Search for Dark Matter Produced in Association With a Higgs Boson in the Four Leptons Final State Using Run II Data at s = 13 TeV With the

UNIVERSITÀ DEGLI STUDI DI BARI

Dipartimento Interateneo di Fisica "M. Merlin"

DOTTORATO DI RICERCA IN FISICA

Ciclo XXXIII - Settore scientifico disciplinare Fis/01

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

Dottoranda:

Dott.ssa **Reham Mohamed Mahmoud Mohamed Aly**

Supervisori:

Prof. **N. De Filippis**

Dott.ssa **F. De Fazio**

Coordinatore:

Ch.mo Prof. **Giuseppe Iaselli**

ESAME FINALE 2021

Abstract

This thesis presents the search for dark matter candidates produced in association with a Higgs boson. The search is performed using proton proton collision data collected by CMS experiment during Run II and corresponding to a total integrated luminosity of 137 fb^{-1} at center of mass energy of 13 TeV. The Higgs boson decay channel considered in the search is the Higgs decay to a pair of Z bosons which decay to four leptons with three possible final states (4μ, 4e and $2e2\mu$). The experimental signature is characterized by significant amount of missing transverse energy p_T^{miss} accompanied with four leptons from the Higgs decay products. This search is referred as "Mono-Higgs" search, where the Higgs boson is used as a candle to tag the presence of the dark matter candidate. The results obtained are interpreted in the context of the two Higgs doublet simplified model extended by light pseudoscalar. This model has been inspected with the Higgs $\rightarrow$ ZZ $\rightarrow$ 4l decay channel in CMS, for the fist time.

Acknowledgments

As the thesis come out to the conclusion, I would like to thank all the people who helped me through my journey in the PhD, and Please forgive me if I missed some names. I would like to express my gratitude to my lovable parents for being the foundation of my life, you made my life beautiful. Many thanks to my two elder sisters for the unceasing encouragement, support and attention during my work.

I would like to thank the referees for accepting to review the thesis. I would like to thank Bari University & INFN to provide me the chance to run my PhD research in Italy. I want to express my grateful to my supervisors Prof. Nicola De Filippis & Dott.ssa Fulvia De Fazio for their suggestion, invaluable discussions, sharing expertise and support they had given me throughout this work. I would like to express deepest gratitude to prof. Giuseppe Iaselli for his help and support, he was like my father in Italy. Many thanks to the people involved in the CMS RPC group especially Prof. Gabriella Pugliese for guiding me through my work at CERN. Special thanks to MonoHiggs group for their valuable discussions and sharing expertise. Many thanks to Guest and visitor program to give me the chance to work in the laboratories at Fermi National Laboratory (Fermilab) and cooperate with expertise. Thanks a lot prof. Pushpalatha C Bhat, you helped me to learn the machine learning techniques, we had a lot of fruitful discussions through our meetings, thanks for your help and a lot of good times we had together at Fermilab.

Many thanks to all the people I met in Bari for their kindness and welcoming me to a lovely city.

Last, certainly not least, Dear father and mother, my mind fails to find words that can show how deeply I am indebted to you for your unconditional love and care. Dear father and mother, you have always been supportive and kind to me, my warm regards.

Reham ...

Contents

List of Figures

List of Tables

Abbreviations

ALICE	Alarge Ion Collider Experiment
AOD	Analysis Object Data
ATLAS	A Toroidal LHC ApparatuS
BDT	Boosted Decision Trees
BPIX	Barrel Pixel detector
BR	Branching Ratio
BSM	Beyond Standard Model
CERN	Conseil Europen pour la Recherche Nuclaire
CL	Confidence Level
CMB	Cosmic Microwave background
CMS	Compact Muon Solenoid
CR	Control Regions
CSC	Cathode Strip Chamber
CTF	Combinatorial Track Finder
DAQ	Data Acquision System
DM	Dark Matter
DT	Drift Tube
EB	ECAL Barrel
ECAL	Electromagnetic Calorimeter
EE	ECAL Endap
EPIX	Endcap Pixel detector
FR	Fake Rate

FSR	Final State Radiation
GSF	The Gaussian Sum Filter
GT	Global Trigger
HB	HCAL Barrel detector
HCAL	Hadronic Calorimeter
HE	HCAL End-cap detector
HEP	High Energy Physics
HF	HCAL Forward detector
HL-LHC	High Luminosity Large Hadron Collider
HLT	High Level Trigger
HM	High Mass region
HO	HCAL Outer calorimeter
IP	Impact Parameter
IRPC	Improved Resistive Plate Chamber
ISR	Initial State Radiation
JEC	Jet Energy Correction
LEP	Large Electron-Positron collider
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty experiment
LHCf	Large Hadron Collider forward
LM	Low Mass region
LS	Long Shutdown
MC	Monte Carlo
MET	Missing Energy Transverse
MVA	Multivariate Approach
NASA	National Aeronautics and Space Administration
NLO	Next to Leading Order
PD	Primary Dataset

PF	Particle Flow
PS	Parton Shower
$\mathbf{p_T^{miss}}$	Missing Transverse Energy
PU	Pile-Up
PV	Primary Vertex
ROC	Receiver Operating Characteristic curve
RPC	Resistive Plate Chamber
SC	Super Cluster
SD	Spin-Dependent
SI	Spin Independent
SIP	Significance of the Impact Parameter
SM	Standard Model
SPS	Super Proton Synchrotron
SUSY	Supersymmetry
SV	Secondary Vertex
TEC	Tracker Endcaps
TIB	Tracker Inner Barrel
TID	Tracker Inner Disks
TOB	Tracker Outer Barrel
TP	Trigger Primitive
UV	Ultra Violet
VBF	Vector Boson Fusion

Introduction

It is our nature "human" to wonder about the universe around us. It is the curiosity of mankind to understand the universe and its constituents. The most important question is "What is the universe made of ?". An experimental branch known as High Energy Physics (HEP) is concerned with answering this question. Over many years, large numbers of particle detectors were built to detect new particles from cosmic rays. Besides that, a large number of particle accelerators and colliders were built to exploit the high energy interactions to allow the production of new particles which generally don't exist in nature under ordinary conditions. A huge number of elementary particles started to show up. All these particles and the interactions between them are described within the theory called **S**tandard **M**odel (SM) of particle physics [28, 30–32].

Particle Physics evolution

Over years, the scientific community searched for the fundamental constituents of the universe "particles that can't be further subdivided", a large number of ideas were exploited until they came up with the idea of an "atom", the smallest possible piece of matter. The idea of the atom was widely accepted till 1897 when J.J. Thomson, while working with cathode rays, discovered that they were made up of negatively light charged particles called electrons [33]. He formulated a model called "plum pudding model" in which the negatively electrons are floating in a sea of positive charge inside the atom so the overall charge is neutral. In 1911, Rutherford's experiment about the scattering of alpha particles from a thin gold foil concluded that the atom consists of hard, positively-charged nucleus orbited by a cloud of negative electrons [34, 35]. The subsequent discovery of neutrons by J. Chadwick in 1932 completed the picture of the atom [36]. Protons, neutrons and electrons are considered as a fundamental particles until Gell-Mann and Zweig postulated the quark model in 1964. This model says that the protons and neutrons, as well as many other particles subsequently discovered (hadrons), consist of more fundamental particles called "quarks". According to the quark model, baryons are made of three quarks while mesons are made of a quark and an anti-quark pair. In 1967, the first evidence of the existence of quarks has been gathered in the Deep Inelastic Scattering experiments, initiated at SLAC (USA). The experiments

showed that the proton consists of three quarks. Afterwards, other elementary particles were identified and their interactions are described within the SM theory. These interactions are classified in four forces. The strong force is felt by the quarks and is mediated by eight massless vector particles called gluons. The electroweak sector of the theory unifies weak and electromagnetic forces. The weak force, such as the one responsible of beta decay, is mediated by vector massive bosons, $W^{\pm}$ and Z^0, while the familiar electromagnetic force is mediated by the (vector, massless) photon. The $W^{\pm}$ and Z^0 bosons have been discovered in 1983 by UA1 and UA2 Collaboration at CERN [37–41]. The discovery of those particles has represented a great success of SM theory which predicted their existence before the discovery. In 1964, the Belgian physicists "Robert Brout" and "Francois Englert', and the Scotsman "Peter Higgs' proposed a mechanism in which particles of the SM gain mass by interaction with so-called the Higgs field [42, 43]. The Higgs boson, quantum of the Higgs field, is one of the building block of the SM. The discovery of Higgs boson "the last missing piece of the SM" on July 2012 by A Toroidal LHC ApparatuS (ATLAS) and Compact Muon Solenoid (CMS) collaborations is considered one of the greatest successes of the SM theory [44, 45].

All those discovered particles represent the visible matter of the universe while on the other hand on the macroscopic scale a series of observations suggest the existence of Dark Matter (DM) particles. Cosmological observations indicates that total mass - energy of the universe contains around 4.8 % ordinary matter, 25.8 % dark matter and 69.4 % dark energy, which means that the great majority of our universe is unseen [46, 47]. Those observations don't provide information about the DM nature, but they provide constraints on its properties. Different experiments have been used for dark matter detection through the interaction of DM and SM particles. No DM signal has been observed till now but experiments put only constraints on the DM properties. Although the SM theory is considered a successful theory, it doesn't provide any possible DM candidate. Many theories beyond the SM propose dark matter candidates such as axions, sterile neutrinos and weakly interacting massive particles (WIMPs). The most widely accepted candidates beyond the SM theories are WIMPs because their relic density naturally has at least the right order of magnitude predicted by the observations. If the DM interacts with SM particles, DM can be produced in proton proton collision at colliders such as Large Hadron Collider and hence their interactions can be studied in details. If the DM is produced in the detector, it will escape the detector without leaving a signature. Therefore, the DM search at colliders uses a SM particle to tag the DM existence through large missing momentum balanced by visible particles produced in association with the DM. This search is called "Mono-X" where X represent the SM particle produced in association with the DM particle. The Higgs boson discovery in 2012 opened a new portal for the dark matter search at LHC; if the dark matter particles have mass and weakly interact they can couple to the Higgs boson.

In this thesis, a search for DM particles produced in association with Higgs boson at LHC by proton-proton collisions at a center of mass energy of 13 TeV is presented. The data used in the analysis have been collected by CMS experiment during Run II with total integrated

luminosity of 137 fb^{-1}. In this analysis, we consider the Higgs boson decay to a pair of Z bosons, which decay to four leptons. This search has been performed for the first time in CMS by using 2016 data were published in [48, 49]. The search has been extended for full Run II and the results have interpreted using a new signal model used for the first time in this decay channel.

Thesis outline

The results presented in this thesis represent the work done during my journey as a Ph.D. student. Along with the physics analysis, I got the chance to participate in hardware activity related to the Resistive Plate Chambers (RPC) installed inside the CMS experiment. The physics analysis part related to the search for DM particles produced in association with Higgs boson using data collected by CMS experiment during Run II at center of mass energy 13 TeV at the LHC represents the main part of the thesis. I got the chance to perform the full Run II analysis starting from working on a framework for the analysis up to the final results obtained in the thesis, which also introduces a multivariate analysis technique to increase the sensitivity of the analysis. The hardware work was related to the aging study on RPC detectors which are currently used in CMS. This work was performed in Gamma Irradiation Facility (GIF++) at CERN where I participated in the installation and test of prototypes and real size RPC detectors for the CMS upgrade projects for the High Luminosity LHC (HL-LHC) scenario. The results obtained were published in [50] and more data have been collected later and added to the published results. The thesis is organized in the subsequent chapters as follows:

- Chapter 1 briefly describes the SM theory and its shortcomings, reviews the main evidence for the existence of the Dark Matter and the possible DM candidates. In addition the DM detection techniques, the current limits and the model used to interpret the results are presented.

- Chapter 2 describes the experimental framework, the LHC accelerator and its experiments. A detailed description of CMS experiment and its detectors is presented as the data used in the thesis were collected using the CMS detector.

- Chapter 3 describes the hardware work related to the aging study being performed on RPC detectors which are currently used in CMS. Since the physics analysis relies on the four muon final state, it was a great chance to work directly on muon detector and study their performance.

- Chapter 4 describes how the informations collected from CMS sub-detectors translated to physical objects. The algorithms used to reconstruct the physics objects used in the

thesis are presented. Besides that, an overview of how physics events are generated and simulated in the detector is provided.

- Chapter 5 describes all the building blocks required to perform the analysis. It describes the data used, the signal samples and the main background processes that affect the analysis. Moreover, I give a detailed description of how the physics objects used in the thesis are selected.
- Chapter 6 describes how the analysis strategy is designed to distinguish the Mono-Higgs signal from the background processes. In addition the background estimation and the analysis flow with results are presented.
- Chapter 7 introduces the multivariate approach used to increase the sensitivity of the analysis and the sources of systematic uncertainties affecting the measurements.
- Chapter 8 present the final yield and distributions and the interpretation of the results in terms of limit settings.
- The thesis ends with conclusion section and a short summary.

1 Search for Dark Matter

This chapter consists of a brief introduction to the Standard Model (SM) of particle physics [28, 30–32], the discovery of the last building block of the SM, "the Higgs Boson", and some open questions not answered by the SM. The nature of the Dark Matter is considered one of the mysteries in our universe that can't be explained within the SM. An overview of those cosmological observations besides different experiments that aim to detect possible DM particles is discussed. The latest results from those experiments and from the previous search of DM particles at the LHC is presented. The last part of the chapter introduces the extension of the SM that is considered for the interpretation of the results reported in this thesis.

1.1 Standard Model of particle physics

The SM of particle physics is the theory describing the elementary constituents of matter and their interactions at the fundamental level. It was developed throughout the twentieth century, and is accepted as the current description of particle physics. The SM is a gauge theory based on the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry group that describe the interaction between the fundamental particles through the strong ($SU(3)_C$) and electroweak forces ($SU(2)_L \otimes U(1)_Y$). The quantum numbers of the elementary particles under these groups are called color charge, weak isospin and the hypercharge respectively. The particle content of the theory consists of matter constituents and gauge bosons, plus the Higgs particle. They are all shown in Figure 1.1

- Matter Constituents.

  Matter constituents are spin 1/2 fermions. Fermions are divided in two categories leptons (colorless) and quarks (color charged). They are grouped in three generations shown in the first three columns of Figure 1.1. In the first two rows of Figure 1.1 the three quark generations are displayed. Each contains two quarks. The quarks up, charm and top are called up-type quarks and have electric charge 2/3 while the quarks down, strange and beauty are called down-type quarks and have charge -1/3. In the case of leptons, each generation consists of a neutrino and of a charged lepton to which it corresponds. Each generation is heavier than the previous one while the lifetime in

general decreases; taus are heavier than muons and electrons. The same relation exists between the quarks also; the top quark is the heaviest SM particle, and the b quark is an exception since it has a longer lifetime than the charm. The heavier generation particles decay via weak interaction to the lower generation particles. So our universe consists only of the first generation particles of the standard model. The SM predicts that the hundreds of particles discovered during the twentieth century are made of only twelve fundamental particles, six quarks (are the constituents of hadrons) and six leptons (the electron, the muon, the tau and their respective neutrinos) summarized in Table 1.1[1].

- Interaction mediators. For the fundamental interactions, the mediators are spin 1 particles called gauge bosons summarized in Table 1.2. The photon (γ) has a null electromagnetic charge and is massless particle that mediates the electromagnetic interaction. The Z and $W^{\pm}$ are the mediators of the weak force and differently from the photon, they are not massless. Those particles have been discovered by the UA1 and UA2 experiments at CERN [37–41]. Fermions (except neutrinos) and the gauge bosons $W^{\pm}$ and Z^0 in the SM acquire their mass through the Brout - Englert-Higgs mechanism [42,43]. In this mechanism, an associated massive scalar boson is predicted, the so-called Brout - Englert Higgs boson, recently discovered by ATLAS and CMS experiments [44, 45]. Gluons are the mediators of the strong force, they interact with quarks and within themselves. They are massless particles.

1.2 The Higgs Boson

The discovery of the Higgs boson was the main target of the first LHC run. On 4^{th} of July 2012 ATLAS and CMS experiments announced the discovery of a new particle with mass near 125 GeV with properties compatible with the SM Higgs boson [44,45]. The discovery of the Higgs boson "the last building block of the SM" represents one of the greatest successes of the SM theory. The Higgs boson production in proton-proton collisions can happen through five main mechanisms: the gluon fusion, the vector boson fusion, the associated production of a Higgs boson with a W or Z boson, and the associated production with a $t\bar{t}$ pair. Figure 1.2 (left) shows the production cross sections for the Higgs boson from different mechanisms. The branching ratios of the different Higgs decay channels as a function of the Higgs mass are shown in Figure 1.2 (right). In the low mass region (up to $\approx$ 150 GeV), the fermions decay modes dominate the branching ratio where the most important contribution comes from the $b\bar{b}$ channel. Then the Higgs decay to vector boson pairs opens up and quickly dominates. The Higgs decay to pair of photons and pair of Z bosons, where each Z decays to pair of

[1]The natural units will be used so that the speed of light c = 1

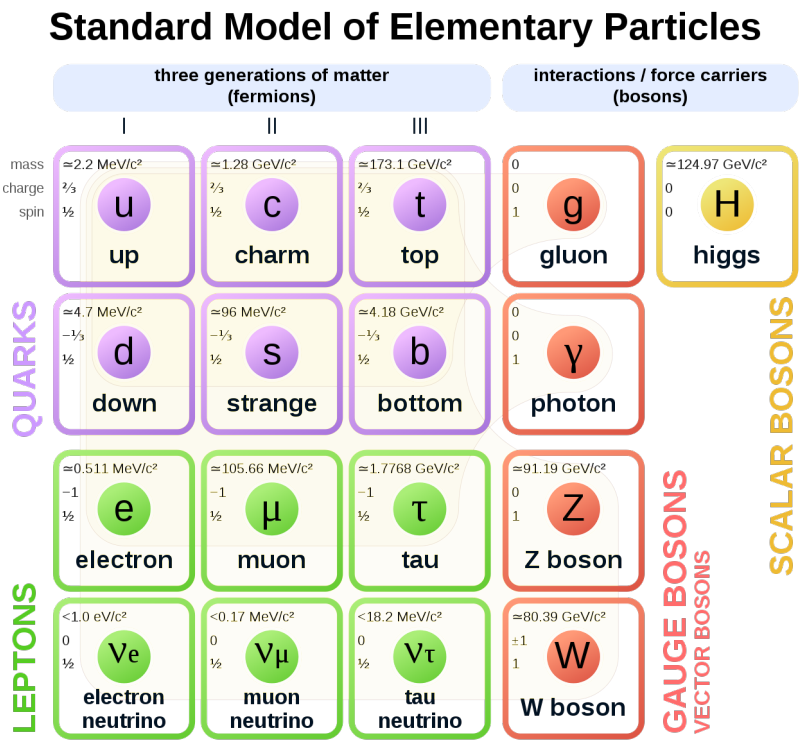


Figure 1.1: The fundamental particles of the Standard Model.

Table 1.1: List of fermions masses, charges and the forces they subject to [28].

Particle	Mass (GeV)	Charge (e)	Weak	EM	Strong
Quarks					
Up, u	2.2×10^{-3}	+2/3	Yes	Yes	Yes
Down, d	4.7×10^{-3}	-1/3	Yes	Yes	Yes
Charm, c	1.27	+2/3	Yes	Yes	Yes
Strange, s	96×10^{-3}	-1/3	Yes	Yes	Yes
Top, t	173.2	+2/3	Yes	Yes	Yes
Bottom, b	4.18	-1/3	Yes	Yes	Yes
Leptons					
Electron, e	0.511×10^{-3}	-1	Yes	Yes	No
Muon, μ	105.6×10^{-3}	-1	Yes	Yes	No
Tau, τ	1.776	-1	Yes	Yes	No
Electron neutrino, ν_e	-	0	Yes	No	No
Muon neutrino, ν_μ	-	0	Yes	No	No
Tau neutrino, ν_τ	-	0	Yes	No	No

Table 1.2: A list of the gauge bosons, their mass, charge and the forces that they mediate [28].

Particle	Mass (GeV)	Charge (e)	Weak	EM	Strong
Photon,γ	0	0	No	Yes	No
$W^{\pm}$	80.38	± 1	Yes	No	No
Z	91.18	0	Yes	No	No
Gluon, g	0	0	No	No	Yes

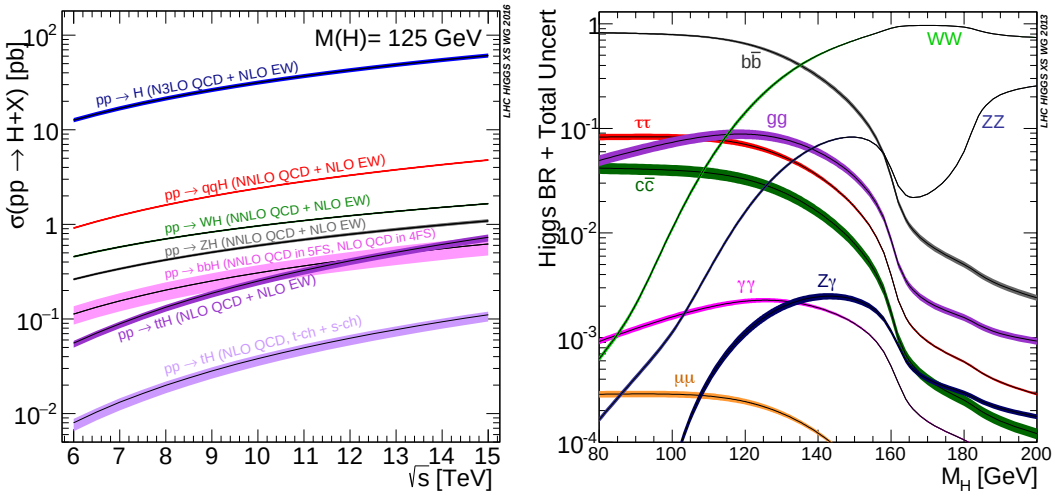


Figure 1.2: Shows the Higgs production cross section as a function of the center-of-mass energy (left) and the branching ratios of the Higgs decay as a function of Higgs mass (right) [1].

leptons, have been used for the Higgs boson discovery. Figure 1.3 shows the distribution of the di-photon invariant mass in the $H \to \gamma\gamma$ search (left) and the four leptons invariant mass in the $H \to ZZ \to 4l$ search (right) at CMS experiment. An excess of events is observed around 125 GeV with respect to the expectations in the background-only hypothesis. The $H \to ZZ \to 4l$ channel is considered as the "golden channel" for the Higgs discovery since it provides a clean signature in the detector. In the thesis, we consider the Higgs decay to four leptons in the search for DM particles.

1.3 Open questions of the SM

Although the SM has proven to be a successful theory, it is considered incomplete. There are still open questions that the SM does not explain, a short list is summarized as:

- **What is the Dark Matter?** Cosmological observations indicate that an important fraction of the mass of the Universe is unseen "Dark Matter" and what we see from galaxies and stars represent $\approx$ 4.8 % of the universe [46, 47]. The SM does not offer any candidate for dark matter. An extension of the SM theory is needed to explain the

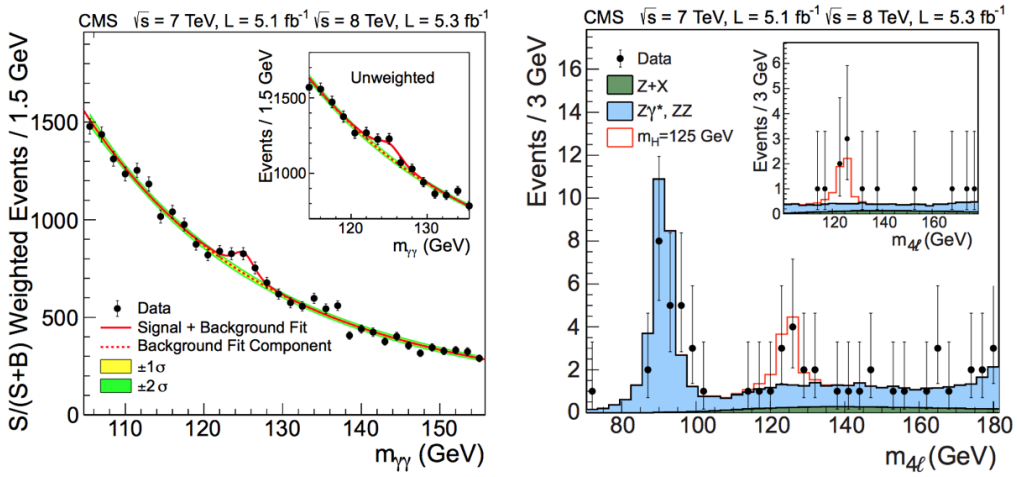


Figure 1.3: Invariant mass distributions in the $H \to \gamma\gamma$ search (left) and in the $H \to ZZ \to 4l$ search (right) at CMS experiment [2].

existence of dark matter. Detailed studies of the dark matter search are presented in this thesis.

- **Gravity** The gravity is the fourth fundamental force which is responsible for the macroscopic structure of the Universe. It is the weakest of the four fundamental interactions. The SM describes the strong and electroweak forces, but it doesn't provide any information about the gravity. A complete theory attempts to unify the four fundamental forces into one.

- **Neutrino masses** Neutrinos in the SM are massless particles. However, experiments showed that neutrinos created with specific flavor can be measured with a different flavor. This neutrino oscillation phenomenon between different flavors can't happen if neutrinos are massless [51–54].

- **Matter anti-Matter asymmetry** Why the universe is made of matter and no antimatter?. We need to know what happened in the early stages of the universe, either more matter was initially created or equivalent amount between matter and antimatter was created but somehow antimatter disappeared. There is no mechanism in the SM to explain the dominance of matter over antimatter.

- **Three generations of matter** The SM doesn't explain why there exist three generations of quarks and leptons, why there are many orders of magnitude in the fermion masses.

1.4 Evidence of Dark Matter

Cosmology indicates that the total mass - energy of the universe contains around 4.8 % ordinary matter, 25.8 % DM and 69.4 % dark energy, which means that the great majority of our universe is unseen [46,47]. Although DM has not been directly observed, its existence and properties are inferred from some cosmological observational evidence listed below. Those observations don't provide information about the DM nature, but they provide constraints on its properties.

1.4.1 Galaxy rotation curve

The first indication of DM existence was introduced by Fritz Zwicky [55] in 1933 when calculating the mass of the Coma Cluster using the virial theorem that relates the kinetic energy of the system to its potential energy. He indicates that the calculated galaxy mass is around two orders of magnitude larger than what is expected from the visible matter which implies the existence of non-luminous (dark) matter. The idea of dark matter was accepted later by scientists after the studies of galaxy rotational curves. The rotational curve describes the velocity of the visible stars or gas in the galaxy versus their radial distance from that galaxy's center. According to Newtonian dynamics, it is expected that the velocity v(r) of the visible matter decreases as $1/\sqrt{r}$, following equation 1.1, while moving from the center of the galaxy where most of the luminous matter is located as shown in Figure 1.4 left by the red curve:

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{1.1}$$

where G is the Newton's constant and M(r) is the mass of the visible matter of the galaxy at a radius r. But Ford and Rubin found that the rotational velocity remains constant while moving from the center of the galaxy as shown in Figure 1.4 left by white curve [56]. This observation leads to the idea of the presence of a non-luminous dark matter halo in the galaxy. Figure 1.4 right shows the rotational curve of the NGC 6503 galaxy which describes the observed rotation velocity for the NGC 6503 galaxy compared to the the expectations assuming a spherical halo of dark matter and no coupling between dark matter and visible matter. A good agreement between data and expectations is found which show that the galaxy center is mainly composed of visible matter, while the dark matter halo dominates at large radii [57].

1.4.2 Gravitational lensing

A simple definition of gravitational lensing is "mass bends light". The gravitational field of a massive object (such as galaxies and clusters of galaxies) will extend far into space, and when light emitted from the distanced object passes close to that massive object (and thus through

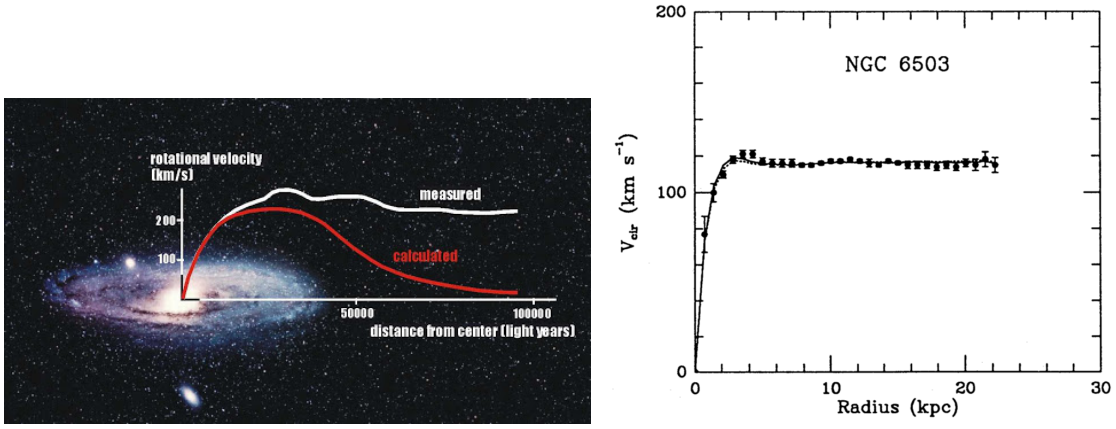


Figure 1.4: Rotation curve of NGC 6503 galaxy.

its gravitational field) the light will bend leading to a distortion of the object's image [58,59]. The bent light will refocus somewhere else producing a single or multiple images of the light source, arcs, or even Einstein rings. The bending of the light depends on the mass of the object, the more massive the objects means stronger gravitational field, more bending of the light and hence more distortion of the image. The intensity of the image distortion gives an estimate of the object's mass generating the lensing. For example, figure 1.5 shows the Abell 370 cluster, where the ratio between the total and the luminous mass has been measured to be 300. This measurement confirms the presence of dark matter in galaxies and in clusters of galaxies. Another evidence of dark matter existence is the Bullet Cluster (a cluster formed out of a collision of two smaller clusters). Figure 1.6 shows an image of the 1E 0657-56 galaxy cluster obtained with lensing and X-ray techniques. During the collision of two galaxies, the gas particles (the "normal" matter) will interact electromagnetically with each other and slowing down. On the contrary, the dark matter interacts only through gravity without undergoing electromagnetic interaction. So the majority of the light is coming from the hot X-ray emitting gas has shown with pink in the image, while in blue is the location of the dark matter in the cluster determined indirectly from the gravitational lensing. The observations show that most of the visible matter is now in the center of the image, but the lensing tells us that most of the mass lies further out which indicates the presence of dark matter.

1.4.3 Cosmic Microwave Background

The previous measurements provide an evidence of the DM existence without giving an information about the amount of DM in the universe. The Cosmic Microwave background (CMB) analysis can provide such information [60, 61]. CMB is the radiation left from the early universe produced in the recombination epoch after the big bang, it is known as "photon decoupling". The CMB has almost uniform temperature T = 2.726 K except some fluctuation of order 10^{-5} K in some places as shown in Figure 1.7 from the observations of the Plank collaboration. Studying this temperature anisotropies allow the estimation of some

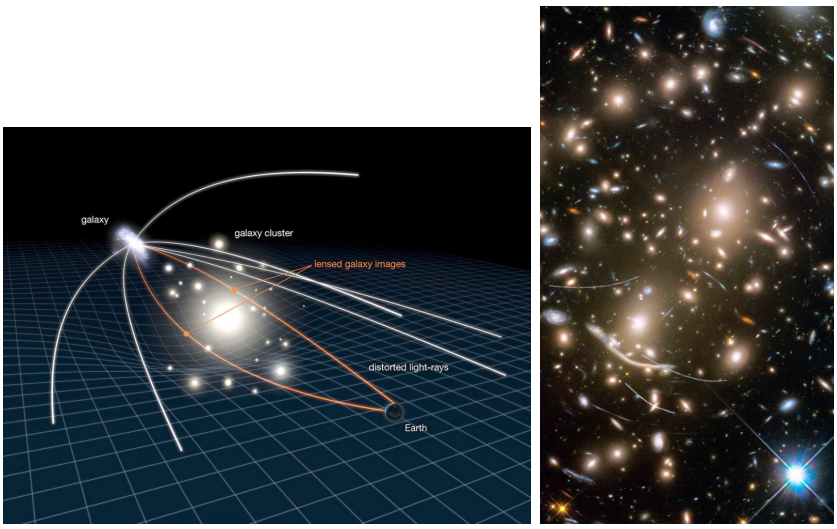


Figure 1.5: Left shows image of gravitational field created by clusters of galaxies and the distortion of light reached to the earth from distanced galaxy, right shows the image of the Abell 370 galaxy cluster taken by the NASA Hubble Space Telescope where the giant arcs seen are due to strong gravitational lensing [3].

Figure 1.6: Shows the galaxy cluster 1E 0657-56, also known as the "bullet cluster." This cluster was formed after the collision of two large clusters of galaxies, the most energetic event known in the universe since the Big Bang where the ordinary matter shown by pink and the dark matter shown by blue as proposed by gravitational lensing [4].

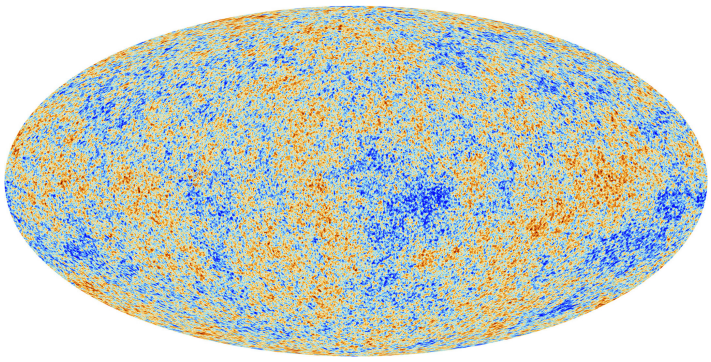


Figure 1.7: Shows the CMB anisotropies as observed by the Planck collaboration [5].

cosmological parameters such as the DM abundance. The results reflect that the universe has 25.8% DM and 4.8% baryonic matter in the universe and the remaining 69.4% represent dark energy [62].

1.5 Dark matter candidates

There are some constraints on the DM properties from the cosmological and astrophysical observations. The DM candidates should satisfy the following characteristics [62, 63]:

- Stable (long lived): the DM particles should be stable and have lifetime comparable to the age of the universe to explain its current existence otherwise they would have decayed by now.
- Right relic density: The DM candidate mass should account for the dark matter abundance observed.
- Interact very weakly with electromagnetic radiation: otherwise we could infer their existence from photons.
- non-relativistic: If the DM is relativistic particle, they would move too fast at the matter dominant epoch to gravitationally cluster and would result in a different distribution of structure inconsistent with what is observed.

There is no such candidate in the SM fulfilling those requirements. The only candidate in the SM could be the neutrino due to its stability, no electric charge and being only weakly and gravitationally interacting. On the other hand, the neutrino mass ($m_\nu < 2$ ev) can't explain the measured DM abundance and neutrinos are relativistic particles [64]. So there is no DM candidate in the SM. There are a lot of theories beyond the SM proposing dark matter candidates such as axions, sterile neutrinos, black holes and Weakly Interacting Massive Particles (WIMPs). The most widely accepted candidates beyond the SM theories are WIMPs because their relic density naturally has at least the right order of magnitude predicted by

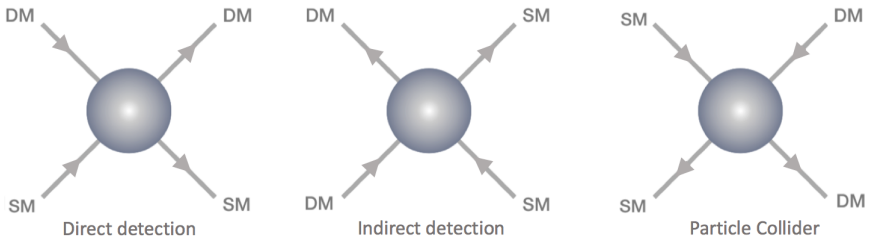


Figure 1.8: Shows a schematic view of the three different approaches for dark matter detection through the interaction of DM and SM particles.

the observations. The WIMP should have a mass in the range 100 GeV to 1 TeV. Moreover, DM is predicted to have a weak interaction with the SM particles which motivates to the following DM searches to observe it.

1.6 Dark matter search

Different approaches have been used for dark matter detection through the interaction of DM and SM particles as shown schematically in Figure 1.8.

1.6.1 Direct Detection

Those experiments search for nuclear recoil produced from the elastic scattering of DM particles from SM nuclei as shown in the schematic diagram 1.8 left. According to cosmological observations that galaxies contains DM particles, many of those particles can pass through the earth and interact with the material nuclei. From the galaxy rotational curves, the velocity of the DM particles is in the order of few hundred kilometers per second. At this velocity, the DM particles are elastically scattered from the SM particles with recoil energies in the range 1 - 100 KeV for DM particles with mass between 10 GeV and 10 TeV. This reflects the need for detectors with low nuclear recoil energy threshold. The expected interaction rate of DM with SM particles depends on many parameters such as the DM density, interaction cross section, the velocity and mass of the DM particles. The only two unknown parameters are the mass of the DM and the interaction cross section, therefore the direct detection results are usually expressed as a function of those parameters. In fact, the DM - SM interaction is a rare process (few events per year per thousands of kilograms of target material) therefore, the experiments are built underground to protect against cosmic ray induced backgrounds. There are a lot of experiments targeting the DM detection in direct way such as DAMA [65, 66], LUX [67, 68], Xenon100 [6], IceCube [69], PICASSO [70, 71] and Super-Kamiokande [72, 73]. Figure 1.9 shows the latest results for direct detection experiments which have constrained the DM-nucleon scattering cross section for different dark matter masses. Figure 1.9 left

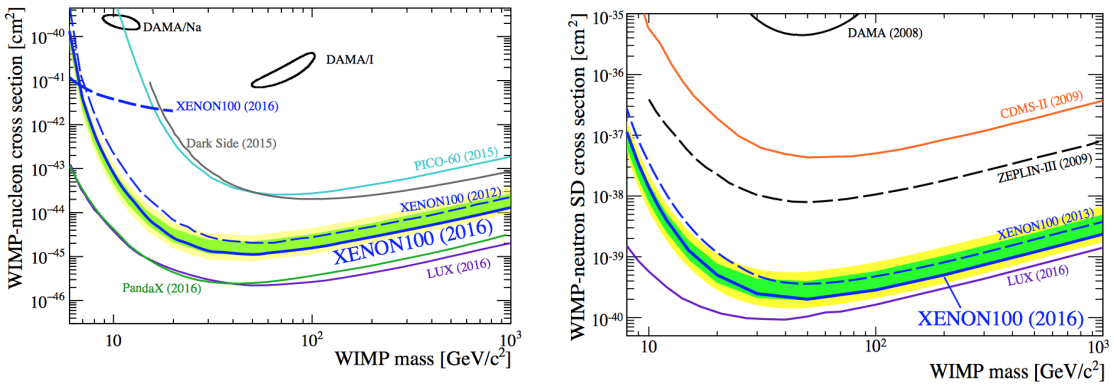


Figure 1.9: Shows cross section limits as a function of the DM mass spin-independent (left) and spin dependent (right) interactions for various experiments [6].

shows the spin independent (SI) limits, for WIMP with masses between 10 GeV and 100 GeV an exclusion reaching cross section of 10^{-46} cm^2. For the spin-dependent (SD) limit shown in Figure 1.9 right, the best sensitivity is given by LUX experiment. Direct detection experiments didn't provide any evidence for dark matter existence.

1.6.2 Indirect Detection

Such experiments study the production of SM particles from the annihilation of DM particles as shown in the schematic view of Figure 1.8 middle. The annihilation products include stable SM particles such as gamma rays, charged particles (positrons, and antiprotons) and neutrinos. The flux of those particles depends on the DM annihilation rate, which in turn is proportional to the DM density. For this reason, the best region for the search is the place where DM densities are high such as the center of the galaxy, the sun and the earth where the DM can be captured due to the energy loss caused by the scattering of WIMPs with the nucleons of those objects. Gamma rays produced from DM annihilation are expected to have mono-energetic lines in the photon spectrum. If such signal is detected and not compatible with astrophysical gamma rays source, it would represent a signal of DM. Gamma rays are detected through large telescopes such as HESS [74], MAGIC [75] and through the space telescope such as Fermi Gamma-ray Space Telescope [76]. Fermi telescope detected an excess of gamma rays with energies between 1- 3 GeV, which is compatible with DM particles of mass in the range 36 - 51 GeV from the center of the Milky Way [77, 78], more data are needed for the results confirmation. Other experiments such as IceCube and SuperKamiokande search for neutrinos produced from the DM annihilation in the sun. Neutrinos will escape the Sun and reach the Earth interacting with detector medium, producing Cherenkov radiation. No excess of events has been observed so far and limits on the WIMP-nucleon scattering cross sections have been set as shown in Figure 1.10. IceCube put the most strict bounds for WIMP masses above 100 GeV. Finally, different experiments such as ATIC (in a balloon) [79], PAMELA [80] AMS (in space) [81] and Auger (on the ground) [82] search for charged particles

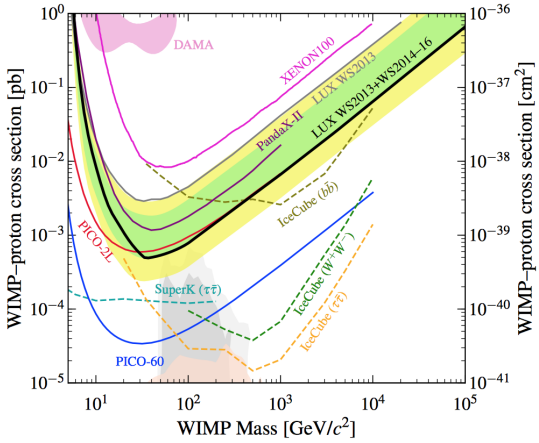


Figure 1.10: Current limits on the spin-dependent WIMP-nucleon cross sections reported by the LUX collaboration [7].

produced from DM annihilation. Several experiments claimed an increase in the positron flux but more statistics is needed to interpret that excess from DM annihilation.

1.6.3 Dark matter detection at particle colliders

These experiments search for DM particles produced from the collision of SM particles as shown in Figure 1.8 right. The produced DM particles will leave the detector without signature, but their existence can be inferred as large missing momentum balanced by visible particles produced in association with the DM. Various DM searches have been performed in ATLAS and CMS experiments, generally referred to as mono-X where X is the visible particle associated with DM production. X particle includes jets [83, 84], vector bosons [49, 85–87], photons [88, 89], leptons [90, 91] and heavy flavor quarks [92–94]. Studying different Mono-X processes allows a complimentarity within collider searches, in fact including different final states in the same process improves the limit setting. The search for the DM can be based on a complete beyond SM theory, such as supersymmetry or on a simplified model being only a minimal extension of the SM theory which include DM. In those models an attention are put in the description of the mediator and its coupling to DM and SM particles. The coupling is chosen to give the highest cross sections without violating any previous results and the mediator can be a scalar, a pseudoscalar, a vector, or an axial-vector. The latest results from CMS collaboration in the DM search have been presented in LHCP 2020 conference and are shown in Figure 1.11 [8, 95]. 95% CL exclusion limits for the scalar model for different DM searches is shown in Figure 1.11 left where the best sensitivity and the largest exclusion is given by Mono-top search , while for the pesudo-scalar model shown in Figure 1.11 right the best sensitivity and the largest exclusion are given by mono-jet followed by Mono-top search. 95 % CL exclusion regions for different DM searches in the vector model is shown in Figure 1.11 bottom left and in the Axial-vector model in Figure 1.11 bottom right. The best sensitivity is given by the di-jet searches. Those searches provide the largest exclusion which

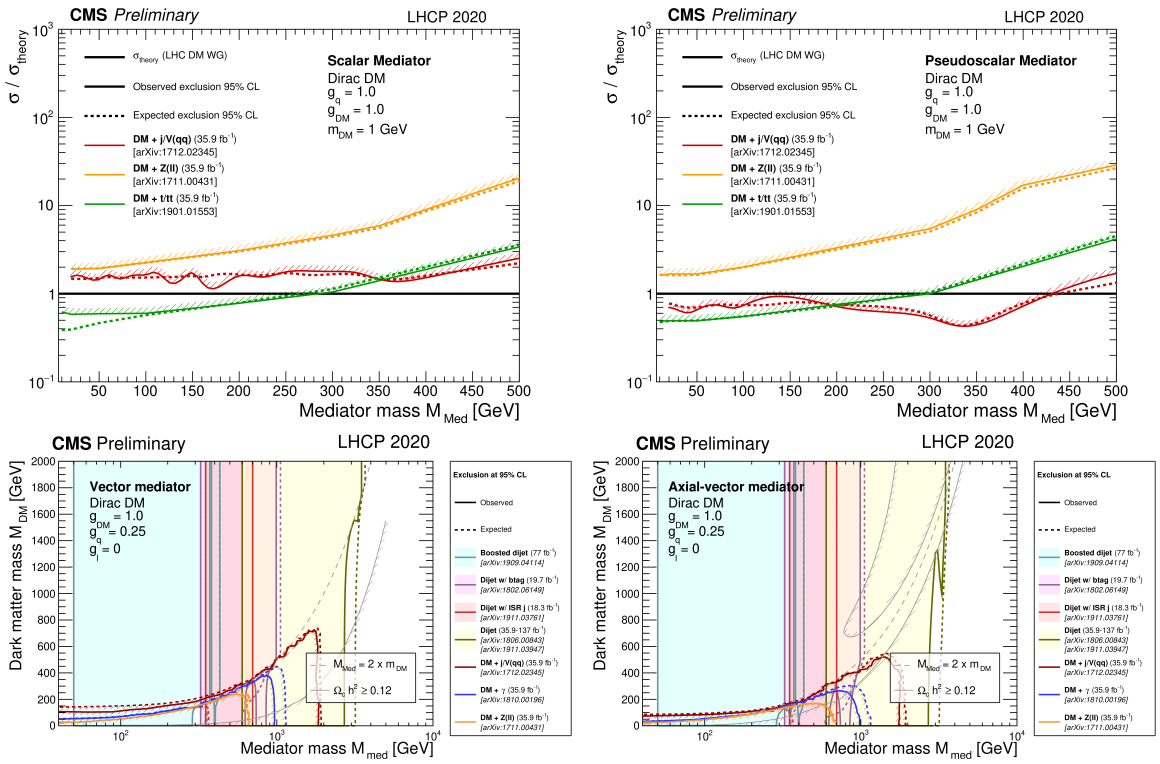


Figure 1.11: Summary plots of the latest results from CMS collaboration in the search for dark matter for (top left) Scalar , (top right) pseudo-scalar (bottom left) vector and (bottom right) axial-vector models [8].

are almost independent of the mass of the dark matter particles.

For the mono-X search the strongest exclusion limits are given by mono-jet followed by mono-γ and mono-Z. The latest results from ATLAS and CMS on the DM search can be found in the dedicated public pages [8, 96].

1.7 Mono-Higgs search

The discovery of the Higgs boson, as discussed in section 1.2, opened a new portal into searching for dark matter at the LHC. The mono-Higgs search has a benefit over mono-X signatures that emitting H as an initial state radiation is highly suppressed due to the small H-quark coupling, or a H-gluon coupling through a loop. This allow the mono-Higgs analysis to directly probe the structure of the effective DM-SM coupling. Therefore, H can be emitted from the new vertex probing directly the effective DM-SM coupling. Mono-Higgs search has been performed in the ATLAS detector at $\sqrt{s} = 8$ TeV for Higgs decaying to photons [97] and two bottom quarks [98] with results consistent with the SM prediction. At 13 TeV, the mono-Higgs search has been performed in ATLAS detector for Higgs decaying to $b\bar{b}$ [98] and at CMS detector for the five H decay modes: $b\bar{b}$ [99,100], ZZ, WW, $\gamma\gamma$ [101] and $\tau\tau$ [101]. There

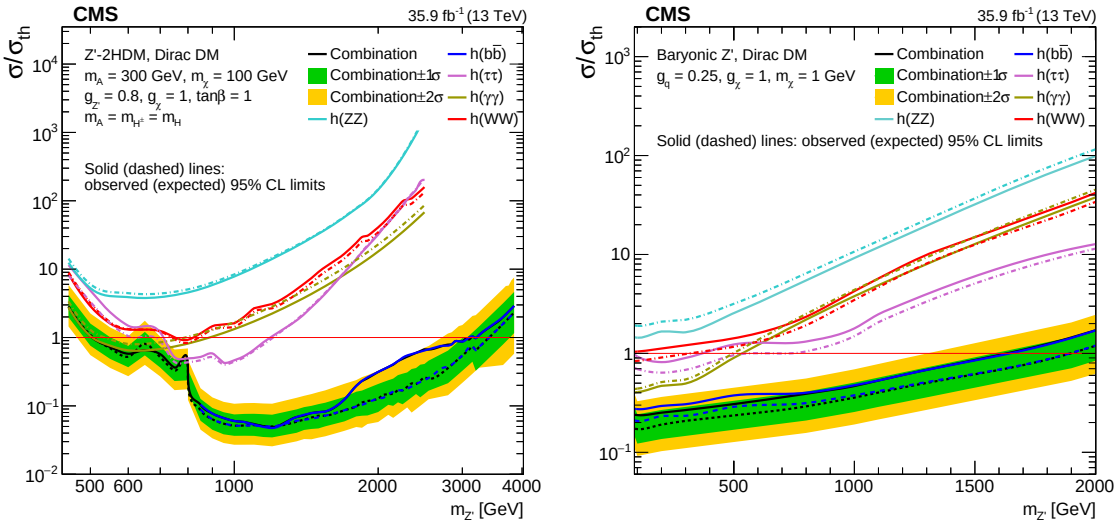


Figure 1.12: Shows the upper limits at 95% CL on the observed and expected signal strength for $Z^{'} - 2HDM$ (left) and $Z^{'}$ Baryonic (right) model for the five individual decay modes of the Higgs boson, and for their combination.

are advantages and disadvantages of each channel, the $b\bar{b}$ channel offers the highest statistical power of those decay modes since it has the highest Higgs boson branching ratio. However, this channel suffers from high multijet background. On the other hand the ZZ analysis has low SM backgrounds, but suffers of lower branching ratio. Each analysis is performed separately and the results are combined to maximize the sensitivity of the analysis. No significant excess is observed over the expected SM backgrounds in any of the five decay channels or in their combination. Therefore, limits are set on DM production in the context of two simplified models [49]. The first signal model is the $Z^{'}$- Two Higgs Doublet Model ($Z^{'} - 2HDM$) simplified model where a $Z^{'}$ mediator decay to a Higgs boson and a pseudoscalar mediator A^0 which in turn can decay to a pair of dark matter particles. The second one is $Z^{'}$ Baryonic simplified model where a $Z^{'}$ mediator radiates a Higgs boson before decaying to dark matter particles. Figure 1.12 shows the upper limit at 95 % CL on the observed and expected signal strength σ/σ_{th} for $Z^{'} - 2HDM$ (left) and $Z^{'}$ Baryonic model (right) for the five decay modes and their combination. It is noted that the sensitivity of the analysis is driven by $b\bar{b}$ while $\gamma\gamma$ and $\tau\tau$ channels play a significant role in the low mass region. Figure 1.13 shows the observed and expected 95% CL exclusion contours on signal strength for $Z^{'} - 2HDM$ (left) and $Z^{'}$ Baryonic (right) model. The range of mass $Z^{'}$ from 500 to 3200 GeV for $m_A = 300$ GeV have been excluded at 95% CL for $Z^{'} - 2HDM$ and range of mass $Z^{'}$ from 100 to 1600 GeV for $m_\chi = 1$ GeV has been excluded at 95% for $Z^{'}$ Baryonic model [48, 49].

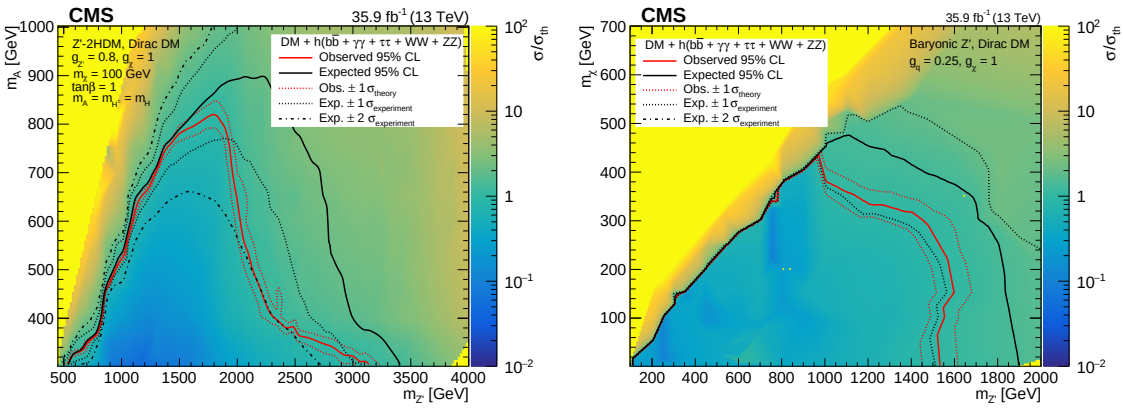


Figure 1.13: Shows the observed and expected 95% CL exclusion contours on signal strength for $Z^{'} - 2HDM$ (left) and $Z^{'}$ Baryonic (right) model. The region enclosed by the contours is excluded using the combination of the five decay channels of the Higgs decay.

1.8 Signal Model

There are many extensions to the SM theory that provide dark matter candidates and new SM-DM interaction. The simplest extensions are the simplified models of DM production. Those models follow the following requirements [102]:

- The DM particle should be stable or lives long enough to escape the LHC detectors.
- The model should contain a mediator that couples the SM - DM sectors.
- The model should fulfill Lorentz invariance.
- The additional interaction terms between the SM and DM sector shouldn't violate the SM symmetries. This means that the baryon and lepton numbers should be conserved.

Simplified models are designed to add a small number of new particles and interactions, therefore the physics of simplified models can be expressed in terms of a small number of parameters such as particle masses and couplings.

1.8.1 Two Higgs Doublet Model with pseudoscalar mediator

As mentioned in section 1.3 the open questions that SM is not able to explain need a presence of new physics. The simplest possible extension of the SM is the two Higgs doublet model (2HDM) [103]. The 2HDM is generated by adding to the complex scalar doublet of the SM another new doublet. There are different types of 2HDM theories depending on which type of fermions couples to which doublet. In the type-II 2HDM, one of the doublet ϕ_u couples to the up-type quarks only while the other doublet ϕ_d couples to the d-type quarks and the leptons. The relevant terms in Lagrangian are given by

$$\mathcal{L} \supset -y_u Q \bar{\phi}_u \bar{u} - y_d Q \phi_d \bar{d} + y_l L \phi_d \bar{l} + h.c. \tag{1.2}$$

where y_u, y_d and y_l represent the yukawa couplings of the Higgs doublets with the fermions and leptons. After the symmetry breaking, the new doublets acquire the VEVs ν_u and ν_d leading to the following parametrization:

$$\begin{aligned} \Phi_d &= \frac{1}{\sqrt{2}} \begin{pmatrix} -\sin(\beta) H^+ \\ \nu_d - \sin(\alpha) h + \cos(\alpha) H - i \sin(\beta) A^0 \end{pmatrix} \\ \Phi_u &= \frac{1}{\sqrt{2}} \begin{pmatrix} \cos(\beta) H^+ \\ \nu_u + \cos(\alpha) h + \sin(\alpha) H + i \cos(\beta) A^0 \end{pmatrix} \end{aligned} \tag{1.3}$$

where h, H are the neutral CP-even scalars with masses m_h and m_H respectively, H^+ is the heavy charged scalar with mass m_{H^+} and its antiparticle H^- with mass m_{H^-}, A^0 is neutral CP odd with mass m_{A^0}. The angle β represents the ratio of the vacuum expectation values of the two CP even Higgs bosons $tan\beta = \nu_u/\nu_d$ and α is the mixing angle between the CP even scalars h and H. The scalar h is considered as SM Higgs boson with mass $m_h = 125 GeV/C^2$.

In the 2HDM with pseudoscalar mediator "denoted 2hDM+a" the interaction between the SM and DM particles is governed by the CP-odd spin-0 mediator [104,105]. This is done by mixing the pseudoscalar P with CP-odd scalar from the Higgs doublet:

$$\mathcal{L} \supset P(i b_p \phi_u^\dagger \phi_d + h.c) + P^2(\lambda_{p1} \phi_u^\dagger \phi_u + \lambda_{p2} \phi_d^\dagger \phi_d) \tag{1.4}$$

where b_p, λ_{p1} and λ_{p2} are the trilinear and quartic portal couplings. The portal coupling b_p mixes the two neutral CP-odd weak eigenstates (A, a) with angle θ representing the associated mixing angle. Here "a" is the extra degree of freedom not present in the 2HDM. The heavy pseudoscalar A^0 couples to the SM and the dark sector, the light pseudoscalar a couples directly to DM particles. In total we have six physical Higgs bosons: a light neutral CP-even scalar h, assumed to be the observed 125 GeV Higgs boson; a heavy neutral CP-even scalar H; a neutral CP-odd scalar A^0; a light CP-odd scalar a; and two charged scalars $H^\pm$. In this search different model parameters have been scanned over while other parameters are fixed. The following parameters have been scanned:

- The heavy pseudoscalar mass m_A^0
- The light pseudoscalar mass m_a
- The ratio of the vacuum expectation values of the light and heavy scalar h and H: $tan\beta$
- The angle $sin\theta$ associated with the mixing between the light and heavy pseudoscalar a and A

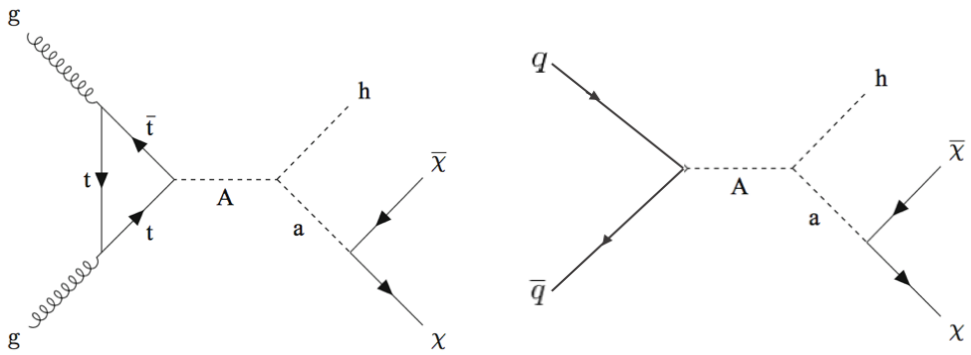


Figure 1.14: Feynman diagrams for the 2HDM+a for gluon-gluon fusion (left) and $q\bar{q}$ annihilation (right).

while the mass of the DM particle m_χ is fixed to 10 GeV and the mass of the charged Higgs and heavy scalar Higgs are assumed as the mass of the heavy pseudoscalar A $m_H = m_{H^\pm} = m_{A^0}$. The model parameter description is summarized in table 1.3. The Feynman diagram leading to Higgs + p_T^{miss} production from 2HDM+a is shown in Figure 1.14. The different scans of the model parameters considered in the thesis and the corresponding cross-sections are reported in section 5.2.1.

This chapter provided a brief description of the SM theory and its shortcomings. The evidence for the existence of the DM from cosmological observations, the possible DM candidates and the detection techniques were presented. Finally, the current limits and the model used to interpret the results were presented.

Table 1.3: Parameters description of 2HDM+a.

Particle	Description
χ	Fermionic DM particle
ϕ_u, ϕ_u	Two Higgs doublets
h, H	Light and heavy neutral CP-even scalars
$H^{\pm}$	Charged heavy Higgs
a, A^0	Light and heavy neutral CP-odd pseudoscalar
Parameter	Description
m_{χ}	DM mass = 10 GeV
m_{A^0}, m_a	mass of the heavy and light pseudoscalars
m_h	mass of the light neutral CP-even scalar, m_h = 125 GeV
m_H	mass of the heavy neutral CP-even scalar
$m_{H^{\pm}}$	mass of the heavy charged CP-even scalar
$tan\beta$	the ratio of the VEVs of the two CP even Higgs bosons
α	the mixing angle between the CP even scalar Higgs bosons
θ	the mixing angle between the two neutral CP-odd Higgs bosons

2 Experimental Setup

This Chapter gives an overview of the experimental setup of the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector used to collect the data analyzed for this thesis. The first half of the chapter includes a description of the design and the performance of the LHC while the other half reviews the design of CMS, its subdetectors, and the trigger system.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [106,107] is currently the largest and highest energy particle accelerator ever built. It has been built between 1998 and 2008 by the European Organization for Nuclear Research (known by its French acronym CERN). It is located approximately 100 m under the surface of the Earth (50 -175 m) in a ring of 27 km in circumference that crosses the French-Swiss border near the city of Geneva as shown in Figure 2.1. The goal of the LHC and its experiments is to search for new physics, study the physics phenomena at high energy scale and test the predictions of the physics models such as the SM and beyond SM theories. On 4^{th} July 2012, LHC achieved one of his goals by announcing the discovery of the last building block of the SM "the Higgs Boson" with a mass around 125 GeV/c^2. This discovery led to the award of Noble Prize in Physics in 2013 to the theoreticians Peter Higgs and Francois Englert, who predicted theoretically the existence of Higgs Boson [108–110]. LHC is designed to accelerate and collide two beams of protons at center-of-mass energy $\sqrt{s}$ = 14 TeV and with instantaneous luminosity of 10^{34} cm^{-2} s^{-1}. LHC not only collide proton-proton, but also can collide 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 and is presented in Section 2.3.

[1] The definition of luminosity and center of mass energy will be introduced in 2.5

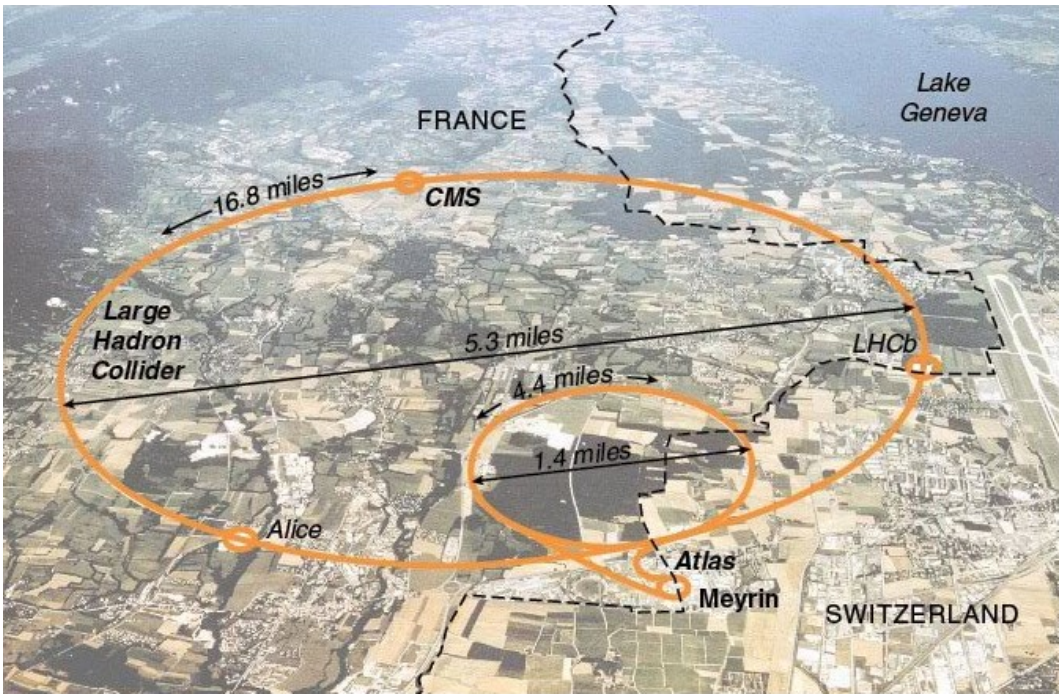


Figure 2.1: LHC overview

2.2 The LHC operation and The HL-LHC

On the 10^{th} of September 2008, LHC succeeded to circulate the proton beam for the first time but after 9 days the operation was stopped due to a serious fault in electrical connection between the superconducting magnets [111, 112]. On the 20^{th} 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 [113](at $\sqrt{s}$ = 900 GeV) and then later at $\sqrt{s}$ = 2.136 TeV. This center of mass energy exceeded the Tevatron energy (1.96 TeV), making the LHC the highest energy collider ever built.

Subsequently, on 30^{th} of March 2010 , LHC increased the energy of proton beams to 3.5 TeV leading to a center-of-mass energy $\sqrt{s}$ = 7 TeV with a new world record for the highest-energy man-made particle collisions [114]. The second period of data collection at $\sqrt{s}$ = 8 TeV started on 5^{th} of April 2012 till 6^{th} of February 2013. This collision time-period (2010-2013) is known as "Run I" of the LHC. After Run I, LHC entered a period of long shutdown for two years (known as LS1) for upgrades. LHC started its operation again in 2015 known as "Run II" with energy of 6.5 TeV per proton beam giving center-of-mass energy $\sqrt{s}$ = 13 TeV. On 3^{rd} of December 2018 Run II ended after collecting a large amount of data "about 150 fb^{-1}" and LHC entered a second long shutdown (LS2) period. After LS2, LHC is scheduled to operate again in 2022 at center of mass energy of $\sqrt{s}$ = 14 TeV, this period known as "Run III". During LS3 shutdown period, the accelerator infrastructure will be improved to cope with the increasing of energy and unprecedented instantaneous luminosity

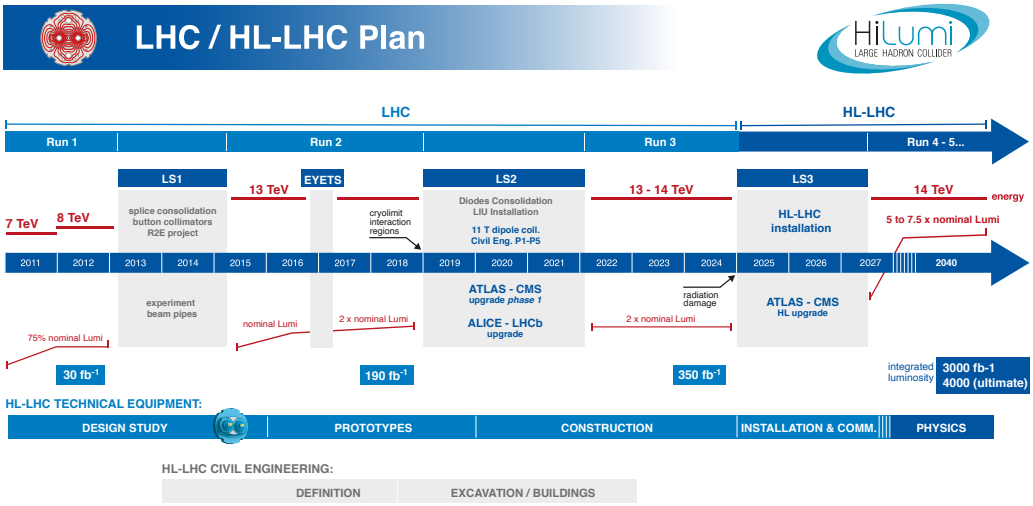


Figure 2.2: Overview of the timeline for the LHC and HL-LHC operation showing the evolution of the beam energy and integrated luminosity [9]

of 5.0 to 7.5 x 10^{34} cm^{-2} s^{-1} resulting in a total of 3000 to 4000 fb^{-1} of collected data for the full accelerator lifetime. This project is called High-luminosity LHC "HL-LHC" . Figure 2.2 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.

2.3 The CERN Accelerator Chain

The accelerator complex at CERN [115] located near Geneva is a chain of machines that accelerate particles to increasingly higher energies. Each machine boosts the energy of a beam of particles, before injecting the beam into the next machine in the sequence. The LHC is the final element in this chain of linear and synchrotron accelerators which uses the previous CERN Large Electron Positron collider (LEP) tunnel. LHC has two main injection lines, one optimized for protons (LINAC 2) and the other one for lead ions (LINAC 3) as shown in Figure 2.3. Protons are extracted by ionizing hydrogen gas by applying a large electric field to strip hydrogen atoms of their electrons to yield protons, those protons are fed into a linear accelerator (LINAC 2) and accelerated to an energy of 50 MeV. The protons are then passed into the first of three pre-LHC synchrotron accelerators, the Proton Synchrotron Booster which accelerates the protons to 1.4 GeV and separates the beam into bunches, their

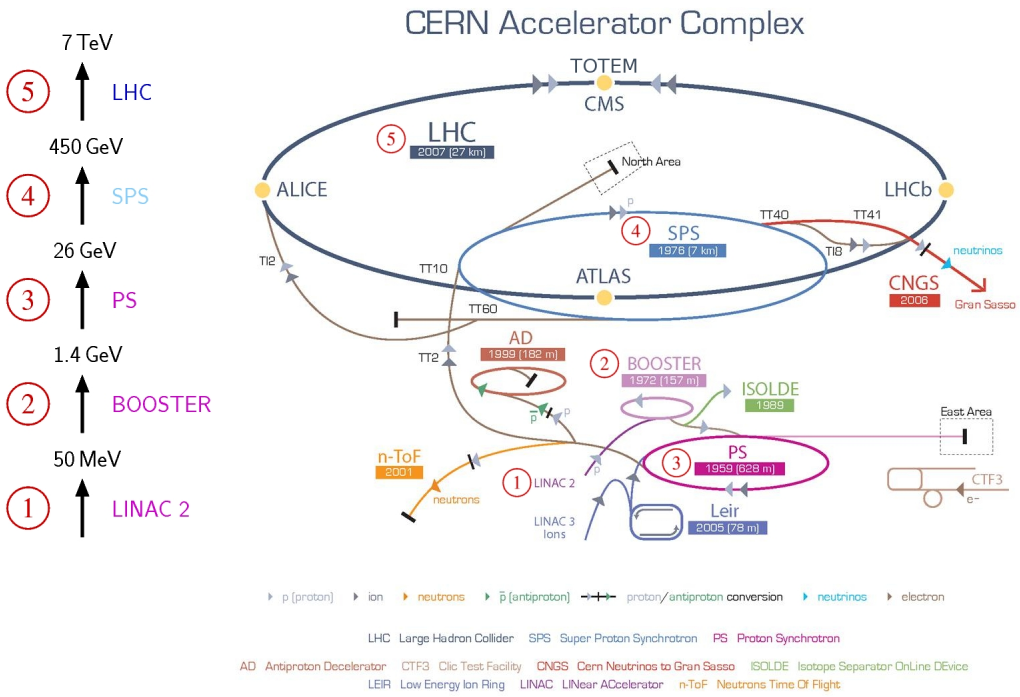


Figure 2.3: Schematic layout of the accelerator chain at CERN [10].

spacing is 300 ns apart. The proton bunches are sent to the Proton Synchrotron (PS), which accelerates the bunches to 25 GeV and splits them with spacing of 25 ns apart, this bunch spacing is the final bunch structure. Afterwards the proton bunches are injected into Super Proton Synchrotron (SPS), which has a 7 km circumference and accelerates the protons to 450 GeV. At this point, the proton bunches are ready to be injected into the LHC via two transfer lines, one for the clockwise and the other for the anti-clockwise beam. Protons beams are accelerated for 30- 45 minutes to reach maximum energy of 7 TeV per beam within LHC. For Run I, the beam energy was 3.5 TeV and 4 TeV and increased to 6.5 TeV during Run II. On the LHC's ring there are four interaction points where 4 huge detectors are installed as described in the next section.

2.4 LHC experiments

Along the accelerator ring, there exist four LHC detectors located at the main four interaction points (IP) as shown in Figure 2.4, where the proton beams collide at the center of each detector. The four detectors are described as:

- **A Toroidal LHC ApparatuS (ATLAS)** [116]: It is a general-purpose particle detector located at IP1 and the largest volume collider detector ever constructed. It was designed to cover wide range of precise measurements of physics processes, such as electroweak and strong interactions. The detector was designed especially to provide

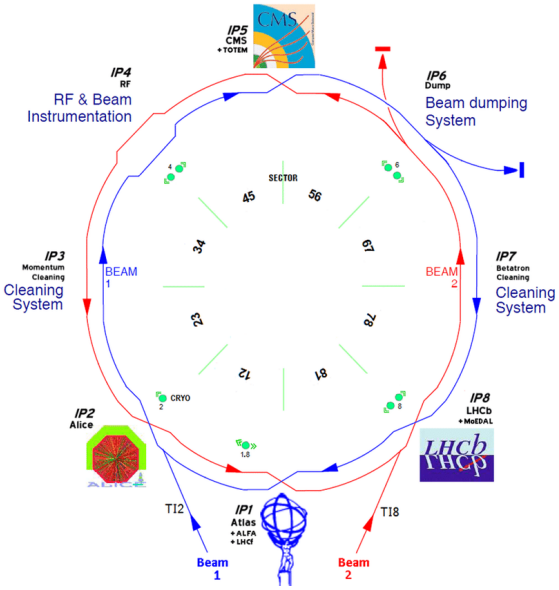


Figure 2.4: Schematic view of LHC with main interaction points housing four LHC detectors.

the best possible sensitivity to the Higgs boson discovery and search for new physics such as Supersymmetry (SUSY), dark matter, extra-dimensions, etc.

- **Compact Muon Solenoid (CMS)** [15]: a multi-purpose particle detector located at IP5, with the same physics research program as ATLAS, but with different technical design. CMS is discussed in details in section 2.6.

- **A Large Ion Collider Experiment (ALICE)** [117]: a heavy-ion detector located at IP2, designed to study the physics of strongly interacting matter at extreme energy densities. The focus of the ALICE detector is the study of Quark-Gluon Plasma (QGP) and different features of heavy-ion physics, such as proton-lead and lead-lead collisions [118].

- **Large Hadron Collider beauty (LHCb)** [119]: It is a single-sided spectrometer located at IP8, built to study matter-antimatter asymmetry puzzle and to measure CP-violation parameters . It aims to study physics with hadrons originating from b quarks. The biggest discovery of LHCb Collaboration at CERN is the observation of a phenomenon, the matter-antimatter asymmetry known as CP violation in the decays of a particle known as D^0 meson for the first time [120].

2.5 The LHC parameters and performance

LHC has many important parameters that can influence its performance such as the beam luminosity, the powerful magnet used to control the focusing and the bending of the pro-

ton beam and the collision energy available to generate new physics processes. The main parameters of LHC are described in the following sections:

2.5.1 The LHC magnet

LHC uses powerful superconducting magnets to keep the protons beams circulating around the LHC ring and focusing at the four LHC collision points. The proton beams are kept in their circular path by using 1232 superconducting dipole magnets (each of the length of 14.3 m). The magnets are kept in superconducting state by using a cryogenic system which uses superfluid helium at the temperature of 1.9^{o}K. At this temperature the dipole magnets are in a superconducting state, providing a magnetic filed of 8.33 T in the +z-direction in one beam- pipe and in the -z-direction in the other. This magnetic field is necessary to bend the proton beams around the LHC ring. The focusing of the beam is done with about 858 superconducting quadrupole magnets which keep the proton beams confined along the four LHC collision points, to increase the probability of bunch interactions.

2.5.2 The center of mass energy

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 ($\sqrt{s}$) is the sum of the energy of the two incoming beams. The LHC nominal collisional energy is $\sqrt{s}$= 7+7 = 14 TeV. Not all this energy contributes in the hard interaction but only a fraction of the protons energy. When two protons collide, two of its partons (quarks and gluons) can participate in the interaction so the effective center-of-mass energy of the hard scattering ($\sqrt{s}$) 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:

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

2.5.3 The Beam Luminosity

The luminosity is one of the most important collider parameters, the higher the luminosity of the collider the higher the probability of particle interactions to happen. Luminosity is a measure of the number of proton collisions at LHC. There are two types of luminosity: instantaneous and integrated luminosity. Instantaneous luminosity can be defined as the number of collisions per unit time and unit cross-sectional area of the beams:

$$L_{inst} = \frac{N_1 N_2 n_b f}{A} \tag{2.2}$$

where N_1 and N_2 are the number of particles in the two colliding bunches, n_b is the number of bunches in each beam, f is the revolution frequency of one bunch and A is the overlap area of the two bunches in the beam. The nominal LHC instantaneous luminosity is L_{inst}

$\approx 10^{34}$ cm^{-2} s^{-1}. Squeezing the protons into a smaller area means increasing the instantaneous luminosity and hence increasing the probability of proton collisions. Therefore the performance of the LHC can be characterized on the basis of the delivered luminosity. The integrated luminosity is the instantaneous luminosity accumulated over a period of time, as in Formula 2.3, effectively the total number of protons that have passed through a unit of area. It is a measure of the total amount of data gathered by the detector and often expressed in inverse femtobarn fb^{-1}.

$$L = \int Ldt \tag{2.3}$$

CMS accumulates higher luminosity every year of data taken during runs. Figure 2.5.(a) shows the cumulative luminosity delivered to the CMS experiment in every year of data taking during Run I (2010-2012) and Run II (2015-2018) [11]. The analysis presented in this thesis is based on the data collected by CMS detector for full Run II during 2016, 2017 and 2018 corresponding to an integrated luminosity of 35.9 fb^{-1}, 41.86 fb^{-1} and 59.74 fb^{-1} as shown in figure 2.5.(b), (c), (d) respectively [12].

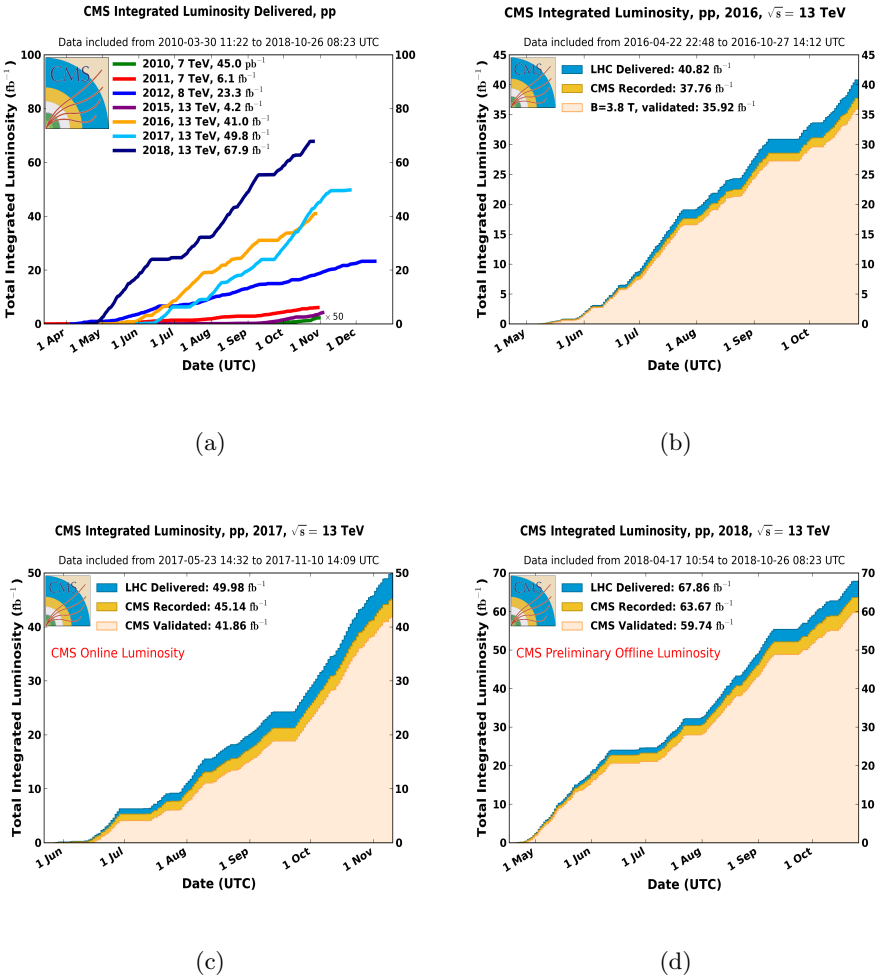
2.5.4 Number of Events at LHC

The number of events of a specific physics process is related to two main parameters, the luminosity L delivered by the machine and the physics process cross-section "$\sigma_{process}$". The number of events is given by:

$$N = L\sigma_{process} \tag{2.4}$$

Interesting physics processes such as the Higgs physics and the physics beyond the SM is predicted to suffer from low production cross section in proton-proton collisions. Figure 2.6 shows the predicted cross sections for some important SM processes at proton-antiproton and proton-proton colliders. As shown in the figure, the Higgs production cross section is several (about 3-4) orders of magnitude lower than W/Z production cross section, but it increases with increasing the center-of-mass energy of the collision. This is the reason why the colliders tend to increase the luminosity and the center-of-mass energy in order to increase the event rate.

Table 2.1 summarizes the nominal designed LHC beam parameters for the CMS and ATLAS experiments [106].



(a) (b)

(c) (d)

Figure 2.5: (a) Total integrated luminosity delivered by the LHC to the CMS detector for proton-proton collisions during Run I and Run II [11]. Total integrated luminosity delivered by the LHC to the CMS, and the total recorded versus validated luminosity by the CMS in (b) 2016, (c) 2017 and (d) 2018 for proton-proton collisions at 13 TeV [12].

Table 2.1: Nominal LHC beam parameters.

Parameter	LHC Design
Center of Mass Energy	14 TeV
Proton Energy per beam	7 teV
Bunch Spacing	25 ns
Nominal Luminosity	10^{34} cm^{-2}s^{-1}
Number of Bunches	2808
Protons per Bunch	1.1×10^{11}
Circonference	27 Km

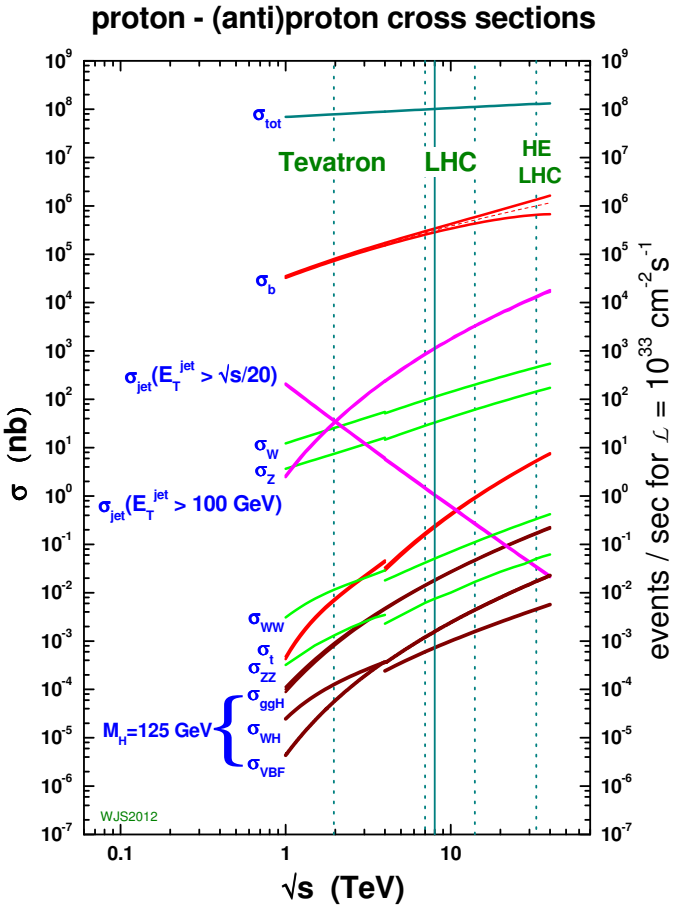


Figure 2.6: Expected cross-section versus energy [13, 14].

2.6 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) is one of two general-purpose detectors built at the Large Hadron Collider (LHC). It is designed not only for studying the SM physics processes (including the Higgs boson) but for beyond SM searches such as search for extra dimensions and DM candidates. Although CMS has the same scientific goals as the ATLAS experiment, but it uses different technical solutions and a different magnet-system design. In the following sections the CMS design and operation is outlined.

2.6.1 CMS detector design

The CMS detector layout is a cylindrical barrel shape around the beam pipe closed with two endcap disks. It has a total length of 21 m, a diameter of 7.3 m and a total weight of 14,000 ton. It is located in a 100 m underground hall below the village of Cessy in France. The detector's name comes from:

- **Compact**: due to the detector's small dimensions compared to its mass and compared to ATLAS detector's size.
- **Muon**: due to the advanced muon system in the outer layer of the detector.
- **Solenoid**: due to a superconducting solenoid magnet.

CMS has an onion structure, it is a multi-filter with different layers, where each layer consists of a subdetector designed to stop, track or measure different types of particle emerging from proton-proton or heavy-ion collisions. It is divided into four major sub-detectors, as shown in Figure 2.7. Three sub-detectors are located inside a superconducting solenoid magnet: the tracking system, the electromagnetic and hadronic calorimeter. The last sub-detector is the muon system interspersed within an iron magnetic flux return yoke on the outside of the solenoid magnet.

These sub-detectors with different materials are used to give informations about different particles and measure their energy, momentum and charge and hence identify them. The high magnetic field (≈ 3.8 T) from the solenoid magnet provides enough bending power for all the charged particles coming from the collision. The precision measurement of the particles track curvatures in the tracking detectors helps in identifying the particle's charge and momentum. Photons and electrons produce electromagnetic showers in the electromagnetic calorimeter, from these showers the particle's energy is measured. The more penetrating hadrons such as charged pions produce hadronic showers in the hadronic calorimeter, which help in measuring their energy. Muons are the only particles can traverse all the subdectors and reach the muon system. Figure 2.8 shows the signatures left when different types of the particles pass through the CMS sub-detector. Combining the informations from different subdetectors helps in discriminating between particles. As we can see from the figure both photons and

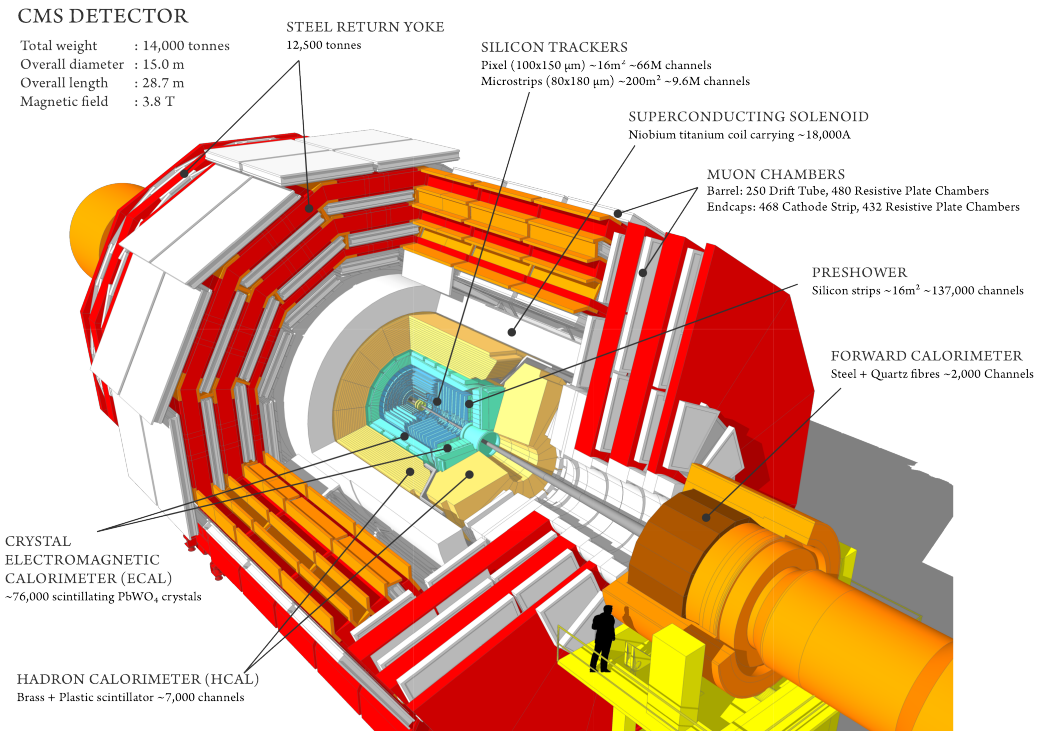


Figure 2.7: An overview of the CMS detector with major sub-detectors [15].

electrons leave showers in the electromagnetic calorimeters but photons don't leave tracks in the tracking system and hence we can distinguish photons and electrons. 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.

2.6.2 The CMS Coordinate Systems

Standard coordinate system is defined to describe the detector. The origin of coordinates is taken to be the interaction point at the center of the detector. The z-axis is coincident with the anticlockwise beam direction , the x-axis is pointing towards the center of the LHC ring and the y-axis is perpendicular to the x-z plane and points vertically upward as shown in figure 2.9. The azimuthal angle called ϕ is defined in the x-y plane such that $\phi = 0$ corresponds to the x-axis directed toward the center of the LHC ring and takes values of $[-\pi, +\pi]$. The polar angle called θ is defined such that $\theta = 0$ lies along the beam-pipe in the positive-z axis and takes values of $[0, +\pi]$. Taking into account that the particles produced by proton collisions are strongly boosted to the collision axis and distributed over the angle θ very unevenly, it is more convenient to use different kinematic variable called pseudorapidity (η), which is a good approximation of the rapidity with avoiding energy dependence. The η value is determined by the Formula 2.5:

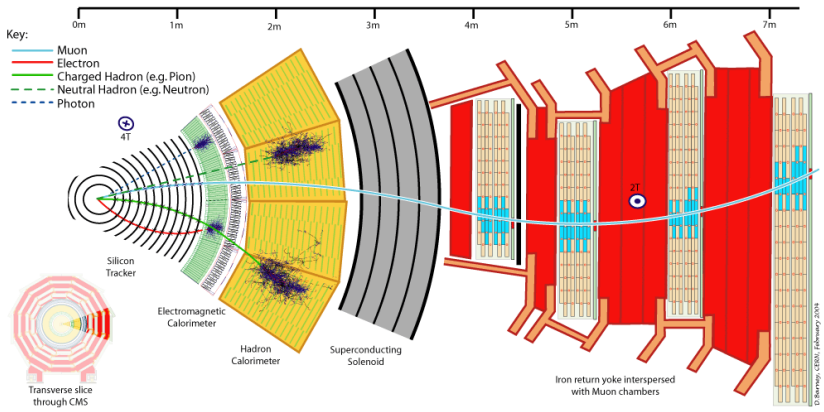


Figure 2.8: The CMS detector transverse section with signatures of different types of detected particles [10].

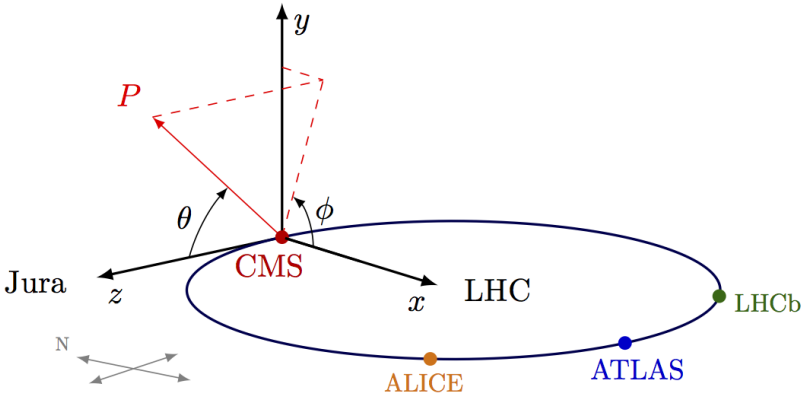


Figure 2.9: The CMS coordinate system.

$$\eta = -ln\left(tan(\frac{\theta}{2})\right) \tag{2.5}$$

Smaller (larger) values of η represent the particles lying in a direction perpendicular (parallel) to the beam axis. The distance between two particles is given in terms of a two dimensional angular distance (ΔR) in the $\eta - \phi$ plane and is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{2.6}$$

Another commonly useful variable is the transverse component of the momentum p_T. Before the collision, the parton momentum is expected to be longitudinal (along the beam axis) while the transverse momentum is expected to be negligible. The transverse momentum is particularly useful because the conservation of momentum can be used to quantify an imbalance in the energies of the outgoing particles and hence predict the outgoing particles escaping the detector without leaving signature such as Dark Matter candidate used in this analysis.

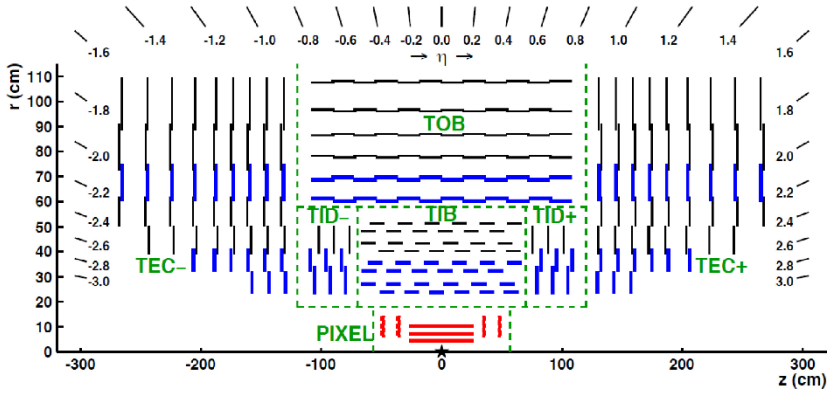


Figure 2.10: Overview of the CMS tracker system [16].

2.7 The Tracking System

The silicon tracker is the first sub-detector near the interaction point. Therefore the tracker material must be very resistant to the radiation. It has a diameter of 2.5 m, a total length of 5.8 m and covers a pseudorapidity range of $\mid \eta \mid < 2.5$. The tracking system is used to measure with high resolution the charge and momentum of all charged particles as they propagate outward from the collision point while minimizing the energy loss that the particles experience from passing through matter. The tracker can reconstruct the trajectory of all the charged particles coming from the collision point besides the tracks coming from the decay of very short lived particles, allowing the identification of primary and secondary vertices with very high resolution [16].

The tracking system is divided into a silicon pixel sub-detector at inner radius and a silicon strip sub-detector at outer radius. A general overview of the CMS tracker system is given in Figure 2.10, with its subsystems: Pixel, Tracker Inner Barrel (TIB) and Tracker Inner Disks (TID), Tracker Outer Barrel (TOB) and Tracker Endcaps (TEC).

2.7.1 The silicon pixel sub-detector

Silicon pixel detector is the first sub-detector lies around the interaction point. It consists of four Barrel Pixel (BPix) layers lies around the beam-pipe, three endcap disks (EPix) in each end-cap region as shown in figure 2.11.

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 data acquisition (DAQ) system or no trigger is received and the data is overwritten.

In order to accurately reconstruct particle tracks and measure their momentum, one must know precisely where the particle originated from primary or secondary vertex. This is especially important in the identification of b quark and τ lepton decays which have lifetimes

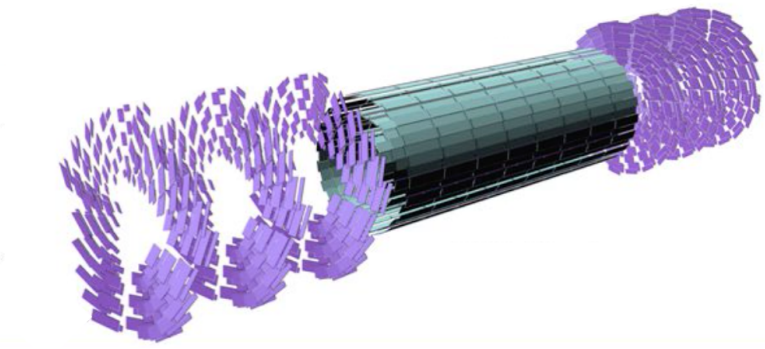


Figure 2.11: View of the CMS silicon pixel sub-detector geometry. In the middle there are three layers of the barrel pixels, and on either end of the barrel are two end-cap disks [17].

sufficient to allow them to travel several millimeters from the primary vertex before decaying. The pixel detector is meant to distinguish these secondary vertices from the primary collision points.

However, over the time, the pixel detector will suffer from degradation of the resolution due to radiation damage. To keep its high performance, an upgrade "Phase Pixel Upgrade" was performed during the technical stop at the end of 2016 [121] [122]. The fourth layer in barrel and the third disk in end-cap were added in the pixel upgrade phase during the technical stop.

2.7.2 The silicon strip sub-detector

After the silicon pixel sub-detector and in the way out of the tracker lies the strip sub-detector. This system uses modules comprised of silicon microstrips instead of pixels. It consists of 10 cylinder layers in the barrel region around the pixel sub-detector: four-layer Tracker Inner Barrel (TIB), six-layer tracker outer barrel (TOB), and 24 disks of endcap modules, 12 on each end of the barrel cylinders: three disk Tracker Inner Disks (TID) and nine disk Tracker Endcaps (TEC). It extends to $|\eta| < 1.6$ in the barrel region and extends to $|\eta| < 2.5$ in the end-cap. The silicon strip detector 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.

2.8 The calorimetry sub-detectors

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 of the incident particle based on the depth of penetration into the calorimeters and the profile of the energy lost. However calorimeter

can't 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 first is the electromagnetic calorimeter (ECAL) which measures the energies of electrons and photons, the second is the hadronic calorimeter (HCAL) which measures the energy of hadrons and the third is the Forward calorimeter which is designed to detect the energies of particles in the very forward regions of the detector.

2.8.1 The Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) [18] provides very precise energy measurements. ECAL is the only sub-detector providing information about photons which necessary for the analysis of the $H \rightarrow \gamma\gamma$ decay. The ECAL is also used in the electron reconstruction: the combination of its information with the tracker provides a very precise measurement of electron position and energy necessary to analyse the multi-lepton final state.

ECAL uses an array of lead tungstate crystals ($PbWO_4$) to detect particle energies. The lead tungstate crystals are characterized by their high density (8.28 g/cm^3) and short radiation length X_0 (0.89 cm) which provide the possibility to absorb electron and photon showers with reasonably short crystals and also characterized by good separation of electromagnetic showers. Finally the fast response time of the crystals is compatible with the LHC collisions, around 80% of the light is emitted in 25 ns.

When an ionizing particle enters the crystal, it will ionize the atoms within the crystal, loosing its energy. As the crystal re-absorbs the released electrons, the atomic transitions result in a release of blue scintillating light that can be measured by attached photo-detectors. The incident particle may radiate photons in the material (bremsstrahlung process) that photo-converts into light particle electron-positron pairs which photo-convert again, and so on in a process known as electromagnetic showering. The total amount of scintillation light produced in this process is proportional to the amount of energy lost by the incident particle in the crystals. If the particle stops completely, then the total energy of the particle can be known.

The ECAL provides a pseudorapidity coverage of $|\eta|< 1.479$ in the barrel which is called the ECAL Barrel (EB), and to $1.479 <|\eta|< 3.0$ in the endcap region named ECAL Endap (EE). Figure 2.12 shows the longitudinal view of the CMS electromagnetic calorimeter with the ECAL barrel, an ECAL endcap and the preshower in front of it. Figure 2.13 shows the ECAL layout with the arrangement of crystal grouped into modules and super-modules in barrel and two half-disks or dees in the endcaps, with preshower in front.

A preshower (ES) detector is installed in front of the ECAL endcaps in the forward region of the detector to distinguish between single photon emission and photons coming from the pion $\pi^0 \rightarrow \gamma\ \gamma$ decay to two photons very close to each other.

The pre-shower detector consists of four layers of alternating lead and silicon strip de-

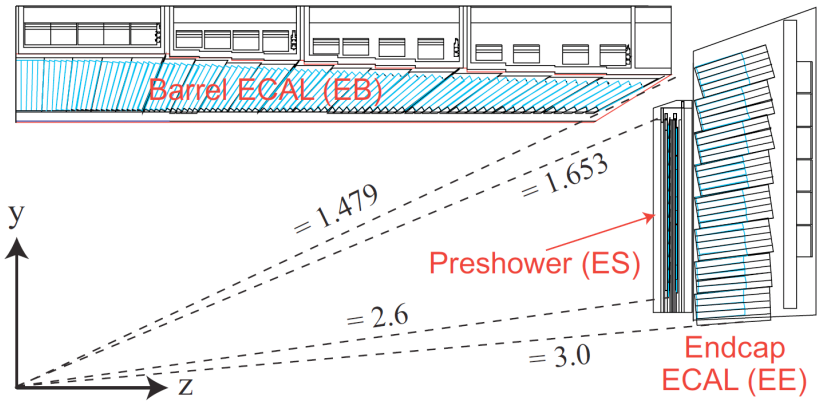


Figure 2.12: Longitudinal view of part of the CMS electromagnetic calorimeter showing the ECAL barrel and an ECAL endcap, with the preshower in front.

tectors. The electromagnetic showers are produced when the incident electrons and photons interact with Lead. 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. In this way the pre-shower detectors are used to measure the shower profile of incident particles in order to distinguish between photons and π^0 mesons.

The energy resolution of the ECAL can be affected by different several sources, such as the electronic noise, calibration errors and the crystal non-uniformities. The ECAL energy resolution has been measured on one barrel super module during test beam [123], it can be expressed as a function of measured energy by:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{(E)}}\right)^2 \oplus \left(\frac{12\%}{\sqrt{(E)}}\right)^2 \oplus (0.3\%) \tag{2.7}$$

The energy here is expressed in GeV. A resolution of 1 % has been achieved for electrons with energy greater than 15 GeV.

2.8.2 The Hadronic calorimeter

Hadronic calorimeter [19] (HCAL) lies between the ECAL and the inner side of the solenoid magnet. It is designed to measure the energy and position of hadronic jets. The HCAL consists of three components: the HCAL Barrel detector (HB), the HCAL End-cap detector (HE), and the HCAL Forward detector (HF) as shown in Figure 2.14.

HCAL is made of dense brass as absorber and plastic fluorescent scintillator which acts as active material. When a hadronic particle hits a plate of absorber, it initiate an interaction producing several secondary particles. When these secondary particles move through further layers of absorber they can interact and result in a cascade or "shower" of particles. As this shower develops, the particles pass through the alternating layers of active scintillation material causing them to emit light. The energy of a particle is then measured as the sum of the light energies emitted by the scintillator over the path of the particle.

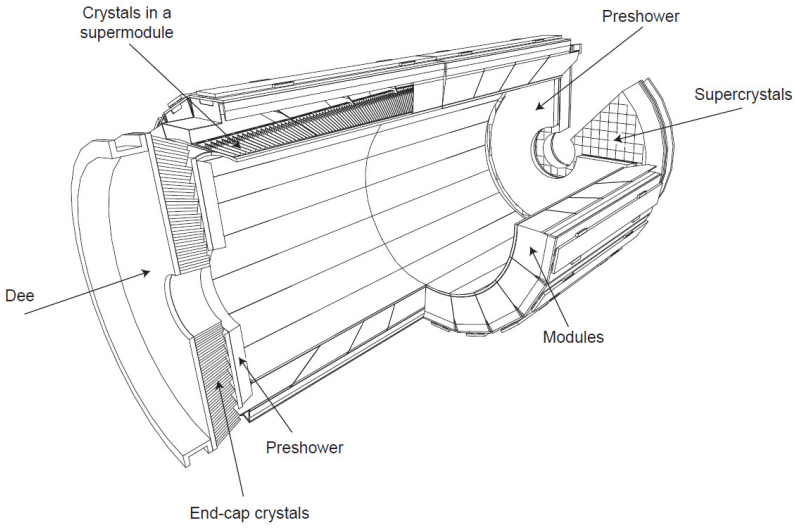


Figure 2.13: Schematic view of the CMS ECAL with the arrangement of crystals into modules and super-modules distributed in the barrel and two half-disks or dees in each endcap. The pre-shower detector covers most of the endcap regions. [18].

The HCAL is consisting of 70,000 tiles in total. It extends to pseudorapidity $\approx \mid \eta \mid = 1.3$ in barrel and to $1.3 < \mid \eta \mid < 3.0$ in endcap.

A layer of scintillation material is attached to the outer layer of the solenoid magnet. This portion of the barrel detector, called the HCAL Outer calorimeter (HO), uses the magnet material as an absorber to "catch" the shower particles that punch through the barrel to ensures that no energy leaks out at the back of the HB undetected.

The forward HCAL detector is mounted on the outside of the iron magnetic return yokes at a distance of 11.2 m along the z-axis, and extented in pseudorapidity of $3 < \mid \eta \mid < 5$ as shown in figure 2.15 and useful to identify and reconstruct very forward jets.

The HCAL energy resolution has been measured during test beam [124] and can be expressed as:

$$\left(\frac{\sigma_E}{E}\right) = \left(\frac{84.7\%}{\sqrt{(E)}}\right) \oplus (7.4\%) \tag{2.8}$$

The resolution is about 30 % for energy of 30 GeV and the resolution goes down to 10 % for higher energies. Compared to the ECAl energy resolution, the energy resolution is dominant by HCAL resolution.

2.9 The Solenoidal Magnet

The CMS solenoid is one of the most powerful superconducting solenoidal magnets [125]. It has an internal diameter of 6 m and length 12.5 m. It is designed to reach a magnetic

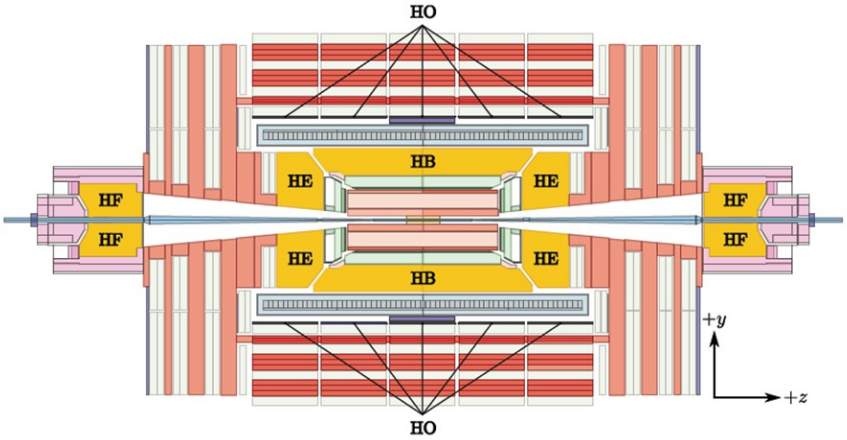


Figure 2.14: A view of the CMS detector in y-z projection with the components of the hadronic calorimeter labeled.

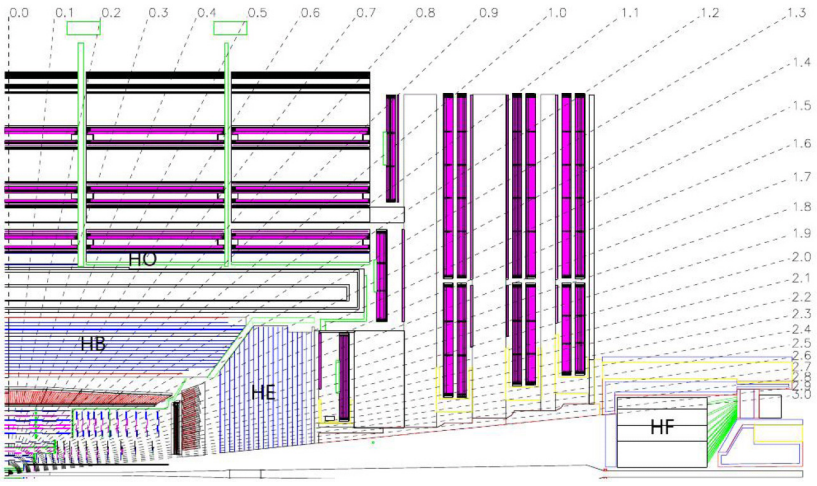


Figure 2.15: Longitudinal view of the CMS detector. The locations of the hadron barrel (HB), the endcap (HE), the outer (HO) and the forward (HF) calorimeters [19].

field density of 3.8 T, approximately 100,000 times stronger than the earth's magnetic field. The magnetic field is essential in identifying the particles emerging from collisions: it curves their trajectory allowing to calculate their momentum and to identify their charge. Stronger is the magnetic field, highly curved will be the particle's trajectory which allows obtaining higher momentum measurements resolution. A 10,000 tonnes iron yoke is used to contain the magnetic field within the detector volume which is capable of stopping almost all particles (except muons and neutrinos).

2.10 The Muon System

On the outer most layer of the CMS detector lies a second set of tracking systems "The muon system" [20]. It has been designed to provide an independent muon identification, a precise measurement of muon momentum and charge as well as an efficient muon trigger. It provides a powerful support for the discovery of new physics with muons in the final state.

The muon chambers are placed at the very edge of the experiment since the muons are the only particles that penetrate several meters of iron without interacting, in contrast to most of the particles that are stopped by the CMS's calorimeters or the iron yoke, and hence they are the only particles capable to provide a signal.

The robustness and redundancy of the muon spectrometer are guaranteed by the presence of three types of gaseous detector: Drift Tubes (DT) in the barrel region, Cathode Strip Chambers (CSC) in the endcap and Resistive Plate Chamber (RPC) in both barrel and endcap regions. The detectors are arranged in a framework of an iron yoke used to guide the field. As the muons pass through the iron yoke, their paths are bent in the focused magnetic field allowing to measure their momentum and charge.

The muon system is composed of a central barrel ($| \eta | < 1.2$) and two closing end-caps ($1.2 < | \eta | < 2.4$), the barrel region is divided into five wheels (Wheel 0, Wheel ± 1, Wheel ± 2). The muon chambers are installed on the outer and inner sides of the yokes arranged in four stations at different radius. Each station consists of 12 chambers, one per ϕ sector, except for the fourth station where 14 chambers are present. The two endcaps are made of 3 iron disks and 4 stations, divided in 2 or 3 rings where the ring is formed by 18 or 36 trapezoidal chambers.

Figure 2.16 A schematic view of the CMS Muon spectrometer. The interaction point is located at the lower left corner where Z=0. The Drift Tubes (DT) are shown in yellow where MB denote Muon Barrel, the Cathode Strip Chambers (CSC) are shown in green where ME denote Muon Endcap and Resistive Plate Chambers (RPC) shown in blue for barrel and endcap regions where the return yoke is shown in dark gray.

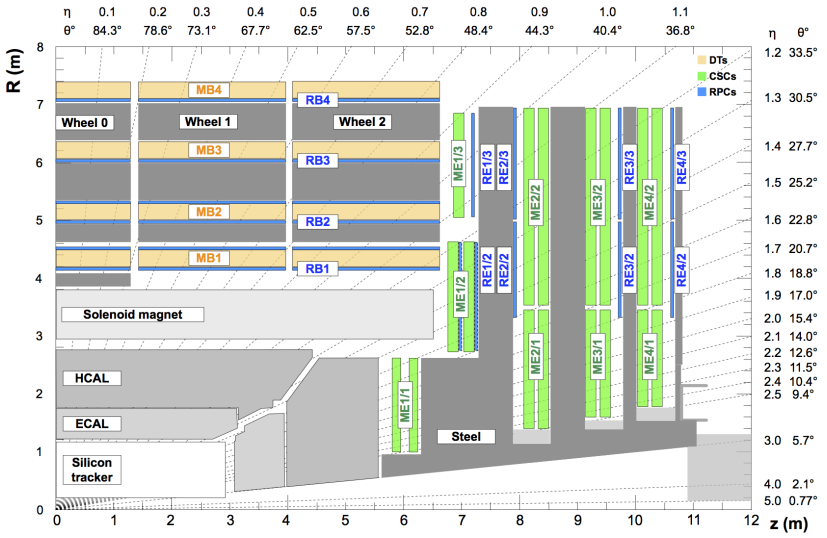


Figure 2.16: Schematic view of the CMS Muon spectrometer.

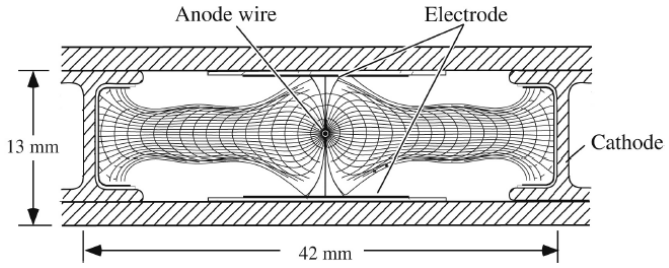


Figure 2.17: schematic view of Drift tube chamber [20].

2.10.1 The Drift Tubes

The Drift tubes (DT) [126] are positioned in the barrel region ($| \eta | < 1.2$) and have four stations named MB1, MB2, MB3 and MB4. In Barrel region the particle rate is low and the magnetic field is well contained in the iron plates of the magnet return yoke which allows the usage of DTs in those uniform conditions.

The basic element of a DT chamber is a drift cell: it is a rectangular tube with cross section of 42 x 13 mm^2 as shown in Figure 2.17. The two shorter sides of the rectangle (I-shaped) represent the cathode while the anode is 50 μm stainless steel wires located in the center of the cell. The chamber is filled with a 85%/15% gas mixture of Ar/CO_2. The working principle is based on the ionization: when a charged particle passes through the DT, it ionizes the gas producing electrons that drift toward the wire and induce a fast signal on the wire.The drift time is measured and converted into distance using the knowledge of drift velocity. A single drift cell has an efficiency of about 99.8 % and a resolution of 180 μm.

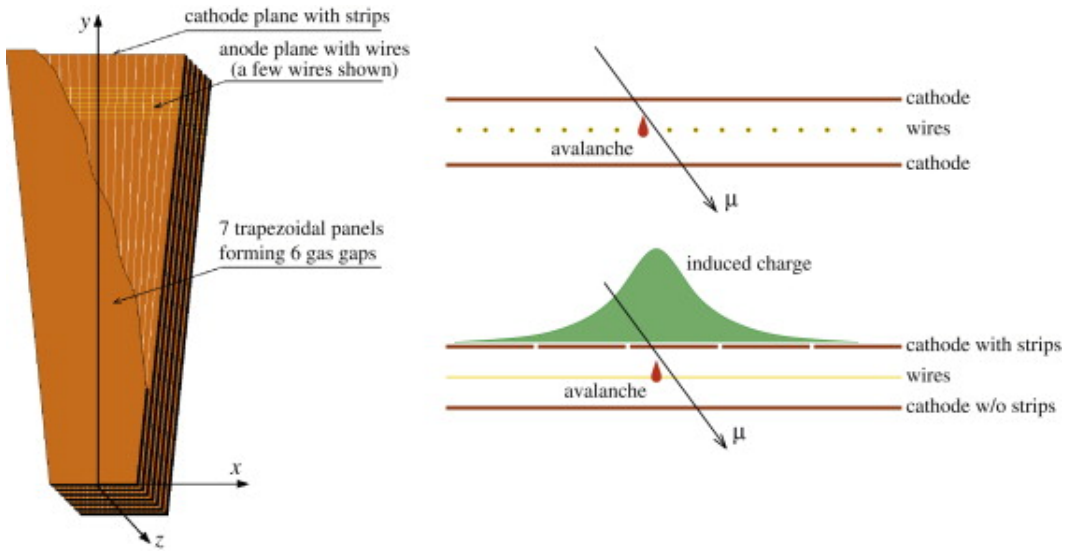


Figure 2.18: schematic view of cathode strip chamber (left) and the principle of operation (right) with cross-section shows the wires (right-top) and cross-section shows the strips (right-bottom) [20].

2.10.2 Cathode Strip Chambers

Cathode strip chambers (CSC) [127] are used in the end cap regions ($0.8 <| \eta |< 2.4$) and have four stations, namely ME1, ME2, ME3, and ME4. The detectors are able to provide precise time and position measurement in the endcap region where the particle rate is high and the magnetic field is inhomogeneous. CSC are used due to their better performance in the non uniform magnetic field, they are also faster and finer segmented than the drift tube.

CSCs are multi-wire proportional chambers consists of two cathode planes, one of them is segmented into strips running along the ϕ direction, and orthogonal to array of anode wires laying between these two planes as shown in figure 2.18 left.

The chamber is filled with 40%/50%/10% mixture of $Ar/CO_2/CF_4$. The working principle is based on the ionization and shown in Figure 2.18 right: when a muon passes through the chamber it ionizes the gas, the electrons drift towards anode wires forming an avalanche that is collected by the wires. This avalanche induces an electrical charge on several adjacent cathode strips, and the position of the muon along the wire can be determined by measuring the distribution of charge picked up by the strips. Each individual chamber has a trapezoidal shape and is made of seven cathode panels stacked together, forming six gas-gaps, each containing an array of anode wires and shown in Figure 2.18 left. The CSC chamber has a spatial resolution of 80 μm.

2.10.3 Resistive Plate Chambers

Resistive plate chambers (RPC) are gaseous detectors developed early 1980s [128] [129]. Similar to other gaseous detectors, the RPC working principle is based on ionization process.

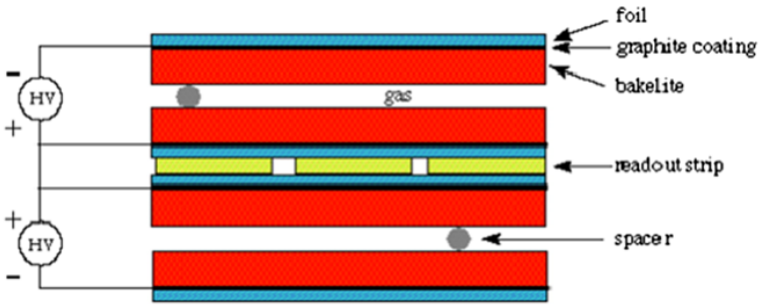


Figure 2.19: The layout of double gap RPC [20].

RPCs are used in CMS detector in both barrel (along with DTs) and end-cap (along with CSCs) regions. RPCs have an excellent time resolution about 1.5 ns, this time is shorter than the 25 ns LHC bunch spacing, therefore RPC detectors are used for muon triggering purpose.

A total of 1056 RPC detectors are present in CMS muon system covering $\mid \eta \mid < 1.9$ region. The RPCs geometry depends on their position in the detector. There are six layers of RPC chambers in the barrel region, four of them are attached to each side of DT chambers (MB1 and MB2), the other two layers are attached to the inner side of MB3 and MB4. In end-caps region; there are four disks of trapezoidal RPCs attached to the CSCs chambers [130].

The CMS RPC chambers consist of two layers of gas gaps "called double gap geometry" with a sheet of segmented copper readout strips sandwiched between them as shown in Figure 3.1.

The chambers operate with a humidified gas mixture composed of 95.2 $\%C_2H_2F_4$(Freon-R134a), 4.5%$iC4H_{10}$ (isobutane) and 0.3%SF_6 (Sulphur Hexafluoride). In the next Chapter, the RPCs detectors are discussed in more detail.

2.10.4 The CMS RPC system upgrade

The muon system worked efficiently for 10 years of operation during Run I and II with the the designed luminosity of the LHC. In the view of HL-LHC project, both the collision energy and the instantaneous luminosity will increase, which means higher pileup and hence higher background rates. Those operating conditions will be a hard challenge for the muon subdectors. To cope with those new operating conditions, the CMS RPC group has planned an extensive upgrade project [131, 132]:

- **Validation of the present RPC system for HL-LHC**: The present RPC system worked efficiently during LHC Run I and II without showing any aging effect or losing the detector performance. However, the present RPC system has to prove that they can sustain the hard background conditions during the HL-LHC running period. An extensive irradiation test on spare RPC chambers is ongoing. The detectors parameters

and performance are monitored periodically to spot any possible aging effect, this test is explained in detail in chapter 3.

- **Installation improved RPC in high η region** : The high η region of the muon system is still empty and not equipped with the existing RPC detectors, since the background rate is expected to be high in this region and the current RPC will not be able to handle with this high background rate. An improved RPC "iRPC", based on the RPC technology, but with improvements that allow to increase the rate capability, will be installed in this high η region in the stations RE3/1 and RE4/1 [133].

2.11 The Trigger System

The trigger system is the first step of the physics event selection process [134]. It performs a fast selection of events that seems to be interesting for physics analysis from the huge amount of events produced by LHC collisions. 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 about 20 proton-proton collisions and hence that result in about 800 million collisions per second. The data event 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 $\sim$ 1 MB of memory to be recorded on the tape and therefore the final memory requirement is 40 TB per second. This huge amount of data can't 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 $\sim$ 25 ns such as per the LHC requirements. The trigger system consists of two independent levels, the Level-1 (L1) and the High Level Trigger (HLT) described in the following:

2.11.1 The Level-1 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 [135,136]. The L1 trigger searches for key signatures of interesting events: leptons, photons, hadronic jets, and p_T^{miss} , 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 L1 Global Trigger (GT) which decides whether to pass or reject the event. The event information is sent to the High Level Trigger (HLT) system if the L1 trigger decision is positive otherwise

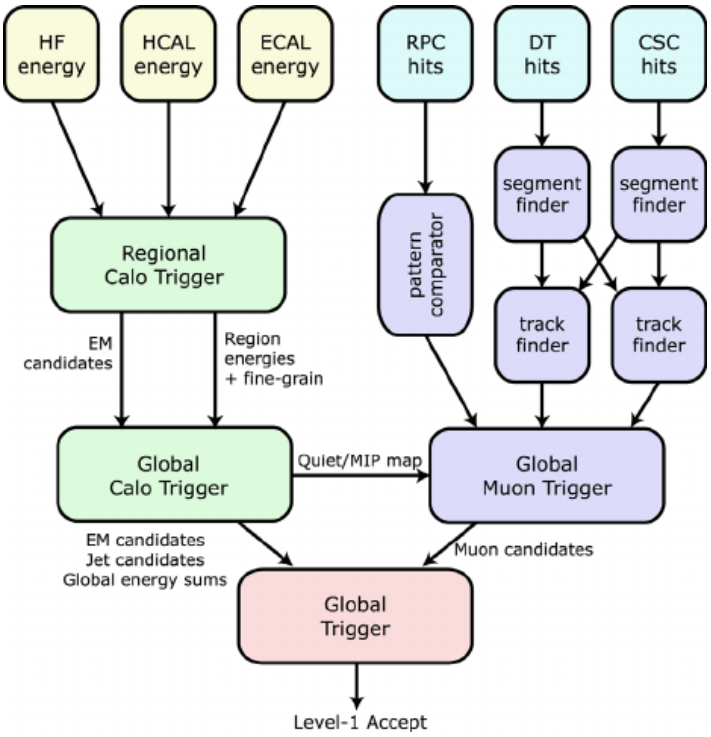


Figure 2.20: Flow diagram of the L1 Trigger system of CMS.

the event will be deleted. Figure 2.20 shows the schematics of the L1 trigger system.

2.11.2 The High Level Trigger

Opposite to the L1 trigger, the High Level Trigger (HLT) system [137] is software based. 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 data acquisition (DAQ) system for further physics analysis and transferred to CERN Tier0 (T0) storage system. Then data quality monitoring (DQM) checks the quality of the recorded data and labeling the datasets either good or bad. At the end of the DQM chain, a list of certified datasets is produced to be used later for physics analysis.

In the case of MonoHiggs analysis which uses four leptons final state, the trigger looks for events containing leptons (electron and muon) signals. For the Level-1Trigger, an electron signature is an energy deposit in the ECAL, and a muon signature is a track segment or a hit pattern in muon chambers. Then the High-Level Trigger combines the L1 trigger information with the tracker and pre-shower information.

This chapter provided a description of the LHC accelerator and its experiments. A detailed description of CMS experiment and its sub-detectors is presented besides the CMS trigger system. At this point, after the description of the CMS sub-detectors especially the muon chambers; it is a good opportunity to describe the study performed on the RPC

chambers which are currently used in CMS detector in the next chapter.

3 Aging Study of Resistive Plate Chambers for HL-LHC

Gas detectors are commonly used as tracking devices in various experiments with different design to cope with the experiment requirements, but with the basic working principle. The gas detector layout and configuration have evolved through the time from the single-wire proportional chambers to a wide range used highly efficient detectors. The basic working principle for gas detectors is the same: the ionization created by the passage of the charged particle through the detector material. This chapter gives a detailed description of Resistive Plate Chamber (RPC) working principle, design and configuration of RPCs inside the CMS experiment, followed by a part dedicated to the aging study being performed on RPC detectors which are currently used in CMS. Four RPC chambers are under continuous gamma radiation at CERN Gamma Irradiation Facility (GIF++) to monitor the detector parameters and performance [1].

3.1 The Resistive plate chamber

The Resistive Plate Chambers (RPC) [129] are parallel-plate gaseous detectors that combine good spatial resolution with a time resolution. RPC's excellent time resolution allows the capability of tagging an ionizing event in time shorter than the 25 ns between two LHC successive bunch crossing. The good performance of RPC's is crucial in assigning the muon to the right bunch crossing. The CMS RPC system is contributing to the trigger, reconstruction and identification of muons.

3.1.1 RPC Design

The CMS RPC chambers consist of two layers of 2 mm gas gaps "called double gap RPC" with a sheet of segmented copper readout strips sandwiched between them. A grid of polycarbonate spacers are distributed inside the gas gap to ensure the rigidity and the thickness of the gap. The gas gap consists of two sheets of high resistive bakelite plates having 2 mm

[1] The work done in this chapter has been published in [50]

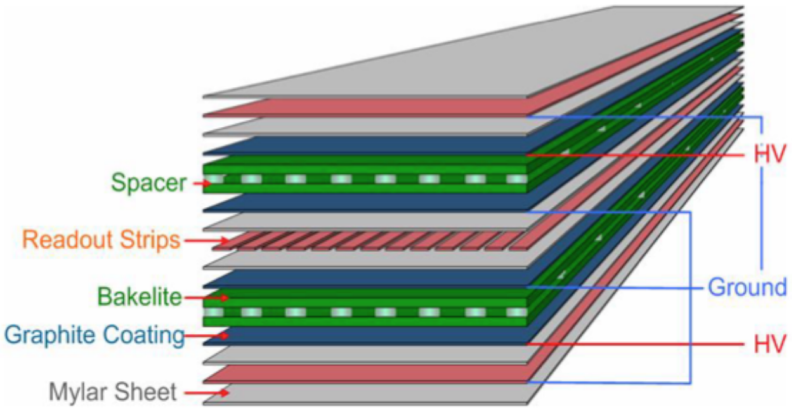


Figure 3.1: The layout of double gap RPC.

thickness and resistivity $1 - 6 \times 10^{10}$ Ω cm. The inner bakelite surfaces are coated with linseed oil, while the outer surfaces are coated with conductive graphite paint to form the high voltage and ground electrodes. The electrodes are connected to a HV supply in order to form a uniform and intense electric field in the gap. The sheet of the readout strips is centered on the bottom gap to read out the signals. The strips are separated from the graphite layers by an insulating PET (polyethylene terephtalate) film. There are 32 strips running the length of the chamber but they are broken up into three sections for the readout purpose. In this configuration, the total induced signal is the sum of the two single-gap signals. A simplified layout of the double gap design is shown in Figure 3.1

3.1.2 The CMS RPC Detector Layout

The CMS RPC system consists of 1056 RPCs distributed in the barrel and endcap regions. The RPCs are organized in 4 stations called RB1 to RB4 covering a pseudorapidity range of $|\eta| < 1.2$ in the Barrel region, and RE1 to RE4 in the Endcap region covering a pseudorapidity range of $0.9 < |\eta| < 1.9$ in the Endcap [20, 130].

- **In the Barrel region:**

  There are 480 RPC installed and distributed equally in 5 wheels along the beam pipe named Wheel 0, Wheel ±1 and Wheel ±2 as shown in Figure 3.2. Each wheel consists of 4 muon stations (called RB1, RB2, RB3 and RB4) from inside to outside the CMS detector and divided into 12 sectors (called S01-S12) in ϕ. The inner two stations consist of 2 layers of RPCs (RB1in, RB1out, RB2in and RB2out) with a DT sandwiched between them and the outer two station consists of a single layer of RB3 and RB4 and one layer of DTs.

- **In the Endcap region:**

  There are 576 RPCs are installed and distributed along 8 disks (4 disks per each endcap side, positive and negative) named RE±1, RE±2, RE±3 , RE±4 as shown in Figure 3.3.

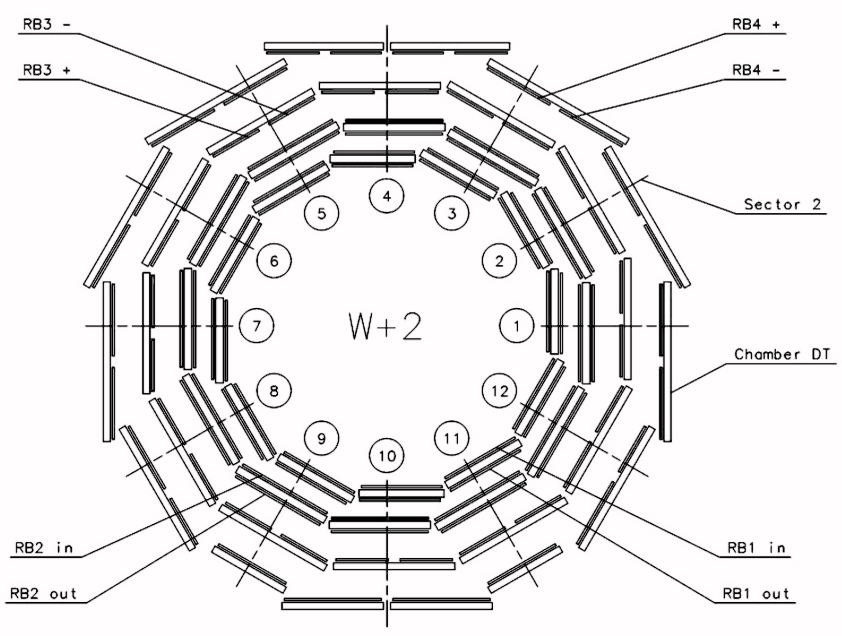


Figure 3.2: CMS RPC system layout in the Barrel for one Wheel W+2.

Endcap RPCs have different geometry than barrel RPC besides the fine segmentation which is essential for higher multiplicity forward regions. Each disk is divided into 36 sectors along ϕ and has two rings called ring 2 (R2) and ring 3 (R3). Each ring consists of 36 trapezoidal chambers, each chamber splits into three rolls, labeled as A, B and C ordered respectively in increasing η. Due to the division of the read-out strips into three rolls, the TOP layer gap is divided into two gaps: Top Narrow (TN) and Top Wide (TW) while the bottom (BOT) layer consists of one gap as shown in Figure 3.4.

3.1.3 The Principle of Operation

RPC working principle is based on the ionization process like all gas detectors. When a charged particle passes through the gas, it ionizes the gas atoms producing electron-ion pairs. Some of the formed electrons have enough energy to produce further ionization, and those formed electrons are grouped into clusters. The applied electric field on the gap causes the electrons to drift towards the anode. Due to their smaller mass, the electrons have higher drift velocity than the ions, they are accelerated towards the anode and ionizing more molecules in their way forming avalanche as shown in Figure 3.5; this process is called "multiplication".

The multiplication process is characterized by a parameter α (first Townsend coefficient), which represents the number of ionizations per unit length, and by a parameter β which is called the attachment coefficient and represents the number of electrons captured by the gas per unit length. The parameter β becomes important in the case where the gas mixture has electronegative gas (the gas having a high tendency to absorb free electrons to form an ion).

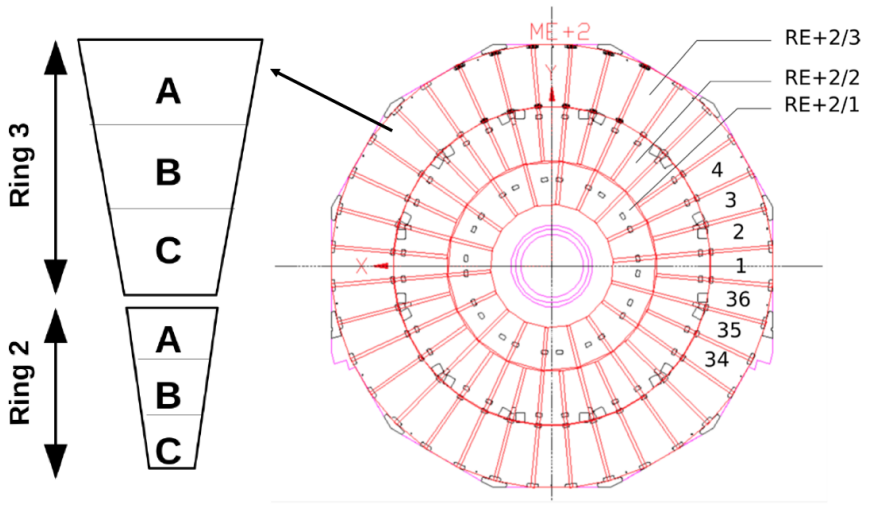


Figure 3.3: CMS RPC system layout for one endcap disk with 3 rolls are shown.

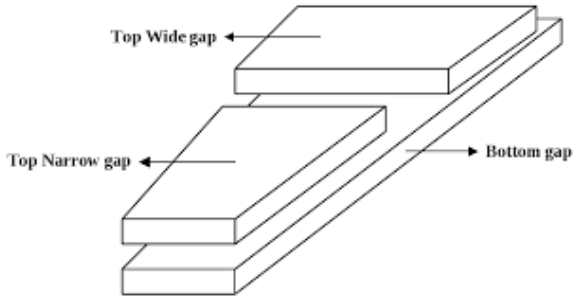


Figure 3.4: CMS RPC gaps layout in the endcap.

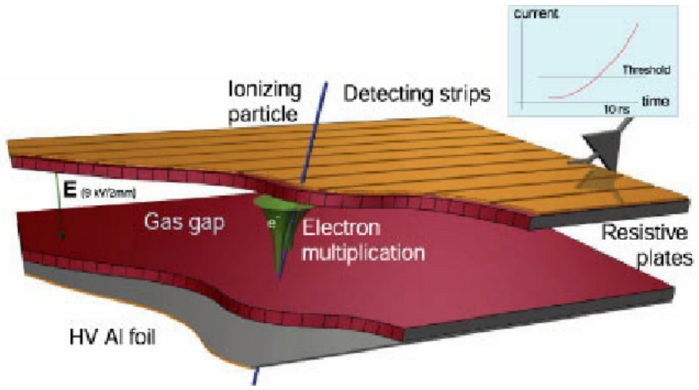


Figure 3.5: RPC working principle.

Let's denote the number of primary electrons created by an ionizing particle by n_0 and the distance that the electrons travel from the point where the cluster is formed to the anode by x. The number of electrons that reach the anode will be given by :

$$n = n_0 e^{(\alpha-\beta)x} \tag{3.1}$$

The gain factor of the detector is defined in terms of the cluster size per ionization as

$$M = n/n_0 \tag{3.2}$$

"M" is used to define the working mode of the RPC detector. There are two modes where RPCs can be operated, the avalanche or streamer mode (discharge mode). If the value of M is of order 10^8 or higher this corresponds to the streamer mode while M $< 10^6$ corresponds to avalanche mode.

The induced signal is collected using the readout strips and transmitted to the readout electronics. Despite the signal strength in the streamer mode is quite large as compared to the one in avalanche mode, the rate capability of an RPC operating in the streamer mode is limited. The streamer mode is suitable for low count rate experiments. The CMS RPC working in avalanche mode has a high rate capability of 300 Hz/cm^2.

3.1.4 The Gas Mixture for RPC

The selection of the gas mixture components play an important role in the detector operation mode. Choosing a suitable gas mixture prevents the shift from avalanche to streamer modes. The RPC uses a Fluorine (F) based gas as a primary component of the gas mixture and uses quenching gas which absorbs the photons and restricts the formation of secondary avalanches and electronegative gas to limit the avalanche development to the streamer. The CMS RPC operates with a humidified gas mixture composed by 95.2% $C_2H_2F_4$ (Freon-R134a) which is the charge carrier gas and has to be ionized by the incident charged particles, 4.5% iC_4H_{10} (isobutane) which used as quencher gas to absorb UV photons from molecule de-excitation and 0.3% of SF_6 (Sulphur Hexafluoride) the electronegative gas absorbing any excess of electrons which will result in streamer mode [20, 138, 139]. The standard gas mixture used in RPC has high Global Warming Potential; based on CERN strategy to reduce green house gases emissions, an intensive R&D is carried to find new environmentally friendly gases to replace the present RPC gas and ensure high detector performance [140, 141].

3.2 RPC Aging study

Gas detectors can suffer from aging effects when exposed to high radiation for long time which result in a degradation of detector performance appearing as loss in detector efficiency, in-

crease in dark current[2] and rise in noise rates. The main reason for this detector performance degradation are the chemical processes that happen inside the electron multiplication region where the gas fragments produced inside avalanches could form polymers accumulating on the electrodes' surface [142]. During Run I and Run II the CMS RPC system provided stable operation, excellent performance and did not show any evident aging effects after collecting $\approx 10\ mC/cm^2$ of integrated charge. Nevertheless, the expected conditions during HL-LHC, in terms of background, pile-up and the probable aging of the present detectors, will make the muon identification and correct p_T assignment a challenge.

3.2.1 Goals of the longevity study

During 10 years of LHC operation, the present CMS RPC system has been certified at a maximum background rate of 300 Hz/cm^2 and a total integrated charge of 50 mC/cm^2 [143–145]. Based on the data collected by CMS during LHC Run II and assuming a linear dependence of the background rates as a function of the instantaneous luminosity, the expected background rates and integrated charge at HL-LHC will be about 600 Hz/cm^2 and 840 mC/cm^2, respectively (including a safety factor of three) [146]. Those operating conditions are much harder with respect to those for which the detectors have been designed and can induce non-recoverable aging effects due to the higher collision rates and pile-up, that can affect the detector performance and properties. Therefore, a long term irradiation test has been carried out in Gamma Irradiation Facility (GIF++) to accumulate the charge expected at the end of HL-LHC and hence study whether the present RPC detectors can survive the hard background conditions during the HL-LHC running period. During the longevity studies, I'm continuously monitoring the main detector parameters and performance as a function of the integrated charge to spot any possible aging effects [50, 147]

3.2.2 The Gamma Irradiation Facility

The Gamma Irradiation Facility known as "GIF++" was designed and built at CERN during the Long Shutdown1 (LS1) period in spring 2015 to allow irradiate and test large size detectors. The GIF++ is a unique place equipped by a gamma source (13 TBq Cs-137) and a system of movable filters for varying the gamma flux which allow to test the detectors in a background condition similar to the ones at LHC/HL-LHC [148]. In addition to the gamma source, a 100 GeV muon beam is provided 3-4 times per year for detector performance studies. The GIF++ provides a fairly realistic simulation of the LHC/HL-LHC conditions. The GIF++ facility provides also a controlled monitor of the environmental parameters during irradiation such as temperature and pressure. The GIF++ layout is reported in Figure 3.6

[2]Dark current is the current produced in the chamber when applying high voltage in the absence of background radiation.

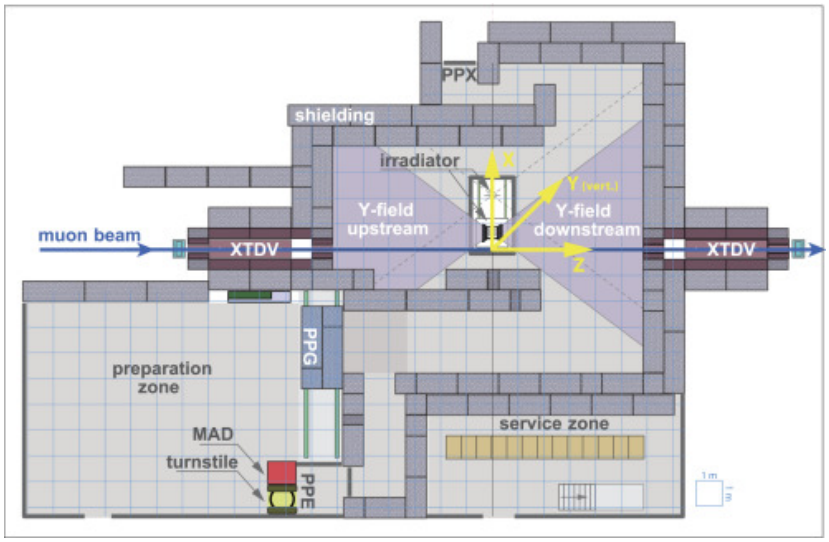


Figure 3.6: CERN Gamma Irradiation Facility (GIF++) layout.

3.2.3 The Longevity test setup and procedure

A long term irradiation test is ongoing since 2016 at GIF++ on four spare end-cap chambers (two RE2/2 and two RE4/2) since the maximum background rate is expected in the end-cap region. The detectors are trapezoidally shaped with height = 1687 mm , long side = 979 mm and short side = 684 mm. Since the endcap RPC production has been done in two periods (all RPCs in the endcap system have been done in 2015, except RE4/2 and RE4/3 chambers, which were made in 2012 - 2013) we irradiate both types in the test. To spot any possible aging effect due to long term irradiation, two chambers (one RE2 and one RE4) are continuously under irradiation while the other two chambers of the same type are kept as reference and switched on from time to time. The setup in GIF++ is shown in Figure 3.7 left while a sketch shows the chambers position with respect to the source (Figure 3.7 right). All the detectors are flushed continuously with the standard gas mixture where the detectors are currently running with gas humidity $\approx$ 60% and 3 gas volume exchange per hour for irradiated chambers and one gas volume exchange per hour for reference chambers. The gas flow and gas humidity were changed during irradiation period as shown in section 3.3.2. The irradiated chambers are continuously operated at high voltage ($\approx$ 9.8 kV) and under gamma radiation of $\approx$ 600 Hz/cm^2. The applied High voltage (HV_{app}) is not exactly equal to the effective high voltage (HV_{eff}) in the gas volume due to the small variations of GIF++ environmental temperature and pressure.

The high voltage applied (HV_{app}) to the detector is calculated by normalizing the effective high voltage (HV_{eff}) wanted in the gas volume at the standard temperature and pressure (T_0 = 293.15 K, and P_0= 990 mbar, and the coefficients α = 0.8) using the formula [149]:

$$HV_{app} = HV_{eff}\left(1 - \alpha + \alpha\frac{P_0}{P}\right)\frac{T}{T_0} \tag{3.3}$$

During irradiation the accumulated integrated charged (Q_{int}) is calculated as the average

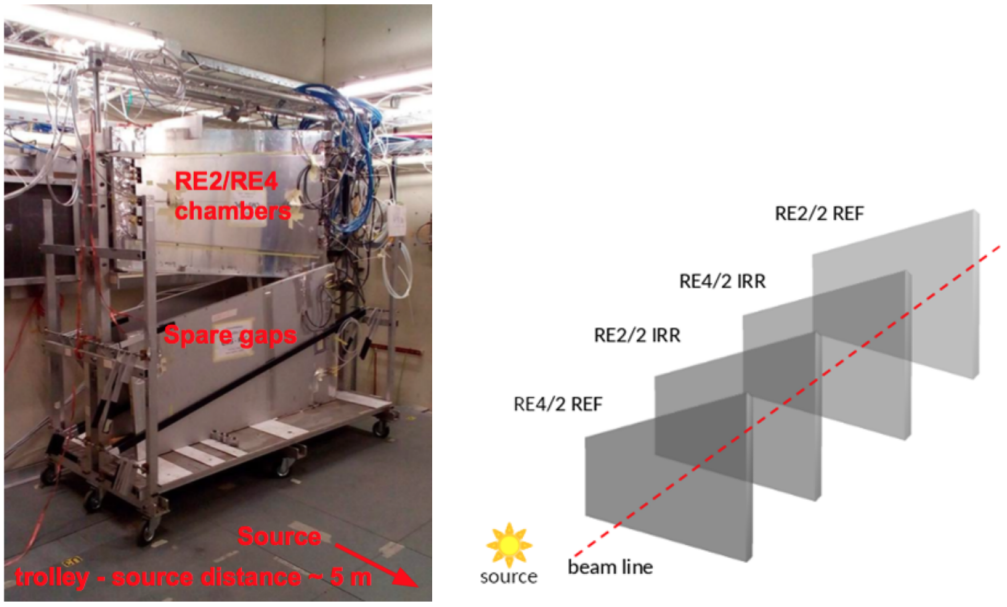


Figure 3.7: The longevity setup at GIF++ (left), a sketch of the chambers' position with respect to the source (right).

current density (J_{mon}) accumulated in time in the three gaps that constitute the detectors as shown in the formula:

$$Q_{int} = \int_{t_i}^{t_f} J_{mon} dt$$
$$J_{mon} = \frac{I^{TW} + I^{TN} + I^{BOT}}{A_{TW} + A_{TN} + A_{BOT}} \tag{3.4}$$

where I^{TW}, I^{TN} and I^{BOT} are the current measured in top wide, top narrow and bottom gaps, A_{TW}, A_{TN} and A_{BOT} are the surface area of each gas gap. The accumulated integrated charge versus time for both RE2/2 and RE4/2 irradiated chambers from the beginning of irradiation up to now is shown in Figure 3.8. At present, about 748 mC/cm^2 and 428 mC/cm^2 integrated charge has been accumulated for RE2/2 and RE4/2 irradiated chambers, which correspond at around 89% and 51% respectively of the expected integrated charge at HL-LHC (840 mC/cm^2). The accumulated integrated charge for RE4/2 chamber is lower than RE2/2 chamber because the RE4/2 chamber has been irradiated a few months later due to limitation in the total gas flow. Taking into account that the detector validation includes a safety factor of three, it means that the present accumulated charge in RE4/2 and RE2/2 already correspond to the maximum expected integrated charge, and twice the maximum expected integrated charge at HL-LHC, respectively.

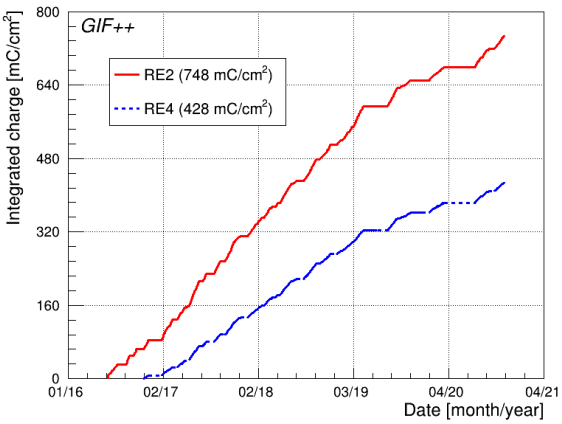


Figure 3.8: Integrated charge versus time, accumulated during the longevity test at GIF++ for RE2/2 (solid red line) and RE4/2 (dashed blue line) chambers.

3.3 Monitoring the detector parameters

The detector parameters (such as dark current, noise rate, current and count rates at several background conditions) are monitored continuously and compared with the measurements from the reference chambers to spot any degradation in the detector parameters due to long term irradiation. Moreover, the detector performance is studied at different irradiation fluxes when the muon beam at GIF++ is available.

3.3.1 Dark current and noise rate

The dark current and noise rate are measured periodically, without background radiation (source is off) in order to spot any possible degradations on the surface of the electrodes due to irradiation. Figure 3.9 (top) shows the dark current density (current normalized to the surface area of the detector) monitored as function of the effective high voltage at different values of collected integrated charge. The dark current density for RE2 both irradiated and reference chambers as a function of collected integrated charge is shown in Figure 3.9 (bottom). The dark current values were measured at 6.5 kV (left), which represent the ohmic current, and at 9.6 kV (right), which includes the gas amplification. The dark current is almost stable with time with small acceptable variations of dark current level since the beginning of irradiation. The dark current values remain low as measured during the chamber production certification. Figure 3.10 shows the average noise rate for RE2 irradiated (blue) and reference (red) chamber as a function of collected integrated charge where the mean noise rate is calculated as the average among all 96 strips distributed over the three detector η partitions, the average noise rate is stable with time and less than 1 Hz/cm^2. The low and stable dark current and noise rate indicate a good state and quality of the inner gaps surface, excluding any possible aging damaging so far.

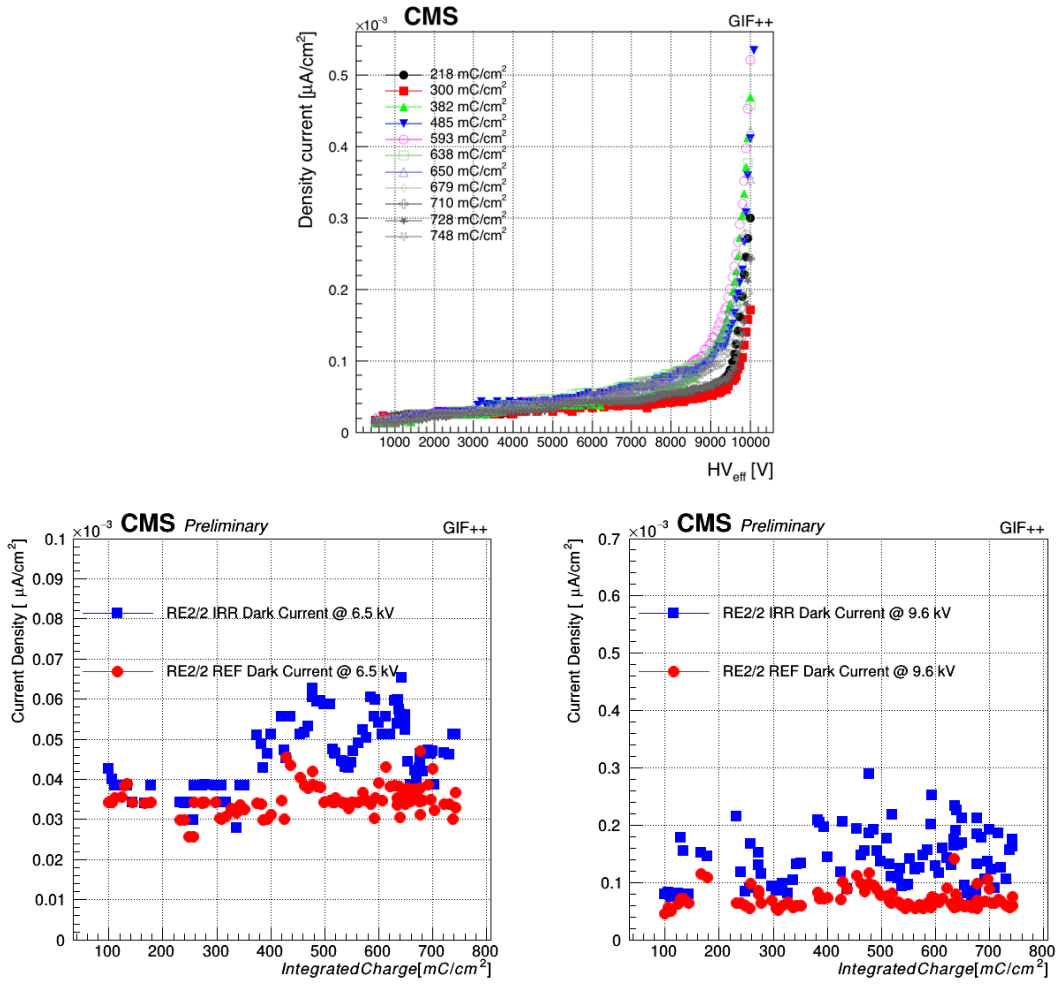


Figure 3.9: Top: Dark current density monitored as a function of the effective high voltage at different values of collected integrated charge for RE2 irradiated chamber. Bottom: Dark current density for RE2 irradiated (blue squares) and reference (red circles) chambers as a function of collected integrated charge at 6.5 kV (left) and at 9.6 kV (right).

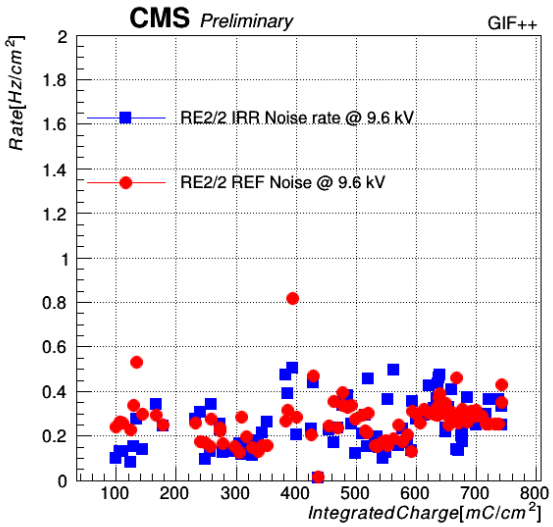


Figure 3.10: Average noise rate as a function of collected integrated charge for RE2 irradiated (blue squares) and reference chambers (red circles).

3.3.2 The Resistivity and the current

The current with the presence of background radiation "source on" is measured periodically as well. In addition, the electrode's resistivity is measured several times per year since it is a crucial parameter that influences the detector performance.

The resistivity is measured by irradiating the detector while filled up with pure argon and operating in a self-sustaining streamer mode (Argon is characterized by a sufficiently high Townsend coefficient even when the applied electric field is relatively low around 2 kV) [150]. The gas is ionized and when the gas quenching components such as the isobutane are removed the streamers propagate over all the detector area causing the appearance of a discharge and so a kind of short circuit is created between the resistive anode and cathode. By measuring the current and the applied high voltage the resistance (R) can be calculated and hence the resistivity (ρ) using the Formula:

$$\rho = \frac{R \times S}{L} \tag{3.5}$$

where R is the resistance measured, S is the bakelite plate surface area and L is the bakelite plate thickness $\approx$ 4 mm. The resistivity for each chamber is calculated as the average with respect to the three gaps which made up a chamber. The measured resistivity values are normalized to 20 ^{o}C to allow comparing the measured resistivity values at different temperature conditions and at different time [151]. Figure 3.11 (left) shows the resistivity measured as a function of the accumulated integrated charge for RE2/2 irradiated (blue) and reference (red) chambers. Both chambers have the same trend which depends on the GIF++ environmental temperature and humidity variation.

Figure 3.11 shows the resistivity ratio and the current ratio "current measured with the

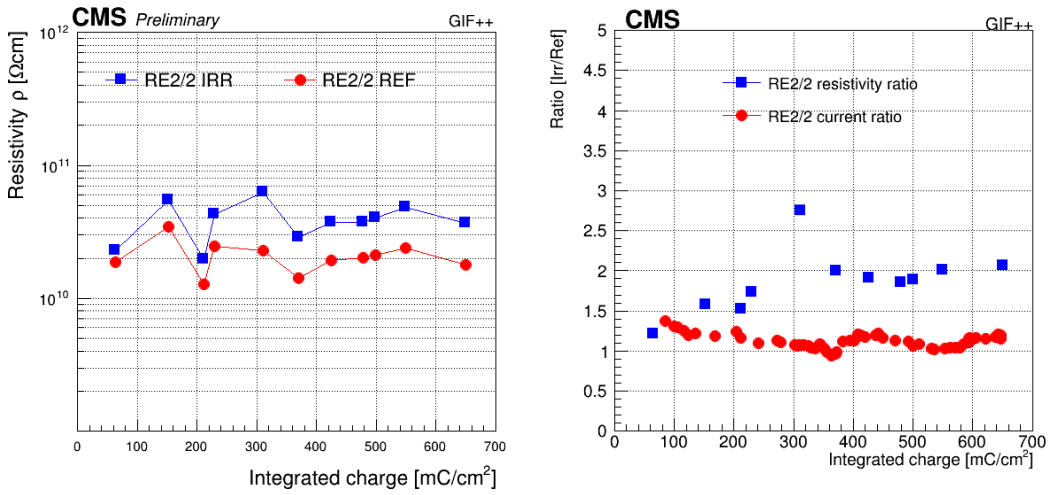


Figure 3.11: Left: Average measured resistivity for RE2/2 irradiated (blue) and reference (red) chambers as a function of the collected integrated charge, Right: Resistivity ratio (blue squares) and current ratio (red circles) between RE2 irradiated and reference chambers as a function of collected integrated charge.

presence of gamma background " between irradiated and reference chambers, taking these ratios to exclude the effect of external parameters. In the first irradiation period up to ≈ 300 mC/cm^2 an increase of resistivity has been observed. At this irradiation period the detectors operated in similar conditions as in CMS, the gas flow is one gas volume exchange per hour and the gas humidity is $\approx$ 35-45%. Those operating conditions were optimized for CMS, but they are not optimal with respect to the high gamma background rate (≈ 600 Hz/cm^2) at GIF++. Those operating conditions led to a drying up of the bakelite plates and hence resistivity increase, which is also observed in the decrease of the current. At ≈ 300 mC/cm^2 the detectors operated with new conditions, the relative gas humidity was increased and maintained at $\approx 60\%$, and the gas flow was increased in the irradiated chamber to three gas volume exchanges per hour. After applying those conditions the bakelite resistivity reduced (observed also in the current increase) proving that the observed resistivity increase was related to the operating conditions and it is a recoverable effect.

3.4 Monitoring of the detector performance

The detector performance has been measured during test beams before irradiating the chamber and at different periods of irradiation. The last measurement was done at 479 mC/cm^2 at the last muon beam in GIF++ in 2018. Figure 3.12 shows the RE2 irradiated chamber efficiency measured as a function of the effective high voltage without background radiation (left) and in the presence of 600 Hz/cm^2 background (right) at different values of collected integrated charge [152]. The efficiency is stable in time in the absence of the background

radiation, while in the case of presence of background, the efficiency is stable at the working point but we observe a working point shift of 100 V after collecting 260 mC/cm^2 of integrated charge. The working point shift is related to the increase of the electrode's resistance (R) observed at 300 mC/cm^2 of integrated charge as shown in Fig. 3.11. This increase of R causes an increase of the voltage drop (RI) on the effective voltage (HV) applied to the electrodes, and the effective voltage on the gas (HV$_{gas}$) is no longer the same [150,151,153]. The HV$_{gas}$ is defined as:

$$HV_{gas} = HV - RI \tag{3.6}$$

where R is the bakelite resistance and I is the total current.

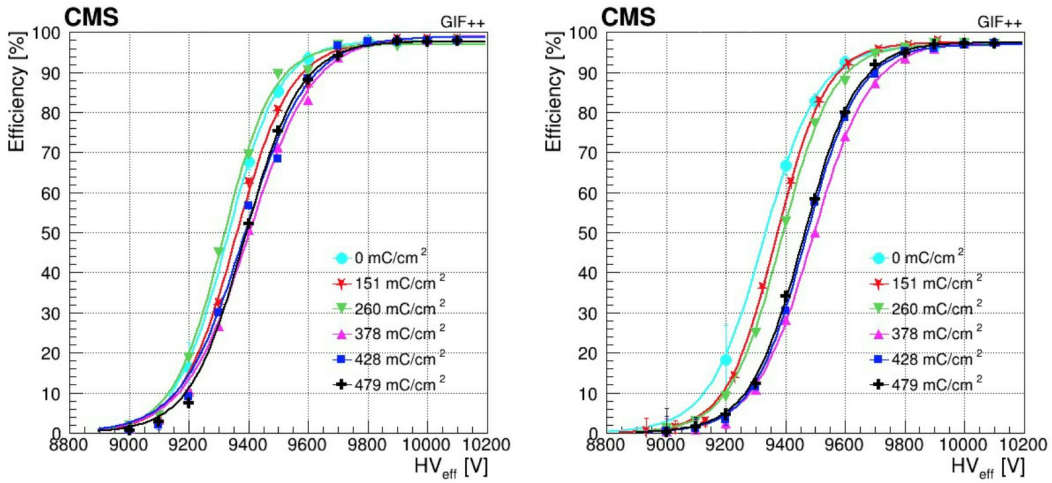


Figure 3.12: RE2/2 irradiated chamber efficiency as a function of the effective HV, with no irradiation (left) and under a gamma background rate of about 600 Hz/cm^2 (right) at different irradiation periods.

The detector operation regime is invariant with respect to HV$_{gas}$, therefore the efficiency as a function of HV$_{gas}$ does not depend anymore on the bakelite resistance as shown in Fig. 3.13 (left) which represents the efficiency at different irradiation periods and different background rates up to 600 Hz/cm^2. All the efficiency curves overlap and we do not observe anymore the working point shift, since the R increase effect on the electrodes has been removed.

The RE2 irradiated chamber efficiency at working point is measured at different background rates (up to 600 Hz/cm^2) and at different integrated charge values as shown in Fig. 3.13 (right). The efficiency is stable in time up to the highest background rate expected at HL-LHC (600 Hz/cm^2).

This result shows no evidence of any aging effect. The main detector parameters and performance are stable. The integrated charge collected up to now represents 89% of the expected integrated charge at High Luminosity Large Hadron Collider. The present accumulated charge in RE2/2 already corresponds to twice the maximum expected integrated

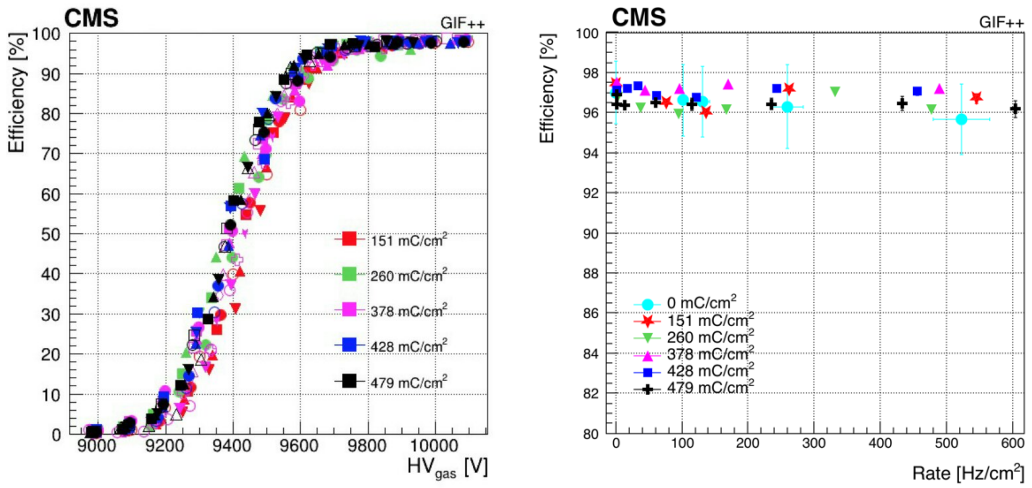


Figure 3.13: RE2/2 irradiated chamber efficiency as a function of the HV$_{gas}$ (left) at different background irradiation rates and different integrated charge values. The RE2 irradiated chamber efficiency at working point as a function of the background rate at different values of collected integrated charge (right) .

charge at HL-LHC, and more irradiation is needed to reach the desired accumulated charge including safety factor three.

This chapter summarized the layout of RPC detector in CMS, the RPCs working principle and design, besides the aging study performed in GIF++ on spare RPC chambers that are currently used in CMS detector. The results show that the main detector parameters and performance are stable and no any evidence of aging effect is observed. After describing the CMS detector and the working principles of its sub-detectors, in the next chapter we will provide a description of how the physics objects are reconstructed from the informations collected by the sub-detectors. Besides, the next chapter includes a description of how the physics processes can be generated and simulated in a monte-carlo simulation program such as the experimental events that happen in real experiments.

4 Event Generators, Simulation and Physics Objects Reconstruction

This Chapter provides the description of the event generation, simulation of the interaction of the particles in the CMS detector and the reconstruction of physics objects used in the analysis.

4.1 Event Generation

The generation of events is the first step of the simulation process chain which uses the monte-carlo (MC) technique to simulate experimental events that happen in real 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 4.1 and are described in the following:

- **Hard Scattering**: Hard processes resulting from the interaction between the constituents of two colliding partons 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 Parton Showers (PS) which is taken into account in the event generation process.

- **Underlying Events**: This is a soft scattering process panying the hard process. This process 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 boson before or after colliding with each other called "Initial State Radiation" (ISR) and "Final State radiation"

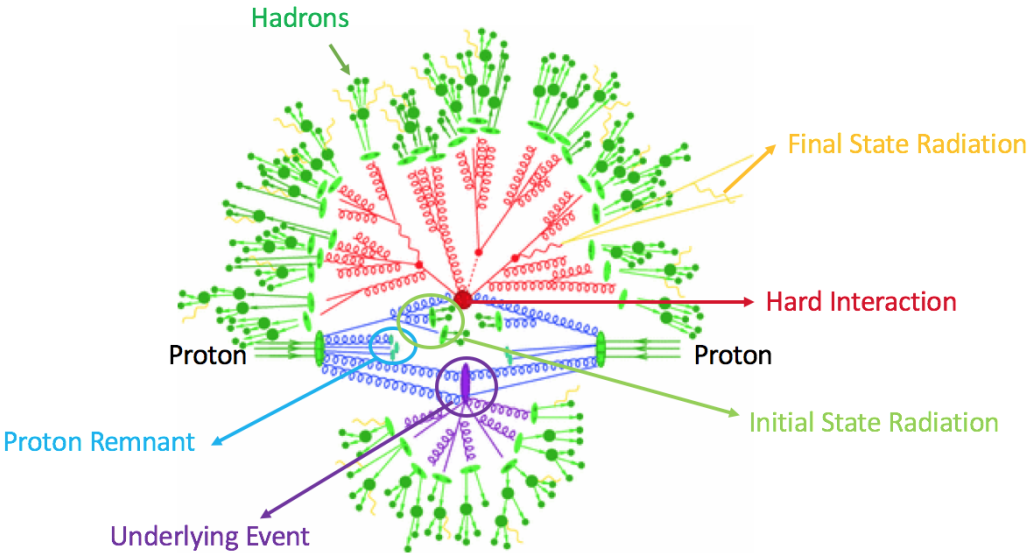


Figure 4.1: Scheme of a proton proton collision.

(FSR) respectively. ISR and FSR result in the creation of jets close to the direction of incoming and outcoming hadrons.

- **Pileup Process**: The LHC accelerates and collides bunches of protons and each bunch contains $\approx 10^{11}$ protons. Pileup refers 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. Figure 4.2 shows the distribution of the average number of interactions per bunch crossing (pileup) in Run II.

- **Hadronization**: After the interaction, the process of combining quarks and gluons forming colorless hadrons that are observed in the final state is known as hadronization process. This process happens when the partons reach hadronization scale of $\approx$ 1 GeV. There are two models describing the transition of a colored partonic system into colorless primary hadrons:

  - The Cluster Model: In this model, gluons are split into $q\bar{q}$ pairs where quarks are clustered together into colorless groups. Clusters formed in this way have large invariant mass which further decay to smaller mass scales suitable to form hadrons [154].

  - The String Model: In this model, 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

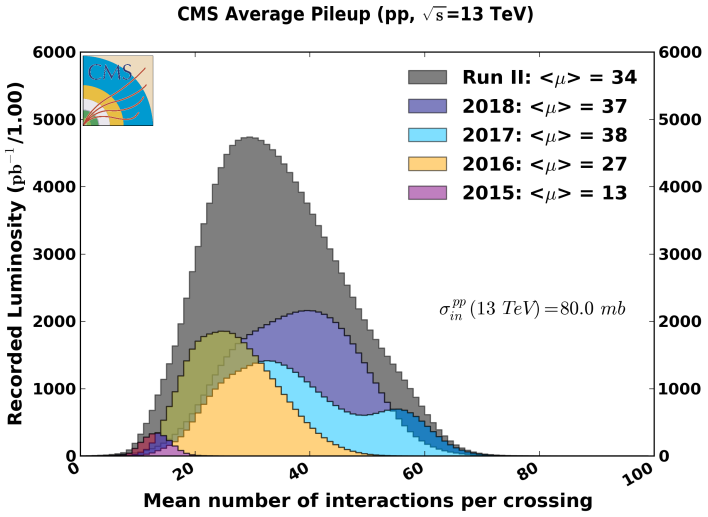


Figure 4.2: Mean number of interactions per bunch crossing for the proton-proton collision in Run II at 13 TeV [11].

$q\bar{q}$ pairs. The process is continued till the energy left to create another such pair is low [155].

The CMS Collaboration uses different MC event generators to generate physics processes for different purposes such as PYTHIA [156], Madgraph [157] and POWHEG [158].

4.2 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 GEANT4 (GEometry ANd Tracking) [159] simulation toolkit. 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".

4.3 The Object reconstruction

The reconstruction process starts with the raw information collected from all sub detectors "such as the hits in the tracker and 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 Particle Flow (PF) algorithm [160]. Before describing the PF algorithm it is important to describe briefly the tracking and clustering processes done in the tracking systems and calorimeters respectively to define tracks and clusters.

4.3.1 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.

4.3.1.1 The Tracking Sequence

The tracking process aims to build tracks from hits inside the silicon detector and hence momentum measurement. CMS uses a Combinatorial Track Finder algorithm (CTF) based on Kalman Filtering (KF) [161, 162] 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 fit χ^2 and the track compatibility to originate from primary vertex [163, 164].

4.3.1.2 The Clustering Sequence

The clustering process done in the calorimeters is used to:

- measure the energy of electrons with their bremsstrahlung photons;
- measure the energy of photons and neutral hadrons;
- measure the energy 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 [165].

4.3.2 Particle Flow Algorithm

The particle flow (PF) algorithm [160] 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 subdectors leaving a signature in a form of hit or energy deposit, the idea of the PF algorithm is to link the information from all subdectors 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 4.3.4;

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 particle flow 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 missing transverse energy will be described in the next sections since they are key ingredients for the analysis.

4.3.3 The Primary Vertex Reconstruction

The aim of the vertex reconstruction [166] 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:

- the selection of the tracks consistent with being produced in 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.
- 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 [16].
- 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.

4.3.4 The Muon Reconstruction

Muons are the only charged particles that can reach to 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 4.3 shows a schematic view for the muon reconstruction algorithms. The muon reconstruction [167–169] 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:

- **Stand-alone 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 the segments are matched together to form seeds. Then

by matchings the seeds with the RPC hits to form the muon track a standalone 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 standalone 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 "standalone 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 5.3.2.

4.3.5 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 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 [29, 165]. 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 x 1 crystals in $\eta \times \phi$ are added around the seed crystal, in a range of N_{steps}

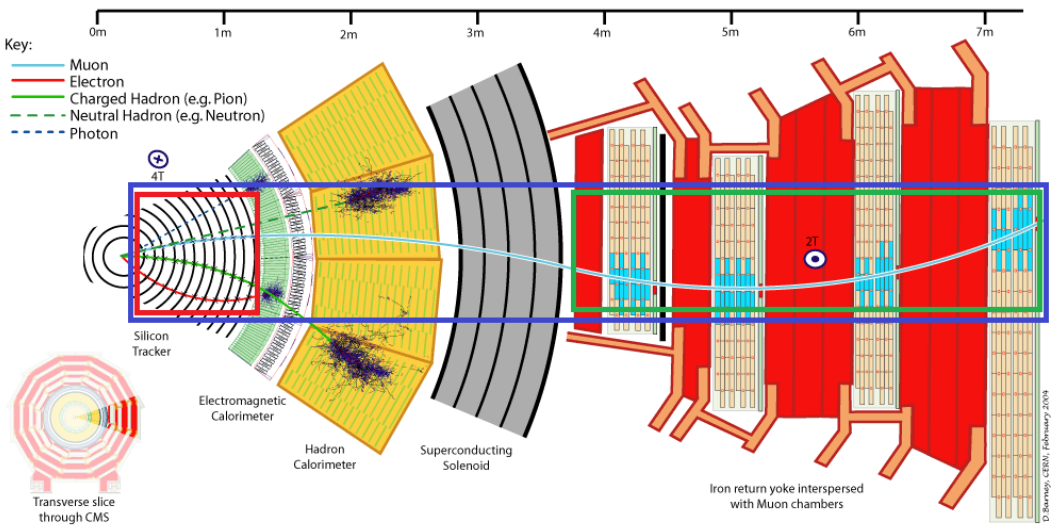


Figure 4.3: Reconstruction of muon objects at the CMS experiment. Tracker track (red box), stand-alone track (green box) and global muon (blue box).

crystals in both directions of ϕ if their energies exceed a minimum threshold as shown in Figure 4.4. 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 fit can be poor. For this reason a dedicated track algorithm is used for electrons track reconstructions. The Gaussian Sum Filter (GSF) tracking [170], 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 fit takes into account the sudden

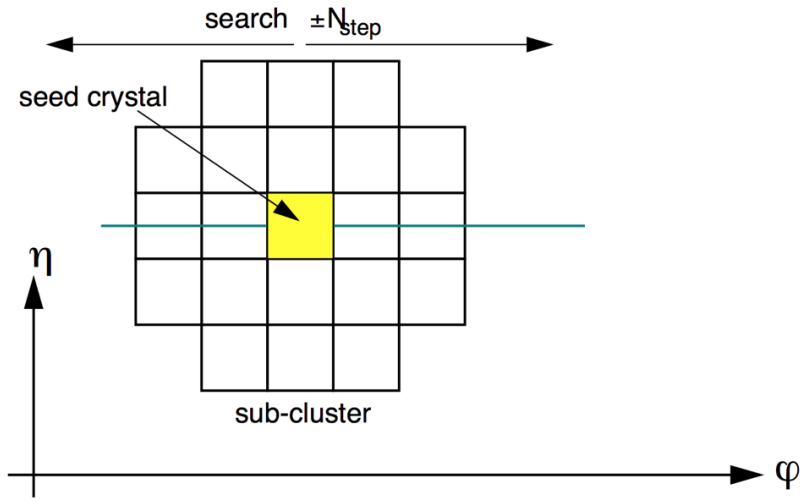


Figure 4.4: The hybrid algorithm with seed crystal shown.

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 5.3.1.

4.3.6 Missing Transverse Energy

All the outgoing particles, "charged and neutral", produced from the collision leave a signature in the detector's material except neutrinos "or hypothetical neutral weakly interacting particles" which leave the detector without interacting. Their presence can be inferred from the momentum imbalance in the transverse plane (plan perpendicular to the beam direction), this quantity known as missing transverse energy [171, 172] and denoted by E_T^{miss} or MET. The MET is not only playing an important role for the SM physics such as the measurements of W boson, top quark, and tau lepton decays where the MET is interpreted as originated by neutrinos but it is a key variable in many searches for physics beyond the standard model such dark matter search and search for supersymmetric particles where the MET is interpreted as new particle due to the presence of hypothetical neutral weakly interacting particles. The MET is defined as the magnitude of the negative vector sum of the transverse momentum of all PF reconstructed particles in an event as shown in equation 4.1. This definition of MET is called "Raw MET".

$$\overrightarrow{\not{E}}_T^{raw} = -\sum_{i\epsilon\, all} \overrightarrow{p}_{T,i} \tag{4.1}$$

where the index i runs over all particle flow candidates.

The raw MET is different from the true MET, i.e., the transverse momentum carried by invisible particles, for many reasons including the non-compensating nature of the calorime-

ters and detector misalignment. To make the MET a better estimate of the true MET, some corrections need to be applied. In addition a set of filters has been derived by CMS to remove some of detector noise, cosmic rays, and beam-halo particles which can cause a fake MET. The MET used in the analysis with the applied correction and filters will be discussed in section 5.3.4.

This chapter provided the basic information about the algorithms used to reconstruct the physics objects for the thesis, besides an overview of how physics events are generated and simulated in the detector. The next chapter will describe all the building blocks required to perform the analysis.

5 Building Blocks for Mono-Higgs Analysis

This chapter describes the building blocks of the monoHiggs analysis: the list of data sets used and of the simulated signal and background samples beside the object selection used to perform the analysis.

5.1 Data-Sets and Triggers

The analysis presented in the thesis uses the data sets recorded by the CMS experiment during run II (2016, 2017, 2018) at $\sqrt{s} = 13$ TeV. All the used data 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 Run II correspond to a total integrated luminosity of 137 fb^{-1}. Table 5.1 lists the integrated luminosity recorded in each year of data taken and the corresponding files which contain the certified runs [173].

The analysis used five primary datasets[1] (DoubleMuon - DoubleEG - MuonEG - SingleElectron - SingleMuon) in 2016 and 2017 and four primary datasets (DoubleMuon - EGamma - MuonEG - SingleMuon) in 2018 where DoubelEG and SingleElectron datasets are merged together in one dataset (EGamma). The data sets are collected by passing single-lepton, dilepton, or trilepton triggers (as noticed in the name of the dataset, as example DoubleMuon (DoubleEG) data set means that this set contains the data passing di-muon (di-electron) or tri-muon (tri-electron) filters with specific p_T thresholds and loose lepton identification and isolation requirements). The High Level Trigger is used to decide whether an event should

Table 5.1: Integrated luminosity recorded in each year of data taken and the certified files.

Year	L (fb^{-1})	ReReco	Json file
2016	35.9	17Jul2018	Cert_271036-284044_13TeV_ReReco_07Aug2017_Collisions16_JSON.txt
2017	41.5	31Mar2018	Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON.txt
2018	59.7	17Sep2018	Cert_314472-25175_13TeV_PromptReco_Collisions18_JSON.txt

[1]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 ,....), tracks, calorimeter clusters, vertices and identification criteria for the physics objects.

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 (as described in section 2.11.1). The list of data sets used in the 3 years are reported in Table 5.2. The data used in the 3 years 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 "promptReco" workflow with preliminary detector calibrations).

In this analysis we rely on multi-lepton HLT paths (single, Di and tri-leptons HLT paths) as listed in Table 5.3[2] for each year. The events are required to pass at least one of those HLT paths. To avoid the duplication of events from different primary datasets, events are taken in the following sequence:

- from DoubleEG dataset if they pass the diElectron or triElectron trigger paths in 2016 and 2017 analysis while in 2018 the events are taken from EGamma if they pass the diElectron or triElectron or singleElectron triggers;
- from DoubleMuon dataset if they pass the diMuon or triMuon trigger paths and fail the above triggers;
- from MuonEG dataset if they pass the MuEle or MuDiEle or DiMuEle trigger paths and fail all the above triggers;
- from SingleElectron dataset if they pass the singleElectron trigger path and fail all the above triggers in case of 2016 & 2017 data while in 2018 the SingleElectron trigger is already used in EGamma dataset in first step;
- finally from SingleMuon dataset if they pass the singleMuon trigger paths and fail all the above triggers;

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 [21].

[2]For an event to pass the HLT-Mu17-TrkIsoVVL-Mu8-TrkIsoVVL-DZ-Mass3p8, as example, it is required that the event has at least two muons with p_T threshold of 17 and 8 GeV passing very loose tracker isolation and transverse impact parameter condition. The definition of isolation and impact parameter will be introduced in 5.3

[3]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.

Table 5.2: Datasets used in the analysis for full Run II.

DataSets		
2016	2017	2018
/DoubleMuon/Run2016B-17Jul2018_ver2_v1/ /DoubleEG/Run2016B-17Jul2018_ver2_v1/ /MuonEG/Run2016B-17Jul2018_ver2_v1/ /SingleElectron/Run2016B-17Jul2018_ver2_v1/ /SingleMuon/Run2016B-17Jul2018_ver2_v1/	/DoubleMuon/Run2017B-31Mar2018_v1/ /DoubleEG/Run2017B-31Mar2018_v1/ /MuonEG/Run2017B-31Mar2018_v1/ /SingleElectron/Run2017B-31Mar2018_v1/ /SingleMuon/Run2017B-31Mar2018_v1/	/DoubleMuon/Run2018A-17Sep2018_v2/ /MuonEG/Run2018A-17Sep2018_v1/ /SingleMuon/Run2018A-17Sep2018_v2/ /EGamma/Run2018A-17Sep2018_v2/
/DoubleMuon/Run2016C-17Jul2018_v1/ /DoubleEG/Run2016C-17Jul2018_v1/ /MuonEG/Run2016C-17Jul2018_v1/ /SingleElectron/Run2016C-17Jul2018_v1/ /SingleMuon/Run2016C-17Jul2018_v1/	/DoubleMuon/Run2017C-31Mar2018_v1/ /DoubleEG/Run2017C-31Mar2018_v1/ /MuonEG/Run2017C-31Mar2018_v1/ /SingleElectron/Run2017C-31Mar2018_v1/ /SingleMuon/Run2017C-31Mar2018_v1/	/DoubleMuon/Run2018B-17Sep2018_v1/ /MuonEG/Run2018B-17Sep2018_v1/ /SingleMuon/Run2018B-17Sep2018_v1/ /EGamma/Run2018B-17Sep2018_v1/
/DoubleMuon/Run2016D-17Jul2018_v1/ /DoubleEG/Run2016D-17Jul2018_v1/ /MuonEG/Run2016D-17Jul2018_v1/ /SingleElectron/Run2016D-17Jul2018_v1/ /SingleMuon/Run2016D-17Jul2018_v1/	/DoubleMuon/Run2017D-31Mar2018_v1/ /DoubleEG/Run2017D-31Mar2018_v1/ /MuonEG/Run2017D-31Mar2018_v1/ /SingleElectron/Run2017D-31Mar2018_v1/ /SingleMuon/Run2017D-31Mar2018_v1/	/DoubleMuon/Run2018C-17Sep2018_v1/ /MuonEG/Run2018C-17Sep2018_v1/ /SingleMuon/Run2018C-17Sep2018_v1/ /EGamma/Run2018C-17Sep2018_v1/
/DoubleMuon/Run2016E-17Jul2018_v1/ /DoubleEG/Run2016E-17Jul2018_v1/ /MuonEG/Run2016E-17Jul2018_v1/ /SingleElectron/Run2016E-17Jul2018_v1/ /SingleMuon/Run2016E-17Jul2018_v1/	/DoubleMuon/Run2017E-31Mar2018_v1/ /DoubleEG/Run2017E-31Mar2018_v1/ /MuonEG/Run2017E-31Mar2018_v1/ /SingleElectron/Run2017E-31Mar2018_v1/ /SingleMuon/Run2017E-31Mar2018_v1/	/DoubleMuon/Run2018D-PromptReco_v2/ /MuonEG/Run2018D-PromptReco_v2/ /SingleMuon/Run2018D-PromptReco_v2/ /EGamma/Run2018D-PromptReco_v2/
/DoubleMuon/Run2016F-17Jul2018_v1/ /DoubleEG/Run2016F-17Jul2018_v1/ /MuonEG/Run2016F-17Jul2018_v1/ /SingleElectron/Run2016F-17Jul2018_v1/ /SingleMuon/Run2016F-17Jul2018_v1	/DoubleMuon/Run2017F-31Mar2018_v1/ /DoubleEG/Run2017F-31Mar2018_v1/ /MuonEG/Run2017F-31Mar2018_v1/ /SingleElectron/Run2017F-31Mar2018_v1/ /SingleMuon/Run2017F-31Mar2018_v1/	
/DoubleMuon/Run2016G-17Jul2018_v1/ /DoubleEG/Run2016G-17Jul2018_v1/ /MuonEG/Run2016G-17Jul2018_v1/ /SingleElectron/Run2016G-17Jul2018_v1/ /SingleMuon/Run2016G-17Jul2018_v1/		
/DoubleMuon/Run2016H-17Jul2018_v1/ /DoubleEG/Run2016H-17Jul2018_v1/ /MuonEG/Run2016H-17Jul2018_v1/ /SingleElectron/Run2016H-17Jul2018_v1/ /SingleMuon/Run2016H-17Jul2018_v1/		

*All datasets used are MINIAOD[3]

Table 5.3: Trigger paths used in the analysis for full full Run II.

Trigger paths		
2016	2017	2018
HLT-Ele17-Ele12-CaloIdL-TrackIdL-IsoVL-DZ	HLT-Ele23-Ele12-CaloIdL-TrackIdL-IsoVL	HLT-Ele23-Ele12-CaloIdL-TrackIdL-IsoVL-v
HLT-Ele23-Ele12-CaloIdL-TrackIdL-IsoVL-DZ	HLT-DoubleEle33-CaloIdL-GsfTrkIdVL	HLT-DoubleEle25-CaloIdL-MW-v
HLT-DoubleEle33-CaloIdL-GsfTrkIdVL	HLT-Ele16-Ele12-Ele8-CaloIdL-TrackIdL	HLT-Ele32-WPTight-Gsf-v
HLT-Ele16-Ele12-Ele8-CaloIdL-TrackIdL	HLT-Mu17-TrkIsoVVL-Mu8-TrkIsoVVL-DZ-Mass3p8	HLT-Mu17-TrkIsoVVL-Mu8-TrkIsoVVL-DZ Mass3p8-v
HLT-Mu17-TrkIsoVVL-Mu8-TrkIsoVVL	HLT-Mu17-TrkIsoVVL-Mu8-TrkIsoVVL-DZ-Mass8	HLT-TripleMu-10-5-5-DZ-v
HLT-TripleMu-12-10-5	HLT-TripleMu-12-10-5	HLT-TripleMu-12-10-5-v
HLT-Mu8-TrkIsoVVL-Ele17-CaloIdL-TrackIdL-IsoVL	HLT-TripleMu-10-5-5-D2	HLT-Mu23-TrkIsoVVL-Ele12-CaloIdL-TrackIdL-IsoVL-v
HLT-Mu8-TrkIsoVVL-Ele23-CaloIdL-TrackIdL-IsoVL	HLT-Mu23-TrkIsoVVL-Ele12-CaloIdL-TrackIdL-IsoVL	HLT-Mu8-TrkIsoVVL-Ele23-CaloIdL-TrackIdL-IsoVL-DZ-v
HLT-Mu17-TrkIsoVVL-Ele12-CaloIdL-TrackIdL-IsoVL	HLT-Mu8-TrkIsoVVL-Ele23-CaloIdL-TrackIdL-IsoVL-DZ	HLT-Mu12-TrkIsoVVL-Ele23-CaloIdL-TrackIdL-IsoVL-DZ-v
HLT-Mu23-TrkIsoVVL-Ele12-CaloIdL-TrackIdL-IsoVL	HLT-Mu12-TrkIsoVVL-Ele23-CaloIdL-TrackIdL-IsoVL-DZ	HLT-Mu23-TrkIsoVVL-Ele12-CaloIdL-TrackIdL-IsoVL-DZ-v
HLT-Mu23-TrkIsoVVL-Ele8-CaloIdL-TrackIdL-IsoVL	HLT-Mu23-TrkIsoVVL-Ele12-CaloIdL-TrackIdL-IsoVL-DZ	HLT-DiMu9-Ele9-CaloIdL-TrackIdL-DZ-v
HLT-Mu8-DiEle12-CaloIdL-TrackIdL	HLT-DiMu9-Ele9-CaloIdL-TrackIdL-DZ	HLT-Mu8-DiEle12-CaloIdL-TrackIdL-DZ-v
HLT-DiMu9-Ele9-CaloIdL-TrackIdL	HLT-Mu8-DiEle12-CaloIdL-TrackIdL	HLT-IsoMu24-v
HLT-Ele25-eta2p1-WPTight	HLT-Mu8-DiEle12-CaloIdL-TrackIdL-DZ	
HLT-Ele27-WPTight	HLT-Ele35-WPTight-Gsf-v	
HLT-Ele27-eta2p1-WPLoose-Gsf	HLT-Ele38-WPTight-Gsf-v	
HLT-IsoMu20 OR HLT-IsoTkMu20	HLT-Ele40-WPTight-Gsf-v	
HLT-IsoMu22 OR HLT-IsoTkMu22	HLT-IsoMu27	

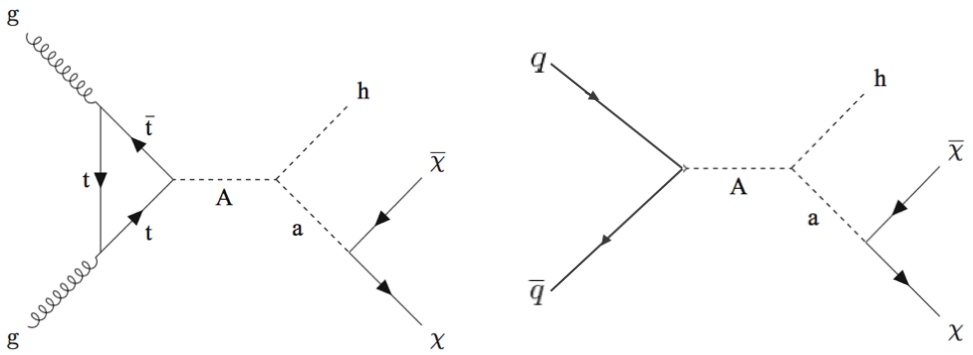


Figure 5.1: Feynman diagrams for the 2HDM+a for gluon-gluon fusion (left) and $q\bar{q}$ annihilation (right).

5.2 The Simulated Samples

The Monte Carlo (MC) simulation plays an important role in every analysis since it is used to simulate the signal and background processes. Also, it is used in various steps in the analysis starting from the event selection optimization, the estimation of the background rates, the comparisons with the real measurements from data, computation of the expected yields up to the production of final results. This section describes the signal and the background samples used in the analysis.

5.2.1 Signal Samples

The Two Higgs doublet model extended by light pseudoscalar a, "2HDM+a", is the first time inspected in the Higgs to ZZ to 4 leptons decay channel in CMS in this thesis. Although the Higgs decay to ZZ channel suffers from a low branching ratio, but it has the advantage of clean final state with relatively small background contamination. Therefore, this Higgs decay channel can provide a contribution in the sensitivity of the Mono-Higgs search when combined with all the possible Higgs boson decay channels. The model is schematically represented by the Feynman diagram in Figure 5.1. Here a heavy pseudoscalar A is produced from proton proton collision by gluon gluon fusion or quark-antiquark annihilation and subsequently decays to SM Higgs boson and a light pseudoscalar a, which on turn decays to a pair of DM particles χ. The resonant decay of the heavy mediator A affects the kinematics associated with the model which can be different from SM processes such as Higgs to ZZ to 4 leptons as shown in Figure 5.2. Moreover, in the heavy mass mediator the missing transverse energy has harder spectrum.

In this search different model parameters have been scanned over while other parameters are fixed. The following parameters have been scanned:

- the heavy pseudoscalar mass m_A;

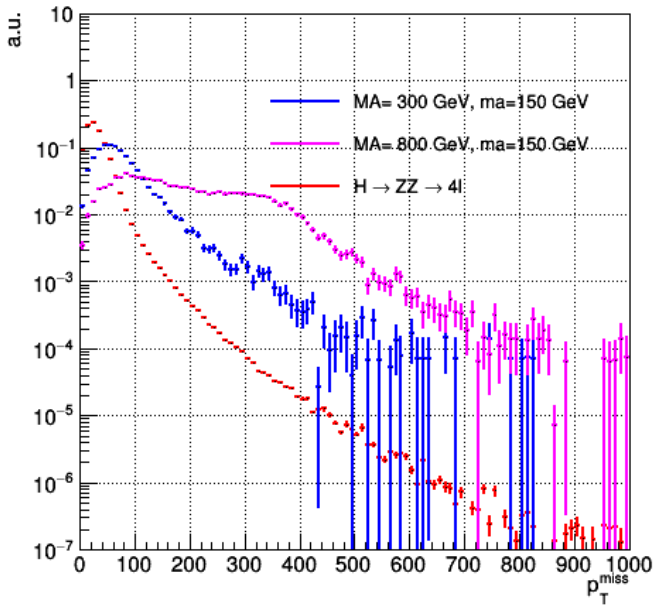


Figure 5.2: Comparison of the p_T^{miss} distribution of low and high mass of heavy pseudoscalar mediator A for the signal model and Higgs to ZZ background.

- the light pseudoscalar mass m_a;
- the ratio of the vacuum expectation values of the light and heavy scalar h and H: $tan\beta$;
- the angle $sin\theta$ associated with the mixing between the light and heavy pseudoscalar a and A;

To study the dependence of the kinematics on the model parameters, different signal samples have been produced using MADGRAPH5_AMC@NLO [157] generator and PYTHIA8 [156] for parton shower. Tables 5.4, 5.5 and 5.6 summarize the different scans considered in the thesis and the corresponding cross-section values. The missing energy distributions for different signal parameters are shown in Figures 5.3, 5.4. Figure 5.3 shows a kinematic dependence on the mass of the heavy pseudoscalar A (left) and on the mass of the light pseudoscalar a (right). The heavier the mass of the pseudoscalar, the harder the missing energy spectrum. While Figure 5.4 (left) shows no kinematic dependence on the choice of the $tan\beta$ value, only the production cross section scales as a function of $tan\beta$. Similarly, the value of the $sin\theta$ doesn't affect the kinematic distributions, but only the production cross sections, as shown in Figure 5.4 (right). No significant change in the kinematics is observed by varying the mass of the DM m_χ as shown in Figure 5.5, so the mass of the DM particle m_χ is fixed to 10 GeV in this search. Also the mass of the charged Higgs and heavy scalar Higgs are assumed equal to the mass of the heavy pseudoscalar $m_H = m_{H^\pm} = m_A$. The search in the context of the thesis has been performed using the scans of the heavy and light pseudoscalar masses with low values since the production cross section is very suppressed with the increase

Table 5.4: Cross section values in pb for different mass points of light and heavy pseudoscalars considering $tan\beta$ =1 and $sin\theta = 0.35$. The branching ratio of $H \to ZZ \to 4l$ is included in the calculation.

			M_A [GeV]				
M_a [GeV]	200	300	400	500	600	700	800
gg fusion							
150	8.7×10^{-5}	4.3×10^{-4}	2.6×10^{-4}	1.3×10^{-4}	7.9×10^{-5}	-	3.1×10^{-5}
250	-	3.5×10^{-5}	6.1×10^{-5}	6.3×10^{-5}	4.4×10^{-5}	2.9×10^{-5}	1.9×10^{-5}
350	-	-	9.4×10^{-6}	1.4×10^{-5}	2.0×10^{-5}	1.8×10^{-5}	1.3×10^{-5}
$q\bar{q}$ annihilation							
150	6.47×10^{-7}	8.28×10^{-7}	2.5×10^{-7}	2.0×10^{-7}	1.82×10^{-7}	-	1.6×10^{-7}
250	-	2.6×10^{-7}	7.95×10^{-8}	7.5×10^{-8}	6.78×10^{-8}	6.2×10^{-8}	5.8×10^{-8}
350	-	-	2.82×10^{-8}	2.48×10^{-8}	2.6×10^{-8}	2.4×10^{-8}	2.3×10^{-8}

Table 5.5: Summary of cross sections in pb for different $tan\beta$ values vs light pseudoscalar mass considering $M_A = 300$ GeV and $sin\theta = 0.35$. The branching ratio of $H \to ZZ \to 4l$ is included in the calculation.

			$tan\beta$		
M_a [GeV]	0.5	1.0	2.0	4.0	8.0
gg fusion					
150	1.7×10^{-3}	4.3×10^{-4}	-	1.8×10^{-5}	2.3×10^{-6}
250	1.4×10^{-4}	3.5×10^{-5}	8.2×10^{-6}	1.6×10^{-6}	2.7×10^{-7}
$q\bar{q}$ annihilation					
150	2.0×10^{-7}	8.28×10^{-7}	-	1.2×10^{-5}	3.9×10^{-5}
250	6.5×10^{-8}	2.6×10^{-7}	1.0×10^{-6}	3.8×10^{-6}	1.2×10^{-5}

of the heavy and light pseudoscalar masses, as summarized in table 5.4. Also the analysis includes the $tan\beta$ scan with two scenarios of the mass of the light pseudoscalar and two scenarios of $sin\theta$, as summarized in tables 5.5 and 5.6.

5.2.2 The Background Samples

Backgrounds are all processes which can mimic the signal process under the study. All the physics processes which can give the same signature as the Mono-Higgs signal process (four leptons plus Missing Energy) are considered as background. We can have four leptons in the final state which could be real leptons coming from decay of particles like ZZ or fake reconstructed objects plus missing transverse energy that could be real object "such as neutrinos or hypothetical dark matter particle" or fake object coming from uninteresting source from the detector. All the background samples are officially produced in CMS and listed below:

Table 5.6: Summary of cross sections in pb for different $tan\beta$ values vs light pseudoscalar mass considering $M_A = 300$ GeV and $sin\theta = 0.7$. The branching ratio of $H \to ZZ \to 4l$ is included in the calculation.

			$tan\beta$			
M_a [GeV]	0.5	1.0	1.5	2.0	4.0	8.0
			gg fusion			
150	6.9×10^{-4}	1.7×10^{-4}	7.4×10^{-5}	4.0×10^{-5}	7.8×10^{-6}	1.4×10^{-6}
			$q\bar{q}$ annihilation			
150	4.0×10^{-7}	1.6×10^{-6}	3.6×10^{-6}	6.4×10^{-6}	2.5×10^{-5}	9.3×10^{-5}

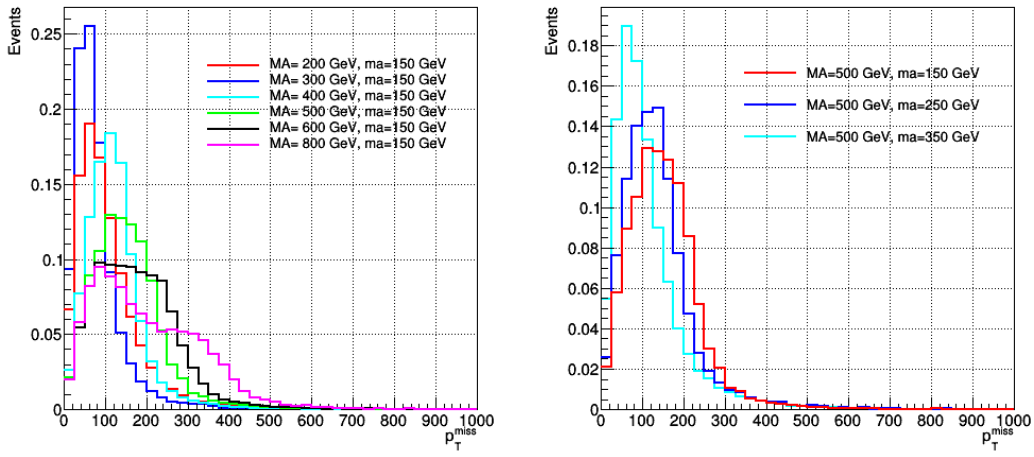


Figure 5.3: p_T^{miss} distributions for signal samples with varying the mass of the heavy pseudoscalar A (left) and the light pseudoscalar a (right).

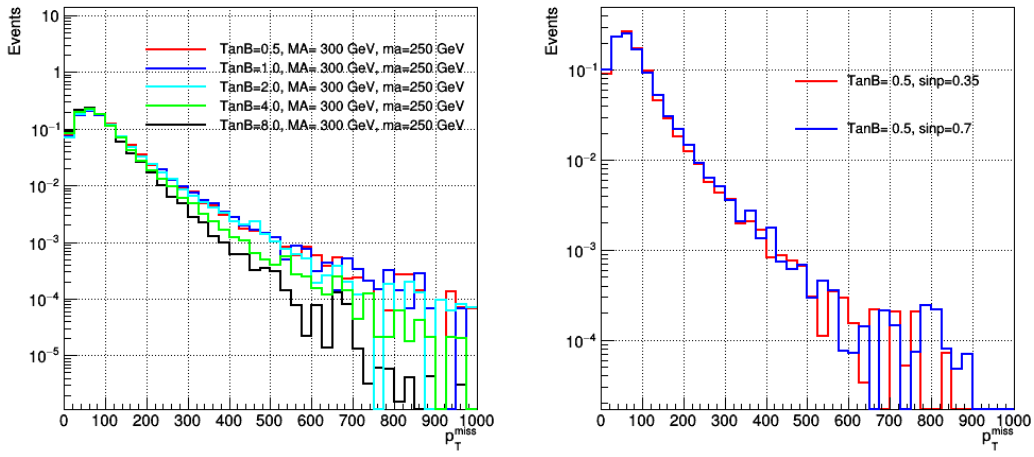


Figure 5.4: p_T^{miss} distributions for signal samples with varying the $tan\beta$ value (left) and the $sin\theta$ value (right).

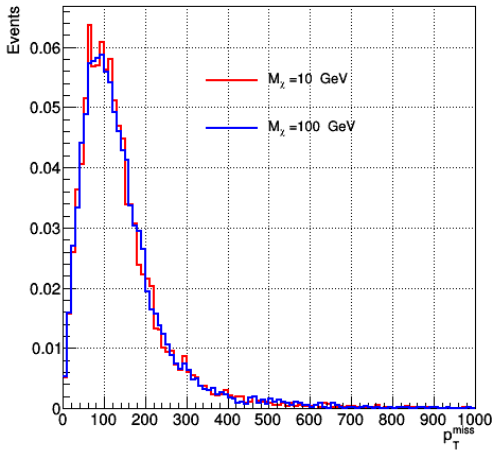


Figure 5.5: p_T^{miss} distributions for signal samples with varying the mass of the DM particle.

- **SM Higgs background**:

  The SM Higgs process is one of the major backgrounds "irreducible" in the analysis considered in the thesis, in particular when the Higgs is produced in association with a Z boson decaying to neutrinos; in this case the undetectable neutrinos mimic the presence of dark matter particles, leading to a kinematics very close to the signal one. This background will survive up to the final step of the analysis. We consider two Higgs decays, the Higgs decay to ZZ which decay to 4l and Higgs decay to WW which decay to $2l2\nu$.

  The SM Higgs simulated samples are produced at NLO accuracy using the POWHEG 2.0 [174–176] generator for the five main production modes (gluon gluon fusion [177] - Vector Boson fusion [178] - associated production of Higgs with W, Z and $t\bar{t}$ [179]) shown in Figure 5.6. For the associated production of Higgs with W and Z the MINLO HVJ [180] extension of POWHEG 2.0 is used for preserving more accuracy in the production of hard jet. All the samples are generated with Higgs mass of 125 GeV. The Higgs boson decay to four leptons is handled by using the JHUGEN[4] generator [181,182] which properly accounts for interference effects associated with permutations of identical leptons in the 4e, 4μ and 4τ final states. In the case of associated production of Higgs with W, Z and $t\bar{t}H$ processes, the Higgs boson decay to $H \to ZZ \to 2l2X$ is allowed such that two leptons in the final state can come from the decay of associated Z, W bosons, or top quarks. On the other hand, the decay of Higgs boson to $2l2\nu$ is done via PYTHIA 8.1 except for vector-boson-fusion (VBF) done via JHUGen v5.2. All the SM Higgs samples with their corresponding cross sections [183, 184] are listed in Table 5.7.

[4]The JHU generator is a model-independent generator for studying spin-parity properties of resonances such as the Higgs boson.

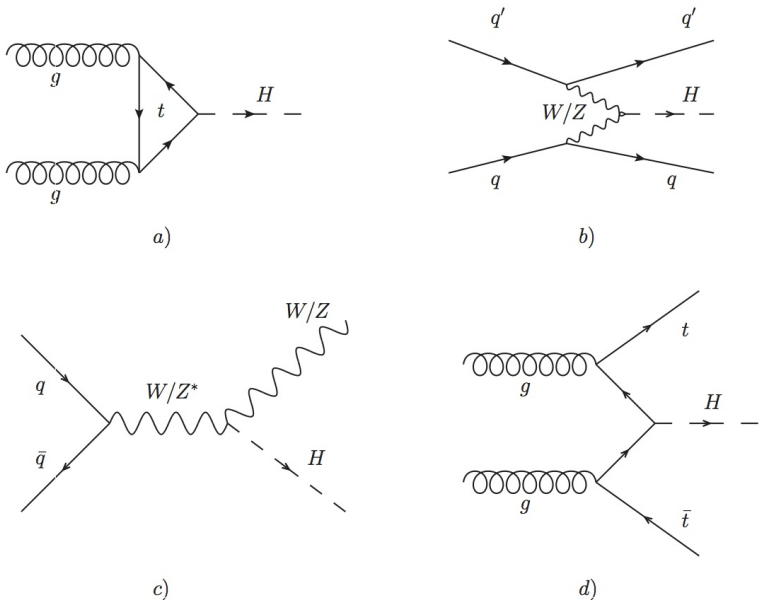


Figure 5.6: Feynam diagrams representing the five SM Higgs production modes at LHC: (a) gluon-gluon fusion; (b) vector boson fusion; (c) W and Z associated production; (d) $t\bar{t}$ associated production.

- **Non-resonant ZZ background**:

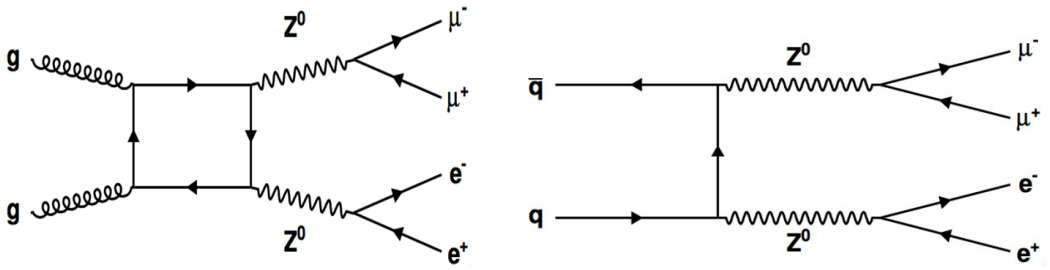
  The non-resonant ZZ background represents an important source of reducible background for the analysis. The ZZ background sample coming from $q\bar{q}$ annihilation is generated at NLO with POWHEG 2.0 and PYTHIA8 (used for the parton shower and hadronization) while the ZZ process coming from the gluon gluon fusion is generated at LO with MCFM [185] as shown in Figure 5.7. The ZZ background samples are listed with their corresponding cross sections in Table 5.7 [186, 187].

- **Additional background**:

  Additional background samples such as di-bosons (WZ, WW), Drell-Yan+jets, W+jets, $t\bar{t}$, $t\bar{t}V(V = W, Z)$ and tri-bosons (ZZZ, WWZ, ZZW) have small contribution in the analysis. The production of those backgrounds at LHC is shown in Figure 5.8. The samples are generated using either POWHEG 2.0 or MADGRAPH5_AMC@NLO [157]. Apart from tri-bosons and $t\bar{t}V$, the other backgrounds will be estimated from data as will be explained in section 6.4 but they are used to validate the method and also used in the early step of the analysis. Those background samples are listed in Table 5.7 with their corresponding cross sections.

5.2.3 Event 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



(a) (b)

Figure 5.7: Feynman diagrams for ZZ background production from (a) $gg \to ZZ$ process and from (b) $q\bar{q} \to ZZ$ process.

parameters such as the beam conditions and the instantaneous luminosity. The pile-up distribution in data and MC is used to compute pileup weights to weight the MC events and hence match the pile up distribution in data. Figure 5.9 shows the distribution of the number of the pileup interactions in 2016 (left), 2017 (middle) and 2018 (right) 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 data the integrated luminosity (L). The MC event is weighted by a factor (w) given by

$$w = \frac{\sigma.L}{N} \tag{5.1}$$

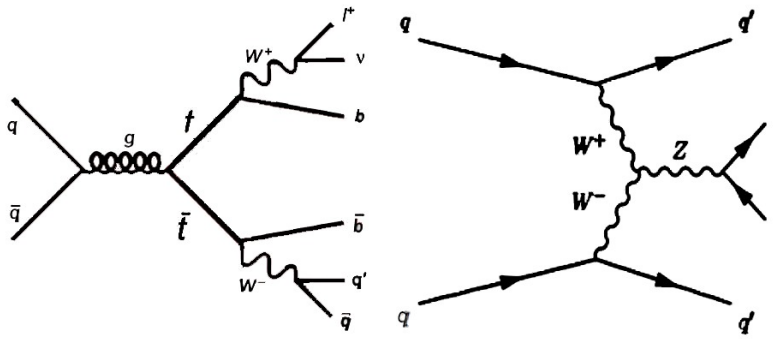
where N represents the total number of MC simulated events for the corresponding process.

5.3 The Object selection for the analysis

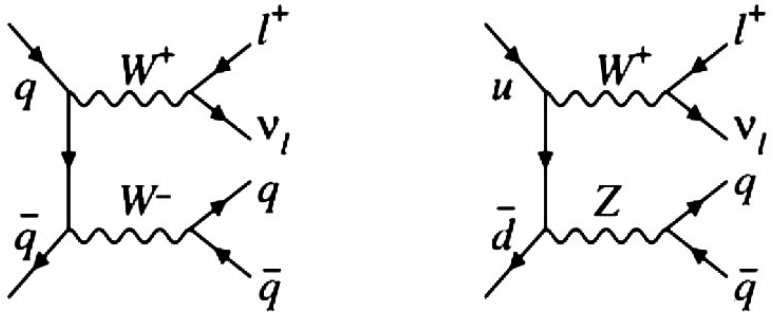
The analysis relies on the following physics objects: muons, electrons, photons and missing energy. A selection criteria is applied to those reconstructed objects to identify a set of selected objects for the analysis. The description of the main selection criteria is reported in the following sections.

5.3.1 Electron selection

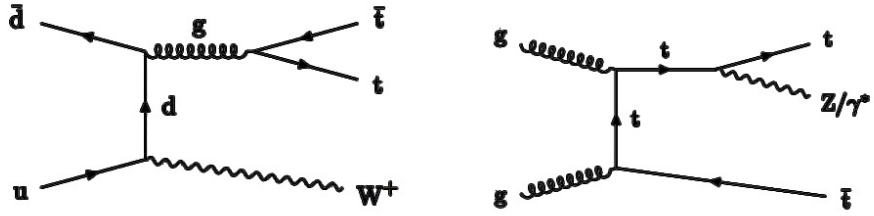
After the reconstruction process of the electrons, we end up with set of electrons that could be real electrons or fake objects that pass the electron reconstruction process and hence are considered as electron candidates. The electron reconstruction efficiency [22] is very high which means that we only miss a small fraction of real electrons but we could have an amount of fake electrons. Therefore, a set of selection cuts are applied to the reconstructed electrons



(a)



(b)



(c)

Figure 5.8: Feynam diagrams for backgrounds production at LHC: (a) $t\bar{t}$ production in left and $Z + Jets$ in right (b) Di-Boson production: WW in left and WZ in right (c) $t\bar{t}V$ production: $t\bar{t}W$ in left and $t\bar{t}Z$ in right.

Table 5.7: Background Monte Carlo samples and cross sections.

Process	Dataset Name	$\sigma\times$ BR
Higgs boson $H \to ZZ^* \to 4l$		
$gg \to H \to ZZ \to 4l$	/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUgenV709_pythia8/[1]	12.18 fb
$q\bar{q} \to Hqq \to ZZqq \to 4lqq$	/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
$gg \to ttH \to ttZZ \to 4l + X$	/ttH_HToZZ_4LFilter_M125_13TeV_powheg_JHUgenV709_pythia8/[1]	0.393 fb
Higgs boson $H \to WW \to 2l2\nu$		
$gg \to H \to WW \to 2l2\nu$	/GluGluHToWWTo2L2Nu_M125_13TeV_powheg_JHUgen_pythia8/[1]	1101.790 fb
$q\bar{q} \to Hqq \to WWqq \to 2l2\nu qq$	/VBFHToWWTo2L2Nu_M125_13TeV_powheg_JHUgen_pythia8/[1]	85.776 fb
$q\bar{q} \to W^+H \to W^+WW \to l\nu 2l2\nu$	/HWplusJ_HToWWTo2L2Nu_WToLNu_M125_13TeV_powheg_pythia8/ [1]	2.138 fb
$q\bar{q} \to W^-H \to W^-WW \to l\nu 2l2\nu$	/HWminusJ_HToWWTo2L2Nu_WToLNu_M125_13TeV_powheg_pythia8/ [1]	1.357 fb
$q\bar{q} \to ZH \to ZWW \to 2l2l2\nu$	/HZJ_HToWWTo2L2Nu_ZTo2L_M125_13TeV_powheg_pythia8/[1]	2.029 fb
ZZ background samples		
$qq \to ZZ \to 4l$	/ZZTo4L_13TeV_powheg_pythia8/[1]	1.256 pb
$qq \to ZZ \to 2l2\nu$	/ZZTo2L2ν_13TeV_powheg_pythia8/[1]	0.564 pb
$qq \to ZZ \to 4e$	/GluGluToContinToZZTo4e_13TeV_MCFM701/[1]	0.00159 pb
$qq \to ZZ \to 4\mu$	/GluGluToContinToZZTo4μ_13TeV_MCFM701/[1]	0.00159 pb
$gg \to ZZ \to 4\tau$	/GluGluToContinToZZTo4τ_13TeV_MCFM701/[1]	0.00159 pb
$gg \to ZZ \to 2e2\mu$	/GluGluToContinToZZTo2e2μ_13TeV_MCFM701/[1]	0.00159 pb
$gg \to ZZ \to 2e2\tau$	/GluGluToContinToZZTo2e2τ_13TeV_MCFM701/[1]	0.00159 pb
$gg \to ZZ \to 2\mu 2\tau$	/GluGluToContinToZZTo2μ2τ_13TeV_MCFM701/[1]	0.00159 pb
Other backgrounds		
$Z \to ll + jets$	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/[1]	6104 pb
$Z \to ll + jets$	/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/[1]	18610 pb
$WZ \to 3l\nu$	/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8/[1]	4.430 pb
$t\bar{t} \to 2l2\nu b$	/TTTo2L2Nu_13TeV-powheg/[1]	87.31pb
$WW \to 2l2\nu$	/WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8/[1]	12.178 pb
$W + jets \to 2l2\nu$	/WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/[1]	61526.7 pb
ZZZ	/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[1]	0.01398 pb
WZZ	/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[1]	0.05565 pb
WWZ	/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[1]	0.1651 pb
$t\bar{t}Z$	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/[1]	0.2529 pb
$t\bar{t}W$	/TTWJetsToLNu_TuneCUETP8M113TeV _amcatnloFXFX _madspin _pythia8/[1]	0.2043 pb
$ST\_t-channel\_top$	/ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8/[1]	44.33pb
$ST\_t-channel\_antitop$	/ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8/[1]	26.38 pb
ST_tW_top	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[1]	35.85 pb
$ST\_tW\_antitop$	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[1]	35.85 pb

[1]= RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v for 2016,

[1]= RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v for 2017,

[1]= RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v1 for 2018.

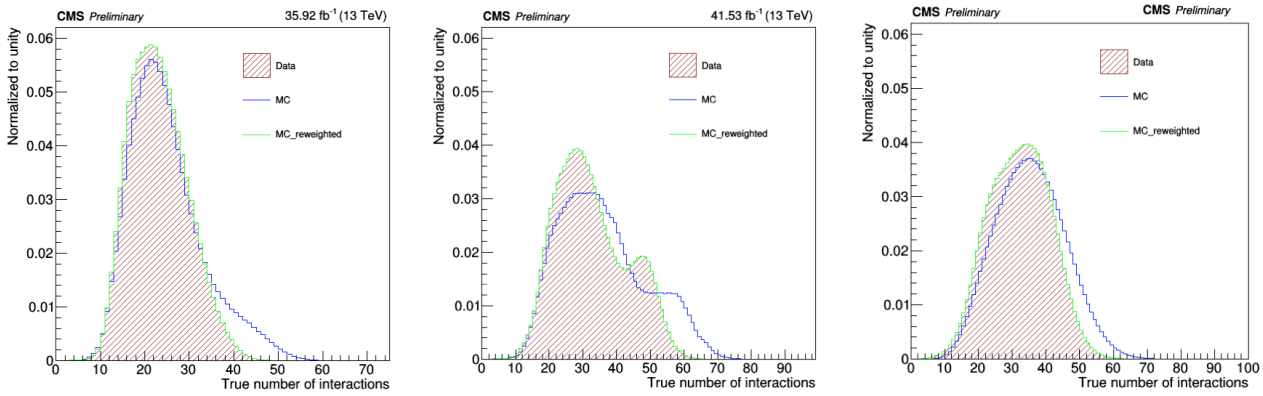


Figure 5.9: The distribution of the number of pileup interactions in 2016 (left) and 2017 (middle) and 2018 (right) in simulation before and after pileup reweighing and compared to that in data.

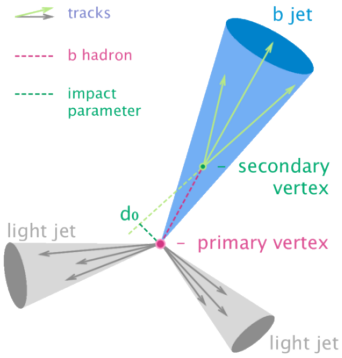


Figure 5.10: Shows a schematic view of a b hadron coming from primary vertex and decay to b jet in a secondary vertex (SV) resulting in charged-particle tracks that are displaced with respect to the primary vertex (PV) with an impact parameter (d_0).

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.

5.3.1.1 Electron kinematics and impact parameter cuts

All the electrons should satisfy a pseudorapidity cut $|\eta^e| < 2.5$ to be in the detector acceptance and have a transverse momentum (p_T^e >7 GeV). 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. The impact parameter is defined as the distance of closest approach between the electron's measured track and the event's primary vertex as shown in figure 5.10. The electrons should satisfy the following constraints:

$$d_{xy} < 0.5\,cm \quad and \quad d_z < 1\,cm \tag{5.2}$$

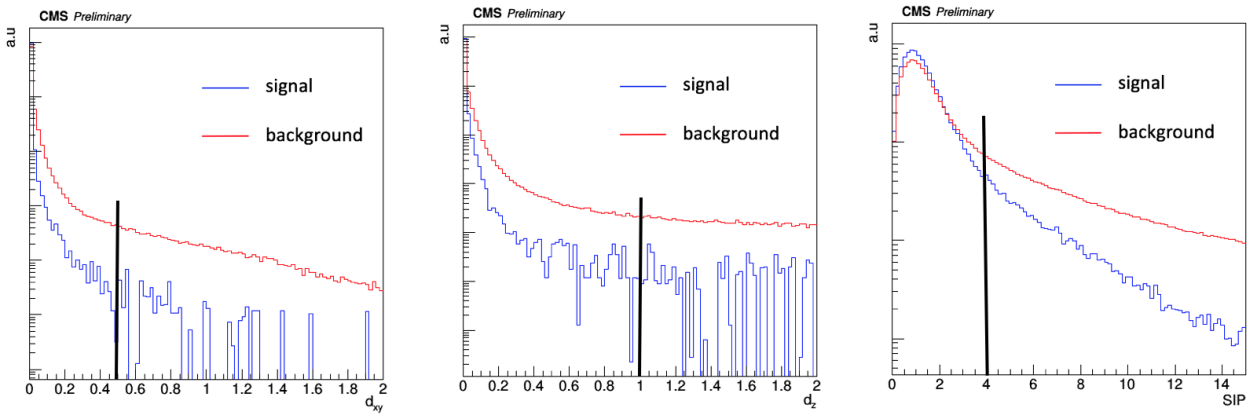


Figure 5.11: Shows the distributions of electron impact parameter variables d_{xy} (left), d_z (middle) and SIP_{3D} (right) for signal and background electrons. Vertical lines represent the cuts used.

where d_{xy} and d_z are the electron transverse and longitudinal impact parameters with respect to the primary vertex respectively. The electrons passing the above cuts are defined as "loose electrons". An additional cut on the three-dimensional significance of impact parameter is required for the selected electrons for the analysis:

$$SIP_{3D} = \frac{|IP_{3D}|}{\sigma_{IP_{3D}}} < 4 \tag{5.3}$$

where $\sigma_{IP_{3D}}$ is the uncertainty in the measurement of the three-dimensional impact parameter. By applying those cuts we can remove some of background electrons coming from the displaced vertices. Figure 5.11 shows the impact parameter distribution used in the electron selection. The plots use signal electrons from one of the signal samples which are geometrically matched to the generated electrons originating from the Higgs boson while electrons from one of background samples are geometrically separated from the generated electrons. The plots show the clear separation between the signal and background electron at the selected cut.

5.3.1.2 Electron Isolation

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 as shown in Figure 5.11. 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 as follows:

$$I_{iso}^{PF} = \frac{\sum P_T^{charged} + \sum E_T^{\gamma} + \sum E_T^{neutral}}{P_T^{l}} \tag{5.4}$$

where $P_T^{charged}$ is the charged hadron transverse momentum, E_T^{γ} and $E_T^{neutral}$ are the photon

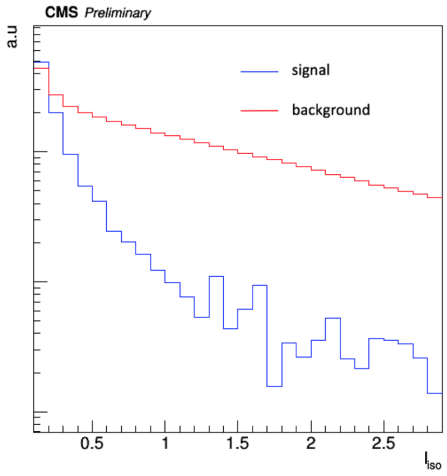


Figure 5.12: Distribution of electron isolation variable for signal and background electrons.

(exclude FSR as will be described in 5.3.3) and neutral hadron transverse energies respectively and p_T^l represents the transverse momentum of the lepton. The isolation variable is sensitive to the pileup since the pileup 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 to take into account the pileup effect:

$$I_{iso}^{PF} = \frac{\sum P_T^{charged} + max\left[0, \sum E_T^{\gamma} + \sum E_T^{neutral} - p_T^{PU}\right]}{P_T^l} \tag{5.5}$$

The pileup contribution in equation 5.5 is defined as:

$$p_T^{PU} = \rho . A_{eff} \tag{5.6}$$

where the effective area A_{eff} is defined as the geometric area of the isolation cone and ρ is the mean of the energy density distribution for the neutral particles in the event. Figure 5.12 shows the isolation variable for the signal and background electron with clear separation between them. The optimal working point (the cut and the cone size) is chosen to ensure higher signal efficiency. The working point for the isolation is

$$I_{iso}^{PF} < 0.35 \tag{5.7}$$

5.3.1.3 Electron Identification

The Identification is another procedure used to distinguish between a good electron and hadronic jets that can mimic the electron signature in the detector. The electron identification

Table 5.8: List of the input variables used for the electron identification BDT classifier [29].

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	$E_{5\times5}$ is the energy computed in the 5×5 block of crystals centered on the highest energy crystal
	$(E_{5\times5} - E_{5\times1})/E_{5\times5}$	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.
	$\Delta\eta_{in} = \lvert\eta_{SC} - \eta_{in}\rvert$	Pseudorapidity distance between the energy weighted position of the supercluster and the track position extrapolated from the innermost track position
	$\Delta\phi_{in} = \lvert\phi_{SC} - \phi_{in}\rvert$	Similar to $\Delta\eta_{in}$ but in the ϕ direction
	$\Delta\eta_{seed} = \lvert\eta_{seed} - \eta_{out}\rvert$	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

exploits a wide range of observables such as the shape of the electromagnetic cluster, the tracking information, the matching between the cluster and the electron track as well as the electron isolation variable. All the electron variables together are fed into a single multivariate classifier using eXtreme Gradient Boosting (XGBoost) techniques. This technique has the advantage over using several cuts on the electron variables "cut-based identification" that have the ability to reduce the amount of fake electrons especially in low p_T region without large impact on the signal efficiency. The list of the variables used is reported in Table 5.8.

The classifier is trained on a Drell-Yan+jets MC sample for both signal and background where the real electrons are reconstructed electrons with a geometrical match with generated electrons originating from Z while the background or fake electrons are reconstructed electrons that are geometrically separated from the generated electrons originating from the Z. 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. The training is done in two bins of electron's p_T ($7 < p_T < 10$ and $p_T > 10$ GeV) to improve the identification performance in the low p_T region. Also the training is done separately for electrons in the barrel, the endcap, and the transition region between barrel in endcap in three η bins ($|\eta| < 0.8$, $0.8 > |\eta| > 1.479$, $|\eta| > 1.479$) since the electron reconstruction is done separately for barrel and endcap regions, as discussed in 4.3.5. An example of the output classifier is shown in Figure 5.13 where blue (red) region represents signal (fake) electrons [21]. A cut on the output of the classifier score has been chosen to distinguish between the signal and the fake electrons while preserving

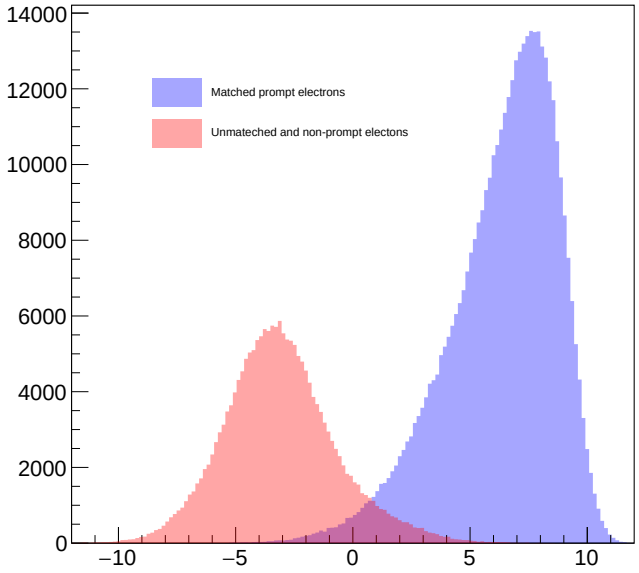


Figure 5.13: The Output of the multiclassifier discriminant for signal electrons (blue) and for fake electrons (red) [21].

a high signal electron efficiency. The values of the score for the different years (2016 -2017 -2018) are given in Table 5.9. Loose electrons passing the identification cut are defined as "tight" electrons for the analysis in this thesis.

5.3.1.4 Electron Efficiency

When applying a set of selections cuts on an object, it is applied to both data and simulation. The efficiency of the cuts could behave differently for data and simulation due to the lack of understanding of the detector in simulation and hence those differences in efficiency need to be studied and quantified. First, we need an accurate and reliable measurement of the efficiency of the object reconstruction, the identification, the isolation and the impact parameter cuts with both data and MC. These measurements are then used to build data/MC scale factors to correct the simulation, by using the tag and probe technique [188]. In the analysis, we have four leptons in the final state which means that the scale factors participate with the power of four in the final event selection therefore it is important to measure them precisely.

The Tag & Probe method uses a mass resonance (e.g. J/ψ, Z) to probe the efficiency of a particular selection criterion on the object. The Z boson is used as a source of electrons and muons while J/ψ is used as a source of low transverse momentum in case of muons. A sample of Z bosons is selected, by finding opposite sign, same flavor leptons with invariant mass consistent with that of a Z boson. One lepton (called the tag) is required to pass a very tight selection criteria to ensure the purity of the Z sample. The other leg, called the probe is initially selected using only a loose criteria and the invariant mass of the tag and

Table 5.9: BDT score required for passing the electron identification, for every year of full Run II samples.

BDT score	$\|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.479$	$\|\eta\| > 1.479$
2016			
$5 < p_T < 10$ GeV	0.9503	0.9461	0.9387
$5 < p_T < 10$ GeV	0.3782	0.3587	-0.5745
2017			
$5 < p_T < 10$ GeV	0.8521	0.8268	0.8694
$5 < p_T < 10$ GeV	0.9825	0.9692	0.7935
2018			
$5 < p_T < 10$ GeV	0.896	0.911	0.94
$5 < p_T < 10$ GeV	0.0424	0.0047	-0.6042

that probe should fit the Z ($60 < M_Z < 120$ GeV). The probe is then passed through the selection criteria and the candidates are sorted into passing and failing collections.

The selection efficiency of a given criteria is measured as the ratio of the number of probes P_{pass} passing the criteria to the total number of probes (P_{all}):

$$\epsilon = \frac{P_{pass}}{P_{all}} \tag{5.8}$$

Those numbers are extracted from fitting the Z peak for the passing and failing probes. The overall electron efficiency can be defined as:

$$\epsilon = \epsilon_{reco} \cdot \epsilon_{sel\,|\,reco} \tag{5.9}$$

where ϵ_{reco} is the probe reconstruction efficiency, $\epsilon_{sel\,|\,reco}$ represents the probe selection efficiency given that it passes the reconstruction step. After extracting the efficiency, the scale factor is built by the ratio of the measured efficiency from data and simulation. The data/MC scale factors can be defined as the overall efficiency measured in data over the overall efficiency measured in MC

$$\text{The scale factor} = C = \frac{\epsilon_{TnP}(Data)}{\epsilon_{TnP}(MC)} \tag{5.10}$$

The electron efficiencies are measured as a function of electron p_T and η in data and MC then a set of scale factors are derived. The electron efficiency is measured in two steps: first, the electron reconstruction efficiency ϵ_{reco} is the GSF tracking efficiency (as defined in 4.3.5). It is measured as the ratio of superclusters that are reconstructed as electrons (passing probes) to the superclusters (all probes). The electron reconstruction efficiency and

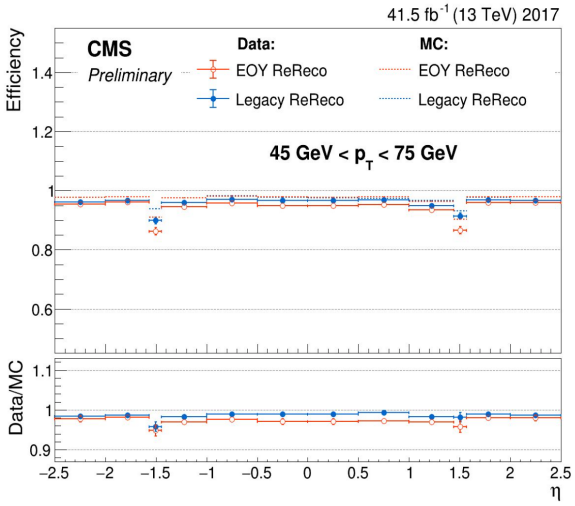


Figure 5.14: Shows the electron reconstruction efficiency measured in data (solid) and MC (dashed) as a function of electron pseudorapidity [22].

the corresponding scale factors are derived centrally by the CMS EGamma physics object group as all the analyses start with the same set of reconstructed electrons. Second, the electron selection efficiency $\epsilon_{sel\,|\,reco}$ that depends on each specific analysis, it is measured as the selection cut efficiency on the electrons that already passed the reconstruction step. The tag needs to satisfy the following quality requirements to ensure high purity:

- $p_T > 30$ GeV, $\eta^{SC} < 2.17$;
- the tag and the probe need to have opposite charge;
- trigger matched to single electron trigger.

Probe electrons only need to be reconstructed as GSF Electron. The electron selection efficiency is measured as a function of the probe electron p_T and its η^{SC}, and separately for electrons falling in the ECAL gaps and non gap. Figure 5.14 shows the electron reconstruction efficiency in data (solid) and MC (dashed) as a function of electron pseudorapidity η for 2017 as an example [22]. Figure 5.15 shows the electron selection efficiency as a function of probe p_T for non-gap electrons (left) and gap electrons (right), together with the corresponding data/MC ratio (bottom), for 2016 (top), 2017 (middle) and 2018 (bottom) samples. Every year of data taken has the corresponding efficiency curves and the scale factors that are used to scale the MC events to data in order to remove any possible discrepancies [21], as shown in Figure 5.16.

5.3.2 Muon selection

The muon reconstruction efficiency [167] is very high which means that we only miss small fraction of muons but we could have an amount of fake muons. A set of selection cuts are applied to the reconstructed muons in order to reduce the amount of fake objects while

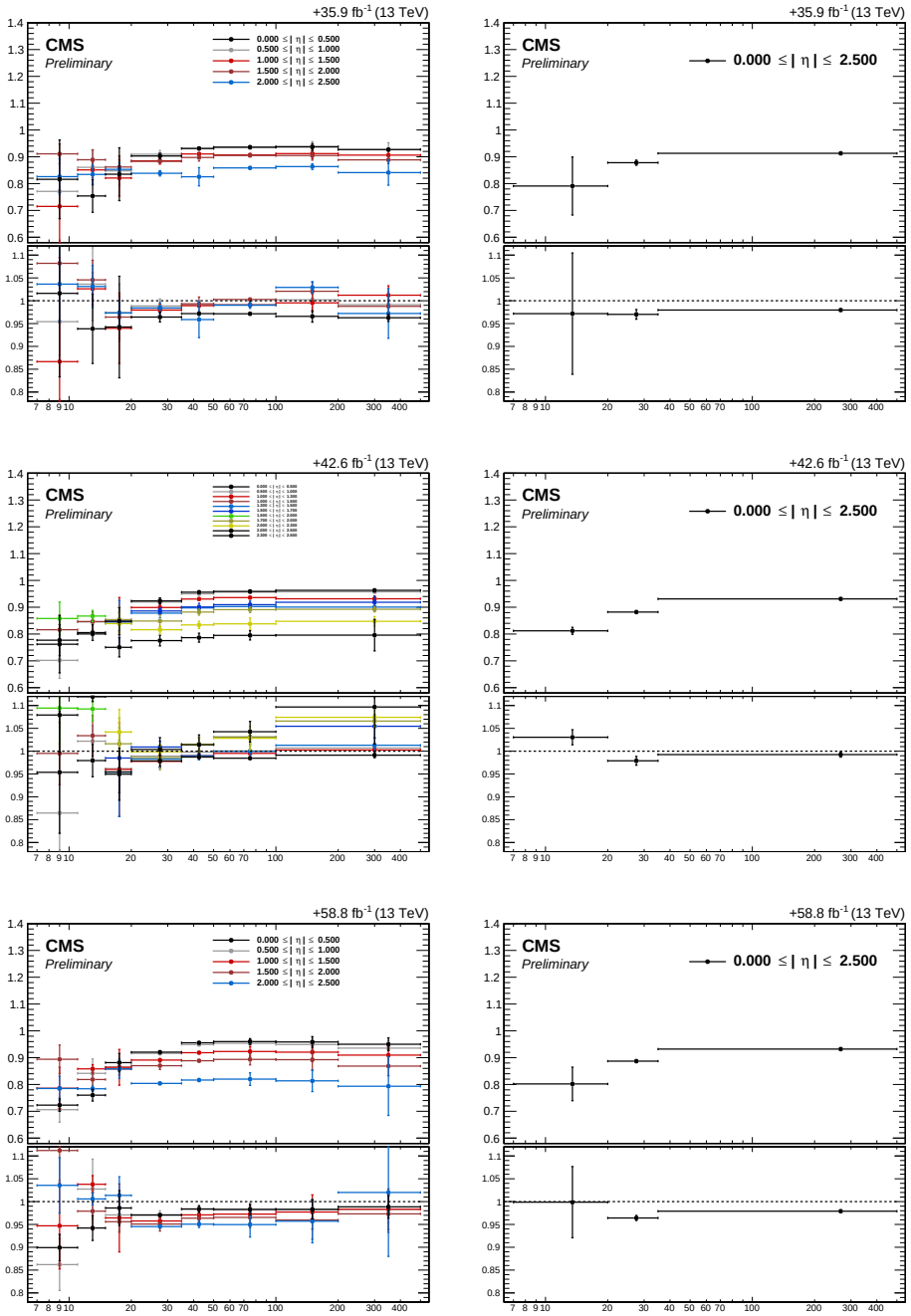


Figure 5.15: Shows the electron selection efficiency as a function of probe p_T for non-gap electrons (left) and gap electrons (right), together with the corresponding data/MC ratio (bottom) for 2016 (top), 2017 (middle) and 2018 (bottom) [21].

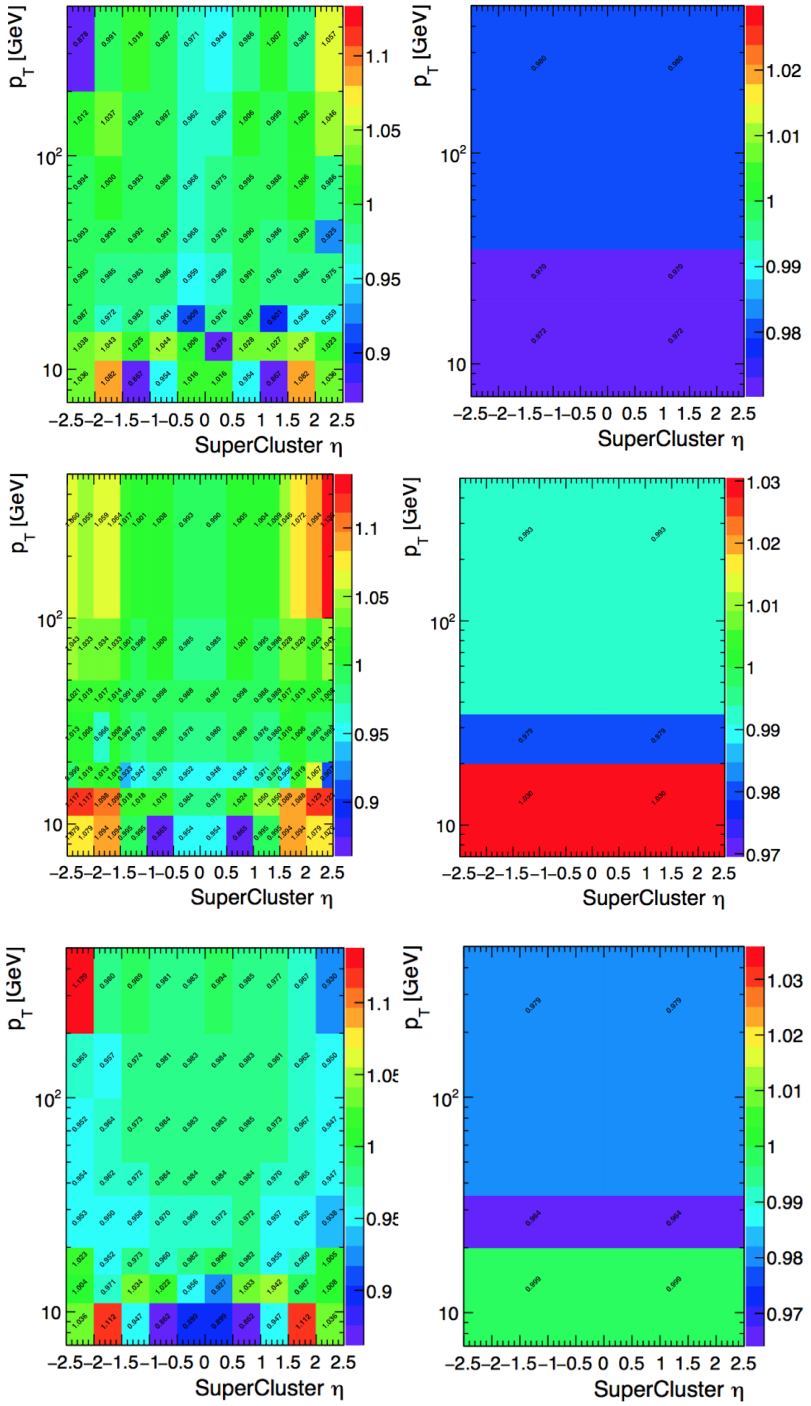


Figure 5.16: Shows the scale factors for non- gap (left) and gap (right) electrons in 2016 (top), 2017 (middle) and 2018 (bottom) data taken [21].

keeping as much as possible the real muons. The criteria applied to the muons are listed below.

5.3.2.1 Muon kinematics and impact parameter cuts

As for the electrons, all muons should be in the detector acceptance $|\eta|$ <2.4 and have $p_T > 5$ GeV. 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) the muons should pass the impact parameter cuts listed below:

$$d_{xy} < 0.5\,cm \quad and \quad d_z < 1\,cm \quad and \quad SIP_{3D} < 4 \tag{5.11}$$

where d_{xy} and d_z are the muon transverse and longitudinal impact parameters with respect to the primary vertex respectively and SIP_{3D} is the 3D muon's significance of impact parameter.

5.3.2.2 Muon Isolation

As already mentioned in 5.3.1 the lepton isolation is a powerful variable to distinguish between real leptons and fake leptons (leptons embedded in jets). The muon isolation can be defined as in equation 5.5 with the difference of the pileup correction of the photons and neutral hadrons. For the muon, the pileup correction is defined by means of the $\Delta\beta$ correction:

$$\Delta\beta = \frac{1}{2}.\sum_{PU}^{ch_had} p_T \tag{5.12}$$

where the summation is over all the transverse momentum of the PF charged hadrons not originating from the primary vertex. In this equation, the pileup contribution in the isolation cone from the neutral particles is considered as half of the charged particles contribution associated to pileup. The isolation cut for muon is the same as for electron ($I_{iso}^{PF} < 0.35$). The final pileup isolation for muons can be defined as:

$$I_{iso}^{PF} = \frac{\sum P_T^{chhad} + max\left[0, \sum E_T^{neuthad} + \sum E_T^{\gamma} - \Delta\beta\right]}{P_T^l} \tag{5.13}$$

5.3.2.3 Muon Identification

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. The muon has to be reconstructed as a Global Muon or TrackerMuon as described in section 4.3.4. If a non global muon is classified as a Tracker muon, it must be "arbitrated" (i.e. tracker track matched with at least one muon segment). Standalone Muon tracks which are only reconstructed in the muon system, must be rejected.

Table 5.10: Requirements of Tracker High p_T muon ID definition.

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)	$d_{xy} < 2$ mm
Vertex compatibility (z)	$d_z < 5$ mm
Pixel hits	At least one pixel hit
Tracker hits	Hits in at least six tracker layers

Muons passing those identification cuts besides the kinematic and impact parameter cuts are marked as "loose muons". An additional cut is added to define tight muons to be used in the analysis. We can define two categories of muons depending on their momentum, loose muons with $p_T < 200$ GeV are considered tight muons if they pass the Particle Flow muon identification criteria as described in section 4.3.2, while loose muons with $p_T > 200$ GeV are considered tight muons if they pass the PF identification criteria or the Tracker High p_T identification criteria as shown in Table 5.10. This high p_T identification criteria is introduced especially to increase signal efficiency for the high-mass search. For example, when a very heavy resonance decays to two Z bosons, both bosons will be very boosted, the leptons coming from the decay of a highly boosted Z will be nearly collinear, and hence the PF identification loses efficiency for muons separated by approximately $\Delta R < 0.4$ which roughly corresponds to muons originating from Z bosons with $p_T > 500$ GeV. So the relaxed definition is introduced in the Tracker High p_T identification instead of using only PF to keep high signal efficiency when dealing with high p_T muons. An additional requirement called "ghost cleaning" is performed to deal with situations when a single muon is incorrectly reconstructed as two or more muons. If two reconstructed muons are sharing 50 % or more of their segments then the muon with lower quality is removed.

5.3.2.4 Muon Efficiency

The same tag and probe technique as mentioned in section 5.3.1.4 is used to drive the muon efficiencies, with the only difference that $J/\psi \to \mu^+\mu^-$ events are used as a source of low p_T muons in addition to the $Z \to \mu^+\mu^-$ sample for high p_T muons. The total efficiency for muons is factorized as the product of:

$$\epsilon = \epsilon_{trk} \times \epsilon_{id|trk} \times \epsilon_{sip|id} \times \epsilon_{iso|sip} \tag{5.14}$$

where:

- ϵ_{trk} is the efficiency to reconstruct a muon track in the tracker detector "tracking efficiency",
- $\epsilon_{id|trk}$ is the PF identification efficiency for muons that have been reconstructed in the tracker detector,
- $\epsilon_{sip|id}$ is the impact parameter efficiency for muons passing the identification, the muon tracking efficiency,
- $\epsilon_{iso|sip3d}$ is the isolation efficiency for muons passing all previous criteria.

The efficiency to reconstruct a muon in the inner tracker ϵ_{track} was measured centrally by the CMS muon Tracking physics object group since all the analysis use the same tracking algorithm. The tracking efficiency in the silicon tracker was measured using a Tag "a tight isolated" muon matching to the HLT object corresponding to Single Muon triggers and a probe which is required to be any stand-alone muon with valid hits in the muon system. The efficiency is calculated by matching the probe to a track with $\Delta R < 0.3$. The tracking efficiency was found to be 99% or higher within the whole acceptance for full Run II and in good agreement with the expectation from simulations as shown in Figure 5.17 for the full Run II data [23]. The muon reconstruction and identification efficiency $\epsilon_{id|trk}$ have been derived centrally by the CMS Muon physics object group since most of the analyses uses the same reconstructed muon. The probes are defined as muon tracks successfully reconstructed in the inner tracker, and the passing probes are those that are also reconstructed as a global or tracker muon and passing a set of loose muon identification cuts defined by physics object group. The efficiency is calculated for every year; for example the muon reconstruction and identification efficiency as a function of muon η where $p_T > 20$ GeV is shown in Figure 5.18 for three years of data taken [24–26], while the muon identification efficiency for $p_T < 20$ GeV is derived using J/ψ as shown in [21]. The identification efficiency is greater than 99 % for the three years. The measurement of the efficiency of the impact parameter cut $\epsilon_{sip|id}$ is done by considering the probe as a muon passing the Loose identification criteria, and it is considered a passing probe if it satisfies the SIP_{3D}, d_{xy}, d_z cuts for this analysis. Finally the muon isolation efficiency, the last component of the offline selection efficiency, is measured for muons that successfully passed both the identification and the significance of the impact parameter criteria. The calculation of the muon efficiency for impact parameter and isolation requirements is done by using Z decays only since the J/ψ decays contain contamination of non-prompt J/ψ's from B hadron decays which are not expected to be isolated. The overall data to MC simulation scale factors is shown in Figure 5.19 and defined as the product of all scale factors components (tracking, reconstruction, identification, impact parameter and isolation) .

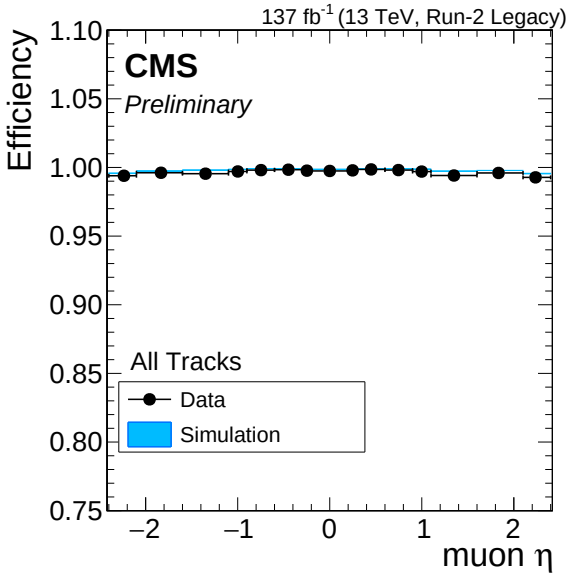


Figure 5.17: Shows the muon tracking efficiency for full Run II data [23]

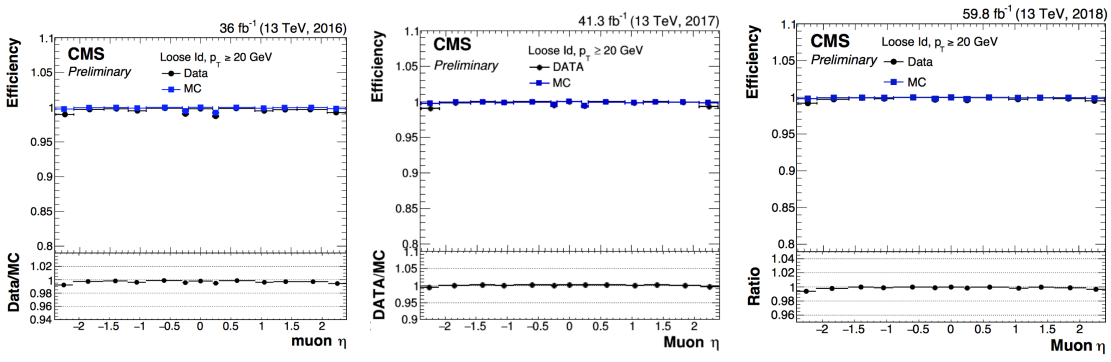


Figure 5.18: Shows the muon reconstruction and identification efficiency for 2016 (left), 2017 (middle) and 2018 (right) [24–26]

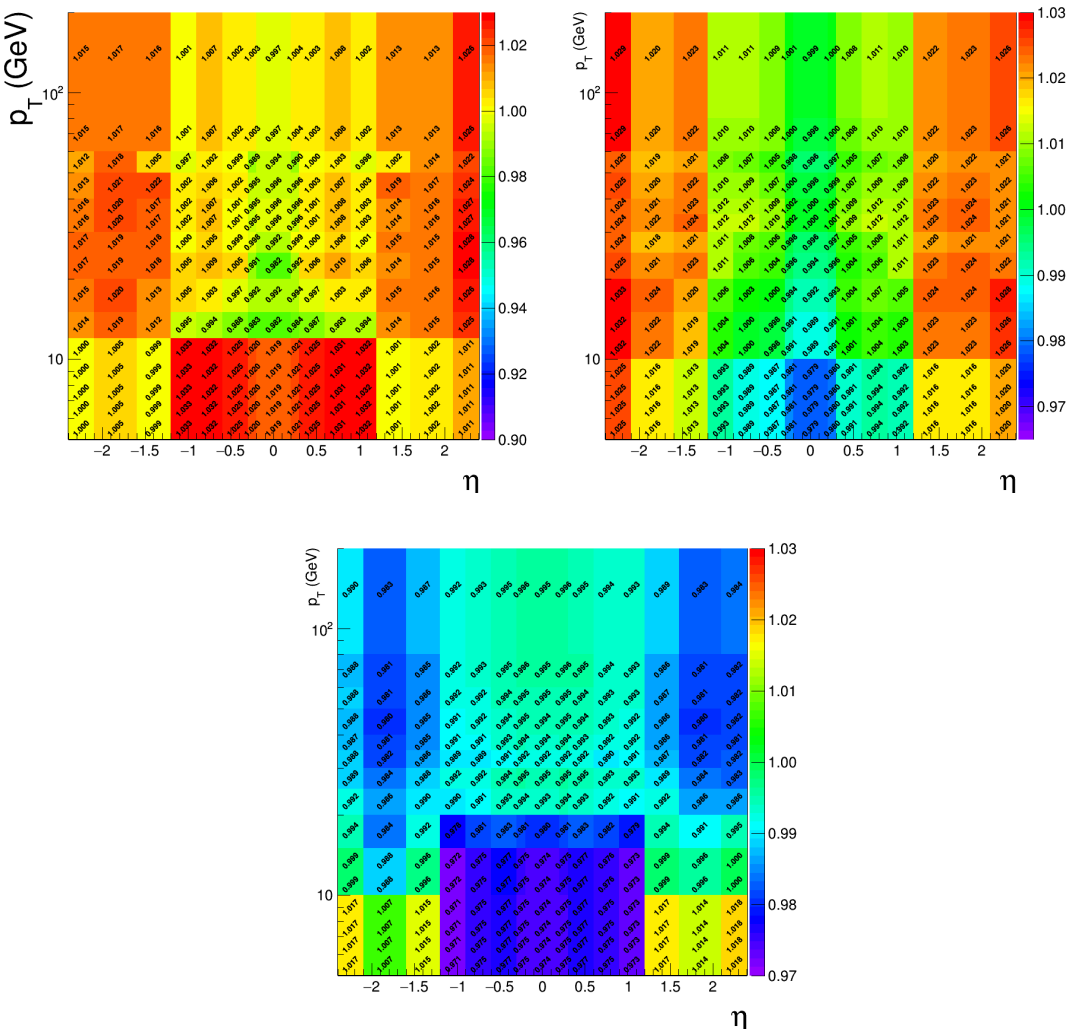


Figure 5.19: Shows the overall data to simulation scale factors for muons, as function of p_T and η for 2016 (top left), 2017 (top right) and 2018 (bottom) [21].

5.3.3 Photon selection

Photons are considered in this analysis as a candidates for the final state radiation (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:

- The preselection cut: The PF photons are required to satisfy a pseudorapidity cut $|\eta^{\gamma}| < 2.4$, to have $p_T^{\gamma} > 2$ GeV and PF isolation $I_{PF}^{\gamma} < 1.8$. 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 $\Delta R > 0.0001$ and the neutral hadrons and photons have a threshold of 0.5 GeV with a veto cone of $\Delta R > 0.01$, 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 event that pass the loose ID and impact parameter cuts.

- Discard all the photons that do not satisfy the cuts $\Delta R(l,\gamma)/E_{T,\gamma}^2 < 0.012$ 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 $\Delta R(l,\gamma)/E_{T,\gamma}^2$.

- 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 [189–192].

Studying the effect of applying the FSR algorithm on different signal samples showed that the effect is less than 1% as shown in the four leptons invariant mass distribution before and after applying the FSR algorithm on one signal sample in Figure 5.20. While the effect of applying FSR algorithm on Higgs to ZZ background sample which is the main background is less than 2 %. The muon channel has a bigger effect than electron channel where the majority of FSR already included in the electromagnetic supercluster.

5.3.4 The Missing Transverse Energy selection

As mentioned in section 4.3.6, large MET values reconstructed in the event could be interpreted as a signal of new physics such as the production of dark matter particles but in

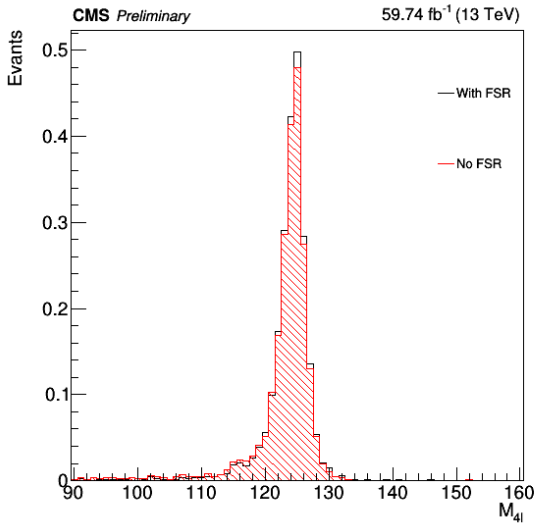


Figure 5.20: Distribution of four-lepton invariant mass before and after applying the FSR algorithm on one signal sample.

reality, large MET values could come from uninteresting sources from the detector. MET with uninteresting causes is called false MET, anomalous MET, or fake MET. The fake MET could come from experimental or instrumental sources as explained in the next section.

5.3.4.1 The Experimental and instrumental sources of p_T^{miss}

Besides the expected missing transverse momentum from neutrinos or dark matter candidate escaping the detector, there are other sources that can lead to reconstruction of high MET.

- **Beam Halo:** The protons in the LHC beam can interact with the collimators that are installed in the LHC ring producing pions and other light hadrons which decay to muons; those particles are called beam halo. Those particles travel nearly parallel to the collision axis and can interact in the calorimeters, leaving energy deposits along a line with constant ϕ and can leave signal in the muon system. Taking into account this fake signal affects the reconstruction of MET. This kind of signal has different properties with respect to the pions and muons coming from p-p interaction. The timing and the shower shape in the calorimeter is different, moreover the muons produced in collisions do not often leave energy in the calorimeters but they are reconstructed using the tracker and the muon chambers. Those beam halo particles are produced in the beam pipe far from the collision point of the detector, so their signal will be out-of-time with respect to the bunch crossing [193].

- **Cosmic rays:** When high energetic particles like protons from outside the solar system interact in the high atmosphere, they produce a cascade or shower of particles that decay to stable particles like muons and neutrinos. Those particles live long enough to reach the earth and can cross the detector. Neutrinos will escape the detector without

interaction while muons can interact and leave a signal in the detector and hence can give a wrong information in the MET calculation. Those particles arrive randomly to the detector, so the time at which they arrive is random and is used to discriminate them against the time of the pp bunch-crossings [193].

- **Electronic Noise:** Electronic noise is one of the sources that can introduce anomalous MET. In both ECAL and HCAL, the crystals started to suffer from radiation damage or the readout electronics don't work as desired which lead to a high energy noise which introduces large anomalous signals. Those signals can mimic physical signals from particles and hence introduce fake MET. The list of such crystals has evolved with time due to their loss of transparency. The events from those crystals are filtered out [194, 195].

- **Reconstruction issue:** The reconstruction algorithm can itself introduce a source of fake MET. The poor reconstruction of muons with high momentum during muon tracking could introduce a poorly reconstructed PF muon or a poorly reconstructed charged hadron, in both cases it introduces fake MET to the event. The events with low quality muon track need to be removed.

In CMS, the MET working group is dedicated to the identification of events containing such spurious MET and design the algorithms to reject those events. Those set of algorithms are the basic rationale for filtering the events, and are referred as "MET filters".

5.3.4.2 The MET Filters

A set of MET filters [196] have been introduced by CMS MET group to identify false MET. After the identified false MET is removed by applying the filters, the agreement of the MET spectrum with MC, for what the causes of false MET are not explicitly simulated, will typically improve significantly. The MET filters are used for the analysis and a description of each filter is reported below:

- **"HBHENoiseFilter"** & **"HBHENoiseIsoFilter"**: Those filters remove noisy events from HCAL "HB, HE stands for HCAL barrel and HCAL endcap respectively" where some scintillator produce anomalous signals with pulse shapes different from those coming from a clean signal.

- **"goodVertices filter"**: This filter removes the events with noisy vertex reconstruction produced from pileup effect. This filter requires the reconstruction of at least one good vertex (PV) with the requirements of high number of degree of freedom ($N_{PV} > 4$), with collisions restricted along the Z- axis ($Z_{PV} < 24$ cm) and with a small radius of the primary vertex ($r_{PV} < 2$ cm).

- **"BadPFMuonFilter"**: This filter removes events with mis-reconstructed muon. It removes events with high p_T muons of low quality (low quality means large uncertainty

on its transverse momentum) passes the PF requirements and enters the computation of the MET as a PF muon.

- **"EcalDeadCellTriggerPrimitiveFilter"**: This filter removes events with non-functioning ECAL data links, comparing the sum of energy deposited in each supercluster cell to the trigger primitive saturation energy.
- **"globalSuperTightHalo2016Filter"**: This filter removes events containing beam halo particles flying with the beam.
- **"ecalBadCalibReducedMINIAODFilter"**: A few endcap ECAL crystals in the very forward region lost their transparency and start to behave sometimes oddly, leading to high energy noise and hence introducing fake MET. This filter dedicated to remove such kind of events depending on a list of such crystals.

Figure 5.21 shows two event displays, the left plot shows an event where four supercrystals in the ECAL endcap were found to produce anomalous high energy signals (red bars) leading to fake reconstructed missing transverse momentum (violet arrow); this event has bad quality hits with energy greater than 1 TeV in an affected ECAL supercrystal, and is rejected by the dedicated ECAL noise filter. The right plot shows an event with mis-reconstructed muon (muon 1) with high p_T leading to fake reconstructed missing transverse momentum (violet arrow). This event is rejected by "BadPFMuonFilter" based on the presence of a high p_T muon with a poor quality track and a large momentum measurement uncertainty [197].

Figure 5.22 left shows the MET distribution before and after the application of the filters described above. Without applying the MET filters, the tail of the distribution is polluted by fake MET that can be described as an excess of events from a signal of new physics while after applying the filters those events are removed.

In the analysis, the MET filters are applied in both data and MC simulated samples, to check if there is a difference or bias from applying the filters, and the MET filters efficiency is calculated from data and all MC samples used in the analysis. The filters efficiency is calculated as:

$$\epsilon = \frac{Number\ of\ events\ after\ applying\ the\ MET\ filters}{Number\ of\ events\ before\ applying\ the\ MET\ filters} \tag{5.15}$$

Figure 5.22 right shows the filters efficiency as a function of MET distribution at an advanced step of the event selection where we reconstruct Higgs boson from four charged leptons, the event selection will be discussed in details in 6.1. The filters don't introduce any bias between data and MC simulated samples.

5.3.4.3 Type I correction

For an accurate reconstruction of MET, it is not sufficient to reconstruct all visible particles produced in collisions as defined for the raw MET in equation 4.1. Some corrections need to

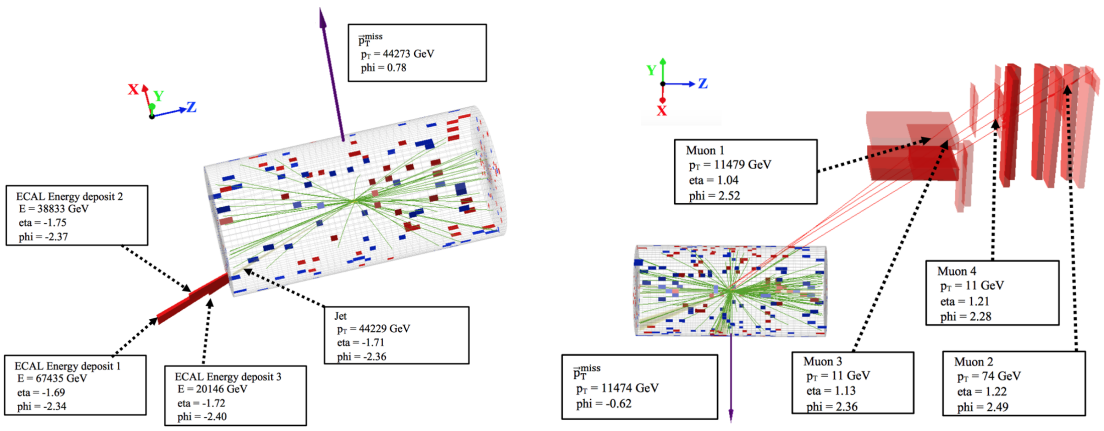


Figure 5.21: Event display for two rejected events by (a) ECAL noise filter due to anomalous high energy signals in ECAL (b) BadPFMuonFilter due to mis-reconstructed high p_T muon with large measurement uncertainty.

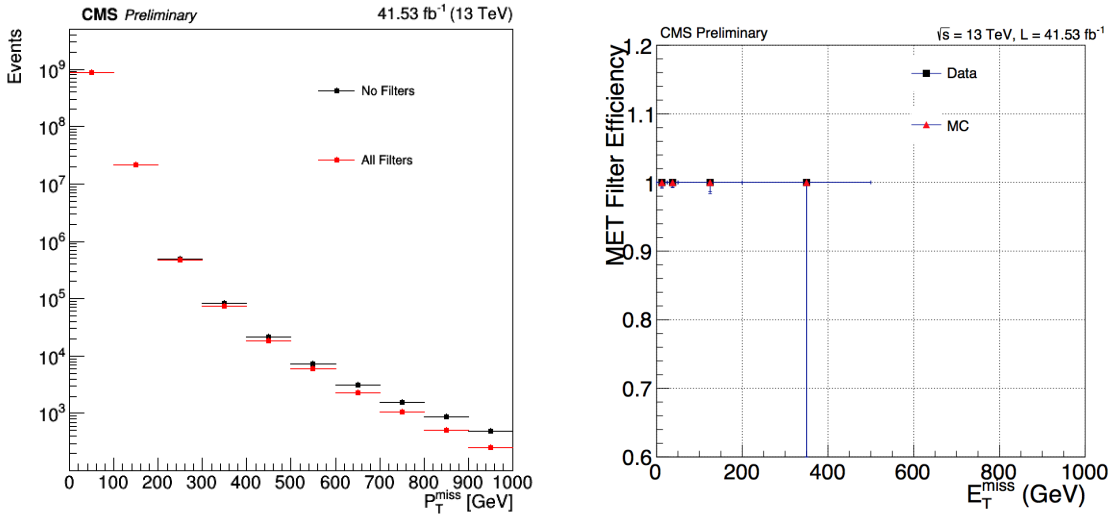


Figure 5.22: Shows the MET distribution before and after applying the MET filters (left), MET filters efficiency as a function of MET distribution after reconstruction of Higgs boson from four charged leptons (right) using dataset recorded in 2017.

be added to the raw MET. Type-I correction [198] is the recommended correction to be used in CMS. This correction propagates the jet energy corrections (JEC)[5] to the MET definition. It replaces the vector sum of the transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC is applied. We can classify the particles into two sets: either clustered as jets or unclustered, so the raw MET definition can be written as:

$$\overrightarrow{\not{E}}_T^{raw} = -\sum_{i\epsilon\, jets} \overrightarrow{p}_{T,i} - \sum_{i\epsilon\, uncl} \overrightarrow{p}_{T,i} \tag{5.16}$$

where the first term is the vector sum of $\overrightarrow{p}_T$ of all jets and can be written as:

$$\sum_{i\epsilon\, jets} \overrightarrow{p}_{T,i} = \sum_{jet} \overrightarrow{p}_{Tjet}^{raw} \tag{5.17}$$

The superscript "raw" indicates that the JEC is not applied to the jets. The Type-I correction replaces the raw jet $\overrightarrow{p}_T$ with the corrected jet $\overrightarrow{p}_T$ so it can be written as the difference between the two vector sum:

$$C_T^{Type-I} = \sum_{jets} \overrightarrow{p}_{Tjet}^{raw} - \sum_{jet} \overrightarrow{p}_{Tjet}^{JEC} \tag{5.18}$$

using equation 5.17 to rewrite equation 5.18 as:

$$C_T^{Type-I} = \sum_{i\epsilon\, jets} \overrightarrow{p}_{Ti} - \sum_{jet} \overrightarrow{p}_{Tjet}^{JEC} \tag{5.19}$$

This is a vector term which can be added to the raw MET in equation 5.16 to produce a corrected MET.

$$\overrightarrow{\not{E}}_T^{Type-I} = \overrightarrow{\not{E}}_T^{raw} + C_T^{Type-I} \tag{5.20}$$

$$\overrightarrow{\not{E}}_T^{Type-I} = -\sum_{i\epsilon\, jets} \overrightarrow{p}_{T,i} - \sum_{i\epsilon\, uncl} \overrightarrow{p}_{T,i} + \sum_{i\epsilon\, jets} \overrightarrow{p}_{Ti} - \sum_{jet} \overrightarrow{p}_{Tjet}^{JEC} \tag{5.21}$$

The first and third sums cancel with each other and as a result the Type-I corrected MET can be written as:

$$\overrightarrow{\not{E}}_T^{Type-I} = -\sum_{jet} \overrightarrow{p}_{Tjet}^{JEC} - \sum_{i\epsilon\, uncl} \overrightarrow{p}_{T,i} \tag{5.22}$$

The index i runs over all particle candidates. Equation 5.22 represents the type-I corrected MET to be used in the analysis.

[5]The measured jet energy is usually different than the energy of the true particle or parton energy because of detector non-linear response. So, jets require some energy corrections. The jet corrections are a set of tools that allows the proper mapping of the measured jet energy deposition to the particle level jet energy. JEC applied depending on the η and p_T values of jets.

5.3.4.4 The XY-Shift correction

The distribution of MET should be independent of azimuthal angle ϕ because of the rotational symmetry of the collisions around the beam axis. However, it was observed that the reconstructed MET depends on ϕ. The MET ϕ distribution has a sinusoidal shape with the period of 2π as shown in Figure 5.23 left. This modulation could be introduced from anisotropic detector responses, some inactive calorimeter cells, the detector misalignment and the displacement of the beam spot. The amplitude of the modulation increases roughly linearly with the number of the pile-up interactions. The xy-Shift correction [199] is introduced by the MET physics object group and is used to reduce this MET ϕ modulation. Applying this correction is also a mitigation for the pile-up effects.

The amplitude of the ϕ modulation can be reduced if the origin of the coordinate in the transverse momentum plane is shifted

$$\overrightarrow{p}_{Ti} \rightarrow \overrightarrow{p}_{Ti} - \overrightarrow{c} \tag{5.23}$$

where $\overrightarrow{c}$ is the shift. Therefore the xy-shift corrected E_T^{miss} becomes:

$$\begin{aligned} \overrightarrow{\not{E}}_T^{miss^{xy}} &= -\sum_{i\epsilon\, all} (\overrightarrow{p}_{Ti} - \overrightarrow{c}) \\ &= -\sum_{i\epsilon\, all} \overrightarrow{p}_{Ti} + \sum_{i\epsilon\, all} \overrightarrow{c} \\ &= \overrightarrow{\not{E}}_T^{miss} + n\overrightarrow{c} \end{aligned} \tag{5.24}$$

where n is the number of particles in the event. This ϕ modulation is applied per year, depending on the run number and the number of vertices in the event; as an example Figure 5.23 shows the MET ϕ distribution before (left) and after (right) applying the xy-shift correction in 2018 datasets where the difference before and after applying the correction is clear. The type-I corrected MET is used in the analysis after applying the MET filters and the XY- shift correction.

5.3.4.5 The L1 pre-firing

During data taken in years 2016 and 2017 the ECAL gradual timing shift was not propagated to the L1 trigger primitives properly. This effect results in a wrong association of a large fraction of high η trigger primitives to the previous bunch crossing. This effect is not described in the MC simulations, therefore a weight is computed for each event and it is applied to the MC simulation for the years 2016 and 2017. Figure 5.24 shows the impact of the L1 pre-firing weights on one signal sample for 2016 (left) and 2017 (right) analysis. The figure shows the p_T^{miss} distribution with (red) and without (black) L1 pre-firing weights, the effect is less than 2 %.

This chapter presented all the cornerstones to perform the analysis, the data sets, the signal and background samples besides the physics objects selections along with their selection

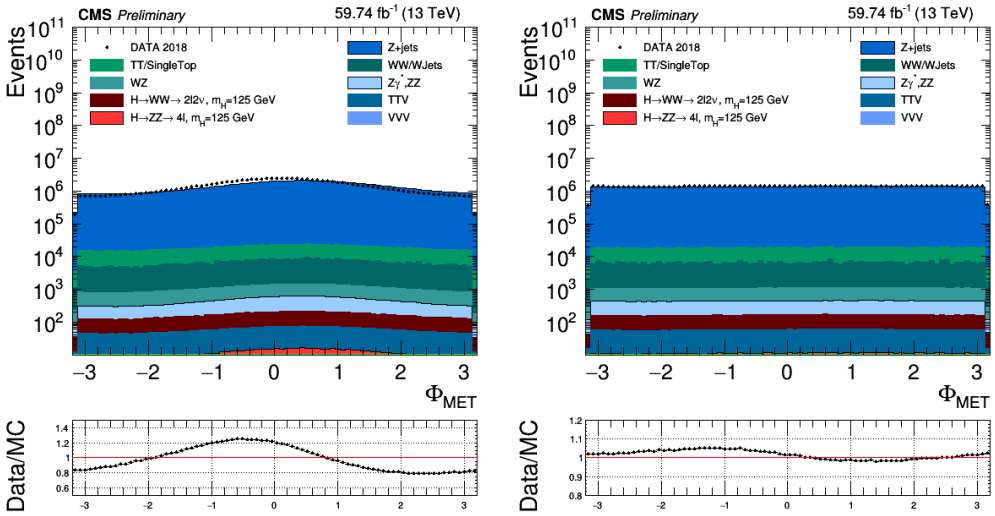


Figure 5.23: The MET ϕ distribution before (left) and after (light) applying the xy-shift correction.

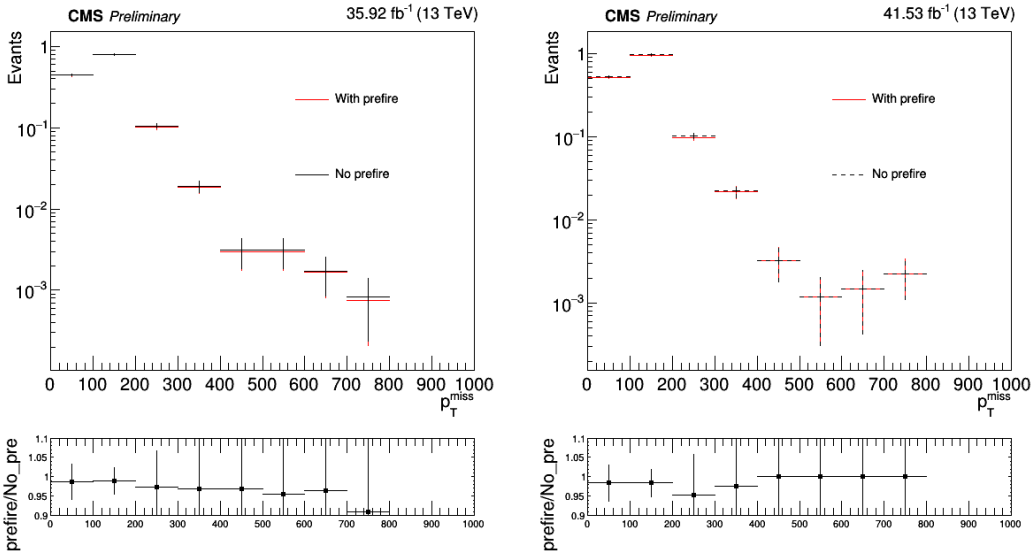


Figure 5.24: Comparison between one signal samples with (red) and without (black) L1 pre-firing weights in 2016 (left) and 2017 (right) analysis.

efficiencies. Those cornerstones will be linked together to select the physics events as will be shown in the next chapter.

6 Analysis Strategy and Events Selection

After selecting all the objects needed to perform the analysis as detailed in the previous chapter, all the stones have to be put together to select the events. In this chapter, we describe the event selection strategy based on a cut based approach applied on some observables that have the ability to separate the Mono-Higgs signal from background processes; in addition the background estimation is presented. Event selection process is a crucial step in the analysis, especially in the physics processes which suffer from low production cross section; moreover without a reliable estimate of the background rates, the analysis would not be reliable to search for new signals. This chapter is organized in the same way the analysis was performed: the event selection along with the choice of the signal region, the estimation of backgrounds and the data to simulation comparison in different steps.

6.1 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:

- Trigger selection;
- Vertex selection;
- Objects selection;
- Higgs candidate selection;
- MonoHiggs selection;

A flow chart describing the selection steps is shown in Figure 6.1. Some of those selection steps have been already discussed in the previous chapter, but I will just flash them in the following points:

1. **Trigger Selection**: In this step, the events that fired at least one of the HLT trigger paths mentioned in 5.1 are saved.

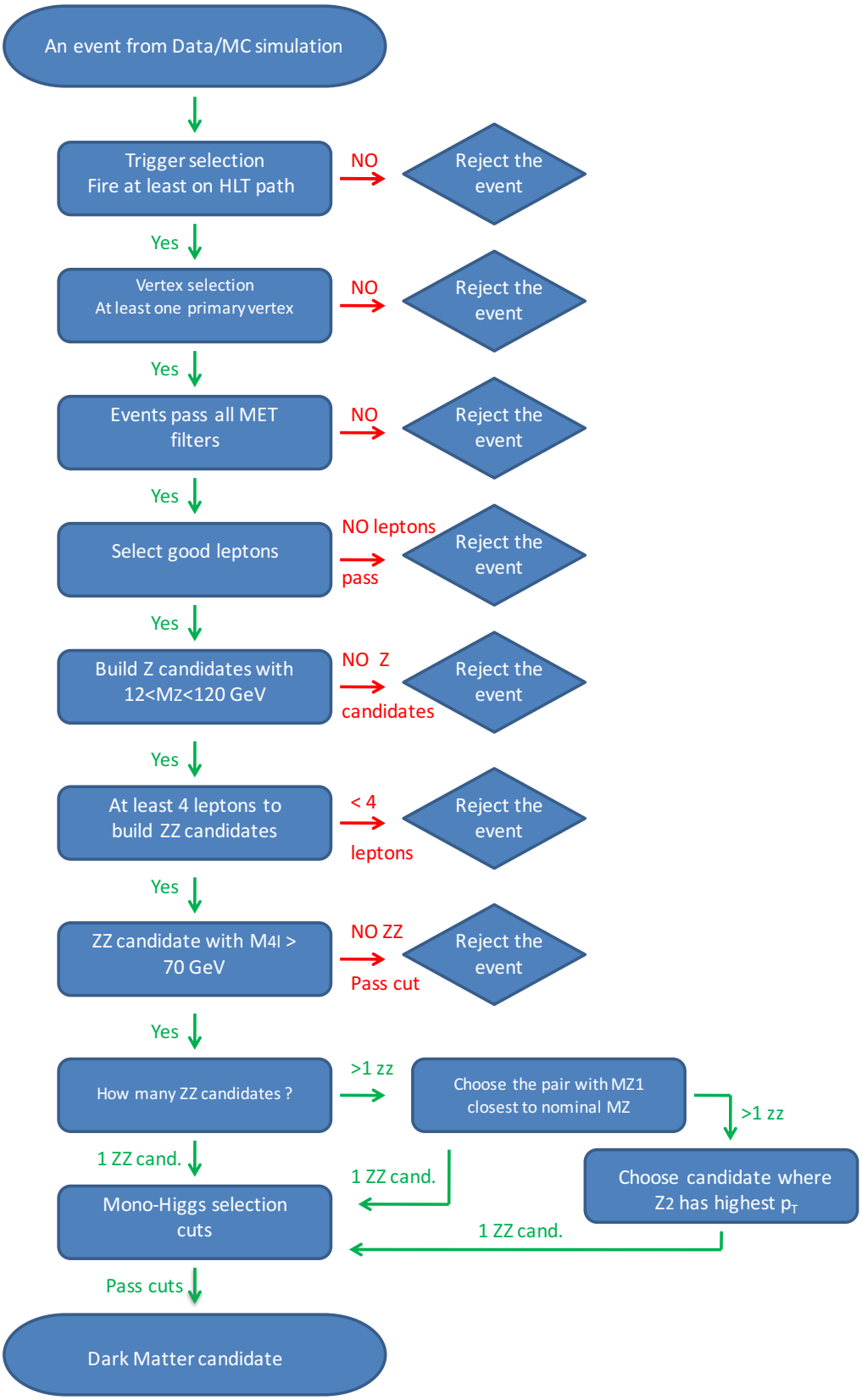


Figure 6.1: A flow chart summarizes the event selection steps.

Table 6.1: Summary of object selection.

Object	Required cuts
Electrons	**- kinematic cuts**: $p_T^e > 7$ GeV , $\|\eta^e\| < 2.5$ **- Impact parameter cuts**: $d_{xy} < 0.5$ cm , $d_Z < 1$ cm , $SIP_{3D} < 4$ **- Identification cut**: BDT cut from table 5.9
Muons	**- kinematic cuts**: $p_T^\mu > 5$ GeV , $\|\eta^\mu\| < 2.4$ **- Impact parameter cuts**: $d_{xy} < 0.5$ cm , $d_Z < 1$ cm , $SIP_{3D} < 4$ **- Identification cut**: PF Muon ID **- ISolation cut**: $I_{PF}^\mu < 0.35$
FSR photons	**- kinematic cuts**: $p_T^\gamma > 2$ GeV , $\|\eta^\gamma\| < 2.4$ **- ISolation cut**: $I_{PF}^\gamma < 1.8$ **-** ΔR cuts: $\Delta R(l,\gamma) < 0.5$, $\frac{\Delta R(l,\gamma)}{(p_T^\gamma)^2} < 0.012$
p_T^{miss}	**- Type**: Type1 correction. **- Filters**: Apply all p_T^{miss} filters discussed in 5.3.4.2. **- Correction**: Apply xy correction discussed in 5.3.4.4.

2. **Vertex selection**: The event is required to have at least one good primary vertex (PV) fulfilling the quality requirements described in 4.3.3.

3. **Object selection**: In this step we distinguish between real and fake leptons by exploiting different observables as already discussed in 5.3. We end up with a set of objects $(\mu, e, \gamma, p_T^{miss})$ to be used in building the Z candidates in the next step. Summary of objects selection is shown in Table 3.

4. **Higgs candidate selection**: The first criterion to select a signal region is to define a preselection phase space enriched with $H \to ZZ \to 4l$ events. To do this the official $H \to ZZ \to 4l$ analysis for the Higgs discovery is taken as reference [21]. In this step the Higgs candidate is built in the event from four selected leptons. The event selection is identical for the three years (2016, 2017, 2018) analysis, and can be summarized in the following steps:

   (a) **Z candidates**: Build Z candidates from all possible pairs of opposite charge and same flavor of selected leptons, such as (e^+e^- and $\mu^+\mu^-$) taking into account the FSR photon if it exists. The distribution of Z candidates for the full run II data is shown in Figure 6.2 left while the p_T^{miss} distribution at the same step is shown in Figure 6.2 right. The solid color distributions in the figure 6.2 represent the different backgrounds stacked after being scaled by the sample cross section and the integrated luminosity, the colored lines represent different signal samples superimposed and are scaled by 50 and the solid points represent the data collected. The dominant background contributing in this step is the Drell-Yan process where we require at least one reconstructed di-lepton pair with an invariant mass in a range around the nominal Z_0 mass; a good agreement between data and background is

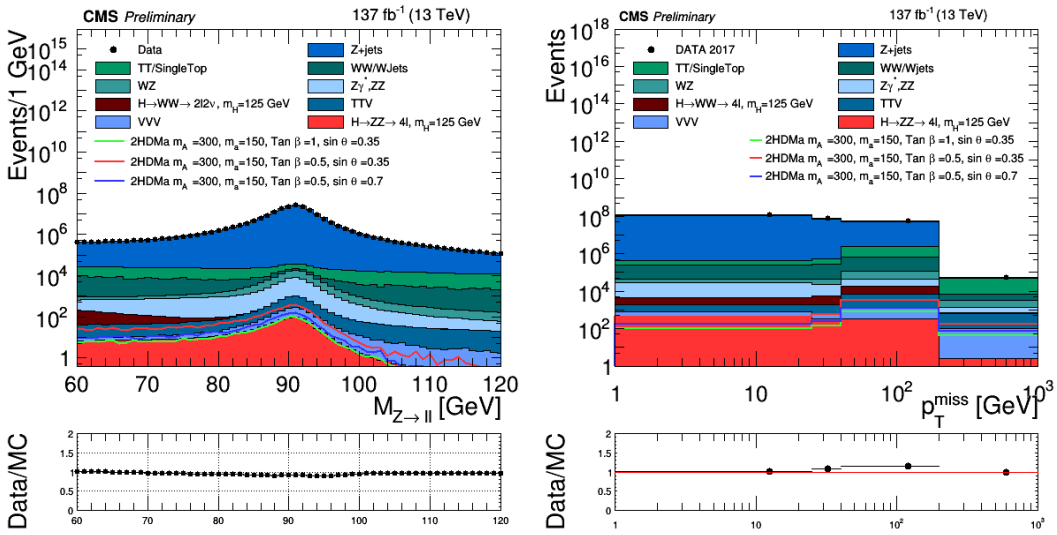


Figure 6.2: Distribution of Z invariant mass (left) and the missing energy (right) for full run II data.

obtained at this step on the analysis.

(b) **ZZ candidates**: Build all possible non overlapping Z candidates which are ZZ candidates that do not share the same lepton. The Z candidate with invariant mass $m_{ll\gamma}$ closest to the nominal Z_0 boson mass is denoted as Z_1, and the seconed one is denoted as Z_2. This step reduces the backgrounds such Z+ jets, WZ and $t\bar{t}$ background.

(c) All ZZ candidates are required to pass the following list of requirements:

(i) **Ghost removal**: $\Delta R > 0.02$ between any of two leptons. This cut is performed to deal with the situation when a single object reconstructed as two or more objects.

(ii) **Lepton** p_T: at least two of the four selected leptons must pass $p_{T_1} > 20$ GeV and $p_{T_2} > 10$ GeV.

(iii) **QCD suppression**: All possible opposite sign pair of leptons (regardless of lepton flavor) that can be built with the four leptons must satisfy $m_{ll} > 4$ GeV. The FSR photons are not used in computing m_{ll}, since a QCD-induced low mass dilepton resonance (such as J/ψ) may have photons nearby (e.g. coming from π^0 decay).

(iv) Z_1 **invariant mass**: $40 < m_{Z1} < 120$ GeV

(v) **Alternative pairing check**: Defining Z_a and Z_b as the mass-sorted alternative pairing Z candidates, with Z_a being the one closest to the nominal Z_0 boson mass, the ZZ candidate is excluded if $m_{Z_b} < 12$ GeV and m_{Z_a} is closer to the nominal Z_0 boson mass than m_{Z_1}. This cut discards 4μ and 4e candidates where the alternative pairing looks like an on-shell Z and a low-mass l^+l^- resonance.

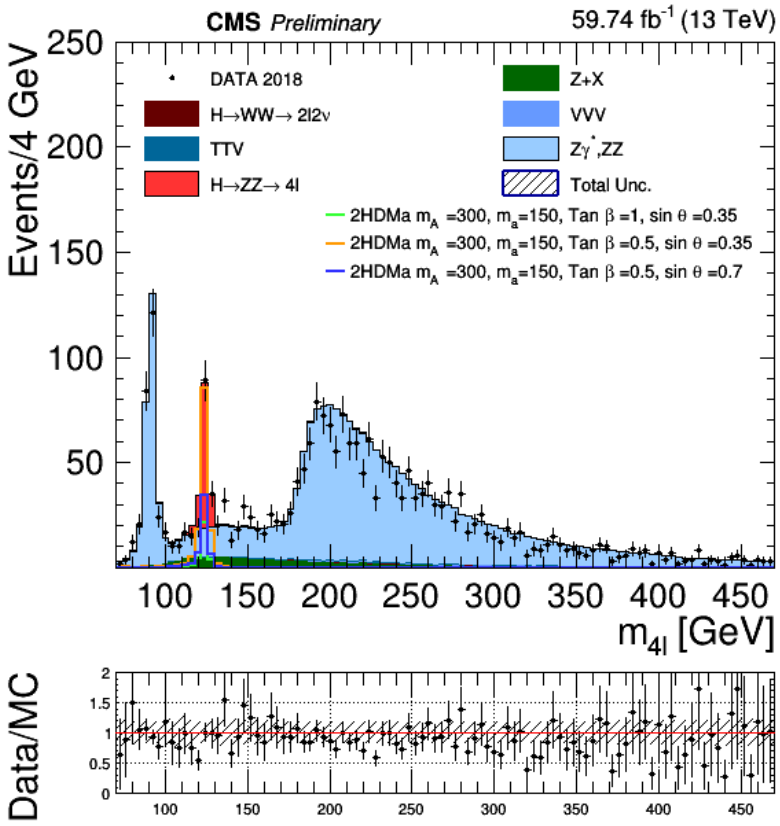


Figure 6.3: Distribution of four-lepton invariant mass after the SM Higgs selection for 2018 datasets ,simulated background and signal events.

(vi) **Four-lepton invariant mass**: $m_{4l} > 70$ GeV.

(d) **Choice of the best ZZ candidate**: if more than one ZZ candidate survives the above selection, we choose the pair where Z_2 has the highest value of the scalar sum of the leptons p_T.

Figure 6.3 shows the four leptons invariant mass distribution for 2018 datasets, as an example, then all the other distributions will be shown in section 6.5. The solid color distributions represent the different backgrounds stacked after scaled by the sample cross section and the integrated luminosity; the colored lines represent different signal samples superimposed and scaled by 10 while the solid points represent the data. We can notice that the distributions of the signal samples are peaked around 125 GeV as expected since we correctly reconstructed the Higgs boson in the signal. Anyway, at this step the signal is completely overwhelmed by the backgrounds such as the non resonant ZZ and SM Higgs background processes. Therefore, we need to define additional cuts to enhance the mono-Higgs signal region.

5. **MonoHiggs Selection**: Additional cuts are added to define the signal region which

consists of four leptons in the final state plus missing energy that represents the main variable of the analysis. The cut optimization and the selection are discussed in section 6.2.

6.2 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 by scanning over a range of cuts for different variables and selecting the set of cuts that maximizes the sensitivity, measured directly by the cross section upper limit. Defining the same signal region for all the signal points could lead to a small loss in sensitivity for some points, but the benefit of simplicity outweighs the cost of this loss. Therefore, we apply different cuts for several observables and choose the cuts that are suitable for all signal models.

- **Four leptons invariant mass M_{4l} :** This cut is powerful to discriminate the mono-Higgs signal from the SM backgrounds such as the non resonant ZZ background. As already mentioned before, the Mono-Higgs signal dark matter particles are produced in association with the Higgs boson, where we consider the Higgs particle produced as the SM Higgs with mass of 125 GeV. Figure 6.3 shows the distribution of four leptons invariant mass where different stacked colors correspond to different backgrounds; part of the background is estimated from data with a data driven method which will be introduced later in section 6.4; different lines represent signal samples with a different value for the scanned parameters and finally data are represented by solid points. We studied different cuts for a mass window between 5 to 15 GeV around the Higgs mass peak at 125 GeV and measured the signal efficiency and background rejection for each cut. We defined the optimal cut that maximizes the sensitivity, measured directly by the cross section upper limit[1], as $|M_{4l} - 125| < 10$ GeV. This cut preserves signal efficiency greater than 95 % for different signal points with greater than 90 % of the background rejection as shown in Figure 6.4. In the next sections data are not shown in this region to avoid any bias in defining the selection of the signal region.

- **Number of leptons**: Since the mono-Higgs signal has four leptons in the final state coming from Higgs decay, if we restrict the number of leptons to exactly four leptons, we can reduce the contribution of $t\bar{t} + V$, VVV, ZH, WH and ttH background events without a relevant impact on the signal, as it can be inferred from Figure 6.5 left. This cuts preserve higher signal efficiency $\approx$ 98 % along with the help in reducing combinatorial backgrounds.

[1]The upper limit will be introduced in 8.1

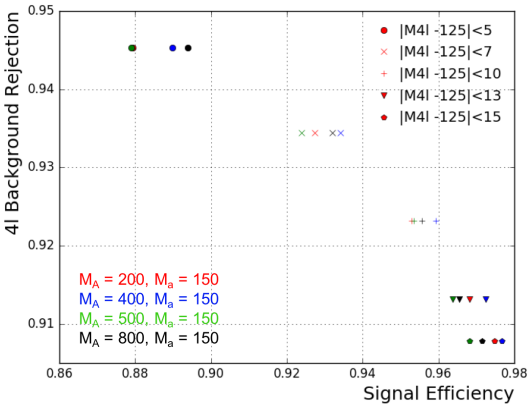


Figure 6.4: Signal efficiency vrs. background rejection for applying different cuts on the 4 leptons invariant mass.

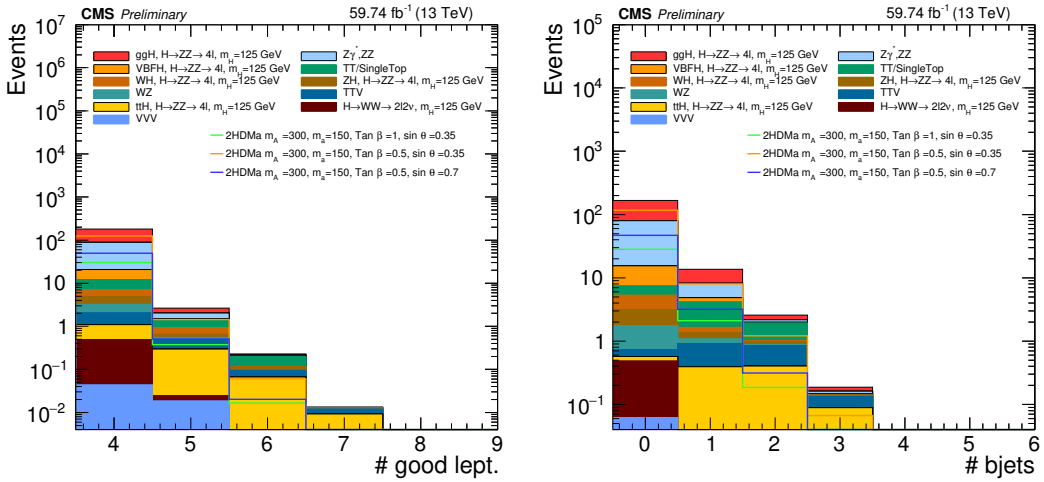


Figure 6.5: Number of selected good leptons (left) and b-tagged jets (right) for simulated background and signal events as derived after applying a cut on the Higgs mass window.

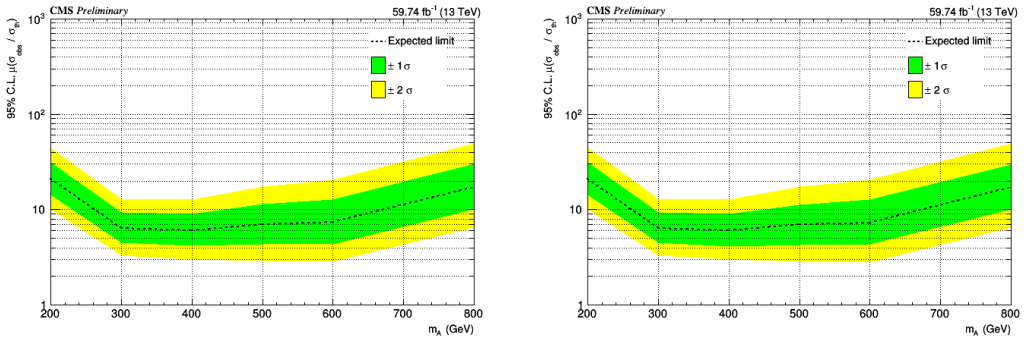


Figure 6.6: The 95 % CL upper limit on the signal strength for the combination of the three final states (4μ, 4e, $2e2\mu$) as a function of the heavy pesudoscalar mass "A" without applying cut on the number of jets (left) and after applying cut on number of b jets < 1 (right).

- **Number of b jets**: A cut on the b tagged jet multiplicity in the event could be useful in the rejection of background processes such as $t\bar{t} + V$ and the associated production of Higgs with $t\bar{t}$ as shown in Figure 6.5 right. We studied the effect of applying a cut on the number of b-jets in the event by measuring the cross section upper limit for different signal points and we found that the cut has no effect as shown in Figure 6.6. Moreover the $t\bar{t} + V$ and $t\bar{t}H$ backgrounds does not represent an overwhelming source of background for this analysis, beside the fact that the jet is affected by large uncertainties, so we decided to not use this cut since it has no effect in the analysis.

- $\Delta\phi(H, p_T^{miss})$: From the event topology one could expect that the dark matter particle would be boosted against the Higgs boson and they will be emitted back to back having a large opening angle between them. Figure 6.7 left shows the azimuthal angle difference between the four-lepton system and the p_T^{miss} for different signal points superimposed over stacked backgrounds; this distribution is quite flat and applying any cut on this variable leads to suppress a large fraction of the signal up to ≈ 50 % as shown in Figure 6.8 left.

- **Transverse mass** M_T: The information of the four-lepton invariant mass and the p_T^{miss} is combined with the definition of the transverse mass:

$$M_T(4l + p_T^{miss}) = \sqrt{m_{4l}^2 + 2 \times p_T^{4l} \times p_T^{miss} - \overrightarrow{p}_T^{4l} \cdot \overrightarrow{p}_T^{miss}} \tag{6.1}$$

  where m_{4l} and p_T^{4l} are is the invariant mass and the transverse momentum of the four lepton candidate system, respectively, and $\overrightarrow{p}_T^{4l}$ is the vector of the four lepton momentum projected on the transverse plane. The distribution of this variable is shown in Figure 6.7 right, the signal distributions are superimposed over the background distribution; introducing any cut on the M_T would lead to suppressing a large fraction of the signal as shown in Figure 6.8 right, moreover applying the same cut has different effect on different signal samples.

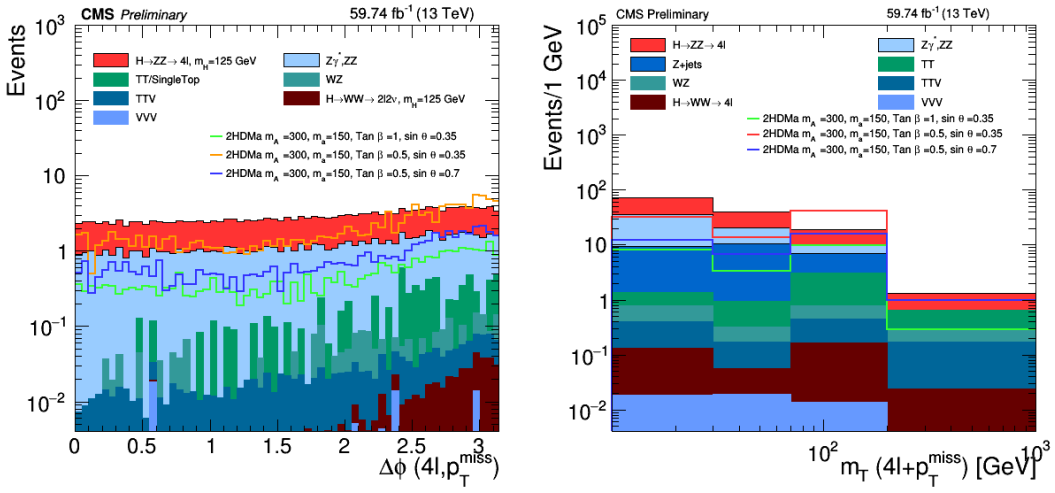


Figure 6.7: Distribution of the azimuthal angle difference between the four-lepton system and the p_T^{miss} (left) and the transverse mass of the four-lepton system and p_T^{miss} (right) for simulated background and signal events as derived after applying a cut on the Higgs mass window.

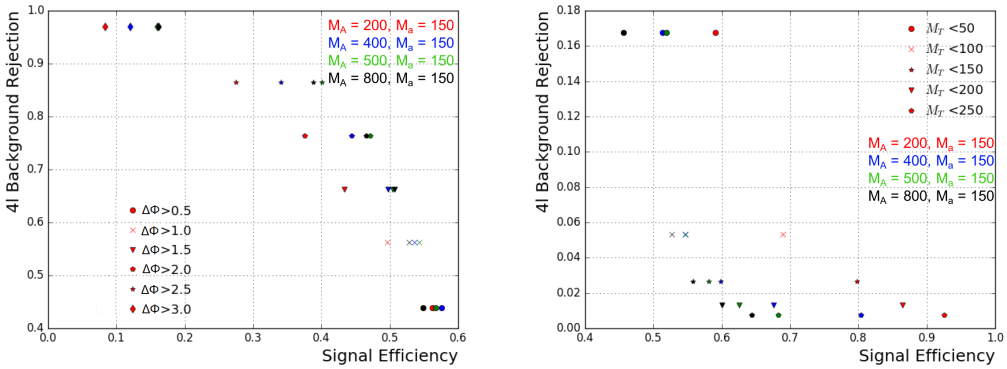


Figure 6.8: Signal efficiency vs background rejection for applying different cuts on the $\Delta\phi$ between the four-lepton system and the p_T^{miss} (left) and applying different cuts on the transverse mass M_T (right).

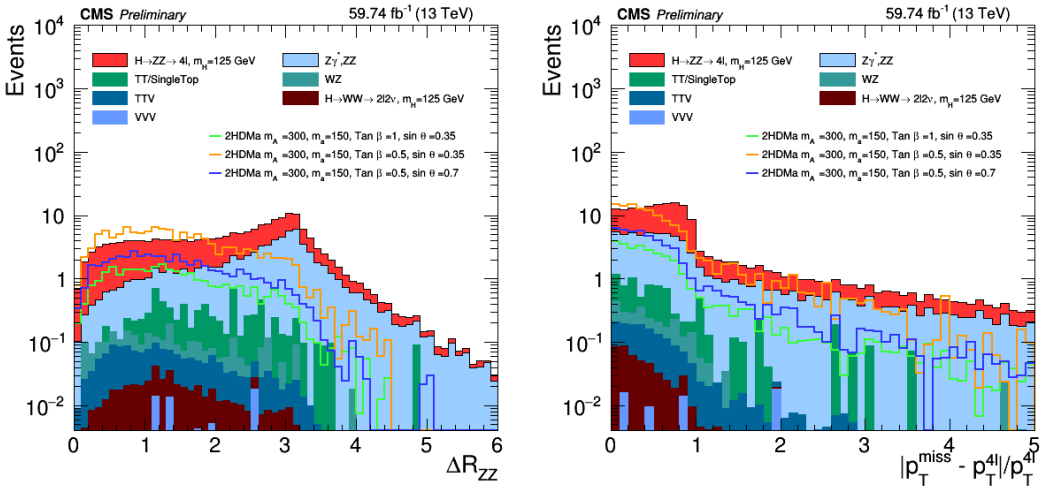


Figure 6.9: Distribution of the angular distance between the two Z's boson coming from Higgs decay (left) and the $|\frac{p_T^{miss}-p_T^{4l}}{p_T^{4l}}|$ distribution (right) for simulated background and signal events as derived after applying a cut on the Higgs mass window.

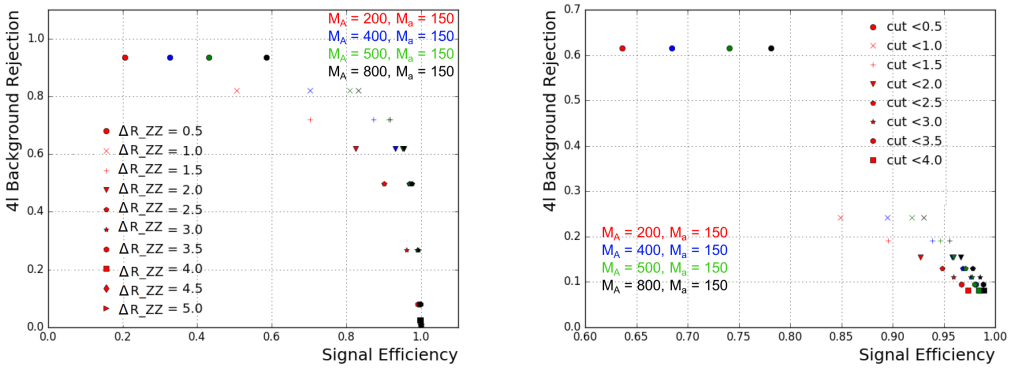


Figure 6.10: Signal efficiency vrs. background rejection for applying different cuts on the ΔR between the two Z's coming from the Higgs decay (left) and cuts on $|\frac{p_T^{miss}-p_T^{4l}}{p_T^{4l}}|$ quantity (right).

- $\Delta R(Z_1, Z_2)$: We studied the angular distance between the two Z's coming from the Higgs decay as shown in Figure 6.9 left, the purpose of this cut is to check if the mono-Higgs and SM Higgs samples having different kinematics in case that the Higgs and p_T^{miss} are emitted back to back. Applying different cuts on the $\Delta R(Z_1, Z_2)$ could remove large fraction of background events while preserving high signal efficiency, but has different effect on different signal samples as shown in Figure 6.10 (left). Moreover, applying different cuts on the $\Delta R(Z_1, Z_2)$ and measuring the cross section upper limit for different signal points doesn't increase much the sensitivity of the analysis as shown in Figure 6.11.

- $|\frac{p_T^{miss}-p_T^{4l}}{p_T^{4l}}|$: the magnitudes of the four lepton's transverse momentum p_T^{4l} and the missing transverse energy p_T^{miss} vectors are expected to have comparable magnitudes, Figure

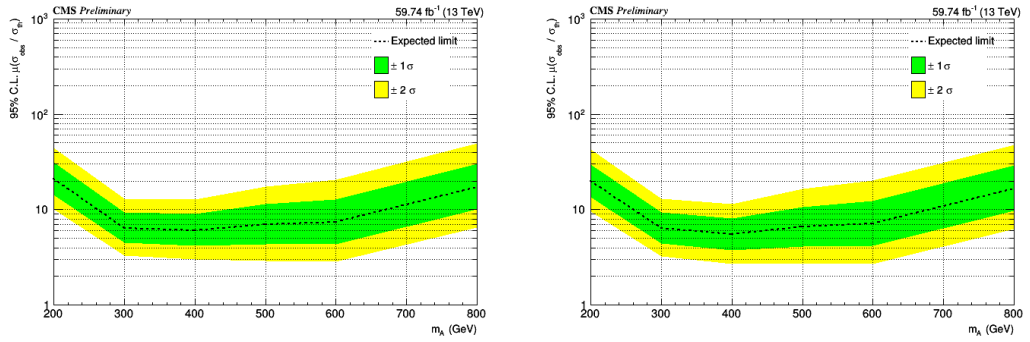


Figure 6.11: The 95 % CL upper limit on the signal strength for for the combination of the three final states (4μ, 4e, $2e2\mu$) as a function of the heavy pesudoscalar mass "A" without applying cut on $\Delta R(Z_1, Z_2)$ (left) and after applying cut on $\Delta R(Z_1, Z_2) < 2.5$ (right).

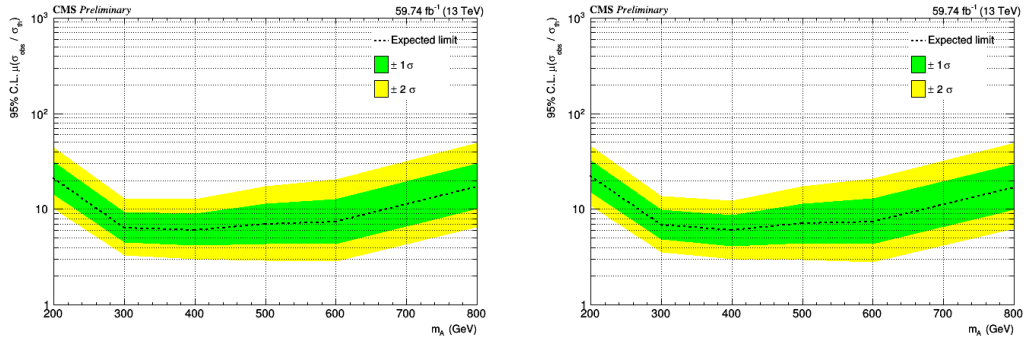


Figure 6.12: The 95 % CL upper limit on the signal strength for the combination of the three final states (4μ, 4e, $2e2\mu$) as a function of the heavy pesudoscalar mass "A" without applying cut on the quantity $|\frac{p_T^{miss} - p_T^{4l}}{p_T^{4l}}|$ (left) and after applying cut on $|\frac{p_T^{miss} - p_T^{4l}}{p_T^{4l}}| < 1.5$ (right).

6.9 shows the distribution of $|\frac{p_T^{miss} - p_T^{4l}}{p_T^{4l}}|$ for different backgrounds and different signal samples. We tried different cuts on this quantity, but applying these cuts doesn't remove much the background events and has different effect on different signal samples as shown in Figure 6.10 (right). So applying these cuts doesn't offer gain in the analysis sensitivity as shown in Figure 6.12.

Among all the variables studied above, the following selection cuts are applied to define the signal region:

- Higgs mass window: $|m_{4l} - 125| < 10$ GeV
- Number of good leptons: $N_{4l} = 4$

Those cuts preserve high signal efficiency for all the signal points while have a good background rejection.

6.3 Background

Background estimation is a crucial step in the analysis since without a reliable estimate of background rates, the predictions about new physics could be false and not reproducible. We need to ensure a precise evaluation of all possible background processes, specially when the analysis suffers from low cross section and hence a small number of expected events in the signal region. We consider two categories of backgrounds:

- Irreducible Background: All the processes that have exactly the same final state as the signal process (four leptons plus MET)
- Reducible Background: All the processes that don't have the same final state as the signal process, but due to the misconstruction of physics objects in the detector, can mimic the signal final state.

6.3.1 Irreducible Background

The background coming from the SM Higgs production (with the five main production modes and the Higgs decays to ZZ and WW) is considered irreducible for the analysis; especially the associated production of Higgs with W/Z boson where we have four real leptons coming from the same primary vertex plus MET. Another source of irreducible background is the triple boson production (ZZZ, WWZ, ZZW) where we have also four real leptons plus MET. Those backgrounds are reduced by the event selection cuts, but they survive to the final step of the analysis. All those backgrounds are estimated from MC simulation and normalized according to their cross sections times the branching ratio as shown in Table 5.7.

6.3.2 Reducible Background

This type of background raises from misidentification or mis-reconstruction of leptons and missing energy in the detector. If we have an ideal detector with ideal reconstruction algorithm, we wouldn't end up with this kind of background in the final step of the analysis. Reducible background event rate is decreased with event selection 6.1 that's designed to decrease as much as possible this kind of background while keeping high signal efficiency, but there is still a significant contribution left after the event selection that needs to be accounted for. The reducible background can arise from fake leptons and fake missing energy besides real objects to have 4 leptons plus missing energy. The sources of fake objects are light flavored hadrons mis-reconstructed as lepton, lepton coming from the decay of heavy flavor jets, photon conversion and fake missing energy shown in section 5.3.4. We can divide this background into two categories:

- ZZ background which is estimated from MC simulation.

- Z+jets/γ, WZ, $t\bar{t}$, W+jets and WW+jets backgrounds which are estimated from data by Fake rate method which will be described in details in section 6.4.

6.3.2.1 ZZ background Modeling

The Non resonant ZZ background comes from $q\bar{q}$ annihilation and gluon gluon fusion as shown in Figure 5.7. This background is estimated from MC but the simulation samples have some missing higher order corrections. To take into account those, the k-factors are applied for the samples.

- $q\bar{q} \rightarrow ZZ$: This background is generated at NLO, while the fully differential cross section has been computed at NNLO [200] which are not yet available in any event generator. Therefore NNLO/NLO k-factors is applied to the sample. The NNLO/NLO k-factors are applied in the analysis differently as a function of M_{ZZ}. Besides, a NLO electroweak correction is applied to the background sample depending on the initial state quark flavor and kinematics in the region $M_{ZZ} > 2M_{Z^0}$ for which is computed. This correction is not important for the thesis as our signal region is defined with a mass window around the SM Higgs mass peak.

- $gg \rightarrow ZZ$: This background process is generated at LO accuracy with MCFM 7.0 [185] generator. Although no exact calculation exists beyond the LO, it has been shown in [201] that the soft collinear approximation is able to describe the background cross section and the interference term at NNLO. Further calculations also show that the k-factors are very similar at NLO for the $gg \rightarrow H$ sample and $gg \rightarrow ZZ$ sample [202] and at NNLO for the $gg \rightarrow H$ sample and interference terms [203]. Therefore, the same k-factor is used for the Higgs sample and ZZ sample [204].

6.4 Reducible Background Estimation

The backgrounds such as Z+jets/γ, WZ, $t\bar{t}$, W+jets and WW+jets are called together "Z+X" and estimated from data rather than MC simulation since some of those samples don't have enough statistics to populate the signal region. Also the jet modeling for those samples is not simulated accurate enough. This background originates from processes which contain one or more than one non-prompt leptons in the final state. The main sources of non-prompt leptons are non-isolated leptons coming from decays of heavy flavor jets, light flavor jets mis-reconstructed as leptons and electrons from photon conversions. These leptons are called fake leptons. We will consider a "fake lepton" any jet misreconstructed as a lepton and any lepton originating from a heavy meson decay. Similarly, any electron originating from a photon conversion will be considered a "fake electron". The rate of these background processes is called fake ratios or fake rates (FR). The background estimation will be done in three steps:

- measuring the lepton's fake rates.
- building Control Regions (CR) orthogonal to the signal region.
- applying the measured fake rate in the control regions to extract the final background yield expected in the signal region.

6.4.1 Fake Rate

The electron (f_e) and muon fake rates (f_μ) are measured as the probability of leptons passing the loose selection criteria to also pass the final selection criteria (as already defined in Section 5.3.1, 5.3.2). To measure the leptons fake rates we build Z+L control region which expected to be dominated by Z boson plus fake lepton (L). This fake lepton is used as a probe for measuring the fake rates. The fake rate is calculated as the fraction of events where the loose lepton passes the tight selection criteria.

$$FR = \frac{N_{pass}}{N_{total}} \tag{6.2}$$

The events that populate the Z+L control region should satisfy the following criteria:

- build a Z boson from two opposite charge same flavor leptons with $p_T > 20/10$ GeV passing the tight selection criteria;
- the FSR photon is included in Z reconstruction if it exists;
- the reconstructed invariant mass (M_{ll}) should be close to nominal Z mass within 7 GeV, $|M_{ll} - M_Z| < 7$ GeV, to reduce the contribution from photon asymmetric conversions populating low masses. Figure 6.13 shows the lepton fake rate measured as a function of M_{ll} for 2017 data set as example. The uncorrected (corrected) measurements means the fake rate is measured in data before (after) subtracting the fake rate contribution from WZ background sample. The fake rate is almost stable in the region around M_Z with 7 GeV;
- the MET in the event is required to be $p_T^{miss} < 25$ GeV to suppress the contribution from WZ processes which have three real leptons in the final state. Also the contribution to the fake rate estimated from WZ simulated sample is subtracted from the measured fake rate from data to remove any remaining contribution from WZ background;
- exactly one lepton passing the loose selection criteria (kinematic and impact parameter cuts); this lepton is used as a probe for fake rate measurement;
- the invariant mass of this loose lepton and the opposite sign lepton from the reconstructed Z candidate should satisfy $m_{2l} > 4$ GeV to suppress QCD contamination;
- $\Delta R > 0.02$ between any of the three leptons.

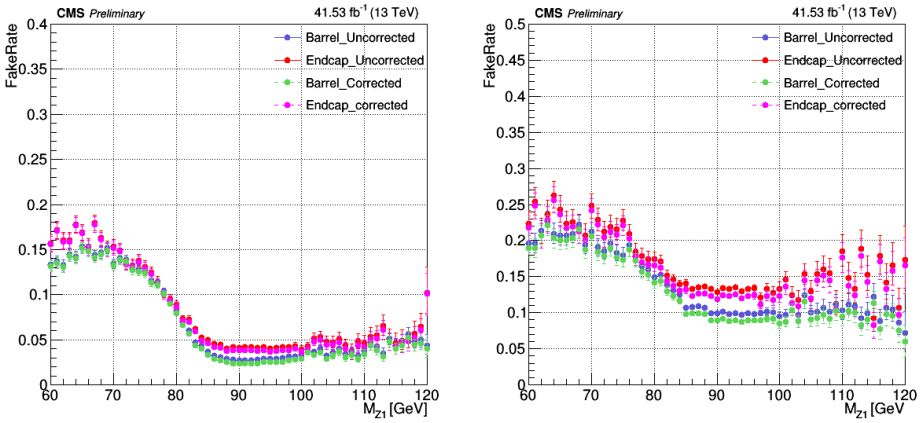


Figure 6.13: Electron (left) and Muon (right) fake rate measured as a function of the reconstructed Z mass in 2017 dataset before (uncorrected) and after (corrected) subtracting the contribution from WZ background sample.

The fake rate is measured in bins of loose lepton's p_T in barrel and endcap regions separately. Figure 6.14 shows the electron (top) and muon (bottom) fake rate as a function of loose lepton p_T for barrel and endcap regions in 2016 (left), 2017 (middle) and 2018 (right) datasets. The uncorrected (corrected) measurements are related to the fact that the fake rate is measured in data before (after) subtracting the fake rate contribution from the WZ background sample. The plots show that around 5 % - 10 % (5 % - 20 %) of fake electrons (muons) can pass the tight selection criteria used to define the signal region. Before using the fake rates to the next step, we studied the dependence of fake rate in the loose lepton charge as shown in Figure 6.15 for the electron fake rate as a function of p_T for positive (top) and negative (bottom) charge for 2016 (left), 2017 (middle) and 2018 (right) datasets. The case of the muon fake rate is shown in Figure 6.16 (top) for positive and negative (bottom) loose muon charge for 2016 (left), 2017 (middle) and 2018 (right) datasets. The plots show no dependence of fake rate on the lepton's charge.

Since our analysis search for high p_T^{miss} events, it is not convenient to measure the fake rate in regions with low p_T^{miss} ($p_T^{miss} < 25$ GeV) and apply to a region of high p_T^{miss}. So we studied if there is a dependence of the measured lepton's fake rate on the p_T^{miss} in the event, as shown in Figure 6.17 for electron (top) and muon (bottom) in 2016 (left), 2017 (middle) and 2018 (right) datasets. We observed a variation of the lepton fake rate as a function of p_T^{miss} in the high p_T^{miss} region after removing the WZ contribution . To take into account any dependence of fake rate on the measured p_T^{miss}, we computed the fake rate as function of lepton's p_T^l, η^l and p_T^{miss}. Figure 6.18 shows the electron fake rate map for barrel (left) and endcap (right) regions for 2016 (top), 2017 (middle) and 2018 (bottom) datasets. The muon fake rate map for barrel (left) and endcap (right) regions for 2016 (top), 2017 (middle) and 2018 (bottom) datasets is shown in figure 6.19. The fake rate values will be used to weight the events populating the control regions defined in the next section.

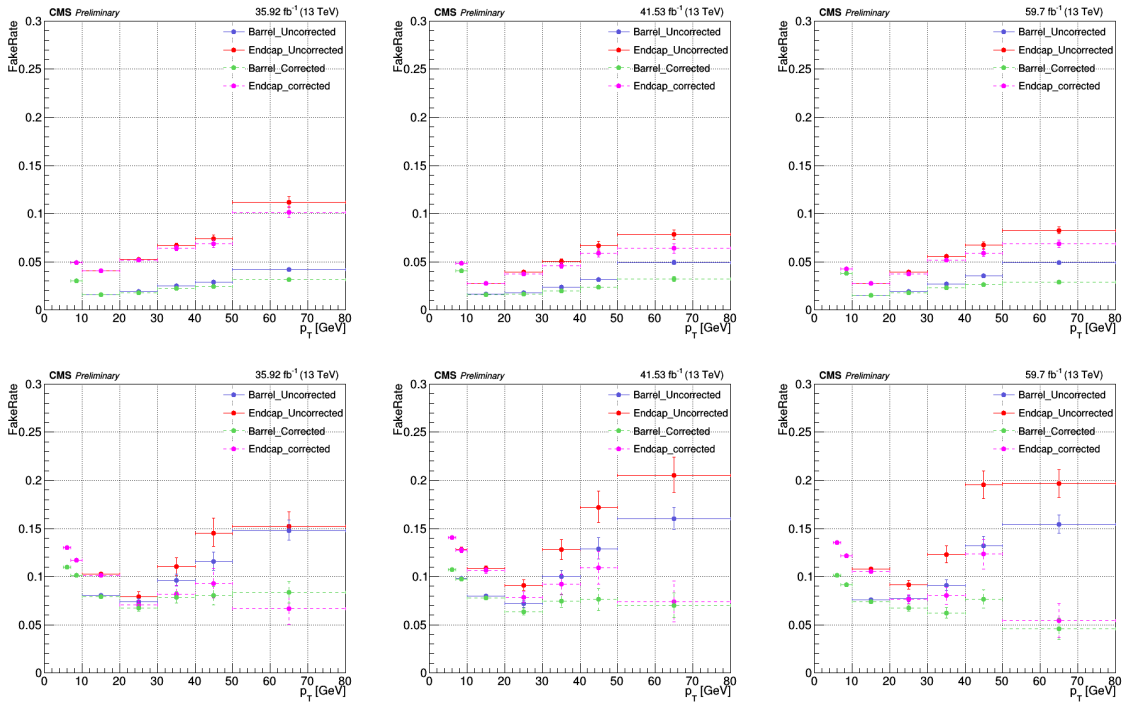


Figure 6.14: Electron (top) and Muon (bottom) fake rates measured in 2016 (left), 2017 (middle) and 2018 (right) as a function of loose lepton p_T for barrel and endcap regions.

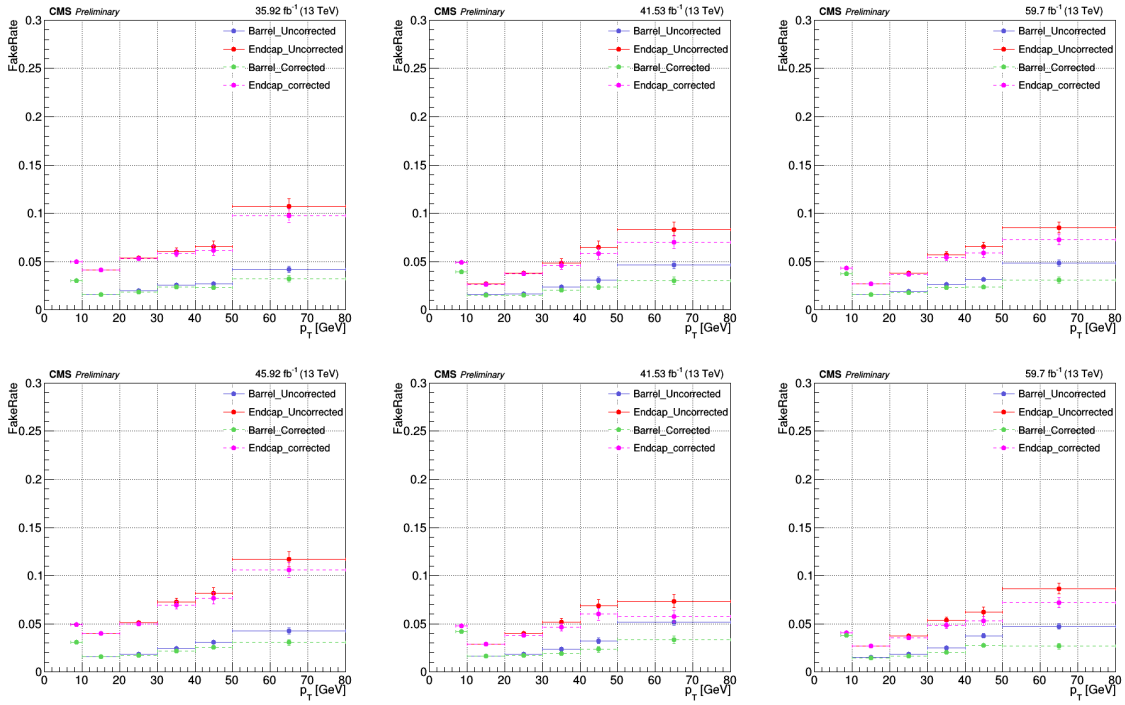


Figure 6.15: Electron fake rate as a function of p_T for positive (top) and negative (bottom) charge for 2016 (left), 2017 (middle) and 2018 (right) datasets.

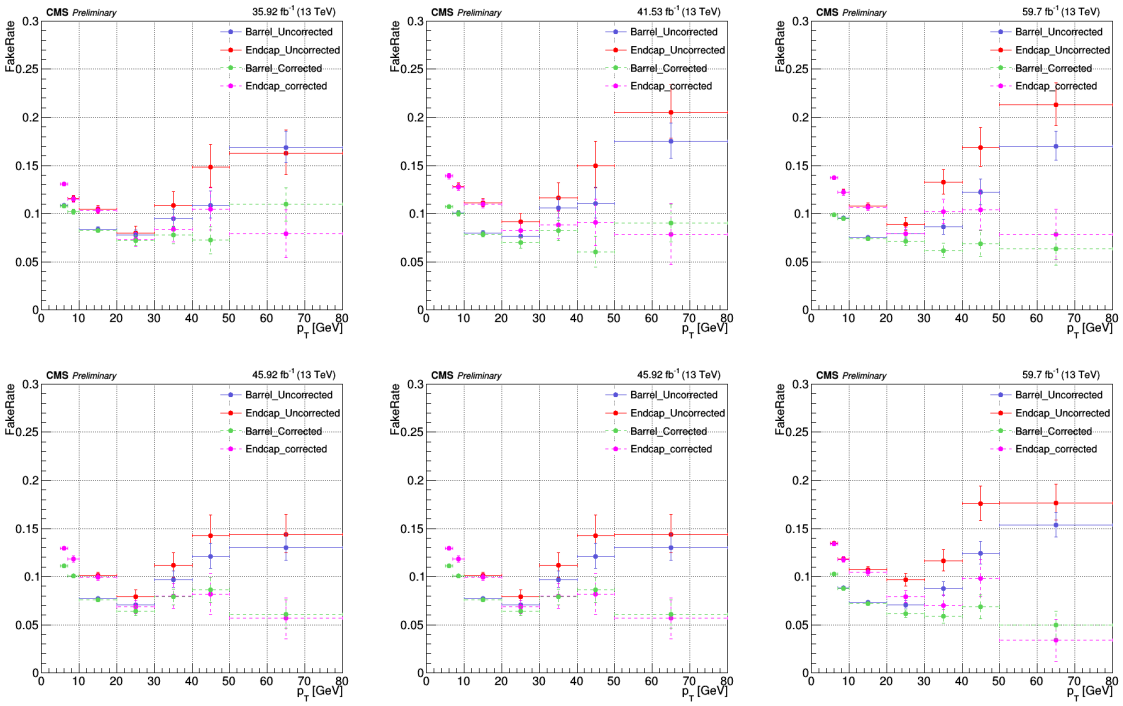


Figure 6.16: Muon fake rate as a function of p_T for positive (top) and negative (bottom) charge for 2016 (left), 2017 (middle) and 2018 (right) datasets

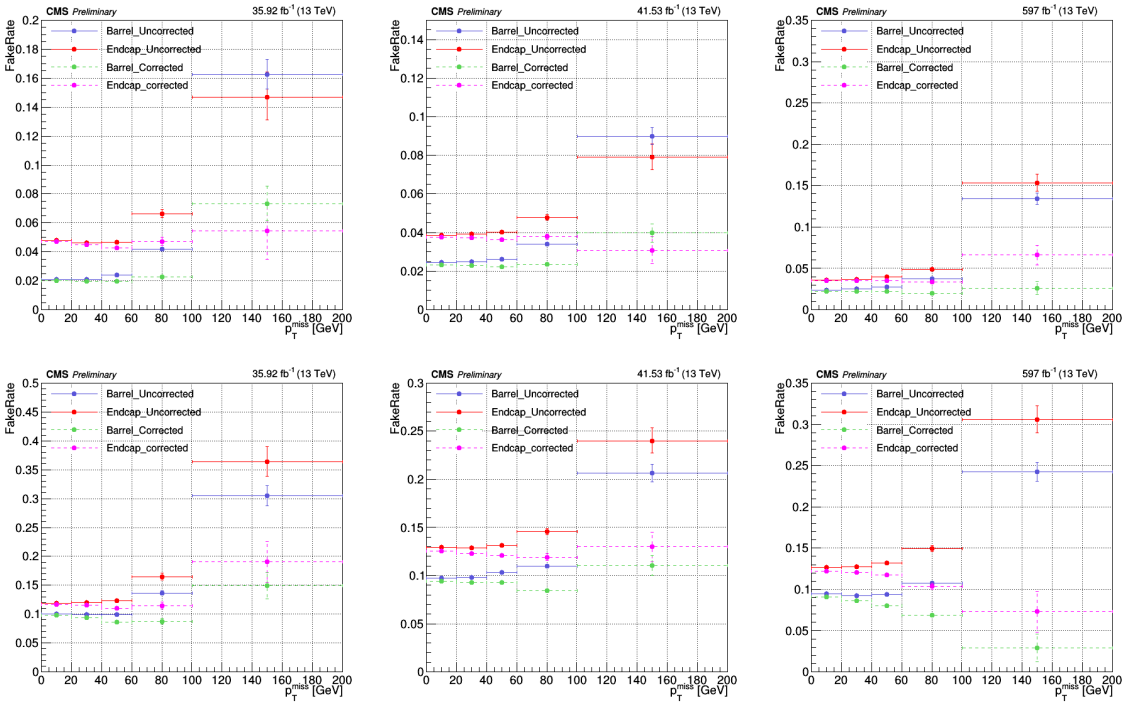


Figure 6.17: Electron (top) and Muon (bottom) fake rate as a function of p_T^{miss} for 2016 (left), 2017 (middle) and 2018 (right) datasets.

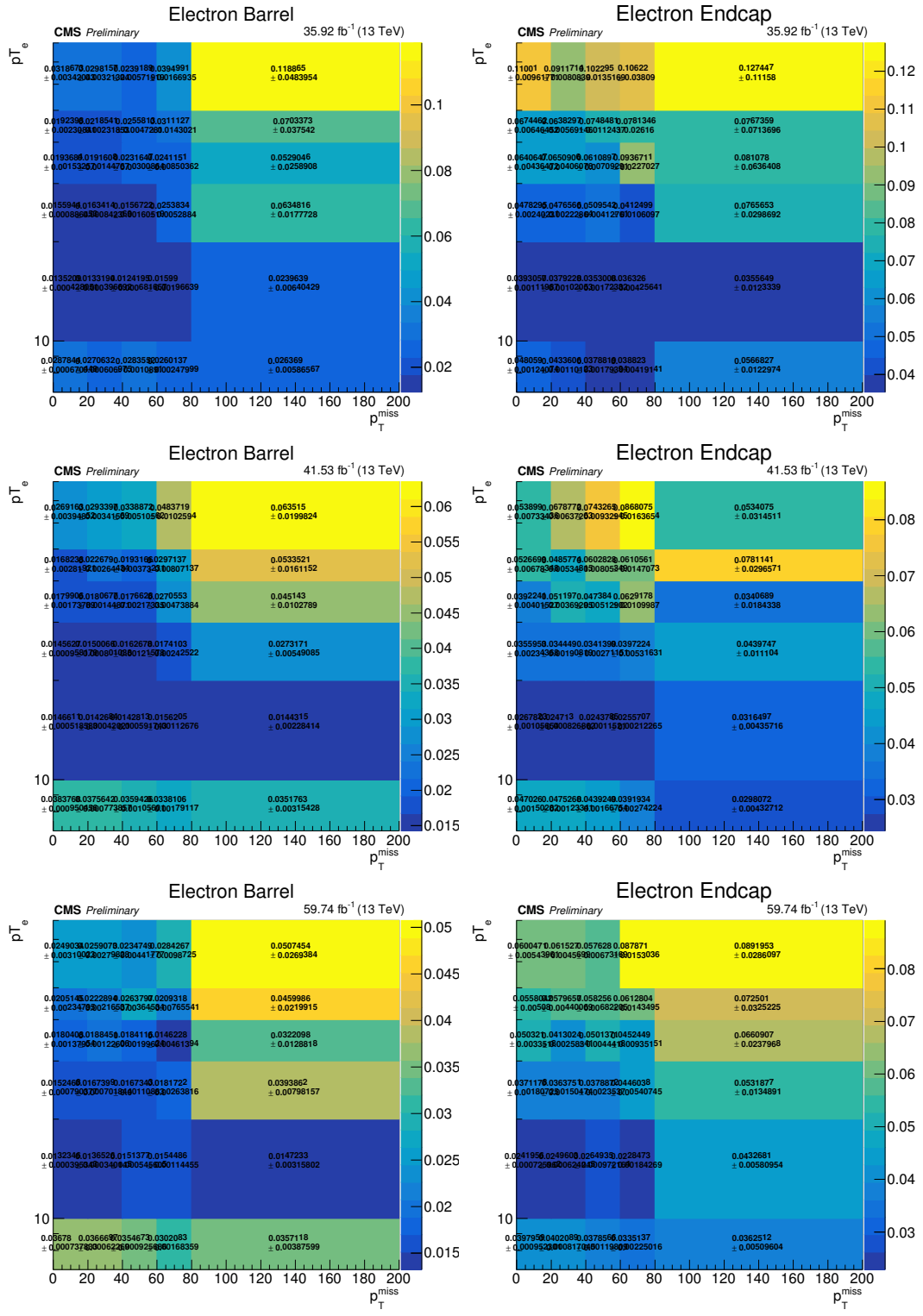


Figure 6.18: Electron fake rate measured as a function of electon's p_T^l and the p_T^{miss} in barrel (left) and endcap (right) regions for 2016 (top), 2017 (middle) and 2018 (bottom) datasets.

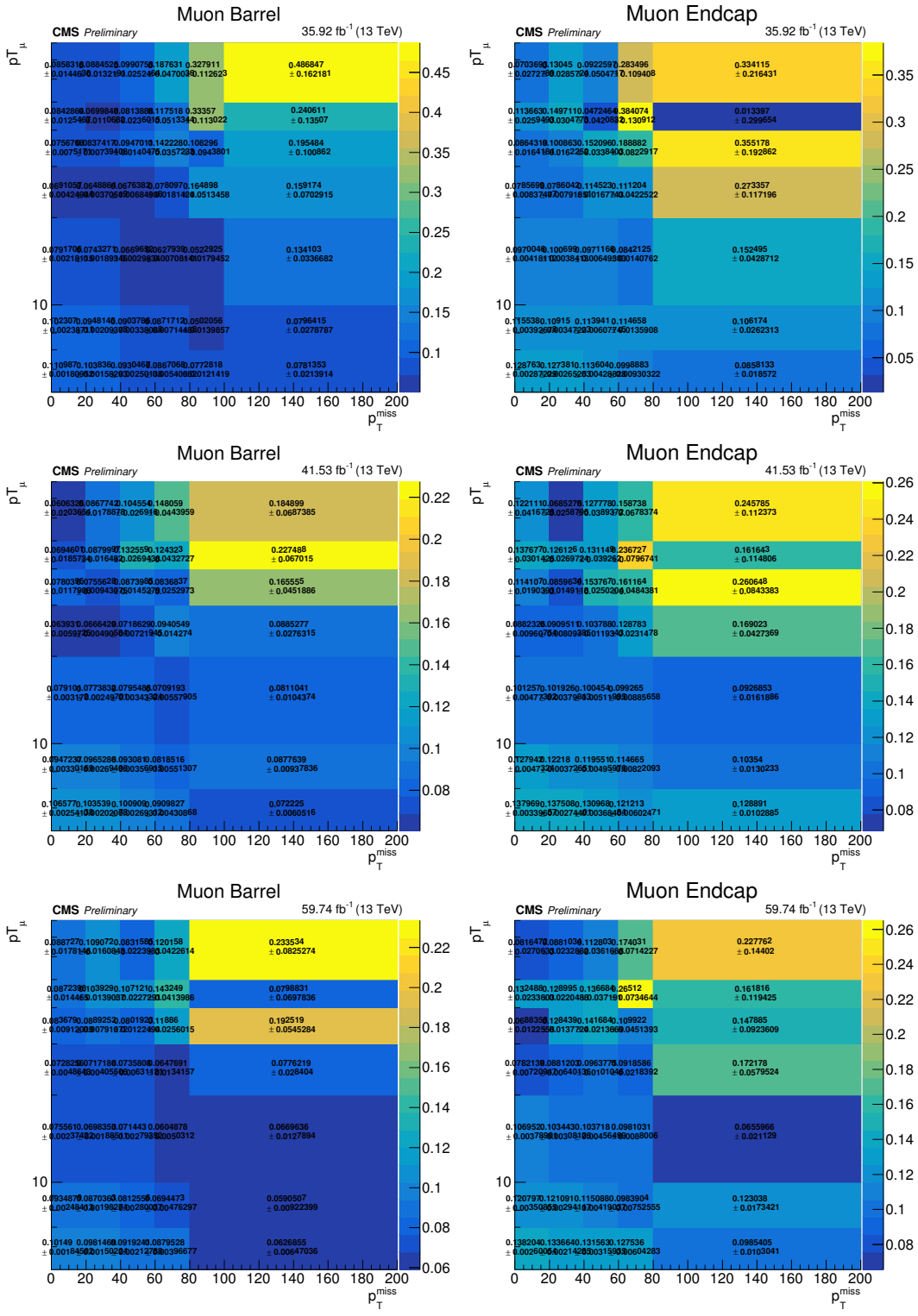


Figure 6.19: Muon fake rate measured as a function of muon's p_T^l and the p_T^{miss} in barrel (left) and endcap (right) regions for 2016 (top), 2017 (middle) and 2018 (bottom) datasets.

6.4.2 Background Control Regions

A control region is a region of phase space that is orthogonal to the signal region and has higher statistics than it. We define two control regions where a di-lepton passes the Z selection in event selections steps in section 6.1 and denoted as "Z_1", plus additional two leptons of the same flavor and opposite charge that pass the loose selection cuts. The events must satisfy all kinematic cuts applied for the Higgs and Mono-Higgs phase space selection (as described in section 6.1, 6.2).The two regions are called 2P2F and 3P1F regions where P denote that lepton "Pass" the final leptons' selection cuts, F denotes that lepton "Fail" either the identification or isolation cuts or both.

- **2P2F region**: This control region is obtained by reconstructing the Z boson from two opposite charge same flavor leptons passing tight selection cuts, "Z_1", with additional two opposite sign same flavor leptons passing the loose selection criteria "Z_2" and failing the identification or isolation cuts or both. This control region is expected to be populated with events that have two prompt leptons such as $Z + jets, t\bar{t}$ and $Z\gamma$.

- **3P1F region**" This control region is obtained by reconstructing Z_1 boson from two opposite charge same flavor leptons passing tight selection cuts and the second Z "Z_2" consist of one lepton passing the final selection cuts while the other one passing the loose selection cuts and failing the identification or isolation cuts or both. It is expected to be populated with the type of events that populate the 2P+2F region, where one of the fake leptons is reconstructed as a prompt lepton, as well as with WZ events that have three prompt leptons.

The control regions built in this way are enriched with fake leptons and are used to estimate the reducible background in the signal region. Figures 6.20, 6.21 show the four leptons invariant mass distribution for events populating the 2P2F and 3P1F control regions respectively for 2016 dataset as an example. The solid points represent the data collected in 2016 and the stacked distributions represent the background contribution expected from main processes. In 3P1F distributions, the magenta line represents the 2P2F events contributing in 3P1F region. The DY sample has limited statistics when requiring on shell Z. From data to simulation comparison, we can see that in the cases where Z_2 is reconstructed from electrons, the simulation can reasonably well describe the data, but it fails in the case where Z_2 is coming from muons. Those plots justify the reason to estimate the Z + X background from data rather than MC simulation.

6.4.3 Final background estimation

From the set of events obtained in the control regions, an inclusive number of reducible backgrounds in the signal region is obtained by taking into account the application of fake rates. The worflow followed to estimate the reducible backgrounds in the signal region by

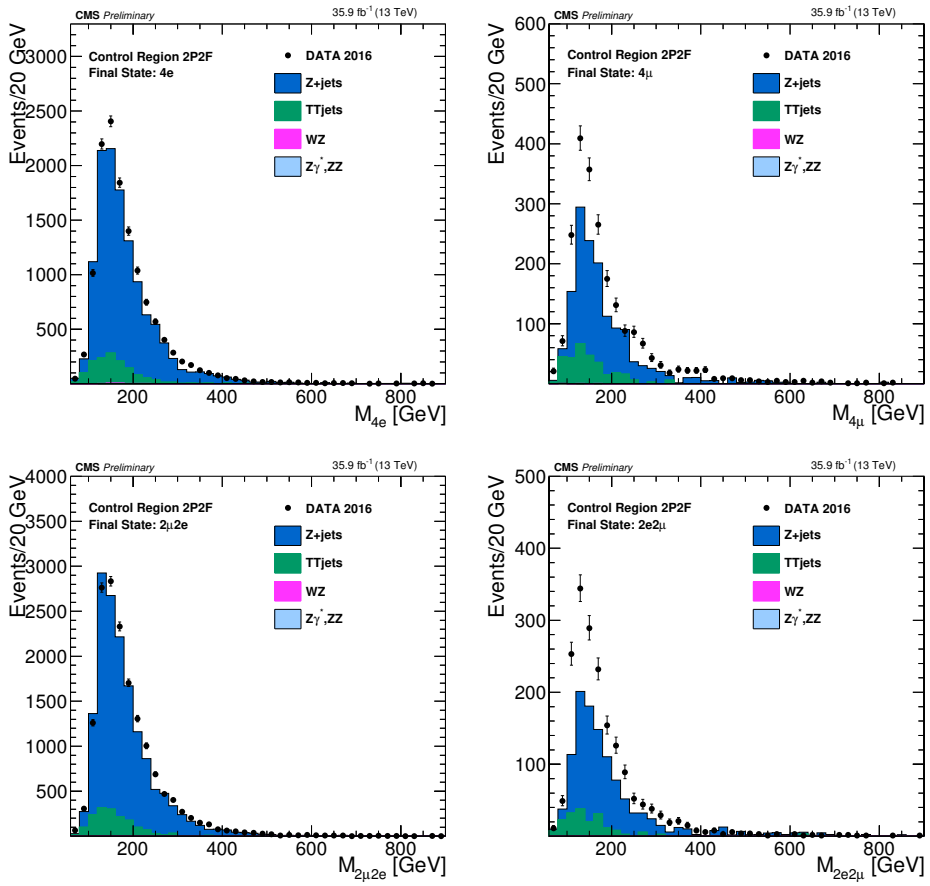


Figure 6.20: Invariant mass distribution of the events populated in the 2P+2F control sample in the 2016 dataset for 4e (top left), 4μ (top right), 2μ2e (bottom left) and 2e2μ (bottom right).

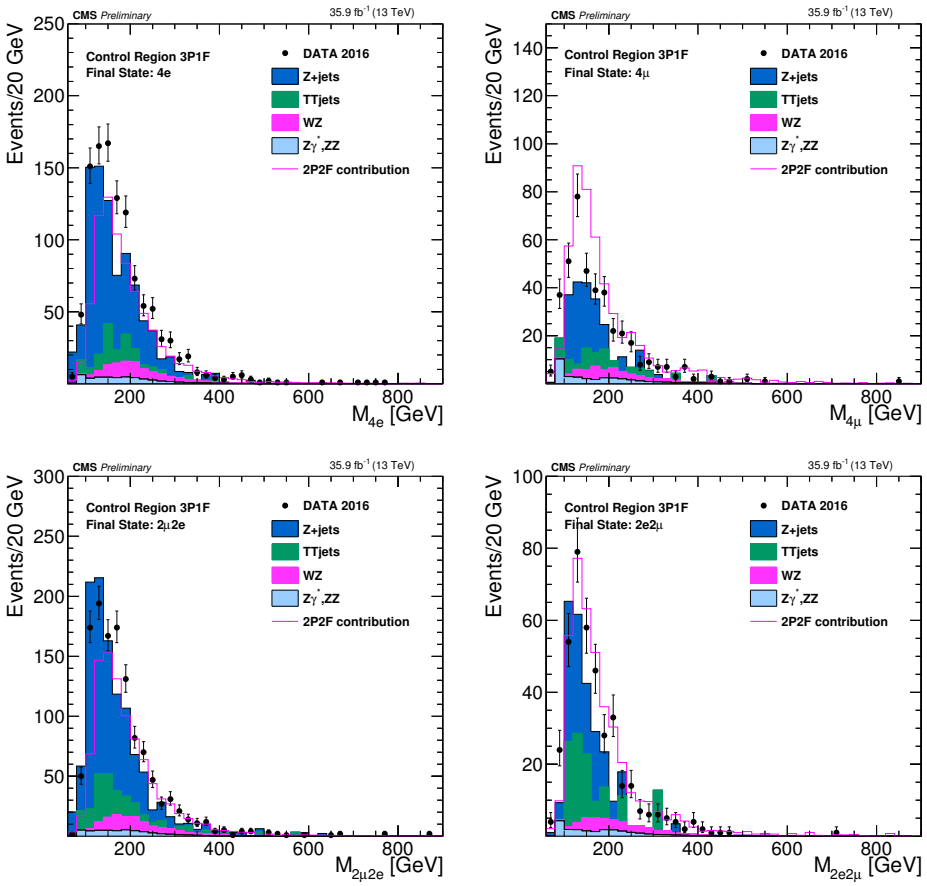


Figure 6.21: Invariant mass distribution of the events populated in the 3P+1F control sample in the 2016 dataset for 4e (top left), 4μ (top right), 2μ2e (bottom left) and 2e2μ (bottom right).

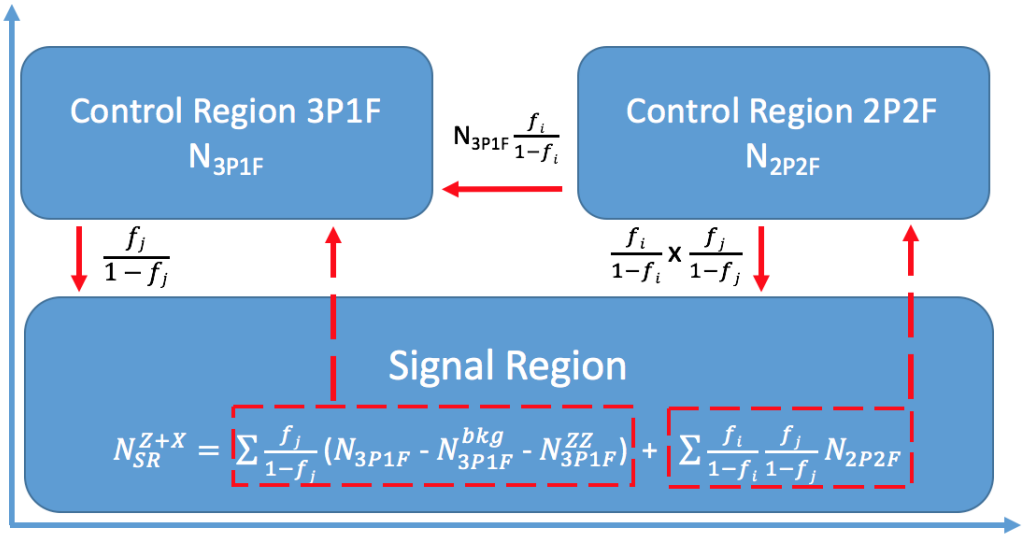


Figure 6.22: Diagram showing the strategy to estimate the Z+X background. P and F represent loose leptons which pass or fail final selections as described in the text. Dashed lines show the contribution of each CR in the final estimation equation.

choosing the CR and applying the measured FR to the CR, is shown in Figure 6.22. The event can be categorized as passing or failing the tight identification or isolation cuts as given in the following equation:

$$N_{total} = N_{pass} + N_{fail} \tag{6.3}$$

where: N_{pass} (N_{fail}) is the total number of events where the probe lepton pass (fails) the isolation and identification criteria.

$$FR = \frac{N_{pass}}{N_{total}} \tag{6.4}$$

From equations (6.3) and (6.4), we get

$$\frac{N_{pass}}{FR} = N_{pass} + N_{fail} \tag{6.5}$$

$$N^{pass} = (N_{pass} + N_{fail})FR \tag{6.6}$$

$$N_{pass} = \frac{FR}{1 - FR} N_{fail} \tag{6.7}$$

The estimation of the reducible background in the signal region N_{SR}^{Z+X} coming from the two categories 2P2F and 3P1F can be written as:

$$N_{SR}^{Z+X} = \sum \frac{f_i}{1 - f_i}(N_{3P1F} - N_{3P1F}^{bkg} - N_{3P1F}^{ZZ}) + \sum \frac{f_i}{1 - f_i}\frac{f_j}{1 - f_j} N_{2P2F} \tag{6.8}$$

where:

- N_{2P2F} represents the number of events obtained in the 2P2F control region. Each event in this region is weighted by $\frac{f_i}{1-f_i}\frac{f_j}{1-f_j}$ to get the contribution to the signal region, where f_i and f_j represent the fake rates for the third and the fourth leptons "loose leptons".

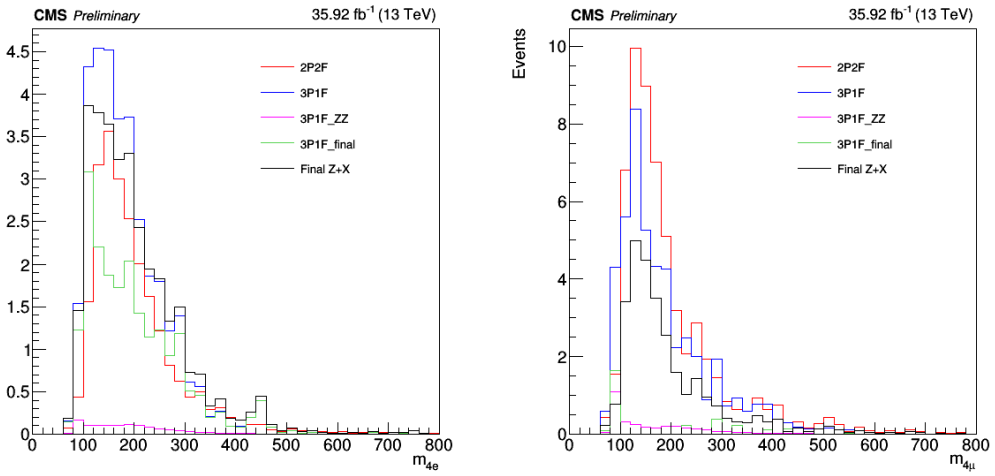


Figure 6.23: Invariant mass distribution for 4e (left) and 4μ (right) for all the components used in Z+X background estimation in 2016 dataset.

- N_{3P1F} represents the number of events obtained in the 3P1F control region. Each event is weighted by $\frac{f_i}{1-f_i}$ to get the contribution in the signal region, where f_i represents the fake rate of the loose lepton.
- N^{bkg}_{3P1F} represents the number of events from 2P2F and contaminating the 3P1F control region. This term is subtracted from N_{3P1F} to avoid the double counting of the event between the two control region (2P2F & 3P1F). This term is calculated as $2 \times \sum \frac{f_i}{1-f_i} \frac{f_j}{1-f_j} N_{2P2F}$.
- N^{ZZ}_{3P1F} represents the number of ZZ events contributing to 3P1F control region which arises from events where a signal lepton fails identification or isolation requirements. Since we rely on MC simulation of ZZ sample, this contribution is subtracted from N_{3P1F} to avoid the double estimation of ZZ background from data and MC simulation. The ZZ background could have also a contribution to 2P2F region where two signal leptons fail identification or isolation requirements , but this probability is very small , so we neglect this term in equation 6.8.

An example of Z+X background estimation together with all the components used in the calculation for 4μ and 4e channels is shown in Figure 6.23. The final value of Z+X background is estimated at the Higgs step and in the Mono-Higgs phase space after adding the two additional selection cuts used to define the signal region. Table 6.2 reports the final value of Z+X contributing to the Higgs and Mono-Higgs signal region for the different channels and different years of data taken with the associated total uncertainties.

6.4.4 Uncertainties in Z+X background

One of the highest source of uncertainties in the analysis comes from the Z+X background estimation from data. This background is not the dominant background after the full selection

Table 6.2: Final estimation of the reducible background by fake rate method for different channels and different years of data taken.

Final Estimation	4e	4μ	$2e2\mu$
SM Higgs region			
2016	22.95 ± 9.9	29.2 ± 16.2	48.6 ± 15.7
2017	15.9 ± 6.2	30 ± 11.2	45.4 ± 12.7
2018	27.5 ± 10.4	44.8 ± 17.1	69.3 ± 18
Mono- Higgs region			
2016	2.7 ± 1.19	5.12 ± 2.25	8 ± 2.4
2017	2.13 ± 0.88	5.5 ± 2.1	7.5 ± 2.2
2018	3.3 ± 1.3	8.1 ± 2.6	11.24 ± 2.8

Table 6.3: Systematic uncertainty associated to the Z+X estimate for each final state in all three years [21].

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

hence it has a small impact on the final results. The source of the statistical uncertainty of the method arises from the limited size of the samples in the control regions where we measure and where we apply the fake ratios , which is typically in the range of 1-13%. On the other hand, the source of systematic uncertainty arises from the difference in composition of background processes in the Z+L region where we measure the fake rate and in the control regions where we apply the fake rates. This uncertainty is estimated in the following steps: first, by measuring the fake ratios for each individual background processes in the Z+ L region in simulation. Then we compute a weighted average of these individual fake ratios according to the exact background composition of the 2P2F region obtained from simulation. The difference between the reweighed fake ratio and the regular measured one is used as an estimate of the uncertainty in the measurement of the fake rates [21]. This uncertainty is of the order or 30 - 41 %, depending on the final state and each year of data taken. The uncertainty for each final state in all three years is reported in Table 6.3.

6.5 Physics observables

After introducing the full event selection, estimating the reducible background from data in control regions, in this section the data to MC simulation comparison is performed for different observables after the Higgs selection step. This phase space is quite close to the signal

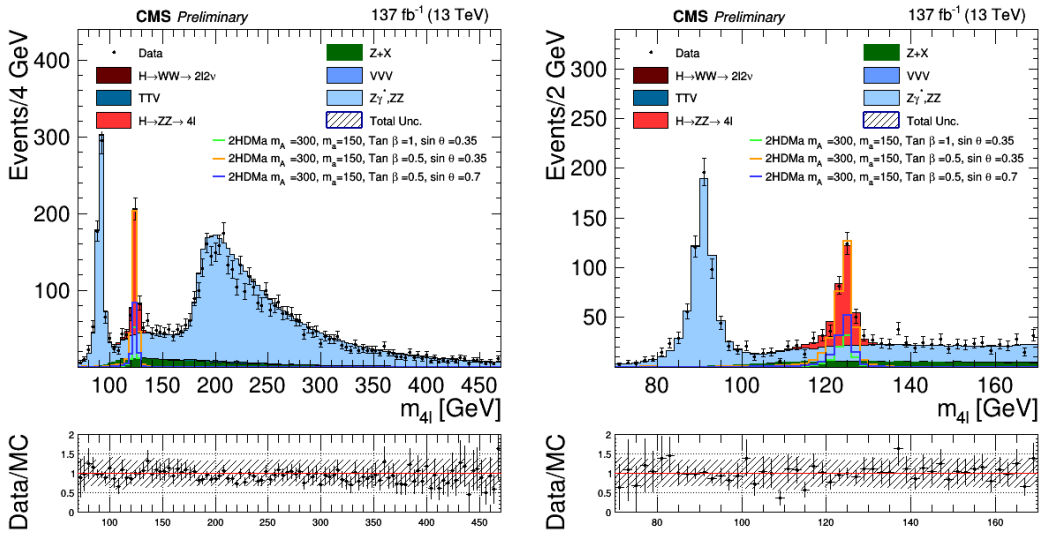


Figure 6.24: Distribution of the four-lepton reconstructed invariant mass m_{4l} in the full run II data for full mass range (left) and small range (right).

region, which is a subset of Higgs phase space. The four lepton invariant mass distribution is shown in Figure 6.24 for full Run II data in full mass range (left) and small mass range (right), while figure 6.25 shows the distribution for 2016 (top), 2017 (middle), 2018 (bottom) for the three channels (4μ, 4e, $2e2\mu$) summed together. The solid color distributions represent the different backgrounds stacked together after scaled by the cross section and the integrated luminosity, the colored lines represent different signal samples superimposed over the background distributions and scaled by 10, while the solid points represent the data collected every year. A boson with a mass around 125 GeV and compatible with the prediction of the SM Higgs boson is clearly observed in data and confirms the discovery made in 2012. The signal samples are peaked also around 125 GeV as expected which can be considered as a sanity check that we are to reconstruct the Higgs boson correctly. At this step the dominant background contributing are the single Higgs events and the non resonant ZZ production. There is a fair agreement between data and MC simulation within the statistical uncertainties, giving confidence in the careful usage of corrections and scale factors. Also, there is a fair agreement between data and MC simulation within the statistical uncertainties in the invariant mass distribution per channel, as shown in Figure 6.26.

The reconstructed invariant masses of lepton pairs selected as Z_1 and Z_2 are shown in Figure 6.27 for full Run II data. For the Z_1 distribution a peak observed around 90 GeV for data, background and signal which ensure the correct reconstruction of Z_1 with mass closest to the nominal Z mass given by PDG, while an off-shell mass is observed in Z_2 distribution for signal and Higgs background samples. Again, a good agreement between data and MC is

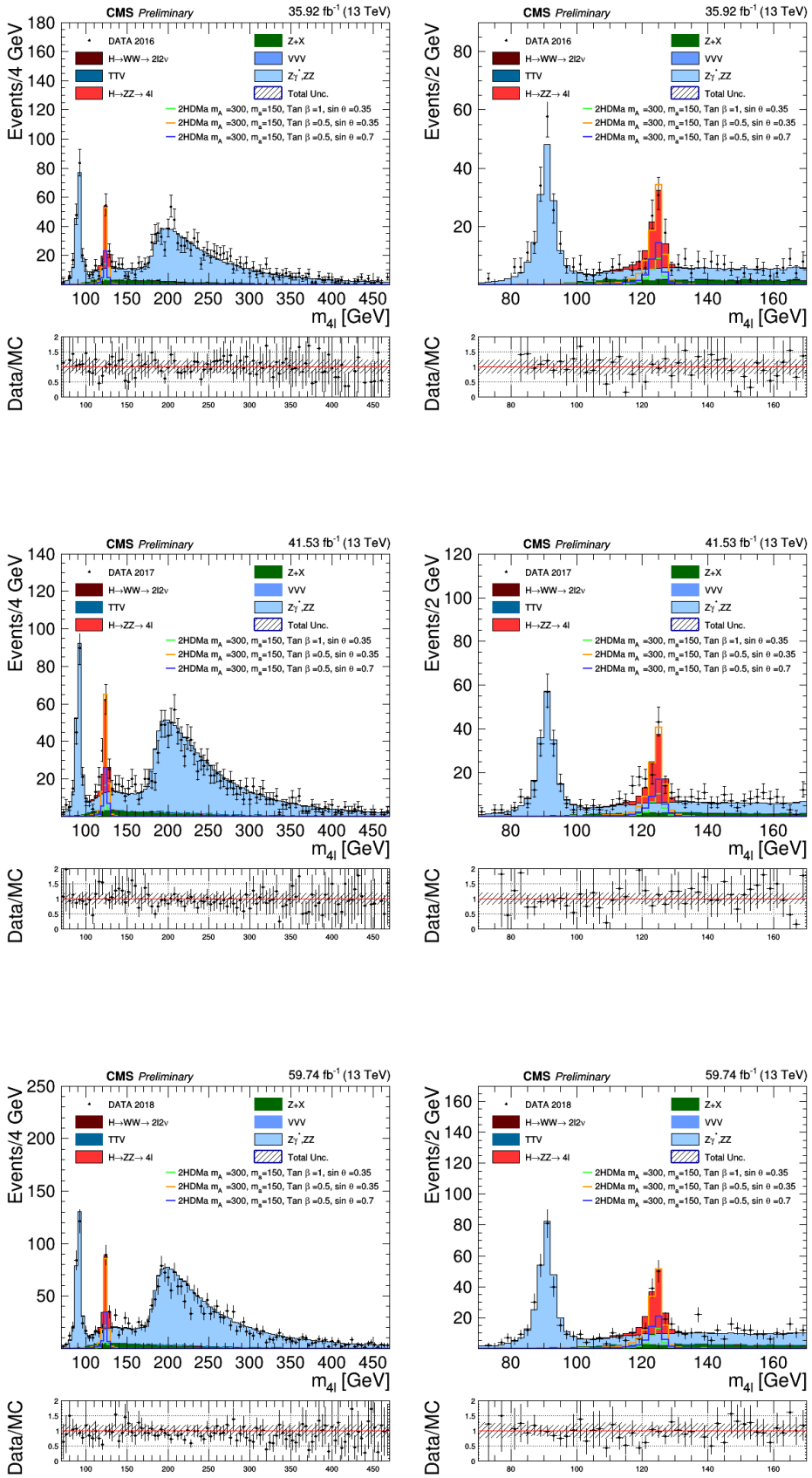


Figure 6.25: Distribution of the four-lepton reconstructed invariant mass m_{4l} in 2016 (top), 2017 (middle), 2018 (bottom) for full mass range (left) and small range (right).

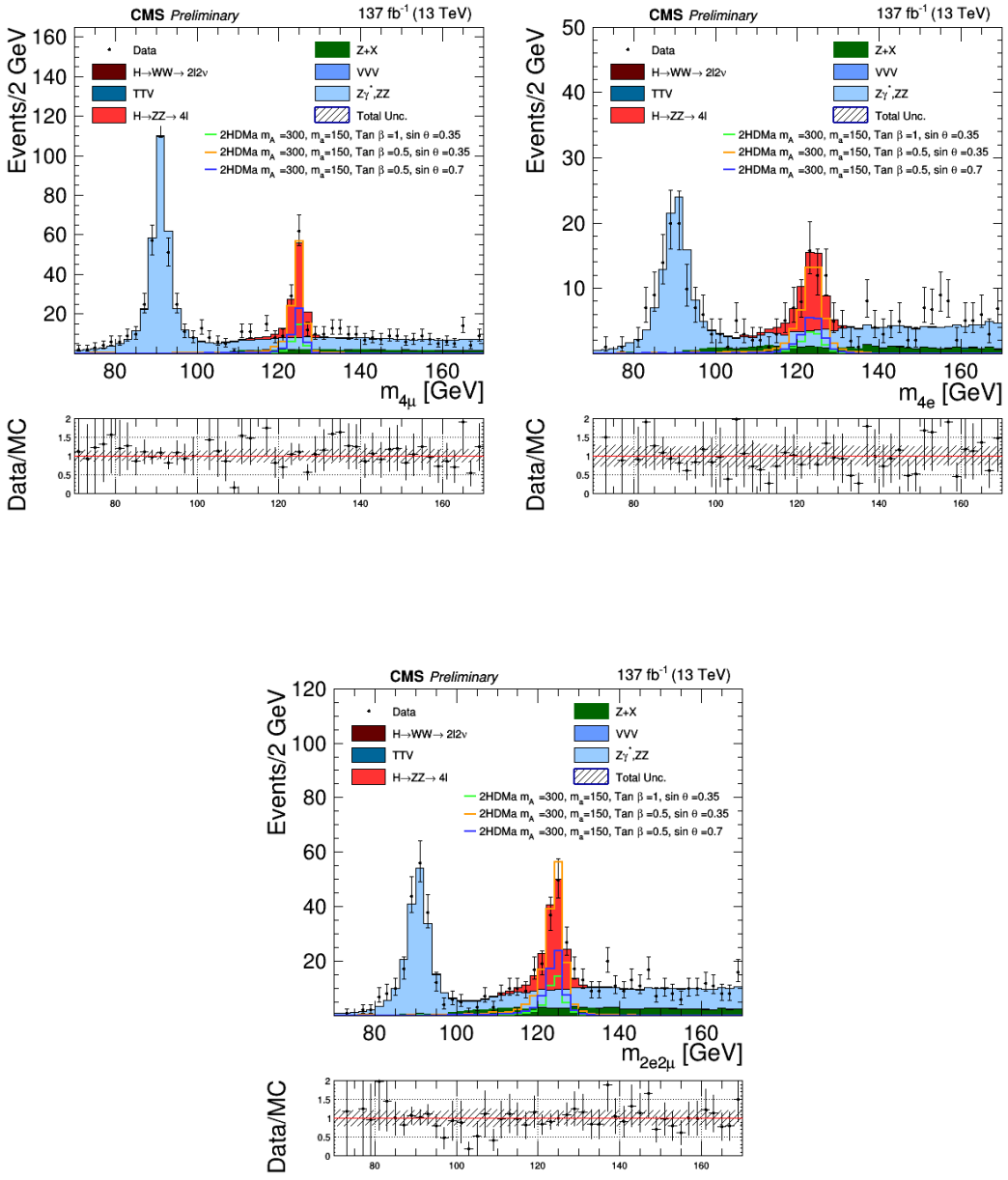


Figure 6.26: Distribution of the four-lepton reconstructed invariant mass m_{4l} for 4μ (top left), 4e (top right) and 2e2μ channel (bottom) for full run II data.

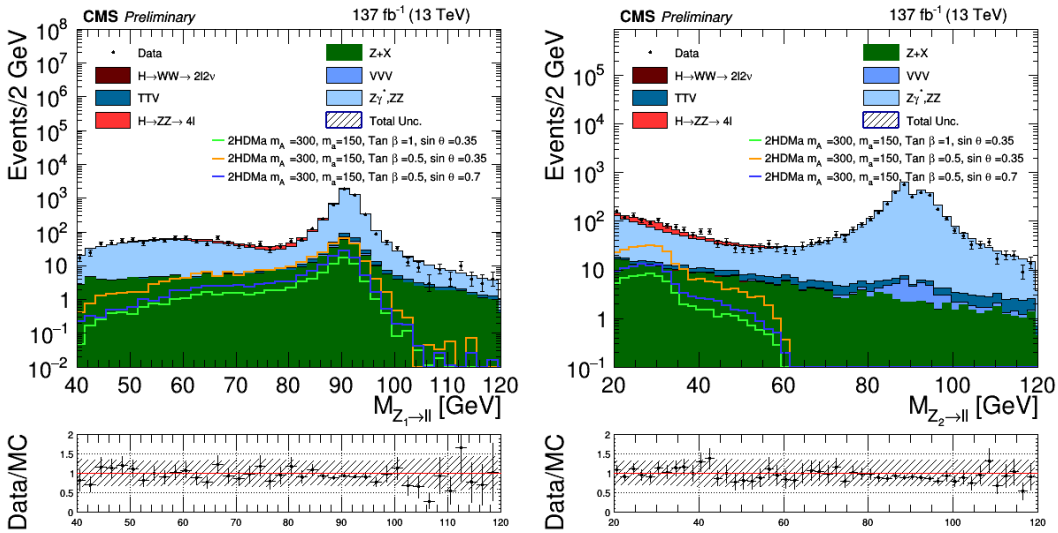


Figure 6.27: Distribution of the reconstructed invariant masses of lepton pairs selected as Z_1 (left) and Z_2 (right) for full Run II datasets.

observed.

The four leptons transverse momentum is shown in Figure 6.28 for 2016 (left), 2017 (middle) and 2018 (right). This plot shows a good agreement between data and backgrounds which ensures the careful usage of the corrections and leptons' scale factors.

The missing energy distribution p_T^{miss} which represents the main variable of the analysis is shown in Figure 6.29 for 2016 (top left), 2017 (top right), 2018 (bottom left) and for full Run II in (bottom right). Since the analysis is still "blind[2]" we restrict the data distribution till 200 GeV and blind the distribution after 200 GeV. The lower plot shows the ratio of the data and the sum of all the backgrounds, while the band corresponds to statistical and systematic uncertainties[3]. The data to MC ratio is quite good while the observed discrepancy is still within the total uncertainty. The main backgrounds contributing are $t\bar{t}V$, VVV, associated production of the Higgs with vector boson (ZH and WH), and non resonant ZZ background.

In order to check the full p_T^{miss} distribution, especially the tail, a control region has been built in the side-band of the four leptons invariant mass distribution, where $|m_{4l} - 125| > 10$ GeV. Figure 6.30 shows the p_T^{miss} distribution in the m_{4l} side band for 2016 (top left), 2017 (top right), 2018 (bottom). The number of events at large p_T^{miss} is small and the statistics is limited; the events contributing to the tail come mostly from $t\bar{t}V$, VVV, associated production of the Higgs with vector boson fusion (ZH and WH), and non resonant ZZ background. The

[2]Blinding means that the analysis validation is done by using only the MC simulation or data in a control regions away from the signal region, since looking at data in the signal region could introduce a bias when defining the event selection.

[3]Systematic uncertainties will be introduced in 7.3

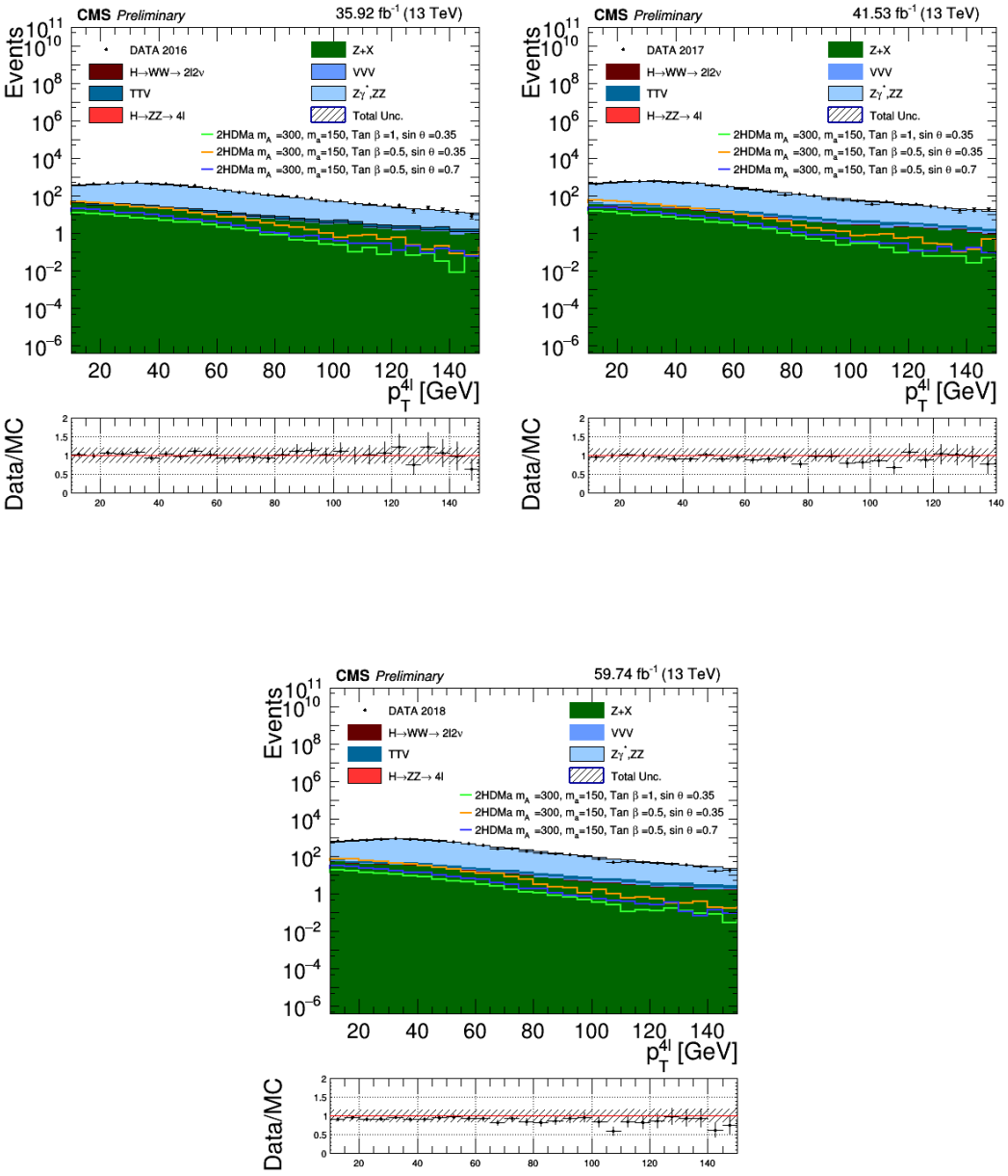


Figure 6.28: Distribution of four leptons's transverse momentum for 2016 (top left), 2017 (top right) and 2018 (bottom).

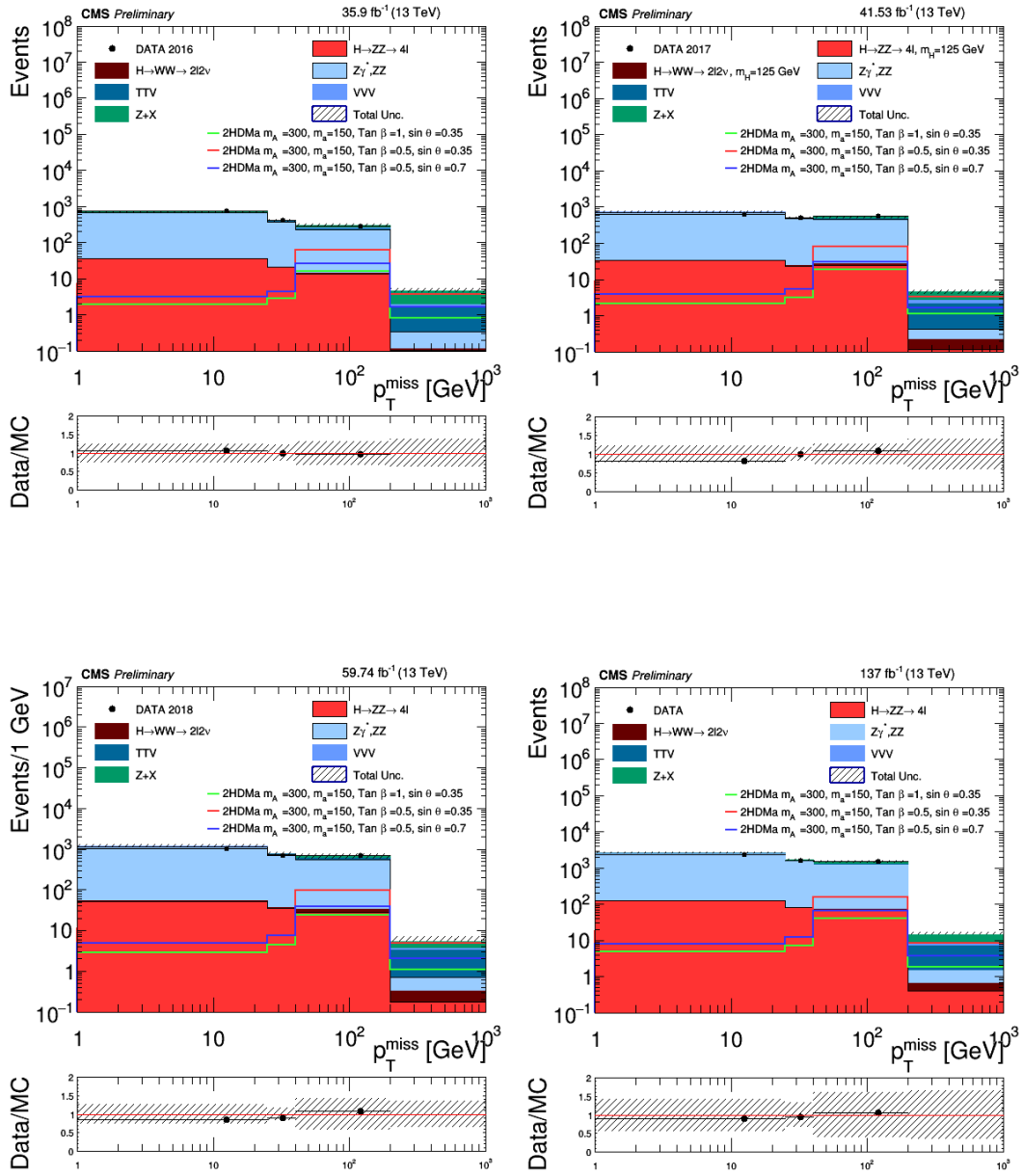


Figure 6.29: Missing transverse energy distribution p_T^{miss} after applying the SM Higgs selection step for 2016 (top left), 2017 (top right), 2018 (bottom left) and full Run II (bottom right).

observed discrepancy is still within the total uncertainty.

The event yields after each selection step for the three final states (4μ, 4e, $2e2\mu$) for each source of background processes, signal sample and data is presented in Table 6.4. The data counter is reported up to the Higgs selection step and blinded for the next steps. Only the statistical uncertainty is quoted in the table. The table shows how the event selection cuts bring the analysis to the signal phase space, requiring at least 2 Z bosons in the event is reducing significantly the contribution of DY+jets, WW, WZ and $\bar{t}t$ backgrounds. While the additional cuts required for MonoHiggs signal region remove most of $t\bar{t}V$ and VVV backgrounds and reduce the non resonant ZZ background while keeping the signal efficiency high. As it can be observed, the number of signal events observed after the full selection step is quite small compared to the main backgrounds (ZZ, Higgs), in order to overcome this limitation by increasing the signal to background separation we used the Multivariate approach (MVA), as will be discussed in next chapter.

This chapter described the selection process of this analysis, targeting the dark matter signal. A cut based approach was used with cuts on a set of observables to distinguish the mono-Higgs signal from different backgrounds and then used for the extraction of the results. The optimization and the definition of the signal region besides to the background estimation were detailed. Most of the discussed observables are used as inputs for the MultiVariate analysis as will be described in the next chapter. All the possible sources of systematic will be introduced in the next chapter. All these pieces will be connected together in a statistical analysis where a likelihood approach will be used to measure physics parameters of interest.

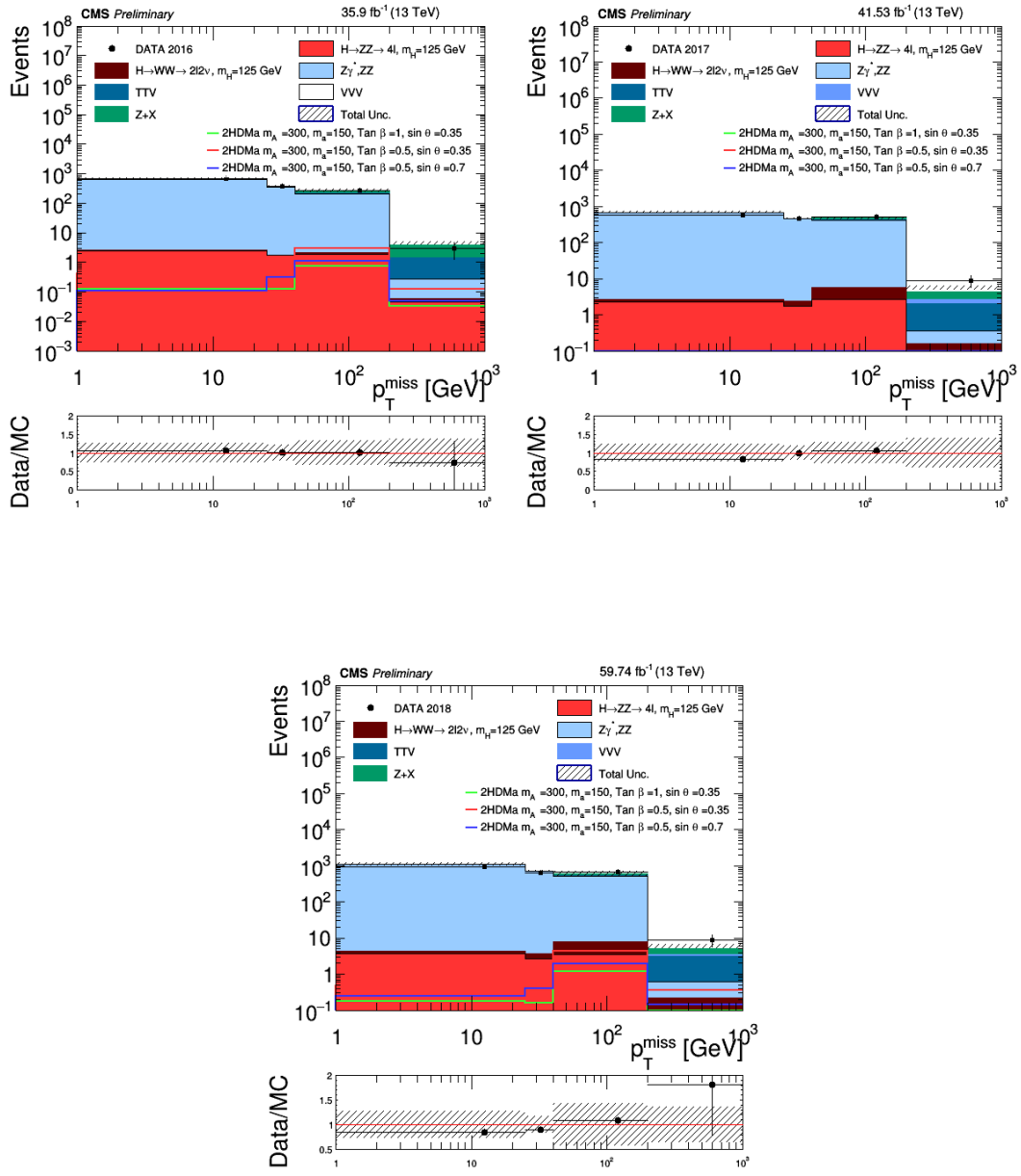


Figure 6.30: Missing transverse energy distribution p_T^{miss} in the side band of four leptons mass where $|m_{4l} - 125| > 10$ GeV for 2016 (top left), 2017 (top right), 2018 (bottom).

Table 6.4: Cut flow tables show the number of events passing the full selection for the 4μ (top), $4e$ (middle) and $2e2\mu$ (bottom) final states, as obtained from simulation for background and signal and from real data with a luminosity of $137\,fb^{-1}$. The signal sample correspond to 2HDM+a with $m_A = 300$ GeV and $m_a = 150$ GeV; only statistical errors are quoted.

Channel: 4μ	$Z\gamma^*, ZZ$	$Z + js$	$WZ, WW, W + js$	$t\bar{t}$	ttV	VVV	SMH	signal: $tan\beta = 0.5, sin\theta = 0.7$	signal $tan\beta = 1.0, sin\theta = 0.35$	*Obs.*
Initial	1.13e+05 ± 1.48e+01	3.10e+08 ± 3.66e+04	1.92e+09 ± 3.58e+05	1.11e+07 ± 1.41e+03	1.77e+04 ± 4.66e+00	5.21e+03 ± 9.55e+00	1.50e+03 ± 9.24e-01	6.70e+01 ± 3.33e-01	4.11e+01 ± 2.03e-01	2.02e+09
At least one Z	32066.47 ± 4.19	85877367.60 ± 825.23	4794301.1 ± 89.39	858062.72 ± 24.15	1662.74 ± 1.15	456.30 ± 0.98	519.95 ± 0.31	35.18 ± 0.13	2.17e+01 ± 8.18e-02	114024397
At least 2 Zs	7855.26 ± 1.03	156373.85 ± 461.47	38154.25 ± 7.15	9876.88 ± 31.94	756.69 ± 0.55	105.48 ± 0.99	231.47 ± 0.19	10.48 ± 0.04	6.43e+00 ± 2.43e-02	435763
$mz_i > 40$	1981.79 ± 0.26	178.94 ± 0.52	8.22 ± 0.01	14.81 ± 0.01	25.18 ± 0.10	9.29 ± 0.21	104.35 ± 0.13	4.85 ± 0.02	3.01e+00 ± 1.9e-02	2849
Choose ZZ pair	1779.38 ± 0.23	112.2 ± 0.33	7.28 ± 0.01	13.69 ± 0.01	24.76 ± 0.10	8.67 ± 0.19	97.40 ± 0.12	4.45 ± 0.02	2.80e+00 ± 1.7e-02	2690
$m_{4l} > 70$	1770.37 ± 0.23	108.33 ± 0.1	7.20 ± 0.001	13.52 ± 0.001	24.65 ± 0.01	8.23 ± 0.18	96.58 ± 0.11	4.41 ± 0.01	2.77e+00 ± 1.31e-02	-
$\lvert m_{4l} - 125\rvert \leq 10$, $N_l = 4$	59.96 ± 0.07	78.92 ± 0.009	1.21 ±0.0	1.29 ± 0.0	0.26 ± 0.0	0.04 ± 0.0	89.13 ± 0.1	4.20 ± 0.01	2.67e+00 ± 1.00e-02	-

Channel: $4e$	$Z\gamma^*, ZZ$	$Z + js$	$WZ, WW, W + js$	$t\bar{t}$	ttV	VVV	SMH	signal: $tan\beta = 0.5, sin\theta = 0.7$	signal $tan\beta = 1.0, sin\theta = 0.35$	*Obs.*
Initial	1.13e+05 ± 1.48e+01	3.10e+08 ± 3.66e+04	1.92e+09 ± 3.58e+05	1.11e+07 ± 1.41e+03	1.77e+04 ± 4.66e+00	5.21e+03 ± 9.55e+00	1.50e+03 ± 9.24e-01	6.70e+01 ± 3.33e-01	4.11e+01 ± 2.03e-01	1.66e+09
At least one Z	34600.63 ± 4.53	81608609.06 ± 9635	393195.78 ± 73.3	971661.11 ± 43.82	1975.30 ± 1.17	661.38 ± 1.25	560.12 ± 0.31	24.90 ± 0.11	1.51e+01 ± 6.61e-02	62178033
At least 2 Zs	5073.06 ± 0.42	146822.67 ± 16.79	13388.06 ± 2.49	11276.57 ± 0.5	629.11 ± 0.00	86.46 ± 0.16	169.31 ± 0.09	7.03 ± 0.03	4.33e+00 ± 1.9e-02	1102203
$mz_i > 40$	1240.07 ± 0.27	199.18 ± 0.022	10.28 ± 0.002	25.40 ± 0.001	19.97 ± 0.00	6.95 ± 0.013	57.39 ± 0.03	2.40 ± 0.01	1.43e+00 ± 0.6e02	1204
Choose ZZ pair	1159.59 ± 0.19	136.70 ± 0.015	9.82 ± 0.002	25.07 ± 0.001	19.72 ± 0.00	6.93 ± 0.013	53.70 ± 0.02	2.20 ± 0.01	1.31e+00 ± 0.57e-02	1157
$m_{4l} > 70$	1066.69 ± 0.139	110.17 ± 0.012	9.22 ± 0.001	23.31 ± 0.001	18.20 ± 0.21	6.41 ± 0.012	47.09 ± 0.02	1.99 ± 0.01	1.18e+00 ± 0.52e-02	-
$\lvert m_{4l} - 125\rvert \leq 10$, $N_l = 4$	25.94 ± 0.00	102.29 ± 0.011	0.71 ± 0.00	2.8 ± 0.0	0.73 ± 0.00	0.05 ± 0.00	43.17 ± 0.02	1.88 ± 0.01	1.11e+00 ± 4.8e-02	-

Channel: $2e2\mu$	$Z\gamma^*, ZZ$	$Z + js$	$WZ, WW, W + js$	$t\bar{t}$	ttV	VVV	SMH	signal: $tan\beta = 0.5, sin\theta = 0.7$	signal $tan\beta = 1.0, sin\theta = 0.35$	*Obs.*
Initial	1.13e+05 ± 1.48e+01	3.10e+08 ± 3.66e+04	1.92e+09 ± 3.58e+05	1.11e+07 ± 1.41e+03	1.77e+04 ± 4.66e+00	5.21e+03 ± 9.55e+00	1.50e+03 ± 9.24e-01	6.70e+01 ± 3.33e-01	4.11e+01 ± 2.03e-01	3.85e+09
At least one Z	53664.59 ± 5.67	1638792?1.96 ± 1013.78	764402.63 ± 967.65	1543365.92 ± 35.34	3057.00 ± 1.59	836.91 ± 1.38	807.03 ± 0.48	53.99 ± 0.21	3.31e+01 ± 1.30e-01	177527759
At least 2 Zs	16576.76 ± 1.67	267577.05 ± 741.32	51658.59 ± 9.63	21381.93 ± 48.09	1450.82 ± 0.77	208.86 ± 1.36	545.87 ± 0.30	25.91 ± 0.01	1.59e+01 ± 0.62e-01	552366
$mz_i > 40$	2392.18 ± 0.32	172.36 ± 0.47	15.97 ± 0.003	57.38 ± 0.13	40.23 ± 0.13	15.60 ± 0.25	130.04 ± 0.15	6.90 ± 0.002	4.21e+00 ± 0.17e-01	1810
Choose ZZ pair	2392.18 ± 0.32	151.3 ± 0.41	15.67 ± 0.002	56.32 ± 0.13	40.23 ± 0.13	15.60 ± 0.25	130.04 ± 0.15	6.70 ± 0.002	4.18e+00 ± 0.16e-01	2661
$m_{4l} > 70$	2296.16 ± 0.3	142.23 ± 0.38	15.43 ± 0.002	55.43 ± 0.12	38.67 ± 0.29	15.06 ± 0.41	120.10 ± 0.14	6.53 ± 0.002	4.00e+00 ± 0.16e-01	-
$\lvert m_{4l} - 125\rvert \leq 10$, $N_l = 4$	67.94 ± 0.01	140.23 ± 0.38	1.49 ± 0.0	4.8 ± 0.01	1.43 ± 0.0	0.09 ± 0.0	113.31 ± 0.14	6.23 ± 0.002	3.79e+00 ± 0.15e-01	-

7 Multivariate Analysis

As shown in the last chapter, the mono-Higgs signal is quite small compared to the main backgrounds (ZZ, Higgs). In order to cope with the limitations due to the low statistics and to enhance the signal to background discrimination the Multivariate analysis approach (MVA) 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. Moreover the exclusion limits are obtained by using the cut-based analysis results and compared to those obtained with the MVA approach.

7.1 The Multivariate analysis approach

It has been shown in the last chapter in section 6.2 that simply applying additional cuts didn't improve the sensitivity to the DM signals but actually could reduce the signal efficiency and hence the sensitivity. 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 $\mathbf{x} = (x_1, x_2, x_3, ..., x_d)$ 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 7.1 where we have two variable x_1 (left), x_2 (middle) shown for signal (blue) and background (red) where there are no clear cuts on x_1 and x_2 that could separate the two classes. However in two dimensions we can find that the two classes are largely separable with a cut applied to the linear function called linear discriminant (right). MVA uses multiple variables to learn from the features of the input data variables to respond correctly to future data. This method is called supervised learning where a training data sets (from MC simulations) are used as inputs (signal and background) to build a mapping function $y = f(x, w)$ from the inputs (x) to the output (y) "where w are some adjustable parameters" and hence can respond correctly to future data. For this analysis I have used the boosted decision trees (BDT) implemented in the TMVA toolkit [205] released with ROOT.

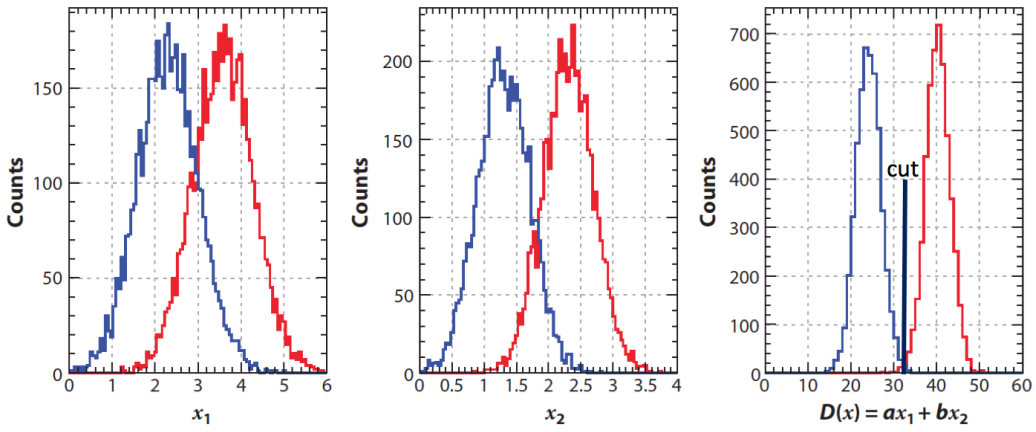


Figure 7.1: Shows the distribution of two variables x_1 (left), x_2 (middle) from two classes signal (blue) and background (red) and the linear discriminant between the two classes (right) with the optimal cut shown with the vertical line [27].

7.2 The Boosted Decision Tree

A decision tree is a binary tree structured classifier such as the one shown in Figure 7.2 which performs sequential cuts to perform the classification task. The process starts from an input training sample usually from MC simulation (signal and background) and finds the variable with the corresponding cut that provides the best separation between signal and background events. This process of finding the cut criteria is known as the training process. The training process starts with the root node as shown in figure 7.2; applying the chosen cut criteria will divide the sample into two subsamples, a signal-like (right) and a background-like (left) sample resulting in creating two nodes. At each node, the split process starts again by finding the variable and corresponding cut value that provides the best separation between signal and background. The determination of the optimal cut is done through the Gini index, defined as G = p(1 - p), where p is the purity, defined as the ratio of signal events to all events in that node p = s/(s+b) (hence pure background nodes have zero purity). The splitting process is repeated until the node has reached either a maximum signal purity or a minimum number of events. At the end the final leave nodes are called "signal" or "background"according to the majority of events belongs to. The idea of "Boosting" process means that the signal events that end up in a background node (and vice versa) are largely weighted than the events that are in the correct node. Then a new decision tree can be started from the re-weighed training event sample. The boosting process can be repeated several times ($\approx$ 500 times) where we end up with a set of decision trees (a forest). At the end of the training process the variables are ranked (means the importance of each variable) by counting how many times the variables are used in decision tree node splitting, and weighting each occurrence by the square of the gain achieved in the separation and by the number of events in the node. The output of the BDT discriminator is a value "score" between -1 to +1 for each event. The signal like events

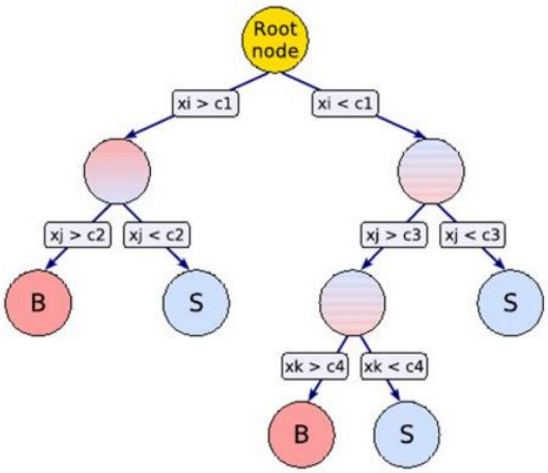


Figure 7.2: Schematic view of a decision tree. Starting from the root node and perform sequent cuts on a discriminating variables (x_i) where the nodes classified as signal-like or background like. The leaf nodes at the bottom of the tree are labeled "S" for signal and "B" for background depending on the majority of events that end up in the respective nodes.

have a score close to +1 while the background like events have a score close to -1.

7.2.1 Input variables

The choice of the BDT input variables are done by exploiting the kinematic differences between signal and background events. The missing transverse energy and the transverse mass of the four-lepton system and p_T^{miss} should give a reasonable separation between mono-Higgs signal and Higgs background process. Also the Higgs mass can give good separation between mono-Higgs sample and the non-resonant ZZ background. To choose the best set of the input variables, a different configurations were performed between different variables combinations and in each time the area under the ROC curve and the overtraining check was reported as a figure of merit. The final choice of the variables is:

- MET (p_T^{miss}) in the event
- Transverse mass of the Higgs and p_T^{miss}
- Higgs Mass (M_{4l})
- Mass of the second Z (Z_2)
- $\Delta\phi$ between the Higgs and the MET ($\Delta\phi(H, p_T^{miss})$)
- Transverse momentum of the four leptons ($p_{T1}, p_{T2}, p_{T3}, p_{T4}$)
- ΔR between Z_1 and Z_2

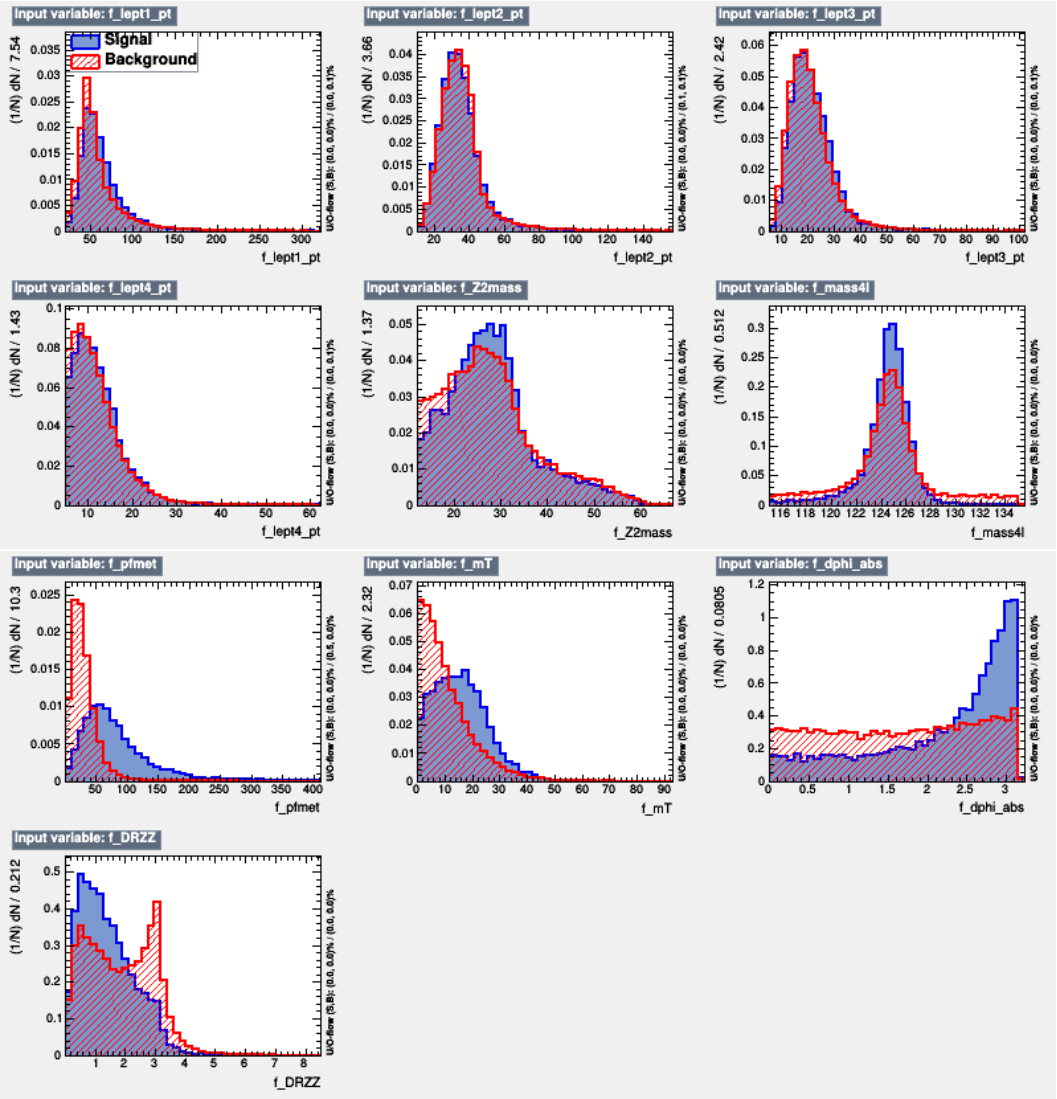


Figure 7.3: Distributions of the BDT input variables.

Figure 7.3 Shows the distributions of the input variables used in the training for both signal and background processes. The majority of the input variables have a lower correlation for both signal and background as shown in Figure 7.4. The importance of each input variable in the training, "the variable ranking", shows that the most discriminant variables are the p_T^{miss}, $\Delta\phi(H, p_T^{miss})$, m_{4l}, m_{Z2} and Δ R (Z_1 , Z_2).

7.2.2 The BDT training

The BDT training is performed in a region enriched with signal events passing all the mono-Higgs event selection cuts defined in 6.2 and listed as follow:

- the Higgs selection cuts;
- $|M_{4l} - 125| < 10$ GeV;
- Number of selected leptons in the event N_{4l}= 4.

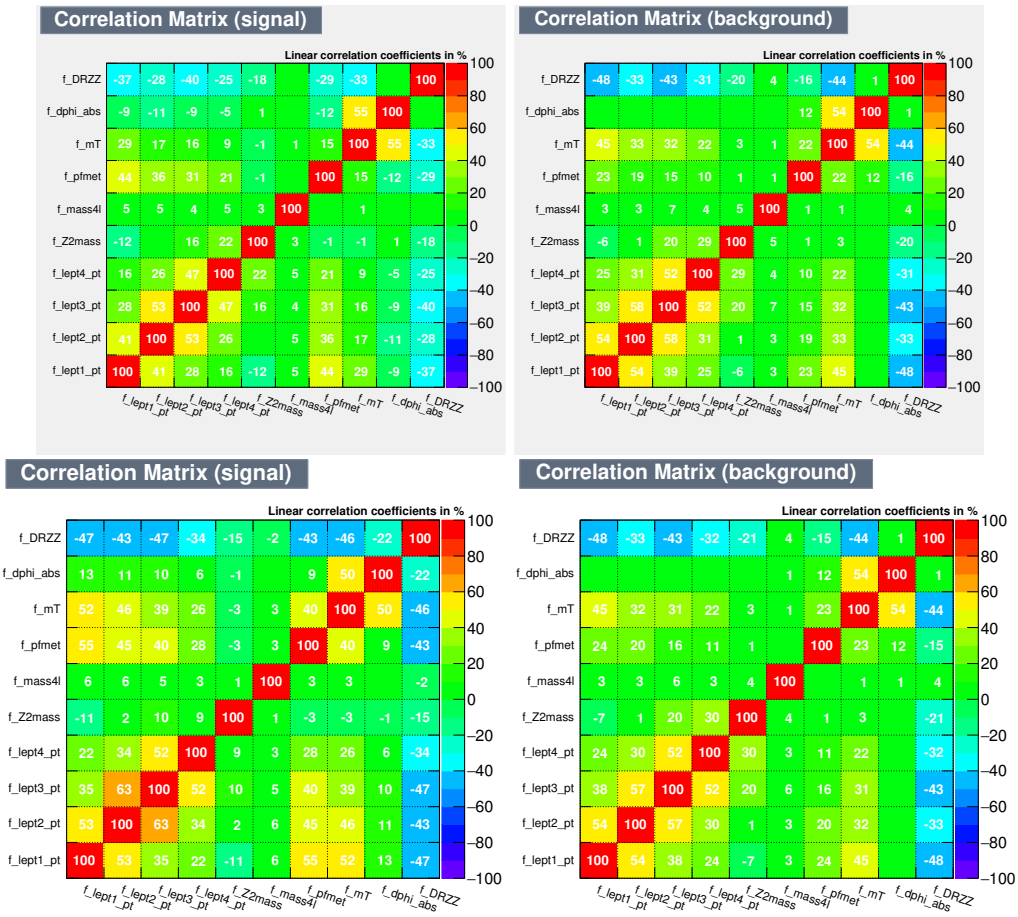


Figure 7.4: Correlation coefficient of the BDT input variables for signal (left) and background (right) events in LM (top) and HM (bottom) regions.

The training is performed for each decay channel (4μ, 4e, $2e2\mu$) and each year. In the training process, the signal samples are divided into low mass region (LM) and high mass region (HM) defined by $m_A \leq$ 400 GeV and $m_A >$ 400 GeV respectively. This chosen separation is used to better ensure the similarity of the input variable distributions which reflects a higher uniformity of the BDT response in those two regions. In principle, the training should be performed for each signal sample but this is very difficult to validate and control, given the large number of signal samples and the different final states. The division of the samples to LM and HM regions is a compromise to overcome this complexity. Also, training the samples together is aimed to increase the number of events available in addition to the purpose to get a BDT response as much independent as possible on a specific sample. The event weights are not applied to the signal events in order to make the BDT insensitive to the cross section of the samples. If the weight is applied, the signal samples having larger cross sections would be seen by the BDT as more interesting events and that introduce a bias. All the signal samples entering the training have approximately the same number of events to receive the same importance from the BDT. To validate the method of training with a combined set of signal samples, a comparison of the performance obtained by training with a specific signal sample with M_x value to the performance obtained by training with a combined set of signal samples is performed. The comparison is shown in Figure 7.5 for different mass points. The Figure shows the ROC curve obtained by training with single sample (blue) vs. the one obtained by training with a combined set of signal samples together (red) depending on the division of LM and HM regions. The blue curves of Figure 7.5 shows the ROC curves for signal samples with m_a = 150 GeV and m_A = 200, 500, 800 GeV in the top row from left to right respectively while the middle row for signal samples with m_a = 250 GeV and m_A = 300, 500, 800 GeV respectively and the bottom row for signal samples with (m_a = 350 GeV - m_A = 800 GeV), (m_a = 150 GeV - m_A = 300 GeV - $tan\beta = 4$) and (m_a = 250 GeV - m_A = 300 GeV - $tan\beta = 4$) respectively. The training performed on single signal sample achieves a background rejection that is similar to the one achieved from the combined training. The plots show that the method used is reasonable and can be used for a variety of signals.

All the background processes contributing to the signal region (such as SM Higgs, ZZ, VVV, $t\bar{t}V$) are used as input for the training. The ROC curves showing the signal efficiency vs. the background rejection obtainable in LM and HM regions are reported in Figure 7.6 for the different final states in 2016 as an example. Table 7.1 reports the area under the ROC curves for different final state and different year which shows high signal efficiency and background rejection for the trained BDT in LM and HM regions. The input sample is divided randomly to two subsets training and testing samples in order to train and verify the BDT output. The overtraining checking plot is shown in Figure 7.7 for the LM (left) and HM (right) regions for 4μ (top). 4e (middle) and $2e2\mu$ (bottom). A comparison between the ROC curves for the training and testing samples are shown in Figure 7.8. The plots in Figures 7.7, 7.8 show the good discrimination reached between the signal (in blue) and the background (in red), and the agreement between the training and test samples.

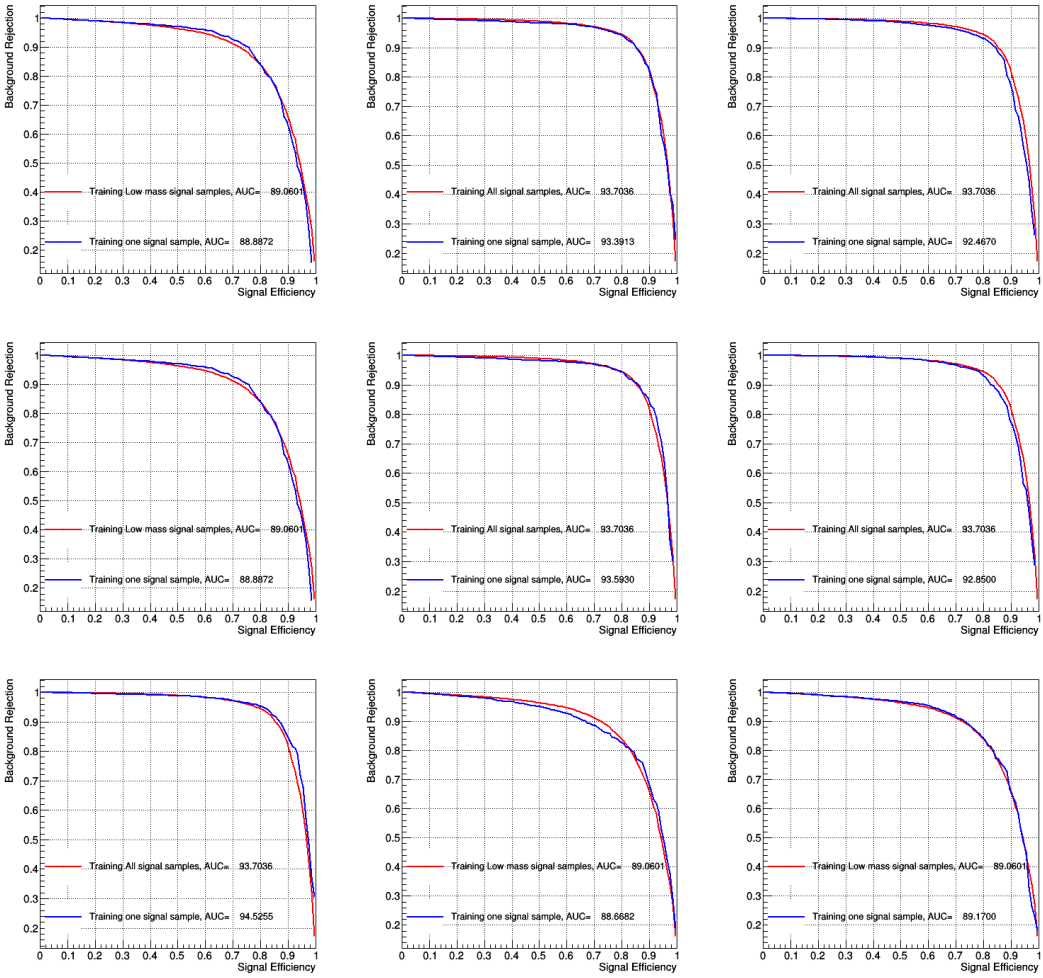


Figure 7.5: Comparison of the ROC curve obtained from training single sample (blue) vrs. the one obtained from training a combined set of signal samples together (red).

Table 7.1: Area under the ROC curves for different final states and different years of data taken.

Year	4μ		$4e$		$2e2\mu$	
	LM	HM	LM	HM	LM	HM
2016	0.89	0.936	0.885	0.935	0.895	0.942
2017	0.888	0.94	0.882	0.941	0.903	0.951
2018	0.865	0.926	0.857	0.92	0.868	0.93

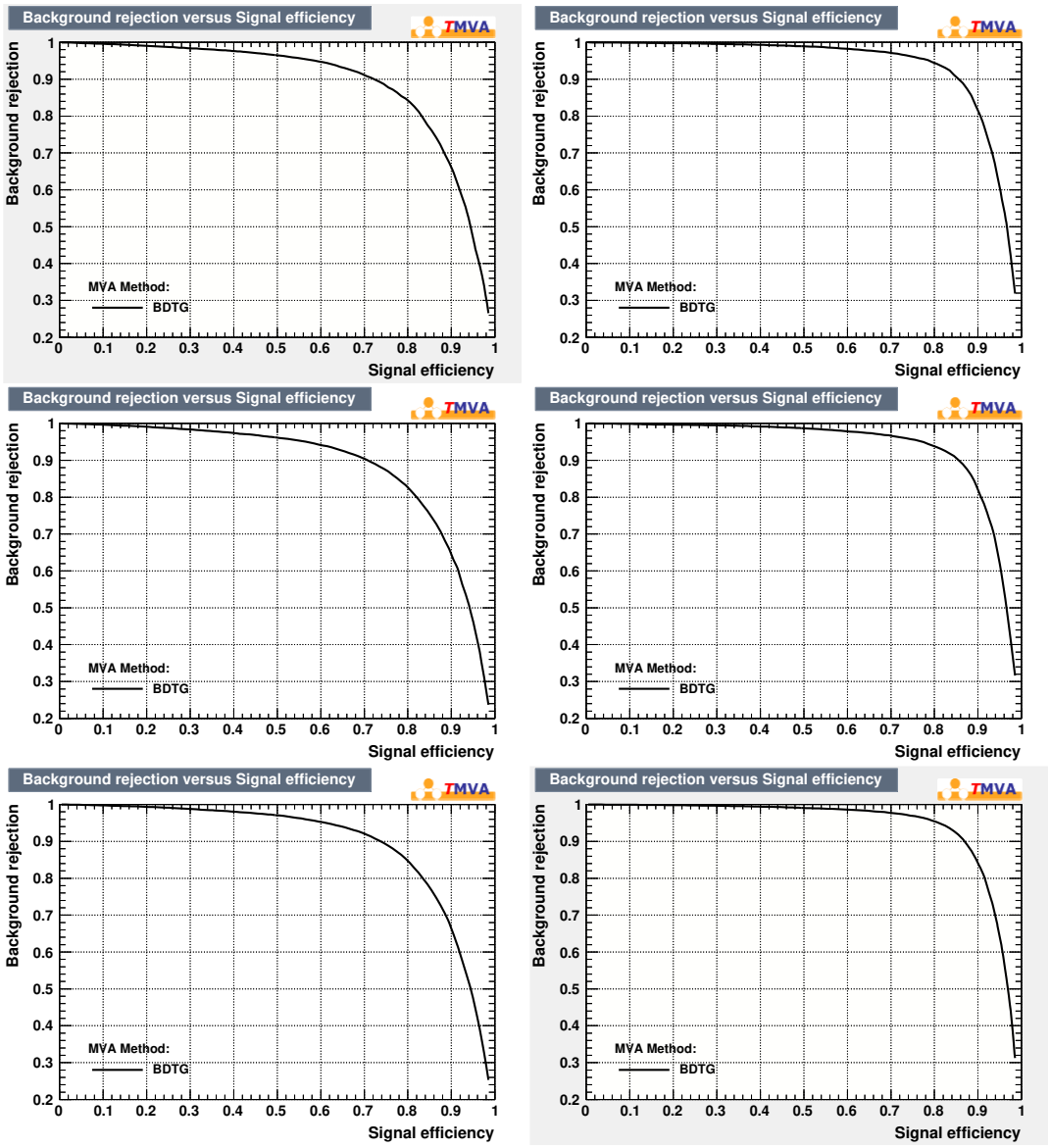


Figure 7.6: ROC curves obtained in the training of LM (left) and HM (right) regions for 4μ (top), 4e (middle) and $2e2\mu$ (bottom) for 2016.

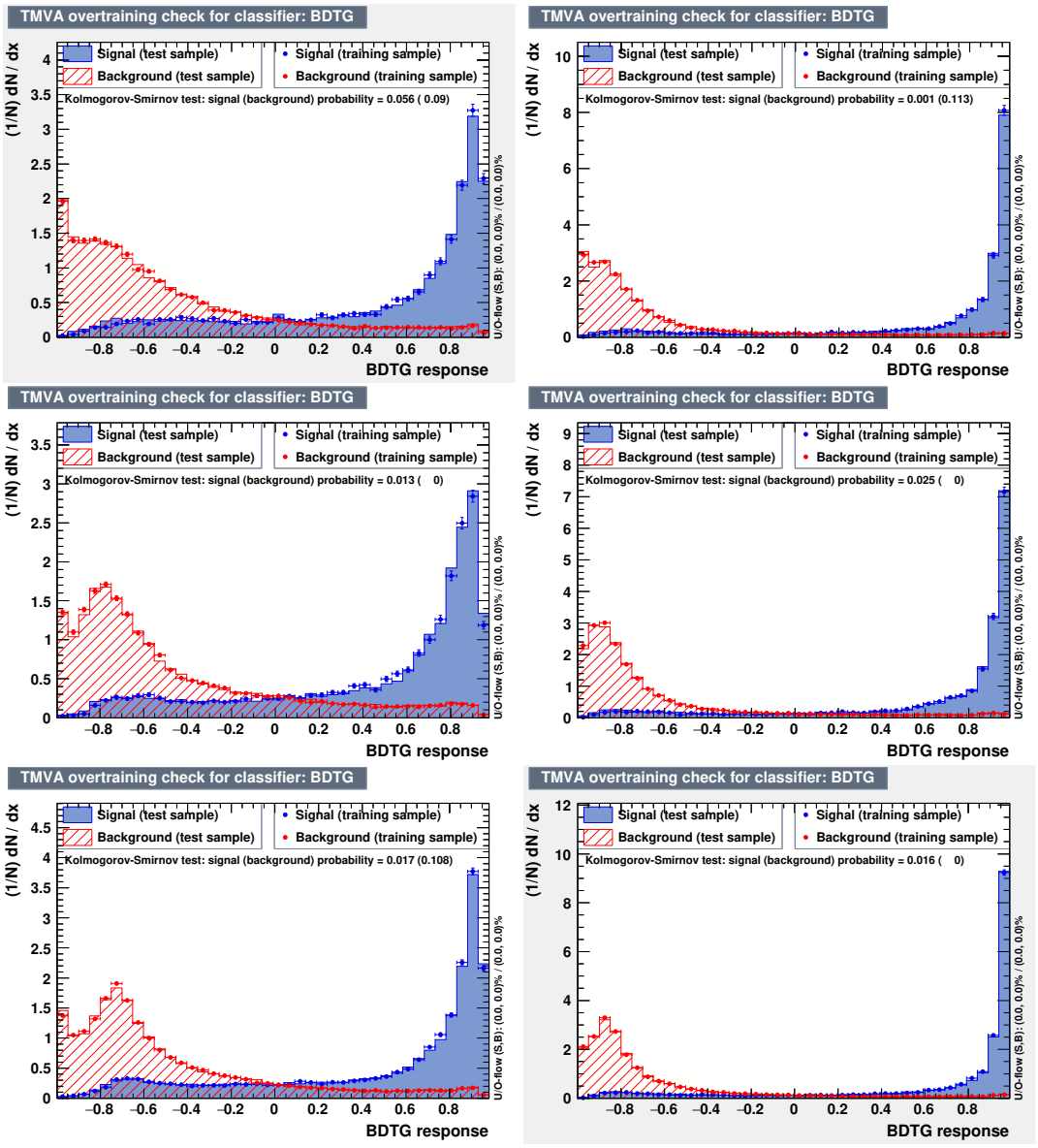


Figure 7.7: BDT distribution for training and testing samples for LM (left) and HM (right) regions for 4μ (top), 4e (middle) and $2e2\mu$ (bottom) channels in 2016 data taken as example.

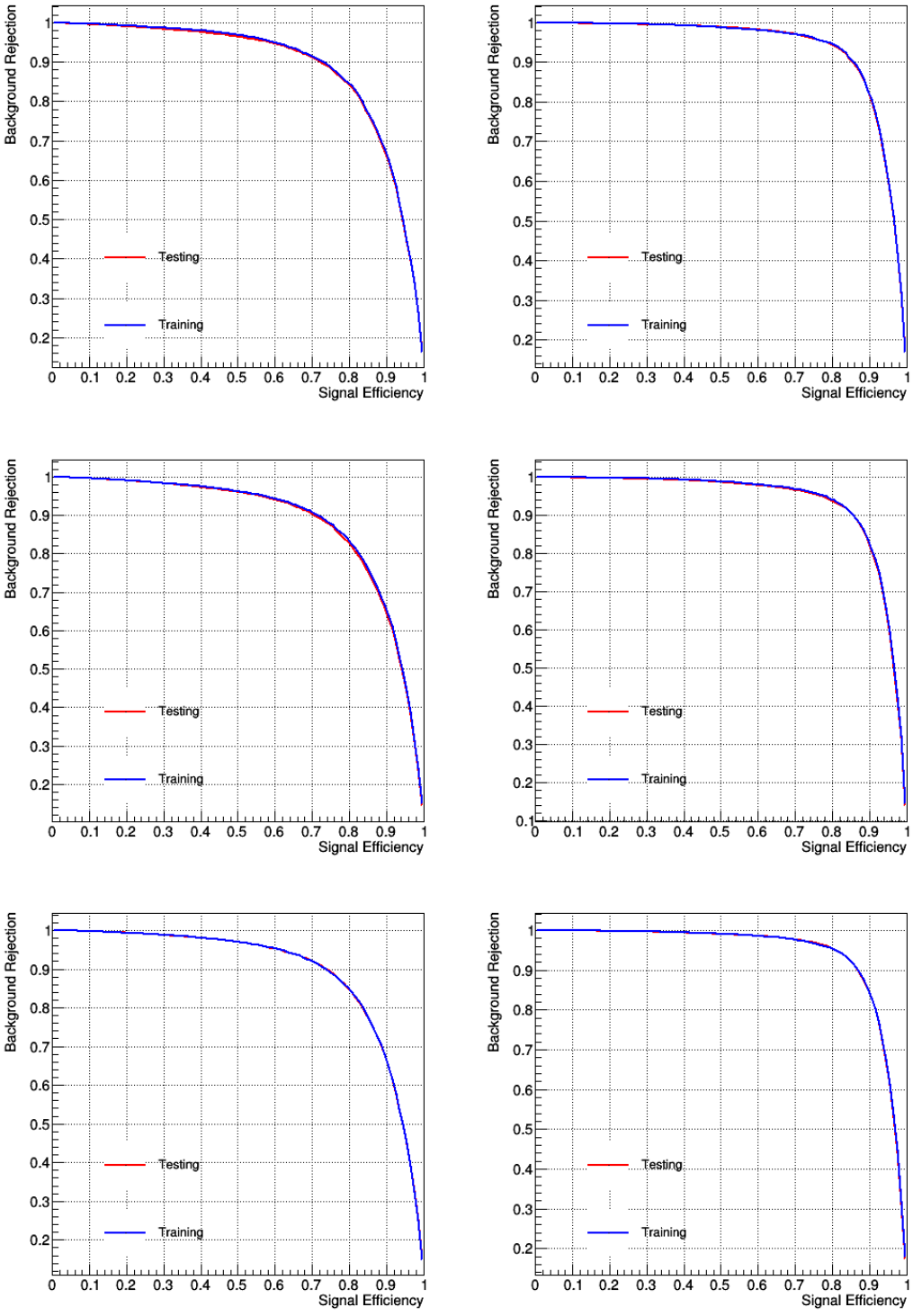


Figure 7.8: Comparison between the ROC curves for the training (blue) and testing (red) samples for LM (left) and HM (right) regions for 4μ (top), 4e (middle) and $2e2\mu$ (bottom) channels.

The BDT score is shown in Figure 7.9 for different signal samples, different background processes stacked together and data point in the signal region for 2016 (top), 2017 (middle) and 2018 (bottom) for LM region (left and HM region (right). The data points is shown in the background side of the BDT score while blinded in the signal side. The signal samples are peaking close to +1 and backgrounds close to -1. The signal sample are scaled up by 20 (50) for LM (HM) region.

7.3 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 8.1. They are treated as a normalization uncertainty which affects the yield or shape uncertainties which affect the distribution of the observables. In this section the sources of uncertainties are described in the following and summarized in table 7.2, 7.3. They are taken into account in the three years of the analysis [21].

7.3.1 The experimental uncertainties

- Uncertainty on the measurement of integrated luminosity: it's value varies from 2.3% to 2.6% depending on the data taking period [206–208]. It affects both signal and background, but doesn't affect the background estimation from data. This uncertainty is used as a normalization uncertainty on the event yield.

- Uncertainty on the lepton identification and reconstruction efficiency: it ranges from 1 to 2.5% and from 11 to 15.5% on the overall yields, in the 4 μ and 4e channels, respectively. It arises from variations of the fit method, tag p_T selection, and Z boson mass window while using the "tag-and-probe" technique [188]. This uncertainty is used as a normalization uncertainty on the event yield.

- Experimental uncertainty for the reducible background estimation discussed in section 6.4.4, that varies between 30% to 41% depending on the final state. This uncertainty is used as a normalization uncertainty on the event yield for the Z+X background.

- The uncertainty on the p_T^{miss} reconstruction: it is estimated by varying the energies of the PF objects used in the p_T^{miss} calculation within their uncertainties as suggested by the JET MET physics object Group and propagating those variations to the final p_T^{miss}

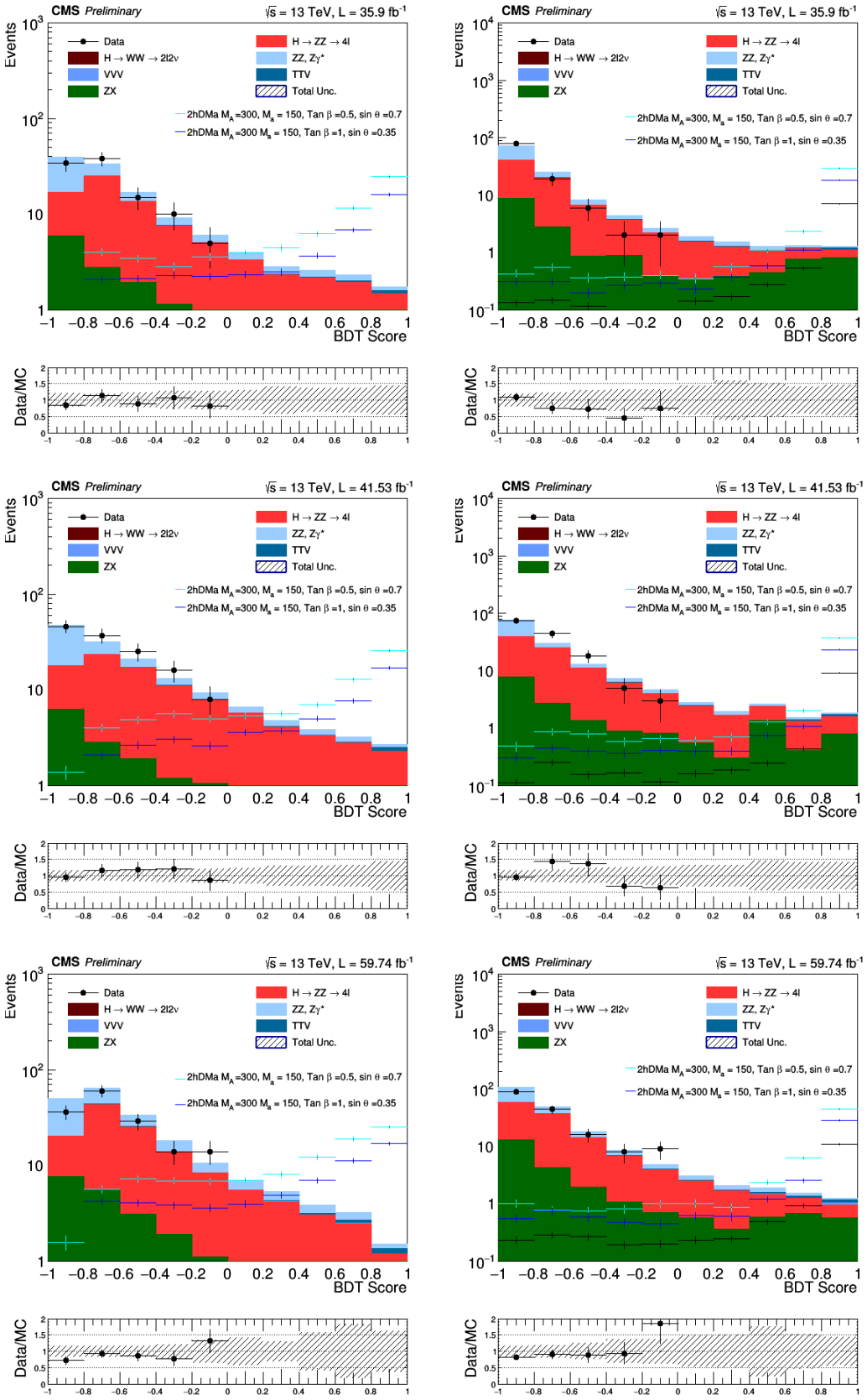


Figure 7.9: BDT score for different signal hypothesis and different stacked backgrounds for LM (left) and HM (right) regions for 2016 (top), 2017 (middle) and 2018 (bottom).

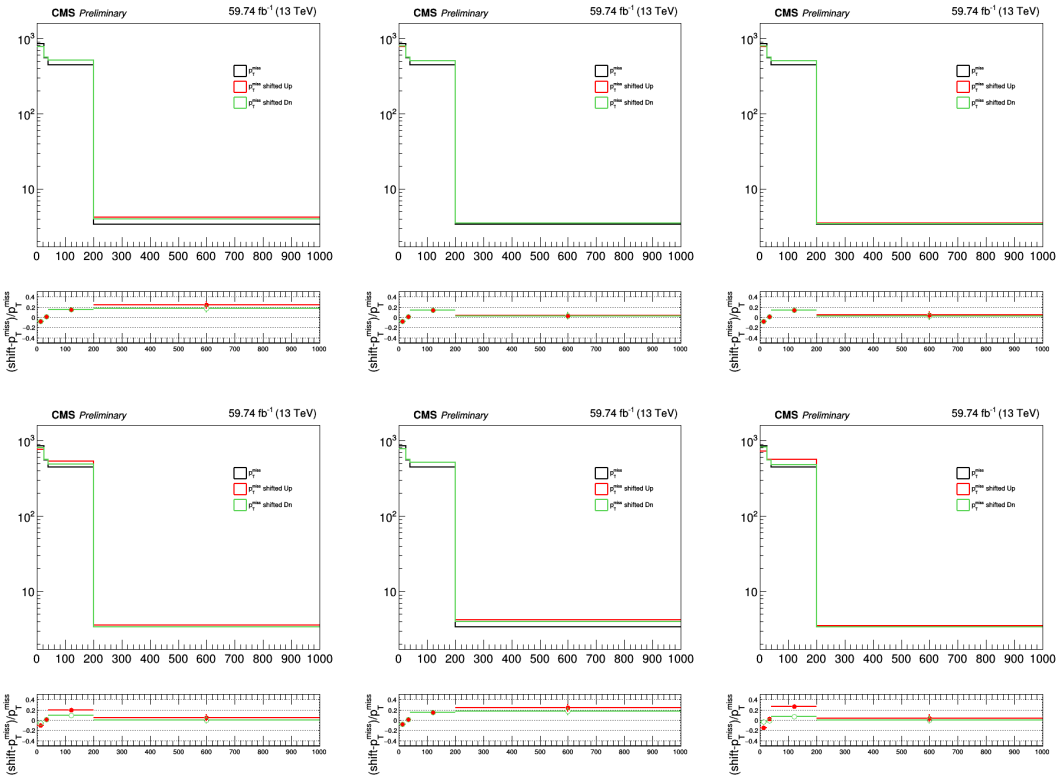


Figure 7.10: The p_T^{miss} distribution for all the background together are shown after passing the mono-Higgs selection and after varying the following uncertainty sources up and down (from left to right, from top to bottom): muon energy, electron energy, photon energy, jet Energy, jet resolution and unclustered jet energy.

distribution after the full mono-Higgs selection as shape uncertainties. The sources of uncertainties are (jet Energy - jet resolution, muon energy, electron energy, photon energy, unclustered jet energy). Each source is varied up and down by one standard deviation of its input distribution. Figure 7.10 shows the p_T^{miss} distribution after passing the mono-Higgs selection where each source of uncertainty is varied up and down for all the background together.

- The BDT discriminator shape uncertainty: it is estimated by checking the impact of systematic uncertainties on the BDT input variables on the shape of the BDT output. The uncertainties on the energy of the PF objects used in the the p_T^{miss} calculation are varied up and down by one standard deviation of its input distribution. The BDT were fed with the nominal inputs (no systematic uncertainty shifts applied) and fed with $\pm\sigma$ shifts from the nominal inputs. Figure 7.11 shows the BDT score distribution after passing the mono-Higgs selection when each source of the p_T^{miss} uncertainty is varied up and down for all the backgrounds together.

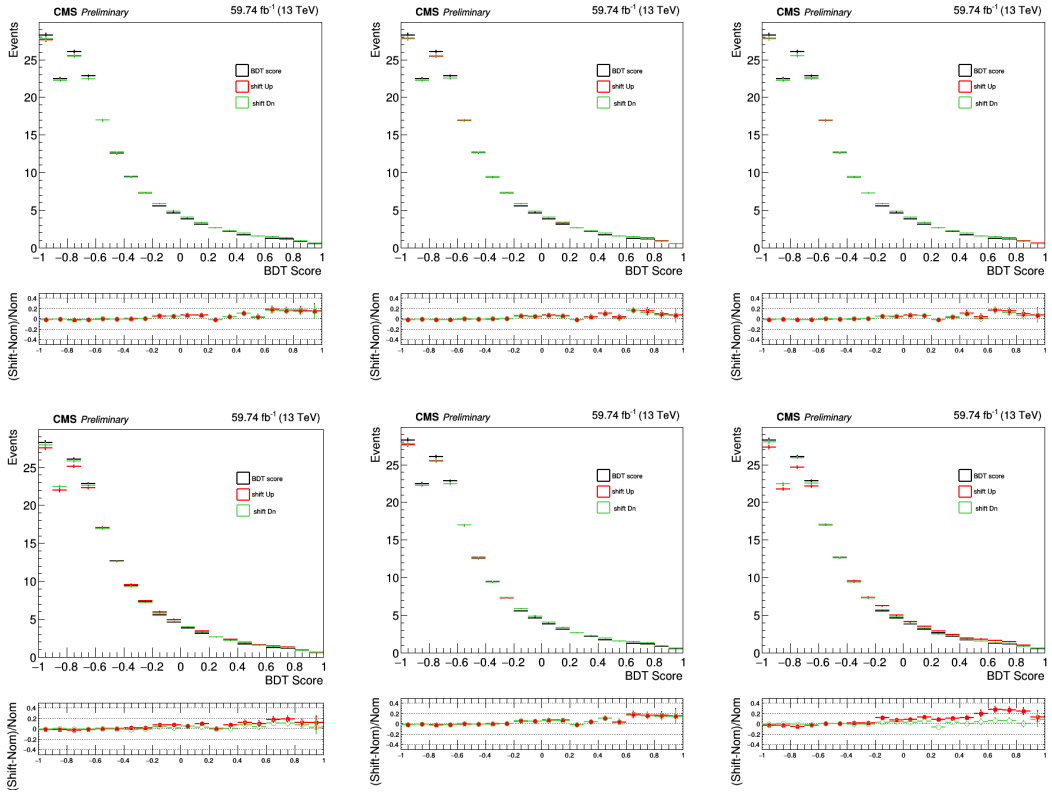


Figure 7.11: The BDT score distribution for all the background together after passing the mono-Higgs selection and after varying the following uncertainty sources up and down (from left to right, from top to bottom): muon energy, electron energy, photon energy, jet Energy, jet resolution and unclustered jet energy.

Table 7.2: Summary of experimental systematic uncertainties.

Experimental uncertainties			
Type	2016	2017	2018
Luminosity	2.6%	2.3%	2.5%
Leptons ID and reco eff.	1.6 - 15.5 %	1.1 - 12.1 %	1.0 - 11 %
p_T^{miss} uncertainties			
Jet energy scale	shape	shape	shape
Jet energy resolution	shape	shape	shape
muon energy	shape	shape	shape
electron energy	shape	shape	shape
photon energy	shape	shape	shape
unclustered jet energy	shape	shape	shape
Reducible background (Z+X)	30 - 41 %	30 - 38 %	30 - 37 %

7.3.2 Theoretical uncertainties

Those uncertainties affect both signal and background processes [209], they are used as a normalization uncertainty. They are described in the following and summarized in table 7.3.

- the uncertainty from the choice of the PDF set is determined by calculating the root mean square of the variation when using different replicas of the default NNPDF set, following the PDF4LHC recommendations [210];
- the uncertainty from the renormalization and factorization scale is determined by varying these scales between 0.5 and 2 times their nominal value [211];
- the uncertainty on the K factor used for the ggZZ samples is applied [21];
- the uncertainty on the branching ratio of Higgs decay to ZZ to four leptons is applied to the yields of all Higgs processes [21].

7.3.3 Impact of systematic uncertainties

In any analysis, an important step is to blind the analysis since looking at data in the signal region could introduce a bias when defining the event selection as mentioned in 6.5. Blinding the analysis is a standard procedure in CMS Collaboration. Once the analysis procedure is well defined, it is possible to look at data in the signal region, the unblinding process. An important step is to check the impact of the systematic uncertainties on the sensitivity of the analysis before looking at data. This effect is measured by measuring the change $\Delta\mu$ induced in the signal strength μ when varying each source of systematic uncertainties θ by 1 σ. This is done using Asimov data from pseudo experiments without looking at the data. Figure 7.12 shows the impact of the systematic uncertainties used in the analysis on the signal strength. The left panel lists the names of the uncertainties, the middle panel shows the pulls of each each source of systematic uncertainties, defined as the difference between its maximum likelihood estimation and its nominal value, divided by the uncertainty, and the right panel shows the impacts of each nuisance parameter on the signal strength value [218]. The middle panel shows that no parameter is significantly pulled away from its nominal value. The largest impact comes from uncertainties related to p_T^{miss}.

This chapter presented the Multivariate approach "Boosted Decision Trees" used to better discriminate between signal and background events exploiting different kinematic observables to improve the sensitivity of the analysis. In addition, the sources of systematic uncertainties affecting the results are presented. The BDT discriminator distributions will be used to extract the 95% C.L. upper limits on the signal strength in the next chapter. Moreover the exclusion limits were driven by using the cut-based analysis results and compared to those obtained from the MVA approach.

Table 7.3: Summary of theory systematic uncertainties.

Theory uncertainties	
PDF set ggH	1.8 %
QCD scale ggH	4.4 - 6.5 %
PDF set VBF	2.1 %
QCD scale VBF	0.3 - 0.4 %
PDF set ZH	1.6 %
QCD scale ZH	2.7 - 3.5 %
PDF set WH	1.3 %
QCD scale WH	0.5 %
PDF set ttH	3.6 %
QCD scale ttH	6.0 - 9.2 %
PDF set qqZZ	3.1 - 3.4 %
QCD scale qqZZ	3.2 - 4.2 %
Electroweak correction qqZZ	0.1 %
PDF set ttW	25 - 37.5 %
QCD scale ttW	3 - 4 %
PDF set ttZ	7 - 14 %
QCD scale ttZ	2 - 3 %
PDF set VVV	2 - 17 %
QCD scale VVV	3 %
PDF set ggZZ	3.2 %
QCD scale ggZZ	4.6 - 6.7 %
Electroweak correction ggZZ	10 %
PDF set Mono-H	1.6 - 3.2 %
QCD scale Mono-H	0.05 - 0.3 %

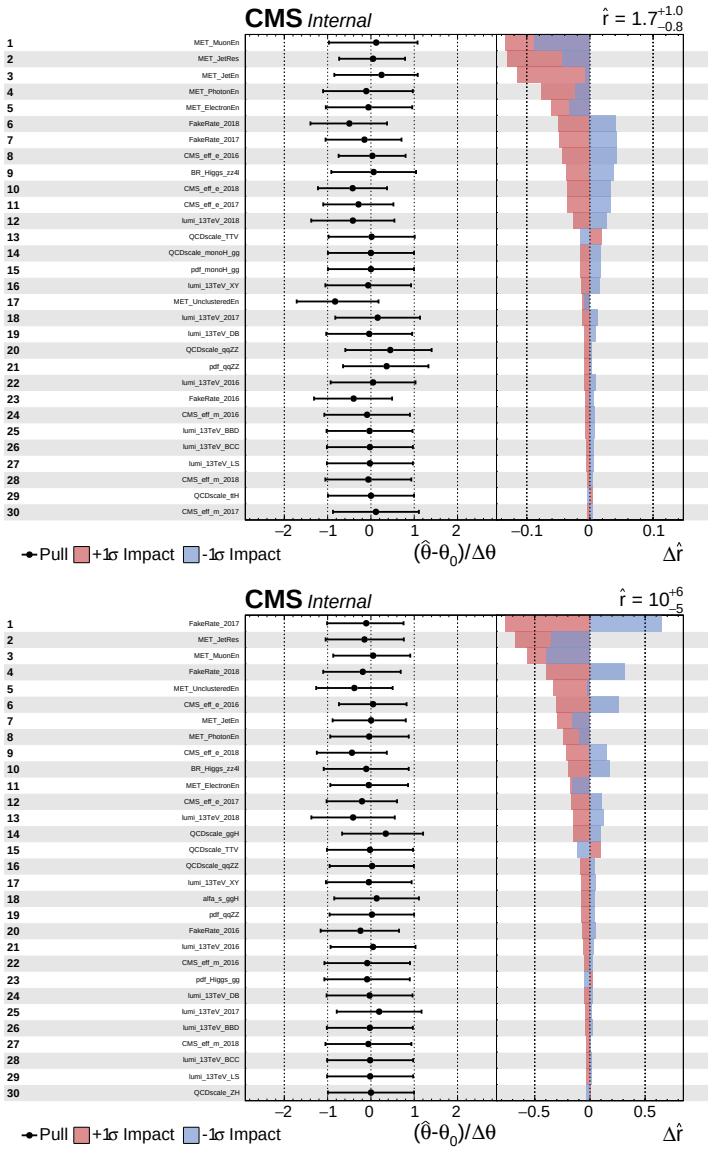


Figure 7.12: Impacts of each source of systematic uncertainties on the signal strength value for signal sample from LM region (top), HM region (bottom) for full RunII.

8 Final Results

At this point, we got all the ingredients to start the statistical analysis. Since we performed the selection of the events in the signal region, the estimation of the backgrounds and a description of systematic uncertainties, we can finally proceed to the interpretation of the results as upper limits on the signal strength for the dark matter model under study. These limits can allow the exclusion of some points of the parameter space of the model chosen. Before the interpretation of the results, next section discusses the statistical method used in the analysis [212, 213].

8.1 The statistical method

The statistical procedure used for this search is the frequentist profile likelihood approach adopted by the CMS and ATLAS collaborations in the context of the LHC Higgs Combination Group [214] during Run I. Starting from the final yield in the signal region, we consider the expected signal yield and total backgrounds yield in the signal region given by s , b respectively. In the analysis binned distributions are used, so s, b represent the event yield obtained in all the bins of the distributions used in the three final states. The values of s and b are affected by the systematic uncertainties that have been already discussed in section 7.3, so s and b are functions of those parameters. The systematic uncertainties are considered as nuisance parameters θ_i called collectively as θ. 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 $\tilde{\theta}_i$. To determine how signal or background events like the experimental observations are, we define a test statistic based on likelihood function $\mathcal{L}$ given by:

$$\mathcal{L}(n, \tilde{\theta}|\mu, \theta) = Poission(n|\mu s + b).p(\tilde{\theta}|\theta) \tag{8.1}$$

where n represent the observed data or Asimov data (from pseudo experiment). The probability density function P$(\tilde{\theta}|\theta)$ is the probability to measure a set of nuisance parameters $\tilde{\theta}$ given its true value θ. The form of the Poisson term in equation 8.1 is given by:

$$Poission(n|\mu s + b) = \prod_j \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \tag{8.2}$$

where the product runs over all bins j for the binned distributions.

A statistic test is performed to determine the existence or absence of a 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 which represents the ratio of the measured cross section σ and the value predicted by the theory model $\mu_s = \sigma/\ \sigma_{th}$). 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. In opposite, if we want to quantify a signal excess, we express the compatibility of the observed data with the H_b hypothesis in terms of a probability or p-value.

The test statistic q_μ used to set exclusion limit is defined from the likelihood given in equation 8.1 and can be expressed as:

$$q_\mu = -2ln\frac{\mathcal{L}(n|\mu, \hat{\theta}_\mu)}{\mathcal{L}(n|\hat{\mu}, \hat{\theta})}, 0 \leqslant \hat{\mu} \leqslant \mu \tag{8.3}$$

where $\hat{\theta}_\mu$ in the numerator represents the value of the nuisance parameter θ that maximizes the likelihood function for a fixed μ, the parameters $\hat{\mu}$ and $\hat{\theta}$ in the denominator are the ones which maximize $\mathcal{L}$ and n refers to the events from the observed data or obtained from a "pseudo-experiment". The likelihood function $\hat{\theta}_\mu$ is the product of Poisson probabilities for number of events (observed or simulated), given the expected signal and background or background only hypothesis as given in equation 8.2 [215].

Exclusion limits are computed in this analysis using q_μ in the modified frequentist criterion CLs [216, 217]; the exclusion limits are calculated at 95 % confidence level (CL), by using equation 8.3 to calculate 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 H_b hypothesis, such as :

$$\begin{aligned} CL_{s+b}(\mu) &= P(q_\mu \geqslant q_\mu^{obs}|H_{\mu_s+b}) \\ CL_b(\mu) &= P(q_\mu \geqslant q_\mu^{obs}|H_b) \end{aligned} \tag{8.4}$$

From the previous equation the CL_s can be defined as:

$$CL_s(\mu) = \frac{CL_{s+b}(\mu)}{CL_b(\mu)} \tag{8.5}$$

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

Table 8.1: Number of expected background and signal events after applying the full Mono-Higgs selection. Signal samples with $M_A = 300$ GeV, $M_a = 150$ GeV, $M_\chi = 10$ GeV.

Process	2016	2017	2018
$H \to ZZ \to 4l$	63.2 ± 5.57	77.9 ± 6.9	104.5 ± 9.2
$Z\gamma, ZZ$	38.5 ± 2.49	47.9 ± 3.1	67.7 ± 4.4
TTV	0.43 ± 0.077	0.55 ± 0.09	0.89 ± 0.13
VVV	0.08 ± 0.04	0.05 ± 0.03	0.04 ± 0.02
$H \to WW \to 2l2\nu$	0.028 ± 0.003	0.32 ± 0.009	0.44 ± 0.008
Z+X	15.54 ± 3.5	15.4 ± 3.17	22.67 ± 4.0
$signal : \tan\beta = 1, \sin\theta = 0.35$	2.04 ± 0.072	2.4 ± 0.08	3.0 ± 0.11
$signal : \tan\beta = 0.5, \sin\theta = 0.35$	7.8 ± 0.28	9.8 ± 0.36	12.5 ± 0.45
$signal : \tan\beta = 0.5, \sin\theta = 0.7$	3.3 ± 0.12	3.9 ± 0.14	4.97 ±0.18

8.2 Event yields and final distributions

The final yields obtained after the full event selection, as expected from signal and from the backgrounds, estimated both from simulation and from data (Z+X processes) are summarized in Table 8.1, the total uncertainty (statistical+systematic) is quoted for the simulated signal and background samples as well as for the Z+X estimation. The final distribution of the p_T^{miss} after applying the mono-Higgs event selection is shown in Figure 8.1 for 2016 (top left), 2017 (top right), 2018 (bottom left) and full run II combined together (bottom right). The background distributions from simulated processes are stacked; the simulated DM signals are also superimposed, we can see that the signal started to peak at the tail of the distribution. The SM Higgs background and the ZZ non resonant production are the most important sources of background in the signal region besides a contribution from the reducible background.

8.3 Exclusion limits

In the analysis the results are extracted using a binned maximum likelihood fit performed both on the p_T^{miss} distribution and also on the MVA output distribution in the signal region. The fit is performed separately per channel (4μ, 4e, 2e2μ), for their combination and for different year of data collection. The sources of systematic uncertainties are treated as nuisance parameters in the fit. Some systematics affect the normalization while other systematics affect the shape of the p_T^{miss} distribution or the MVA output distribution. The shape uncertainty affecting the distribution can be modeled by changing in the simulation the parameter affected by the systematic to its most probable value plus (minus) its uncertainty. Two new spectrums are obtained corresponding to $\pm 1\sigma$ above (below) the distribution obtained with the best

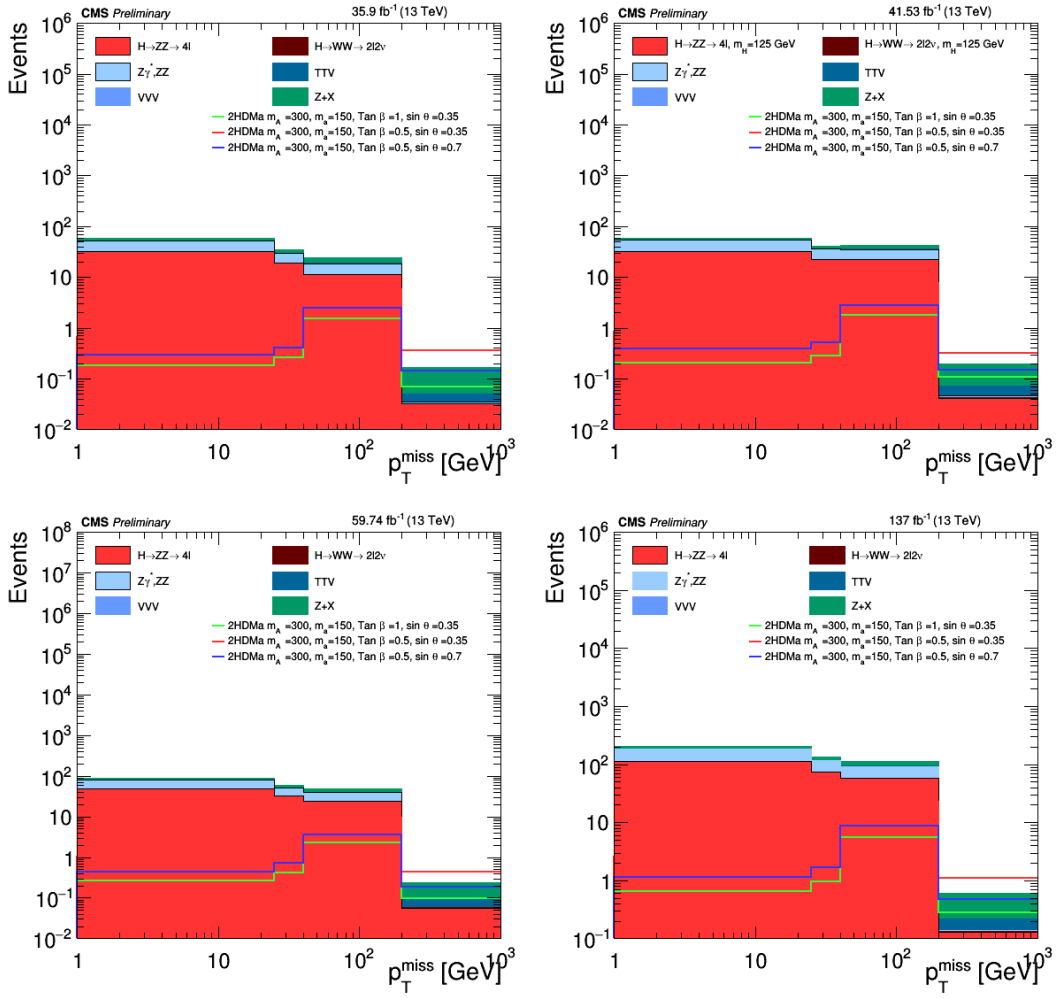


Figure 8.1: Missing transverse energy distribution p_T^{miss} in the Mono-Higgs signal region after applying all the selection cuts for 2016 (top left), 2017 (top right), 2018 (bottom left) and full Run II (bottom right).

estimate of the parameter (central distribution) as described in 7.3.

The number of events for signal and background processes in the mono-Higgs signal region for the three years are summarized in table 8.1. Those results and the p_T^{miss} distribution are used to derive the upper limit on the signal strength. Figure 8.2 shows the 95 % CL upper limit on the signal strength for the combination of the three final states in the full Run II analysis. The top figure shows the 95 % CL upper limit on the signal strength as a function of the heavy pesudoscalar mass "A", and different curves correspond to different mass values of the light pesudoscalar "a". Better sensitivity is observed in lower values of the mass of the light and heavy pesudoscalars, while the sensitivity of the analysis is limited in the high mass region. The left bottom plot shows the 95 % CL upper limit on the signal strength as a function of $tan\beta$ values, and different curves correspond to the different mass values of the light pesudoscaler "a", while the bottom right plot shows the 95 % CL upper limit on the signal strength as a function of $tan\beta$ values, and different curves correspond to different values of $sin\theta$. The analysis can exclude the signal sample with the parameters $tan\beta = 0.5$ for $m_a = 150$ GeV, $m_A = 300$ GeV and $sin\theta = 0.35$.

To increase the sensitivity of the analysis the results extracted using BDT distributions instead of p_T^{miss} distribution. Figure 8.3 top shows the 95 % CL upper limit on the signal strength as a function of the heavy pesudoscaler mass "A", different curves corresponding to different mass values of the light pesudoscalar "a". The best sensitivity is observed in lower values of the mass of the light and heavy pesudoscalars, while the sensitivity of the analysis is limited in the high mass region. The left bottom plot shows the 95 % CL upper limit on the signal strength as a function of $tan\beta$ values, where the different curves correspond to the different mass values of the light pesudoscaler "a", while the bottom right plot shows the 95 % CL upper limit on the signal strength as a function of $tan\beta$ values, where the different curves correspond to different values of $sin\theta$. An increase of the sensitivity is observed in some points of the parameter space after using the BDT distribution. The gain in the sensitivity after using the BDT approach is shown in the next section.

8.3.1 Limit comparison

After performing the shape analysis and extracting the upper limits by using both the BDT shapes and the p_T^{miss} distribution as input, a comparison between the two methods is performed to check the gain obtained with using the BDT discriminator. The gain is defined as the ratio between the upper limit obtained using the BDT shapes over the one obtained by using the p_T^{miss} distribution, if the gain value > 1 this means a better sensitivity. Figure 8.4 top shows the gain as a function of the heavy pseudo-scalar mass M_A which indicates that the BDT approach improves the sensitivity between 20% and 40% in the low mass region for different mass points while a loss in the sensitivity is reached in the high mass region. This decrease in the sensitivity comes from the fact that the statistics of the signal sample at high mass region is limited. Anyway, the analysis has a limited sensitivity in the high mass

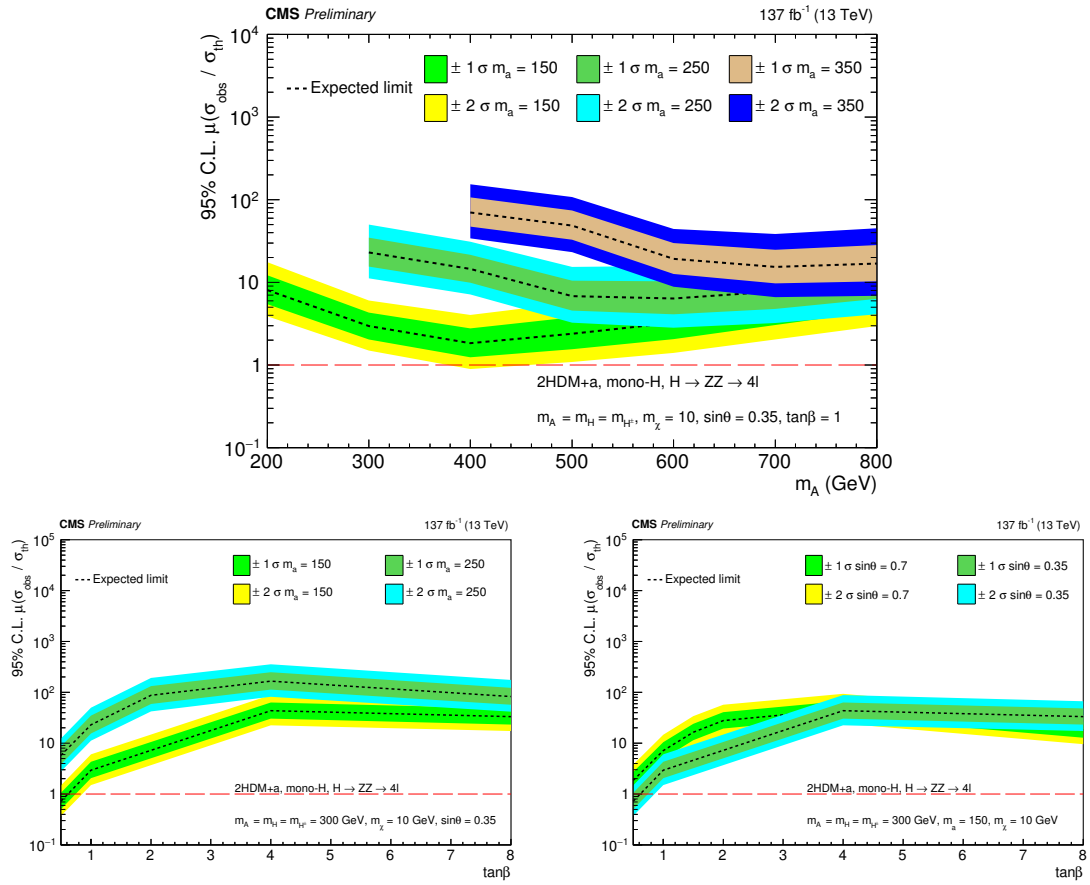


Figure 8.2: The 95 % CL upper limit on the signal strength for the combination of the three final states and three years of data taken as a function of: (top) the heavy pesudoscaler mass "A" and different mass values of the light pesudoscaler "a", (bottom left) $tan\beta$ values and different mass values of the light pesudoscaler "a", (bottom right) $tan\beta$ values and different values of $sin\theta$.

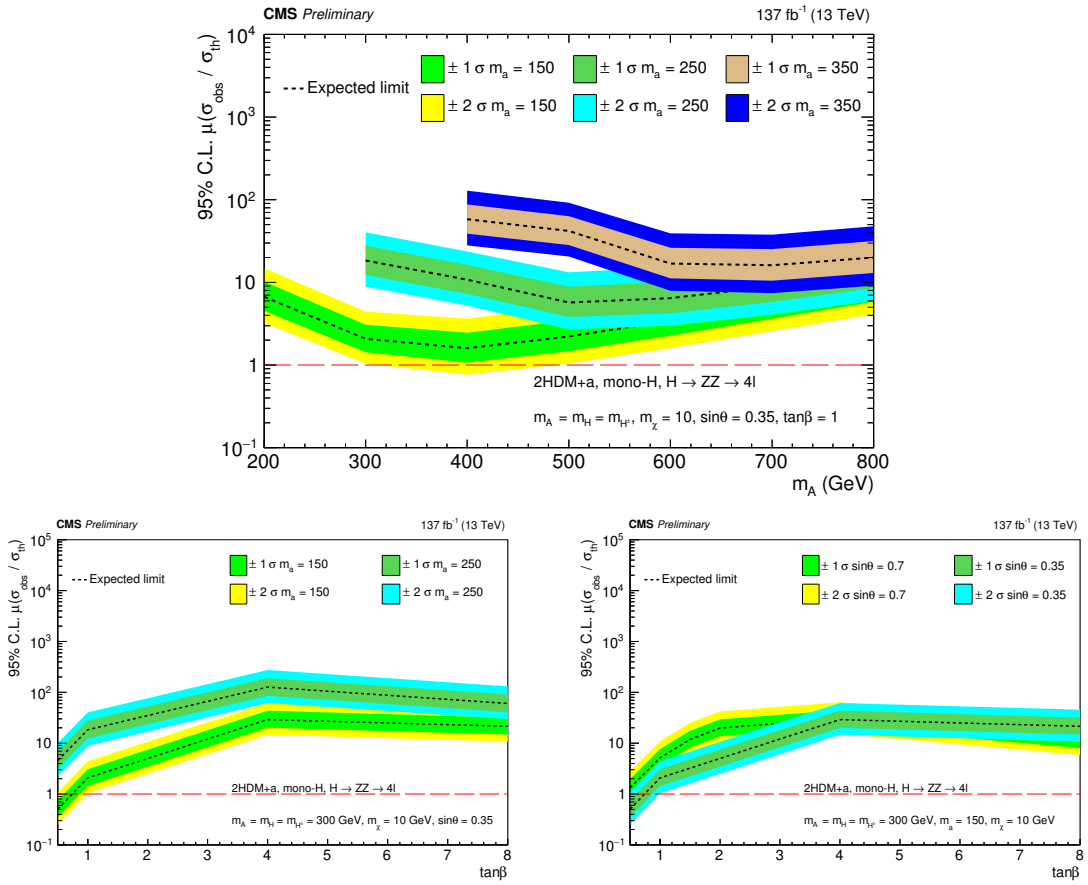


Figure 8.3: The 95 % CL upper limit on the signal strength for the combination of the three final states and three years of data taken as a function of: (top) the heavy pesudoscaler mass "A" and different mass values of the light pesudoscaler "a", (bottom left) $tan\beta$ values and different mass values of the light pesudoscalar "a", (bottom right) $tan\beta$ values and different values of $sin\theta$.

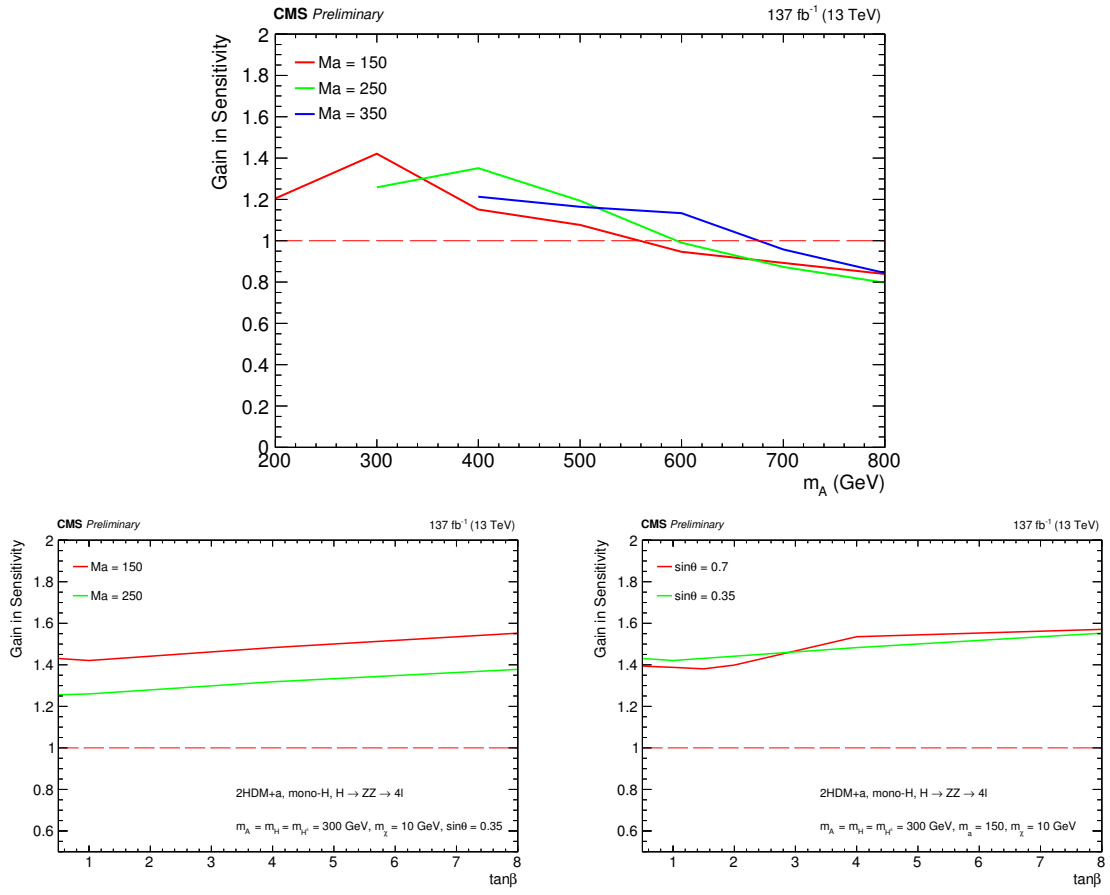


Figure 8.4: The gain obtained in the sensitivity from using BDT approach in shape analysis rather than using p_T^{miss} distributions.

region compared to the low mass region. The bottom plots shows the gain as a function of the $tan\beta$ values, and the BDT approach improves the sensitivity between 20% and 40% for all the $tan\beta$ values.

As it can be observed, the sensitivity of the Mono-Higgs, $H \rightarrow ZZ \rightarrow 4l$, search to the predictions of the 2HDM+a model is limited and only one point of the parameter space of the model can be excluded at 95% CL. The limitation of the sensitivity reflects the fact that the $H \rightarrow ZZ \rightarrow 4l$ branching ratio is very small, but when the analysis will be combined with other Higgs decay channels analysis, it will become possible to exclude more points of the parameter space of the 2HDM+a model.

8.4 Perspectives for RunIII and HL-LHC

According to the timeline for the LHC and HL-LHC operation shown in Figure 2.2, the LHC is scheduled to operate in 2022 after long shut down "LS2"; this operation period is known as "Run III" where the expected integrated luminosity is 300 fb^{-1}. This period of data taken

will be followed by a period of long shutdown for the LHC upgrade known as "LS3". After LS3 the LHC operation will start again with a goal of increasing the integrated luminosity by factor 10 and expected to collect 3000 fb^{-1} of data at center of mass energy $\sqrt{s}$ =14 TeV. This increase of the luminosity is a good opportunity to study rare physics processes. During the LS3 phase, the CMS subdetectors will be upgraded, in addition to an upgrade of the trigger system which will lead to an improvement of the CMS overall performance.

The aim of this section is to check the sensitivity of the Mono-Higgs analysis during Run III and HL-LHC periods after collecting a huge amount of data. The results will be measured as an upper limit on monoHiggs signal strength after extrapolating the results obtained in 2018 analysis to higher luminosity values. We will present the results according to two scenarios as follow:

- Senario 1 "S1": this is a pessimistic scenario where we assume that the systematic uncertainty will not be changed, only an increase in the integrated luminosity is considered.

- Senario 2 "S2": this is an optimistic scenario where we consider that the systematic uncertainties will decrease after collecting large amount of data and the sub-detectors upgrade. In this scenario, we assume that the systematic uncertainty on lepton reconstruction and identification to be decreased to 1 %, the uncertainty on the integrated luminosity to be decreased to 1 %, the uncertainty on the background estimated from data driven method to be decreased to half of its value in 2018 analysis and the shapes uncertainties on the missing energy to be neglected. We will leave the theoretical uncertainties untouched.

In this study, we didn't take into account the increase of the production cross section of the physics processes after increasing the center of mass energy from 13 to 14 TeV. This increase in the energy can result on a limited gain but we will keep the normalization of the processes to the corresponding cross section at 13 TeV. The results are extracted using a binned maximum likelihood fit performed on the p_T^{miss} distribution.

Figure 8.5 shows the 95 % CL upper limit on the signal strength as a function of the heavy pseudoscalar mass "A" for different years of data taken and for RunIII, HL-LHC periods taking into account scenario 1. The figure shows the gain in sensitivity obtained with the increase of the integrated luminosity. According to S1, the increase in the analysis sensitivity is limited during Run III while it improves significantly during HL-LHC period. Figure 8.6 shows the results taking into account S2 where a reduction of systematic uncertainties have been assumed. In this plot, the S2 is only valid for RunII and HL-LHC phases. Also the results are pessimistic for Run III and a significant gain obtained in HL-LHC period. To have an idea about the impact of the reduction of systematics, a comparison between S1 and S2 results is shown in Figure 8.7. The reduction of the systematics uncertainties reflects a very limited gain in sensitivity between 2- 4 % in the low mass region, which indicates that the

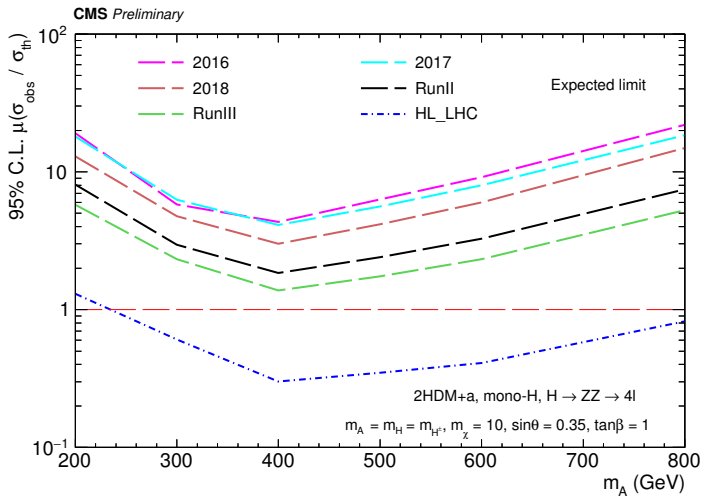


Figure 8.5: The 95 % CL upper limit on the signal strength for different data taken period according to scenario 1 as described in the text.

gain in sensitivity obtained in HL-LHC phase is totally driven by the statistical uncertainties. The results of Run III are discouraging because they indicate a little improvements in the sensitivity while a significant gain is expected in the HL-LHC phase. We still can have a hope of an improvement with increasing the center of mass energy and using a multivariate analysis that already reflected an increase of sensitivity in Run II analysis.

This chapter provided a description of the statistical procedure used to extract the final results. The results of the MonoHiggs ZZ $\rightarrow$ 4l analysis have been presented as an upper limit on the signal strength using the cut based analysis and MVA approaches moreover; a comparison between the obtained results has been presented. A prospective view on the MonoHiggs analysis during RunII and HL-LHC is presented. At this point, we reached the end of our journey in this thesis and I will summarize the results obtained during this journey in the next chapter.

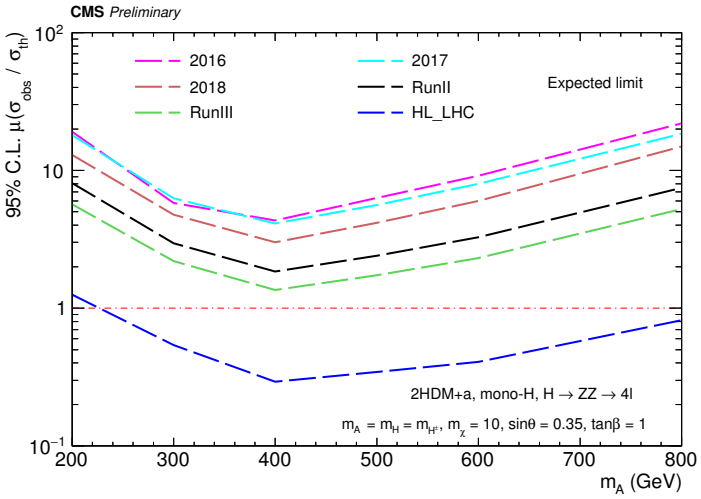


Figure 8.6: The 95 % CL upper limit on the signal strength for different data taken period according to scenario 2 as described in the text.

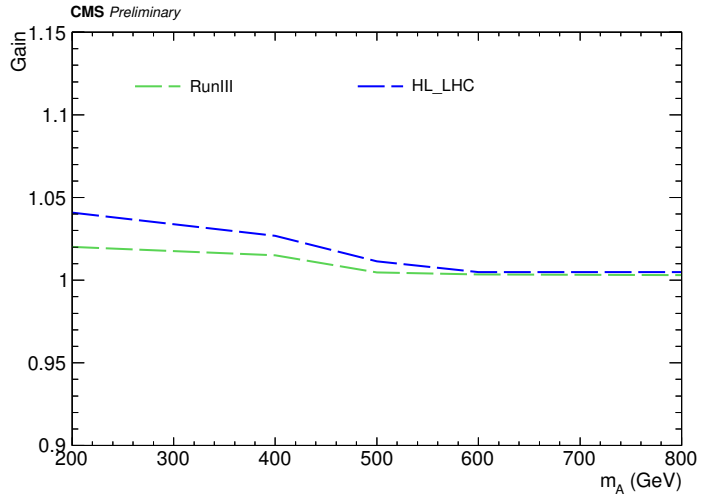


Figure 8.7: The expected gain in sensitivity from improving the systematic uncertainties during Run III and HL-LHC phase.

Summary and Conclusions

Cosmological observations indicate the existence of the dark matter in our universe. Those measurements don't provide information about the Dark Matter (DM) nature, but they provide constraints on its properties. The Standard Model (SM) of particle physics which is considered as a successful theory doesn't provide any dark matter candidate. Therefore, new physics models beyond the SM theory are needed to overcome this limitation. If the DM interacts with SM particles, it can be produced in proton proton collision at colliders such as Large Hadron Collider (LHC). This provides a good opportunity to study the dark matter interactions in detail. The discovery of the Higgs boson opened a new window for the dark matter search at LHC, if the DM particles have mass and interact through the weak force they can couple to the Higgs boson.

In this thesis, the search for DM particles produced in association with a Higgs boson has been presented. The search is performed using proton proton collision data collected by CMS experiment during Run II and corresponding to total integrated luminosity of 137 fb^{-1} at center of mass energy 13 TeV. The Higgs boson decay channel considered in the search is the Higgs decay to a pair of Z bosons which decay to four leptons with three possible final states (4μ, 4e and $2e2\mu$) plus significant amount of missing transverse energy. This search is called "Mono-Higgs" search, where the Higgs boson is used to tag the reconstructed event. Although this decay channel suffers from low branching ratio, it has the advantage of clean final state with relatively small background contamination. The analysis strategy consists of defining a signal region where dark matter signal is enhanced over the standard model backgrounds. The signal region selection is performed by exploiting the kinematic difference between the dark matter signal and the background processes. One signal model (Two Higgs doublet model extended by light pseudoscalar denoted as "2HDM+a") has been used as benchmark model to interpret the results. This model has been inspected in the Higgs $\rightarrow$ ZZ $\rightarrow$ 4l decay channel in CMS in this thesis, for the first time. In this model, a heavy pseudoscalar A is produced from proton proton collision by gluon gluon fusion or quark-antiquark annihilation and decays to SM Higgs boson and a light pseudoscalar a, which in turn decays to a pair of DM particles χ. To increase the analysis sensitivity, multivariate analysis techniques, specifically boosted decision trees, were used. The results obtained are interpreted as an upper limit on the Dark Matter signal strength. By using the boosted decision trees the analysis sensitivity is increased by 20 - 40 %. Even if the analysis is limited by the low branching fraction of

the Higgs decay, it is able to exclude one signal point of the parameter space. The analysis showed high sensitivity in low mass region of the heavy pseudoscalar A; when the analysis will be combined with other Higgs decay channels analyses, it will become possible to rule out more points of the parameter space of the 2HDM+a model.

Besides the dark matter analysis, an aging study is presented to certify the present CMS Resistive Plate Chambers (RPC) system for High Luminosity Large Hadron Collider (HL-LHC) operation. During the HL-LHC period, the present CMS RPC system will be subjected to background rates higher than those for which the detectors have been designed, which could affect the detector properties and induce aging effects. To study whether the present RPC system can sustain the hard background conditions during the HL-LHC running period, a dedicated longevity test is ongoing at the CERN Gamma Irradiation Facility, where a few spare RPCs are exposed to high gamma radiation for a long term period to mimic the HL-LHC operational conditions. During the longevity test, the main detector parameters were continuously monitored as a function of the integrated charge. After collecting a significant amount of the total irradiation, around 89 % (this presents accumulated charge already correspond to twice the maximum expected integrated charge at HL-LHC), the detector parameters and performance are stable and no evidence of any aging effect has been observed. Within few months, the analysis will conclude the certification of 1056 Resistive Plate Chambers for operation at High Luminosity Large Hadron Collider period after collecting the missing amount of irradiation to complete the test.

Final remark

The analysis presented in this thesis represents the search for dark matter particles in the mono-Higgs (ZZ) decay channel for the CMS full Run II data. Despite the low branching ratio of the Higgs decay to four leptons, the analysis is able to to rule out one signal point of the parameter space. An official signal sample with high statistics would improve the analysis sensitivity especially in the high mass region of the heavy pseudoscalar A where the analysis loses a part of sensitivity due to the low statistics of the private signal samples. The combination of the mono-Higgs (ZZ) analysis with the other mono-Higgs channels (bb, $\gamma\gamma$, $\tau\tau$ and WW) will increase the analysis sensitivity and would be able to exclude more points of the parameter space. The analysis will continue during Run III and with the huge data expected at High Luminosity LHC, we hope to obtain new results and observe new physics.

Bibliography

[1] LHC Higgs Cross Section Working Group. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

[2] CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". *JHEP*, (2013) 081. Doi:https://doi.org/10.1007/JHEP06(2013)081.

[3] "Hubble site". *Phys. Rept.*, https://hubblesite.org/image/4024/gallery.

[4] NASA. "1e 0657-56: Nasa finds direct proof of dark matter". http://chandra.harvard.edu/photo/2006/1e0657/index.html.

[5] Planck collaboration. https://sci.esa.int/web/planck/-/51553-cosmic-microwave-background-seen-by-planck.

[6] XENON100 Collaboration and E. Aprile et al. "XENON100 Dark Matter Results from a Combination of 477 Live Days". arXiv:1609.06154.

[7] D. S. Akerib et al. "Limits on spin-dependent WIMP-nucleon cross section obtained from the complete LUX exposure". *Phys. Rev. Lett*, 118 (2018) 251 - 302.

[8] https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryPlotsEXO13TeVTest.

[9] "the hl-lhc project". https://project-hl-lhc-industry.web.cern.ch/sites/project-hl-lhc-industry.web.cern.ch/files/inline-images/HL-LHC-plan-2020-Plan-2.pdf.

[10] Jason Nielsen. "Fundamentals of LHC Experiments". arXiv:1106.2516 [hep-ex] (2011). doi: 10.1142/9789814350525_0003.

[11] "CMS Luminosity - Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[12] "Public CMS Data Quality Information". https://twiki.cern.ch/twiki/bin/view/CMSPublic/DataQuality.

[13] http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html.

[14] J.M. Campbell, J.W. Huston, and W.J. Stirling. "Hard Interactions of Quarks and Gluons: a Primer for LHC Physics". *Rept.Prog.Phys.*, 70:89(2007).arXiv:hep-ph/0611148. Doi:10.1088/0034-4885/70/1/R02.

[15] CMS Collaboration. "The CMS experiment at the CERN LHC". *JINST*, 3 (2008) S08004.

[16] CMS Collaboration. "Description and performance of track and primary vertex reconstruction with the CMS tracker". *JINST*, 9 (2014) CMS-TDR-005, arXiv:1405.6569v2, https://dx.doi.org/10.1088/1748-0221/9/10/p10009.

[17] CMS Collaboration. "The CMS tracker system project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[18] CMS Collaboration. "The CMS electromagnetic calorimeter project : Technical Design Report". CERN-LHCC-97-033, CMS-TDR-4, (1997). https://cds.cern.ch/record/349375?ln=en.

[19] CMS Collaboration. "The Hadronic Calorimeter Technical Design Report". CERN-LHCC-97-031, CMS-TDR-2, (1997). https://cds.cern.ch/record/357153?ln=en.

[20] CMS Collaboration. "The CMS muon project : Technical Design Report". CERN-LHCC-97-032, CMS-TDR-3, (1997). https://cds.cern.ch/record/343814?ln=en.

[21] CMS Collaboration. "Measurements of properties of the Higgs boson in the four lepton final state at sqrt(s) = 13 TeV with full Run II data". CMS AN -2019/139.

[22] CMS Egamma Results. "Performance of Electron and Photon reconstruction in the 2017 Legacy dataset of the CMS Experiment". CMS DP-2020/024.https://cds.cern.ch/record/2718815/files/DP2020_024.pdf.

[23] CMS Collaboration. "Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-probe technique". *CMS Physics Analysis Summary*, CMS DP -2020/035.

[24] CMS Performance Note. "Muon Identification and Isolation efficiency on full 2016 dataset". CMS DP -2017/007.https://cds.cern.ch/record/2257968/files/DP2017_007.pdf.

[25] CMS Performance Note. "Muon identification and isolation efficiencies with 2017 and 2018 data". CMS DP -2018/042.https://cds.cern.ch/record/2629364/files/DP2018_042.pdf.

[26] CMS Performance Note. "Muon reconstruction performance during Run II". CMS DP -2019/022.https://cds.cern.ch/record/2682902/files/DP2019_022.pdf.

[27] Pushpalatha C. Bhat. "Multivariate Analysis Methods in Particle Physics". *Ann.Rev.Nucl.Part.Sci.*, 61 (2011) 281-309. Doi:10.1146/annurev.nucl.012809.104427.

[28] F.Halzen and A.D.Martin. "Quarks and Leptons: An introductory course in modern particle physics". *John Wiley and Sons,1984.*

[29] S. Baffioni et al. "Electron reconstruction in CMS". Technical Report CMS-NOTE-2006-040, CERN, Geneva, Feb 2006.

[30] S. L. Glashow. "Partial Symmetries of Weak Interactions". *Nucl. Phys.*, 22 (1961) 579-588.

[31] A. Salam and J. C. Ward. "Electromagnetic and weak interactions". *Phys. Lett.*, 13 (1964) 168-171.

[32] S. Weinberg. "A Model of Leptons". *Phys. Rev. Lett.*, 19 (1967) 1264-1266.

[33] J. J. Thomson. "XL. Cathode Rays.". *Philosophical Magazine*, 44(269):293 - 316, 1897. Doi:10.1080/14786449708621070.

[34] E. Rutherford. "The scattering of alpha and beta particles by matter and the structure of the atom". *Philosophical Magazine*, Ser.6, 21:669 - 688, 1911. Doi:10.1080/14786440508637080.

[35] E. Rutherford. "The structure of the atom". *Nature*, 92, 423 (1913). Doi:https://doi.org/10.1038/092423a0.

[36] J. Chadwick. "Possible Existence of a Neutron". *Nature*, 129, 312 (1932). Doi:https://doi.org/10.1038/129312a0.

[37] UA1 Collaboration. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s = 540 GeV". *Phys.Lett.B*, 122 (1983) 103-116 .Doi: https://doi.org/10.1016/0370-2693(83)91177-2.

[38] UA2 Collaboration. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". *Phys.Lett.B*, 122 (1983) 476-485 .Doi: https://doi.org/10.1016/0370-2693(83)91605-2.

[39] UA1 Collaboration. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider". *Phys.Lett.B*, 126 (1983) 398-410 .Doi: https://doi.org/10.1016/0370-2693(83)90188-0.

[40] UA1 Collaboration. "Further evidence for charged intermediate vector bosons at the SPS collider". *Phys.Lett.B*, 129 (1983) 273-282 .Doi: https://doi.org/10.1016/0370-2693(83)90860-2.

[41] UA2 Collaboration. "Evidence for $Z^0 \to e^+e^-$ at the CERN $p\bar{p}$ collider". *Phys.Lett.B*, 129 (1983) 130-140 .Doi: 10.1016/0370-2693(83)90744-X.

[42] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett.*, 13 (1964) 508-509.

[43] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". *Phys. Rev. Lett.*, 13 (1964) 321-323.

[44] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys.Lett. B*, 716 (2012) 1- 29, arXiv:1207.7214v2.

[45] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Phys. Lett. B*, 716 (2012) 30 - 61, arXiv:1207.7235v2.

[46] Planck Collaboration. "Planck 2018 results.I. Overview and the cosmological legacy of Planck". (2018). arXiv:1807.06205.

[47] P. J. E. Peebles and B. Ratra. "The Cosmological constant and dark energy". *Rev.Mod.Phys*, 75 (2003) 559 - 606. arXiv:astro-ph/0207347. Doi: 10.1103/RevModPhys.75.559.

[48] R.aly on behalf of CMS Collaboration. "Search For Dark Matter produced in association With a Higgs Boson in the four lepton final state at 13 TeV With CMS Experiment". *PoS EPS-HEP2019 (2020) 617*, Doi:10.22323/1.364.0617.

[49] R.aly and CMS Collaboration. "Search for dark matter particles produced in association with a Higgs boson in proton-proton collisions at $\sqrt{s} = 13$ TeV.". *JHEP*, 03 (2020) 025. arXiv:1908.01713v2 . Doi: 10.1007/JHEP03(2020)025.

[50] R.Aly on behalf of the CMS Muon Group. "Aging Study on Resistive Plate Chambers of the CMS Muon Detector for HL-LHC". *JINST 15, (2020)*, arXiv:2005.11397v2. Doi:https://iopscience.iop.org/article/10.1088/1748-0221/15/11/C11002.

[51] T2K Collaboration. "Precise Measurement of the Neutrino Mixing Parameter θ_{23} from Muon Neutrino Disappearance in an Off-axis Beam.". *Phys.Rev.Lett.*, 112 (2014) 18. arXiv:1403.1532. Doi: 10.1103/PhysRevLett.112.181801.

[52] M. P. Decowski. "KamLAND's precision neutrino oscillation measurements". *Nucl. Phys. B*, 908 (2016) 52 - 61. Doi: https://doi.org/10.1016/j.nuclphysb.2016.04.014.

[53] Y. Fukuda et al. "Evidence for oscillation of atmospheric neutrinos". *Phys.Rev.Lett*, 81:1562–1567, 1998.

[54] T. Adam et al. "Measurement of the neutrino velocity with the OPERA detector in the CNGS beam". *arXiv:1109.4897*.

[55] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln". Helvetica Physica Acta 6 (1933) 110-127.

[56] Rubin Vera C. and Ford W.Kent Jr. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". *The Astrophysical Journal*, 159 (1970) 379. Doi:10.1086/150317.

[57] K. G. Begeman, A. H. Broeils, and R. H. Sanders. "Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics". *Mon. Not. Roy. Astron. Soc*, 249 (1991) 523. Doi:10.1086/150317.

[58] K. Freese. "Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars". EAS Publ. Ser. 36 (2009) 113-126, arXiv:0812.4005.

[59] M. Bartelmann and P. Schneider. "Weak gravitational lensing". *Phys. Rept.*, vol. 340 (2001), 291-472. Doi: 10.1016/S0370-1573(00)00082-X.

[60] N. Sugiyama W. Hu and J. Silk. "The Physics of microwave background anisotropies". *Nature*, 386 (1997) 37-43. arXiv:astro-ph/9604166v1.

[61] W. Hu and S. Dodelson. "Cosmic microwave background anisotropies". *Ann. Rev. Astron. Astrophys*, 40 (2002) 171-216, arXiv:astro-ph/0110414.

[62] C. Patrignani et al. Particle Data Group Collaboration. "Review of Particle Physics". *Chin. Phys*, C40 (2016), no. 10, 100001.

[63] Ernst Fischer. "The properties of dark matter". arXiv:1104.2525v1.

[64] G. Bertone, D. Hooper, and J. Silk. "Particle dark matter: Evidence, candidates and constraints". *Phys. Rept.*, 405 (2005) 279-390, arXiv:hep-ph/0404175.

[65] DAMA Collaboration and R. Bernabei et al. "First results from DAMA/LIBRA and the combined results with DAMA/NaI". *Eur. Phys. J*, C56 (2008) 333-355, arXiv:0804.2741.

[66] DAMA Collaboration and R. Bernabei et al. "DAMA/LIBRA results and perspectives". *Journal of Physics: Conference Series*, CNNP2017 (2017).

[67] D. S. Akerib et al. "Results from a search for dark matter in the complete LUX exposure". arXiv:1608.07648.

[68] LUX collaboration and D. Woodward et al. "Latest results of the LUX dark matter experiment". *International Journal of Modern Physics: Conference Series*, DOI: 10.1142/S2010194520600022.

[69] IceCube Collaboration and M. G. Aartsen et al. "Constraints on ultra-high-energy cosmic ray sources from a search for neutrinos above 10 PeV with IceCube". arXiv:1607.05886.

[70] PICASSO Collaboration and S. Archambault et al. "Constraints on Low-Mass WIMP Interactions on 19F from PICASSO". *Phys. Lett. B*, 711 (2012) 153-161, arXiv:1202.1240.

[71] PICASSO Collaboration and S. Archambault et al. "Final results of the PICASSO dark matter search experiment". *Astrophys*, 90 (2017) 85-92.

[72] Super-Kamiokande Collaboration and T. Tanaka et al. "An Indirect Search for WIMPs in the Sun using 3109.6 days of upward-going muons in Super-Kamiokande". *Astrophys*, J. 742 (2011) 78, arXiv:1108.3384.

[73] Katarzyna Frankiewicz for the Super-Kamiokande collaboration. "A Dark matter searches with the Super-Kamiokande detector". *Journal of Physics: Conference Series*, 888 (2016) 4- 9.

[74] H.E.S.S. collaboration and F. Aharonian et al. "Spectrum and variability of the Galactic Center VHE gamma-ray source HESS J1745-290". *Astronomy and Astrophysics*, 503 (3) 817 - 825 (2019). arXiv:0906.1247v2. Doi: `https://doi.org/10.1051/0004-6361/200811569`.

[75] J. Aleksic et al. "MAGIC Gamma-Ray Telescope Observation of the Perseus Cluster of Galaxies: Implications for Cosmic Rays, Dark Matter and NGC 1275". *Astrophys. J.*, 710 (2010) 634 - 647. arXiv:0909.3267v3 . Doi: `10.1088/0004-637X/710/1/634`.

[76] Fermi-LAT Collaboration and M. Ackermann et al. "Measurement of separate cosmic-ray electron and positron spectra with the Fermi Large Area Telescope". *Phys. Rev. Lett.*, 108 (2012) 011103, arXiv:1109.0521 . Doi: `https://doi.org/10.1103/PhysRevLett.108.011103`.

[77] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer. "The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter". *Phys. Dark Univ.*, 12 (2016) 1-23, arXiv:1402.6703.

[78] M. Ackermann et al. "The Fermi Galactic Center GeV Excess and Implications for Dark Matter". *Astrophys. J.*, 840 (2017) 43 , arXiv:1704.03910. Doi: `10.3847/1538-4357/aa6cab`.

[79] ATIC collaboration. Atic web page. `https://cosmicray.umd.edu/atic-home`.

[80] PAMELA Collaboration and O. Adriani et al. "An anomalous positron abundance in cosmic rays with energies 1.5 -100 GeV". *Nature*, 458 (2009) 607- 609.

[81] M. e. a. Aguilar. "Antiproton flux, antiproton-to-proton flux ratio, and properties of elementary particle fluxes in primary cosmic rays measured with the alpha magnetic spectrometer on the international space station". *Phys. Rev. Lett.*, 117 (2016) 117. Doi: 10.1103/PhysRevLett.117.091103.

[82] Auger collaboration. Atic web page. https://www.auger.org/.

[83] ATLAS Collaboration. "Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector". *Phys. Rev. D*, 4, 032005 (2016). arXiv:1604.07773v2. Doi: 10.1103/PhysRevD.94.032005.

[84] CMS Collaboration. "Search for dark matter, extra dimensions, and unparticles in monojet events in protonproton collisions at $\sqrt{s} = 13$ TeV.". *Eur. Phys. J.*, C75 (2015). arXiv:1408.3583. Doi:https://doi.org/10.1140/epjc/s10052-015-3451-4.

[85] ATLAS Collaboration. "Search for dark matter in events with a hadronically decaying vector boson and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *JHEP*, 10 (2018) 180. arXiv:1807.11471v1. Doi: 10.1007/JHEP10(2018)180.

[86] CMS Collaboration. "Search for dark matter produced in association with a leptonically decaying Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV.". arXiv:2008.04735v1.

[87] ATLAS Collaboration. "Search for dark matter in association with a Higgs boson decaying to b-quarks in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.". *Phys.Lett.B*, 765 (2017) 11-31. arXiv:1807.11471v1. Doi:https://doi.org/10.1016/j.physletb.2016.11.035.

[88] ATLAS Collaboration. "Search for dark matter at $\sqrt{s} = 13$ TeV in final states containing an energetic photon and large missing transverse momentum with the ATLAS detector". *Eur. Phys. J. C*, 77 (2017) 393. arXiv:1704.03848v2. Doi: 10.1140/epjc/s10052-017-4965-8.

[89] CMS Collaboration. "Search for new physics in the monophoton final state in proton-proton collisions at sqrt(s) = 13 TeV.". *J. High Energy Phys.*, 10 (2017) 073. arXiv:1706.03794v2. Doi:10.1007/JHEP10(2017)073.

[90] ATLAS Collaboration. "Search for new resonances in events with one lepton and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.". *Phys.Lett.B*, 762 (2016) 334-352. arXiv:1606.03977v4. Doi:10.1016/j.physletb.2016.09.040.

[91] CMS Collaboration. "Search for heavy gauge W' bosons in events with an energetic lepton and large missing transverse momentum at $\sqrt{s} = 13$ TeV". *Phys.Lett.B*, 770 (2017) 278-301. Doi:http://dx.doi.org/10.1016/j.physletb.2017.04.043.

[92] ATLAS Collaboration. "Search for dark matter produced in association with bottom or top quarks in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector.". *Eur. Phys. J. C*, 78 (2018) 1, 18. arXiv:1710.11412v2. Doi: 10.1140/epjc/s10052-017-5486-1.

[93] ATLAS Collaboration. "Search for dark matter associated production with a single top quark in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector.". *ATLAS-CONF-2020-034*.

[94] CMS Collaboration. "Search for dark matter produced in association with a single top quark or a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV.". *JHEP*, 03 (2019) 141. arXiv:1901.01553v2 . Doi: 10.1007/JHEP03(2019)141.

[95] http://www.lhcp2020.fr.

[96] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/EXOTICS/.

[97] ATLAS Collaboration. "Search for Dark Matter in Events with Missing Transverse Momentum and a Higgs Boson Decaying to Two Photons in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.". *Phys. Rev. Lett.*, 115(13):131801. arXiv:1506.01081. Doi:10.1103/PhysRevLett.115.131801.

[98] ATLAS Collaboration. "Search for dark matter produced in association with a Higgs boson decaying to two bottom quarks in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector.". *Phys. Rev. D*, 93(7):072007, 201. arXiv:1510.06218. Doi:10.1103/PhysRevD.93.072007.

[99] CMS Collaboration. "Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos and b quarks at $\sqrt{s} = 13$ TeV". *JHEP*, 11 (2018) 172. arXiv:1807.02826. Doi:10.1007/JHEP11(2018)172.

[100] CMS Collaboration. "Search for dark matter produced in association with a Higgs boson decaying to a pair of bottom quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Eur. Phys. J.*, C 79 (2019) 280. arXiv:1811.06562. Doi:10.1140/epjc/s10052-019-6730-7.

[101] CMS Collaboration. "Search for dark matter produced in association with a Higgs boson decaying to $\gamma\gamma$ or $\tau\tau$ at $\sqrt{s} = 13$ TeV". *JHEP*, 09 (2018) 046. arXiv:1806.04771. Doi:10.1007/JHEP09(2018)046.

[102] J .Abdallah et al. "Simplified Models for Dark Matter Searches at the LHC". *Phys. Dark Univ.*, 9-10 (2015) pp. 8-23, arXiv: 1506.03116. Doi:http://dx.doi.org/10.1016/j.dark.2015.08.001.

[103] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". *Phys. Rept.*, 516 (2012) pp. 1-102, arXiv: 1106.0034. Doi:10.1016/j.physrep.2012.02.002.

[104] Martin Bauer, Ulrich Haisch, and Felix Kahlhoefer. "Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators". *JHEP 05 (2017) p. 138*, arXiv:1701.07427. Doi:10.1007/JHEP05(2017)138.

[105] Giorgio Arcadi, Giorgio Busoni, Thomas Hugle, and Valentin Titus Tenorth. "Comparing 2HDM + scalar and pseudoscalar simplified models at LHC". *JHEP 89 (2020)*, arXiv: 2001.10540. Doi:https://doi.org/10.1007/JHEP06(2020)098.

[106] L. Evans et al. "LHC Machine". *JINST*, 3 (2008) S08001.

[107] The LHC Study Group. "The Large Hadron Collider : conceptual design". CERN/AC-95-05, 1995, LHC - CERN Geneva.

[108] CMS Collaboration. "Evidence for the direct decay of the 125 GeV Higgs boson to fermions". *Nature Phys.*, 10 (2014).

[109] CMS Collaboration. "Observation of a new boson at amass of 125 GeV with the CMS experiment at the LHC". *Phys. Lett. B*, 716 (2012), https://dx.doi.org/10.1016/j.physletb.2012.08.021.

[110] CMS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". *Phys. Lett. B*, 716 (2012), https://dx.doi.org/10.1016/j.physletb.2012.08.020.

[111] CERN Press Office (10 September 2008). "First beam in the LHC - Accelerating science". https://home.cern/news/press-release/cern/first-beam-lhc-accelerating-science.

[112] Paul Rincon (23 September 2008). "Collider halted until next year. http://news.bbc.co.uk/2/hi/science/nature/7632408.stm.

[113] CERN Press Office (23 November 2009). "Two circulating beams bring first collisions in the LHC". https://home.cern/news/press-release/cern/two-circulating-beams-bring-first-collisions-lhc.

[114] CERN Press Office (31 January 2011). "CERN announces LHC to run in 2012". Press release: https://home.cern/news/press-release/cern/cern-announces-lhc-run-2012.

[115] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. "LHC Design Report Vol. 3: The LHC Injector Chain". CERN-2004-003-V3, 2004.

[116] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". *JINST*, 3 (2008) S08003.

[117] ALICE Collaboration. "The ALICE experiment at the CERN LHC". *JINST*, 3 (2008) S08002.

[118] R. S. Bhalerao and R. V. Gavai. "Heavy Ions at LHC: A Quest for Quark-Gluon Plasma". arXiv:0812.1619v1 [hep-ph] (2008). doi: 10.1007/978-81-8489-295-6_8.

[119] LHCb Collaboration. "The LHCb Detector at the LHC". *JINST*, 3 (2008) S08005.

[120] LHCb Collaboration. "Observation of CP violation in charm decays". *Phys. Rev. Lett.*, 122, 211803 ,arXiv:1905.05428v1 [hep-ex]. doi: 10.1103/PhysRevLett.122.211803.

[121] CMS Collaboration. "CMS Technical Design Report for the Pixel Detector Upgrade". Techincal Design Report CMS CERN Geneva, (2012).

[122] CMS Collaboration. "Upgrade of the CMS tracker". *JINST*, 9 (2014), https://iopscience.iop.org/article/10.1088/1748-0221/9/03/C03041.

[123] P.Adzic et al. "Energy resolution of the barrel of the CMS electromagnetic calorimeter". *JINST 2, P04004, 2007.*

[124] The CMS Collaboration. "CMS experiment at the CERN LHC". *JINST, vol. 3 (2008)*, Doi: http://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[125] CMS Collaboration. "The CMS Magnet Project: Technical Design Report". CERN-LHCC-97-010, (1997), https://cds.cern.ch/record/331056.

[126] CMS Collaboration. "Performance of the CMS drift tube chambers with cosmic rays". Journal of Instrumentation, vol. 5, T03015, (2010).

[127] CMS Collaboration. "Performance of the CMS cathode strip chambers with cosmic rays". Journal of Instrumentation, vol. 5, T03018, (2010).

[128] R.Santonico and R.Cardarelli. "Development of resistive plate counters". Nuclear Instruments and Methods in Physics Research 187, 1981.

[129] M. Abbrescia, P. Fonte, and V. Peskov. "Resistive Gaseous Detectors: designs, performance, and perspective". Wiley-VCH, 2018.

[130] G. Pugliese and The CMS Muon Collaboration. "The RPC system for the CMS experiment". 2006 IEEE NSS Conference Record N24-3, (2007).

[131] CMS Muon Collaboration. "Technical Proposal for the Phase-II Upgrade of the CMS Detector". Technical Report CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02, 2015.

[132] CMS Muon Collaboration. "The Phase-2 Upgrade of the CMS Muon Detectors". CERNLHCC-2017-012. CMS-TDR-016, 2017.

[133] K.Shchablo et al. "Performance of the CMS RPC upgrade using 2D fast timing readout system". Nuclear Instruments and Methods in Physics Research Section A, 2019.

[134] V. Khachatryan et al. "The CMS trigger system". JINST 12, P01020 (2017).

[135] CMS Collaboration. "CMS The TriDAS Project:Technical Design Report". CERN-LHCC-2002-026, 2002.https://cds.cern.ch/record/578006.

[136] A. Tapper et al. [CMS Collaboration]. "CMS Technical Design Report for the Level-1 Trigger Upgrade". CERN-LHCC-2013-011. CMS-TDR-12, 2013.

[137] CMS Trigger and Data Acquisition Group. "The CMS high level trigger". Eur.Phys. J.C, vol. 46, (2006). https://dx.doi.org/10/1140/epjc/s2006-02495-8.

[138] M. Abbrescia, P. Fonte, and V. Peskov. "Resistive Gaseous Detectors: designs, performance and perspectives". Wiley-VCH, 2018.

[139] M. Abbrescia et al. "Properties of C2H2F4 based gas mixtures for avalanche mode operation of Resistive Plate Chambers". CMS Note 97/004.

[140] M. Abbrescia et al. "Eco-friendly gas mixtures for Resistive Plate Chambers based on tetrafuoropropene and Helium". *JINST*, vol. 11, P08019 (2016).https://doi.org/10.1088/1748-0221/11/08/p08019.

[141] B. Mandelli R. Guida and G. Riogoletti. "Performance studies of RPC detectors with new environmentally friendly gas mixtures in presence of LHC-like radiation background". *Nucl. Instr. Meth.*, (2019).https://doi.org/10.1016/j.nima.2019.04.027.

[142] M. Abbrescia et al. "HF production in CMS Resistive Plate Chambers". *Nucl. Phys. Proc. Suppl. 158 (2006) 30. doi: 10.1016/j.nuclphysbps.2006.07.002.*

[143] G. Pugliese et al. "Long-term performance of double gap resistive plate chambers under gamma irradiation". Nuclear Instruments and Methods in Physics Research A 477, 2002.

[144] M. Abbrescia et al. "Study of long-term performance of CMS RPC under irradiation at the CERN GIF". Nuclear Instruments and Methods in Physics Research A, 2004. https://doi.org/10.1016/j.nima.2004.07.009.

[145] H. C. Kim et al. "Quantitative aging study with intense irradiation tests for the CMS forward RPCs". Nuclear Instruments and Methods in Physics Research A, 2009. https://doi.org/10.1016/j.nima.2008.12.140.

[146] CMS Collaboration. "The Phase-2 Upgrade of the CMS Muon Detectors". *CMS-TDR-016*, CERN/LHCC 2017-12, (2017).

[147] A. Gelmi. "Long-term aging studies on Resistive Plate Chambers (RPC) of the CMS muon system for HL-LHC". *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, (2018). Doi:10.1109/NSSMIC.2018.8824516.

[148] EN R. Guida, EP, and AIDA GIF++ Collaborations. "A new CERN Irradiation Facility to test large-area detectors for the HL-LHC program". *PoS, ICHEP*, vol. 2016, (2016). https://doi.org/10.22323/1.282.0260.

[149] S. Colafranceschi et al. "Resistive plate chambers for 2013-2014 muon upgrade in CMS at LHC". *Journal of Instrumentation 9(10)*, (2014). 10.1088/1748-0221/9/10/C10033.

[150] G. Aielli et al. "Further advances in aging studies for RPCs". Nuclear Instruments and Methods in Physics Research A, 2003. https://doi.org/10.1016/j.nima.2003.09.020.

[151] D. Domenici et al. "An extensive aging study of bakelite Resistive Plate Chambers". Nuclear Instruments and Methods in Physics Research A 518 (2004) 82-85. https://doi.org/10.1016/j.nima.2003.10.030.

[152] Filip Thyssen. "Performance of the Resistive Plate Chambers in the CMS experiment". 2012 JINST 7 C01104. doi:10.1088/1748-0221/7/01/C01104.

[153] G. Pugliese et al. "Aging study for resistive plate chambers of the CMS muon trigger detector". Nuclear Instruments and Methods in Physics Research A, 2003. https://doi.org/10.1016/j.nima.2003.09.021.

[154] B. R. Webber. "A QCD Model for Jet Fragmentation Including Soft Gluon Interference". *Nucl. Phys. B*, vol. 238 (1984). Doi: 10.1016/0550-3213(84)90333-X.

[155] B.Andersson, G.Gustafson, G.Ingelman, and T.Sjstrand. "Parton fragmentation and string dynamics". *Phys. Rept.*, vol. 97 (1983). Doi: 10.1016/0370-1573(83)90080-7.

[156] T. Sjostrand, Stephen Mrenna, and Peter Skands. "A brief introduction to PYTHIA 8.1". *Comput.Phys.Commun.*, 187 (2008), doi:10.1016/j.cpc.2008.01.036.

[157] J. Alwall, R. Frederix, and S. Frixione et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". *JHEP*, 97 (2014), doi:https://doi.org/10.1007/JHEP07(2014)079.

[158] C. Oleari. "The powheg box". *Nuclear Physics B*, 205 (2010) 36- 41. Doi: https://doi.org/10.1016/j.nuclphysbps.2010.08.016.

[159] S. Agostinelli et al. "Geant4 : a simulation toolkit". *Nucl. Instum. Meth. A,*, 506(3).

[160] CMS Collaboration. "Particle flow reconstruction and global event description with the CMS detector". *JINST*, 12 (2017),https://doi.org/10.1088/1748-0221/12/10/P10003.

[161] P. Billoir. "Progressive track recognition with a Kalman like fitting procedure". *Comput. Phys. Commun*, 57:390–394, 1989.

[162] R.Fruhwirth. "Application of Kalman Filtering to Track and Vertex Fittting". *Nucl.Instrum.Meth.*, A262:444–450, 1987.

[163] CMS Collaboration. "Track Reconstruction Performance in CMS". *Nucl. Phys. Proc.Suppl.*, 197:275-278, 2009.

[164] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". *JINST*, 9(10):P10009, 2014.

[165] CMS Collaboration. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV". *JINST*, 10(06):P06005, 2015.

[166] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". *JINST*, Volume 9, October 2014.

[167] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST*, 13 (2018) P06015. doi:`0.1088/1748-0221/13/06/P06015`.

[168] CMS Collaboration. "Performance of Muon Identification in pp collisions at sqrt(s)= 7 TeV". *JINST*, 7 (2012), `https://doi.org/10.1088/1748-0221/7/10/P10002`.

[169] N. Adam et al. G. Abbiendi. "Muon Recostruction in the CMS Detector". MS AN 2008/097, July 2009.

[170] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov. "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at LHC". arXiv:physics/0306087v1. doi: `10.1088/0954-3899/31/9/N01`.

[171] CMS Collaboration. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". *JINST*, 14 (2019):P07004.

[172] CMS Collaboration. "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV". *Journal of Instrumentation*, 10 (2015).

[173] Luminosity Physics Object Group. `https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM#SummaryTable`.

[174] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". *JHEP*, 11:070, 2007.

[175] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. "NLO vector-boson production matched with shower in POWHEG". *JHEP*, 807:060(2008). Doi:10.1088/1126-6708/2008/07/060.

[176] Paolo Nason. "A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms". *JHEP*, 0411:040(2004). Doi:10.1088/1126-6708/2004/11/040.

[177] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini. "Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM". *JHEP*, 02 (2012) 088.

[178] P. Nason and C. Oleari. "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG". *JHEP*, 02 (2010) 037. doi:10.1007/JHEP02(2010)037.

[179] Heribertus B. Hartanto, Barbara Jager, Laura Reina, and Doreen Wackeroth. "Higgs boson production in association with top quarks in the POWHEG BOX". *Phys. Rev. D*, 91, 094003 (2015). doi:10.1103/PhysRevD.91.094003.

[180] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano. "$HW^{\pm}/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO". *JHEP*, 101310 (2013) 083. doi:10.1007/JHEP10(2013)083,arXiv:1306.2542.

[181] Y. Gao et al. "Spin determination of single-produced resonances at hadron colliders". *Phys. Rev. D*, 81 (2010) 075022. doi:10.1103/PhysRevD.81.075022.

[182] A. V. G. et. al., S. Bolognesi, and Y. Gao. "JHUGen". https://spin.pha.jhu.edu.

[183] Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. CERN Yellow Reports: Monographs Volume 2/2017 (CERN–2017–002-M). Doi:10.23731/CYRM-2017-002.

[184] C. Anastasiou et al. "High precision determination of the gluon fusion Higgs boson cross-section at the LHC". *JHEP*, 05:058 (2016).

[185] J. M-Campbell and R. K. Ellis. "MCFM for the Tevatron and the LHC". *Nucl.Phys.Proc.Suppl.*, 205-206:10-15 (2010). doi:10.1016/j.nuclphysbps.2010.08.011.

[186] "Summary table of samples produced for the 1 Billion campaign, with 25ns bunch-crossing". https://twiki.cern.ch/twiki/bin/viewauth/CMS/SummaryTable1G25ns.

[187] CMS Collaboration. "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Journal of High Energy Physics*, 11 (2018).arXiv:1711.02547v2. Doi:10.1007/JHEP08(2018)011.

[188] CMS Collaboration. "Performance of CMS muon reconstruction in pp collision events at sqrt(s) = 7 TeV". *Technical report*, Muo-10-004, 2010.

[189] CMS Collaboration. "Measurement of the properties of the Higgs boson in the four lepton final state at $\sqrt{s}$ =13 TeV". CMS Physics Analysis Note CMS-AN-15-277, 2016.

[190] R. Salerno and M. Ahmad. "Measurement of the properties of the Higgs boson in the four lepton final state at $\sqrt{s}$ =13 TeV". CMS AN 2016/442 (2016).

[191] T. Sculac and D. Sperka. "Measurement of the properties of the Higgs boson in the four lepton final state at $\sqrt{s}$ =13 TeV with 2017 data". CMS AN 2017/342 (2017).

[192] CMS Collaboration. "New results on the study of Higgs boson production in the four lepton final state at $\sqrt{s}$ =13 TeV". CMS Physics Analysis Note CMS-AN-16-217, 2016.

[193] Chang Liu and et al. "Reconstruction of cosmic and beam-halo muons with the CMS detector". *Eur. Phys. J. C*, (2008) 56: 449-460. Doi:10.1140/epjc/s10052-008-0674-7.

[194] W. Bialas and D. A. Petyt. "Mitigation of anomalous APD signals in the CMS ECAL". *Journal of Instrumentation*, 8(03):C03020, 2013. Doi:10.1088/1748-0221/8/03/C03020.

[195] CMS Collaboration. "Identification and Mitigation of Anomalous Signals in CMS Hadronic Calorimeter". *Journal of Physics: Conference Series*, 404(1): 012044, 2012. Doi:10.1088/1742-6596/404/1/012044.

[196] JetMET POG. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2.

[197] CMS Collaboration. "Mitigation of anomalous missing transverse momentum measurements in data collected by CMS at $\sqrt{s}$ = 13 TeV during the LHC Run 2". CMS Performance Note, CMS DP -2020/018. https://cds.cern.ch/record/2714938/files/DP2020_018.pdf.

[198] JetMET POG. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis#Type_I_Correction.

[199] JetMET POG. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis#xy_Shift_Correction.

[200] M. Grazzini, S. Kallweit, and D. Rathlev. "ZZ production at the LHC: Fiducial cross sections and distributions in NNLO QCD". *Phys. Lett B*, 750:407-410, 2015. doi: https://doi.org/10.1016/j.physletb.2015.09.055.

[201] M. Bonvini, F. Caola, Fabrizio, S. Forte, K. Melnikov, and G. Ridolfi. "Signal background interference effects in $gg \to H \to W^+W^-$ beyond leading order". *PhysRev.*, D88(3):034032, 2013. doi: https://doi.org/10.1103/PhysRevD.88.034032.

[202] K. Melnikov and M. Dowling. "Production of two Z-bosons in gluon fusion in the heavy top quark approximation". *Phys. Lett., B*, 744:43-47, 2015. doi: 10.1016/j.physletb.2015.03.030.

[203] C.S. Li et al. "Soft gluon resummation in the signal-background interference process of $gg \to H \to ZZ$". *JHEP*, 08:065, 2015. doi: 10.1007/JHEP08(2015)065.

[204] G. Passarino. "Higgs CAT". *Eur. Phys. J.*, C74:2866, 2014. doi: 10.1140/epjc/s10052-014-2866-7.

[205] et al. A. Hoecker. "TMVA - Toolkit for Multivariate Data Analysis. arXiv:physics/0703039v5.

[206] CMS Collaboration. "CMS luminosity measurements for the 2016 data-taking period". CMS-PAS-LUM-17-001. CERN, Geneva, 2017.

[207] CMS Collaboration. "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV". CMS-PAS-LUM-17-004. CERN, Geneva, 2018.

[208] CMS Collaboration. "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV". CMS-PAS-LUM-18-002. CERN, Geneva, 2019.

[209] LHCHXSWG. "SM Higgs production cross sections at $\sqrt{s} = 13$, 14 and 27 TeV (update in CERN HL-LHC YR 2019)". CMS Twiki (2019). https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNHLHE2019.

[210] J. Butterworth et al. "PDF4LHC recommendations for LHC Run II". *J. Phys.*, G43:023001, 2016.

[211] J. C. Collins, D. E. Soper, and G. F. Sterman. "Factorization of Hard Processes in QCD". *Adv. Ser. Direct. High Energy Phys.*, 5 (1989) 1-91. arXiv:hep-ph/0409313.

[212] G. Cowan. "Statistical data analysis". Clarendon (Oxford), Oxford, 1998.

[213] The CMS Collaboration T. L. H. C. G. The ATLAS Collaboration. "Procedure for the LHC Higgs boson search combination in Summer 2011". Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011.

[214] ATLAS, CMS Collaborations, and LHC Higgs Combination Group. "Procedure for the LHC Higgs boson search combination in Summer 2011". *CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, 2011.*

[215] K. Cranmer. "Practical Statistics for the LHC". in Proceedings, 2011 European School of High-Energy Physics (ESHEP 2011): Cheile Gradistei, Romania, September 7-20, 2011, pp. 267-308. 2015. arXiv:1503.07622.

[216] A. L. Read. "Presentation of search results: The CL(s) technique". *J. Phys.*, G28 (2002) 2693-2704, [11(2002)].

[217] T. Junk. "Confidence level computation for combining searches with small statistics". *Nucl. Instrum.*, Meth. A 434 (1999) 435, arXiv:hep-ex/9902006. Doi:10.1016/S0168-9002(99)00498-2.

[218] T. M. Karbach and M. Schlupp. "Constraints on Yield Parameters in Extended Maximum Likelihood Fits". arXiv:1210.7141.