$ up to $|\eta| < 5.2$ and it is useful to identify and reconstruct very forward jets (it is not shown in Figure 2.18). Because of the higher radiation levels in the forward region, the HF uses a radiation-hard technology and it is composed of steel absorbers and

---

9 The nuclear interaction length λ_I is the mean distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction. It depends on the inelastic cross section σ like : $\lambda_I = \frac{A}{N_A \sigma \rho}$ where A,ρ are the atomic weight and the density of the target, N_A is the Avogadro number.

quartz fibres that produce Cherenkov light, which is measured by **P**hoto**m**ultiplier **T**ubes (PMTs). Fibres of two different lengths are installed to distinguish the electromagnetic component of the shower which deposit a large fraction of their energy in the first 22 cm of the calorimeter, from those generated by hadrons, which produce, on average, nearly equal signals in both calorimeter segments (respectively long 22 and 143 cm).

The global layout of the HCAL is illustrated in Figure 2.17 while the location of HCAL and ECAL with respect to the CMS magnet is shown in Figure 2.18. The overall HCAL performance is dominated by the imperfect containment of the hadronic shower, which results in a resolution sampling term of about 110% and a constant term of 9%, as measured in pion test beams [115].
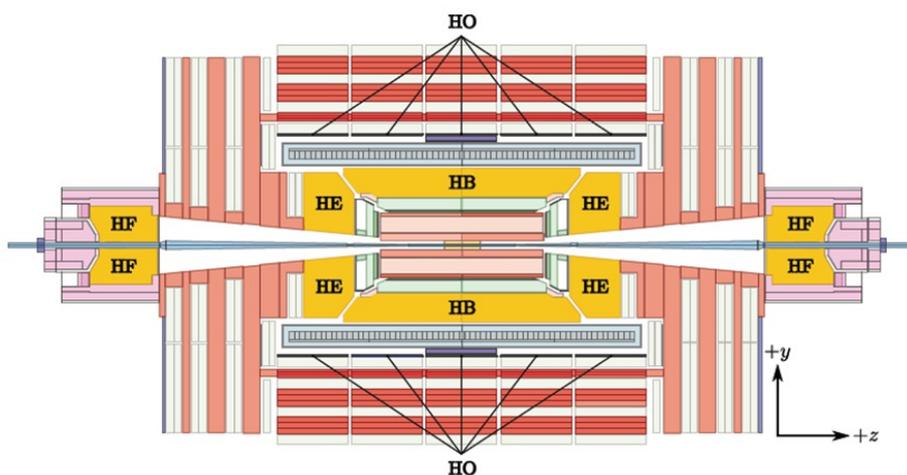
Figure 2.17: A view of the CMS detector in the (y,z) plane with the structure of the hadronic calorimeter labeled. [116].

2.2.5 The CMS muon system

Because of their typical energy, muons produced in collisions at the LHC have minimal energy loss rates. As a consequence, they traverse the ECAL, the HCAL, and the solenoid volumes without being stopped, and are identified and measured in the muon detectors located in the outermost part of CMS. Indeed, the main tasks of the CMS muon system are the muon identification and the measurement of p_T and charge of muons with energies ranging from few GeV up to few TeV. Additionally, it provides a robust trigger for events that involve these particles and a precise time measurement of the bunch crossing [118].

The system is placed outside the magnet and the detector stations are integrated into the iron return yokes so that the 3.8 T magnetic field, inside the solenoid,
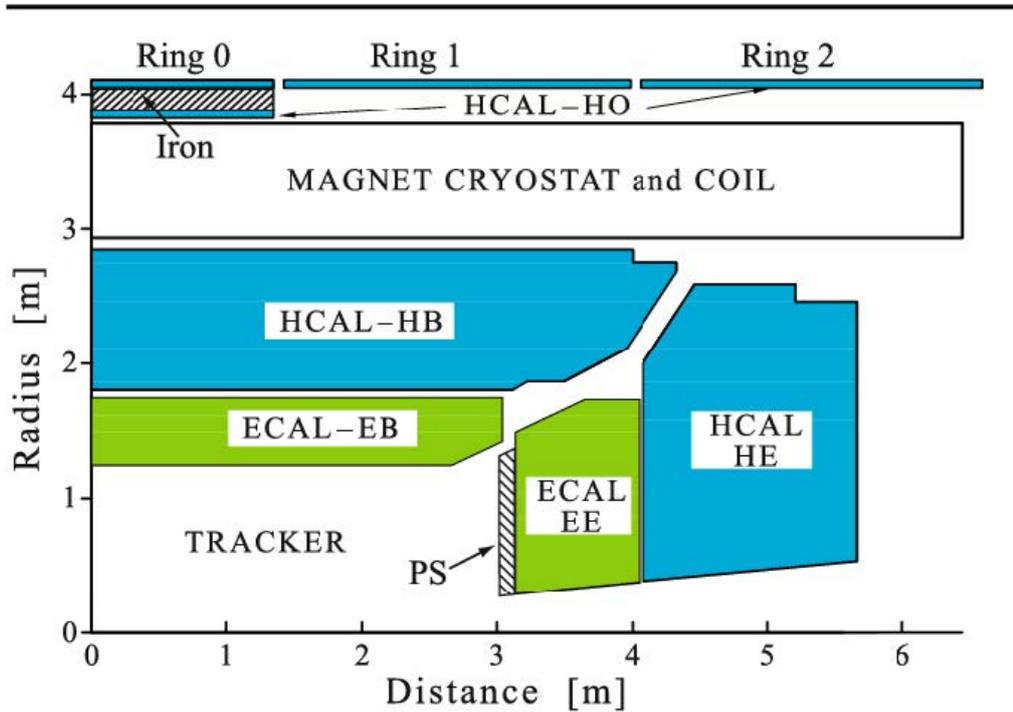
Figure 2.18: Location of the hadron calorimeter in and around the CMS magnet. HCAL HB and HCAL HE are the hadron calorimeters. ECAL EB and ECAL EE are the electromagnetic calorimeters and **P**arton **S**hower (PS) is pre-shower detector [117].

and the 1.8 T average return field, bend the muon tracks in the transverse plane, thus allowing the measurement of their p_T. Furthermore, because of the large amount of material in front of the muon chambers, also due to the presence of the magnet, the muon system is well shielded from charged particles other than muons,making their identification easier [119]. The R-z cross section of a quadrant of the CMS muon spectrometer is shown in Figure 2.19.

The muon system is made up of three different kinds of gaseous detectors, which assure robustness and redundancy. The detectors are the DT, in the barrel ($|\eta| < 1.2$), CSC in the endcaps ($0.9 < |\eta| < 2.4$) and RPC, in the barrel and in the endcaps (pseudorapidity region with $|\eta| < 1.9$). Each of these detector technologies is installed in a specific region of the experiment (characterized by its own value of particle flux and magnetic field), because their have different rate capabilities[10] and not all of them are able to work properly inside a strong magnetic field. In particular, the DT, which do not have a high rate capability and require that the trajectory of the particles inside them is modified as less

---

10 In experimental particle physics, the rate capability is a important detector parameter being inversely proportional to the time needed, after the arrival of a particle to get ready to receive the next.

as possible by the magnetic field, are used only in the barrel region, where the residual magnetic field and the muon and neutron induced background rate are low.

On the contrary, CSC have been installed in the endcaps, where there is a higher residual magnetic field and a large particle rate, because they are most suitable for these radiation conditions and they can work without any problem inside the magnetic field. Both DT and CSC provide a very good spatial resolution for the measurement of the p_T of charged particles.

In addition to them, RPC are placed in both regions (barrel and endcaps), in order to guarantee the redundancy of the system. Owing to their very good timing, these detectors mainly contribute to the trigger. Moreover, DT, CSC and RPC have different sensitivity to the backgrounds, assuring the robustness of the system. In this region, the background is composed mainly by secondary muons produced in π and K decays, or coming from punch-through hadrons (due to hadron shower remnants penetrating through the calorimeters and reaching the muon system) and from low energy electrons originating after slow neutron capture by nuclei, with subsequent photon emission.
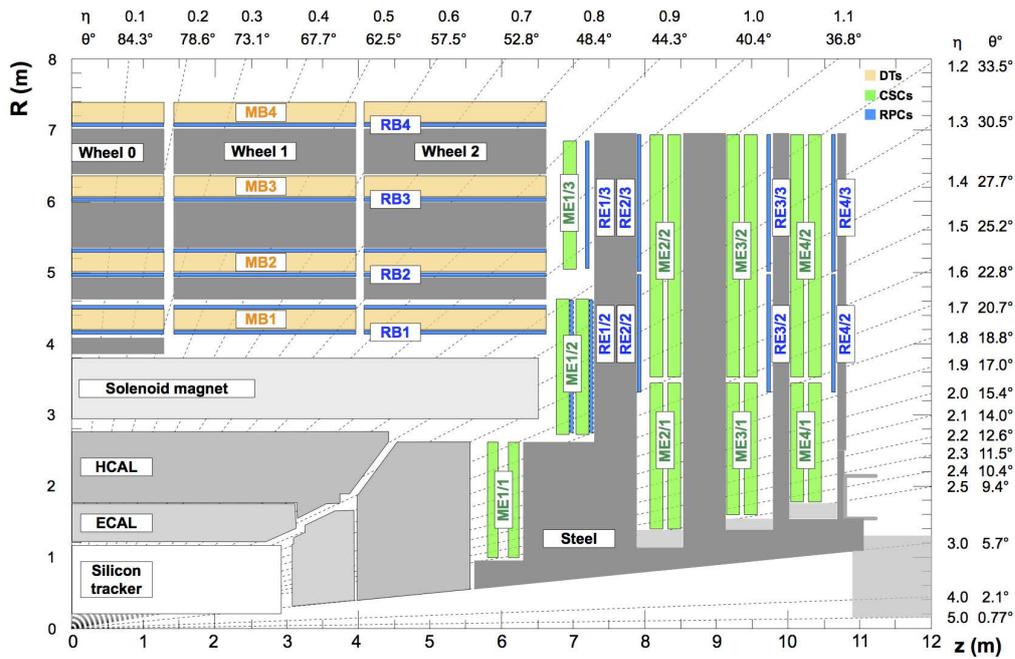
Figure 2.19: A quadrant of CMS muon system with the axis parallel to the beam (z) running horizontally and the radius (R) increasing upward. The three different sub-detectors are highlighted: Drift Tubes (in yellow) are installed in the Muon Barrel (MB), Cathode Strip Chambers (in green) are placed in the Muon Endcap (ME) and Resistive Plate Chambers (in blue) are present in both, barrel and end-caps (and labelled as RB and RE). The dark grey areas are the steel flux-return disks of the magnet.

Drift tubes

The **M**uon **B**arrel (MB) system of detectors is made up of four stations, arranged in coaxial cylinders around the beamline and interleaved with the iron yoke. It is also divided into five wheels along the beam direction following the five wheels of the return yokes. In this region, there are in total 250 drift chambers. The basic element of a DT is the drift cell, shown in Figure 2.20. It is a tube with a rectangular cross section, filled with an Ar/CO_2 mixture (85/15) and operating at a gas gain of 10^5. The cathodes stripes are placed along the shorter sides of the rectangle, while the anode wire is in the middle of the cell. A charged particle, passing through the detector, ionizes the gas and the produced primary electrons drift towards the anode wire. Since the drift velocity in the operating conditions is known and constant (because the geometry of the cell guarantees a uniform electric field), from the measurement of the electrons drift time is possible to obtain the position of the ionizing particle. A single drift cell has a cross-section of $42 \times 13\ mm^2$ and wire length 2-3 m. It is characterized by a maximum drift time of ~ 400 ns and a single point resolution of about 200 μm, resulting in a resolution of 80 - 120 μm for the global chamber measurement [119]. Each DT is composed of two or three Super-Layers (SL), each made of four stacked layers of drift cells. The orientation of the anode wires differs among the SLs, in order to provide information regarding different coordinates. In the outer SL, the wires are parallel to the beamline, while in the inner one they are orthogonal to the beamline. The former allows a track measurement in the plane (r,ϕ), in which the low residual magnetic field bends the tracks, while the latter measures the z-coordinate.

Cathode Strip Chambers

The tracking measurement of muons in the two end-caps is the main task of the Cathode Strip Chambers (CSC), which are arranged in the **M**uon **E**ndcap (ME) system in four stations. The CSC is a multi-wire proportional chamber, in which the cathode plane is segmented into strips perpendicular to the wire direction. These detectors are operated at a gain of 7×10^4, using a gas mixture of $Ar/CO_2/CF_4$ (40/50/10). Each chamber has a trapezoidal shape and is made of seven cathode planes stacked together, forming six gas gaps $\sim$10 mm thick, each containing a plane of anode wires, as displayed in Figure 2.21. In the same Figure is illustrated the mechanism of formation of the signal: when a muon passes through the chamber, it produces an avalanche in the gas by producing electron-ion pairs,and it induces signals both on the wires and on the cathode strips. These two contributions are combined in order to obtain the position of the ionizing particle, since the wires provide information on the radial coordinate,
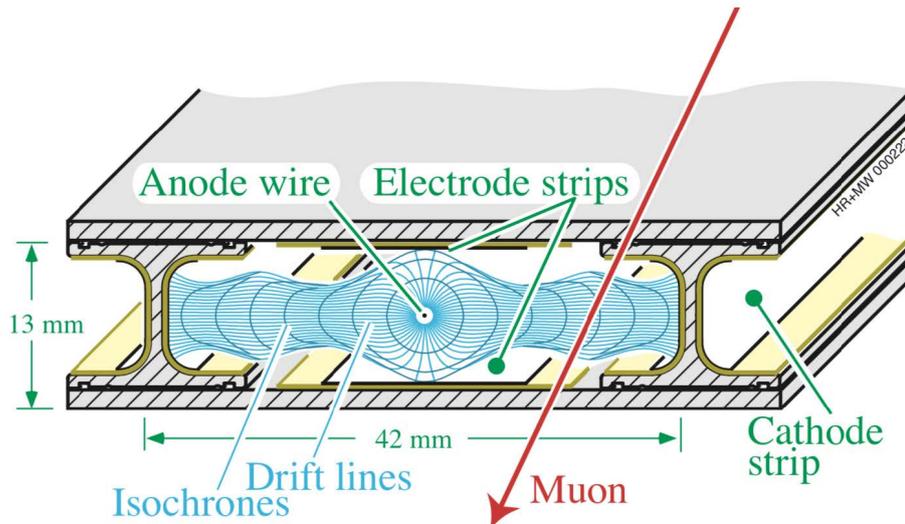
Figure 2.20: Section of a drift cell of a Drift Tube detector, showing the anode wire and the cathode strips, as well as the drift lines and the isochrones (contours of equal drift times) [120].

while the cathode planes, segmented into radial strips orthogonal to the wires, measure the z-coordinate. The CSC is a fast detector, capable of identifying the bunch crossing of a pp collision, and achieves a spatial resolution of 40 - 150 μm depending on the CSC station in consideration [119].

Resistive Plate Chambers

The main goal of the 1056 RPC, installed both in the barrel and in the endcaps of CMS, is to provide a fast trigger signal, while adding, at the same time, redundancy to the muon spectrometer. The RPC are gaseous parallel-plate detectors characterized by a modest spatial resolution of 0.8 - 1.2 cm but an excellent time resolution of the order of tens of ns. Indeed, while the CMS RPC detectors have an intrinsic time resolution of the order of 1.5 ns [121], the electronics system records the RPC hits information in steps of one **B**unch **C**rossing (BX) (25 ns), degrading the full timing resolution of the detector, but still able to assign the hits in the correct pp BX. A single RPC consist of two parallel planes made of Bakelite (a very resistive resin) of 2 mm thickness, externally coated with graphite and separated by a 2 mm wide gas gap,filled with a gas mixture of $C_2H_2F_4$ (Freon)/i - C_4H_{10} (isobutane)/SF_6 (96.2, 3.5, 0.3). In CMS two RPC are combined in order to improve their efficiency of detection. The signals produced by the avalanches, generated by the ionization of the gas during the passage of a charge particle, are collected on a set of readout aluminium strips, placed between the two chambers, as shown in Figure 2.22. RPC can operate in
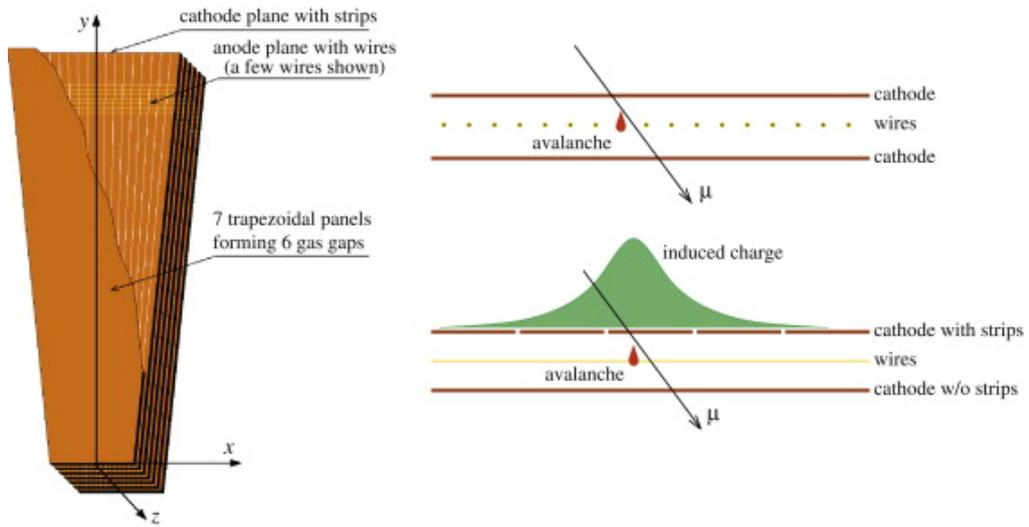
Figure 2.21: Schematic view of cathode strip chamber (left) and the principle of operation (right) with cross-section shows the wires (right-top) and cross-section shows the strips (rightbottom) [118].

two different modes: a *streamer mode*, with a strong electric field that produces localized gas discharges in the region near the passage of the ionizing particle, or an *avalanche mode*, in which the electric field is weaker than the previous one. The former mode allows only few counts per unit area, while the latter one, because of the reduced charge generated in the ionization, is characterized by an increased counting capacity of the chamber. For this reason inside CMS the RPC operate in avalanche mode, allowing the detectors to sustain higher rates [119].
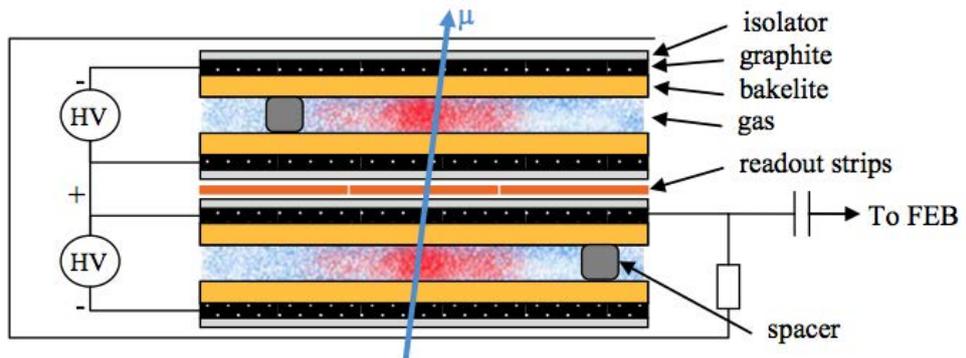
Figure 2.22: Schematic view of a dual RPC detector [121].

2.3 The CMS trigger system

Proton-proton collisions occur in the centre of the CMS detector every 25 ns, and generate a huge amount of information in the detector, corresponding to about 70 Terabytes of data every second. No technology exists nowadays to read out, store and analyse such volumes of data. However, the large majority of the collisions result in low-energy proton-proton interactions that are not interesting to pursue the physics programme of CMS. As it can be observed in Figure 2.23, even the most frequent processes studied at the LHC have a cross section of the order of 10^5 pb, that is 5 orders of magnitude smaller than the total proton-proton interaction cross section $\sigma_{pp} \sim 10^{11}$ pb [122].
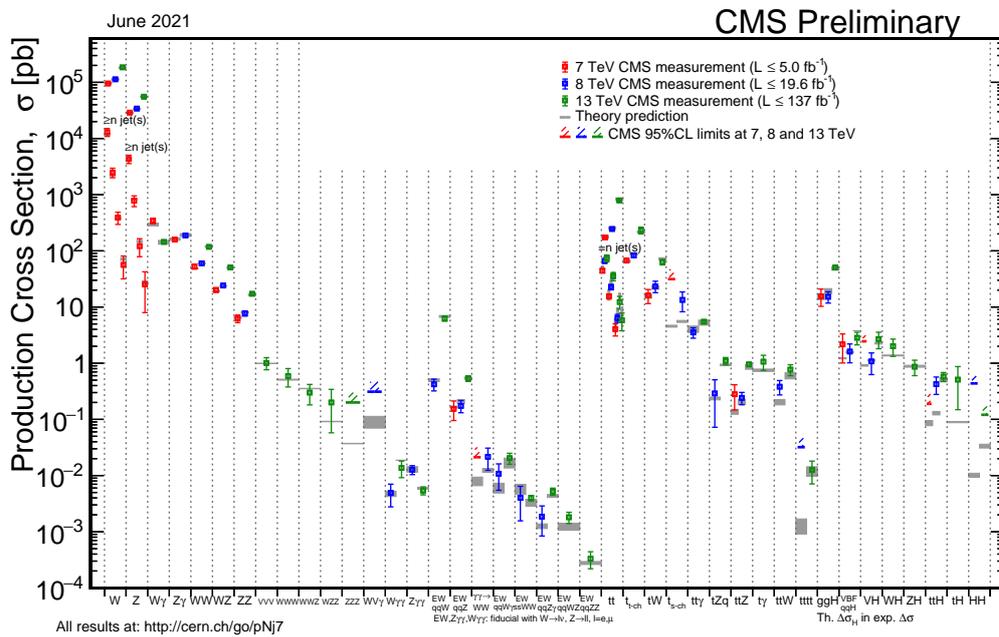
Figure 2.23: Summary of the cross section measurements of Standard Model processes at CMS. Values are to be compared to the total pp interaction cross section of about 10^{11} pb [123].

The role of the trigger system of the CMS experiment is to identify and select the interesting collision events for physics analysis from the huge amount of events produced by LHC collisions, thus reducing the acquisition rate by a factor of about 10^5. The trigger lies between the **online** data-taking and the **offline** data analysis ,and must at the same satisfy the technical constraints of the former and ensure a high efficiency for the latter [124]. More specifically, at the nominal operational conditions of LHC, the bunch crossing rate is 40 MHz (corresponding to 25 ns) and every bunch crossing gives rise to ∼20 proton-proton collisions and hence that result in about 800 million collisions per second. The

data corresponding to events is stored in pipelines that can retain and process information from many interactions at the same time. To identify the particles from the same event and not confuse particles from two different events, the detectors must have very good time resolution and the signals from the millions of electronic channels must be synchronized. One event needs ~ 1 MB of memory to be recorded on the tape and therefore the final memory requirement is 70 TB per second. This huge amount of data cannot be stored and a reduction process is needed. The decision of the selection of events to be stored or not is taken by the trigger system within a short time ~ 25 ns such as per the LHC requirements. The trigger system consists of two independent levels, the **L**evel-**1** (L1) and the HLT described in the following.

L1 trigger

The L1 trigger is completely hardware-based system. It reduces the event rate from from the initial 40 MHz to 100 kHz. The L1 takes the decision depending on the informations from the calorimeter and the muon system while the tracker information doesn't participate because the track reconstruction time exceeds the time limits of L1 decision [125, 126]. The L1 trigger searches for key signatures of interesting events: leptons, photons, hadronic jets, and p_T, known as trigger objects or trigger primitives (TPs). Those trigger objects are produced based on the energy deposit in the calorimeters, and track segments or hit patterns in the muon chambers. The L1 triggers from muon and calorimeter are then combined in the **L**1 **G**lobal **T**rigger (L1 GT) which decides whether to pass or reject the event. The event information is sent to the HLT system if the L1 trigger decision is positive otherwise the event will be deleted. Figure 2.24 shows the schematics of the L1 trigger system.

High Level Trigger

Opposite to the L1 trigger, the HLT system is software-based [127]. It uses advanced software system consisting of an array of multi-core computers. It is able to reduce the event rate from 100 kHz to $\sim$1 kHz that can be written on tape. In order to pass the HLT, an event needs to satisfy the requirements of at least one of the HLT menu paths. The events which passed the HLT are recorded permanently on the disk by the DAQ system for further physics analysis and transferred to CERN Tier0 (T0) storage system. Then **D**ata **Q**uality **M**onitoring (DQM) checks the quality of the recorded data and labeling the data-sets either good or bad. At the end of the DQM chain, a list of certified data-sets is produced to be used later for physics analysis.
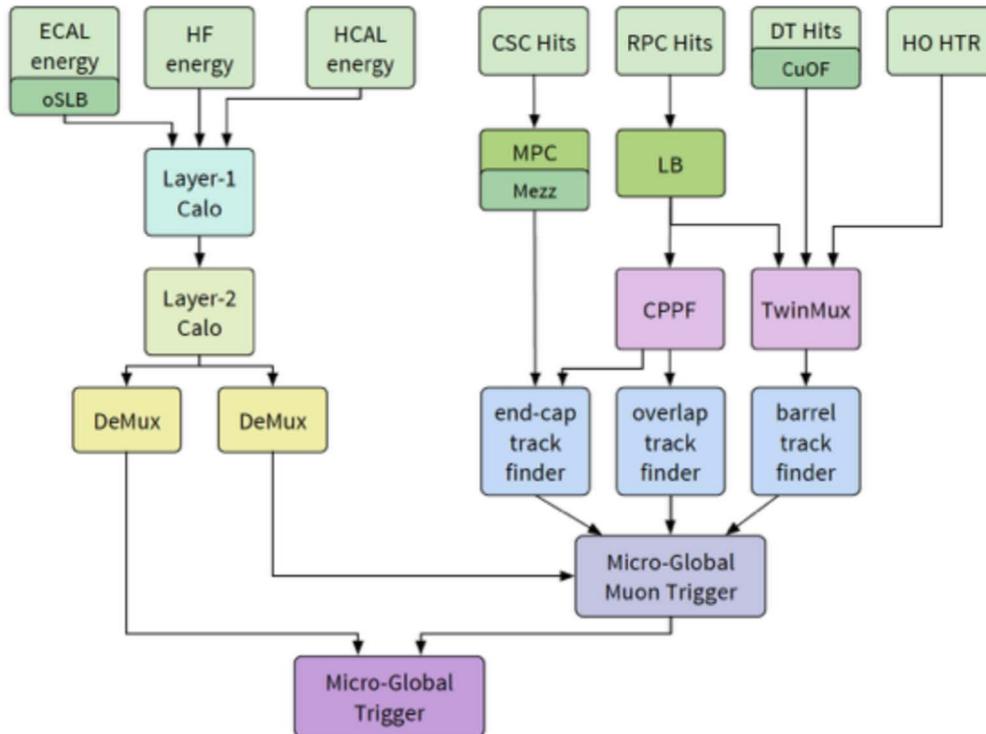
Figure 2.24: Overview of the CMS L1 trigger.

2.4 The WorldLHC Computing Grid (WLCG)

Events that have fired the HLT are stored and then reprocessed in order to be analysed. In order to deal with the very demanding requirements of storing, distributing and analysing the $\sim 50 - 70$ Petabytes of data expected every year of operations from the LHC at CERN on the Franco-Swiss border, an infrastructures worldwide distributed, called **W**orld LHC Computing Grid (WLCG) has been created [128] in the 1990s. The WLCG combines the computing resources of about 900 000 computer cores and 1 Exabyte of storage from over 170 sites in 42 countries, producing a massive distributed computing infrastructure that provides more than 12 000 physicists around the world with near real-time access toLHC data, and the power to process it [130]. The WorldwideLHC Computing Grid is partnered with **E**uropean **G**rid **I**nfrastructure (EGI), **O**pen **S**cience **G**rid (OSG), and **N**ordic e-**I**nfrastructure **C**ollaboration (NEIC). Data pours out of the LHC detectors at a blistering rate. Even after filtering out 99% of it, in 2018 we gathered 88 petabytes of data. That's 88 million gigabytes, the equivalent to around 22 million high-definition (HD) movies. It runs over 2 million tasks per day and, at the end of the LHC's Run II, global transfer rates regularly exceeded
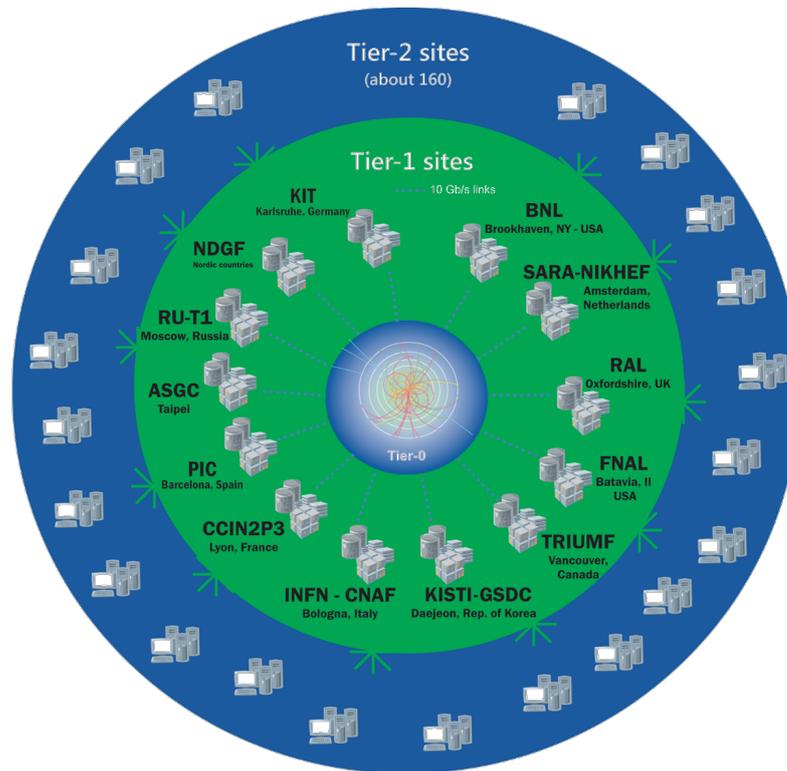
Figure 2.25: Schematic representation of the WLCG distributed infrastructure. It is organized in different levels, called Tiers, connected via high-speed networks. The figure is taken from The figure is taken from Ref. [129].

60 GB/s. These numbers will increase as time goes on and as computing resources and new technologies become ever more available across the world.

The WLCG is arranged in four layers, or "Tiers", called 0, 1, 2 and 3. Each Tier is made up of several computer centres and provides a specific set of services. Between them the tiers process, store and analyse all the data from the Large Hadron Collider LHC. Tier 0 is the CERN Data Centre. All of the data from the LHC passes through this central hub, but it provides less than 20% of the Grid's total computing capacity. CERN is responsible for the safe keeping of the raw data (millions of digital readings from across the detectors), and performs the first pass at reconstructing the raw data into meaningful information organised in different groups (*Primary Datasets*) according to the trigger path with which they were acquired. Moreover, it converts these data into data formats useful for analysis: **RECO**nstructed (RECO) and **A**nalysis **O**bject **D**ata (AOD)[11]. Tier 0

---

11 The output of the Reconstruction phase is a set of low/medium/high level physics objects. Currently, the total amounts to 4 MB/ev of content, and is defined as the RECO data tier. This is too much for the present disk capabilities, and mostly useless: only very few analysis need all the Tracker Silicon hits, all ECAL Clusters etc. CMS collaboration do not save RECO any more apart from for a few debug samples. Analysis objects are selected out of RECO, and saved as the AOD Tier corresponding to < 500 kB/ev. This data format still contains most of the objects needed for analysis. For those few analyses which need more, CMS defines the SKIMS, very specific samples defined specifically for one analysis. The AOD was the most

distributes the raw data and the reconstructed output to Tier 1s, and reprocesses data when the LHC is not running.

Tier 1 consists of 13 computer centres (see Figure 2.25) large enough to store LHC data. They provide round-the-clock support for the Grid, and are responsible for storing a proportional share of raw and reconstructed data, as well as performing large-scale reprocessing and storing the corresponding output, distributing data to Tier 2s, and storing a share of the simulated data that the Tier 2s produce. Optical-fibre links working at 10 gigabits per second connect CERN to each of the 13 major Tier 1 centres around the world. This dedicated high-bandwidth network is called the **LHC O**ptical **P**rivate **N**etwork (LHCOPN).

Individual scientists can access the Grid through local (or Tier 3) computing resources, which can consist of local clusters in a university department or even an individual PC. There is no formal engagement between WLCG and Tier 3 resources.

Tier 2s, the core of the Grid, are typically universities and other scientific institutes that can store sufficient data and provide adequate computing power for specific analysis tasks. They handle a proportional share of the production and reconstruction of simulated events. There are around 155 Tier 2 sites around the world. In particular, the Tier 2 which is in Bari, called **R**ete di **Ca**lcolo per **S**uperB (ReCaS), has been extensively used for the analysis described in this thesis.

---

widespread data tier used in Run I analysis. Still, 500 kB/ev turned out to be too big when projected to the Run II. **MiniA**nalysis **O**bject **D**ata (MiniAOD) are the default / advised data tier for Run II analysis. W.r.t. to AOD, they drop track hits, very small p_T tracks, reduce precision of objects (we do not need 12 digits on an off diagonal term of an error matrix). They are ~ 50 kB/ev, they would be "technically ok" for storage space for the next 10 years and they can currently cover 95% of the analyses; the rest (5%) still needs either AOD or even Skims or NanoAOD. The latter are the last incarnation of centrally produced data tiers. They consist of a ROOT flat ntuple which mostly drops tracks by knowing only their number and are ~ 1 kB/ev. It is our bet for a faster , easier and less error prone analysis global effort.

Part III

DATA ANALYSIS

Event Generators, Simulation and Physics Objects Reconstruction at CMS

3

"It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong."

– Richard Feynman

This chapter is intended to give an overview of the basics of the description of the event generation, simulation of the particles' interaction in the CMS detector starting from pp collisions to the reconstruction of the physics objects used in the analysis [131, 132].

3.1 Event Generation

The generation of events is the first step of the simulation process chain which uses MC technique to simulate experimental events that happen in actual collider experiments. At LHC, the protons are accelerated to high energy moving in opposite directions to collide at the center of the detector. The interaction takes place between the proton constituents, the partons, which is a term that globally indicates quarks and gluons inside the proton. In each collision, not all the interactions produce high energy particles in the detector. However, sometimes two partons from separate protons interact very strongly and produce an interesting event for the study. The types of the interactions that can take place in the collision are shown in Figure 3.1. The **hard scattering** processes result from the interaction between the constituents of two colliding protons having high momentum exchange between them and having product particles with high momentum in direction perpendicular to the protons' original direction of motion. The incoming and outgoing partons of the hard scattering process may also emit initial and final state radiations, creating PS which is taken into account in the event generation process. An **underlying event** is a soft scattering process panying the hard process. It happens between the proton's remnants

that were not involved in the hard scattering. These partons may do relevant soft scattering between themselves (called Multiple Parton Scattering), may also radiate gauge bosons before or after colliding with each other called **I**nitial **S**tate **R**adiation (ISR) and **F**inal **S**tate **R**adiation (FSR) respectively. Quarks and gluons can emit additional radiation via the strong interaction and ISR and FSR result in the creation of jets close to the direction of incoming and outcoming hadrons. The LHC accelerates and collides bunches of protons and each bunch contains 10^{11} protons. **Pileup processes** refer to any interactions that happen between protons which were not involved in the hard scattering within the bunch crossing. LHC is running with an increase of the instantaneous peak luminosity which increases the chance of the number of interactions per bunch crossing and hence the pileup interactions. I have already showed in Figure 2.7 the distribution of the average number of interactions per bunch crossing (pileup) in Run II. After the interaction, the process of combining quarks and gluons forming colourless hadrons that are observed in the final state is known as **hadronization process**. This process happens when the partons reach hadronization scale of ~ 1 GeV. There are two models describing the transition of a coloured partonic system into colourless primary hadrons:
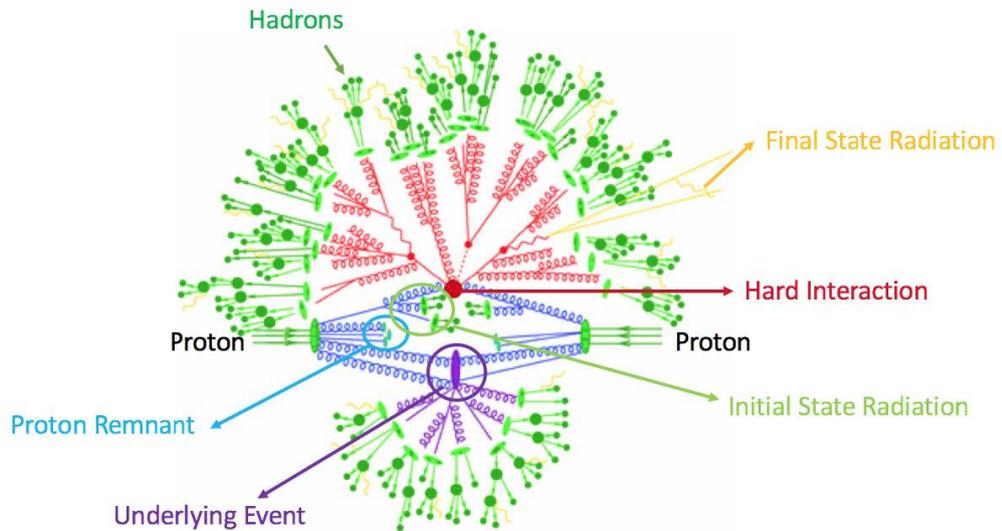
Figure 3.1: Representation of a proton-proton collision event.The red part includes the hard interaction and the decay of the products. Initial (final) state radiation are in blue (yellow). A secondary interaction can take place, in purple, before the final-state partons hadronize. The hadronization is represented by the green blobs, and the hadron decay in dark green [133].

1. **The Cluster Model**: in this model, gluons are split into $q\bar{q}$ pairs where quarks are clustered together into colourless groups. Clusters formed in this

way have large invariant mass which further decay to smaller mass scales
suitable to form hadrons [134].

2. **The String Model**: in this theoretical approach, gluons are split into
quarks, the produced quark and antiquark move out in opposite directions
from their production vertex and lose their energy. During the motion, a
string-like configuration between them is formed. On stretching the string,
the potential energy stored in the string increases and the kinetic energy
decreases, which breaks the string into two parts forming $q\bar{q}$ pairs. The
process is continued until the energy left to create another such pair is low
[135].

The CMS Collaboration uses different MC event generators to generate physics
processes for different purposes such as PYTHIA, Madgraph and POWHEG
[62, 136, 137]. Several theoretical, phenomenological and experimental inputs are
necessary to build a simulation of the proton-proton collisions. Different techniques
are used in particular to describe the QCD processes, whose phenomenology
varies greatly at different energy scales [138]. The hadronic cross section σ_{pp}
is calculated based on the QCD factorization theorem [139]. The factorization
theorem states that the hadronic cross-section pp is a convolution of the partonic
cross section $\hat{\sigma}_{ij}$ with the PDFs $f_i(x)$:

$$\sigma_{pp} = \int_{x_{min}}^{1} \sum_{i,j} f_i(x_1)f_j(x_2)\hat{\sigma}_{ij}(x_1p_1, x_2p_2)dx_1dx_2, \qquad (3.1)$$

where the function $f_i(x)$ is the probability density that a parton of type i has a
fraction x of the hadron energy. The elements involved in the calculations of a
process can be summarized as:

1. The PDFs that are phenomenological functions computed using experimen-
tal information,

2. the **hard scattering**, computed perturbatively order by order,

3. the **parton showering**, used to simulate additional emissions in perturba-
tive QCD,

4. the **hadronization**, describing the transition from coloured particles to
hadrons, treated using phenomenological models,

5. the **decay of unstable particles**, modeled based on experimental data.

The first two are usually included in Matrix Elements generators, while the
last three are included in Parton Showering programs. Both use Monte Carlo
techniques. The matching between these Matrix Elements generators and Parton
Showering should be done in a way to avoid double counting of QCD radiation.

3.2 The Detector simulation

The detector simulation step is an essential part of each high energy experiment as much as the detectors themselves. The detector simulation is the way to mimic the interaction of particles with detector material such as it realistically happens in the detector. This step comes after the MC event generation, by passing the final state particles through the CMS detector for the reconstruction of an event. The detector simulation is based on a full description of the CMS detector geometry which is implemented in the **G**Eometry **AN**d Tracking (GEANT4) simulation toolkit [140]. An accurate description of the full geometry of the detector and the materials of the detecting devices are provided to GEANT4 to simulate the particle response. The particles from the generator level are propagated through the detector materials taking into account the measured magnetic field map (for charged particles) and the interactions processes between the particle and the detector material such as the bremmstralung, the multiple scattering and the photon conversions. As a result of the interactions, GEANT4 produces a set of simulated hits in the active material such as the energy loss of a particle in the sensitive volume of a detector. Then those set of informations (hits, energy deposit, etc) from various sub-detectors are used to reconstruct high-level objects such as jets and leptons in a process known as **reconstruction**.

Pileup interactions are also added at this stage. A library of simulated hits of minimum bias events is used to add a number of extra interactions onto the signal event according to a specified pileup scenario. Out-of-time pileup is modeled by modifying the timing of the detector hits when adding a minimum bias interaction. Once the simulation of the detector is run, all the detector signals are converted to electronic signals in a format identical to the one used for data. From this point onwards the simulated events go through the same reconstruction steps as the collision data.

3.3 Reconstruction of physical objects

The reconstruction process starts with the raw information collected from all sub-detectors such as the hits in the tracker and in the muon system, and the energy deposit in the calorimeters, to build up high-level objects such as jets, tracks, vertices and leptons etc.. For each collision, a list of final-state particles are identified and reconstructed by an algorithm called PF algorithm [141]. Before describing the PF algorithm it is important to describe briefly the tracking and clustering processes done in the tracking system and calorimeters respectively to define tracks and clusters.

3.3.1 The Tracking and Clustering

After the collision, the outgoing particles leave their signatures in various sub-detectors. The charged particles leave hits in the silicon detector and deposit an amount of energy in the calorimeters, while the neutral particles deposit energy only in the calorimeters without hits in the tracker. The muons are the only particles that can reach the muon system leaving hits in the muon chambers which help in the identification and the reconstruction of the muons.

The Tracking Sequence

The tracking process aims to build tracks from hits inside the silicon detector and hence to make momentum measurement. CMS uses a **C**ombinatorial **T**rack **F**inder (CTF) algorithm based on Kalman Filtering (KF) [142, 143] to reconstruct the tracks across a wide range of particle momentum (100 MeV to 1 TeV). The tracking process starts with generating **initial seeds** from few hits compatible with charged-particle trajectory, then building a trajectory by gathering hits from all tracker layers along the charged-particle trajectory.Then fitting all the hits to build a **track** and determine the charged particle's properties such as the origin, the transverse momentum, and the direction.The reconstructed tracks are required to have at least eight hits in total and in a distance of few mm from the beam axis. This process is iterated many times and in each time the hits used in the selected tracks are masked and the remaining hits are used to form new seeds and tracks. At the end, all the reconstructed tracks should pass quality requirements based on number of hits, the track χ^2 and the track compatibility to originate from primary vertex [105, 143].

The Clustering Sequence

The clustering process done in the calorimeters is used to measure the energy of electrons with their bremsstrahlung photons, of photons and neutral hadrons and of charged hadrons. The clustering process starts with defining a **cluster seed** which is a calorimeter cell with energy deposit higher than threshold and higher than the energy deposit of neighboring cells. Then the neighboring cells are added to the cluster until reaching a cell without any signal or have energy deposit lower than the threshold "noise level". For a cell to be added to the cluster, it should not have been already assigned to another cluster and the previous crystal added (in the same direction) has higher energy. Then the formed clusters are combined to form **superclusters** (clusters of clusters) to initiate the building of the trajectories in the inner tracker. The clustering process is done separately for ECAL and HCAL and separately for the barrel and endcap regions. The energy

response of the calorimeter clusters is calibrated from test beam data, radioactive sources, and early collision data events [144].

3.4 The Particle Flow Algorithm

The PF algorithm [141] is developed by the CMS experiment and used to reconstruct and identify all the outgoing particles from the collision. Since the particles interact with various sub-dectors leaving a signature in a form of hit or energy deposit, the idea of the PF algorithm is to link the information from all sub-dectors related to single particle. The identification and reconstruction sequence proceeds in the following order:

1. muon candidates are identified and reconstructed firstly before other particles. The PF muon is reconstructed if the momentum of the global muon track matches the momentum of the muon track in the tracker within three standard deviations. If the tracks pass the criteria for PF muon they are removed from the PF block. Detailed description of global and tracker muons is provided in Section 3.6;

2. electron candidates are identified and reconstructed by carefully taking into account the energy of all bremsstrahlung photons accompanied with electrons. Energetic photons are identified also in this step. If a track in the tracker matches an ECAL cluster it will be identified as an electron but if the ECAL cluster has no track in the tracker this will correspond to a photon. Again the associated tracks and ECAL clusters are excluded from PF block.

3. The remaining tracks in the PF block that are linked to energy deposit in the HCAL are used to create the charged hadron candidates.

4. The left-over cluster in the ECAL without tracks in the tracker gets assigned to photons and similarly to the neutral hadrons in the HCAL. The output of the PF algorithm is a list of PF candidates with their four-momenta, which are then used for further processing to reconstruct jets and to calculate the missing transverse energy from the unbalance of the vectorial sum of the transverse momenta of all reconstructed particles. The reconstruction of leptons (electrons and muons) and jets will be described in the next sections since they are key ingredients for the analysis.

3.5 The Primary Vertex Reconstruction

The aim of the vertex reconstruction [105] is to measure the location of all proton-proton interaction vertices in each event, including the primary vertex PV and all the vertices from pileup collisions, using the available reconstructed tracks. The PV refers to the exact point where the hard scattering or proton-proton collisions happen while the secondary vertices arise from the decay of long lived particles that originated from PV such as b-jets. The vertex reconstruction consists of three steps:

1. the selection of the tracks consistent with the production at the primary interaction region by imposing some requirements on the tracks such as on the track impact parameter relative to the centre of the beam spot, on the number of strips and pixel hits associated with a track and on the track fit χ^2;

2. the clustering of the tracks that appear to originate from the same interaction vertex based on the basis of z-coordinates at their point of closest approach to the centre of the beam spot. The clustering is performed using a Deterministic Annealing (DA) algorithm [105].

3. the fitting of the position of each vertex using its associated tracks. All the reconstructed vertices should have high number of degree of freedom (N_{PV}).

All the reconstructed vertices get stored and further requirements are imposed on the primary vertex:

- its z-position should be within 24 cm of the nominal detector center;

- a small radius of the PV ($r_{PV} < 2$ cm). If more than one primary vertex is reconstructed, the vertex with associated tracks with the highest sum of the square of the transverse momenta is considered as the primary vertex.

3.6 The Muon Reconstruction

Muons are the only charged particles that can reach the muon system leaving signature in the muon chamber. Combining the information from the inner tracker and the muon system provides highly efficient reconstructed muon tracks with good resolution for momentum measurement. Figure 3.2 shows a schematic view for the muon reconstruction algorithms. The muon reconstruction [119, 145, 146] chain starts with the local reconstruction of the muon tracks independently in

the inner tracking (tracker track) and in the muon systems (stand-alone muon track) as follow:

- **Standalone tracks**: those are muons reconstructed using the informations from muon system only. The hits from DT and CSC chambers are clustered to form track segment "track-stubs" then they are matched together to form seeds. Then by matchings the seeds with the RPC hits to form the muon track a stand-alone muon track is built.

- **Tracker-tracks**: those are muons reconstructed using the informations from inner tracker only. The hits in the silicon tracker are clustered to form a tracker track also called a "inner track" or a "silicon track".

Based on those two local reconstruction processes two reconstruction algorithms are defined:

- **Global Muon Reconstruction (outside-in)**: for each stand-alone muon track, a search for matching tracker track reconstructed in the inner tracking system is performed, and the best-matching tracker track is selected. For each "stand-alone muon"-"tracker track" pair, the track fit using all hits in both tracks is performed based on the Kalman filter technique. For muons with $p_T > 200$ GeV, the global muon reconstruction provides better momentum resolution than tracker muon reconstruction.

- **Tracker Muon Reconstruction (inside-out)**: in this approach, all tracker tracks with $p_T > 0.5$ GeV and total momentum p > 2.5 GeV are considered to be potential muon candidates by searching for at least one compatible segment in the muon stations. Tracker tracks identified as muons by this method are referred to as "tracker muons". This method has a better energy resolution for low momentum muons having $p_T < 5$ GeV, those muons do not always able to traverse the CMS detector up to the muon system. This criteria used to define a tracker track as "tracker muon" are very loose and should not be used without further requirements.

The selection of muons used in the analysis will be discussed in Section 4.4.2 .

3.7 Electron Reconstruction

Electrons are reconstructed by combining the informations from the tracker and the ECAL sub-detectors by associating a track reconstructed in the silicon detector with a cluster of energy deposited in the ECAL. Electrons loose part of their energy when passing the silicon material by Bremsstrahlung radiation and ionization process. To measure accurately the energy of the electron, it is necessary
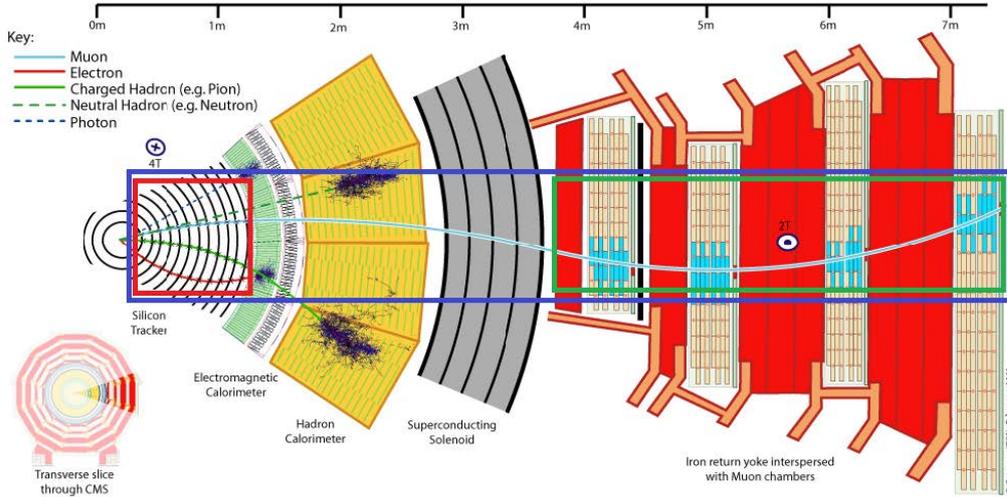
Figure 3.2: Reconstruction of muon objects at the CMS experiment. Tracker track (red box), stand-alone track (green box) and global muon (blue box) [131].

to collect the energy of the radiated photons. The radiated photon spreads its energy on various crystals along the ϕ direction because of the bending of the electron trajectory in the magnetic field with negligible spread in η direction. The clustering of the energy of the electrons and the photons is done by two clustering algorithms: the "hybrid" and the "multi-5×5" algorithms in barrel and endcap regions respectively [144, 147]. The hybrid algorithm starts by finding a seed crystal, defined as the one containing most of the energy deposited in any considered region and had energy greater than threshold value. An arrays of 5×1 crystals in $\eta \times \phi$ are added around the seed crystal, in a range of N_{steps} crystals in both directions of ϕ if their energies exceed a minimum threshold as shown in Figure 3.3. The contiguous arrays are grouped into clusters, with each distinct cluster required to have a seed array with energy greater than a threshold in order to be collected in the final global cluster, called the supercluster seed-array (SC). In the endcap region the multi-5×5 algorithm is used where crystals are not arranged in an $\eta \times \phi$ geometry. In the same way the algorithm starts with finding the seed crystals defined as the ones with local maximal energy relative to their four direct neighbors and exceeds a threshold value. Around these seeds and beginning with the largest E_T, the energy is collected in clusters of 5×5 crystals, that can partly overlap. These clusters are then grouped into an SC if their total transverse energy is greater than a threshold energy in a range of η and ϕ around each seed crystal. The SC energy corresponds to the sum of the energies of all its clusters. The SC position is calculated as the energy-weighted mean of the cluster positions. When the electron crosses the tracker detector

it loses an amount of its energy because of the significant amount of material budget in the pixel detector. This loss of energy, especially for bremsstrahlung photon emission, causes a kink in the electron trajectory making the electron track reconstruction by KF approach not suitable. If the standard KF approach succeeds to follow the electron path the quality of the track χ^2 can be poor. For this reason a dedicated track algorithm is used for electrons track reconstructions. The Gaussian Sum Filter (GSF) tracking [148], based on the KF approach, is used for the electron tracking. The KF approach relies only on the Gaussian probability density functions while the bremsstrahlung energy loss distribution is highly non Gaussian which is used in the GSF algorithm. The GSF algorithm models the bremsstrahlung energy loss distribution by a Gaussian mixture rather than a single Gaussian function. In this way the GSF track χ^2 takes into account the sudden curvature in the electron track caused by the Bremsstrahlung photon emission. A final step in the electron reconstruction is the association of a GSF track with a cluster in the ECAL reconstructed by the hybrid or the multi 5×5 algorithms. The selection of electrons used in the analysis will be discussed in Section 4.4.1.
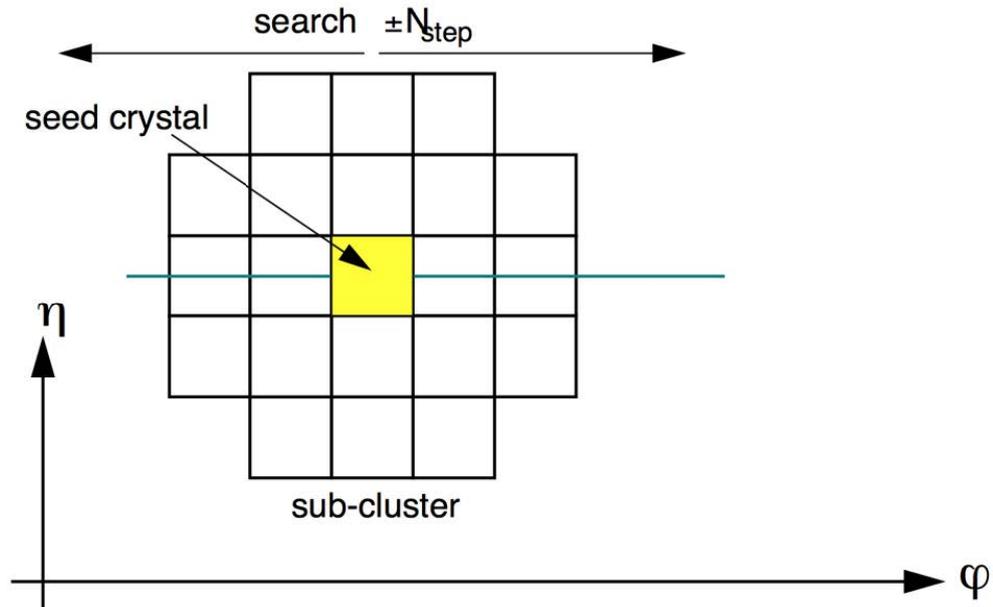
Figure 3.3: The hybrid algorithm with seed crystal shown [131].

3.8 Jet Reconstruction

As quarks and gluons undergo an hadronization process, the estimation of their initial momentum requires the recollection and measurement of the hadronization products. Jets are thus reconstructed by clustering the PF candidates with the **anti-kT algorithm** [149, 150] as implemented in the FASTJET package . The algorithm iteratively combines PF candidates that are close to each other according to a metric, that is defined to produce jets of an approximate conic shape clustered around the hardest (the highest momentum) particles in the event. The size of the jet cone is determined by the distance parameter ΔR at which the algorithm is operated. The standard clustering distance used at the CMS is ΔR = 0.4 since Run II, while it used to be 0.5 at 8 TeV. At the same time, larger jets, with a ΔR of 0.8, are used when looking for boosted heavy particles decaying into hadrons. The value ΔR = 0.4 is used for the search presented in this thesis.

The kT algorithm is infrared and collinear safe: infrared safeness means that the results of the jet clustering are not altered if an arbitrary number of extra particles with momentum that tends to zero is included in the jet clustering; collinear safeness means that the clustering is not sensitive to a splitting of a particle into two collinear ones each taking a fraction of the momentum.

The jet four-momentum is computed as the vector sum of the clustered PF candidates four momenta, and a set of corrections, the so-called **J**et **E**nergy **C**orrection (JEC), are applied to calibrate the jet response using the information of generated particles in a simulation. These corrections of the jet energy scale take into account the contribution from pileup in the event, non-linearities in the detector response to hadrons, and residual differences between the data and the simulation used for the method. They are validated using dijet, multijet, γ+jets and leptonic Z+jets events [151–153]. The jet calibration is repeated for each data-taking year, in order to take into account the different conditions. Typical jets resolutions achieved are of about 15–20% for at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV.

3.8.1 Identification of b-jets

Jets are by far the most common objects at hadron colliders. Jets from b quarks look overall very similar to the other jets and need no special treatment in their reconstruction. However, a B hadron is produced in the hadronization process of a b quark, and the reconstruction of its decay products inside the jet is key to tag the b-jets. Jets containing B hadrons can be distinguished thanks to the B hadron long lifetime:$c\tau \sim 500\ \mu$m. A B hadron with p_T = 50 GeV flies on average

almost half a centimeter ($L \sim \gamma c\tau$) after being produced. The relatively long lifetime of B hadrons is due to the need for b quarks to decay weakly into lighter quarks. The top quark final state would be favored, but it cannot be accessed kinematically due to the mass of $\sim$170 GeV of the top quark. The transition to lighter quarks, belonging to the second or the first family, comes with a sizeable suppression factor[1] and results in a longer lifetime. Often B hadrons decay into charmed D hadrons, which have in turn non negligible lifetimes (c$\tau \sim 300\ \mu$m), so full decay chains can be found and in some cases reconstructed within b-jets. The B hadrons' long lifetime results in a sizeable impact parameter of the decay products with respect to the primary vertex, which are reconstructed, if charged, as tracks. A secondary vertex or multiple secondary vertices can be reconstructed and their properties are highly discriminating variables. Another property of the B hadron decay is the relatively high rate of lepton production from semileptonic decays (around 25%). These leptons can be identified thanks to their relatively high $\boldsymbol{p_T}$ relative to the B flight direction, approximated by the b-jet direction: leptons from B decays have order of GeV momenta relative to b-jet direction, because of the B hadron mass (~ 5 GeV), while leptons in jets of other flavours tend to be closely aligned with the jet.

CMS standard b-tagging algorithms

Jet b-tagging is one of the areas where **M**achine **L**earning (ML) is fundamental to have optimal performances. CMS standard algorithms, optimized with Machine Learning, rely both on secondary vertices and tracks. These algorithms were developed after and in parallel with simpler algorithms, based on a single observable or one type of observables, which are still useful to monitor the main observables. A description of those is first given. Among the single-variable based algorithms the **T**rack **C**ounting (TC) algorithm sorts tracks in a jet by decreasing values of the IP significance[2]. A natural extension of the TC algorithms is the combination of the IP information of several tracks associated with a jet in a cone of ΔR < 0.3. Two discriminators are computed: the **J**et **P**robability (JP) algorithm uses an estimate of the likelihood that all tracks associated with the jet come from the primary vertex while the **J**et **B** **P**robability (JBP) algorithm gives more weight to the four tracks with the highest IP significance, with the

---

1 Weak decays of the quarks can result in different family in the final state, because of the mixing of weak interaction eigenstates and mass eigenstates. The mixing components, i.e. the out of diagonal elements in the CKM matrix, are however small, i.e. $< 10^{-2}$, hence the reduced decay rate.

2 A commonly used variable to tag b-jets is the **S**ignificance of the **I**mpact **P**arameter (SIP): SIP $= \frac{IP}{\sigma_{IP}}$ where σ_{IP} is the IP uncertainty. This observable takes care also of mis-measured tracks, which can have artificially large impact parameters.

number four chosen as it matches the average number of reconstructed charged particles from B hadron decays.

Other single observable based discriminators use the secondary vertices. The **S**imple **S**econdary **V**ertex (SSV) algorithms use the significance of the SV flight distance as discriminating variable.

This set of algorithms was developed before the beginning of the LHC data taking. Already during Run I multivariate discriminators with Machine Learning techniques were employed: the **CSV** algorithm, which combines secondary vertices and tracks via a likelihood ratio was first developed. The performance of these algorithms is shown in Figure 3.4. A **R**eceiver **O**perating **C**haracteristic curve (ROC) curve[3], comparing the efficiency of b-jets and the mis-tag of light flavour jets for the $t\bar{t}$ 7 TeV simulation is shown. The CSV algorithm outperforms all the algorithm, including JP.

Better performing versions of the CSV algorithms were developed afterward. The **CSVv2** requires at least two tracks per jet compatible with the primary vertex. Additionally, any combination of two tracks compatible with the mass of the K_S^0 meson is rejected. The training of the algorithm is then performed in three independent vertex categories. The first vertex category contains jets with at least one associated reconstructed secondary vertex. The second, called "pseudo vertex", contains jets whose tracks with an IP significance larger than two can be combined in a pseudo-vertex, allowing for the computation of a subset of SV observables. Otherwise, a "no vertex" category with track-based variables only is defined.

The variables used for the training include secondary vertex observables (2D flight distance significance, mass, number of tracks, energy and transverse momentum ratio with respect to the jet, etc.), variables relative to the track with the highest 2D SIP (η_{rel}, p_T^{rel}, decay length, etc.), the 3D SIP of the first four tracks, variables relative to the sum of the selected tracks, and the jet, p_T.

In Run II [154], the training was performed using a shallow neural network for each category, and separately for b-jets versus light flavour jets and b-jets versus c-jets. The outputs were then combined via likelihood ratios among the categories and a final re-weighting, with relative weights of 1:3 for the b versus c and b versus light flavour jets, respectively.

Finally, during Run II Deep Learning was introduced for b-tagging. The **DeepCSV** algorithm was developed using a deep feed-forward neural network. The DeepCSV algorithm uses the same information as the CSVv2 one, but the training is performed using more events and a more flexible algorithm. This solves the entire b-tagging problem in one step, i.e. a training including all categories and all jet flavours, and allows improved performances. The performance of b-tagging

---

3 Machine Learning terms will be introduced in Chapter 4.

algorithms used at CMS at 13 TeV are shown in Figure 3.4. The CMVA algorithm, which uses also leptonic decays information to improve on top of the CSVv2 but is outperformed by DeepCSV for high purity working points, is also shown.
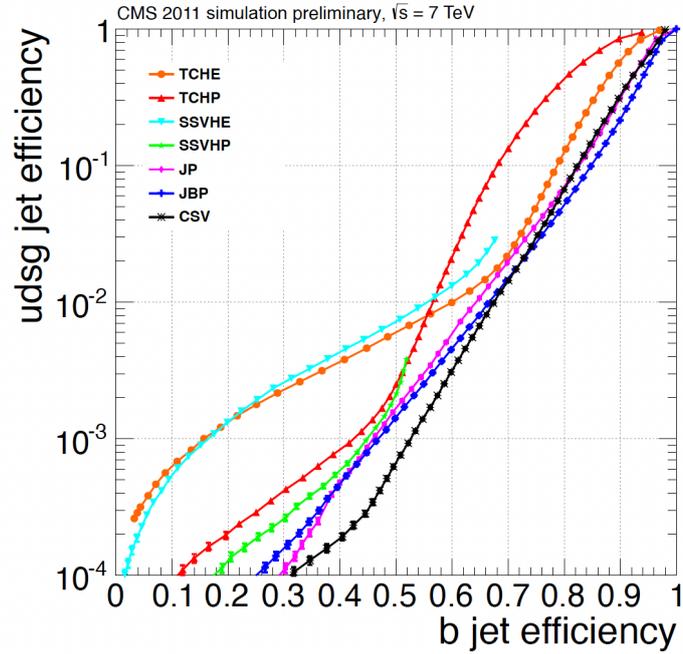
(a)
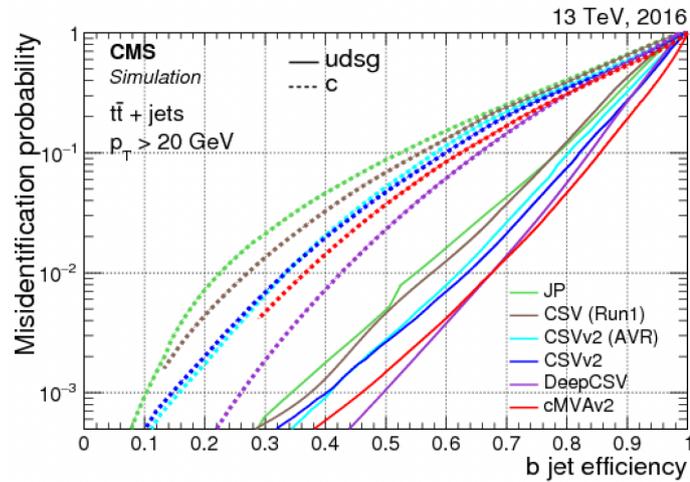
(b)

Figure 3.4: Mis-identification probability for c and light flavour jets versus b-jet identification efficiency for various b-tagging algorithms applied to jets in $t\bar{t}$ events [70, 154].

The Double-Higgs Analysis

> " The best that most of us can hope
> to achieve in physics is simply to
> misunderstand at a deeper level."
> – Wolfgang Pauli to Jagdish Mehra

This chapter describes the analysis steps for the exploration of double Higgs production via vector boson fusion in the $ZZ^*b\bar{b} \to b\bar{b}4l$ decay channel in data collected by the CMS experiment in pp collisions during the 2018 Run II data-taking. It requires the experimental capability to identify and reconstruct several different types of final state objects and to use them for the selection of signal-like events and rejection of background ones. These selections and techniques globally constitute the **analysis strategy**, the backbone of the VBF $HH \to b\bar{b}l^+l^-l'^+l'^-$ search to which this chapter is devoted.

The first part of this chapter introduces the experimental challenges of the studied decay channel. The discussion focuses on the experimental signature of the signal processes under study and on the main background sources. Next sections focus on the trigger , object pre-selections, definition of a SR and presentation of the results. Each part relies on the previous one and improves the discrimination from background processes with an increasing complexity of the selections applied. Trigger represent the first level of event selection and are necessary to store events offline for subsequent analysis.

After a brief description of systematic uncertainties, I proceed to the interpretation of the results as upper limits on the signal strength for SM and BSM double Higgs production. The Di-Higgs signal is quite small compared to the main backgrounds. In order to cope with the limitations due to the low statistics and to enhance the signal to background discrimination the **MultiVA**riate analysis (MVA) approach is presented in this chapter. This method is used to integrate various kinematic observables into a single discriminant to better discriminate between signal and background events and to improve the sensitivity of the analysis.

The output distribution of this method is used to extract the 95% C.L. upper limits on the signal strength. At this point, I got all the ingredients to start the statistical analysis. These limits can allow the exclusion of some points of the

parameter space of the model chosen. In Section 4.8 a discussion of the statistical method used in the analysis precedes the comment on the results obtained [155].

4.1 The VBF bb̄4l decay channel

The $HH \rightarrow b\bar{b}ZZ^* \rightarrow b\bar{b}4l$, $(l = e, \mu)$ channel has a large signal-to-background ratio because it is possible to completely reconstruct the final state leptons in CMS, which present excellent momentum resolution. This feature makes this decay channel very promising in the studies of the VBF double Higgs boson's properties along with the $b\bar{b}\gamma\gamma$ channel mode [63]. Several different measurements have been performed with the data collected during the LHC Run I using the single-Higgs channel in four charged leptons [36, 156].

Since the Z gauge boson is neutral, the sum of the charges of its decay products must be 0. This is because in nature charge is conserved. Therefore Z must decay into a particle, antiparticle pair. The 100% probability of Z to decay is divided between groups of particles according to additional conservation's laws:

1. In 10% of the Z-decays, charged lepton-antilepton pairs are produced. The three possible charged lepton pair types are electron-positron e^-e^+, muon-antimuon $\mu^+\mu^-$, and tau-antitau $\tau^+\tau^-$ pairs. Each pair is approximately equally probable (respectively with BR[1] of 3.363 ± 0.004, 3.366 ± 0.007, 3.370 ± 0.008 [157]).

2. The Z boson decays in 20% of the cases into a neutrino-antineutrino pair. The CMS detector is not capable of detecting neutrinos since they almost do not interact because of no electric charge. The neutrinos are therefore invisible and the only way to "see" them is when we measure that there is some energy or transverse momentum missing after the collision (transverse momentum and energy should be conserved in the collision). The neutrino decays give another 3 possibilities.

3. In 70% of Z decays, a quark-antiquark pair is produced. These appear as particle showers, the "jets" in the detector.

In my study, the signal officially produced by CMS includes one Higgs boson (with hypothesis mass of 125.09 GeV) into two b-quarks with a SM BR of 5.75 $\times 10^{-1}$ and one Higgs boson in final state of four charged leptons $4l$ including electrons,muons and taus under the mass hypothesis of 125.09 GeV (present best fit of the Higgs boson's mass) with a BR of 2.79×10^{-4} [158]. I will perform on

---

1 In particle physics and nuclear physics, the branching fraction (or branching ratio) BR for a decay is the fraction of particles which decay by an individual decay mode with respect to the total number of particles which decay.

these samples a selection for events of four muons 4μ, four electrons $4e$ and two electrons and muons $2e2\mu$. In this context there could be a contamination from events of four taus 4τ or $2\mu2\tau$ with the lepton tau decaying into muons.

The difficulty in discriminating this physical process to SM ones relies in its topology complexity, i.e. the presence of many physical objects involved (four jets and four charged leptons) which are common to many background processes as it will be discussed in Section 4.5.

4.2 Data-sets and Triggers

The analysis presented in the thesis is intended to be applied to the data-sets recorded by the CMS experiment during 2018 Run II campaign of data-taking at $\sqrt{s} = 13$ TeV. All the MINIAOD[2] data used in this analysis are certified by a CMS Official Group, which ensures high data quality with a good functioning operation of the various sub-detectors. The recorded data during 2018 Run II correspond to a total integrated luminosity of 59.74 fb^{-1}. They are listed in Table 4.1 along with the integrated luminosity of each Run-range.

The analysis used four primary data-sets[3] (**DoubleMuon - EGamma - MuEG - SingleMuon**) in 2018 each of which combines a certain collection of high-level trigger HLT paths [159]. The data-sets are collected by passing single-lepton, di-lepton, or tri-lepton triggers .

The High Level Trigger (HLT) is used to decide whether an event should be kept for an offline analysis or not and hence reduce the number of events to be analyzed instead of repeating the analysis million times on uninteresting events. The data used in the 2018 year have been re-reconstructed (called *ReReco*) with the most recent detector calibrations, alignment and fine-tunes with respect to those used during the data-taking (except Run period D in 2018 data which used prompt reconstruction workflow *promptReco* with preliminary detector calibrations). In this analysis, I rely on multi-lepton HLT paths (single, di and tri-leptons HLT paths, in particular the tri-leptons triggers are used to compensate the inefficiencies at low p_T of single and di-leptons) as listed in Table 4.2 [4]. The events are required

---

2 MINIAOD format, introduced by CMS experiment, is a small and quickly derived data format where the majority of CMS analysis users can start their analysis work. The format has size smaller than the AOD format. It has a sufficient information to serve the CMS analysis, while dramatically simplifying the disk and resources needed for analysis.

3 Primary data-sets are Analysis Object Data files AOD that contain all the information about the full reconstructed collision data with no other selections. Those files contain all the informations needed to perform the analysis, such as the high level physics objects (muons, electrons, missing energy, etc.), tracks, calorimeter clusters, vertices and identification criteria for the physics objects.

4 As an example, for an event to pass the `HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*` (see Figure 2.1), as example, it is required that the event has at least two muons with p_T

Run-range	Data-set	$\mathcal{L}_{int}$
315252-316995	/DoubleMuon/Run2018A-17Sep2018-v2/MINIAOD /MuonEG/Run2018A-17Sep2018-v1/MINIAOD /SingleMuon/Run2018A-17Sep2018-v2/MINIAOD /EGamma/Run2018A-17Sep2018-v2/MINIAOD	14.00 fb^{-1}
317080-319310	/DoubleMuon/Run2018B-17Sep2018-v1/MINIAOD /MuonEG/Run2018B-17Sep2018-v1/MINIAOD /SingleMuon/Run2018B-17Sep2018-v1/MINIAOD /EGamma/Run2018B-17Sep2018-v1/MINIAOD	7.10 fb^{-1}
319337-320065	/DoubleMuon/Run2018C-17Sep2018-v1/MINIAOD /SingleMuon/Run2018C-17Sep2018-v1/MINIAOD /SingleMuon/Run2018C-17Sep2018-v1/MINIAOD /EGamma/Run2018C-17Sep2018-v1/MINIAOD	6.94 fb^{-1}
320673-325175	/DoubleMuon/Run2018D-PromptReco-v2/MINIAOD /SingleMuon/Run2018D-PromptReco-v2/MINIAOD /MuonEG/Run2018D-PromptReco-v2/MINIAOD /EGamma/Run2018D-PromptReco-v2/MINIAOD	31.93 fb^{-1}

Table 4.1: List of the data-sets analysed for the $4l$ final state. The first column reports the run intervals associated to a specific data-set. The second column reports the CMS-specific data-set name, where the set of HLT triggers (as noticed in the name of the primary data-set, as example *DoubleMuon* data-set means that this set contains the data passing di-muon or tri-muon filters with specific p_T thresholds and loose lepton identification and isolation), the data-taking period (the year, the Era - A, B, C, D - of the data-taking, the date of reprocessing and the version), and the data tier as explained in Section 2.4. Finally, the third column presents the corresponding integrated luminosity.

to pass at least one of those HLT paths. To avoid the duplication of events from different primary data-sets, events are taken in the following sequence:

- from **EGamma**, if they pass the **diEle**ctron (DIELE) or **triEle**ctron (TRIELE) or **singleEle**ctron (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 **Mu**on**Ele**ctron (MUELE) or **Mu**on**Di**Electron (MUDIELE) or **DiMu**on**Ele**ctron (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.

---

threshold of 17 and 8 GeV passing very loose tracker isolation and transverse impact parameter condition.

This trigger menu is also applied in all simulated samples as on data in order to correct any possible mismatches. The trigger efficiency measured using 4l events is found to be $> 99\%$ for each final state [160].

HLT path	Primary data-set
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v*	EGamma
HLT_DoubleEle25_CaloIdL_MW_v*	EGamma
HLT_Ele32_WPTight_Gsf_v*	EGamma
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v*	DoubleMuon
HLT_TripleMu_10_5_5_DZ_v*	DoubleMuon
HLT_TripleMu_12_10_5_v*	DoubleMuon
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v*	MuEG
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*	MuEG
HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v*	MuEG
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v*	MuEG
HLT_DiMu9_Ele9_CaloIdL_TrackIdL_DZ_v*	MuEG
HLT_Mu8_DiEle12_CaloIdL_TrackIdL_DZ_v*	MuEG
HLT_IsoMu24_v*	SingleMuon

Table 4.2: Trigger paths used in the analysis for the full 2018 CMS data-taking.

4.3 The simulated samples

An accurate modelling of the signal and background processes is crucial to explore the selected data events. Simulations performed with the MC technique are an essential tool it is used in various steps in the analysis starting from the optimization of the analysis strategy, the comparison of the observed data to the predictions, and up to the production of final results. This section describes the signal and the background samples used in the analysis. All generated samples are processed through GEANT4 simulating the CMS detector and then reconstructed through the official production chain.

4.3.1 Signal Samples

Seven samples (s_i) of the $qq' \to HHqq' \to ZZb\bar{b}qq' \to b\bar{b}qq'4l$ process have been generated privately ($\sim 100k$ events) with the corresponding cross sections per BR ($\sigma_i \times BR$) provided by MadGraph5_aMC@NLO LO generator [62] for the SM description and an effective field theory description of the fermion loops. Showering of parton-level events is done using *PYTHIA8* with no specific matching requirement while hadronization refer to *TuneCP5*. Signal samples for BSM hypotheses with different value of the tri-linear Higgs boson self-coupling

(HHH) κ_λ , the quartic (HHZZ) coupling C_{2V} and the (HZZ) C_V one are reported in Table 4.3 [63, 161]. It has to be noticed that the cross section values is much

$\kappa_\lambda, C_V, C_{2V}$	$\sigma \cdot BR$ [fb]	Signal samples name	ϵ_{eff}
1,1,1	0.00053581950	VBFHHTo2B2Z_CV_1_C2V_1_C3_1_TuneCP5_13TeV-madgraph-pythia8	0.73
1,1,2	0.0044084790	VBFHHTo2B2Z_CV_1_C2V_2_C3_1_TuneCP5_13TeV-madgraph-pythia8	0.82
2,1,1	0.00044277300	VBFHHTo2B2Z_CV_1_C2V_1_C3_2_TuneCP5_13TeV-madgraph-pythia8	0.81
0,1,1	0.0014277825	VBFHHTo2B2Z_CV_1_C2V_1_C3_0_TuneCP5_13TeV-madgraph-pythia8	0.75
1,0.5,1	0.0033560910	VBFHHTo2B2Z_CV_0.5_C2V_1_C3_1_TuneCP5_13TeV-madgraph-pythia8	0.82
1,1.5,1	0.020470230	VBFHHTo2B2Z_CV_1.5_C2V_1_C3_1_TuneCP5_13TeV-madgraph-pythia8	0.78
1,1,0	0.0014277825	VBFHHTo2B2Z_CV_1_C2V_0_C3_1_TuneCP5_13TeV-madgraph-pythia8	0.79

Table 4.3: In the first column the values of the effective Lagrangian couplings that define the seven shape benchmarks are reported. In the second and last column the SM and BSM HH samples used for BSM and SM HH signal hypotheses along with their VBF HH cross sections per branching fractions [63] and HLT trigger efficiency are shown. Finally, part of the MC's names are reported. Indeed in the CMS convention the complete name would report respectively the process simulated, generator and hadronizer name, the data-taking campaign, the PileUp, CMSSW Release - *CMSSW* is the third incarnation - C++ language, from 2005 - of CMS software after the *CMSIM* - Fortran language, until 2000 - and *ORCA* - in C++ language, during the period 2000-2005, the detector configuration and the type of conditions.

larger than the SM value (in the latter case Higgs couplings are all equal to one). All the samples are needed for the scan of the parameters κ_λ and C_{2V} as well as for the extraction of upper limits on cross section.

Many studies were conducted on the SM sample in order to understand better the signal topology w.r.t. background ones. Firstly, a comparison in terms of the kinematic observable distributions between SM VBF HH and HH gluon-gluon fusion production mode - one of the main VBF HH backgrounds - by using LHE[5] files from *MadGraph5_ aMC@NLO* is performed and shown in Figure 4.1.

---

5 The LHE is an agreement between Monte Carlo event generators and theorists to define *Matrix Element* level event listings in a common language. Indeed, HEP event generation can typically be split into the following steps: Matrix Element calculation, Parton Shower, Underlying Event and Hadronisation. Usually the physics event of interest, as well as the cross-section information is done with the computation of the Matrix Element (PDF evaluation, phase space, amplitudes, spin correlations, etc...) and the remaining steps are used to evolve the parton-level event to its final state. All these secondary steps rely heavily on models and are generally independent from the Matrix Element calculation. Therefore only few, typically multi-purpose event generators, implement those additional steps. Examples are Pythia (6 and 8), Herwig (Fortran and C++ versions) and Sherpa [162].

(a)

(b)

(c)
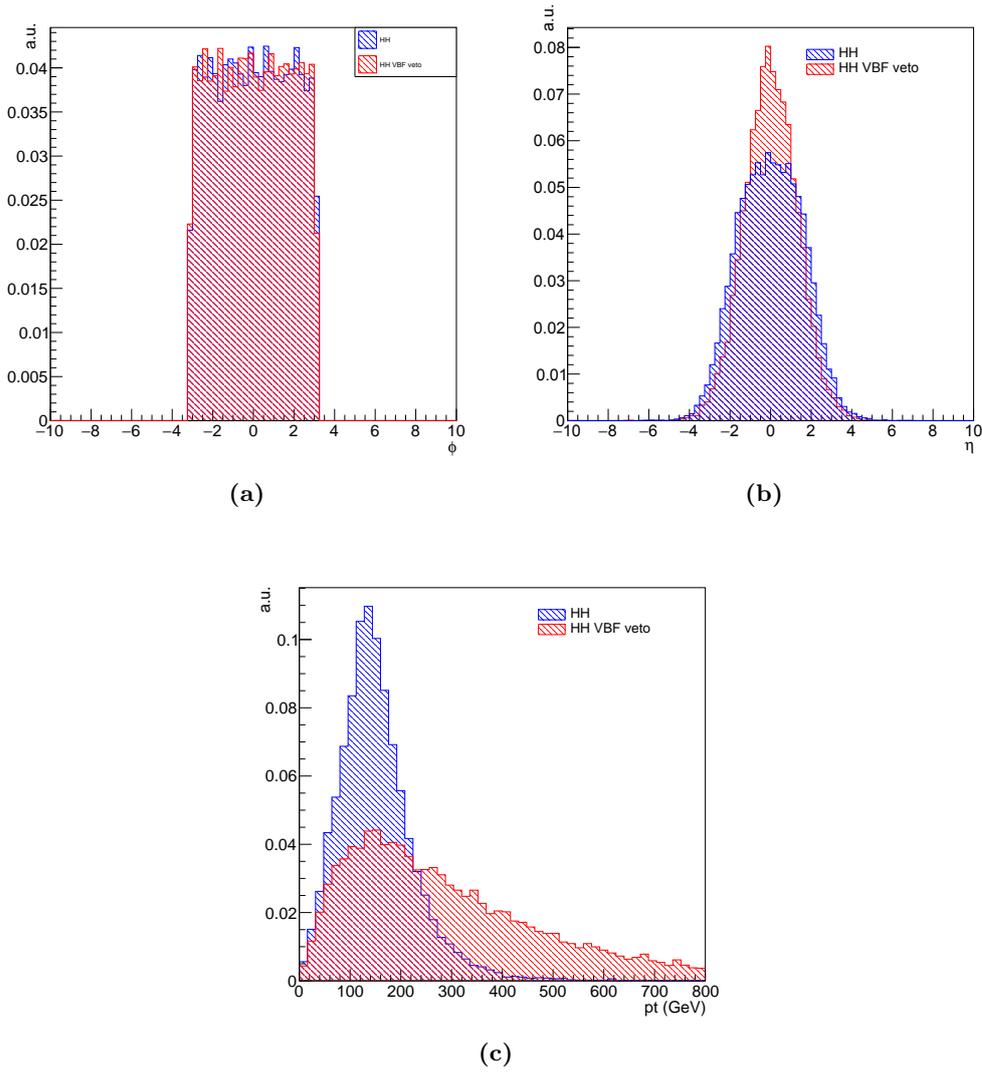
Figure 4.1: Normalized distributions in ϕ (Figure 4.1a), η (Figure 4.1b) and p_T (Figure 4.1c) of the two Higgs bosons by using LHE files for the VBF HH SM signal and the gluon-gluon fusion HH background produced with *MadGraph5_aMC@NLO* . The label *VBF veto* is related to the input command `generate p p > h h j j $$ z w+ w- / a j QED<=4` to Mad-Graph5 that forbids any Feynman diagram with a photon, a gluon or a quark before the propagation and decay of the intermediate bosons $Z/W^{\pm}$. This indication is used by CMS for producing this particular simulated sample in order to face gauge invariance (Lorentz invariance) issues [62].

Moreover, some of the main VBF physical observable distributions at generator level stored in LHE files (without propagation/ hadronization /decay into the CMS detector) from *MadGraph5_aMC@NLO* is reported in Figure 4.3 with a comparison between the simulated samples *VBF HH veto* and *VBF HH no veto*. They actually do not differ too much as it can be seen from the ratio plots.

By following the CMS official indications, the *VBF HH veto* simulated sample, which excludes any Feynman diagrams with a photon,a gluon or a quark before the propagation and decay of the intermediate bosons $Z/W^{\pm}$, will be used in this analysis. Some examples of the Feynman diagrams, which are generated, are reported in Figure 4.2.

Figure 4.2: Feynman diagrams at LO for the SM VBF HH production by using *MadGraph5_aMC@NLO*. The input command `generate p p > h h j j $$ z w+ w- / a j QED<=4` both QCD and pure EW production.
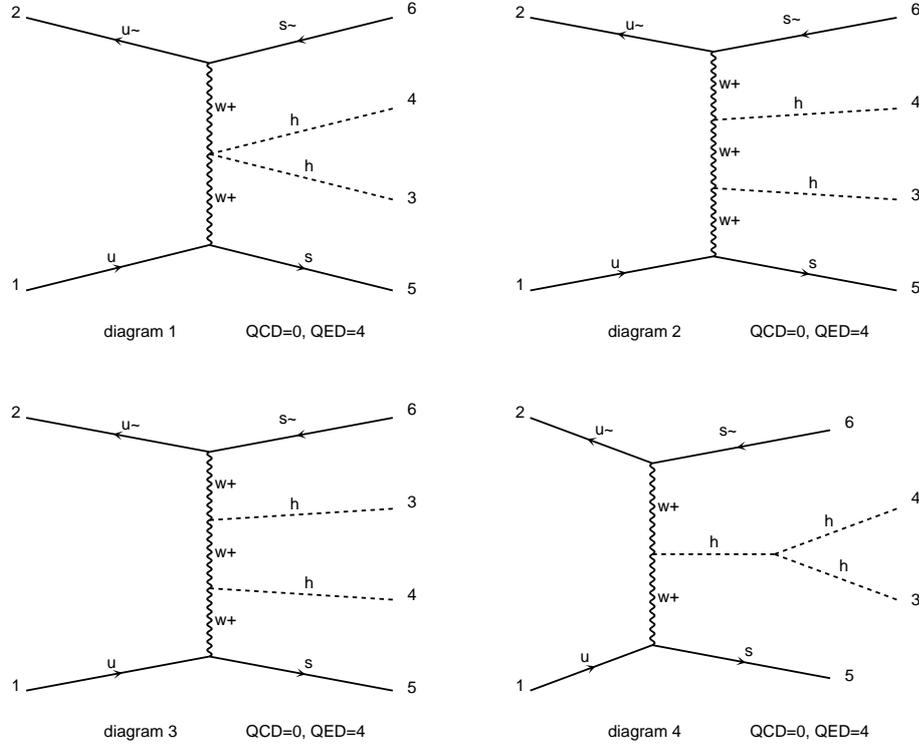
Background Samples

In this analysis, events with double Higgs bosons produced by vector boson fusion into two b-jets and four charged leptons are considered as the signal, while the other HH production modes (I take into account only of the gluon-gluon fusion mode since the cross sections at LHC is much smaller w.r.t. VBF one whose features are described in the caption of Table 4.4.), along with the single Higgs boson production channels and SM backgrounds, constitute the total background.

The simulation of the SM Higgs is obtained by using the POWHEG V2 generator for the five main production modes [163] : gluon fusion (ggH) including quark mass effects, vector boson fusion VBF, and associated production (WH,
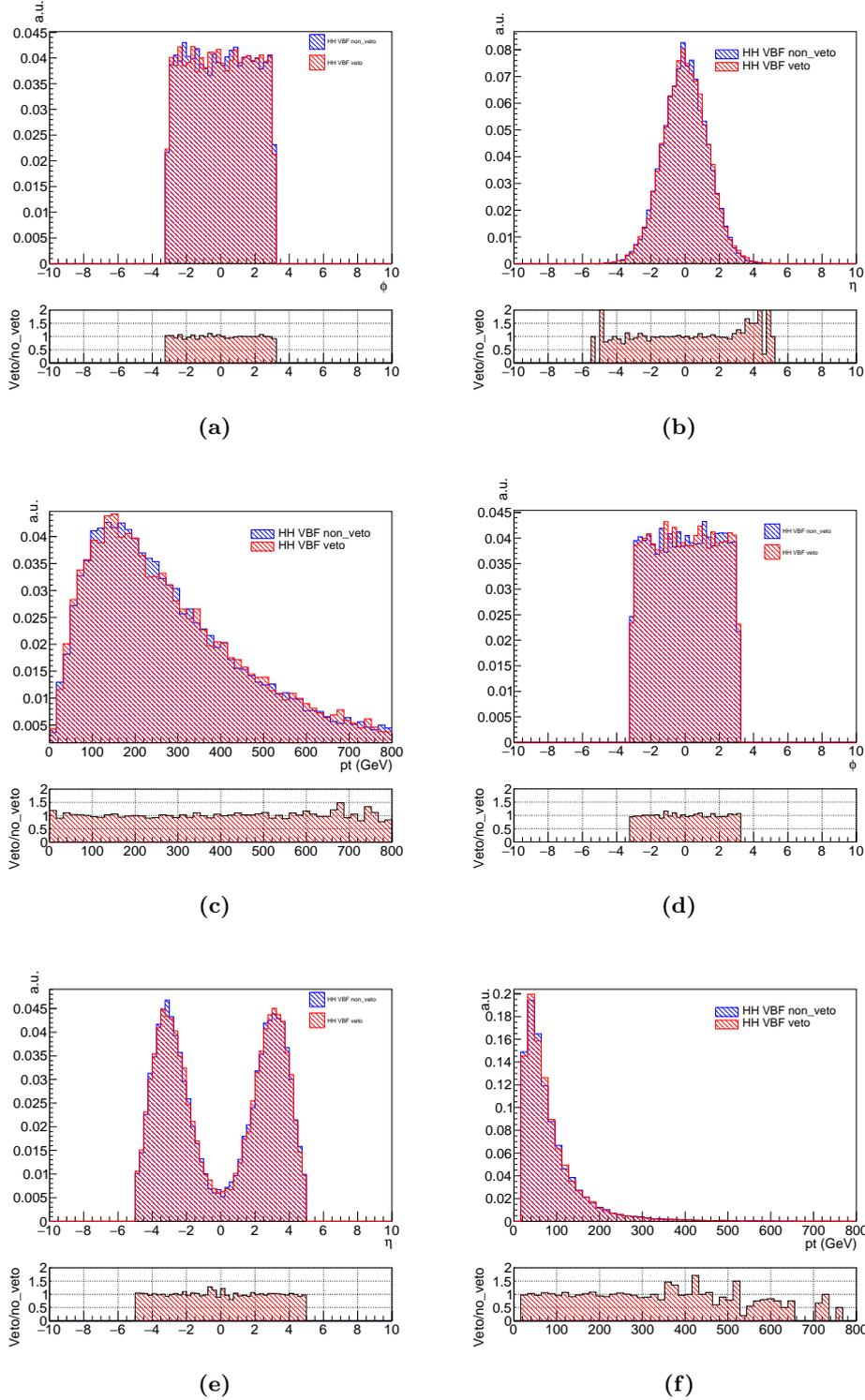
Figure 4.3: Normalized distributions in ϕ (Figure 4.3a), η (Figure 4.3b) and p_T (Figure 4.3c) of the two Higgs bosons by using LHE files for the VBF HH SM with and without VETO produced with *MadGraph5_aMC@NLO*. The same normalized distributions are generated for the VBF jets respectively in Figure 4.3d, Figure 4.3e and Figure 4.3f. The label *VBF veto* is related to the input command `generate p p > h h j j $$ z w+ w- / a j QED<=4` to MadGraph5 that forbids any Feynman diagram with a photon, a gluon or a quark before the propagation and decay of the intermediate bosons $Z/W^{\pm}$. This indication is used by CMS for producing this particular simulated sample in order to face gauge invariance (Lorentz invariance) issues [62].

Process	Dataset Name	$\sigma \cdot BR$
SM Higgs MC Samples		
$gg \to H \to ZZ \to 4l$	/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUgenV709_pythia8/[1]	12.18 fb
$q\bar{q} \to Hq\bar{q} \to ZZq\bar{q} \to 4lq\bar{q}$	/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUgenV709_pythia8/[1]	1.044 fb
$q\bar{q} \to W^+H \to W^+ZZ \to 4l + X$	/WplusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUgenV709_pythia8/[1]	0.232 fb
$q\bar{q} \to W^-H \to W^-ZZ \to 4l + X$	/WminusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUgenV709_pythia8/[1]	0.147 fb
$q\bar{q} \to ZH \to ZZZ \to 4l + X$	/ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUgenV709_pythia8/[1]	0.668 fb
$gg \to t\bar{t}H \to t\bar{t}ZZ \to 4l + X$	/ttH_HToZZ_4LFilter_M125_13TeV_powheg_JHUgenV709_pythia8/[1]	0.393 fb
$gg \to b\bar{b}H \to b\bar{b}ZZ \to 4l + X$	/bbH_ToZZTo4L_M125_13TeV_JHUGenV7011_pythia8/[1]	0.135 fb
SM Backgrounds MC Samples		
$q\bar{q} \to ZZ \to 4l$	/ZZTo4L_13TeV_powheg_pythia8/[1]	1.256 pb
$q\bar{q} \to ZZ \to 4l + jets(EWK)$	/ZZJJTo4L_EWK_13TeV_madgraph-pythia8/[1]	0.0004404 pb
$gg \to ZZ \to 4e$	/GluGluToContinToZZTo4e_13TeV_MCFM701/[1]	0.00159 pb
$gg \to ZZ \to 4\mu$	/GluGluToContinToZZTo4mu_13TeV_MCFM701/[1]	0.00159 pb
$gg \to ZZ \to 4\tau$	/GluGluToContinToZZTo4tau_13TeV_MCFM701/[1]	0.00159 pb
$gg \to ZZ \to 2e2\mu$	/GluGluToContinToZZTo2e2mu_13TeV_MCFM701/[1]	0.00319 pb
$gg \to ZZ \to 2e2\tau$	/GluGluToContinToZZTo2e2tau_13TeV_MCFM701/[1]	0.00319 pb
$gg \to ZZ \to 2\mu2\tau$	/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701/[1]	0.00319 pb
$Z \to ll + jets$	/DYJetsToLL_M-10to50TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/[1]	18610 pb
	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[1]	6104 pb
$t\bar{t}$	/TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8/[1]	815.96 pb
$WZ \to 3l\nu$	/WZTo3LNu_TuneCP5_13TeV-powheg-pythia8/[1]	4.430 pb
ZZZ	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/[1]	0.01398 pb
WWZ	/WWZ_TuneCP5_13TeV-amcatnlo-pythia8/[1]	0.1651 pb
WZZ	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/[1]	0.05565 pb
$t\bar{t}+ZZ$	/TTZZ_TuneCUETP8M2T4_13TeV-madgraph-pythia8/[1]	0.001572 pb
$t\bar{t}+WW$	/TTWW_TuneCP5_13TeV-madgraph-pythia8/[1]	0.007883 pb
$t\bar{t}+Z$	/ttZJets_TuneCP5_13TeV_madgraphMLM/[1]	0.259 pb

[1] = RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v1* for 2018.

Table 4.4: MiniAOD MC simulated samples used in this analysis and their respective cross sections where They are the same used in the single Higgs VBF production into four charged leptons analysis [163]. For this work, we add two processes to the list : the $gg \to HH \to ZZ^*b\bar{b} \to 4lb\bar{b}$ background stored as SM-HH_2018_non-resonant_PowhegV2_pythia8 with a $\sigma \times$ BR of 0.00448548 fb and a HLT efficiency of 0.96; the background sample stored as tqH_HToZZTo4L_M125_13TeV with a $\sigma \times$ BR of 0.021288 fb and an HLT efficiency of 0.74.

ZH, and ttH). For the ggH mode the MiNLO HJJ extension of POWHEG is used, while in the case of WH and ZH is used the MiNLO HVJ extension. The description of the decay of the Higgs boson into ZZ and finally into four leptons is obtained using the JHUGEN generator. In the case of WH, ZH, and ttH, the Higgs boson is allowed to decay to $H \to ZZ \to 2l2X$ such that 4-lepton events where two leptons originate from the decay of associated Z or W bosons, or from the top quarks are also taken into account in the simulation. The SM Higgs boson is also simulated when produced through ggH, VBF and associated production (WH, ZH, $\bar{b}b$H), while decaying through $H \to WW \to 2l2\nu$ (in the case of $\bar{b}b$H the MadGraph5 (aMC@NLO) generator is used). The parton shower and the hadronization are carried out by PYTHIA8.209. The ZZ production via $q\bar{q}$ annihilation is generated at NLO using POWHEG V2 and PYTHIA8, with the same settings as for the Higgs signal. The $gg \to ZZ$ process is simulated at LO with MCFM. Additional MC samples of WZ, Drell-Yan(DY)+jets, $t\bar{t}$, $t\bar{t}$ + vector bosons (V), and tri-bosons ZZZ,WWZ,WZZ are generated using MadGraph5 (aMC@NLO) either inclusively or merging several jet multiplicities. The list of MC simulated data-sets along with its generator level cross section times the branching ratios of the decay channel, which are used in this analysis, is summarised in

Table 4.4. All generated samples is processed through GEANT4 simulating the CMS detector and then reconstructed through the official production chain.

4.3.2 Event Pileup Reweighting

The MC samples are generated with a reasonable idea of pileup distributions but do not exactly match the pileup distribution during data-taking which depends on experimental parameters such as the beam conditions and the instantaneous luminosity. A matching of the pile-up distribution between data and MC is done computing MC pileup weights. Figure 4.4 shows the distribution of the number of the pileup interactions in 2018 in data and MC simulation before and after pileup reweighing application. In addition to the pileup weight, each MC event needs to be assigned a weight according to the process production cross section σ and the data integrated luminosity $\mathcal{L}$. The MC event is weighted by a factor w given by

$$w = \frac{\sigma \cdot \mathcal{L}}{N} \tag{4.1}$$

where N represents the total number of MC simulated events for the corresponding process.

4.4 Objects selection for the analysis

The analysis relies on muons, electrons and jets physics objects. A selection criterion is applied to those reconstructed objects to identify a set of objects for the analysis. The description of the main selection criteria for such objects is reported in part of the following sections while a detailed description can be found in Ref. [163]. The selection of objects inherited from the SM VBF $H \rightarrow ZZ \rightarrow 4l$ analysis.

4.4.1 Electron selection

After the reconstruction process of the electrons, we end up with set of electrons that could be real electrons or fake objects that pass the electron reconstruction process and hence are considered as electron candidates. The electron reconstruction efficiency [164] is very high which means that we only miss a small fraction of real electrons but we could have an amount of **F**ake **R**ate (FR) of electrons. Therefore, a set of selection cuts are applied to the reconstructed electrons in order to decrease the amount of fake objects while keeping as much as possible the real electrons. The criteria applied to the electrons are detailed below.
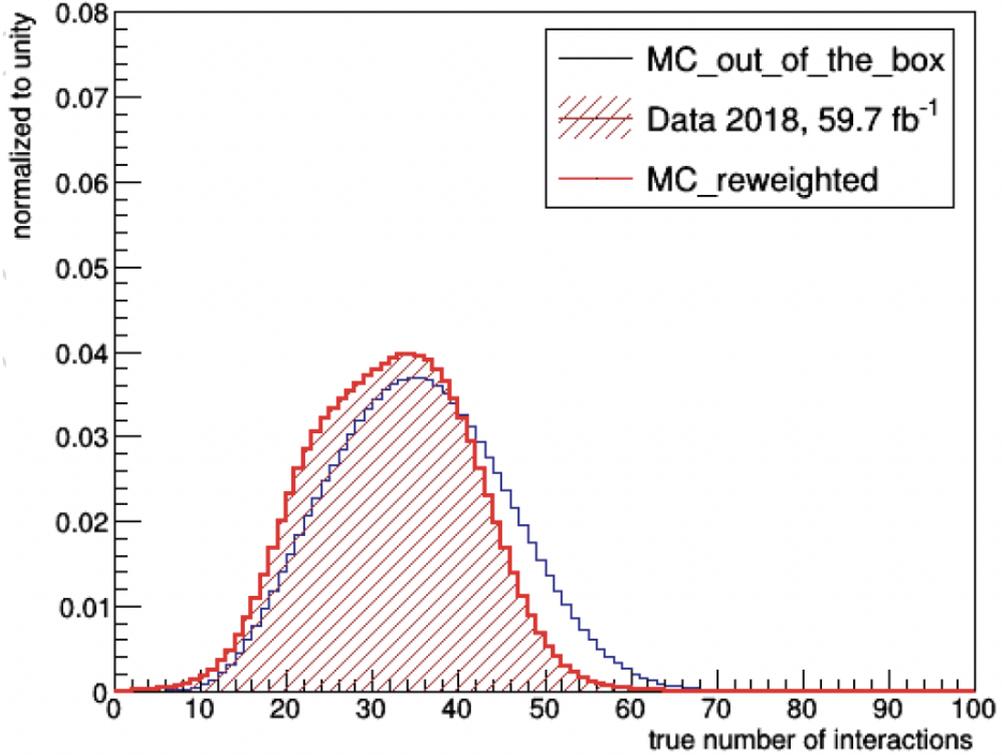
Figure 4.4: Distribution of pileup in 2018 Data and MC, before and after the application of PU weights.[163]

Electron kinematics and impact parameter

Electrons are required to have transverse momentum $p_T > 7\text{GeV}$ and $|\eta| < 2.5$ to be in the detector acceptance. Those criteria are used to reduce the contribution of the QCD background while preserving the highest possible signal efficiency. To identify the electrons coming from the primary vertex or from a secondary vertex (such as electrons coming from B meson decay or photon conversion) a cut on the electron's impact parameter is required, by satisfying a loose primary vertex PV constraint defined as $|d_{xy}| < 0.5$ cm and $|d_z| < 1$ cm, respectively the electron transverse and longitudinal impact parameters with respect to the primary vertex. In order to ensure that the leptons are consistent with a common primary vertex PV we require that they have an associated track with a small impact parameter with respect to the event primary vertex. We use the significance of the impact parameter to the event vertex, $|\text{SIP}_{3D}| = \frac{IP}{\sigma_{IP}}$ where IP is the lepton impact parameter in three dimensions at the point of closest approach with respect to the primary interaction vertex, and σ_{IP} the associated uncertainty. Therefore, a "primary lepton" is a lepton satisfying $|\text{SIP}_{3D}| < 4$.

Electron Identification and Isolation

Reconstructed electrons are identified and isolated by means of a Gradient Boosted Decision Tree (GBDT) multivariate classifier algorithm, which exploits observables from the electromagnetic cluster, the matching between the cluster and the electron track, observables based exclusively on tracking measurements as well as PF isolation sums[6]. The classifier was trained on Drell-Yan plus jets MC sample in the context of the $H \to ZZ \to 4l$ analysis. Indeed, the signal electrons from the Higgs sample and the DY sample have the same kinematics since in both samples the electrons come from Z decay. A cut on the output of the classifier score has been chosen to distinguish between the signal and the fake electrons while preserving a high signal electron efficiency. Loose electrons passing the identification cut are defined as "tight" electrons for the analysis in this thesis.A list of the input variables used for the electron identification BDT classifier is reported in Figure 4.5.

Obsevable Type	Observable symbole	Description
Cluster shape variables	$\sigma_{i\eta i\eta}$	Energy-weighted standard deviation of single crystal η within a 5×5 block of crystals centered on the highest energy crystal of the cluster seed.
	$\sigma_{i\phi i\phi}$	Similar to $\sigma_{i\eta i\eta}$ but in ϕ direction.
	η width	Supercluster width along η.
	ϕ width	Supercluster width along ϕ.
	Shower Circularity $(E_{5\times5} - E_{5\times1})/E_{5\times5}$	$E_{5\times5}$ is the energy computed in the 5×5 block of crystals centered on the highest energy crystal of the cluster seed, and $E_{5\times1}$ is the energy computed in the strip of crystals containing it.
	$E_9 = E_{3\times3}/E_{supercluster}$	Energy sum of the 3×3 block of crystals centered on the highest energy crystal, divided by the supercluster energy.
	H/E	Ratio of the hadronic energy behind the electron supercluster to the supercluster energy.
	E_{ps}/E_{raw}	Energy fraction deposited in pre-shower detectors divided by the untransformed supercluster energy.
Tracking variables	$f_{brem} = 1 - p_{out}/p_{in}$	Fractional momentum loss of the GSF track which measures the amount of bremsstrahlung emission.
	N_{KF}	Number of hits of the KF track
	N_{GSF}	Number of hits of the GSF track
	χ^2_{KF}	Goodness of fit of the KF track
	χ^2_{GSF}	Goodness of fit of the GSF track
	$N_{miss\ hits}$	Number of expected missing inner hits in the first tracker layers
	P_{conv}	Fit probability for a conversion vertex associated with the electron track
Track-cluster matching variables	E_e/p_{out}	Ratio between the energy of the ECAL cluster closest to the electron track extrapolated to ECAL and the track momentum at the outermost track position
	E_{sc}/p_{in}	Ratio between the supercluster energy and the track momentum at the innermost track position.
	$\frac{1}{E_{tot}} - \frac{1}{p_{in}}$	Deviation of the supercluster energy from the electron momentum obtained by combining ECAL and tracker information.
	$\Delta\eta_{in} = |\eta_{SC} - \eta_{in}|$	Pseudorapidity distance between the energy weighted position of the supercluster and the track position extrapolated from the innermost track position
	$\Delta\phi_{in} = |\phi_{SC} - \phi_{in}|$	Similar to $\Delta\eta_{in}$ but in the ϕ direction
	$\Delta\eta_{seed} = |\eta_{seed} - \eta_{out}|$	Pseudorapidity distance between the seed cluster position and the electron track extrapolation to the ECAL.
Isolation	$\sum P_T^{charged}$	Sum of all PF charged hadron momentum in a cone with $\Delta R < 0.3$ around the electron
	$\sum E_T^{neutral}$	Sum of all PF neutral hadron momentum in a cone with $\Delta R < 0.3$ around the electron
	$\sum E_T^\gamma$	Sum of all PF photons energy in a cone with $\Delta R < 0.3$ around the electron

Figure 4.5: List of the input variables used for the electron identification BDT classifier.

---

6 The idea of the isolation variable is to identify the real electron from fake electron (mostly coming from hadronic jet) by requiring that a real electron is not surrounded by other activity within a cone. The PF isolation computes the contribution of all the charged and neutral particles energy surrounding the electron in a cone of $\Delta R < 0.3$ and compare their sum to the electron's energy. The isolation variable is sensitive to the pileup since the pile-up events give rise to extra energy deposits that affect this quantity. The effect of pile-up in isolation variable can be decreased by associating the charged particle candidates to the primary vertex while discarding the charged particle candidates associated with the other vertices. The assignment of charged particles to a vertex is more reliable than for neutral particles (photons, neutral hadrons) which makes their contributions highly affected by the amount of pileup in the event. So the neutral component of the isolation variable needs to be corrected take into account the pileup effect.

4.4.2 Muon selection

Since the muon reconstruction is easier than for the electrons, the muons can be identified efficiently by using some cuts and without using a multivariable classifier as in the electron case.

Two types of muon selections are performed. We define *loose muons* as the muons satisfying $p_T > 5$GeV, $|\eta| < 2.4$, $|d_{xy}| < 0.5$ cm and $|d_z| < 1$cm. The muon p_T cut is slightly lower than for electrons because the low p_T reconstruction of muons is slightly more efficient, allowing to reach lower values of momentum. To distinguish between real and fake muons (coming from cosmic rays or from in-flight decays of hadrons) muons have to be reconstructed by either the *GlobalMuon* or *TrackerMuon* algorithm. Standalone muon tracks reconstructed only in the muon system or marked as global or tracker muons are rejected. Loose muons with $p_T < 200$GeV are considered tight muons if they also pass the Particle Flow muon identification (PF ID) criteria. Loose muons with $p_T > 200$GeV are considered *tight muons* if they pass the PF ID or the more relaxed definition Tracker High-pT ID whose are shown in Table 4.5.

Requirement	Description
Muon station matching	Muons are matched to segments in at least two muon stations "arbitrated tracker muon"
Good p_T measurement	$\frac{p_T}{\sigma_{p_T}} < 0.3$
Vertex compatibility (x-y)	$\left|d_{xy}\right| < 2$ mm
Vertex compatibility (z)	$\left|d_z\right| < 5$ mm
Pixel hits	At least one pixel hit
Tracker hits	Hits in at least six tracker layers

Table 4.5: The requirements for a muon to pass the tracker high-p_T ID.

An additional "ghost-cleaning" step is performed to deal with situations when a single muon can be incorrectly reconstructed as two or more muons:

- Tracker Muons that are not Global Muons are required to be "arbitrated", i.e. associated to the segments in the outer muon detectors;

- If two muons are sharing 50% or more of their segments, the muon with lower quality is removed.

Muons are required to be isolated similarly as described for electrons,but the pileup contribution subtraction is performed in a different way for muons, and the working point is chosen to be equal to the one for electrons, that is, $IP^{PF}(\Delta R = 0.3) < 0.35$. Moreover,we apply the same selection on the muon significance of impact parameter as for the electrons as described previously.

4.4.3 FSR selection

Photons are considered in this analysis as a candidates for the FSR from the leptons. Since we have four leptons in the final state, leptons can radiate energetic photons, and not taking into account this photon could affect the accuracy of the reconstruction of the four leptons in the final state. Also the FSR can affect the lepton isolation calculation if the emitted photon is located in the lepton isolation cone. In this case, the photon needs to be subtracted from the lepton isolation cone. In the analysis we use PF photons passing the following criteria:

- **Preselection cut**: the PF photons are required to satisfy a pseudorapidity cut $|\eta| < 2.4$, to have $p_T > 2$ GeV and PF isolation $I_{PF} < 1.8$. The photon isolation is computed in a cone of radius R = 0.3 with the requirement that charged hadrons have a threshold of 0.2 GeV with a veto cone of R > 0.0001 and the neutral hadrons and photons have a threshold of 0.5 GeV with a veto cone of R > 0.01, also including the contribution from pileup vertices (with the same radius and threshold).

- Supercluster veto: discard all the PF photons that match with any electron that pass both the loose ID and SIP cuts.

- Photons leptons association: photons are associated to the closest lepton in the eventthat pass the loose ID and impact parameter cuts.

- Discard all the photons that do not satisfy the cuts $\Delta R(l, \gamma)/E^2_{T,\gamma} < 0.012$ and $\Delta R(l, \gamma) < 0.5$;

- After passing the above cut, if more than one photon is associated to the same lepton, we select the photon with the lowest $\Delta R(l, \gamma)/E^2_{T,\gamma}$

- For each FSR photon that was selected, we exclude that photon from the isolation calculation of all the leptons in the event that pass both the loose ID and SIP cuts.

The muon channel has a bigger effect than electron channel, where the majority of FSR already included in the electromagnetic supercluster.

4.4.4 Jet selection

Jets are reconstructed through the anti-kT clustering algorithm using PF candidates after rejecting the charged hadrons that are associated to a pileup primary vertex. We use a distance parameter R = 0.4. To reduce instrumental background, the tight working point for the jet identification suggested by the JetMET Physics

Object Group (POG) is applied. In addition, jets from Pile-Up are rejected using the PileUp jet ID criteria suggested by the JetMeET POG.

In this analysis, jets are required to have $p_T > 20$ GeV and $|\eta| < 4.7$ and are cleaned from any tight leptons and FSR photons by a separation criterion of $\Delta R(jet, l\gamma) > 0.3$ as it was done in [165]. Since the calorimeter response to particles is not linear, JEC are needed to translate the measured jet energy to the true particle/parton energy. Standard JEC are applied on reconstructed jets, which consist of L1 Pileup, L2 Relative Jet Correction, L3 Absolute, Jet Correction for both MC samples and data, and also residual calibration for data.

b-tagging

In order to properly select the signal final state, we need to properly select b-jets. For this purpose the DeepCSV algorithm is used as b-tagging algorithm. It combines impact parameter significance, secondary vertex and jet kinematics using information of more tracks. Furthermore,the b tag output discriminator is computed with a Deep Neural Network. Data to simulation scale factors for b-tagging are provided for the entire shape of the DeepCSV b-tagger as a function of jet p_T,η and flavour. Such scale factors are applied to simulated jets by downgrading (upgrading) the b-tagging status of a fraction of the b-tagged (untagged) jets that have a scale factor smaller (larger) than one.

Quark-gluon jet discrimination

This analysis is based on a signal in which the jets originate from quarks, while the jets in the background are gluon enriched. Because of the different colour interaction and hadronization, gluon jets are wider, with higher multiplicities, and contain a more uniform energy fragmentation. In contrast, quark jets are more likely to produce narrow jets with hard constituents that carry a significant fraction of the energy. These differences can be used to construct a probability tagger capable of discriminating jets initiated by light-flavour quarks from those initiated by gluons. The quark-gluon likelihood discriminant is constructed out of three variables, each accessing the particle flow composition of the jet:

- The multiplicity, i.e., the total number of particle flow candidates reconstructed within the jet.

- The jet energy sharing variable,

- The angular spread is measured by a minor axis of the jet in the (η, ϕ) plane.

The set of variables used in the discriminator have been chosen by studying their discrimination performance on the simulation. Their discrimination power

is compared with ROC curves[7]. These curves show the efficiency of selecting a quark jet for a selection cut on the variable that rejects a given fraction of gluon jets. In addition to the ROC curves, correlations are studied to obtain the minimal set of uncorrelated variables with the highest discrimination power. A better discrimination power and stability to pileup effects are found by restricting the charged particle flow candidates to those linked to tracks compatible with the primary interaction vertex and restricting the neutral particle flow candidates to those which have $p_T > 1\text{GeV}$. The jet quark-gluon likelihood of the leading
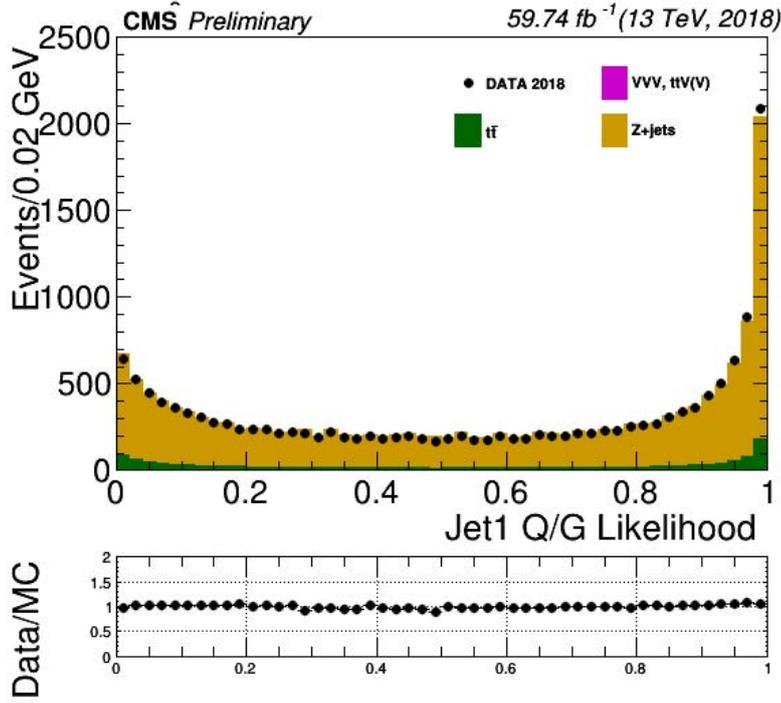
(a)

Figure 4.6: The quark-gluon likelihood of the leading jet for the 2018 data-sets used in this analysis in the ($m_{4l} < 118 \,||\, m_{4l} > 130$) with Z+$\geq$ 2 jets **C**ontrol **R**egion (CR).

jet for the 2018 data-taking used in the analysis after MC reweighting is shown

---

7 Introducing the concepts of **True Positive (TP)** - the event is signal, the prediction is signal (correct result), **FP (false positive)** - the event is background, but the prediction is signal (unexpected result), **TN (true negative)** - the event is background, the prediction is background (correct absence of signal), **FN (false negative)** - the event is signal, the prediction is background (missing a true signal event), **TPR (true positive rate)** - how often the network predicts a positive outcome (signal), when the input is positive (signal) and **FPR (false positive rate)** - how often the network predicts a positive outcome (signal), when the input is negative (background), the **ROC curve** is created by plotting the **true positive rate (TPR)** against the **false positive rate (FPR)** at various threshold settings. The true-positive rate is also known as sensitivity, probability of detection, or signal efficiency in high energy physics. The false-positive rate is also known as the probability of false alarm or fake rate in high energy physics.

in Figure 4.6 in a well-defined CR[8]. The Z + 2 jets control region is built by requiring the presence of two leptons forming a Z candidate and at least two jets, whose selection is added on top of the Z selection. This control region is chosen in order to study the agreement between data and simulation for jet observables used in the Neural network training (see Section 4.6.1)in the region with enough statistics.

4.5 Event selection

The event selection strategy is designed to gradually constrain the phase space, while having a good control of background at each step. The event selection consists of five major steps:

1. **Trigger selection**: the events that fired at least one of the HLT trigger paths mentioned in Section 4.2 are saved;

2. **Vertex selection**: the event is required to have at least one good PV fulfilling the quality requirements described in Section 3.5;

3. **Objects selection**: the events with four lepton candidates are selected from what is called *selected leptons*, which are the tight leptons defined in Section 4.4. Such leptons have $SIP_{3D} < 4$ as a vertex constraint and isolation cuts, where the FSR photons are removed from the isolation cone. A lepton cross cleaning, which discards electrons with $\Delta R \leq 0.05$ from tight muons, is also applied.

4. **Z candidates**: are built from a pair of selected leptons of opposite charge and same flavour (e^+e^-, $\mu^+\mu^-$), having invariant mass satisfying $12 < m_{ll(\gamma)} < 120$ GeV, where the Z candidate mass takes into account (if any) a selected FSR photon;

5. **ZZ candidate selection**: are built from a pair of Z candidates which do not have common leptons (non-overlapping). The Z with closest $m_{ll(\gamma)}$ to the nominal Z boson mass is denoted as Z_1 and the second one is the Z_2. The built ZZ system must satisfy the following requirements:

   - **Ghost removal**: any two leptons must have $\Delta R(\eta, \phi) > 0.02$;

   - **Lepton** p_T: at least two out of the four leptons must have $p_T^i > 10$ GeV/c and $p_T^j > 20$ GeV/c ;

   - **QCD suppression**: all opposite-sign lepton pair that can be built out of the four leptons (regardless of lepton flavour) must satisfy m_{ll}

---

8 A control region is a region of phase space that is orthogonal to the signal region and has higher statistics than it.

> 4 GeV/c^2. Here, the selected FSR photons are not included in the mass computation, since QCD-induced low mass di-lepton (e.g. J/Ψ) may have photons nearby (e.g. from π_0);

- **mass Z$_1$**: $m_{Z_1} > 40$ GeV/c^2.

- **Smart cut**: defining Z_a and Z_b as the mass-sorted alternative pairing Z candidate (Za being the closest one to the nominal Z mass), require NOT($|m_{Z_a} - m_Z| < |m_{Z_1} - m_Z|$ AND $m_{Z_b} < 12$). Selected FSR photons are included in the m_Z's computation. This cut discards 4μ and $4e$ candidates which have the alternative pairing similar to an on-shell Z + a low-mass pair l^+l^-;

- $m_{4l} > 70$ GeV/c^2

- **Choice of the best ZZ candidate**: if more than one ZZ candidate survives the previous selections, the one with the highest four leptons p_T scalar sum is chosen.

6. **SM Higgs selection**: events containing at least one ZZ system satisfying all the previous selections are then used to tag the Higgs decaying into four lepton final state, as it is done for the SM Higgs analysis.

7. **VBF signal region SR**: in order to enhance the presence of VBF HH events over the other SM Higgs production modes and the SM background, additional cuts are applied as discussed in the next section.

4.5.1 VBF HH Signal Region definition

In order to test physics models one has to define a signal region, which is expected to be populated with events from the model of interest while having low background rates. One can define the signal region by choosing observables that are sensitive enough to discriminate the signal from backgrounds. In this section, a cut-based selection is performed for several observables and choose the cuts that are suitable for all signal models.

Specifically, in order to optimize the analysis for the VBF production mode few additional requirements are imposed to events selected after the steps described above (until step 4). I require that:

- **Full selection** of $H \rightarrow 4l$;

- **Number of (reconstructed) jets** must be more than *four jets*;

- **Number of leptons** equals exactly four leptons;
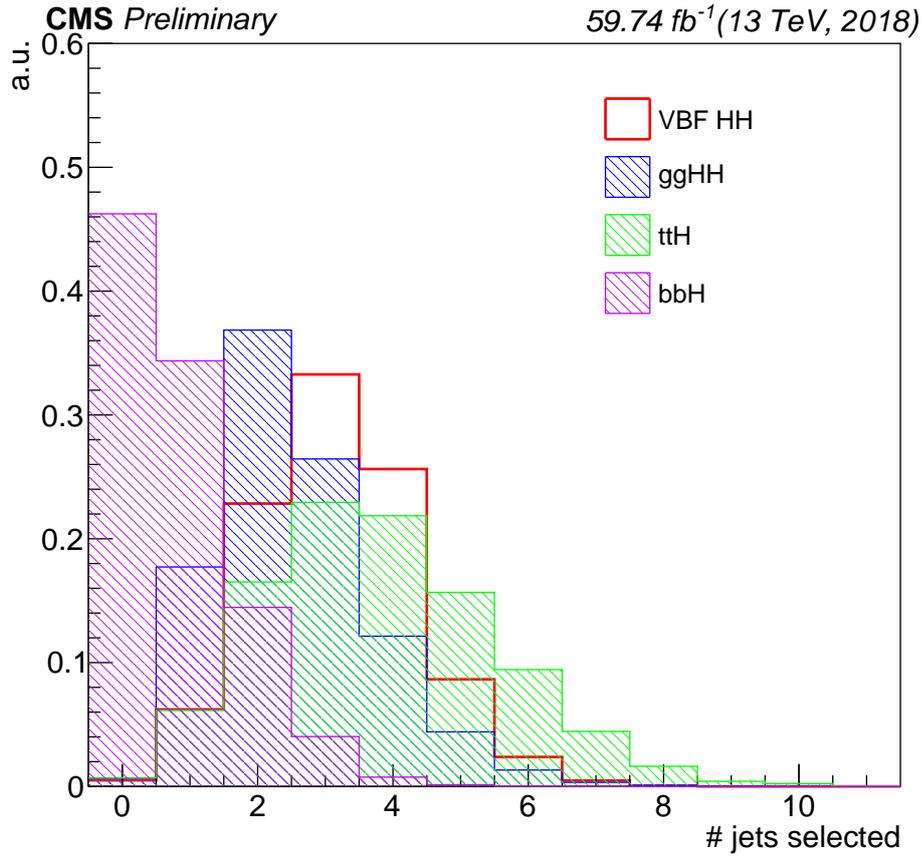
Figure 4.7: Normalized distribution of the RECO jets number for the signal and main backgrounds of the VBF HH analysis. The choise of the definition of a VBF HH SR with more than four jets is related to the presence of less background events in the phase-space selected.

- **Four-lepton invariant mass**: events must have $115 < m_{4l} < 135$ GeV/c^2 since the significant fraction of VBF yields it is contained within that range.

The choice of selecting at least four jets in the event can be simply explained looking at the distributions of the RECO jets for the signal and main backgrounds of the analysis in Figure 4.7. Indeed, it starts presenting differences for $N_{RECOjets} \geq 4$. In Figure 4.8 some of the variables that will be used in the training of the algorithms are presented for the signal and the main backgrounds.

After these event selections, information about the two jets with the highest **DeepCSV** b-discriminator are stored along with the first six highest p_T jets (if they do not coincide with the b-tagged jets) as it will be motivated in Section 4.5.2.

The strategy has been training the signal versus all the backgrounds in the signal region (ttZ, ttW, SM Higgs, VBF, WH, ZH and ZZ) - except for the $t\bar{t}H$ whose similarity with signal events would not permit a good training of a

Deep **N**eural **N**etwork (DNN) - by exploiting the different Higgs decay modes: $H \rightarrow ZZ \rightarrow 4\mu$, $H \rightarrow ZZ \rightarrow 4e$ and $H \rightarrow ZZ \rightarrow 2e2\mu$; the shape of the DNN for signal and background will be used as input in the statistical analysis as explained in details in Section 4.6.1.
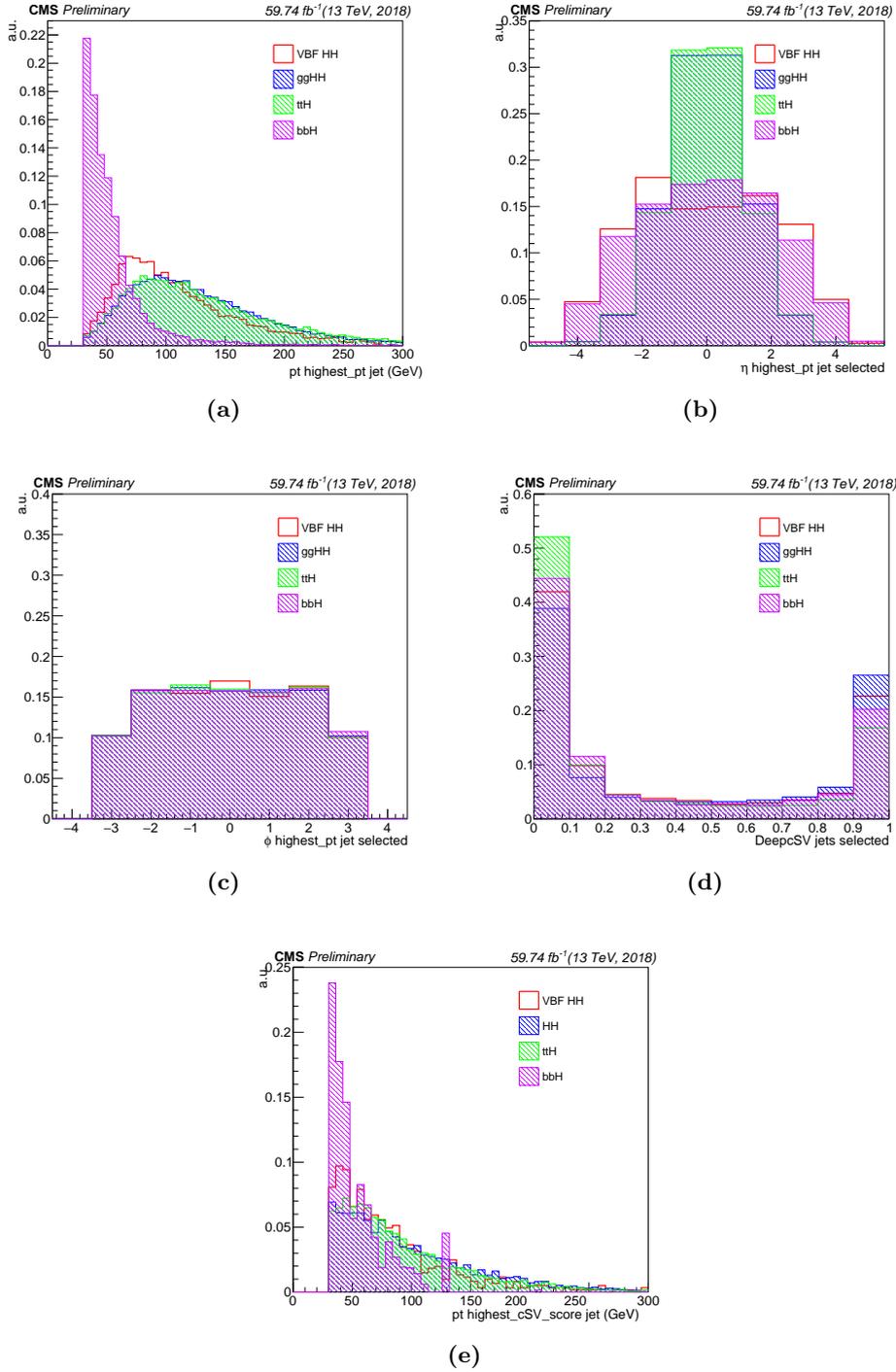
Figure 4.8: Normalized distributions of some physical observables for the HH VBF signal and the main background processes in the VBF SR.

4.5.2 Jets Candidate Selection

Events with less than four reconstructed jets are discarded and the pairs of jets candidates (bb-candidates) for the reconstruction of the second Higgs boson in the HH pair are built from the jets that pass all criteria described in Section 4.4.4 with $|\eta| < 4.7$ (in order to not reject events in which b-jets occupy the tails of the η distributions). In particular, as a first approach, if there are only *four jets* in the event, the bb candidate is formed with jets having the highest b-tagger score in the event. The b-tagger requirement is not used to further reject events.

An alternative procedure of selecting b-jets for the bb candidate would have been using official working points for the jets' *DeepCsV* score by defining the categories *loose*, *medium* and *tight* b-jets and selecting only *tight* jets, but the former approach is much more general. Moreover, as shown in Figure 4.9, the distributions of the number of tight, medium and loose jets do not present great differences between the signal and the main backgrounds and these information cannot be used to select the physical process of interest.

bb and VBF jets candidate selection study

A study was performed on the SM VBF HH signal Monte Carlo sample in order to verify that the way chosen to select the two jets is the most efficient and to persevere in the categorization-direction to build a bb candidate. A cone of radius $\Delta R = 0.4$ is built around each RECO jet selected; then we search for a b-quark coming from the decay of one Higgs boson and we check if the considered RECO jet is matching with the b-parton, inside the cone previously defined. For this study many different ways for selecting the two jets for building the di-jet candidate are considered:

- select the two highest p_T jets in the event;

- select the two highest DeepCsV b-tagger score jets in the event;

- select the highest b-tagger score jet and the highest p_T jet in the event (verifying that they are not the same jet);

- select the two highest m_{jj};

- select the two highest $\Delta\eta$;

The same study about b-jets matching efficiency has been performed without imposing any category selection but only kinematic cuts which were described in the previous sections as shown (cuts on p_T, η variables and cleaning from leptons). We observed a negligible change in the results by requiring different values of ΔR and η for the jets. Moreover, the same check has been performed matching
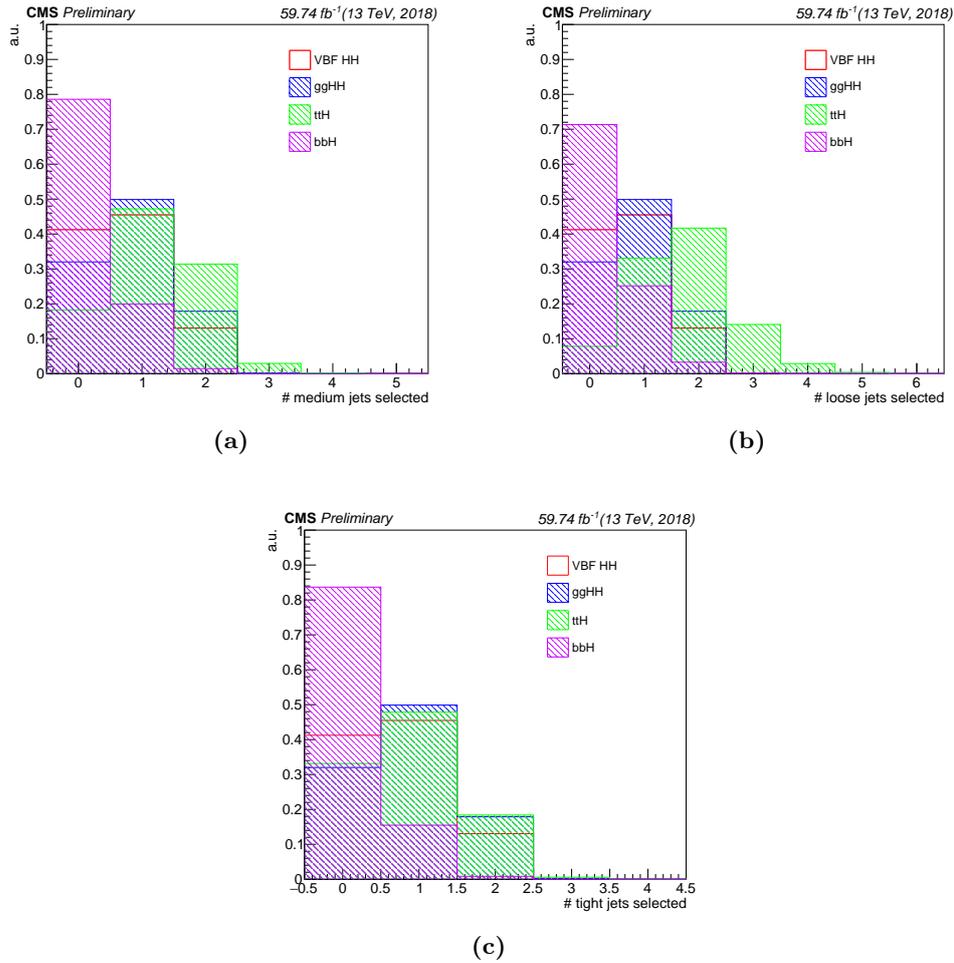
Figure 4.9: Normalized distributions of the number of loose, medium and tight RECO jets for the simulated signal and the main backgrounds. Working points are established imposing a DeepCsV cut of 0.1241, 0.4184,0.7527 respectively. The study is done out of the VBF SR.

the LHE partons of the two VBF forward jets to the same different categories of RECO jets in the event. Figure 4.10 and Figure 4.11 report respectively the result of the categorization (purity) and reconstruction (efficiency) study which brings us to not identify Higgs(bb)-candidate and VBF jets by choosing one of these categories since the low efficiency. Indeed, making a choice would bring us to discard a lot of events.
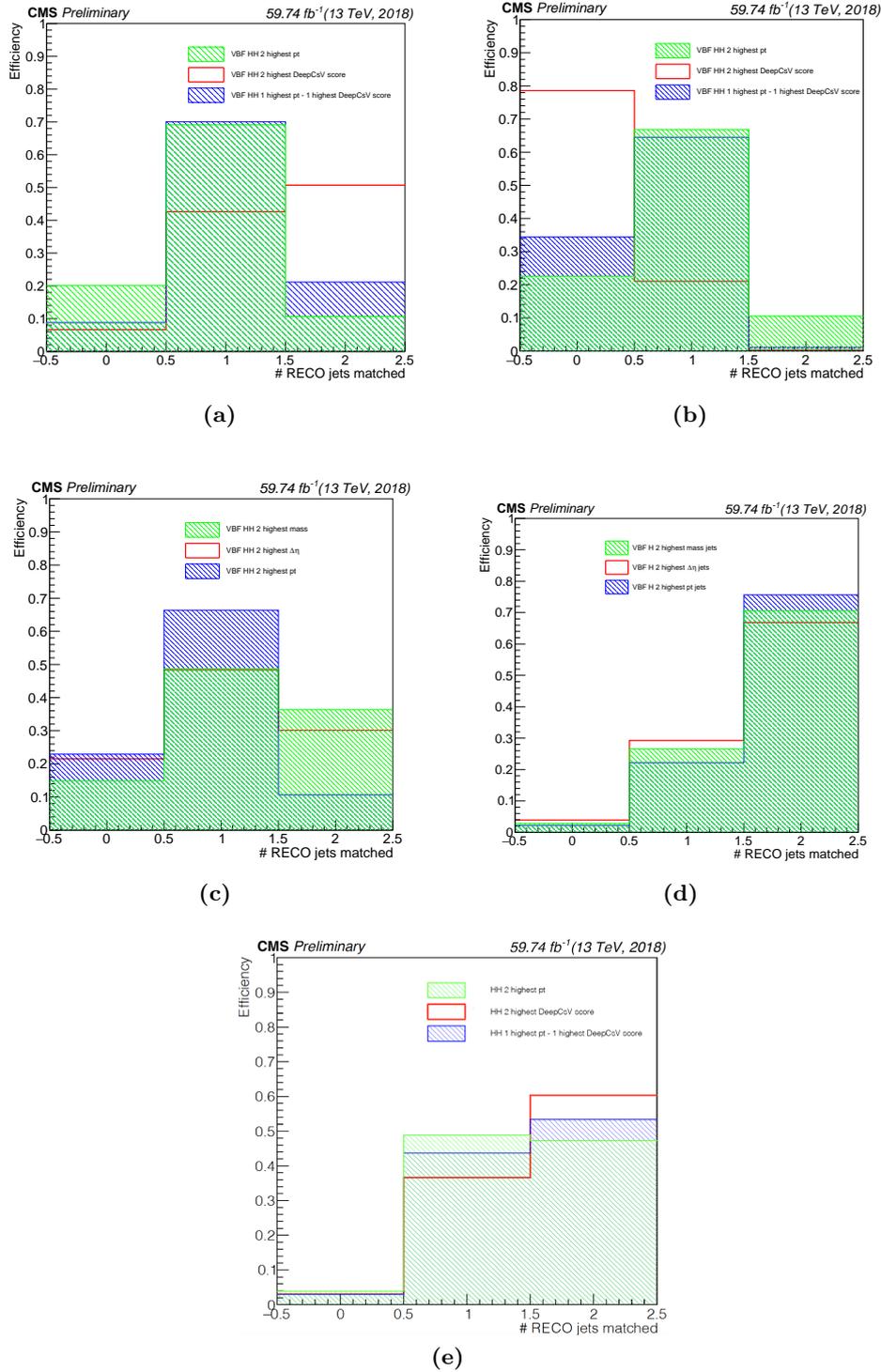
Figure 4.10: RECO-GEN jet matching study with the definition of categories (purity checking). Figure 4.10a. Efficiency of RECO jets matching with the two b-partons coming from the decay $H \rightarrow b\bar{b}$ for the three categories (two highest p_T jets,two highest b-tagging score etc.). The presence of one mis-matched jet as a prevalent behavior does not permit to choose one of these categorizations for building a bb-candidate. Figure 4.10b. Efficiency of RECO jets matching with the two LHE VBF partons for the three categories. The presence of one mis-matched jet as a prevalent behavior does not permit to choose one of these categorizations for identify the VBF jets. Figure 4.10c Efficiency of RECO jets matching with the two LHE VBF partons for other three categories. The presence of one mis-matched jet as a prevalent behavior does not permit to choose one of these categorizations for identify the VBF jets. Figure 4.10d. Matching for the Higgs VBF background of the two VBF jets with the LHE VBF partons for three categories. The choice of the two highest p_T jets in the analysis [163] can be motivated by this matching study.Figure 4.10e.Matching for the HH background of the two b-jets coming from the decay $H \rightarrow b\bar{b}$ with the b-partons for three categories. The choice of the two highest b-tagging score jets in the analysis [165] can be motivated by this matching study.

(a)

(b)

(c)
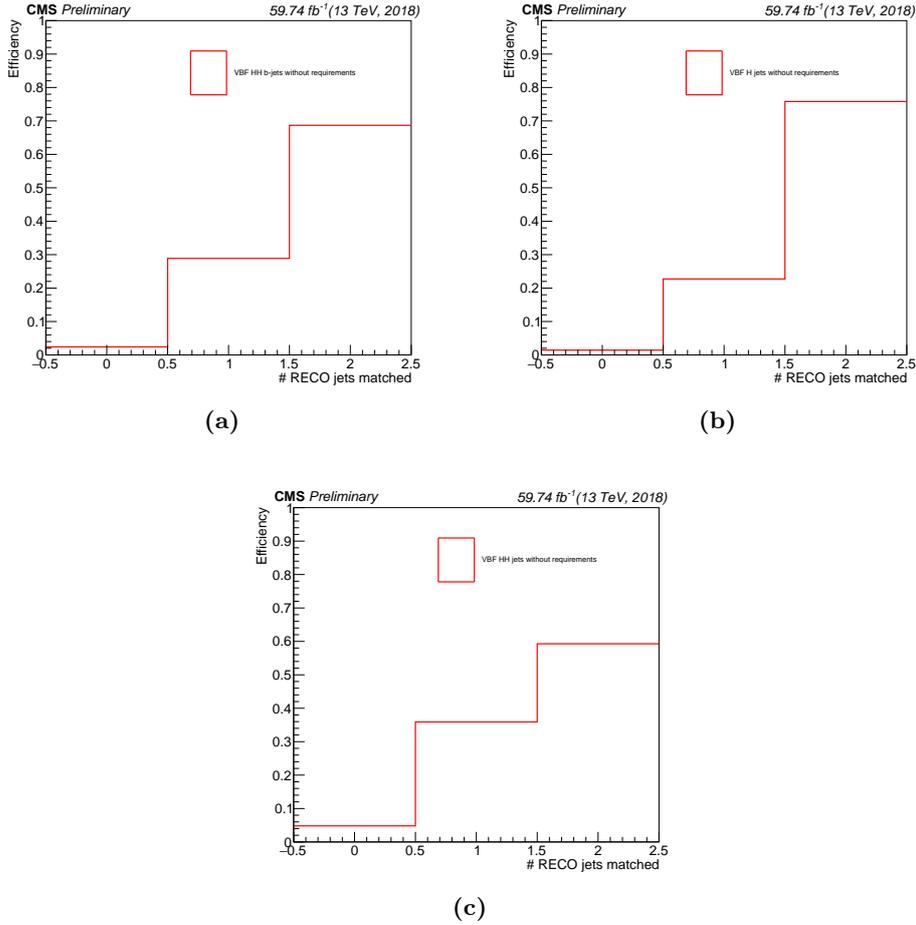
Figure 4.11: RECO-GEN jet matching study without the definition of categories (reconstruction efficiency studies). Figure 4.11a. Efficiency of RECO jets (without categorization) matching with the two b-partons LHE partons in the VBF HH Higgs bosons events. We analysed b-jets categories in Figure 4.10a. Figure 4.11b. Efficiency of RECO jets (without categorization) matching with the two VBF LHE partons in the VBF Higgs boson events. The high efficiency allows to define categories as done in Figure 4.10d. Figure 4.11c. Efficiency of RECO jets (without categorization) matching with the two VBF LHE partons in the VBF HH Higgs bosons events. We analysed VBF jets categories in as shown in Figure 4.10b and Figure 4.10c.

However, we report in Figure 4.12 the invariant mass m_{jj} and $\Delta\eta_{jj}$ built from the two (VBF or b-quark) jets selected with the five methods. It is possible to notice that the invariant mass built from the RECO jets selected with the 2-highest-b-tagger-jets method has less events in the right tail of the distribution w.r.t. the other methods, since with this method less backgound jets (meaning, jets not coming from the Higgs boson decay in two b-jets) are selected.
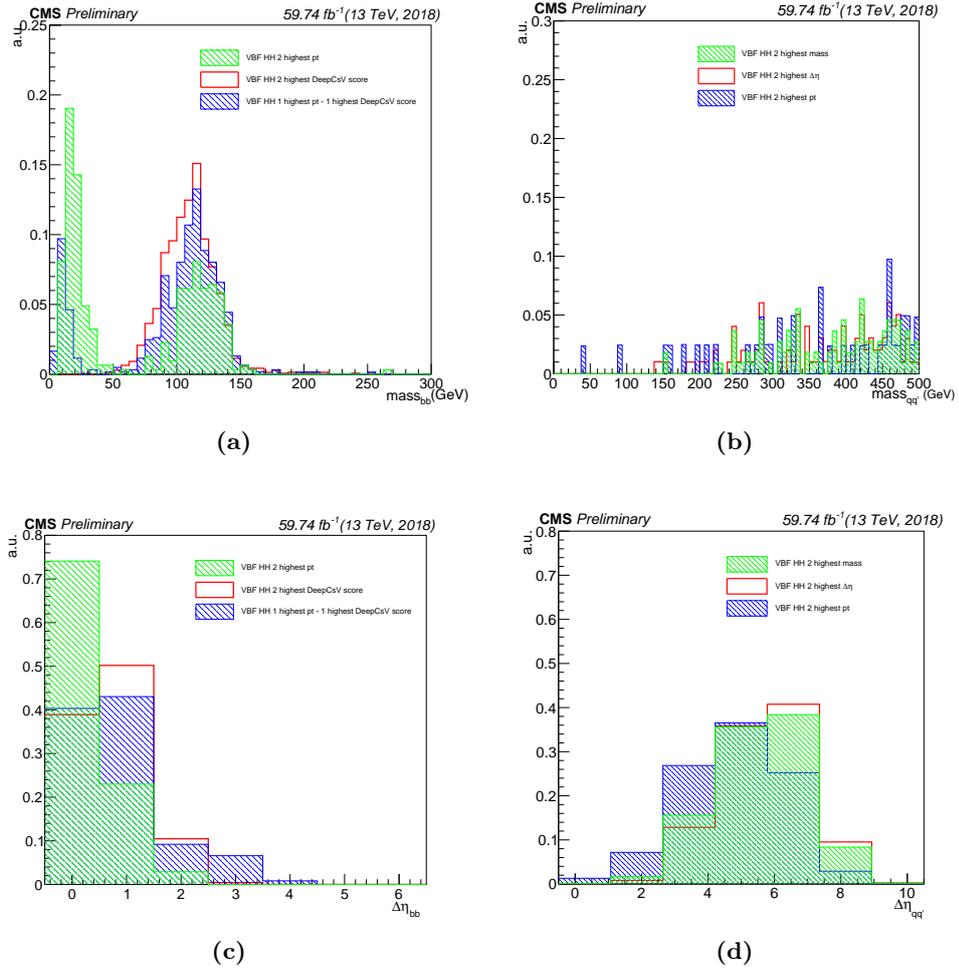
Figure 4.12: Normalized distributions of physical observables (invariant mass and $\Delta\eta_{jj}$) for different categories of RECO-GEN matching jets in case of b-jets (bb) and VBF (qq') jets.

4.6 The Multivariate Analysis approach

As shown in the previous sections, the Di-Higgs signal is quite small compared to the main backgrounds (ZZ+jets, single-Higgs,HH ggF). In order to cope with the limitations due to the low statistics and to enhance the signal to background discrimination the MVA approach is used in this chapter. This method is used to integrate various kinematic observables into a single discriminant to better discriminate between signal and background events and to improve the sensitivity of the analysis. The output distribution of this method is used to extract the 95% C.L. upper limits on the signal strength and to obtain the exclusion limits on the Higgs couplings.

The Multivariate analysis approach exploits the features of different variables to build one single classifier able to discriminate between two categories such as signal and background events. Suppose we have two classes of events, signal and background, described by the same set of variables $\vec{x} = (x_1, x_2, x_3, ..., x_d)$ in a d-dimensional space, the goal of MVA is to construct a function that is able to take subsequent decision. A simple two dimensional example is shown in Figure 4.13 where for the two variables x_1 (left), x_2 (middle) shown for signal (blue) and background (red) it is not possible to choose clear cuts on x_1 and x_2 that could separate the two classes. Alternatively, in two dimensions the two classes are largely separable with a cut applied to the linear function of the two variables which define a linear discriminant $D(x)$ (right).

The $D(x)$ is an example of a MVA discriminator, also called *discriminator*, *score* , or *classifier*, which uses multiple variables to learn from the features of the input data variables to respond correctly to never-seen-before data by using its *output*. The MVA discriminator score is used as a test statistic and is then adopted to perform the signal selection. It could be used as a variable on which a cut can be applied under a particular hypothesis test.

In particular, ML tools are models which have enough capability to define their own internal representation of data to accomplish two main tasks: learning from data and make predictions without being explicitly programmed to do so. This method is called **supervised learning** where training data-sets (from MC simulations) are used as inputs (signal and background) to build a mapping function $\vec{y} = f(\vec{x}; \vec{w})$ from the inputs $\vec{x}$ to the output $\vec{y}$ where $\vec{w}$ are some adjustable parameters that permits to match correctly the examples in the data-sets with their true labels. A good habit is training multiple models with various hyper-parameters on a "reduced" training set (i.e. the full training set minus the so-called *validation set*), and then select the model that performs best on the validation set. Once, the validation process is over, one can re-train the best model on the full training set (including the validation set) and receive the final model. Once the training has been performed, the discriminator score is computed in a separated, independent data-set for the two hypothesis (in a binary classification task) . A comparison is made between test and training classifier and their performances (in terms of ROC curves) are evaluated. If the test fails and the performance of the test and training are different, this could be a symptom of over-training and our model should be discarded.

For this analysis, I have used a DNN implemented by using the Keras **A**pplication **P**rogramming **I**nterface (API)[9] [166].

---

9 Keras is a widely used and powerful set of python libraries that allows one to quickly create any type of known **N**eural **N**etwork (NN). It contains sets of modules that simplifies much of the coding needed by the user.

Keras has recently been integrated into **T**oolkit for **M**ultivariate **D**ata **A**nalysis (TMVA) [167] and the user has the flexibility to either prepare a C++ code version or to transfer a python version to TMVA which plugs it into Keras. In this analysis an entire Keras-python framework was developed to make the studies.

(a)
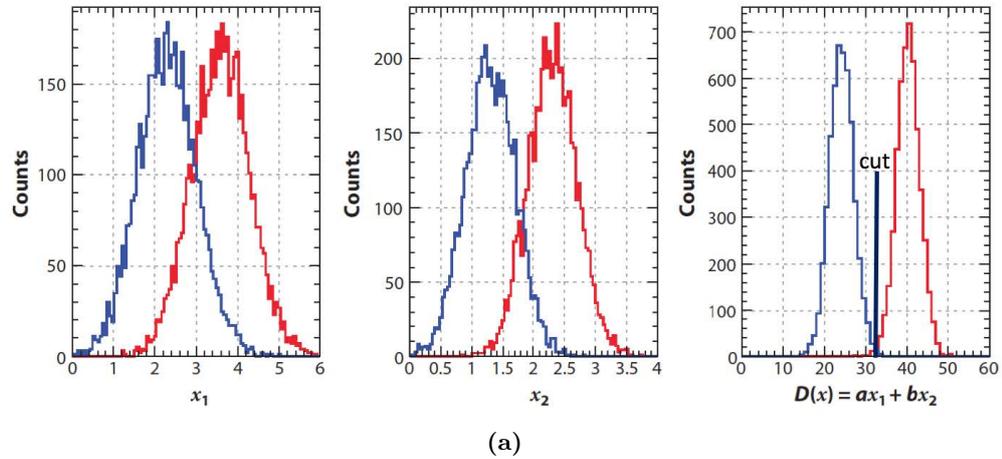
Figure 4.13: Distribution of two variables x_1 (left), x_2 (middle) from two classes signal (blue) and background (red) and the linear discriminant between the two classes (right) with the optimal cut shown with the vertical line [168].

4.6.1 Deep Neural Network

An ANN (frequently referred just as NN) is a biology-inspired analytical model composed by interconnected assembly of processing units usually called *neurons* (also *activations* or *perceptrons*). Each basic element receives input, changes their state according to the input and produces an output. Such a design was motivated by analogy with the brain, which can be thought as a highly complex, non-linear and parallel computer. It is estimated that the human brain has 100 billion neurons and each one of them typically receives thousands of connections from other neurons. The inter-neuron connections are mediated by electrochemical junctions, the so called synapses, which happens on branches of the cell called dendrites.

These thousand of signals are then combined and depending of the result of this combination, the neuron outputs a signal to its neighbourhood [169]. One major point of difference between an ANN and the brain is that for the same input the NN will give the same output but the brain may falter. It may not always give the same response to the same input. For this reason the NN is referred as a not perfect bio-mimetic model. On the other hand, similarly to the brain, NNs have the capability to learn from patterns and generalize the modelling of such patterns. This generalization means the possibility to correctly identify a

pattern that was not seen before by the NN during the learning process. The neurons contain a function which can receives also several arguments (inputs) and combine them into a single output value as shown in Figure 4.14a.

There are several types of functions that can be used in the neurons and the choice of the best one might depend on each type of problem. The NN architecture depends on the neurons' connections. But a typical and well representative NN representation can be seen in Figure 4.14b [169]. This kind of NN is usually called DNN due to its many hidden (intermediate between the input and output) layers. Such NN has shown capability to learn very complex tasks and its success comes from relatively recent improvements in the ML field [170]. Theoretically, it is able to do the feature extraction by itself becoming a Deep Learning algorithm.
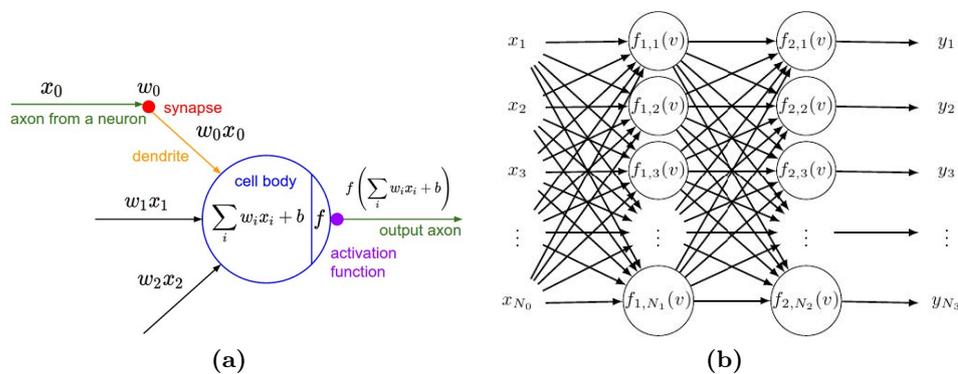
Figure 4.14: Figure 4.14a Graphical representation of an artificial neuron used in ANNs. Each input x_i (given by the user or coming from the hidden neurons) is weighted by w_i and summed to a bias (threshold) b. The sum is passed to the activation function which outputs a vector of values corresponding to each input. Figure 4.14b Graphical representation of a ANN with two hidden layers. In this type of ANN, the outputs of all neurons from a backward layer are fed into the neurons of a forward layer, making the information moved in one direction. It is also named *Feedforward Multilayer Perceptron* (MLP).

The Learning Process

As mentioned in the previous section, the way NNs can be trained allows one to classify the learning process basically in two types: unsupervised and supervised. The former includes the problems in which a given data-set contains some pattern that is not known a prior, in other words, it is not possible to label the objects (in a binary classification we define signal and background events) in the data-set. The NN can be still trained to find out possible patterns existing in the data-set. The supervised procedure is used when patterns within the available data-sets are known. The NN is then trained to correct assign a given label to a given example. The labels can arbitrarily be chosen and usually for classification task of

two classes, the denominations 0 or 1 are used. In order to verify the performance of the labels assignment the loss (cost/risk) function is used, which measures the error between the NN response and what is expected [169]. The output of the loss function is the main input used to modify the NN parameters in order to optimize the modelling of data.

The modification in the NN weight parameters w_i is done by an algorithm called *back-propagation* which consists of an application of the chain rule for partial derivatives. As an example, let j be a neuron in a intermediate layer of a neural network and $w_{ji}(n)$ the weight assigned by the neuron j to an input $y_i(n)$ at a given iteration n. The neuron produces a *local field* which is given by

$$v_j(n) = \sum_i^m w_{ji}(n)y_i(n) + b_j(n) \cdot 1 \tag{4.2}$$

in which m indicates the total number of inputs received by the neuron and $b_j(n) = w_{j0}(n)$ is the bias (which is just one-value per neuron). The local field is given as an argument to an activation function which produces the signal function $y_j(n)$ of the neuron j at iteration n,

$$y_j(n) = f[v_j(n)] \tag{4.3}$$

It is possible to compute an instantaneous error for the neuron j

$$\epsilon(n) = \frac{1}{2}e_j^2(n) \tag{4.4}$$

where $e_j(n)$ is the signal error which is defined as the difference between the neuron signal function $y_j(n)$ (its output after the activation function) and the expected response $d_j(n)$, $e_j = y_j(n) - d_j(n)$. The expected response $d_j(n)$ can be the label identifying a training example - for neurons in the output layer - or an actual response that a hidden neuron - in the intermediate layers - should have based on the example label. The previous equation represents the *loss* or *empirical risk* and it is an example of loss function used here in order to exemplify the NN learning process. The processes described so far can be graphically represented as shown in Figure 4.15a and replacing the derivatives accordingly like so,

By taking the partial derivative of the loss ϵ by application of the chain rule one gets,

$$\frac{\partial \epsilon(n)}{\partial w_{ji}} = \frac{\partial \epsilon_n}{\partial e_j(n)} \frac{\partial e_j}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \tag{4.5}$$

and substituting the derivatives :

$$\frac{\partial \epsilon(n)}{\partial w_{ji}} = -e_j(n)f'[v_j(n)]y_i(n). \tag{4.6}$$

(a)

(b)
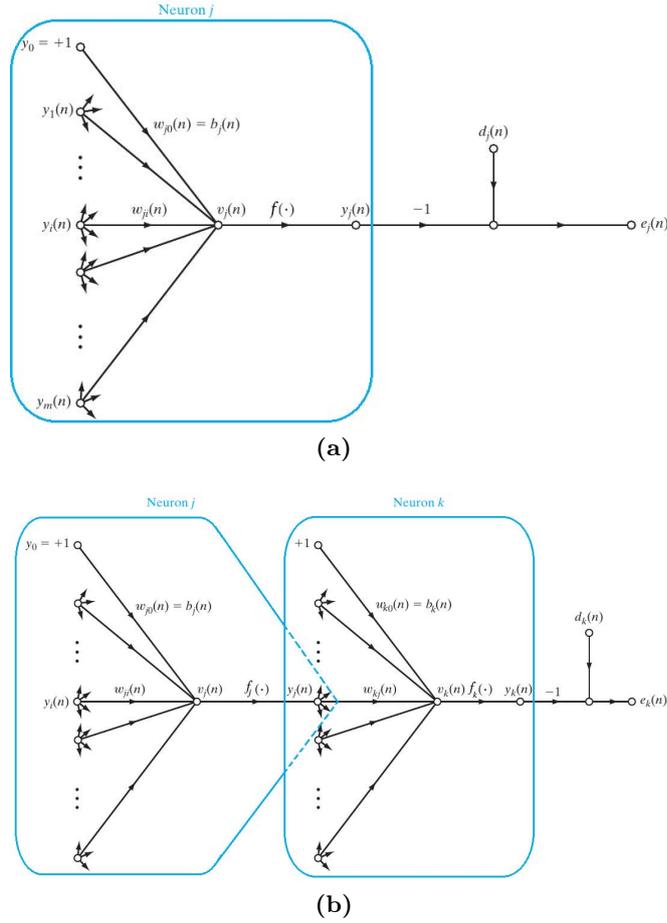
Figure 4.15: Figure 4.15a The flow of signal through the neuron j up to its instantaneous error e_j. Figure 4.15b The flow of signal through a hidden neuron j up to its signal function $y_j(n)$ which is forwarded as an input to an neuron k in an output layer. The activation function can vary between the layers and they are represented as f_j and f_k for the neurons j and k, respectively.

The correction applied to the weights w_{ji} at the iteration n is given by the *delta rule* :

$$\Delta w_{ji}(n) = -\eta \frac{\partial \epsilon(n)}{\partial w_{ji}} \tag{4.7}$$

where the negative sign accounts for the descending gradient in the weights space in order to move the search in a direction in which the weight change leads to decreasing of $\epsilon(n)$. The parameter η is the *learning rate* of the back-propagation algorithm. By replacing the Equation 4.6 in Equation 4.7 the expression of the weight update becomes:

$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n) \tag{4.8}$$

where $\delta_j = \frac{\partial \epsilon(n)}{\partial v_j(n)}$ is the *local gradient*. A recursive formula for the local gradient can be retrieved starting from the differentiation w.r.t. the neuron output $y_j(n)$

from the signal error expression $\epsilon(n) = \frac{1}{2}\sum_k e_k^2(n)$ where the sum is extended also to the neuron of internal layers:

$$\delta_j(n) = f_j'[v_j(n)]\sum_k \delta_k(n)w_{kj} \tag{4.9}$$

where the neuron j belongs to a hidden layer.

Preparing of the Data-sets

In order to better discriminate between signal and background events and reach a good sensitivity for this analysis, a DNN is trained by exploiting different kinematic variables. The network is trained in a region enriched with signal events passing, the VBF SR which has been defined in Section 4.5.1 using the simulated samples described in Section 4.2.

In order to optimize the analysis and properly handle the usage of the BSM signal samples, seven Higgs couplings-based categories have been defined. For each category a different NN was defined.

In order ensure an adequate distribution of events after splitting to form the exclusive training and testing sets, the following procedure was adopted for the HH VBF-SR events:

1. Each simulated process has three samples which has been obtained by applying the objects selection requirements described in Section 4.4 - one for each final state (4μ, $4e$ and $2e2\mu$). Those three samples were merged into just a single sample with the events from each final state being randomized, such that the final states randomly populated the merged sample;

2. The merged and randomized samples for each physical process were divided into two parts: one containing 80% of the total events and one containing the remaining 20%;

3. The first and second parts from each process were merged and once more the events were randomized in order to randomly populate each process inside the two independent parts, the **test** and **training set**.

Moreover, the process $t\bar{t}H$ which has very similar topology w.r.t. our signal as shown in Figure 4.8,is not included in the training set or in the testing set. Indeed, this background would require a separate training as was done in the bb$\gamma\gamma$ analysis [63]. Excluding this sample from the training, there will be a probability for the NN to recognize these events as signal with no negligible probability.

Scaling Events Contribution in the Training

Keras has some features that allows one to include scale factor (example weight) which are in general used to balance the training when the number of events from different classes are very different. In the physics scenario one could use the individual MC event weights of each event or yet the sum of those weights (which constitutes the expected yields, here not reported).

The advantage of the first approach is that some events, from signal or background, crossing the classes zones defined by the NN could have small weights and would be worthy to allow them to come in with the benefit to possibly improve the final discrimination. The disadvantage is that the individual weights might not keep the hierarchical contribution of each process since the expected cross-section is divided by the number of events (such that if one have a large number of events the weights might become smaller for important processes than the ones for other small processes). Both approaches were tested. The weight is used by Keras as a scale factor for the loss function such that during the training each example is seen by the NN with a different importance. That affects the direction in which the minimizer computes the gradients of the loss function. The advantage of doing so is that even if a process has few events, which is very frequent in many analysis, it will be properly taken into account with respect to the other process since it has more events.

The Training Procedure

Training MVA methods is a procedure that always need to be done in multiple fronts. It is hard to guarantee that for a set of inputs a set of MVA hyper-parameter[10] are the optimal choice, since there is an infinity of combinations that one can build. Those configurations (including the training data-set) can strongly affect the evolution of the MVA training.

In order to optimize the study, the framework developed in Python has the feature to build up through Keras different NN architectures and perform multiple parallel scans. These scans can run over different sets of inputs and the several parameters that needs to be configured for the training. The result of each NN training can be retrieved to produce plots which are used to validate and classify the quality of each training in terms of a metrics. In this analysis, the maximum of the efficiency times purity metrics $(\epsilon_S \cdot \pi)$[11] is used to choose the best model

---

10 The *hyper-parameters* are the variables that determine the network structure and how the network is trained. Hyper-parameters are set before training.

11 The **precision**/**purity** is the ratio $\frac{TP}{TP+FP}$ where TPR is the number of true positives and FPR the number of false positives. The **precision** is intuitively the ability of the classifier not to label as positive a sample that is negative. The **recall**/**sensitivity**/**TPR**/**signal efficiency** is the ratio $\frac{TP}{TP+FN}$ where TP is the number of true positives and FN the number of false negatives. The recall is intuitively the ability of the classifier to find all the positive samples. **Accuracy** is

for each signal vs background discrimination. As an example, the results for the
SM VBF HH signal are reported in Figure 4.16.
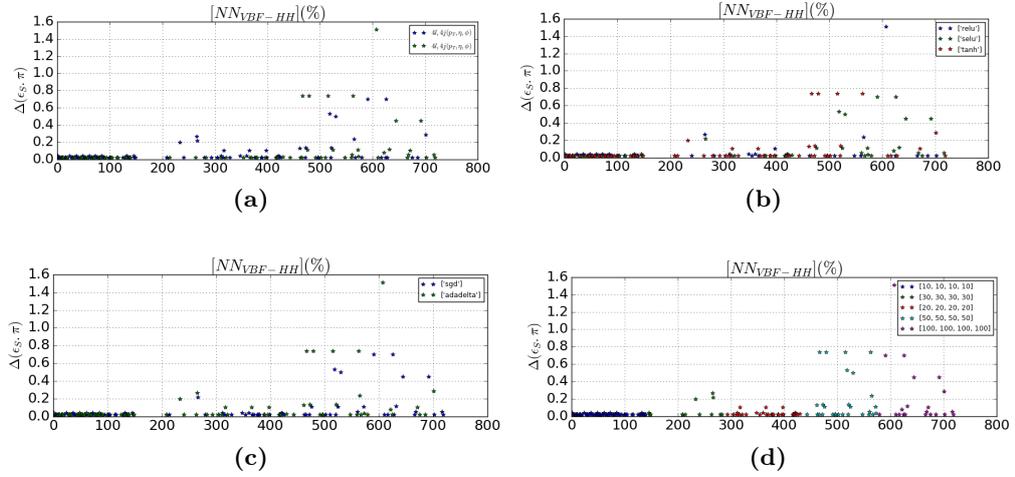
Figure 4.16: Summary of the scanning procedure for finding the best NN model in
terms of the metric $\max(\epsilon_S \cdot \pi)$ in order to discriminate the SM VBF HH
signal versus the main backgrounds of the analysis. The same procedure
has been applied for the discrimination of the BSM VBF HH simulated
signals vs backgrounds. The scanning is performed as a function of the
hyper-parameters of the NN, such as the input variables (Figure 4.16a
- the blue star indicates the absence of QG/Likelihood variables in the
input features to the NN), the activation function (Figure 4.16b), the
minimizer (Figure 4.16c) and the layers topology (Figure 4.16d).

NN hyper-parameters	Tested options
Input variables	leptons/jets (p_T, η, ϕ), jets (Q/G Likelihood, DeepCsV)
Dropout rate	10%, 20%
Topologies	10:10:10:10, 30:30:30:30, 20:20:20:20, 50:50:50:50, ...
Early stop	50, 100, 600, 3000
Minimizer SGD,	SGD, Adadelta
Batch size	5, 32, 64, 128, 786
Neuron	ReLU, SeLU, Tanh
Loss Scaling	MC event weight

Table 4.6: Summary of the scanned hyper-parameters during NN trainings.

The set of tested inputs includes combinations of the leptons and jets kine-
matic variables (p_T, η and ϕ), the jets quark-gluon likelihood and the b-tagger

---

defined as the number of good matches between the predictions and the true labels. It is always
possible to achieve high accuracy on skewed/unbalanced data-sets by predicting the most the
same output (the most common one) for every input. Thus another metric, **F1** can be used
when there are more positive examples than negative examples. It is defined in terms of the
precision and recall as $\frac{2 \cdot precision recall}{precision + recall}$. In our case, we will use a simplification of this metric
that is the product signal times efficiency $\epsilon \cdot \pi$.

discriminant associated to each RECO jets. The scanned NN hyper-parameters are summarized in Table 4.6.

These parameters are:

- **Topology**: the architecture of the NN, that is, the number of hidden layers, neurons per each layer and the neuron type (identified with their activation function). In Keras there are several types of neuron which even includes possible learning parameters (during training). The Rectified Linear Unity (ReLU) is the most recommended due to its property of non-vanishing gradient.

- **Batch size**: the number of events in a subset from the training set used to compute the gradients and update the neuron weights w_i's in each neuron.

- **Epochs**:the number of iterations over the full training set. The total number of iterations is a combination of the size of the training set, the batch size and the number. For instance, setting a training of 10 epochs and a batch size of 10 for a training set size of 100 means that Keras performs 102 updates on the neuron weights.

- **Dropout rate**: in the training with dropout [171], a fraction of the nodes of a layer are randomly zeroed at each iteration. adding dropout to the DNN ensure the regularization and make the network more robust.

- **Early stop**: a parameter to set the number of epochs which Keras should wait if not improvement in the loss function is observed. If still not improvement is seen after that number of epochs the training is stopped.

- **Minimizer**: the algorithms used to minimize the loss and based on stochastic gradient descent, where the parameters are updated at each step by the gradient multiplied by the learning rate indicated here as ϵ.

- **Scaling**: Keras permits to scale the loss function by some weight that can be independent for each training example or the same for a entire class (signal or background, for instance). In this analysis the individual weights of each event are used (w).

The Deep Neural Network results

The different NNs models which are generated through the training procedure which also validates the choice of the best NN architecture, are produced together with some plots as shown in Figure 4.17. They comprehend the metrics adopted

(the area under the ROC curve[12], $\epsilon \cdot \pi$) for the hyper-parameters' choice, and the plots which help to check if there was over-fitting for a given configuration: the NN distributions for the training and test sets,accuracy, and loss. The ROCs curve from training and test sets are also used to check if there was over-fitting. In particular, in Figure 4.17c I report the results of another multivariate discrimination algorithm, the **M**atrix **E**lement **L**ikelihood **A**nalysis (MELA)[13], which was largely used in Run I CMS analysis. The final NN for each of the signal-based categories were chosen to be the one with highest $max(\epsilon_S \cdot \pi)$.

It is important to stress that the whole analysis and optimization procedure has been performed without looking at the data in the signal region in the invariant mass spectrum. This procedure, known as blinding, is used to avoid to introduce biases in the workflow. This is the reason why the data points are not shown in the signal region plots. The un-blinding of the data in the signal region can be performed only after the final approval by the CMS collaboration.

4.7 Systematic uncertainties

The uncertainty on the measurements has two components: **statistical** and **systematical** one. The statistical uncertainty depends on the amount of data collected while the systematic uncertainty arises from the inaccuracy of the measurements coming from the detector or from theoretical calculations. We can divide the systematic uncertainties affecting the results into two categories: **experimental sources** due to the insufficient knowledge of the detector response and the theoretical sources which affect the modeling of signal and backgrounds. Those uncertainties are taken as nuisance parameters in the statistical analysis described in Section 4.8. They are treated as a **normalization uncertainty** which affects the signal/background yield or **shape uncertainties** which affect the distribution of the observables. In this analysis, the general strategy has been to rely firstly on the studies done for the SM VBF H→ZZ→4l search [163] since the selections applied in the present analysis are based on it. Then, the uncertainties associated directly to the signal and background modeling by the Neural Network, due to the uncertainties associated with the objects used as inputs, are discussed.

---

12 By computing the area under the ROC curve (AUC), the curve information is summarized in one number. The AUC is the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. The higher the AUC, the better the performance of the classifier. If the AUC is 0.5, the classifier is uninformative, i.e., it will rank equally a signal or a background observation.

13 MELA uses kinematic inputs to build a kinematic discriminant for signal to background discrimination using two invariant mass and five angles. It was largely used in the $H \rightarrow ZZ^* \rightarrow 4l$ analysis [172].

(a)

(b)

(c)

(d)

(e)

(f)
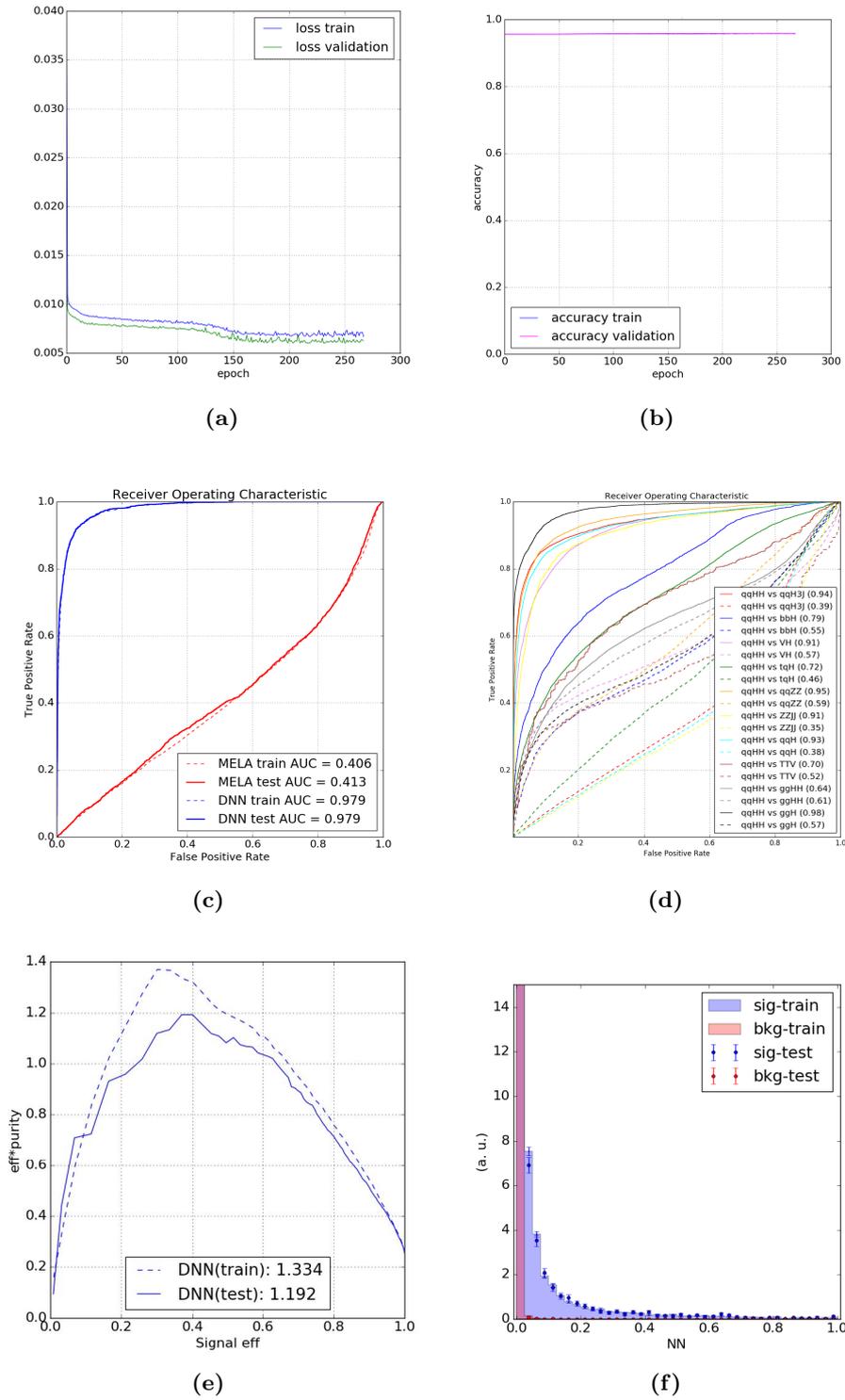
Figure 4.17: Deep Neural Network evaluation results in terms of metrics for one of the BSM (the one with the highest cross section with Higgs couplings C_V equals 1.5, C_{2V} and C_3 equal 1 vs backgrounds training. Figure 4.17a and Figure 4.17b show the behaviour of the loss (accuracy) metric as a function of the number of epochs. The super-imposition of the training loss (accuracy) and the loss (accuracy) applied to the validation set are symptom of no over-training. Moreover, the MELA discriminator's poor performances are also shown. Figure 4.17c ROC curve plots for the training and test data-sets for discriminating signal (VBF HH) and its backgrounds (excluding the ttH). The same NN has been trained independently on the single backgrounds vs the signal in Figure 4.17d, showing worst performances due to the low statistics of the single simulated samples. In Figure 4.17e and Figure 4.17f other metrics (efficiency per purity and the NN score for signal and background) to make over-training checks are shown.

4.7.1 Experimental Uncertainties

There are different experimental systematic uncertainties sources that affect both signal and background processes:

- the uncertainty on the integrated luminosity, that results in a lnN contribution;

- the lepton identification and reconstruction efficiency, that results in a lnN contribution;

- b-tagging scale factors related systematics, that result in 18 shape contributions due to the hadronic composition of the jet, the jet energy and the jet η;

- jet energy scale (JES) and jet energy resolution (JER). This systematic uncertainties are computed by propagating the up and down variation of the jet energy through the event reconstruction chain up to the DNN; this uncertainty results in a shape contribution; JES uncertainties are splitted in 11 different sources, as recommended by the POG [173].

The summary of experimental systematic uncertainty is reported in Table 4.7.

Experimental uncertainties	
type	2018
Luminosity	2.5%
Leptons ID and reco eff	$1.0 - 11\%$
b tagging SF	shape
Jet energy scale	shape
Jet energy resolution	shape
Z+X uncertainties	$30 - 37\%$

Table 4.7: Summary of experimental systematic uncertainties. The Q/G Likelihood uncertainty like a shape contribution should be added in the list.

4.7.2 Theoretical Uncertainties

Theoretical uncertainties sources are the choice of PDF set, the uncertainty on α_S, the renormalization and factorization QCD scale. These uncertainties affect both signal and background processes. For the HH signal, in addition to the uncertainty sources just described, also an uncertainty related to missing finite top-quark mass effects gives a contribution. Values for the systematic uncertainties affecting the HH signal are taken from the related Twiki page [173]. For the single

Higgs backgrounds, the values of the systematic uncertainties are taken from the related Twiki page [174]. For all the other backgrounds, the values of systematic uncertainties are computed by varying the QCD scale and the PDF set used for computing the sample cross section. An additional uncertainty of 10% on the k-factor is used for the gg → ZZ prediction and of 0.1% for the qq → ZZ prediction. The summary of theory systematic uncertainty is reported in Table 4.8.

Theory uncertainties	
PDF set and α_s HH → 4ℓb	3.0%
QCD scale HH → 4ℓb	2.2 − 5%
m_{top} unc HH → 4ℓb	2.6%
PDF set ggH	1.8%
α_s ggH	2.59 − 2.62%
QCD scale ggH	4.27 − 6.49%
PDF set and α_s VBFH	2.1%
QCD scale VBFH	0.3 − 0.4%
PDF set and α_s ZH	1.6%
QCD scale ZH	2.7 − 3.5%
PDF set and α_s WH	1.3%
QCD scale WH	0.5%
PDF set and α_s bbH	3.2%
QCD scale bbH	4.6 − 6.7%
PDF set and α_s ttH	3.6%
QCD scale ttH	6.0 − 9.2%
PDF set and α_s qqZZ	3.1 − 3.4%
QCD scale qqZZ	3.2 − 4.2%
Electroweak correction qqZZ	0.1%
PDF set and α_s ttW	25 − 37.5%
QCD scale ttW	3 − 4%
PDF set and α_s ttZ	7 − 14%
QCD scale ttZ	2 − 3%
PDF set and α_s VVV	2 − 17%
QCD scale VVV	3%
PDF set and α_s ggZZ	3.2%
QCD scale ggZZ	4.6 − 6.7%
Electroweak correction ggZZ	10.0%

Table 4.8: Summary of theory systematic uncertainties.

4.7.3 Discriminants Shape Uncertainties

The systematic uncertainties in data classification using the NNs were estimated by checking the impact of systematic uncertainties in the input variables on the shape of the NN output. The main types of systematic uncertainties are the ones that changes the values of the NN output, which can cause event migration across the signal and background regions defined by the discriminant. That is an issue one should take into account when applying some specific cut in the discriminant. For shape (binned) analysis the effect is amplified since the bins will shift up and down due to the uncertainties and the bin width has some impact: a smaller bin width increase the probability of event migration across the bins and one would expect larger fluctuations than in the case with larger bin width.

The systematic uncertainties of the objects used as inputs for the NNs were applied to the nominal input value. Those are the uncertainty in the leptons energy, the uncertainties on the jet energy scale, jet energy resolution, and quark-gluon Likelihood. The NNs were fed with the nominal inputs (no systematic uncertainty shifts applied) and width $\pm 1\sigma$ shifts from the nominal inputs. These shifts are done one variable at a time such that after all shifts have been done, there are $N_{Inputs} \cdot [1 + 2 \cdot N_{InputUncertainties}]$ output values for each event (and thus the same amount of NN distributions).

4.8 The statistical method and results

The discriminant D_{VBFHH}^{NN} for each VBF HH signal (computed for each final state 4μ, $4e$ and $2e2\mu$), as well their statistical and systematic uncertainties, derived for signal, background, and observed data as explained in previous sections, constitute the inputs (in the form of histograms, more precisely the integral of the DNN distributions corresponding to the expected signal and background yields is used) to perform the statistical analysis.

Such analysis is carried out by using the *CMS Higgs Combined Limit* package , which assembles a collection of RooStats-based software [175]. The upper limit at different C.L.s of the HH VBF signal strength (μ_{HHVBF}), which is the ratio between the observed and the theoretical cross-sections, is computed as a function of the Higgs couplings parameters. The statistical and systematic uncertainties are properly taken into account by defining log-normal and shape uncertainties that affect the shape analysis of the discriminants.

In order to extract the VBF HH signal strength a multi-dimensional likelihood-fit is performed. The systematic uncertainties enter in such likelihood as nuisance parameters, which are left to float during the fit. The likelihood implemented

within the Higgs Combine tool is a function of the probabilities related to the parameters of interests $\vec{\mu}$ and the systematic uncertainties in the measurements (the nuisance parameters θ_i called collectively as $\vec{\theta}$). The estimation of the nuisance parameters is done from a priori separate measurements that are independent of the measurement of the signal region and are denoted as $\vec{\tilde{\theta}}$. It can, thus, be expressed as [176],

$$\mathcal{L}(n, \tilde{\theta}|\mu, \theta) = p(n|\vec{\mu}, \vec{\theta}) \cdot \pi(\vec{\tilde{\theta}}|\vec{\theta}) \sim p(n|\vec{\mu}, \vec{\theta}) \cdot \pi(\vec{\theta}) \tag{4.10}$$

where n represent the observed data or Asimov data (from pseudo-experiment). The probability density function $\pi(\vec{\tilde{\theta}}|\vec{\theta})$ is the probability to measure a set of nuisance parameters $\tilde{\theta}$ given its true value $\vec{\theta}$.

The essence of this likelihood analysis is based on the number of observed events n or Asimov data (from pseudo experiment), such that the probability term is just a Poisson probability which has the simplest form of

$$p(n|\lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \tag{4.11}$$

where the $\lambda = \mu \cdot s + b$, being s and b the representation of the expected yields for signal and background, respectively. In the case of a binned shape analysis (such as the present one), the probability p(n|λ) becomes the product of the individual bin probabilities (product rule of probability) p_i:

$$p \rightarrow \prod_i p_i = \prod_i \frac{\lambda_i^{n_i} e^{-\lambda_i}}{n_i!} \tag{4.12}$$

where the product runs over all bins i for the binned distributions. The log-normal probability density function is assumed for the nuisance parameters affecting the signal yields:

$$\pi(\vec{\theta}) = e^{-\frac{1}{2}\theta^2} \tag{4.13}$$

The value of a likelihood itself does not have any meaning (it can be larger than 1, so it clearly can not be a probability). However, the relative values of likelihoods are useful and one can use as test statistics the so called likelihood ratio. This ratio is computed between the Likelihood maximum value and its variation over different values of the parameter of interests. In order to avoid very small or very large values of the likelihoods one takes their negative logs and thus, the maximum likelihood value corresponds to the minimum negative log-likelihood. Mathematically the used test statistics is defined as :

$$q_\mu = -2ln\frac{\mathcal{L}(n|\mu, \hat{\theta}_\mu)}{\mathcal{L}(n|\hat{\mu}, \hat{\theta})} \tag{4.14}$$

where $\hat{\theta}_\mu$ in the numerator represents the value of the nuisance parameter θ that maximizes the likelihood function for a fixed μ with with $0 \leq \mu \leq \hat{\mu}$ and $\hat{\mu}$ is the parameter of interest point at which the likelihood reaches its global maximum.

This statistic test is performed to determine the existence of the VBF HH signal in the observed data. The test is performed under two hypotheses: the presence of the signal H_{s+b} or the absence of the signal H_b where s, b denote the expected signal and the total background. The test statistic is used to set an upper limit on the signal production cross section (or, equivalently, on the signal strength μ_s which represents the ratio of the measured cross section and the value predicted by the theory model $\mu_s = \sigma/\sigma_{th}$) using the modified frequentist C.L.s criterion. To set a limit on the signal existence, we need to find the value of μ for which the H_{s+b} hypothesis can be excluded in favor of the H_b.

The numerator likelihood function $\mathcal{L}(n|\mu, \hat{\theta}_\mu)$ in Equation 4.14 is the product of Poisson probabilities for number of events (observed or simulated), given the expected signal and background (in this case $\mu = 1$ for the VBF signal expected). The interesting point of the likelihood ratios comes from Wilk's theorem , which states that a 68% C.L. interval can be found for the parameter of interest by looking at the region for which -2 $ln\mathcal{L}(\mu)$< 1.0. The minimum value of the likelihood ratio gives the best estimation of a given parameter of interest.

The exclusion limits are calculated at 95% C.L., by using Equation 4.14 to compute the observed value of the test statistic q_μ^{obs} (from data n), and to calculate the probability for q_μ to be equal or larger than q_μ^{obs} under the hypothesis of signal and background $H_{\mu s+b}$ or background-only hypothesis H_b, such as :

$$CL_s = \frac{P(q_\mu \geq q_\mu^{obs}|\mu \cdot s + b)}{P(q_\mu \geq q_\mu^{obs}|b)} \tag{4.15}$$

A signal of strength is said to be excluded at a confidence level (CL) of α if $CL_s(\mu) < 1 - \alpha$. In this analysis we are using $\alpha = 0.95$ so the 95 % CL upper limit on μ is defined as the signal strength value that gives $CL_s = 0.05$.

In this analysis, the expected best fit for Higgs VBF signal strength have been measured performing a Higgs coupling scanning as explained in the next section.

κ_λ, C_{2V} and C_V Higgs couplings scan

Under the assumption that no VBF HH signal exists, $qq'HH \rightarrow b\bar{b}4lqq'$ C.L.s upper limits on the VBF HH production signal strength are derived as a function of $\kappa_\lambda = \lambda_{HHH}/\lambda_{SM}$, where λ_{SM} denotes the SM prediction and λ_{HHH} denotes the measured value. The same scan is also done for the C_{2V} and C_V parameters. In particular, κ_λ, C_{2V} and C_V scan is performed for the full 2018 MC data-set in

the range [0, 2], [0,2] and [0.5,1.5] respectively while fixing the other parameters value to their SM one-value. The results are reported in the Brazil-band[14] plots in Figure 4.18.

From these results, the analysis of the signal $qq'HH \rightarrow b\bar{b}4lqq'$ done performing a shape analysis using the 2018 data-set is not able to constraint the Higgs couplings to a certain range of values. Indeed, the predicted number of SM VBF HH signal events is roughly less than the uncertainty in the background and in the signal and it is enough small to be compatible with the expected number of background events within the errors ($\sigma_{95\%CL}/\sigma_{SM} > 1$).

An approximated extrapolation of the upper limits to the larger data-set of the full Run II, corresponding to an integrated luminosity of 137 fb^{-1} is reported in Figure 4.19 with a result which is near to exclude some Higgs couplings values. The projection is done assuming that the signal strength is improved by scaling his value according to the square root of the integrated luminosity.

---

14 Brazil-band plots are used to present exclusion limits for the existence of new signals.
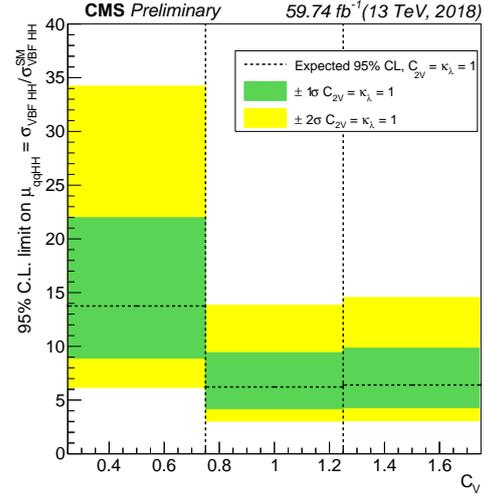
(a)
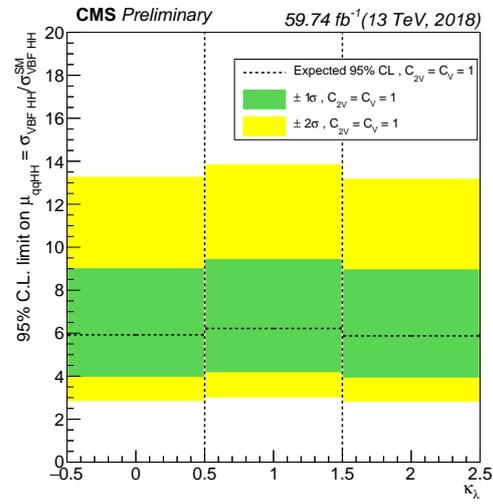
(b)
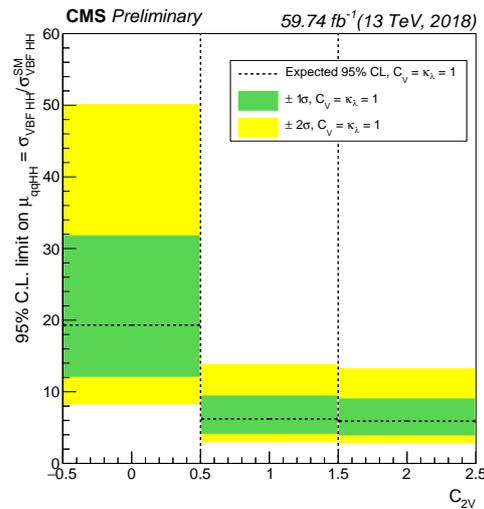
(c)

Figure 4.18: $qq'HH \rightarrow b\bar{b}4lqq'$ expected C.L. upper limits on the VBF HH signal strength derived as a function of Higgs couplings $\kappa_\lambda, C_{2V}, C_V$ for the full 2018 data-set. The results for each final state (4μ, $4e$ and $2e2\mu$) are already combined.
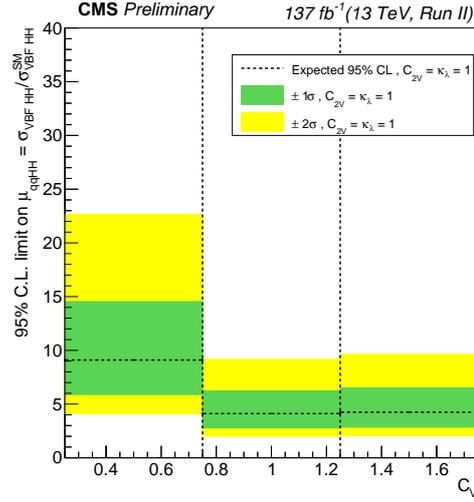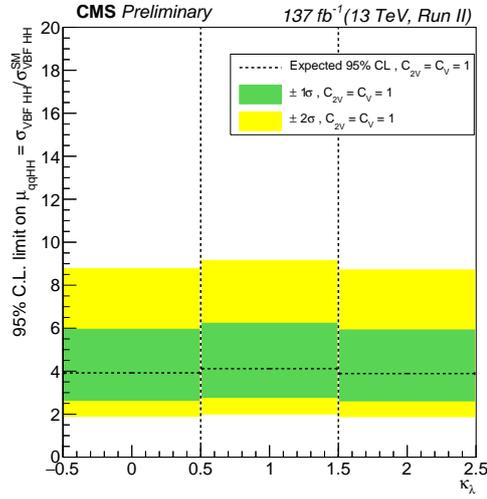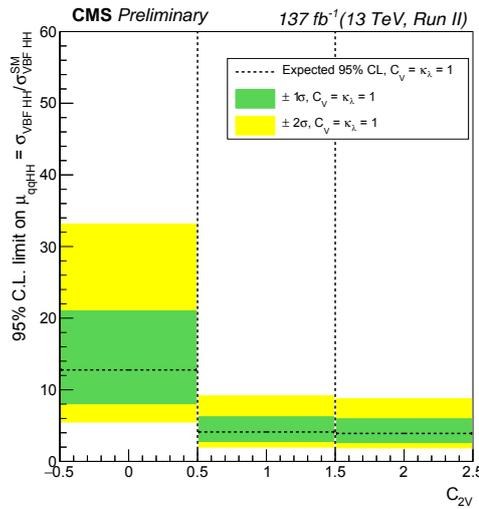
Figure 4.19: $qq'HH \to b\bar{b}4lqq'$ expected C.L. upper limits on the VBF HH signal strength derived as a function of Higgs couplings $\kappa_\lambda, C_{2V}, C_V$ for the full Run II data-set. The results for each final state (4μ, $4e$ and $2e2\mu$) are already combined.

Conclusions

W ITH the mass of the Higgs boson precisely known and its properties in good agreement with the theoretical predictions, VBF HH production is becoming increasingly important as a probe of the scalar sector of the SM. The presence of new resonances, anomalous couplings, and eventually the Higgs boson self-coupling HHH, the quartic HHVV can be simultaneously studied in VBF HH searches. The restart of the LHC, marked by a centre-of-mass energy increase from 8 to 13 TeV and a higher instantaneous luminosity represent an unprecedented possibility to explore HH production.

With this work, the search for VBF HH production in the $b\bar{b}l^+l^-l^+l^-$ decay channel has been developed and optimized for Run II conditions. Benefiting of the rare collision signature having two pairs of leptons, but loosing sensitivity due to the contamination from SM backgrounds, this decay channel is one of the most interesting HH production channels. The search covered the three main decay modes containing two pairs of electrons, of muons or one pair of both. A multivariate method was introduced for the first time for this di-Higgs decay channel and it can be surely be improved by making an a preliminary discrimination of the simulated signal from the $t\bar{t}H$ background as done in the VBF HH $\to b\bar{b}\gamma\gamma$ analysis.

Neverthless, the expected results were found in agreement with the combined results of affiliated analysis, and the expected 95% CL upper limits that were set establish a first step done on the 2018 data-set that can be completed applying the same study to the full Run II data-sets with the definition of proper control regions before performing an un-blind analysis. These results should be regarded in the broader context of CMS HH searches, with four main separate decay channels currently probed at 13 TeV, a few others HH searches in preparation, and an upcoming combination of the results.

As this thesis comes to its end after seven months, the exploration of the TeV scale has just started. The data collected during the LHC Run II and III, and the subsequent operations of the HL-LHC, will elucidate many open questions on the SM and on its possible BSM extensions. We do not know what these data will reveal us, and whether the SM will continue to resist to experimental tests or if clear signs of BSM physics will appear. For sure, HH searches will be at the forefront of the rich physics programme of the CMS experiment. Run III data will give an unprecedented insight on the presence of resonant BSM physics and

on anomalous coupling structures of the Higgs boson. If no deviations from the SM are found, the large data-sets collected during the HL-LHC operations will ultimately open the way to the measurement of σ_{HH} and the determination of Higgs self-coupling constants.

Extrapolations of the current 2018 results to the full Run II and HL-LHC luminosities are being performed addressing the latter scenario within the researches of the Double Higgs group at the time of writing this thesis. Further improvements of the sensitivity are possible and are expected to be achieved with the larger data-set available. Most importantly, these extrapolations show that no "golden channel" exists for a stand-alone HH production observation, and that several decay channels will need to be explored and combined.

From the near future of Run III operations to the end of HL-LHC operations, HH searches will continue to be one of the main paths to the exploration of the SM scalar sector and of the physics at the TeV scale.

Bibliography

[1] Jonathan Richard Ellis, Mary Katherin Gaillard, and Dimitri V Nanopou-
     los. «A phenomenological profile of the Higgs boson.» In: *Nucl. Phys. B*
     106 (1976), 292–340. 64 p. DOI: 10.1016/0550-3213(76)90382-5. URL:
     https://cds.cern.ch/record/87404.

[2] Isaac Newton and Herbert Westren Turnbull. *The Correspondence of Isaac
     Newton: 1661–1675.* London, UK: Cambridge University Press for the
     Royal Society, 1959, Volume 1, 416 p.

[3] J. J. Thomson. «XL. Cathode Rays.» In: *Philosophical Magazine* 5 (2009),
     44 –269, 293 –316 p. DOI: http://dx.doi.org/10.1080/14786449708621070.

[4] University of Zurich. *CMS Wiki Pages.* https://wiki.physik.uzh.ch/
     cms/latex:tikz. Accessed: 2021-09-30.

[5] The UA1 Collaboration. «Further evidence for charged intermediate vector
     bosons at the SPS collider.» In: *Phys. Lett. B* 129 (1985), 273–282. 17
     p. DOI: 10.1016/0370-2693(83)90860-2. URL: http://cds.cern.ch/
     record/163856.

[6] The UA2 Collaboration. «Evidence for $Z^0 \rightarrow$ e+e- at the CERN $\bar{p}p$
     collider.» In: *Phys. Lett. B* 129 (1983), 130–140. 21 p. DOI: 10.1016/0370-
     2693(83)90744-X. URL: http://cds.cern.ch/record/146503.

[7] The UA1 Collaboration. «Experimental observation of isolated large trans-
     verse energy electrons with associated missing energy at $\sqrt{s} = 540 GeV$.»
     In: *Phys. Lett. B* 122 (1983), 103–116. 31 p. DOI: 10.5170/CERN-1983-
     004.123. URL: https://cds.cern.ch/record/142059.

[8] The UA2 Collaboration. «Observation of single isolated electrons of high
     transverse momentum in events with missing transverse energy at the
     CERN $\bar{p}p$ collider.» In: *Phys. Lett. B* 122 (1983), 476–485. 15 p. DOI:
     10.1016/0370-2693(83)91605-2. URL: https://cds.cern.ch/record/
     142759.

[9] UA1 Collaboration. «Experimental observation of lepton pairs of invariant
     mass around $95 GeV/c^2$ at the CERN SPS collider.» In: *Phys. Lett. B*
     126 (1985), 398–410. 17 p. DOI: 10.1016/0370-2693(83)90188-0. URL:
     https://cds.cern.ch/record/163857.

[10] The CDF Collaboration. «Evidence for top quark production in $\overline{p}p$ collisions at $\sqrt{s}$ =1.8 TeV.» In: *Phys. Rev. D* 50 (5 1994), pp. 2966–3026. DOI: 10.1103/PhysRevD.50.2966. URL: https://link.aps.org/doi/10.1103/PhysRevD.50.2966.

[11] The CDF Collaboration. «Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab.» In: *Phys. Rev. Lett.* 74 (14 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[12] The DØ Collaboration. «Observation of the Top Quark.» In: *Phys. Rev. Lett.* 74 (14 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[13] Otto Nachtmann. *Elementary particle physics : concepts and phenomena / Otto Nachtmann ; translated by A. Lahee and W. Wetzel.* English. Springer-Verlag Berlin ; New York, 1990, xix, 559 p. : ISBN: 0387516476.

[14] Peter W. Higgs. «Broken symmetries, massless particles and gauge fields.» In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[15] Peter W. Higgs. «Broken Symmetries and the Masses of Gauge Bosons.» In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[16] Peter W. Higgs. «Spontaneous Symmetry Breakdown without Massless Bosons.» In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[17] F. Englert and R. Brout. «Broken Symmetry and the Mass of Gauge Vector Mesons.» In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. «Global Conservation Laws and Massless Particles.» In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[19] A. Salam and J.C. Ward. «Electromagnetic and weak interactions.» In: *Physics Letters* 13.2 (1964), pp. 168–171. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)90711-5. URL: https://www.sciencedirect.com/science/article/pii/0031916364907115.

[20] G. 't Hooft and M. Veltman. «Regularization and renormalization of gauge fields.» In: *Nuclear Physics B* 44.1 (1972), pp. 189–213. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(72)90279-9. URL: https://www.sciencedirect.com/science/article/pii/0550321372902799.

[21] «Search for the Standard Model Higgs Boson at LEP.» In: *CERN-ALEPH-2002-024, CERN-ALEPH-CONF-2002-013, DELPHI-2002-088-CONF-621, CERN-DELPHI-2002-088-CONF-621, L3-Note-2766, OPAL-TN721, LHWG-Note-2002-01* (2002). The Results quoted in this paper are not final, 28 p. URL: https://cds.cern.ch/record/2310899.

[22] Kyle J. Knoepfel. *Standard Model Higgs Boson Searches at the Tevatron.* 2013. arXiv: 1305.1530 [hep-ex].

[23] Luca Cadamuro. «Search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel with the CMS detector at the LHC. Recherche de la production de paires de bosons de Higgs dans le canal de désintégration $b\bar{b}\tau^+\tau^-$ avec le détecteur CMS auprès du LHC.» Presented 05 Oct 2017. 2017. URL: https://cds.cern.ch/record/2292733.

[24] The CMS Collaboration. «Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.» In: *Phys. Lett. B* 716 (2012), 30–61. 32 p. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235. URL: https://cds.cern.ch/record/1471016.

[25] The ATLAS Collaboration. «Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.» In: *Phys. Lett. B* 716 (2012). Comments: 24 pages plus author list (38 pages total), 12 figures, 7 tables, revised author list, 1–29. 29 p. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214. URL: https://cds.cern.ch/record/1471031.

[26] Biagio Di Micco,Maxime Gouzevitch,Javier Mazzitelli and Caterina Vernieri. *Higgs boson potential at colliders: status and perspectives.* Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2690841.

[27] F. Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course In Modern Particle Physics.* 1984. ISBN: 978-0-471-88741-6.

[28] J. Goldstone. «Field Theories with Superconductor Solutions.» In: *Nuovo Cim.* 19 (1961), pp. 154–164. DOI: 10.1007/BF02812722.

[29] «Measurements of properties of the Higgs boson in the four lepton final state at 13 TeV.» In: *CMS-PAS-HIG-18-001* (2018).

[30] «Measurements of the Higgs boson inclusive differential and production cross sections in the 4ℓ decay channel at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *ATLAS-CONF-2019-025* (July 2019).

[31] «Measurements of Higgs boson production via gluon fusion and vector boson fusion in the diphoton decay channel at 13 TeV.» In: *CMS-PAS-HIG-18-029* (2019).

[32] «Measurements and interpretations of Higgs boson fiducial cross sections in the diphoton decay channel using 139 inverse femtobarn of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *ATLAS-CONF-2019-029* (July 2019).

[33] M. Tanabashi et al. «Review of Particle Physics.» In: *Phys. Rev. D* 98.3 (2018), p. 030001. DOI: `10.1103/PhysRevD.98.030001`.

[34] Albert M Sirunyan et al. «Observation of $t\bar{t}$H production.» In: *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: `10.1103/PhysRevLett.120.231801`. arXiv: `1804.02610 [hep-ex]`.

[35] Serguei Chatrchyan et al. «Combined results of searches for the standard model Higgs boson in *pp* collisions at $\sqrt{s} = 7$ TeV.» In: *CMS-HIG-11-032, CERN-PH-EP-2012-023* 710 (2012), pp. 26–48. DOI: `10.1016/j.physletb.2012.02.064`. arXiv: `1202.1488 [hep-ex]`.

[36] Serguei Chatrchyan et al. «Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs.» In: *CMS-HIG-12-041, CERN-PH-EP-2012-372* 110.8 (2013), p. 081803. DOI: `10.1103/PhysRevLett.110.081803`. arXiv: `1212.6639 [hep-ex]`.

[37] Vardan Khachatryan et al. «Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV.» In: *CMS-HIG-14-009, CERN-PH-EP-2014-288* 75.5 (2015), p. 212. DOI: `10.1140/epjc/s10052-015-3351-7`. arXiv: `1412.8662 [hep-ex]`.

[38] Georges Aad et al. «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.» In: *CERN-EP-2016-100, ATLAS-HIGG-2015-07, CMS-HIG-15-002* 08 (2016), p. 045. DOI: `10.1007/JHEP08(2016)045`. arXiv: `1606.02266 [hep-ex]`.

[39] «Observation of the SM scalar boson decaying to a pair of τ leptons with the CMS experiment at the LHC.» In: *CMS-PAS-HIG-16-043* (2017).

[40] CERN Press Office. *CMS sees evidence for the Higgs boson decaying into muons.* `https://cms.cern/news/cms-sees-evidence-higgs-boson-decaying-muons`. Accessed: 2021-09-30.

[41] D. de Florian et al. «Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector.» In: *CERN-2017-002-M, CERN-2017-002* 2/2017 (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[42] Giuseppe Degrassi, Marco Fedele, and Pier Paolo Giardino. «Constraints on the trilinear Higgs self coupling from precision observables.» In: *JHEP* 04 (2017), p. 155. DOI: 10.1007/JHEP04(2017)155. arXiv: 1702.01737 [hep-ph].

[43] Giuseppe Degrassi, Pier Paolo Giardino, Fabio Maltoni, and Davide Pagani. «Probing the Higgs self coupling via single Higgs production at the LHC.» In: *CP3-16-38, RM3-TH-16-8* 12 (2016), p. 080. DOI: 10.1007/JHEP12(2016)080. arXiv: 1607.04251 [hep-ph].

[44] Ulrich Haisch. «Yukawas and trilinear Higgs terms from loops.» In: *52nd Rencontres de Moriond on EW Interactions and Unified Theories*. 2017, pp. 9–16. arXiv: 1706.09730 [hep-ph].

[45] «Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment.» In: *ATLAS-CONF-2019-049* (Oct. 2019).

[46] E. W. Nigel Glover and J. J. van der Bij. «HIGGS BOSON PAIR PRODUCTION VIA GLUON FUSION.» In: *CERN-TH-4934/87* 309 (1988), pp. 282–294. DOI: 10.1016/0550-3213(88)90083-1.

[47] J. Baglio, A. Djouadi, R. Gröber, M. M. Mühlleitner, J. Quevillon, and M. Spira. «The measurement of the Higgs self-coupling at the LHC: theoretical status.» In: *KA-TP-44-2012, SFB-CPP-12-102, LPT-ORSAY-12-124, PSI-PR-12-10* 04 (2013), p. 151. DOI: 10.1007/JHEP04(2013)151. arXiv: 1212.5581 [hep-ph].

[48] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli, E. Vryonidou, and M. Zaro. «Higgs pair production at the LHC with NLO and parton-shower effects.» In: *CERN-PH-TH-2014-014, CP3-14-07, ZU-TH03-14, LPN14-055* 732 (2014), pp. 142–149. DOI: 10.1016/j.physletb.2014.03.026. arXiv: 1401.7340 [hep-ph].

[49] Oscar J. P. Eboli, G. C. Marques, S. F. Novaes, and A. A. Natale. «TWIN HIGGS BOSON PRODUCTION.» In: *MIT-CTP-1376A* 197 (1987), pp. 269–272. DOI: 10.1016/0370-2693(87)90381-9.

[50] Daniel de Florian, Massimiliano Grazzini, Catalin Hanga, Stefan Kallweit, Jonas M. Lindert, Philipp Maierhöfer, Javier Mazzitelli, and Dirk Rathlev. «Differential Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD.» In: *DESY-16-107, FR-PHENO-2016-007, ICAS-*

08-16, MITP-16-061, ZU-TH-20-16 09 (2016), p. 151. DOI: `10.1007/JHEP09(2016)151`. arXiv: `1606.09519 [hep-ph]`.

[51] Jonathan Grigo, Kirill Melnikov, and Matthias Steinhauser. «Virtual corrections to Higgs boson pair production in the large top quark mass limit.» In: *TTP14-024, LPN14-101* 888 (2014), pp. 17–29. DOI: `10.1016/j.nuclphysb.2014.09.003`. arXiv: `1408.2422 [hep-ph]`.

[52] Fabio Maltoni, Giovanni Ridolfi, and Maria Ubiali. «b-initiated processes at the LHC: a reappraisal.» In: *CP3-12-15, TTK-12-11* 07 (2012). [Erratum: JHEP 04, 095 (2013)], p. 022. DOI: `10.1007/JHEP04(2013)095`. arXiv: `1203.6393 [hep-ph]`.

[53] Joan Elias-Miro, Jose R. Espinosa, Gian F. Giudice, Gino Isidori, Antonio Riotto, and Alessandro Strumia. «Higgs mass implications on the stability of the electroweak vacuum.» In: *Phys. Lett. B* 709 (2012), pp. 222–228. DOI: `10.1016/j.physletb.2012.02.013`. arXiv: `1112.3022 [hep-ph]`.

[54] Fedor L. Bezrukov and Mikhail Shaposhnikov. «The Standard Model Higgs boson as the inflaton.» In: *Phys. Lett. B* 659 (2008), pp. 703–706. DOI: `10.1016/j.physletb.2007.11.072`. arXiv: `0710.3755 [hep-th]`.

[55] Fedor Bezrukov. «The Higgs field as an inflaton.» In: *RBRC1042* 30 (2013), p. 214001. DOI: `10.1088/0264-9381/30/21/214001`. arXiv: `1307.0708 [hep-ph]`.

[56] Sally Dawson and Christopher W. Murphy. «Standard Model EFT and Extended Scalar Sectors.» In: *Phys. Rev. D* 96.1 (2017), p. 015041. DOI: `10.1103/PhysRevD.96.015041`. arXiv: `1704.07851 [hep-ph]`.

[57] Hermès Bélusca-Maïto, Adam Falkowski, Duarte Fontes, Jorge C. Romão, and João P. Silva. «Higgs EFT for 2HDM and beyond.» In: *Eur. Phys. J. C* 77.3 (2017), p. 176. DOI: `10.1140/epjc/s10052-017-4745-5`. arXiv: `1611.01112 [hep-ph]`.

[58] R. Grober and M. Muhlleitner. «Composite Higgs Boson Pair Production at the LHC.» In: *JHEP* 06 (2011), p. 020. DOI: `10.1007/JHEP06(2011)020`. arXiv: `1012.1562 [hep-ph]`.

[59] Alexandra Carvalho, Martino Dall'Osso, Pablo De Castro Manzano, Tommaso Dorigo, Florian Goertz, Maxime Gouzevich, and Mia Tosi. «Analytical parametrization and shape classification of anomalous HH production in the EFT approach.» In: (July 2016). arXiv: `1608.06578 [hep-ph]`.

[60] Steven Weinberg. «Phenomenological Lagrangians.» In: *Physica A* 96.1-2 (1979). Ed. by S. Deser, pp. 327–340. DOI: `10.1016/0378-4371(79)90223-1`.

[61]   Aielet Efrati and Yosef Nir. «What if $\lambda_{hhh} \neq 3m_h^2/v$.» In: (Jan. 2014). arXiv: 1401.0935 [hep-ph].

[62]   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.» In: *CERN-PH-TH-2014-064, CP3-14-18, LPN14-066, MCNET-14-09, ZU-TH-14-14* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[63]   CMS collaboration. «Search for non resonant H(bb)H($\gamma\gamma$) for Run 2.» In: *CMS Analysis Notes 286* (CERN, 2017).

[64]   Georges Aad et al. «Searches for Higgs boson pair production in the $hh \to bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector.» In: *Phys. Rev. D* 92 (2015), p. 092004. DOI: 10.1103/PhysRevD.92.092004. arXiv: 1509.04670 [hep-ex].

[65]   Albert M Sirunyan et al. «Search for Higgs boson pair production in the $bb\tau\tau$ final state in proton-proton collisions at $\sqrt{(s)} = 8$ TeV.» In: *CMS-HIG-15-013, CERN-EP-2017-104* 96.7 (2017), p. 072004. DOI: 10.1103/PhysRevD.96.072004. arXiv: 1707.00350 [hep-ex].

[66]   «Search for resonant Higgs boson pair production in four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *CMS-PAS-B2G-20-004* (2021). URL: https://cds.cern.ch/record/2777083.

[67]   «Search for heavy resonances decaying to a pair of boosted Higgs bosons in final states with leptons and a bottom quark-antiquark pair at $\sqrt{s} = 13$ TeV.» In: *CMS-PAS-B2G-20-007* (2021). URL: https://cds.cern.ch/record/2777173.

[68]   Armen Tumasyan et al. «Search for a heavy Higgs boson decaying into two lighter Higgs bosons in the $\tau\tau$bb final state at 13 TeV.» In: *CMS-HIG-20-014, CERN-EP-2021-094* (June 2021). arXiv: 2106.10361 [hep-ex].

[69]   *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.* Geneva, 2021. URL: https://cds.cern.ch/record/2759683.

[70]   «Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector.» In: *ATLAS-CONF-2021-030* (2021). URL: https://cds.cern.ch/record/2777236.

[71] «Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: (2021). All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-035. URL: http://cds.cern.ch/record/2777861.

[72] Georges Aad et al. «Search for the $HH \to b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *JHEP* 07 (2020). [Erratum: JHEP 01, 145 (2021), Erratum: JHEP 05, 207 (2021)], p. 108. DOI: 10.1007/JHEP07(2020)108. arXiv: 2001.05178 [hep-ex].

[73] «Search for nonresonant Higgs boson pair production in the 4 leptons plus 2 b jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *CMS-PAS-HIG-20-004* (2020). URL: https://cds.cern.ch/record/2725691.

[74] «Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in pp collisions at $\sqrt{s} = 13\,\text{TeV}$.» In: *CMS-PAS-HIG-19-018* (2020).

[75] «Search for Higgs boson pair production in the four b quark final state.» In: *CMS-PAS-HIG-20-005* (2021). URL: https://cds.cern.ch/record/2771912.

[76] «Search for Higgs boson pair production via vector boson fusion with highly Lorentz-boosted Higgs bosons in the four b quark final state at $\sqrt{s} = 13$ TeV.» In: *CMS-PAS-B2G-21-001* (2021). URL: https://cds.cern.ch/record/2776802.

[77] Georges Aad et al. «Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV.» In: *CERN-EP-2019-143* 801 (2020), p. 135145. DOI: 10.1016/j.physletb.2019.135145. arXiv: 1908.06765 [hep-ex].

[78] «Summary of non-resonant and resonant Higgs boson pair searches from the ATLAS experiment.» In: *ATL-PHYS-PUB-2021-031* (2021).

[79] Luis W. Alvarez. *The hydrogen bubble chamber and the strange resonances.* Pions to Quarks: Particle Physics in the 1950s. Cambridge: Cambridge University Press, 1989, pp. 299–306.

[80] Yves Baconnier, Giorgio Brianti, P Lebrun, A G Mathewson, R Perin, and Yves Baconnier. *LHC: the Large Hadron Collider accelerator project.* Geneva: CERN, 1993. URL: https://cds.cern.ch/record/257706.

[81] Thomas Sven Pettersson and P Lefèvre. *The Large Hadron Collider: conceptual design.* Tech. rep. 1995. URL: https://cds.cern.ch/record/291782.

[82] CERN Press Office. *First beam in the LHC Accelerating science.* `https://home.cern/news/press-release/cern/first-beam-lhc-accelerating-science`. Accessed: 2021-09-30.

[83] Paul Rincon. *Collider halted until next year.* `http://news.bbc.co.uk/2/hi/science/nature/7632408.stm`. Accessed: 2021-09-30.

[84] CERN Press Office. *Two circulating beams bring first collisions in the LHC.* `https://home.cern/news/press-release/cern/two-circulating-beams-bring-first-collisions-lhc`. Accessed: 2021-09-30.

[85] CERN Press Office. *CERN announces LHC to run in 2012.* `https://home.cern/news/press-release/cern/cern-announces-lhc-run-2012`. Accessed: 2021-09-30.

[86] CERN Press Office. *The HL - LHC project.* `https://hilumilhc.web.cern.ch/content/hl-lhc-project`. Accessed: 2021-09-30.

[87] Stephen Myers. «The LEP collider, from design to approval and commissioning.» In: *CERN-91-08, CERN-YELLOW-91-08* (Oct. 1991). DOI: `10.5170/CERN-1991-008`.

[88] «LHC Machine.» In: *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: `10.1088/1748-0221/3/08/S08001`.

[89] Esma Mobs. «The CERN accelerator complex. Complexe des accélérateurs du CERN.» In: (2016). General Photo. URL: `https://cds.cern.ch/record/2197559`.

[90] «LHC Design Report. 3. The LHC injector chain.» In: *CERN-2004-003-V-3, CERN-2004-003* (Dec. 2004). Ed. by M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. DOI: `10.5170/CERN-2004-003-V-3`.

[91] The ATLAS Collaboration. «The ATLAS Experiment at the CERN Large Hadron Collider.» In: *Journal of Instrumentation* 3.08 (2008), S08003–S08003. DOI: `10.1088/1748-0221/3/08/s08003`. URL: `https://doi.org/10.1088/1748-0221/3/08/s08003`.

[92] S. Chatrchyan et al. «The CMS Experiment at the CERN LHC.» In: *JINST* 3 (2008), S08004. DOI: `10.1088/1748-0221/3/08/S08004`.

[93] The LHCb Collaboration. «The LHCb Detector at the LHC.» In: *Journal of Instrumentation* 3.08 (2008), S08005–S08005. DOI: `10.1088/1748-0221/3/08/s08005`. URL: `https://doi.org/10.1088/1748-0221/3/08/s08005`.

[94] The ALICE Collaboration. «The ALICE experiment at the CERN LHC.» In: *Journal of Instrumentation* 3.08 (2008), S08002–S08002. DOI: `10.1088/1748-0221/3/08/s08002`. URL: `https://doi.org/10.1088/1748-0221/3/08/s08002`.

[95] Siona Ruth Davis. «Interactive Slice of the CMS detector.» In: (2016). URL: https://cds.cern.ch/record/2205172.

[96] CERN Press Office. *Physics at LHC - Taking a closer look at LHC*. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/1.physics_at_lhc. Accessed: 2021-09-30.

[97] J. Gareyte. «The Large Hadron collider in the LEP tunnel.» In: *CERN-SPS-89-29-AMS, CERN-LHC-NOTE-98* (July 1989).

[98] *CMS Luminosity Public Results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2021-09-30.

[99] James Stirling. *Parton luminosity and cross section plots*. http://www.hep.ph.ic.ac.uk/{~}wstirlin/plots/plots.html. Accessed: 2021-09-30.

[100] *Public CMS Data Quality Information*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality. Accessed: 2021-09-30.

[101] Tai Sakuma. «Cutaway diagrams of CMS detector.» In: (2019). URL: http://cds.cern.ch/record/2665537.

[102] Martina Ressegotti. «Overview of the CMS Detector Performance at LHC Run 2.» In: *CMS-CR-2018-383* 5.1 (2019), p. 18. DOI: 10.3390/universe5010018.

[103] *The CMS magnet project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: http://cds.cern.ch/record/331056.

[104] University of Zurich. *CMS Wiki Pages*. https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates. Accessed: 2021-09-30.

[105] Serguei Chatrchyan et al. «Description and performance of track and primary-vertex reconstruction with the CMS tracker.» In: *CMS-TRK-11-001, CERN-PH-EP-2014-070* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[106] V Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: http://cds.cern.ch/record/368412.

[107] *The CMS tracker: addendum to the Technical Design Report*. Technical design report. CMS. Geneva: CERN, 2000. URL: http://cds.cern.ch/record/490194.

[108] Vardan Khachatryan et al. «CMS Tracking Performance Results from Early LHC Operation.» In: *CERN-PH-EP-2010-019, CMS-TRK-10-001* 70 (2010), pp. 1165–1192. DOI: 10.1140/epjc/s10052-010-1491-3. arXiv: 1007.1988 [physics.ins-det].

[109] János Karancsi. «Operational Experience with the CMS Pixel Detector.» In: *JINST* 10.05 (2015), p. C05016. DOI: 10.1088/1748-0221/10/05/C05016. arXiv: 1411.4185 [physics.ins-det].

[110] «CMS Technical Design Report for the Pixel Detector Upgrade.» In: *CERN-LHCC-2012-016, CMS-TDR-011* (Sept. 2012). Ed. by David Aaron Matzner Dominguez et al. DOI: 10.2172/1151650.

[111] A. Tricomi. «Upgrade of the CMS tracker.» In: *JINST* 9 (2014). Ed. by Pietro Govoni, Pier Simone Marrocchesi, Francesco-Luigi Navarria, Marco Paganoni, Andrea Perrotta, and Tiziano Rovelli, p. C03041. DOI: 10.1088/1748-0221/9/03/C03041.

[112] *The CMS electromagnetic calorimeter project: Technical Design Report.* Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375.

[113] P. Adzic et al. «Energy resolution of the barrel of the CMS electromagnetic calorimeter.» In: *CERN-CMS-NOTE-2006-148* 2 (2007), P04004. DOI: 10.1088/1748-0221/2/04/P04004.

[114] *The CMS hadron calorimeter project: Technical Design Report.* Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/357153.

[115] Efe Yazgan. «The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c.» In: *CMS-CR-2008-040* 160 (2009). Ed. by M. Fraternali, Gabriella Gaudio, and Michele Livan, p. 012056. DOI: 10.1088/1742-6596/160/1/012056.

[116] «Detector Drawings.» CMS Collection. 2012. URL: https://cds.cern.ch/record/1433717.

[117] S. Abdullin et al. «Design, performance, and calibration of the CMS Hadron-outer calorimeter.» In: *CERN-CMS-NOTE-2008-020* 57 (2008), pp. 653–663. DOI: 10.1140/epjc/s10052-008-0756-6.

[118] J. G. Layter. *The CMS muon project: Technical Design Report.* Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[119] A. M. Sirunyan et al. «Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *CMS-MUO-16-001, CERN-EP-2018-058* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[120] Serguei Chatrchyan et al. «Performance of CMS Muon Reconstruction in *pp* Collision Events at $\sqrt{s} = 7$ TeV.» In: *CMS-MUO-10-004, CERN-PH-EP-2012-173* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].

[121] M. Abbrescia et al. «Local and global performance of double-gap resistive plate chambers operated in avalanche mode.» In: *Nucl. Instrum. Meth. A* 434 (1999), pp. 244–253. DOI: 10.1016/S0168-9002(99)00534-3.

[122] «First measurement of the total proton-proton cross section at the LHC energy of $\sqrt{s}$=7 TeV.» In: *EPL* 96 (2011). PACS 13.60.Hb: Total and inclusive cross sections, 21002. 11 p. DOI: 10.1209/0295-5075/96/21002. arXiv: 1110.1395. URL: https://cds.cern.ch/record/1383030.

[123] CMS Collaboration. *CMS Wiki Pages.* https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined. Accessed: 2021-09-30.

[124] Vardan Khachatryan et al. «The CMS trigger system.» In: *CMS-TRG-12-001, CERN-EP-2016-160* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[125] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project.* Technical design report. CMS. Geneva: CERN, 2002. URL: https://cds.cern.ch/record/578006.

[126] «CMS Technical Design Report for the Level-1 Trigger Upgrade.» In: *CERN-LHCC-2013-011, CMS-TDR-12, CMS-TDR-012* (June 2013). Ed. by A. Tapper and Darin Acosta.

[127] W. Adam et al. «The CMS high level trigger.» In: *Eur. Phys. J. C* 46 (2006), pp. 605–667. DOI: 10.1140/epjc/s2006-02495-8. arXiv: hep-ex/0512077.

[128] CMS Collaboration. *CMS computing: Technical Design Report.* Technical design report. CMS. Submitted on 31 May 2005. Geneva: CERN, 2005. URL: http://cds.cern.ch/record/838359.

[129] *Worldwide LHC Computing Grid Tiers infrastructure.* https://wlcg-public.web.cern.ch/tier-centres. Accessed: 2021-09-30.

[130] Michael et al. Aderholz. *Models of Networked Analysis at Regional Centres for LHC Experiments (MONARC), Phase 2 Report, 24th March 2000.* Tech. rep. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/510694.

[131] Reham Aly. «Search for Dark Matter Produced in Association With a Higgs Boson in the Four Leptons Final State Using Run II Data at $\sqrt{s} = 13$ TeV With the CMS Experiment.» 2021. URL: https://cds.cern.ch/record/2777608.

[132]   Leonardo Giannini. «Deep Learning techniques for the observation of the Higgs boson decay to bottom quarks with the CMS experiment.» Presented 28 Jul 2020. 2020. URL: https://cds.cern.ch/record/2730094.

[133]   Ruth Pöttgen. «Proton-Proton Collisions.» In: *Search for Dark Matter with ATLAS: Using Events with a Highly Energetic Jet and Missing Transverse Momentum in Proton-Proton Collisions at √s = 8 TeV* . Cham: Springer International Publishing, 2016, pp. 45–60. ISBN: 978-3-319-41045-6. DOI: 10.1007/978-3-319-41045-6_4. URL: https://doi.org/10.1007/978-3-319-41045-6_4.

[134]   B. R. Webber. «A QCD Model for Jet Fragmentation Including Soft Gluon Interference.» In: *CERN-TH-3713* 238 (1984), pp. 492–528. DOI: 10.1016/0550-3213(84)90333-X.

[135]   Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. «Parton Fragmentation and String Dynamics.» In: *LU-TP-83-10* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[136]   Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. «An introduction to PYTHIA 8.2.» In: *LU-TP-14-36, MCNET-14-22, CERN-PH-TH-2014-190, FERMILAB-PUB-14-316-CD, DESY-14-178, SLAC-PUB-16122* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[137]   Carlo Oleari. «The POWHEG-BOX.» In: *Nucl. Phys. B Proc. Suppl.* 205-206 (2010). Ed. by Johannes Blümlein, Sven-Olaf Moch, and Tord Riemann, pp. 36–41. DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph].

[138]   Peter Z. Skands. «QCD for Collider Physics.» In: *2010 European School of High Energy Physics.* Apr. 2011. arXiv: 1104.2863 [hep-ph].

[139]   John C. Collins, Davison E. Soper, and George F. Sterman. «Factorization of Hard Processes in QCD.» In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.

[140]   S. Agostinelli et al. «GEANT4–a simulation toolkit.» In: *SLAC-PUB-9350, FERMILAB-PUB-03-339, CERN-IT-2002-003* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[141]   A. M. Sirunyan et al. «Particle-flow reconstruction and global event description with the CMS detector.» In: *CMS-PRF-14-001, CERN-EP-2017-110* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[142] Pierre Billoir. «Progressive track recognition with a Kalman like fitting procedure.» In: *Comput. Phys. Commun.* 57 (1989), pp. 390–394. DOI: `10.1016/0010-4655(89)90249-X`.

[143] Paolo Azzurri. «Track Reconstruction Performance in CMS.» In: *CMS-CR-2008-110* 197 (2009). Ed. by Pietro Govoni, Pier Simone Marrocchesi, Francesco-Luigi Navarria, Marco Paganoni, and Andrea Perrotta, pp. 275–278. DOI: `10.1016/j.nuclphysbps.2009.10.084`. arXiv: `0812.5036 [physics.ins-det]`.

[144] Vardan Khachatryan et al. «Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} =$ 8 TeV.» In: *CMS-EGM-13-001, CERN-PH-EP-2015-004* 10.06 (2015), P06005. DOI: `10.1088/1748-0221/10/06/P06005`. arXiv: `1502.02701 [physics.ins-det]`.

[145] «Performance of muon identification in pp collisions at s**0.5 = 7 TeV.» In: *CMS-PAS-MUO-10-002, CMS-PAS-MUO-10-002* (2010).

[146] M. Mulders. «Muon reconstruction and identification at CMS.» In: *CMS-CR-2006-101* 172 (2007). Ed. by P. S. Marrocchesi, F. L. Navarria, M. Paganoni, and P. G. Pelfer, pp. 205–207. DOI: `10.1016/j.nuclphysbps.2007.08.049`.

[147] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. «Electron reconstruction in CMS.» In: *CERN-CMS-NOTE-2006-040* 49 (2007), pp. 1099–1116. DOI: `10.1140/epjc/s10052-006-0175-5`.

[148] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov. «Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC.» In: *CHEP-2003-TULT009* C0303241 (2003), TULT009. DOI: `10.1088/0954-3899/31/9/N01`. arXiv: `physics/0306087`.

[149] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. «The anti-k_t jet clustering algorithm.» In: *LPTHE-07-03* 04 (2008), p. 063. DOI: `10.1088/1126-6708/2008/04/063`. arXiv: `0802.1189 [hep-ph]`.

[150] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. «FastJet User Manual.» In: *CERN-PH-TH-2011-297* 72 (2012), p. 1896. DOI: `10.1140/epjc/s10052-012-1896-2`. arXiv: `1111.6097 [hep-ph]`.

[151] Vardan Khachatryan et al. «Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV.» In: *CMS-JME-13-004, CERN-PH-EP-2015-305* 12.02 (2017), P02014. DOI: `10.1088/1748-0221/12/02/P02014`. arXiv: `1607.03663 [hep-ex]`.

[152] Serguei Chatrchyan et al. «Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS.» In: *CERN-PH-EP-2011-102, CMS-JME-10-011* 6 (2011), P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].

[153] *Jet algorithms performance in 13 TeV data.* Tech. rep. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2256875.

[154] A. M. Sirunyan et al. «Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV.» In: *CMS-BTV-16-002, CERN-EP-2017-326* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[155] G. Cowan. *Statistical data analysis.* 1998. ISBN: 978-0-19-850156-5.

[156] Serguei Chatrchyan et al. «Measurement of the Properties of a Higgs Boson in the Four-Lepton Final State.» In: *Phys. Rev. D* 89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[157] Sylvie Braibant, Giorgio Giacomelli, and Maurizio Spurio. *Particles and fundamental interactions: Supplements, problems and solutions: A deeper insight into particle physics.* Undergraduate Lecture Notes in Physics. Dordrecht, Netherlands: Springer, 2012. ISBN: 978-94-007-4134-8. DOI: 10.1007/978-94-007-4135-5.

[158] CERN Twiki. *SM Higgs Branching Ratios and Partial-Decay Widths (CERN Report 3, 2013 update).* https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR2014. Accessed: 2021-09-30.

[159] Laurent Thomas. «CMS High Level Trigger performance at 13 TeV.» In: *PoS* ICHEP2018 (2019), p. 226. DOI: 10.22323/1.340.0226.

[160] Marko Kovac. «Measurements of properties of the Higgs boson in the four-lepton final state at $\sqrt{s} = 13$ TeV.» In: *PoS* LHCP2018 (2018), p. 067. DOI: 10.22323/1.321.0067.

[161] CERN Yellow Report. *SM Higgs Branching Ratios and Total Decay Widths.* https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR. Accessed: 2021-09-30.

[162] CERN Twiki. *SWGuideLHEInterface.* https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideLHEInterface. Accessed: 2021-09-30.

[163] CMS collaboration. «Measurement of Higgs Production Cross Section via Vector Boson Fusion in H → ZZ → 4l final state at 13 TeV using Artificial Neural Networks.» In: *CMS Analysis Notes 238* (CERN, 2017).

[164] Shilpi Jain. «Lepton and Photon reconstruction and identification performance in ATLAS and CMS in Run II.» In: *PoS* LHCP2020 (2021). Ed. by Bruno Mansoulie, Giovanni Marchiori, Roberto Salern, and Tulika Bos, p. 046. DOI: `10.22323/1.382.0046`.

[165] CMS collaboration. «Double Higgs boson search in the 4-lepton plus 2 b jets final states at $\sqrt{s} = 13$ TeV with full Run II data.» In: *CMS Analysis Notes 117* (CERN, 2019).

[166] Keras. *About Keras.* `https://keras.io/about/`. Accessed: 2021-09-30.

[167] A. Bagoly, A. Bevan, A. Carnes, S. V. Gleyzer, L. Moneta, A. Moudgil, S. Pfreundschuh, T. Stevenson, S. Wunsch, and O. Zapata. «Machine learning developments in ROOT.» In: *J. Phys. Conf. Ser.* 898.7 (2017). Ed. by Richard Mount and Craig Tull, p. 072046. DOI: `10.1088/1742-6596/898/7/072046`.

[168] Pushpalatha C. Bhat. «Multivariate Analysis Methods in Particle Physics.» In: *Ann. Rev. Nucl. Part. Sci.* 61 (2011), pp. 281–309. DOI: `10.1146/annurev.nucl.012809.104427`.

[169] Carsten Peterson and Thorsteinn Rognvaldsson. «An Introduction to artificial neural networks.» In: *1991 CERN School of Computing.* Sept. 1991.

[170] Xavier Glorot and Y. Bengio. «Understanding the difficulty of training deep feedforward neural networks.» In: *Journal of Machine Learning Research - Proceedings Track* 9 (Jan. 2010), pp. 249–256.

[171] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. «Dropout: A Simple Way to Prevent Neural Networks from Overfitting.» In: *J. Machine Learning Res.* 15 (2014), pp. 1929–1958.

[172] Andre Sznajder. *LHC Searches and Higgs Results.* `https://www.ictp-saifr.org/wp-content/uploads/2014/05/AndreSznajder_ICTP-SAIFR_SchoolQCDLHCPhysics2015.pdf`. Accessed: 2021-09-30.

[173] CMS Twiki. *Jet energy scale uncertainty sources.* `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources`. Accessed: 2021-09-30.

[174] LHCHXSWG. *HH cross-sections.* `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/`. Accessed: 2021-09-30.

[175] CERN TWiki. *CMS Higgs Combined Limit.* `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit`. Accessed: 2021-09-30.

[176]   CERN Indico. *Higgs Combine Tool Workshop at the LPC.* https://
indico.cern.ch/event/747340/timetable/. Accessed: 2021-09-30.