

Tau Lepton: as a Tool for Hunting Standard Model and Beyond Standard Model Higgs

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik, Informatik und Naturwissenschaften

der Universität Hamburg

vorgelegt von

Yiwen Wen

aus Shaoguan, China

Hamburg

2019



Gutachter/in der Dissertation:	Prof. Dr. Elisabetta Gallo Dr. Alexei Raspereza
Mitglieder/in der Prüfungskommission:	Prof. Dr. Dieter Horns Prof. Dr. Elisabetta Gallo Prof. Dr. Gudrid Moortgat-Pick Dr. Alexei Raspereza Dr. Christian Sander
Vorsitzende des Prüfungsausschusses:	Prof. Dr. Dieter Horns
Datum der Disputation:	25 Februar 2019
Vorsitzender Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Michael Potthoff
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang Hansen
Dekan der Fakultät MIN:	Prof. Dr. Heinrich Graener

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hamburg, den 22 January 2019

ABSTRACT

This thesis presents a study of the Standard Model (SM) Higgs boson in the decay channel of $\tau\tau \rightarrow e\mu$, as well as a search for neutral Minimal Supersymmetric Standard Model (MSSM) Higgs boson in the same decay channel, using data corresponding to an integrated luminosity of 35.9 fb^{-1} in proton-proton collisions with a center-of-mass energy of 13 TeV, collected by the CMS detector at the Large Hadron Collider. The results of $e\mu$ channel in both analyses are combined with other channels of Higgs to $\tau\tau$ decay: $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$.

An observation of the SM Higgs boson to $\tau\tau$ is established with a significance of 4.9 standard deviations, to be compared to an expected significance of 4.7 standard deviations. The product of the observed SM Higgs boson to $\tau\tau$ production cross section times the branching ratio is $1.09^{+0.27}_{-0.26}$ times the SM prediction. This is the first observation of Higgs boson decays to a pair of τ leptons by a single experiment.

A search for neutral MSSM Higgs boson is performed in a mass range of 90 GeV to 3.2 TeV. No significant deviations from the SM background expectation are observed. Model independent upper limits on the production cross section times branching ratio are set. The results are interpreted into selected MSSM benchmark scenarios.

ZUSAMMENFASSUNG

Diese Arbeit präsentiert eine Studie des Higgs-Bosons des Standardmodells (SM) im Zerfallskanal $\tau\tau \rightarrow e\mu$, sowie die Suche nach einem neutralen Higgs-Boson des minimalen supersymmetrischen Standardmodells (MSSM) im gleichen Zerfallskanal unter Verwendung von Daten, die einer integrierten Luminosität von $35,9 \text{ fb}^{-1}$ entsprechen und in Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von 13 TeV vom CMS-Detektor am Large Hadron Collider erfasst wurden. Die Ergebnisse der $e\mu$ -Kanäle beider Analysen werden jeweils mit anderen Endzuständen der Tauzerfälle aus dem Higgs-Boson-Zerfall kombiniert: $e\tau_h$, $\mu\tau_h$ und $\tau_h\tau_h$.

Das SM-Higgs-Boson wird mit einer Signifikanz von 4,9 Standardabweichungen im Zerfallskanal in ein Tau-Lepton-Paar beobachtet. Die erwartete Signifikanz beträgt dabei 4,7 Standardabweichungen. Das Produkt aus dem beobachteten Produktionswirkungsquerschnitt des SM-Higgs-Bosons mit dem Verzweigungsverhältnis in ein Tau-Lepton-Paar beträgt $1,09^{+0,27}_{-0,26}$ mal der SM-Vorhersage. Dies ist die erste Beobachtung von Higgs-Boson-Zerfällen in ein Tau-Lepton-Paar durch ein einziges Experiment.

Die Suche nach einem neutralen MSSM-Higgs-Boson wird in dem Massenbereich von 90 GeV bis 3,2 TeV durchgeführt. Es werden keine signifikanten Abweichungen von der SM-Untergrunderwartung beobachtet. Daher werden modellunabhängige Obergrenzen auf den Produktionswirkungsquerschnitt eines neutralen MSSM-Higgs-Bosons bestimmt, wobei das Verzweigungsverhältnis in ein Tau-Lepton-Paar festgelegt wird. Die Ergebnisse werden in ausgewählten MSSM-Referenzszenarien interpretiert.

CONTENTS

1	Introduction	13
2	The Standard Model of Particle Physics	17
2.1	Elementary Particles and Forces	18
2.1.1	Force mediators	18
2.1.2	Leptons	19
2.1.3	Quarks	19
2.1.4	The Higgs boson	19
2.2	The Standard Model Lagrangian	19
2.2.1	Electromagnetic interaction	20
2.2.2	Strong interaction	22
2.2.3	Weak interaction and electroweak unification	22
2.2.4	The Higgs mechanism and spontaneous symmetry breaking	25
2.2.5	The mass of fermions	27
2.3	SM Higgs Boson Production Modes	28
2.4	SM Higgs Boson Decay Modes	28
2.4.1	SM Higgs boson decays to fermions	29
2.4.2	SM Higgs boson decays to bosons	30
3	The Minimal Supersymmetric Standard Model (MSSM)	33
3.1	Motivations for Supersymmetry	33
3.1.1	The hierarchy problem	34
3.1.2	The dark matter	35
3.1.3	Unification of the forces	35
3.2	The Minimal Supersymmetric Standard Model	35
3.3	The Higgs Sector of MSSM	37
3.3.1	The MSSM Higgs parameters	37
3.3.2	The MSSM Higgs benchmark scenarios	39

4	Statistics	41
4.1	Likelihood	42
4.1.1	Basics	42
4.1.2	Introducing nuisance parameters	42
4.2	Maximum Likelihood Fit	45
4.3	Significance	45
4.4	Exclusion Limits	47
4.4.1	Observed Limits	47
4.4.2	Expected Limits	48
4.5	Boosted Decision Tree	49
5	The CMS experiment at the CERN Large Hadron Collider	53
5.1	The CERN Organization	54
5.2	The Large Hadron Collider	56
5.2.1	Proton production and acceleration	57
5.2.2	Experiments at the LHC	60
5.2.3	LHC data taking schedule	60
5.3	The CMS Experiment	61
5.3.1	Overview of the CMS detector	61
5.3.2	Tracker	62
5.3.3	Electromagnetic calorimeter	65
5.3.4	Hadronic calorimeter	66
5.3.5	Magnet	67
5.3.6	Muon system	67
5.3.7	Trigger and data acquisition system	69
6	Event Simulation and Reconstruction	71
6.1	Event Generation	72
6.1.1	Monte Carlo method	72
6.1.2	Structure of an event	72
6.1.3	Hard scattering	74
6.1.4	Parton showering	75
6.1.5	Hadronization	75
6.1.6	Underlying event	75
6.1.7	Event generators	76
6.1.8	Detector simulation	77
6.2	Object Reconstruction	77
6.2.1	Particle-flow Algorithm	78
6.2.2	Primary vertex	78
6.2.3	Electrons	80
6.2.4	Muons	82
6.2.5	Jets	84
6.2.6	B-tagging	85
6.2.7	Missing transverse energy $\vec{E}_T$	86

7	Performance of τ_h Reconstruction and Identification	87
7.1	The τ Lepton Properties	88
7.2	Reconstruction and Identification of τ_h	88
7.2.1	The hadron-plus-strip algorithm	88
7.2.2	Discrimination between τ_h and jets	92
7.2.3	Discrimination against electron	94
7.2.4	Discrimination against muon	96
7.3	Measurement of the $e \rightarrow \tau_h$ Misidentification Rate	97
7.3.1	Datasets and Monte Carlo samples	97
7.3.2	Data-to-Monte Carlo corrections	98
7.3.3	Signals and background modeling	98
7.3.4	Event selection	101
7.3.5	Measurement	101
7.3.6	Systematic uncertainties	102
7.3.7	Results	103
7.4	Measurement of the $\mu \rightarrow \tau_h$ Misidentification Rate	104
7.4.1	Datasets and Monte Carlo samples	105
7.4.2	Data-to-Monte Carlo corrections	105
7.4.3	Background modeling	106
7.4.4	Event selection	106
7.4.5	Measurement	107
7.4.6	Systematic uncertainties	108
7.4.7	Results	109
7.5	Measurement of the Efficiencies of Discriminators against Electron and Muons	111
7.5.1	Strategies	111
7.5.2	Results	111
8	Analysis of SM Higgs to $\tau \tau$ in the $e\mu$ final state	117
8.1	Analysis Overview	118
8.2	Datasets and Monte Carlo Samples	118
8.3	Data-to-Monte Carlo Corrections	119
8.3.1	Pileup	119
8.3.2	Triggers	119
8.3.3	Electrons and muon identification and isolation scale factors	122
8.3.4	Recoil corrections	122
8.3.5	Reweighting of LO Madgraph Drell-Yan Z +jets events	123
8.3.6	Top p_T reweighting	123
8.4	Event Selection	124
8.4.1	Baseline selection	124
8.4.2	The D_ζ variable to suppress top pair production background	125
8.4.3	$\tau\tau$ mass reconstruction	126
8.4.4	Event categorization	128
8.5	Background Modeling	129
8.5.1	Drell-Yan Z +jets	129
8.5.2	$t\bar{t}$ events	129
8.5.3	QCD multijet events	130

8.5.4	Other backgrounds	132
8.6	Control Plots	132
8.7	Signal Extraction	132
8.8	Systematic Uncertainties	132
8.8.1	Uncertainties related to object reconstruction and identification	132
8.8.2	Background estimation uncertainties	136
8.8.3	Signal prediction uncertainties	136
8.8.4	Other uncertainties	137
8.9	Results of $e\mu$ Final State	137
8.10	Combination with $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ Final States	137
8.10.1	Brief description of $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states	137
8.10.2	Results	141
9	Analysis of MSSM Higgs to $\tau\tau$ in the $e\mu$ final state	147
9.1	Analysis Overview	148
9.2	Datasets and Monte Carlo Samples	148
9.3	Data-to-Monte Carlo Corrections	149
9.3.1	Triggers	149
9.3.2	B-tagging	149
9.4	Event Selection	150
9.4.1	Categorization	150
9.5	Background Modeling	150
9.5.1	Drell-Yan Z +jets	150
9.5.2	Refined modeling of QCD multijet events	151
9.5.3	Modeling of $\text{jet} \rightarrow e/\mu$ events	154
9.6	Signal Extraction	156
9.7	Systematic Uncertainties	156
9.7.1	Uncertainties related to object reconstruction and identification	156
9.7.2	Background estimation uncertainties	158
9.7.3	Signal prediction uncertainties	159
9.7.4	Other uncertainties	159
9.8	Results of $e\mu$ Final State	160
9.9	Combination with $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ Final States	160
9.9.1	Brief description of $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states	160
9.9.2	Results	164
10	Summary and conclusions	171
	Appendices	173
A	Publications	175
11	Bibliography	177

CHAPTER

1

INTRODUCTION

The universe as we know today can be described as the composition of elementary particles. These particles together with their interactions form every substance we see, from macroscopic objects like stars to microscopic bacterium and ultimately, intelligent life form: the human beings. The standard model (SM) of particle physics describes the elementary particles and three interactions between them: the electromagnetic, the weak and the strong forces. Results from past high energy experiments, such as experiments at the Large Electron-Positron Collider (LEP) at CERN between 1989 and 2000, at Tevatron at Fermilab between 1983 and 2011, and at the Large Hadron Collider (LHC) since 2010, have shown significant good agreements with the SM predictions. It proves that the SM is a great success, especially after the discovery of the 125 GeV Higgs boson, the final piece of a puzzle in the SM, by the ATLAS and CMS experiments at CERN LHC in July 2012. The Higgs boson was introduced to explain why particles have mass. The electroweak bosons $W^\pm$ and Z^0 acquire mass through the electroweak symmetry breaking in the Higgs mechanism, and the fermions obtain mass from Yukawa couplings with this boson. Francois Englert and Peter Higgs were awarded the 2013 Nobel Physics Prize because of their contribution to the understanding of the origin of mass of elementary particles which was confirmed by the discovery of the Higgs boson in both ATLAS and CMS experiments.

The LHC is a circular proton-proton collider twenty-seven kilometers in circumference, located in a tunnel 150 meters deep below the ground near the Swiss-French border. The high performance of its two general-purpose detectors, ATLAS and CMS, made the discovery of the Higgs boson possible. In the first running period, Run-1 from 2010 to 2012, both detectors collected roughly 25 fb^{-1} of data at a center-of-mass energy 7 TeV and 8 TeV. It allowed studying more details of the properties and couplings of the new particle. In Run 1, no significant deviations from SM were observed. All measurements are compatible with the existence of the new particle within the uncertainties. Run-2 started from summer 2015 at a center-of-mass energy of 13 TeV and it extended to the end of 2018 before a long shutdown.

In the SM, the production of the Higgs boson goes through several mechanisms. At

the LHC, the largest cross-section corresponds to gluon-gluon fusion production, with vector boson fusion production coming next. Although the gluon-gluon fusion production dominates, it is also important to study other production modes to test the compatibility of the boson discovered with the SM prediction. During the LHC Run-1 era, a large amount of decay modes of SM Higgs boson were accessible for experimentalists which allowed the studies of a wide range of physics signatures. The discovery in 2012 was declared by the studies of the bosonic channels ($H \rightarrow \gamma\gamma$, $H \rightarrow W^\pm W^\mp$ and $H \rightarrow ZZ$). However, fermionic decay modes ($H \rightarrow b\bar{b}$ and $H \rightarrow \tau^+\tau^-$) are also important to study the Yukawa coupling of the Higgs boson and to test whether it is compatible with the SM prediction. Among the fermionic decay modes, the most promising channel is $\tau^+\tau^-$, due to the large event rate expected in the SM compared to the $\mu^+\mu^-$ decay channel ($B(H \rightarrow \tau^+\tau^-) = 6.3\%$ for a mass of 125.09 GeV), and to the smaller contribution from background events with respect to the $b\bar{b}$ decay channel.

The SM is a successful theory which is not contradicted by the experimental results within experimental accuracy. However, several questions remain, for example, the underlying mechanism responsible for the symmetry breaking and the existence of dark matter. Theorists have proposed several models to address the shortcomings of the SM. Many of these models predict an extra sector of Higgs bosons and one of the main tasks of the LHC is to further explore it. The Minimal Supersymmetric Standard Model (MSSM) introduces a second scalar doublet in addition to the SM one. It leads to five eigenstates after symmetry breaking: two charged Higgs bosons $H^\pm$, a light and a heavy CP-even (scalar) boson h and H , and a CP-odd (pseudo-scalar) boson A . At tree level, the properties of these particles are described by two free parameters, conventionally chosen as the mass of the pseudoscalar m_A and the ratio of the vacuum expectation values of the two doublets $\tan\beta$. For low and medium $\tan\beta$, the gluon-gluon fusion process dominates, while for high $\tan\beta$, associated production with b-jet is the dominant one. With respect to SM Higgs to $\tau\tau$, the branching ratios of $H \rightarrow \tau\tau$ and $A \rightarrow \tau\tau$ are enhanced in the high $\tan\beta$ parameter space. This thesis reports on a search of MSSM and SM Higgs into $\tau\tau$, using the 2016 data recorded by the CMS detector, in particular in the decay channel in which one τ lepton decays to an electron and the other one to a muon.

The tau lepton, with a mass of $m_\tau = 1.777$ GeV, is the only lepton heavy enough to decay into hadrons. In about one third of the case, taus decay leptonically to a muon or an electron, plus two neutrinos. The muons and electrons produced in these decays are reconstructed and identified with the standard techniques for muons and electrons in CMS. In two thirds of the cases, tau leptons decay hadronically, to a combination of charged and neutral mesons, plus one neutrino. Hadronically decayed taus, τ_h , are reconstructed and identified in CMS with the Hadrons-Plus-Strips (HPS) algorithm, which uses the trajectories measured in the tracker detector and energy deposits in the electromagnetic calorimeter to build τ_h candidates. The algorithm also provides discriminators to separate hadronic decay taus from quark and gluon jets, electrons and muons. The algorithm identifies successfully 65% of hadronically decaying taus, while less than 1% of quark and gluon jets are misidentified as τ_h . This thesis is also devoted to the study of the performance of hadronic tau identification against electrons and muons.

In Chapter 2, the Standard Model, the Higgs boson and the electroweak symmetry breaking theories are presented. Chapter 3 continues with the motivations for supersymmetry as well as MSSM with an extended Higgs sector. In Chapter 4, statistical tools and their working

principles are described. In Chapter 5, the experimental setup is presented with descriptions of the LHC machine and the CMS detector is described. In Chapter 6, the simulations and the physics object reconstruction are presented. The HPS algorithm, which reconstructs and identifies τ_h , is documented in Chapter 7, as well as its performances measured in data. Searches for the decay of the SM Higgs boson to tau leptons in final states are presented in Chapter 8. Searches for the MSSM Higgs boson decaying to tau leptons are presented in Chapter 9. All the physics analyses shown in this thesis used data collected by the CMS detector in 2016 during Run-2 with an integrated luminosity of 35.9 fb^{-1} . I have analyzed the $e\mu$ channel in 2016 for SM Higgs to $\tau\tau$ contributing to its first observation, and put the most stringent limits on MSSM Higgs to $\tau\tau$ up to date. I have also contributed to the τ identification used by the whole CMS collaboration.

CHAPTER

2

THE STANDARD MODEL OF PARTICLE PHYSICS

Contents

2.1	Elementary Particles and Forces	18
2.1.1	Force mediators	18
2.1.2	Leptons	19
2.1.3	Quarks	19
2.1.4	The Higgs boson	19
2.2	The Standard Model Lagrangian	19
2.2.1	Electromagnetic interaction	20
2.2.2	Strong interaction	22
2.2.3	Weak interaction and electroweak unification	22
2.2.4	The Higgs mechanism and spontaneous symmetry breaking	25
2.2.5	The mass of fermions	27
2.3	SM Higgs Boson Production Modes	28
2.4	SM Higgs Boson Decay Modes	28
2.4.1	SM Higgs boson decays to fermions	29
2.4.2	SM Higgs boson decays to bosons	30

This chapter describes the standard model (SM) of elementary particle physics. In Section 2.1, the properties of these particles and the interactions between them are described. The SM is a gauge theory based on the $SU(3) \times SU(2) \times U(1)$ symmetry group. The mathematical framework of it is presented in Section 2.2. Section 2.3 and 2.4 describe the main production modes and decay modes that are in the LHC current sensitivity scope.

2.1 Elementary Particles and Forces

As we know today [1], there are four interactions in nature: the electromagnetic force, the weak force, the strong force and the gravitational force. A theory that describes the first three forces was developed between 1960s and 1970s [2–10] which is commonly known as the Standard Model (SM) of particle physics. It has passed all the tests performed by various experiments within their accuracies. Currently, it is the most successful attempt to describe elementary particles and their interactions.

The forces are mediated by bosons, which have an integer spin. The gravitational force is not included in the SM. The work to combine gravity with quantum field theory (quantum gravity) is still ongoing. The effect of excluding gravitational force is negligible if the energy scale is lower than the Planck scale (1.22×10^{19} GeV), a scope which is much larger than the current state-of-the-art particle colliders (for the LHC, 1.3×10^4 GeV) and cosmic ray measured in laboratory frame (10^{11} GeV).

2.1.1 Force mediators

All forces are mediated by particles with an integer spin, bosons. The mediators of the strong interaction are eight gluons. They mediate the strong forces between color charged particles (the quarks). Gluons are massless. Because of their effective color charge, they can also interact among themselves. The gluons and their interactions are described by the theory of quantum chromodynamics (QCD) [11–13].

Photons mediate the electromagnetic force between electrically charged particles. The photon is massless and is well-described by the theory of quantum electrodynamics (QED) [11]. The $W^\pm$ and Z gauge bosons mediate the weak interactions between particles of different flavors (all quarks and leptons). They are massive, with masses shown in Table 2.1¹. The weak interactions involving the $W^\pm$ exclusively act on left-handed particles. Furthermore, the $W^\pm$ carries an electric charge of $+1$ and -1 and couples to the electromagnetic interaction. The electrically neutral Z boson interacts with both left-handed particles and antiparticles. These three gauge bosons along with the photons are grouped together, as collectively mediating the electroweak (EW) interaction as detailed in table 2.1.

Mediators	Interaction	Range (m)	Relative strength	Mass (GeV)
8 gluons (g)	Strong	10^{-15}	1	0
photon (γ)	Electromagnetic	∞	10^{-3}	0
$W^\pm, Z$	Weak	10^{-18}	10^{-14}	80.38, 91.18

Table 2.1: The properties of the three force mediators in the SM [14].

¹The gluons are massless, so the range is ∞ . The range is limited only by confinement.

2.1.2 Leptons

Leptons are fermions with half-integer spin and do not interact strongly. Two categories of leptons exist: charged leptons, and neutrinos, classified in three generations. Electron e and electron neutrino ν_e are the first generation. The muon μ and the muon neutrino ν_μ are the second generation. For the third generation, there are τ lepton and the tau neutrino ν_τ . The mass of electron is four orders of magnitude smaller than that of the heaviest charged lepton τ . The mass of neutrino is small and upper limits are set. The details are summarized in table 2.2.

Generation	Particle	Charge	Lifetime(s)	Mass
1st	electron (e)	$-q_e$	∞	0.511 MeV
	electron neutrino(ν_e)	0	∞	< 2 eV
2nd	muon (μ)	$-q_e$	2.20×10^{-6}	105.7 MeV
	muon neutrino(ν_μ)	0	∞	< 2 eV
3rd	tau (τ)	$-q_e$	2.91×10^{-13}	1777 MeV
	tau neutrino(ν_τ)	0	∞	< 2 eV

Table 2.2: The properties of leptons in the SM. q_e is the elementary charge [14].

2.1.3 Quarks

Quarks are fermions and classified in three generations. There are six flavors of quarks. Quarks interact through the strong interaction and carry color charges. They have either $-1/3$ or $2/3$ of the elementary charge. Among the six quarks, the top quark is the heaviest. It was discovered by the proton-proton collider Tevatron in Fermilab in 1995. Quarks have never been observed as individual particles. They are bound together as composite particles called hadrons. Three quarks (qqq) or three antiquarks ($\bar{q}\bar{q}\bar{q}$) form baryons, quark antiquark pairs ($q\bar{q}$) form mesons. The quarks and their properties are shown in table 2.3.

2.1.4 The Higgs boson

Finally, the last piece of the SM, the Higgs boson, was discovered in 2012 by the ATLAS and CMS experiments in LHC. It is the only scalar (spin-0) elementary particle we have observed so far. According to the SM, the presence of the Higgs boson is responsible for the masses of the $W^\pm$ and Z bosons, and of the fermions.

2.2 The Standard Model Lagrangian

This section describes the formalism of the Standard Model of particle physics, a gauge quantum field theory that is based on symmetry groups $SU(2)_L \times U(1)_Y \times SU(3)_C$. According to Noether's theorem [15], there are conserved charges associated to each of these symmetry.

Generation	Particle	Charge	Mass
1st	up	$2/3q_e$	$2.3^{+0.7}_{-0.5}$ MeV
	down	$-1/3q_e$	$4.8^{+0.5}_{-0.3}$ MeV
2nd	charm	$2/3q_e$	1.275 ± 0.025 GeV
	strange	$-1/3q_e$	95 ± 5 MeV
3rd	top	$2/3q_e$	$173.21 \pm 0.51 \pm 0.71$ GeV
	bottom	$-1/3q_e$	4.66 ± 0.03 GeV

Table 2.3: The properties of quarks in the SM [14]. q_e is the elementary charge.

The $SU(2)_L \times U(1)_Y$ describes the electroweak interaction which has four massless vector fields $W^{1,2,3}$ and B . It conserves weak hypercharge Y and weak isospin I . The index L refers to the left chiral nature of the $SU(2)$ coupling. The index Y refers to the weak hypercharge. The $SU(3)_C$ describes the strong interaction which is mediated by eight gluons. It conserves the color charge C .

Particles gain mass through the Higgs mechanism which spontaneously breaks the electroweak symmetry. As a result, there are four physical vector bosons $W^\pm, Z$ and the photon γ which are linear combinations of the $W^{1,2,3}$ and B fields. The $U(1)_{EM}$ symmetry group describes the electromagnetic interaction with photon γ as the mediator and conserves the electric charge (Q).

The theories of the strong and electroweak interactions with the addition of the Higgs sector form together the framework of the SM. A Lorentz invariant Lagrangian density $\mathcal{L}$ describes the dynamics of the system, which can be expressed by three separate terms:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{strong}} + \mathcal{L}_{\text{electroweak}} + \mathcal{L}_{\text{Higgs}} \quad (2.1)$$

The following subsections will describe the Lagrangians of electromagnetic, strong and weak interactions as well as the Higgs mechanism. Natural units and the Einstein summation convention are used throughout the thesis.

2.2.1 Electromagnetic interaction

The electromagnetic interaction is described by QED [11], which is an abelian gauge theory. By making Schrödinger equation relativistic, Dirac introduced his equations [11] to describe the free propagation of a fermion with mass m .

$$(i\gamma^\mu \partial_\mu - m)\psi = 0, \quad (2.2)$$

where ψ is the four-component spinor representing the fermionic field as a function of space-time coordinate $x^\mu = (t, x, y, z)$, where $\mu \in \{0, 1, 2, 3\}$ is the index of the four space-time components. The symbol γ^μ indicates the Dirac matrices.

The Lagrangian density of the Dirac equation is given by

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \quad (2.3)$$

where $\bar{\psi} = \psi^\dagger\gamma^0$ is the adjoint spinor. It is possible to perform the following transformation,

$$\psi \rightarrow e^{iq\alpha(x)}\psi, \quad (2.4)$$

where the phase $\alpha(x)$ is a function of time-space coordinate x^μ and q is the electric charge of the fermion. Transformations of this type build up the $U(1)$ symmetry group.

The free Dirac equation 2.2 is not invariant under the local phase transformation. However, the invariance can be achieved by substituting ∂_μ with the following new covariant derivative $\mathcal{D}_\mu$ in equation 2.2

$$\mathcal{D}_\mu = \partial_\mu + iqA_\mu. \quad (2.5)$$

The new vector field A_μ is the electromagnetic field and is required to transform as

$$A_\mu \rightarrow A_\mu - \partial_\mu\alpha(x). \quad (2.6)$$

As suggested by Noether's theorem [15], this local invariance results in the conservation of electric current

$$J^\mu = q\bar{\psi}\gamma^\mu\psi. \quad (2.7)$$

With such new set of locally invariant transformations, the Lagrangian density of Dirac equation (equation 2.2) can be rewritten as

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\mathcal{D}_\mu - m)\psi = \bar{\psi}i\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu A_\mu\psi, \quad (2.8)$$

To obtain Maxwell's equations, one can use the following Lagrangian density:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - J^\mu A_\mu, \quad (2.9)$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the electromagnetic field tensor. The first term in equation 2.9 describes the kinetic energy of the electromagnetic field. With this term added, one can complete the Lagrangian density of the electromagnetic interaction:

$$\mathcal{L}_{\text{EM}} = \underbrace{\bar{\psi}i\gamma^\mu\partial_\mu\psi}_{\text{fermion kinetic term}} - \underbrace{m\bar{\psi}\psi}_{\text{fermion mass term}} - \underbrace{\frac{1}{4}F^{\mu\nu}F_{\mu\nu}}_{\text{field kinetic term}} - \underbrace{q\bar{\psi}\gamma^\mu A_\mu\psi}_{\text{interaction term}}. \quad (2.10)$$

The interaction term describes the fermion-photon interaction with the electric charge q of the fermion as the coupling strength. Electromagnetism is an abelian gauge field theory and it is commutative, implying that the result of applying multiple transformations of the type in equation 2.4 does not depend on the order in which they are applied. A result of the abelian nature of electromagnetism is that photons have no self-interactions. Thus such kind of term would not appear in the Lagrangian density above. A mass term for the photon of the form $m^2 A^\mu A_\mu$ would break gauge invariance. However, photons are experimentally observed to be massless.

2.2.2 Strong interaction

The strong interaction is described by QCD [11–13]. Mathematically, it is a non-abelian Yang-Mills gauge theory based on the $SU(3)_C$ symmetry group which conserves color charge. There are three quantum numbers called colors: red (r), green (g) and blue (b). The "color" of quarks and gluons is completely unrelated to the everyday meaning of color. It became popular simply because of the analogy to the primary colors.

Similar to how electrons have negative electric charge while positrons have positive electric charge, quarks have color charges and antiquarks have anticolor charges ($\bar{r}$, $\bar{g}$ and $\bar{b}$). Gluons either have one unit of color or one unit of anticolor. Leptons and bosons (except gluons) do not carry any color charge thus they don't participate in the strong interaction.

In order to get the Lagrangian density for the gluon field, similarly to electromagnetic field, one can replace the electromagnetic field strength tensor $F^{\mu\nu}$ with a gluon field strength tensor $G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu - g_3 f_{abc} G_b^\mu G_c^\nu$ where the index a refers to the color charge. As q in the electromagnetic interaction, g_3 is the coupling constant of the strong interaction. The coefficients f_{abc} are the structure constants. The Lagrangian density of the gluon field is given as

$$\mathcal{L}_{\text{gluons}} = -\frac{1}{4} G_a^{\mu\nu} G_{a\mu\nu}. \quad (2.11)$$

The presence of trilinear and quartic terms here allows for gluon self-interactions with vertices connecting three or four gluons. It is a consequence of the non-abelian nature of QCD. The interaction between quarks and gluons can be introduced by the replacement of the derivative ∂_μ in the Dirac equation with the new covariant derivative

$$\mathcal{D}_\mu = \partial_\mu + ig_3 T_a G_{a\mu}, \quad (2.12)$$

where T_a are the eight generators of the $SU(3)_C$ symmetry group which satisfy the commutation $[T_a, T_b] = if_{abc} T_c$. The Lagrangian for quarks is

$$\mathcal{L} = \bar{\psi}_f (i\gamma^\mu \mathcal{D}_\mu - m) \psi_f = \bar{\psi}_f i\gamma^\mu \partial_\mu \psi_f - g_3 \bar{\psi}_f \gamma^\mu T_a G_{a\mu} \psi_f - m \bar{\psi}_f \psi_f \quad (2.13)$$

where the summation is taken for the six quark flavors $f \in [1, 6]$. Thus, the Lagrangian density of the strong interaction is constructed by the kinematic term of the gluon fields, which includes the free propagator of gluons and their self-interaction; and the fermionic terms, which includes the quark kinetic energy, the interaction between quark and gluon and the quark mass term:

$$\begin{aligned} \mathcal{L}_{\text{strong}} &= \mathcal{L}_{\text{gluons}} + \mathcal{L}_{\text{quarks}} \\ &= - \underbrace{\frac{1}{4} G_a^{\mu\nu} G_{a\mu\nu}}_{\text{field kinetic term}} + \underbrace{\bar{\psi}_f i\gamma^\mu \partial_\mu \psi_f}_{\text{quark kinetic term}} - \underbrace{g_3 \bar{\psi}_f \gamma^\mu T_a G_{a\mu} \psi_f}_{\text{interaction term}} - \underbrace{m \bar{\psi}_f \psi_f}_{\text{quark mass term}}. \end{aligned} \quad (2.14)$$

2.2.3 Weak interaction and electroweak unification

The weak interaction is described by the $SU(2)_L$ symmetry group which conserves the weak isospin third component (I_3) [11]. L refers to the fact that the weak interaction only

acts on left-handed fermions. Thus, the left-handed fermions are ordered into doublets of weak isospin

$$\psi_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L \quad (2.15)$$

where d' , s' and b' are the weak eigenstates of the physical quark states of definite mass d , s and b given by the mixing terms in the Cabibbo-Kobayashi-Maskawa (CKM) matrix

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.16)$$

The CKM matrix element describes the probability of a transition from a specific up-type quark to another specific down-type quark when interacting with a $W^\pm$ boson. The values of these matrix elements are not predicted by the SM. However, the requirement that the matrix be unitary allows the matrix to be parametrized by three mixing angles and a phase.

Similarly to equation 2.3 for the electromagnetic interaction, the Lagrangian of the weak interaction for a pair of massless left-handed fermions can be written as

$$\mathcal{L} = \bar{\psi}_L (i\gamma^\mu \partial_\mu u) \psi_L \quad (2.17)$$

The left-handed fermionic field should have this type of transformation:

$$\psi_L \rightarrow e^{i\frac{g_2}{2}\alpha^a(x)\sigma^a} \psi_L \quad (2.18)$$

where σ^a are the Pauli matrices:

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.19)$$

and $\alpha^a(x)$ are three arbitrary space-time dependent functions. Similar to the charge q in the electromagnetic interaction, g_2 is the coupling strength of the weak interaction. To preserve the invariance of the Lagrangian under local $SU(2)_L$ transformations, the covariant derivative is introduced to replace the derivative ∂_μ in equation 2.17:

$$\mathcal{D}_\mu = \partial_\mu + i\frac{g_2}{2}\sigma^a W_\mu^a. \quad (2.20)$$

The three vector fields $W_\mu^a (a = 1, 2, 3)$ are massless and are required to transform as

$$W_\mu^a \rightarrow \partial_\mu \alpha^a(x) - g_2 \epsilon^{abc} \alpha^b(x) W_\mu^c, \quad (2.21)$$

where ϵ^{abc} is the Levi-Civita tensor. Similarly to equation 2.9 for the electromagnetic interaction, the kinetic energy of these new gauge fields is given by $\frac{1}{4}W^{a\mu\nu}W_{\mu\nu}^a$, where $W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g_2 \epsilon^{abc} W_\mu^b W_\nu^c$ is the field strength tensor. The complete Lagrangian

density of $SU(2)_L$ is given by

$$\mathcal{L} = \underbrace{\bar{\psi}_L i \gamma^\mu \partial_\mu \psi_L}_{\text{fermion kinetic}} - \underbrace{\frac{1}{4} W^{a\mu\nu} W_{\mu\nu}^a}_{\text{field kinetic and self-interaction}} - \underbrace{\frac{g_2}{2} \bar{\psi}_L \gamma^\mu \sigma^a W_\mu^a \psi_L}_{\text{interaction}} \quad (2.22)$$

The Lagrangian density for $U(1)_Y$ is almost identical to that of the electromagnetic interaction. Unlike $SU(2)_L$, $U(1)_Y$ couples to both the left-handed and the right-handed isospin singlets, where the latter ones are indicated as

$$\psi_R = e_R, \mu_R, \tau_R, u_R, d_R, c_R, s_R, t_R, b_R. \quad (2.23)$$

The local gauge transformations have the following type

$$\psi \rightarrow e^{i \frac{g_1}{2} \alpha(x)} \psi, \quad (2.24)$$

where g_1 is the coupling strength of the interaction and $\alpha(x)$ is an arbitrary space-time dependent function. The covariant derivative used to preserve local gauge invariance is given by

$$\mathcal{D}_\mu = \partial_\mu + i \frac{g_1}{2} B_\mu, \quad (2.25)$$

where the new vector field B_μ has been introduced. The field strength tensor is then defined as $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$. The complete Lagrangian of $U(1)_Y$ is then given by

$$\mathcal{L}_Y = \underbrace{\bar{\psi} i \gamma^\mu \partial_\mu \psi}_{\text{fermion kinetic term}} - \underbrace{\frac{1}{4} B^{\mu\nu} B_{\mu\nu}}_{\text{field kinetic term}} - \underbrace{\frac{g_1}{2} \bar{\psi} \gamma^\mu B_\mu \psi}_{\text{interaction term}}. \quad (2.26)$$

Electroweak unification requires $SU(2)_L \times U(1)_Y$ group symmetry [2–4]. The electroweak covariant derivative combines equation 2.20 and equation 2.25, given by

$$\mathcal{D}_\mu = \partial_\mu + i \frac{g_2}{2} \sigma^a W_\mu^a + i \frac{g_1}{2} B_\mu. \quad (2.27)$$

Thus, the complete Lagrangian of the electroweak interaction is given by

$$\begin{aligned} \mathcal{L}_{\text{electroweak}} = & \underbrace{\bar{\psi}_L i \gamma^\mu \partial_\mu \psi_L}_{\text{left-handed fermion kinetic term}} + \underbrace{\bar{\psi}_R i \gamma^\mu \partial_\mu \psi_R}_{\text{right-handed fermion kinetic term}} \\ & - \underbrace{\frac{g_2}{2} \bar{\psi}_L \gamma^\mu \sigma^a W_\mu^a \psi_L - \frac{g_1}{2} \bar{\psi}_L \gamma^\mu \sigma^0 B_\mu \psi_L}_{\text{left-handed interaction terms}} - \underbrace{\frac{g_1}{2} \bar{\psi}_R \gamma^\mu B_\mu \psi_R}_{\text{right-handed interaction term}} \\ & - \underbrace{\frac{1}{4} W^{a\mu\nu} W_{\mu\nu}^a - \frac{1}{4} B^{\mu\nu} B_{\mu\nu}}_{\text{field kinetic and self-interaction terms}}. \end{aligned} \quad (2.28)$$

where σ^0 is the 2×2 unit matrix

$$\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (2.29)$$

A basis change of the fields is needed to relate these four vector fields to the physically gauge bosons responsible for mediating the charged and neutral currents of the electroweak interaction. The fields W_μ^1 and W_μ^2 mix to produce the charged $W^\pm$ bosons

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (2.30)$$

The fields W_μ^3 and B_μ mix to produce the neutral Z boson and the photon

$$\begin{pmatrix} Z^\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \quad (2.31)$$

which corresponds to a rotation by the Weinberg angle θ_W where

$$\cos \theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \text{ and } \sin \theta_W = \frac{g_1}{\sqrt{g_1^2 + g_2^2}}. \quad (2.32)$$

The presence of trilinear and quartic terms in the electroweak Lagrangian density is the results of the non-abelian nature of the electroweak interaction. It implies for $W^\pm$ and Z^0 self-interactions with vertices connecting three or four $W^\pm$ and Z^0 bosons.

2.2.4 The Higgs mechanism and spontaneous symmetry breaking

Mass terms for fermions and gauge fields are not present in the Lagrangian introduced in the previous section. The requirement of local gauge invariance requires massless particles and would no longer be valid if mass terms were inserted into the electroweak Lagrangian by hand. This violates the experimental observation of massive fermions and $W^\pm$ and Z^0 bosons. The problem is solved by introducing the spontaneous symmetry breaking where the electroweak symmetry is broken as intrinsic feature of the fields rather than the explicit introduction of symmetry breaking mass terms into the Lagrangian density. Electroweak symmetry breaking in the SM is achieved by the Higgs mechanism [5–10]. It introduces a new $SU(2)_L$ doublet of complex scalar fields

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \quad (2.33)$$

where ϕ_i are four real scalar fields. The corresponding Lagrangian density contains a kinetic term and a potential term, given by

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}^\mu \Phi)^\dagger (\mathcal{D}_\mu \Phi) - V(\Phi^\dagger \Phi), \quad (2.34)$$

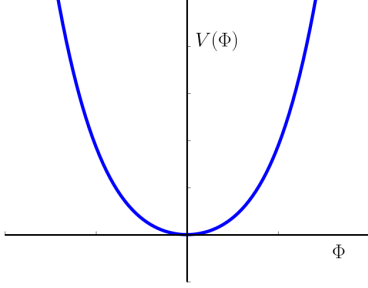
where $\mathcal{D}_\mu$ is the electroweak covariant derivative from equation 2.27 and $V(\Phi^\dagger \Phi)$ is the potential term which is selected to be

$$V(\Phi^\dagger \Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad (2.35)$$

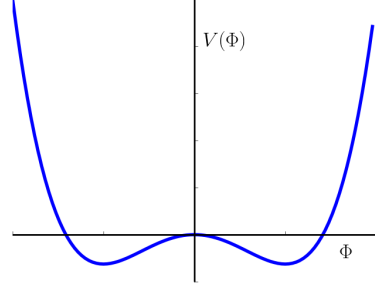
where μ^2 and λ are real parameters. The parameter λ must be positive to make sure that a global minimum exists. When μ^2 has a positive value, a global minimum occurs at $\Phi^\dagger \Phi = 0$.

When μ^2 (Figure 2.1b) has a negative value, the global minimum is non-zero and occurs at

$$(\Phi^\dagger \Phi)_{\min} = (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2)_{\min} = -\frac{\mu^2}{2\lambda}. \quad (2.36)$$



(a) $\mu^2 > 0, \lambda > 0$



(b) $\mu^2 < 0, \lambda > 0$

Figure 2.1: The symmetric quartic Higgs potential, $V(\Phi) = \mu^2\phi^2 + \lambda\phi^4$, of a scalar field ϕ .

It is allowed to choose any point on the four-dimensional hypersphere since the minimum of the potential, the vacuum ground state, has infinite solutions. The chosen point is

$$\phi_{1,\min}^2 = \phi_{2,\min}^2 = \phi_{4,\min}^2 = 0, \quad (2.37)$$

$$\phi_{3,\min}^2 = -\frac{\mu^2}{2\lambda} = v^2, \quad (2.38)$$

where v is the vacuum expectation value of the field ϕ_3 , about 246 GeV. The choice of this minimum leads to $SU(2)$ symmetry spontaneously broken. The vacuum expectation value of the complex scalar doublet is

$$\langle 0 | \Phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (2.39)$$

The physical interpretation is found by expanding the kinetic term of the Lagrangian around this minimum

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \quad (2.40)$$

which gives

$$(\mathcal{D}^\mu \Phi)^\dagger (\mathcal{D}_\mu \Phi) = \frac{1}{2} \partial^\mu H \partial_\mu H + \frac{1}{4} g_2^2 W_\mu^- W^{+\mu} (v + H)^2 + \frac{1}{4} (g_1^2 + g_2^2) Z_\mu Z^\mu (v + H)^2, \quad (2.41)$$

where H is a massive scalar field, the Higgs field. The first term is the kinetic energy term of the Higgs field. The other terms which contain H are the interaction terms with the $W^\pm$ and Z bosons. The mass parameters of the gauge bosons can be rebuilt from the bilinear

terms as:

$$m_W = \frac{1}{2}vg_2, \quad m_Z = \frac{1}{2}v\sqrt{g_1^2 + g_2^2}, \quad m_A = 0. \quad (2.42)$$

The Higgs kinetic term, $\frac{1}{2}\partial^\mu H\partial_\mu H$, indicates the existence of a new scalar particle, the Higgs boson. The mass of the Higgs boson can be found by expansion of the potential around the vacuum expectation value

$$V = \frac{1}{2}\mu^2(v + H)^2 + \frac{1}{4}\lambda(v + H)^4. \quad (2.43)$$

The bilinear term in this expansion is $\frac{1}{2}\lambda v^2 H^2$, which gives the mass of the Higgs boson to be $m_H = \sqrt{\lambda}v$. Although the vacuum expectation value of the Higgs field is known, the value of the parameter λ is not predicted by the theory. For this reason, the mass of the Higgs boson is also a free parameter. In summary, the introduction of a doublet complex scalar field, along with a symmetric potential, breaks the $SU(2)_L \times U(1)_Y$ symmetry which leads to the existence of a massive scalar - the Higgs boson and gives masses to the weak gauge bosons $W^\pm$ and Z while keeping the photon massless.

2.2.5 The mass of fermions

As mentioned in the previous section, adding mass terms for the fermions would break local gauge invariance, as in the case for the gauge bosons. However, additional terms can be added to the SM Lagrangian which couple the Higgs field to the fermions, these terms are called Yukawa couplings. After expansion around the ground state, the Lagrangian density has the form

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}}^f &= -\frac{\lambda_f}{\sqrt{2}}(v + H)\bar{\psi}_L\psi_R - \frac{\lambda_f}{\sqrt{2}}(v + H)\bar{\psi}_R\psi_L \\ &= -\underbrace{\frac{\lambda_f v}{\sqrt{2}}(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)}_{\text{fermion mass term}} - \underbrace{\frac{\lambda_f}{\sqrt{2}}H(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L)}_{\text{interaction term}}. \end{aligned} \quad (2.44)$$

The first term gives mass to each of the fermions, given by

$$m_f = \frac{\lambda_f v}{\sqrt{2}} \quad (2.45)$$

which means that mass is generated through fermions coupling to the Higgs field. The second term corresponds to interactions between a Higgs boson and fermions with a coupling coefficient given by

$$g_{Hff} = i\frac{m_f}{v}. \quad (2.46)$$

This indicates that the strength of the coupling between fermions and the Higgs boson is proportional to the mass of the fermion. The introduction of Yukawa couplings is required to account for experimental observation of massive fermions but not for the theory. Because of this, it is of particular interest to study the Higgs boson coupling to fermions.

2.3 SM Higgs Boson Production Modes

There are four main dominant production modes of the Higgs boson at the LHC in proton-proton collision at $\sqrt{s} = 13$ TeV:

- The gluon-gluon fusion (ggH) production has the largest cross section at the LHC. It goes via a quark loop or a W loop contribution.
- The vector boson fusion (VBF) production has a cross section an order of magnitude smaller than the ggH production. Two high-momentum quarks are present in the final state which will form into jets that can be observed in the detector. The kinematic characteristics of these jets, for example their forward direction or their large invariant mass, cause the VBF production to have a clear signature in the experiment.
- The associated production with vector boson (VH) has a production of a virtual boson V^* that splits into a real boson V and a Higgs boson H . The cross section is smaller than VBF production. However, the presence of leptons or quarks coming from vector boson decays can provide an extra handle for discriminating a Higgs boson VH signal from backgrounds.
- The production in association with a pair of top quarks $t\bar{t}H$ has even a smaller cross section than VH productions. By requiring a top quark pair, the SM background can be significantly reduced. Although the cross section is small, this production mode has been observed in Run-II [16].

The Feynman diagrams of the three dominant production modes of the SM scalar boson at the LHC are shown in Figure 2.2. Figure 2.3 shows the cross-section values of each of production modes as a function of the mass of the Higgs boson.

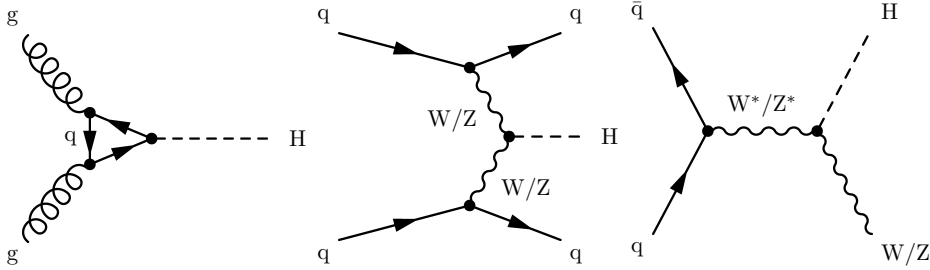


Figure 2.2: Dominant Feynman diagrams for the SM Higgs boson production at the LHC: gluon-gluon fusion (left), vector boson fusion production (center) and associated production with a vector boson (right) [17].

2.4 SM Higgs Boson Decay Modes

For the SM Higgs boson with a particular mass, it is possible to compute its partial decay width to any combination of other SM particles. The branching ratio is obtained by dividing the partial decay width by the sum of the partial decay widths of all possible decays:

$$\mathcal{B}(H \rightarrow XX) = \frac{\Gamma(H \rightarrow XX)}{\sum_{Y \in SM} \Gamma(H \rightarrow YY)}. \quad (2.47)$$

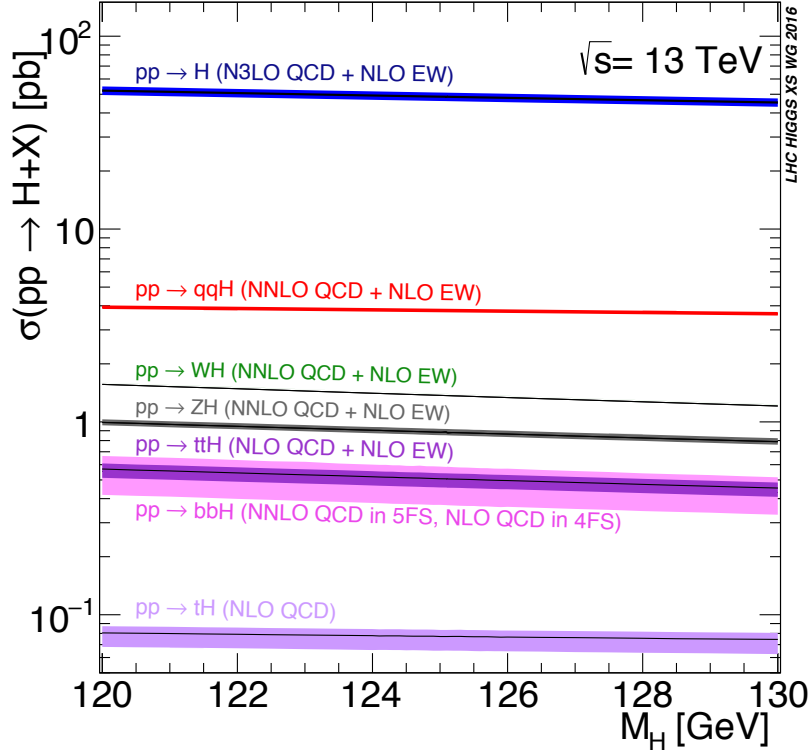


Figure 2.3: Cross-sections of the five dominant Higgs boson production modes in proton-proton collisions at the LHC at $\sqrt{s} = 13$ TeV center-of-mass energy [18].

The partial decay widths can be computed using the prescriptions in [19, 20], and are shown in Figure 2.4 for scalar boson masses from 80 to 200 GeV.

2.4.1 SM Higgs boson decays to fermions

In the SM, Higgs boson couplings to fermions are directly proportional to fermion masses (equation 2.46). The Born approximation gives the partial decay width of the scalar boson to fermion pairs. It can be written as:

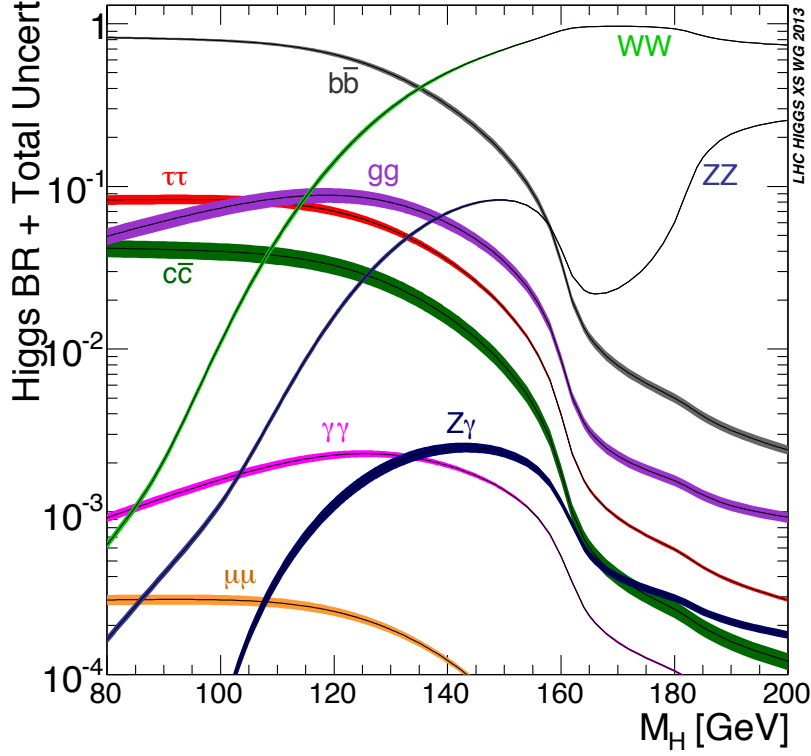
$$\Gamma_{\text{Born}}(H \rightarrow f\bar{f}) = \frac{G_F N_C}{4\sqrt{2}\pi} m_H m_f^2 \beta_f^3, \quad (2.48)$$

where m_f is the fermion mass, N_C is a factor which equals to one for lepton and to three for quarks, β is the velocity of the fermions in the final state and can be expressed as:

$$\beta = \sqrt{1 - \frac{4m_f^2}{m_H^2}}, \quad (2.49)$$

and $G_F \approx 1.17 \times 10^{-5} \text{ GeV}^{-2}$ is the Fermi coupling constant [14]. As the partial decay width is proportional to the square of the fermion mass, the branching ratio of the Higgs boson to tau leptons is approximately two orders of magnitude larger than its branching ratio to muons, while the branching ratio to electrons is negligible.

In the case of Higgs boson decays to quarks, QCD corrections cannot be neglected. For


 Figure 2.4: Branching ratios of the SM scalar boson for m_H between 80 and 200 GeV. [18]

Higgs boson masses much larger than the quark mass, the NLO decay width, including Feynman diagrams with gluon exchange and the emission of a gluon in the final state, can be expressed as:

$$\Gamma_{\text{NLO}}(H \rightarrow q\bar{q}) \simeq \frac{3G_F}{4\sqrt{2}\pi} m_H m_q^2 \left[1 + \frac{4g_s}{3\pi} \left(\frac{9}{4} + \frac{3}{2} \ln \frac{m_q^2}{m_H^2} \right) \right] \quad (2.50)$$

2.4.2 SM Higgs boson decays to bosons

For masses larger than the WW and ZZ kinematical thresholds, the Higgs boson decays essentially to electroweak boson pairs. The partial decay widths to a pair of V (W or Z boson) is given by

$$\Gamma(H \rightarrow VV) = \frac{G_F m_H^3}{16\sqrt{2}\pi} \delta_V \sqrt{1-4x} (1-4x+12x^3) \quad (2.51)$$

with $x = \frac{m_V^2}{m_H^2}$, $\delta_W = 2$ and $\delta_Z = 1$.

Below the WW and ZZ kinematical thresholds, the two-body decay as described above is not allowed. The Higgs boson can still decay to a pair of electroweak gauge bosons, with one or two of them being off-shell (three-body and four-body decays respectively). For $m_H = 125$ GeV, the three-body decay dominates, and its partial decay width can be expressed as,

assuming massless fermions f ,

$$\Gamma(H \rightarrow VV) = \frac{3G_F^2 m_V^4}{16\pi^3} m_H \delta'_V R_T(x) \quad (2.52)$$

with $\delta'_W = 1$, $\delta'_Z = \frac{7}{12} - \frac{10}{9} \sin^2 \theta_W + \frac{40}{9} \sin^4 \theta_W$, and

$$R_T(x) = \frac{3(1-8x+20x^2)}{(4x-1)^{1/2}} \arccos\left(\frac{3x-1}{2x^{3/2}}\right) - \frac{1-x}{2x} (2-13x+47x^2) - \frac{3}{2} (1-6x+4x^2) \ln x \quad (2.53)$$

Even if massless particles do not couple the Higgs boson, Higgs boson decaying to gg , $Z\gamma$ and $\gamma\gamma$ are allowed through massive particle loops. The $H\gamma\gamma$ and $HZ\gamma$ couplings are mediated by W boson and charged fermion loops, and the Hgg couplings by quark loops.

CHAPTER

3

THE MINIMAL SUPERSYMMETRIC STANDARD MODEL (MSSM)

Contents

3.1	Motivations for Supersymmetry	33
3.1.1	The hierarchy problem	34
3.1.2	The dark matter	35
3.1.3	Unification of the forces	35
3.2	The Minimal Supersymmetric Standard Model	35
3.3	The Higgs Sector of MSSM	37
3.3.1	The MSSM Higgs parameters	37
3.3.2	The MSSM Higgs benchmark scenarios	39

The discovery of a Higgs boson is a success of the SM. As the SM has a number of shortcomings, it is not considered as the ultimate theory of elementary particles. The concept of Supersymmetry is proposed to address these shortcoming (Section 3.1). The simplest supersymmetry model is the minimal supersymmetry extension of SM (MSSM) which is introduced in Section 3.2. It introduces the minimal amounts of new parameters while still keeping compatibility with the SM. In Section 3.3, we discuss the Higgs sector of the MSSM. The two favored phenomenological benchmark scenarios are also described.

3.1 Motivations for Supersymmetry

Although the SM has been demonstrated as a successful theory by many measurements, it has a number of shortcomings. The SM does not explain the existence of dark matter and dark energy, unification of the forces and the large hierarchy between the Planck scale and

the EW scale. Several models beyond the standard model (BSM) have been proposed to solve the above issues. One of them is called Supersymmetry (SUSY) [21–23] which is a set of theories with a general assumption: every Standard Model particle has one "superpartner". It is a symmetry that relates bosons and fermions. The SUSY operator Q , an anticommuting spinor, transforms a fermionic field F into a bosonic field B and vice versa:

$$Q|B\rangle = F \text{ and } Q|F\rangle = B \quad (3.1)$$

In the SM, baryon number and lepton number are conserved. In SUSY models, processes where baryon number and lepton number are not conserved can happen. A new quantum number R has been introduced to maintain the baryon number and lepton number conservation:

$$R = (-1)^{2S+3(B-L)} \quad (3.2)$$

where S , B and L are the spin, baryon and lepton numbers, respectively. The SM particles have $R = 1$ while superpartners have $R = -1$.

The reasons why supersymmetry can solve some of the issues of the SM are discussed in the following subsections.

3.1.1 The hierarchy problem

One of the most important questions that the SM fails to address is the huge gap between the electroweak scale (vacuum expectation value ≈ 246 GeV) and the Planck scale ($M_P = 10^{19}$ GeV/ c^2). The presence of a "desert" across 17 orders of magnitude in energy scale introduces problems related to the naturalness of the parameters of the SM. In particular, the radiative corrections to the Higgs boson squared mass m_H^2 are quadratically proportional to the ultraviolet cutoff energy λ_{UV} . It should be at least equal to the energy scale to which the SM is valid without any addition of new physics. If one assumes the SM to be valid up to the Planck energy scale M_P , the correction term to m_H^2 is enormously larger than m_H^2 which indicates that some extraordinary cancellations of terms should happen.

Figure 3.1 (a) shows the correction to m_H^2 coming from a Dirac fermion f that couples directly to the scalar field ϕ with a coupling λ_f .

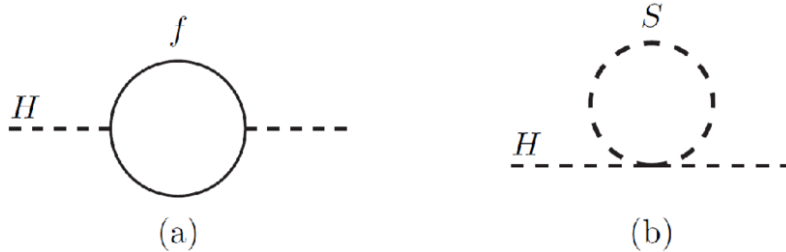


Figure 3.1: One-loop Feynman diagram of correction of Higgs squared mass m_H^2 , due to (a) Dirac fermion f , (b) a scalar S .

This correction term is much larger than the squared mass itself:

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{UV}^2 \quad (3.3)$$

In SUSY models, when another scalar couples to the SM Higgs boson as in Figure 3.1 (b), by assuming that this additional scalar particle S with a mass m_S couples to the Higgs field with the Lagrangian density term $-\lambda_S|\Phi|^2|S|^2$, a correction term is introduced:

$$\Delta m_H^2 = -\frac{\lambda_f}{16\pi^2}[\Lambda_{UV}^2 - 2m_s^2 \ln(\Lambda_{UV}/m_S) + \dots] \quad (3.4)$$

If supersymmetry provides new scalar particles at the TeV scale, they could couple to the SM scalar boson to avoid fine-tuning and cancel the Λ_{UV}^2 divergence. The cancellation of quadratic divergencies is achieved by introducing scalar superpartners of fermions, spin 1/2 superpartners of gauge bosons and spin 1/2 superpartners of Higgs fields.

3.1.2 The dark matter

Measurement of velocities of galaxies [24] shows that there exists non-baryonic matter that does not radiate light in our universe. Further experiments on measurements of the temperature and polarization anisotropies of cosmic microwave background [25, 26] provide results on the percentage of energy and matter of universe. It shows that the baryonic matter, which can be described by the Standard Model, only contributes to 5% of the matter in universe. Dark matter, which participates only in gravitational and weak interactions, constitutes up to 27%. The remaining 68% is assumed to be dark energy, which is responsible for the accelerated expansion of the universe.

The Standard Model does not describe the dark matter, even with the massive neutrinos. It requires beyond Standard Model physics. R-parity conservation in SUSY models implies that supersymmetric particles can only be produced in pairs and that the lightest supersymmetric particle (LSP) is stable. This LSP could be the candidate for the composition of dark matter.

3.1.3 Unification of the forces

In the previous chapter, the unification of electromagnetic and weak forces is shown within the context of the SM. However, the running of the coupling constants of the electromagnetic, weak and strong forces increases as the energy scale increases. It indicates a possibility that these forces will unify at a high energy scale. The presence of supersymmetry can modify the running of the coupling constants to reach a grand unification by introducing new particles, as illustrated in Fig. 3.2.

3.2 The Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) [27–31] adds minimal amount of new parameters, while keeping all the current symmetries and observables of the SM. It is the simplest way in which SUSY can be implemented while being consistent with experimental observation. The MSSM extends particle content of the SM by introducing for each particle a superpartner differing in spin by 1/2.

The representation of fields in supersymmetric theories are the supermultiplets. In the scheme of the MSSM, there are two types of supermultiplets: vector supermultiplets which describe spin-1 vector bosons in SM and spin-1/2 Weyl fermions, gauginos; and chiral supermultiplets which describe spin-1/2 Weyl fermions in SM and spin-0 scalars. Scalar fields are

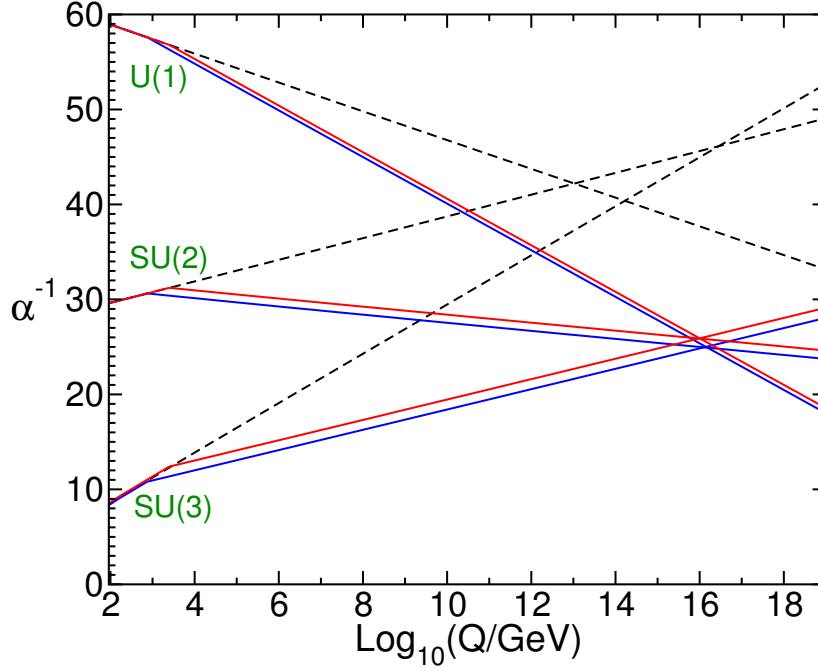


Figure 3.2: Evolution of the electromagnetic ($U(1)_Y$), weak ($SU(2)_L$) and strong ($SU(3)_C$) couplings in the SM (dashed lines), and in the MSSM cases (solid lines). In the MSSM, the three forces can be unified at a high energy scale. [23]

assigned superpartners called higgsinos. The detailed chiral and gauge supermultiplets and higgsinos of the MSSM are shown in Tables 3.1 and 3.2.

Particles	Supermultiplet	spin-0	spin-1/2	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
quarks and squarks	Q	$(\tilde{u}_{iL}, \tilde{d}_{iL})$	(u_{iL}, d_{iL})	3	2	+1/6
	$\bar{u}$	$\tilde{u}_{iR}^*$	$u_{iR}^\dagger$	3	1	-2/3
	$\bar{d}$	$\tilde{d}_{iR}^*$	$d_{iR}^\dagger$	3	1	+1/3
lepton and slepton	L	$(\tilde{e}_{iL}, \tilde{\nu}_i L)$	$(e_{iL}, \nu_i L)$	1	2	-1/2
	$\bar{e}$	$\tilde{e}_{iL}^*$	$e_{iR}^\dagger$	1	1	+1
Higgs, higgsino	H_u	(H_u^+, H_u^0)	$(\tilde{H}_u^+, \tilde{H}_u^-)$	1	2	+1/2
	H_d	(H_d^+, H_d^0)	$(\tilde{H}_d^+, \tilde{H}_d^-)$	1	2	-1/2

Table 3.1: The MSSM chiral supermultiplets and their quantum numbers.

In general, superpartners have a symbol " $\sim$ " above their symbols. The superpartners of quarks are called squarks, while the superpartners of leptons are called sleptons. With similar convention, the superpartners of gauge bosons are called gauginos, while the superpartners of the gauge boson fields W are the winos and the superpartner of the gauge field B is the bino. The fermionic superpartners of the Higgs fields are denoted as higgsinos. Four

Particles	spin-1/2	spin-1	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
gluino, gluon	$\tilde{g}$	g	8	1	0
wino, W boson	$\tilde{W}^\pm, \tilde{W}^0$	$W^\pm, W^0$	1	3	0
bino, B boson	$\tilde{B}^0$	B^0	1	1	0

Table 3.2: The MSSM vector supermultiplets and their quantum numbers.

neutralinos ($\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0$, and $\tilde{\chi}_4^0$) are formed by the mixing of the bino, the neutral wino and the higgsinos, while four charginos ($\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^\pm$) are formed by the mixing of the winos and charged higgsinos.

3.3 The Higgs Sector of MSSM

The Higgs sector of the MSSM is a type-2 two Higgs doublets model (2HDM) [32]. The two Higgs doublets are indicated by ϕ_1 and ϕ_2 , where ϕ_1 couples to the up-type quarks, and ϕ_2 couples to down-type quarks and leptons. The Higgs potential of MSSM, assuming the CP is conserved, can be described as

$$\begin{aligned}
 V = & m_{11}^2 \phi_1^\dagger \phi_1 + m_{22}^2 \phi_2^\dagger \phi_2 - m_{12}^2 (\phi_1^\dagger \phi_2 + \phi_2^\dagger \phi_1) + \frac{\lambda_1}{2} (\phi_1^\dagger \phi_1)^2 + \frac{\lambda_2}{2} (\phi_2^\dagger \phi_2)^2 \\
 & + \lambda_3 \phi_1^\dagger \phi_1 \phi_2^\dagger \phi_2 + \lambda_4 \phi_1^\dagger \phi_2 \phi_2^\dagger \phi_1 + \frac{\lambda_5}{2} [(\phi_1^\dagger \phi_2)^2 + (\phi_2^\dagger \phi_1)^2],
 \end{aligned}
 \tag{3.5}$$

where all the parameters are real.

The minima of the ϕ_1 and ϕ_2 doublets are

$$\langle \phi_1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 \\ 0 \end{pmatrix} \quad \text{and} \quad \langle \phi_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix},
 \tag{3.6}$$

where v_1 and v_2 are the vacuum expectation values of the ϕ_1 and ϕ_2 doublets.

There are eight degrees of freedom of the two Higgs doublets. Three are the physical states of $W^\pm$, Z bosons and the photon γ which are the results of the Goldstone bosons mix with the gauge field W and B . The remaining five degrees of freedom give the five physical Higgs boson states. Two are electrically neutral and CP-even scalar: the lighter h , which is often assumed to be the SM-like Higgs boson observed at 125 GeV, and the heavier H . Another one is the electrically neutral and CP-odd pseudoscalar A . The last two are the charged Higgs H^+ and H^- .

3.3.1 The MSSM Higgs parameters

At tree level, the behavior of the Higgs bosons can be characterized by only two free parameters, which are conventionally chosen to be the mass of the CP-even Higgs m_A and $\tan \beta$. The mass of the neutral Higgs and of the charged Higgs, and the angle α can be

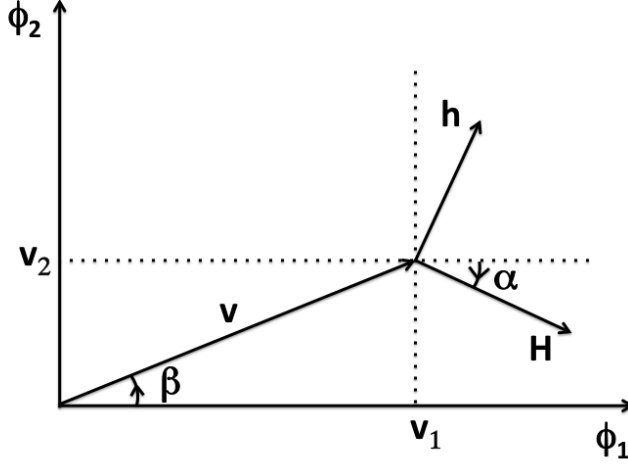


Figure 3.3: Schematic plot [33] of the two angle parameters in the MSSM Higgs sector. The parameter $\tan \beta$ is the ratio of the vacuum expectation values of the doublets ϕ_1 and ϕ_2 . The parameter α is the mixing angle between the neutral Higgs bosons.

expressed as [34]:

$$m_{h/H}^2 = \frac{1}{2} \left(m_A^2 + m_Z^2 \mp \sqrt{(m_A^2 + m_Z^2)^2 - 4m_A^2 m_Z^2 \cos^2 2\beta} \right) \quad (3.7)$$

$$m_{H^\pm}^2 = m_A^2 + m_W^2, \text{ and} \quad (3.8)$$

$$\tan 2\alpha = \tan 2\beta \left(\frac{m_A^2 + m_Z^2}{m_A^2 - m_Z^2} \right) \quad \text{with } -\frac{\pi}{2} \leq \alpha \leq 0. \quad (3.9)$$

These relations bring several constraints on the value of the masses, namely:

$$m_{H^\pm}^2 > m_W^2 \quad (3.10)$$

$$m_H \geq \max(m_A, m_Z) \geq \min(m_A, m_Z) \cos 2\beta \quad (3.11)$$

$$m_H^2 + m_h^2 > m_A^2 + m_Z^2 \quad (3.12)$$

The mass of the lightest Higgs boson could not in principle be larger than the Z boson mass. However, this is violated by LEP bounds and the observation of the 125 GeV Higgs boson. If high order corrections are included, several additional parameters such as the stop mixing parameter or the average scale of SUSY can also appear in the computations. This enables the lightest Higgs h to be as heavy as approximately 135 GeV. In the case where m_A is much larger than the Z boson mass, known as decoupling limit, the relations above will give:

$$m_H \simeq m_{h^\pm} \simeq m_A \quad \text{and} \quad \alpha \simeq \beta - \frac{\pi}{2}. \quad (3.13)$$

The lightest Higgs boson h takes the role of the observed 125 GeV SM-like Higgs boson and the H and A bosons are nearly degenerate in mass. At leading order, the coupling of the H and A bosons to down-type fermions is enhanced with respect to the SM expectation as $\tan \beta$ increases, while the coupling to vector bosons and up-type fermions is suppressed by the same amount. As the τ lepton is a down-type fermion, the $\tau\tau$ channel is of particular interest in searches for additional heavy neutral Higgs bosons at high values of $\tan \beta$. Table 3.3 shows the MSSM modified couplings to gauge bosons and charged fermions.

	h	H	A
W^+W^-	$\sin(\beta - \alpha)$	$\cos(\beta - \alpha)$	0
ZZ	$\sin(\beta - \alpha)$	$\cos(\beta - \alpha)$	0
$u\bar{u}$ (up-type quarks)	$\cos \alpha / \sin \beta$	$\sin \alpha / \sin \beta$	$\cot \beta$
$d\bar{d}$ (down-type quarks)	$\sin \alpha / \cos \beta$	$\cos \alpha / \cos \beta$	$\tan \beta$
$\ell\bar{\ell}$ (charged leptons)	$\sin \alpha / \cos \beta$	$\cos \alpha / \cos \beta$	$\tan \beta$

Table 3.3: The MSSM modified couplings to gauge bosons and charged fermions.

3.3.2 The MSSM Higgs benchmark scenarios

The searches in the $m_A - \tan \beta$ parameter plane can be determined for two representative benchmark scenarios of the MSSM, the $m_h^{\text{mod}+}$ [35] and the hMSSM scenario [36–38]. The m_h^{mod} scenario is a modification of the m_h^{max} scenario [39], which was originally defined to give conservative exclusion bounds on $\tan \beta$ from the LEP searches. The modification reduces the contribution of the stop mixing parameter to make the value of m_h compatible with the mass of the observed Higgs boson. It can be done for different signs of stop mixing parameter. The $m_h^{\text{mod}+}$ scenario is established by taking the positive sign. The hMSSM scenario describes the MSSM sector assuming the validity of the tree level parameters and setting m_h as exactly the observed value of 125 GeV. In addition to this, it also supposes that the corrections to the masses are only affected by the leading logarithmic terms arising from top and stop loops. Finally, it assumes that the masses of the rest of SUSY particles are high enough to escape detection at the LHC, neglecting their effect on the Higgs sector.

In the MSSM, the neutral Higgs boson $\Phi = H/A/h$ can be produced via two production modes: gluon-gluon fusion ($gg\Phi$) and production in association with b quarks ($bb\Phi$). The enhanced coupling to down-type fermions has also influence on the production: firstly, b-associated production dominates over the production via gluon-gluon fusion for large values of $\tan \beta$. Secondly, in gluon-gluon fusion production the kinematic distributions of the Higgs boson change as a function of $\tan \beta$ due to the increasing b quark contribution to the fermion loop. The Feynman diagrams for the h/A production processes are shown in Fig 3.4.

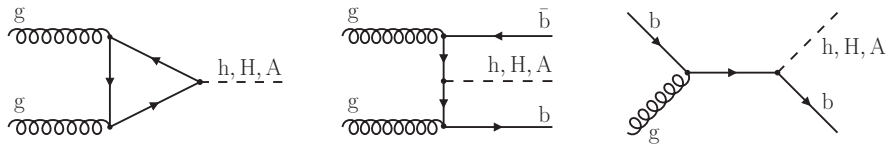


Figure 3.4: Leading order diagrams for the production of Higgs bosons (left) via gluon-gluon fusion and (center and right) in association with b quarks. In the middle figure a pair of b quarks is produced from two gluons (leading order in the four-flavor scheme). In the right the Higgs boson is radiated off from a b quark in the proton (leading order of five-flavor scheme).

CHAPTER

4

STATISTICS

Contents

4.1 Likelihood	42
4.1.1 Basics	42
4.1.2 Introducing nuisance parameters	42
4.2 Maximum Likelihood Fit	45
4.3 Significance	45
4.4 Exclusion Limits	47
4.4.1 Observed Limits	47
4.4.2 Expected Limits	48
4.5 Boosted Decision Tree	49

The physics analyses in high energy experiment heavily rely on statistical tools and methods. This chapter will document the basic concepts of statistics which will be used in the next chapters. In Section 4.1, the concept of likelihood is introduced and the way of introducing systematic uncertainties as nuisance parameters are described. In Section 4.2, the maximum likelihood fit which is used to get the value of parameter of interest, such as signal strength of new physics, is described. The computation of the significance of the observed signal is discussed in Section 4.3. If the observation is compatible with the background expectation, limits on model parameters can be derived. For that, the widely used CL_s method is introduced in Section 4.4. The last section concerns the most commonly used multivariate analysis method in CMS, the boosted decision tree (BDT) method.

4.1 Likelihood

4.1.1 Basics

In a counting experiment, data events follow the Poisson distribution, which is a discrete probability distribution that describes the probability to observe n events in a given time period if these events occur with a known constant rate λ . The Poisson distribution of parameter λ is:

$$f(n) = \frac{e^{-\lambda} \lambda^n}{n!}. \quad (4.1)$$

For a simple counting experiment, where n events are observed while b events are expected, the likelihood function, which describes the level of agreement between the expectation and the observation, is given by the Poisson distribution:

$$\mathcal{L}(n|b) = \frac{e^{-b} b^n}{n!}. \quad (4.2)$$

If data are binned in a histogram and the N bins can be considered as independent of each other, the likelihood is defined as the product of the Poisson distributions for each bin, given by

$$\mathcal{L}(\mathbf{n}|\mathbf{b}) = \prod_{i=1}^N \frac{e^{-b_i} b_i^{n_i}}{n_i!}, \quad (4.3)$$

where $\mathbf{n}$ is the vector of the observed data events in different bins, and $\mathbf{b}$ is the expected number of events in different bins. If data are not binned in a histogram but are described by a probability density function (PDF) $f_b(x)$ of some observables x , k events are observed, and b events are expected in the full x range, the likelihood is [40] :

$$\mathcal{L}(\mathbf{x}|b, f_b(x)) = k^{-1} \prod_{i=1}^k b f_b(x_i) e^{-b} \quad (4.4)$$

4.1.2 Introducing nuisance parameters

Systematic uncertainties in the measurement are handled by introducing nuisance parameters. Nuisance parameters are not the direct parameter of interest but they influence the likelihood model. In the analyses, there are three different sources:

- Theoretical uncertainties, for example, cross section and parton distribution function uncertainties.
- Statistical uncertainties, such as limited number of events in the Monte Carlo (MC) simulations or limited number of observed events in a control region used to model some background processes.
- Experimental uncertainties, such as trigger efficiency or physics object identification efficiency, eg. electron identification efficiency.

The nuisance parameters represented by the vector $\boldsymbol{\theta}$ affect the number of expected events, $b(\boldsymbol{\theta})$. One can introduce the probability density function (PDF) $p(\tilde{\theta}|\theta)$, where $\tilde{\theta}$ is the assumed

value of the nuisance parameter that reflects the degree of belief on what the real value of the parameter θ is. Supposing that there are K nuisance parameters, in the case of the binned histogram, the likelihood becomes:

$$\mathcal{L}(\mathbf{n}|\mathbf{b}) = \prod_{i=1}^N \frac{e^{-b_i} b_i^{n_i}}{n_i!} \prod_{j=1}^K p(\tilde{\theta}_j|\theta_j). \quad (4.5)$$

According to Bayes' theorem, the Bayesian probability $p(\tilde{\theta}|\theta)$ can be expressed as a function of the frequentist probability $\rho(\theta|\tilde{\theta})$.

Systematic uncertainties can be described by log-normal PDFs of parameter κ which can overcome the inconvenient side-effect of the Gaussian PDF, that causes problems for positively defined observables. The Gaussian PDF reads:

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta - \tilde{\theta})^2}{2\sigma^2}\right), \quad (4.6)$$

The log-normal PDF has a longer tails than the Gaussian distributions and goes to zero at $\theta = 0$. It avoids negative values of the observables:

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi \ln(\kappa)}} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln \kappa)^2}\right) \frac{1}{\theta} \quad (4.7)$$

For small uncertainties ϵ , the Gaussian distribution is asymptotically identical to log-normal distribution with $\kappa = 1 + \epsilon$. But it is invalid for large uncertainties in case of positively defined observables. Fig. 4.1 (left) illustrates log-normal distributions for different parameters. In the analysis, systematic uncertainties affecting the normalization of predicted events are described by the log-normal PDF.

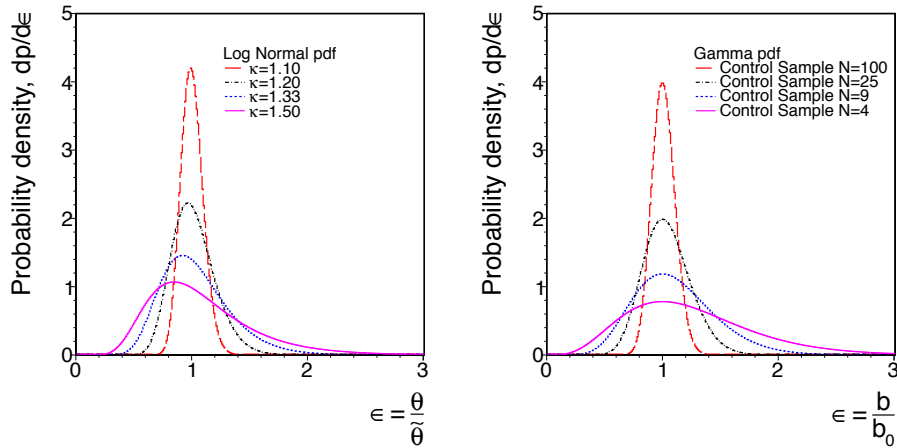


Figure 4.1: Log-normal (**left**) distributions with various κ parameters. Gamma distributions (**right**) with the number of events for the control sample.

Gamma distribution can be used for the case of uncertainties coming from statistically limited numbers of events, for two cases: Monte Carlo events in simulation or a number of observed events in data control sample. The event rate n in the signal region can be related to the number of events N in MC or in data control region via a simple relation $n = \alpha \dot{N}$. The

uncertainties on the predicted rate n associated with the observation of N events is described by the Gamma distribution given by:

$$\rho(n) = \frac{1}{\alpha} \frac{(n/\alpha)^N}{N!} \exp(-n/\alpha). \quad (4.8)$$

The Gamma distributions are illustrated in Fig. 4.1 (right) for different parameter values of N .

Systematics uncertainties that influence shape of variables used to extract the signal are modeled with a linear extrapolation method [41]. In an analysis, the shape uncertainty is implemented by having one up (+1) standard deviation of nuisance parameter shape and one down (-1) standard deviation shape. In the likelihood function, a parameter θ is added to interpolate smoothly between the alternative shapes with a "vertical template morphing" technique. If there are N shape nuisance parameters, modeled with the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N)$, and if h_0, h_j^+ and h_j^- are, respectively, the nominal distribution, the +1 standard deviation of the j th nuisance parameter histogram and the -1 standard deviation of the j th nuisance parameter histogram, then the histogram distribution as a function of the shape nuisance parameter $\boldsymbol{\theta}$ is given by:

$$h(\boldsymbol{\theta}) = h_0 + \sum_{j=1}^N (a(\theta_j)h_j^+ + b(\theta_j)h_0 + c(\theta_j)h_j^-), \quad (4.9)$$

with

$$a(\theta) = \begin{cases} \theta(\theta + 1)/2 & \text{if } |\theta| \leq 1 \\ 0 & \text{if } \theta < -1 \\ \theta & \text{if } \theta > 1 \end{cases}, \quad (4.10)$$

$$b(\theta) = \begin{cases} -\theta^2 & \text{if } |\theta| \leq 1 \\ -(|\theta| - 1) & \text{if } |\theta| > 1 \end{cases}, \quad (4.11)$$

and

$$c(\theta) = \begin{cases} \theta(\theta - 1)/2 & \text{if } |\theta| \leq 1 \\ 0 & \text{if } \theta > +1 \\ |\theta| & \text{if } \theta < -1 \end{cases}. \quad (4.12)$$

A special type of shape uncertainty, the bin-by-bin uncertainty, is associated to the uncertainty on the limited number of MC events or on the limited number of events from a data control region used to estimate the background in each bin of distribution. These uncertainties affect the shape of the distribution but are uncorrelated across the bins. The approach to deal with such cases was proposed by Barlow and Beeston [42]. A separate nuisance parameter is introduced for every bin of each process, scaling the number of expected events in this particular bin for this given process. A large number of nuisance parameters is added to the likelihood function, but it is possible to limit their number by setting a threshold in the ratio of uncertainty of the bin divided by the content of the bin.

4.2 Maximum Likelihood Fit

One can write a model which contains the known physics and new phenomena predicted by the new theory. The first case usually is referred to as background, and the latter as signal. The measured variables has expected number of events $\mathbb{E}_i$ in each bin $i = 1, 2, \dots, N$ of a histogram. The expectation value for the number of events in a particular bin is:

$$\mathbb{E}_i = \mu s_i + b_i, \quad (4.13)$$

where s_i is the expected number of signal events in the bin, b_i is the expected number of background events in the bin, and μ is the signal strength modifier which is referred to as parameter of interest. As mentioned in the previous section, the nuisance parameter $\boldsymbol{\theta}$ has an impact on the expected events as $\mathbb{E}_i(\boldsymbol{\theta}) = \mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})$.

One can perform a maximum likelihood fit to find the optimal values of the parameter of interest and nuisance parameters that give the best agreement between expectation and observation. Fit can be performed under two hypotheses:

- Background-only fit: by taking $\mu = 0$, only those nuisance parameters that affect the expected background distribution $\mathbf{b}$ are varied to the optimal values $\hat{\boldsymbol{\theta}}$ that maximize the likelihood $\mathcal{L}(\mathbf{n}|\mathbf{b}, \boldsymbol{\theta})$.
- Signal-plus-background fit: the nuisance parameters along with the freely floating signal strength modifier μ of the expected signal distribution $\mathbf{s}$ are varied to the optimal values $\hat{\boldsymbol{\theta}}$ and $\hat{\mu}$ that maximize the likelihood $\mathcal{L}(\mathbf{n}|\mu\mathbf{s} + \mathbf{b}, \boldsymbol{\theta})$.

After the maximum likelihood fit, the variations of the nuisance parameter, the pulls defined as

$$\text{pull} = \frac{\hat{\theta} - \theta_0}{\Delta\theta} \quad (4.14)$$

are checked, where $\hat{\theta}$ is the post-fit value of nuisance parameter, θ_0 is the pre-fit value and $\Delta\theta$ is the size of uncertainty returned by the fit. The pull indicates the difference between prefit and postfit values of nuisance in units of its postfit uncertainty. The consistency of the model is tested by inspecting postfit values of nuisance parameters and their uncertainties. Large pull values, for example $> 2\sigma$, and constrained uncertainties, for example $\ll 1\sigma$, usually indicate problems in the signal or background modeling or inconsistent uncertainty model.

4.3 Significance

If an excess of data is observed compared to the background expectation, the presence of the signal is quantified by the background-only p-value which can be interpreted as the probability for the background to fluctuate and give an excess of events as large or larger than the observed one. The test statistic is:

$$q_0 = -2 \ln \frac{\mathcal{L}(\mathbf{n}|\mathbf{b}, \hat{\boldsymbol{\theta}}_0)}{\mathcal{L}(\mathbf{n}|\hat{\mu}\mathbf{s} + \mathbf{b}, \hat{\boldsymbol{\theta}})}, \quad \text{with } \hat{\mu} \geq 0. \quad (4.15)$$

The constraint $\hat{\mu} \geq 0$ is set to prevent downward fluctuations due to the fact that we are not interested in interpreting a deficit of events, but only a positive signal. Having constructed

a distribution $f(q_0|\mathbf{b}, \hat{\boldsymbol{\theta}}_0^{obs})$ by generating pseudo data toys from sampling the MC samples with the probability density function, the p-value corresponding to an observation is:

$$p_0 = P(q_0 \geq q_0^{obs}) = \int_{q_0^{obs}}^{\infty} f(q_0|\mathbf{b}, \hat{\boldsymbol{\theta}}_0^{obs}) dq_0. \quad (4.16)$$

One can also convert the p-value to the notion of significance Z for one-sided Gaussian tail:

$$p_0 = \int_{Z_0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx = \frac{1}{2} P_{\chi_1^2}(Z^2). \quad (4.17)$$

where $P_{\chi_1^2}$ is the survival function of χ^2 for one degree of freedom. A 5σ significance ($Z=5$) corresponds to $p_b = 2.8 \times 10^{-7}$ and it is a convention to claim a discovery in high energy physics. However, it is very CPU-consuming to evaluate such low probabilities. One can use an asymptotic method to estimate the p-value (and corresponding significance) from the observed value q_0^{obs} without generating toy pseudo data:

$$p_0 = \frac{1}{2} \left[1 - \operatorname{erf} \left(\sqrt{q_0^{obs}/2} \right) \right]. \quad (4.18)$$

If the size of the searched parameter phase space is increasing, the chance of statistically significant observation will also increase. This phenomenon is called look-elsewhere effect [43]. In the case of Higgs boson search, if the Higgs boson is searched over a wide mass range, the probability of observing an excess increases. In this case, the local significance for a single Higgs boson mass m_H as previously mentioned should be modified to a global significance. There is another factor which is also affecting the magnitude of the look-elsewhere effect: the mass resolution. A poor mass resolution will give a rather large p-value for a given mass range. Due to this reason, the relatively poor di-tau mass resolution makes the Higgs to a pair of taus searches not so sensitive to look-elsewhere effect.

The global test statistic for search in a broad mass range can be written as:

$$q_0(\hat{m}_H) = \max_{m_H} q_0(m_H). \quad (4.19)$$

For small p-values, an asymptotic procedure can be used [43,44] and the p-value of the global test statistic is:

$$p_0^{global} = P(q_0(\hat{m}_H) > Z_0) \leq N_{Z_0} + p_0^{local}, \quad (4.20)$$

where N_{Z_0} is the number of up-crossing of the likelihood ratio scan $q_0(m_H)$ for a significance Z_0 . N_{Z_0} can be computed by counting the number of times, $N_{Z_{up}}$, the local significance up-crossing the observed test statistic distribution with the following formula:

$$N_{Z_0} = N_{Z_{up}} \exp \left(-\frac{Z_0^2 - Z_{up}^2}{2} \right). \quad (4.21)$$

The up-crossing is illustrated in Figure 4.2.

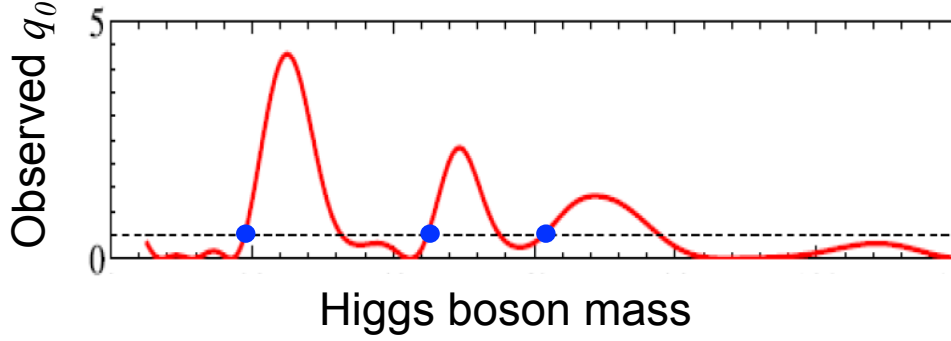


Figure 4.2: An illustration of a scan of the test statistic q_0 vs. m_H [40]. Up-crossings at a certain level of significance Z_{up} are shown by the blue points, where in this example the up-crossing number $N_{Z_{up}}$ is three.

4.4 Exclusion Limits

4.4.1 Observed Limits

If no excess of data is observed on top of the expected backgrounds, exclusion limits can be set on hypothetical model parameters. Many procedures have been developed over the past few years to estimate exclusion limits. In CMS, a modified frequentist method, the CL_s method is used [40, 45–47]. To compare the compatibility of the data with the background-only and signal+background hypotheses, the following profile likelihood ratio is constructed as test statistic q_μ :

$$q_\mu = -2 \ln \frac{\mathcal{L}(\mathbf{n}|\mu\mathbf{s} + \mathbf{b}, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\mathbf{n}|\hat{\mu}\mathbf{s} + \mathbf{b}, \hat{\boldsymbol{\theta}})}, \quad \text{with } 0 \leq \hat{\mu} \leq \mu. \quad (4.22)$$

The signal strength $\hat{\mu}$ maximizes the likelihood. In the numerator, the signal strength and the nuisance parameters may float to reach the maximum. In the denominator, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ give the global maximum of likelihood. The lower constraint $0 \leq \hat{\mu}$ is required to impose that the signal rate is positive. The upper constraint $\hat{\mu} < \mu$ ensures a one-sided (not detached from zero) confidence interval.

The observed value of the test statistic q_μ^{obs} can be determined for the assumed signal strength μ by using equation 4.22. The values of the nuisance parameters $\hat{\boldsymbol{\theta}}_0^{obs}$ and $\hat{\boldsymbol{\theta}}_\mu^{obs}$ that maximize the likelihood, i.e. best describing the experimentally data, for the background-only and signal+background hypotheses can also be determined.

The probability density functions of the test statistics $f(q_\mu|\mathbf{b}, \hat{\boldsymbol{\theta}}_0^{obs})$ and $f(q_\mu|\mu\mathbf{s} + \mathbf{b}, \hat{\boldsymbol{\theta}}_\mu^{obs})$, which assume a signal strength μ in the signal+background hypothesis and for background-only hypothesis ($\mu = 0$) can be generated using Monte Carlo pseudo-data. These distributions are illustrated in figure 4.3.

After constructing the $f(q_\mu|\mathbf{b}, \hat{\boldsymbol{\theta}}_0^{obs})$ and $f(q_\mu|\mu\mathbf{s} + \mathbf{b}, \hat{\boldsymbol{\theta}}_\mu^{obs})$ distributions, one can define two p-values to be associated with the observation for the signal+background and background-

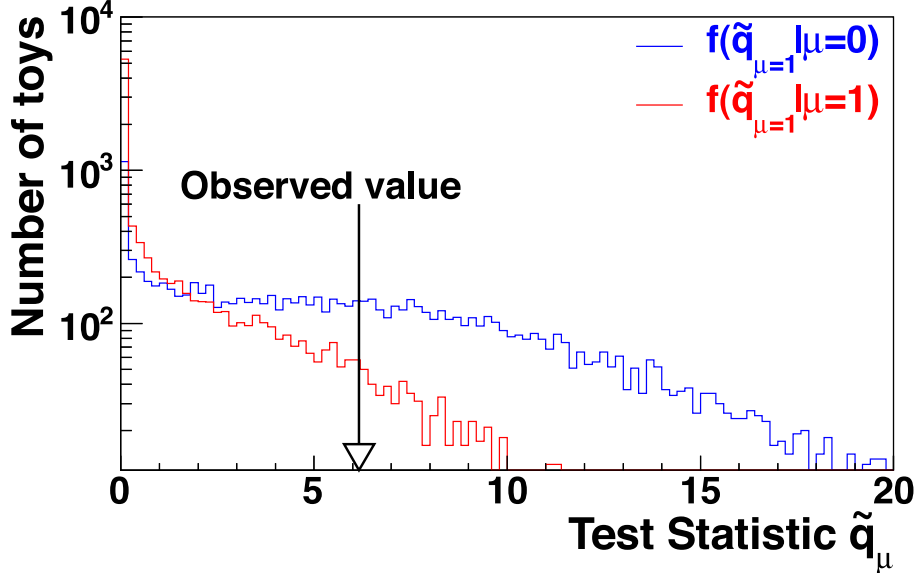


Figure 4.3: Test statistic distributions in the signal+background (red) and background-only (blue) hypotheses. The arrow is shown for the observed values where one can integrate the distributions to compute the p_b and $p_{\mu s+b}$ probability to get $CL_s(\mu)$ value. [40]

only hypotheses, $p_{\mu s+b}$ and p_b :

$$p_{\mu s+b} = P(q_\mu \geq q_\mu^{obs} | \text{signal+background}) = \int_{q_\mu^{obs}}^{\infty} f(q_\mu | \mu s + \mathbf{b}, \hat{\theta}_\mu^{obs}), \quad (4.23)$$

and

$$1 - p_b = P(q_\mu \geq q_\mu^{obs} | \text{background-only}) = \int_{q_\mu^{obs}}^{\infty} f(q_\mu | \mathbf{b}, \hat{\theta}_0^{obs}). \quad (4.24)$$

The $CL_s(\mu)$ value for a given signal strength μ is defined as the ratio of these two probabilities:

$$CL_s(\mu) = \frac{p_{\mu s+b}}{1 - p_b}. \quad (4.25)$$

The 95% confidence level upper limit for the signal strength is μ such that $CL_s(\mu)=0.05$.

4.4.2 Expected Limits

To get the expected median upper limit and $\pm 1\sigma$ and $\pm 2\sigma$ bands, one can also use the background-only hypothesis to generate a large set of background-only pseudo-data and calculate $CL_s(\mu)$ value, as if they were real data. Figure 4.4 (left) illustrates the differential distributions for this case. Then, one can build a cumulative probability distribution based on these results to extract the median (50% quantile), $\pm 1\sigma$ (16% and 84% quantiles), and $\pm 2\sigma$ (2.5% and 97.5% quantiles) bands, as shown in Figure 4.4 (right).

However, generating toys of pseudo-data for every signal strength is not always feasible as it is very CPU-consuming. For large enough expected number of events, one can use the asymptotic method [48] which is an excellent approximation of "true" CL_s limits. A

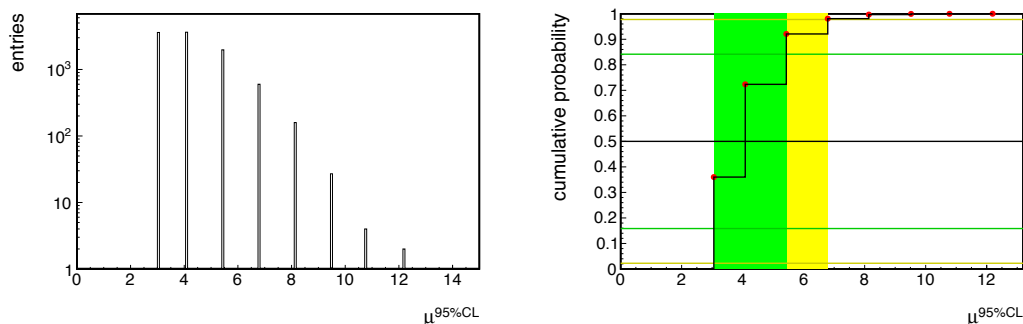


Figure 4.4: Signal strength that give $CL_s = 0.05$ for a set of pseudo-data generated in the background-only hypothesis (left). Cumulative probability distribution (right) of the signal strengths, and extraction of the median and $\pm 1\sigma$ (68%) and $\pm 2\sigma$ (95%) uncertainty bands based on the 2.5%, 16%, 50%, 84% and 97.5% quantiles (horizontal lines). [40]

distribution can be used to approximate the test statistics from pseudo data:

$$f(q_0|\mu') = -2 \ln \lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}). \quad (4.26)$$

In the expression, μ is the test signal strength, $\hat{\mu}$ follows a Gaussian distribution with a mean μ' and standard deviation σ , and N represents the number of events.

The simulated set of dataset to replace pseudo-data is a single dataset called Asimov dataset, defined as the one in which all observed quantities are set equal to their expected values. The computation of the median expected limit and of the uncertainty band is also possible. The test statistic distribution in the Asimov dataset is very close to the toy pseudo-data as shown in Figure 4.5.

4.5 Boosted Decision Tree

The CMS experiment adopts various kinds of multivariate analysis (MVA) methods to perform physics analysis and object reconstruction. MVA methods allow to find a classifier that separates the background and signal events by making use of training of a set of signal and background samples. The most commonly used MVA method is Boosted Decision Trees (BDT) [49, 50]. A typical BDT multivariate analysis include two independent steps: the training step and application step.

In the training, the purpose is to build a single classifier to separate the signal and background events, usually containing the following steps:

1. Divide both signal and background samples into two parts. One is used for training, while the other is used to test the training. Choose appropriate variables that can distinguish signals from background.
2. Determine the cut threshold for a variable which gives the best separation between the signal and the background. Two branches with different signal purities are obtained.

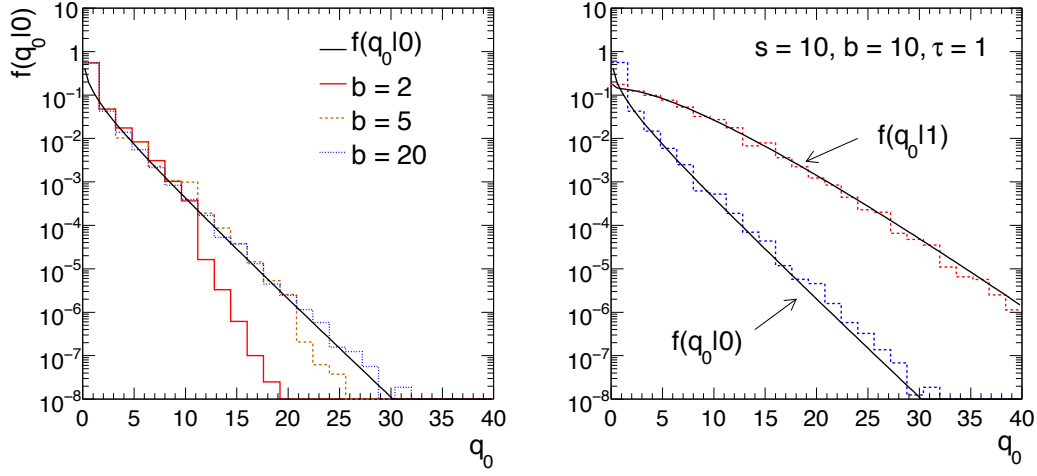


Figure 4.5: **Left:** Test statistic distributions for a background-only hypothesis, in a simple counting experiment with expected background events between 2 and 20. Histograms are test statistic distributions for toy pseudo-data. The curve is obtained by using the asymptotic method. Good agreement starts to appear for five expected events. **Right:** Test statistic distributions in background-only (blue) and signal-plus-background (red) hypotheses. The histograms are toy pseudo data and the curves from the asymptotic method. The agreement is shown for both hypotheses using ten signal events and ten background events. The parameter τ is a scale factor for the number of background events which is set to 1 in this example. [48]

3. The branches of the tree are further split by repeating the steps until a certain number of depths has been created or until all branches have reached a required purity.

These steps are illustrated in Figure 4.6. The most widely used way to estimate the purity is called "Gini". The branch purity is defined as:

$$p = \frac{\sum_{\text{signal}} w_i}{\sum_{\text{signal}} w_i + \sum_{\text{bkg}} w_i}, \quad (4.27)$$

where w_i is the weight of an event i in a branch of n events. One can use it to define the Gini weight:

$$\text{Gini} = \left(\sum_{i=1}^n w_i \right) p(1-p). \quad (4.28)$$

The Gini weight of two daughter branches should be minimized in order to decide the variable's cut value to divide a branch. To consider a leaf as signal leaf, the leaf's purity should be greater than 0.5, otherwise it is a background leaf.

The use of only one decision tree is sensitive to fluctuations in the training samples. A way to prevent it and enhance the performance is to use a "forest" of trees. Boosted decision trees method can combine a set of decision trees together by putting weight to each individual tree. The commonly used boosting algorithm is called adaptive boosting, also known as "AdaBoost". The principle of AdaBoost is that it assumes a larger weight to those misclassified events so that the future trees can learn better. The AdaBoost weight α that

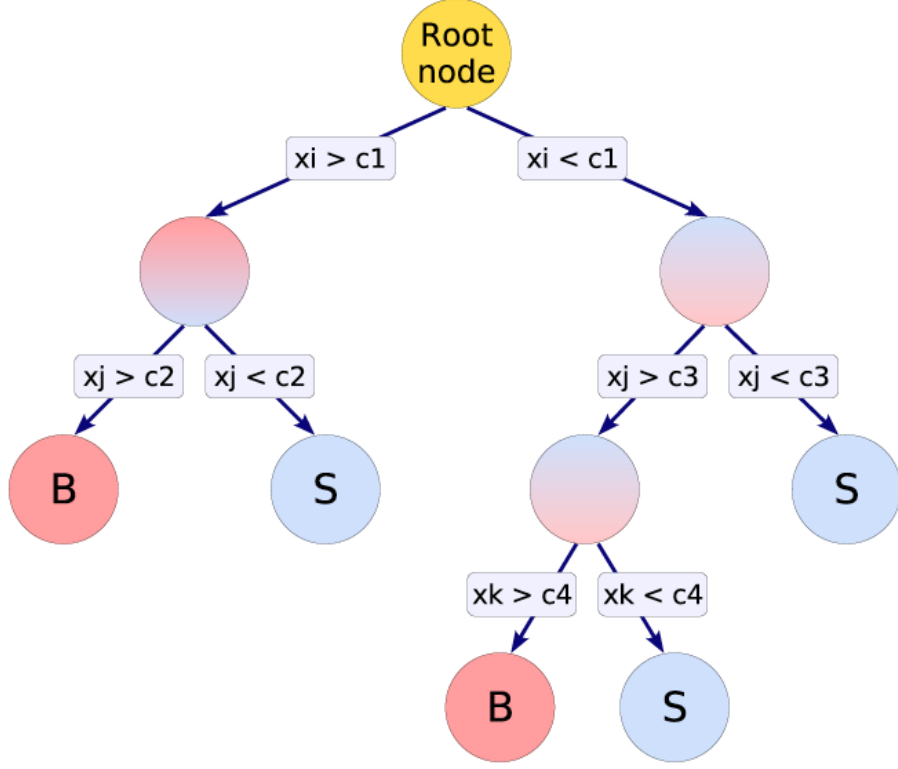


Figure 4.6: Illustration of a decision tree. Branches are created from the root node. Variables are cut by thresholds to provide the best signal and background separation power. [49]

multiplies the weight w_i of all misclassified events of a tree is defined as:

$$\alpha = \frac{1}{2} \ln \frac{1 - \epsilon}{\epsilon}, \quad (4.29)$$

where

$$\epsilon = \frac{\sum_{\text{misclassified}} w_i}{\sum_{\text{all}} w_i}. \quad (4.30)$$

The final score of BDT of an event has the form:

$$Y_{BDT}(\mathbf{x}) = \sum_{m=1}^N \alpha h_m(\mathbf{x}), \quad (4.31)$$

where $\mathbf{x}$ are input values of the variables, m is the index of tree, N is the total number of trees and h_m is the decision of the tree m (1 in signal leaf and 0 in background leaf). One can check if the training is overtrained or not by looking at the agreement of BDT scores of the training and testing samples.

After this, in application stage, the BDT weights computed will be used on observed data and Monte Carlo simulation events not used in the training stage. The cut on the BDT score is chosen to have the best signal and background separation power.

CHAPTER

5

THE CMS EXPERIMENT AT THE CERN LARGE HADRON COLLIDER

Contents

5.1	The CERN Organization	54
5.2	The Large Hadron Collider	56
5.2.1	Proton production and acceleration	57
5.2.2	Experiments at the LHC	60
5.2.3	LHC data taking schedule	60
5.3	The CMS Experiment	61
5.3.1	Overview of the CMS detector	61
5.3.2	Tracker	62
5.3.3	Electromagnetic calorimeter	65
5.3.4	Hadronic calorimeter	66
5.3.5	Magnet	67
5.3.6	Muon system	67
5.3.7	Trigger and data acquisition system	69

One approach to study the properties of elementary particles is to use particle colliders. In this chapter, Section 5.1 introduces the CERN organization and its accelerator complex, including the LHC machine. Section 5.2 describes the LHC machine. Finally, Section 5.3 presents the technical design of the CMS experiment, one of two general purpose detectors at the LHC.

5.1 The CERN Organization

The European Organization for Nuclear Research, known as CERN, is an organization that operates the largest particle physics laboratory in the world. CERN is situated on the borders between Switzerland and France. It was founded in 1954 by 12 European countries with the purpose of stimulating the European nuclear and atomic science. Now, CERN has expanded and it is supported by 22 member states, additional collaboration of observer states and institutions around the world, involving around 10000 scientists.

The fields in which CERN is involved cover a broad range of areas, not only in physics (both theoretical and experimental), but also in engineering and computing fields. Regarding physics, the main subject is particle physics, to which the state-of-the-art machine of CERN, Large Hadron Collider (LHC), belongs. Besides, CERN has also contributed to nuclear physics, plasma physics and astroparticle physics. In the engineering area, the CERN experiments have contributed to expand the knowledge of a wide range of technological applications like electronics, cryogenics, superconducting and construction techniques. In the area of computing, a well-known contribution from CERN is the invention of World Wide Web.

CERN currently is operating several machines. A schematic view of the accelerator complex is illustrated in Figure 5.1. Currently active machines are:

- Two linear accelerators (LINAC 2 and LINAC 3) that generate low energy particles. LINAC 2 accelerates protons to 50 MeV for injection into the Proton Synchrotron Booster (PSB), and LINAC 3 provides heavy ions at 4.2 MeV/ u for injection into the Low Energy Ion Ring (LEIR).
- The Proton Synchrotron Booster (PSB) increases the energy of particles generated by the proton linear accelerator before they are transferred to the other accelerators.
- The Low Energy Ion Ring (LEIR) accelerates the ions from the ion linear accelerator LINAC 3, before transferring them to the Proton Synchrotron (PS).
- The 28 GeV Proton Synchrotron (PS), built during 1954-1959 and still operating as a pre-accelerator to the more powerful Super Proton Synchrotron (SPS).
- The Super Proton Synchrotron (SPS), a circular accelerator with a diameter of 2 kilometers, which started operation in 1976. It was designed to deliver an energy of 300 GeV and was gradually upgraded to 450 GeV. As well as having its own beam lines for fixed-target experiments (currently COMPASS and NA62), it has been operated as a proton-antiproton collider (the $S\bar{p}pS$ collider), and for accelerating high energy electrons and positrons which were injected into the Large Electron-Positron Collider (LEP). Since 2008, it has been re-used to inject protons and heavy ions into the LHC.
- The Isotope Separator Online (ISOLDE), which is used to study unstable nuclei. The radioactive ions are produced by the impact of protons at an energy of 1.0-1.4 GeV from the PSB. Radioactive EXperiment ISOLDE (REX-ISOLDE) accelerates radioactive ion beams and thus the full variety of beams available at ISOLDE become accessible as accelerated beams for experiments. The High Intensity and Energy ISOLDE (HIE-ISOLDE) project aims at several important upgrades of the present ISOLDE facility.

The main goal is to increase the energy of the radioactive ion beams from 3 MeV/u up to 10 MeV/u through the construction of a superconducting accelerator.

- The Antiproton Decelerator (AD), which reduces the velocity of antiprotons to about 10% of the speed of light for research of antimatter.
- The neutron Time-of-Flight (n-TOF) facility, which consists of a pulsed source, a flight path of 200 m length, and a detector systems. It is designed to study neutron-nucleus interactions for neutron kinetic energies ranging from a few MeV to several GeV.
- The CERN neutrino platForm (CENF) supports generic detector, neutrino beams R&D and large detector prototypes or demonstrators.
- The Gamma Irradiation Facility (GIF++) is a new irradiation facility to test large-area detectors for the future high luminosity LHC program. It is a unique place where high energy charged particle beams from SPS are combined with a 14 TBq $^{137}\text{Cesium}$ source.
- The Advanced WAKefield Experiment (AWAKE) is an accelerator R&D project based at CERN. It is a proof-of-concept experiment investigating the use of plasma wakefields driven by a proton bunch to accelerate charged particles.
- The High-Radiation to Materials (HiRadMat) is a users facility at CERN, designed to provide high-intensity pulsed beams to an irradiation area where material samples as well as accelerator component assemblies can be tested.
- The Extra Low ENergy Antiproton ring (ELENA) is a compact ring for cooling and further deceleration of 5.3 MeV antiprotons delivered by the AD.
- The Proton Irradiation Facility (IRRAD) and the Cern High energy AcceleRator Mixed field facility (CHARM) are radiation test facilities. Most of the protons pass through the IRRAD facility without interacting, while the mixed-field environment in CHARM resembles that present in the vicinity of a high-energy accelerator and can be adapted to the application conditions by selecting different test configurations. They both make use of proton beams from the PS.
- The CERN Linear Electron Accelerator for Research (CLEAR) is a facility for general accelerator R&D and component studies. The main goals are the prototyping and validation of accelerator components for the LHC upgrade and studies of high-gradient acceleration methods.

Since the founding of CERN, the CERN experiments have made many discoveries in the fields of fundamental physics. The Gargamelle bubble chamber led to the discovery of the neutral currents in 1973 [52, 53]. The UA1 and UA2 experiments in SPS collider, discovered the $W^\pm$ and Z bosons in 1983 [54–57]. Direct CP-violation was observed by the NA 31 and NA48 experiments [58–60]. In 1989, the Large Electron-Positron Collider (LEP) has started. From 1989 to 2000, LEP has performed high precision measurements on the parameters of the Standard Model and provided hints to the search for the Higgs boson. Finally, at the LHC, in 2012 ATLAS and CMS observed a neutral scalar particle which matched the long sought-after standard model Higgs boson [61, 62].

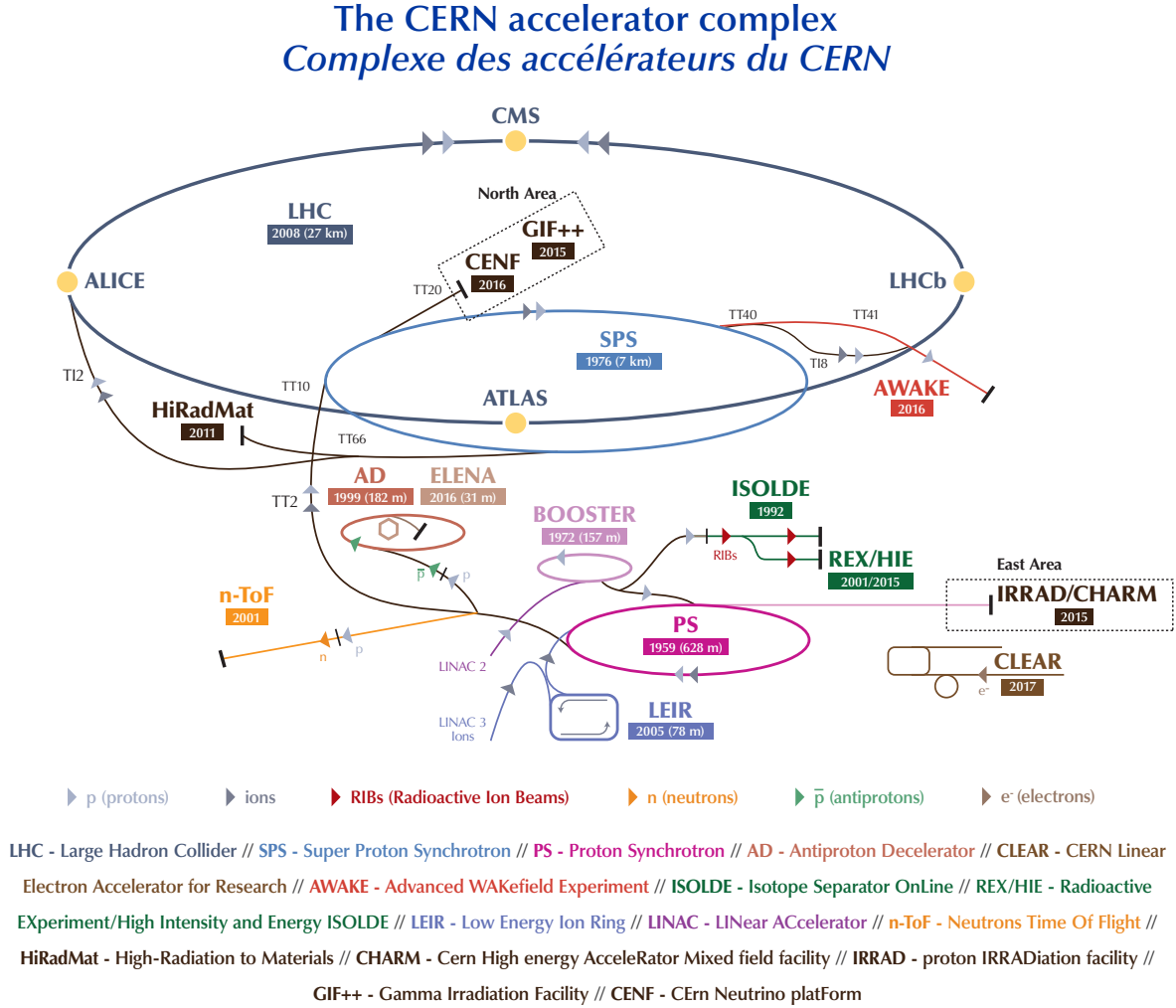


Figure 5.1: Accelerator complex at CERN, some of which are used to inject particle beams to the LHC accelerator [51].

5.2 The Large Hadron Collider

The Large Hadron Collider (LHC) [63] is a superconducting accelerator built to explore physics at the TeV scale. It is operated by CERN, located in a 26.7 km circular underground tunnel located near Geneva, crossing border between Switzerland and France. The tunnel was constructed between 1984 to 1989 for the use of LEP. The layout of the LHC machine is shown in Figure 5.2 which consists of eight 528 m long straight sections and eight 2.45 km long arcs. Each of the eight straight sections serves as site for experiments or machine utilities. They are numbered 1-8 in the clockwise direction. The four main detectors: ATLAS, ALICE, CMS and LHCb are located at Point 1, 2, 5 and 8, respectively. The two particle beams are injected into the LHC at Points 2 and 8 by the SPS and are accelerated in opposite directions. It requires the existence of two rings with opposite magnetic dipole fields and separate vacuum chambers. The eight arc sections contain a total of 1232 niobium-titanium

superconducting dipole (Figure 5.3) magnets providing the 8.3 T magnetic field required to deflect the particles beam trajectory and 392 quadrupole (Figure 5.4) magnets for their collimation. The superconducting magnets are cooled to a temperature of 1.9 K by a liquid helium cryogenic system.

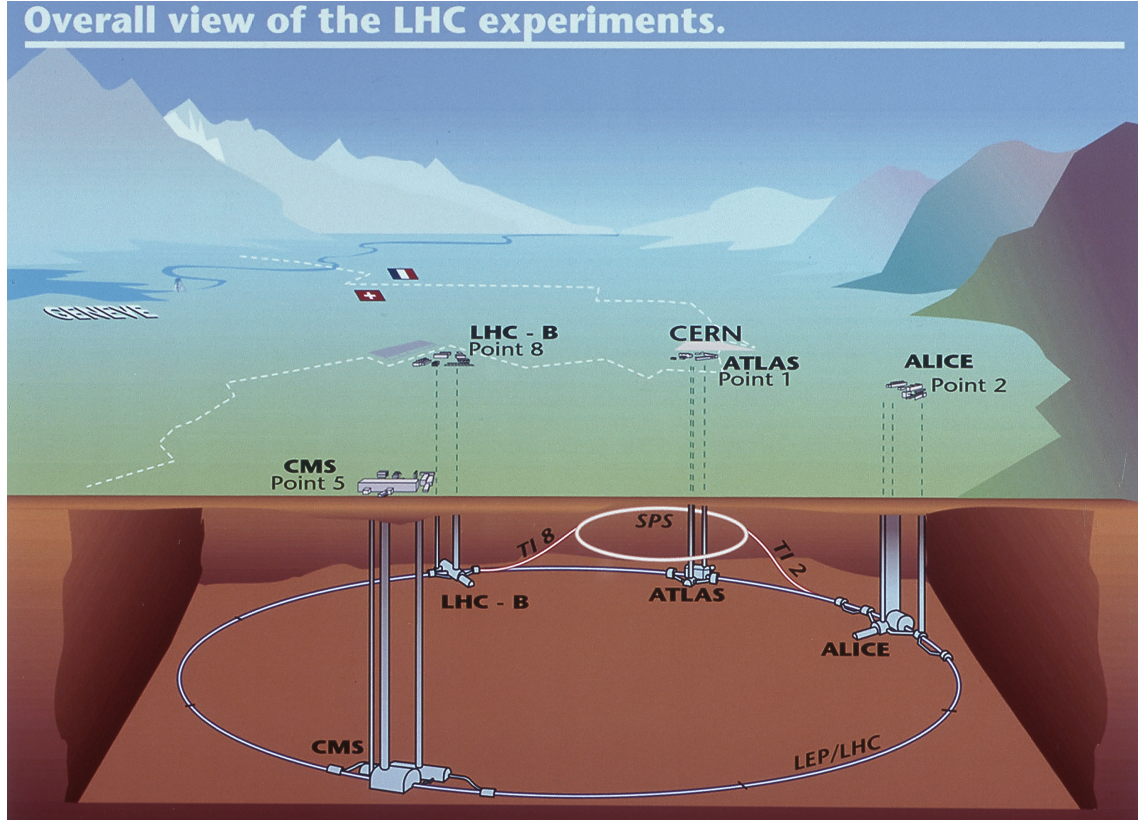


Figure 5.2: Schematic of the LHC machine layout [64].

The design of LHC supports both heavy ions and proton-proton collisions. Thus, a wide range of physics topics can be explored. In proton-proton collision mode, Higgs physics is a the main focus along with searches for new physics beyond standard model. It also serves as a top quark factory so that to measure precisely the properties of top quark. In addition, it also allows the investigation of electroweak physics, heavy flavor physics and QCD. In the heavy ions collision mode, LHC allows the study of quark-gluon plasma physics. The following sections will describe only the proton-proton collisions as they correspond to the data used in the next chapters.

5.2.1 Proton production and acceleration

The protons are obtained from ionized hydrogen in duoplasmatron. They are first accelerated to 50 MeV in the linear accelerator, the LINAC2. Then, they are injected in a circular accelerator, the Proton Synchrotron booster, which increase the proton energy to 1.4 GeV. After that, they enter the Proton Synchrotron (PS), which pushes their energy to 26 GeV. From that, they are sent to the Super Proton Synchrotron (SPS). The SPS is used to accelerate the proton energy to 450 GeV, ready to be transferred to the LHC. The LHC, then, starts

LHC DIPOLE : STANDARD CROSS-SECTION

CERN AC/DI/MM - HE107 - 30 04 1999

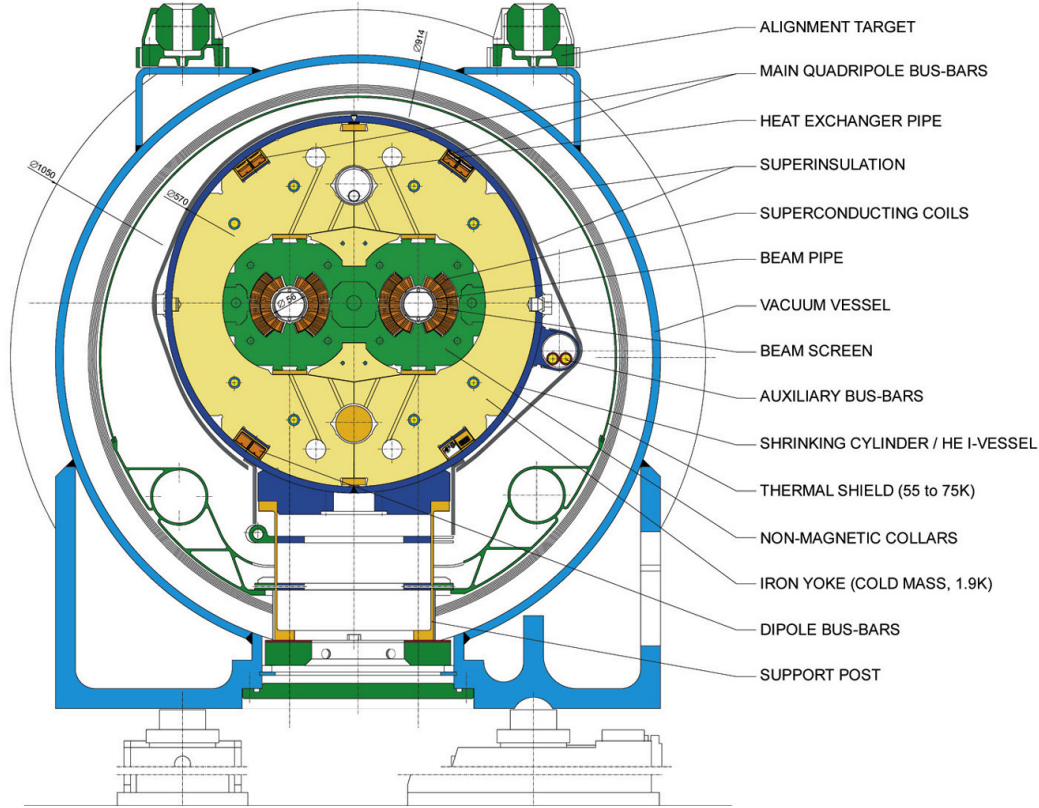


Figure 5.3: Diagram of the cross-section of a dipole of LHC [63].

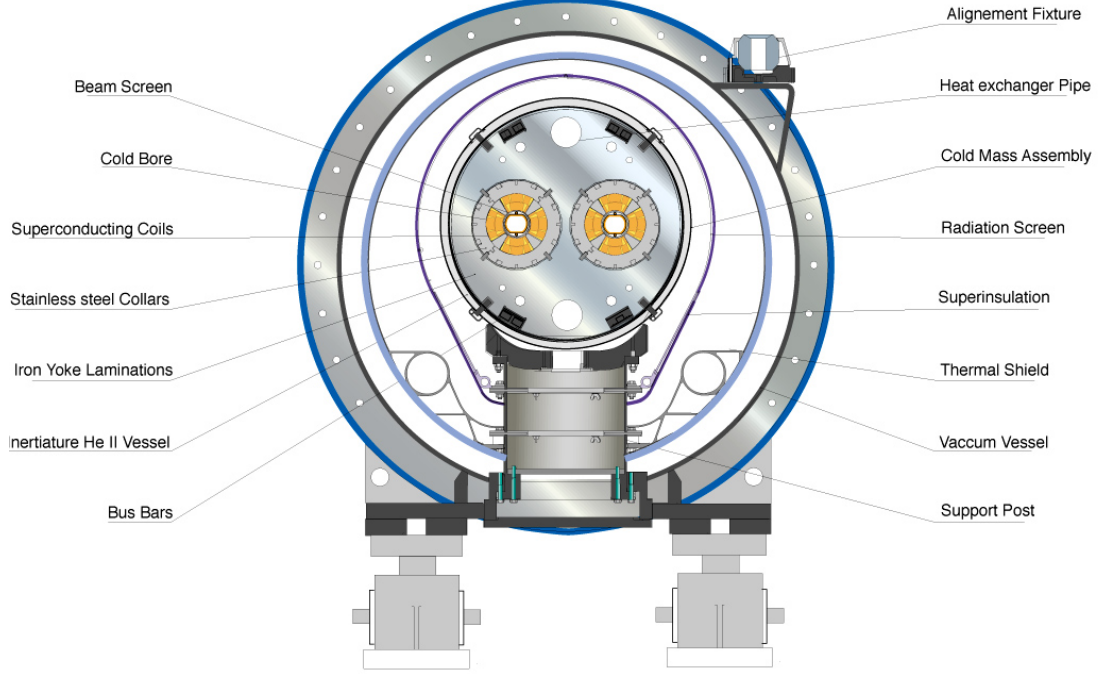
the final acceleration of the protons, it squeezes the bunches, increasing the instantaneous luminosity, and adjust the beams position, triggering the collisions at the crossing points. Three to four cycles of the PS synchrotron are required to fill the SPS. Twelve cycles of the SPS are needed to fill the LHC. Totally, it takes twenty minutes to inject proton beams to the LHC, and twenty more minutes are required to increase the proton energy from 450 GeV to 6.5 TeV. When completed, the LHC is capable of storing 2808 proton bunches that consist of approximately 10^{11} protons. The distance between two successive bunches, the bunch spacing, is 25 ns, which corresponds to a bunch crossing frequency of 40 MHz.

The large amount of bunches and the short spacing in between them lead to a huge number of collisions taking place in a given time interval. One can quantify the interaction rate, $\frac{dN}{dt}$, which has the following relation:

$$\frac{dN}{dt} = \mathcal{L}\sigma, \quad (5.1)$$

where σ is the process cross section and $\mathcal{L}$ is the instantaneous luminosity which is a measure of how many particles cross a given area per time interval. The nominal design instantaneous luminosity is $10^{34}\text{cm}^{-2}\text{s}^{-1}$. One can use the beam parameters to determine the instantaneous

LHC quadrupole cross section



CERN AC - SQ1 - 12/97

Figure 5.4: Diagram of the cross-section of a quadrupole of LHC [63].

luminosity:

$$\mathcal{L} = \frac{N_p^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F, \quad (5.2)$$

where N_p is the number of protons in each bunch, n_b is the number of bunches in each beam, f_{rev} is the revolution frequency, γ is the relativistic gamma factor of the protons, ϵ_n is the normalized transverse beam emittance, which is a parameter to quantify the spread of the beam, β^* is a parameter to quantify how much the beam is squeezed toward the interaction point, and finally F is the geometric luminosity reduction factor due to the crossing angle at the interaction point.

The integral of the instantaneous luminosity over time gives the integrated luminosity, later simply referred to luminosity, expressed in units of inverse cross section, given by

$$L = \int_0^T \mathcal{L} dt. \quad (5.3)$$

The integrated luminosity quantify how much data are collected during a certain data taking periods. The number of expected events of a process with cross section σ in certain time period is given by:

$$N = L\sigma. \quad (5.4)$$

5.2.2 Experiments at the LHC

At the LHC, there are four major experiments built at each collision point and other more specific experiments located close to the major ones. The major four experiments are:

- CMS (Compact Muon Solenoid) [65] is a general-purpose detector located in Point 5, which will be described in details in Section 5.3. The aim was to search for the Higgs boson, to perform precision measurements of SM parameters and to probe new physics beyond the SM.
- ATLAS (A Toroidal LHC ApparatuS) [66] shared the same physics scopes as CMS but exploits a different technical structure, including an innovative toroidal magnet, situated at Point 1.
- ALICE (A Large Ion Collider Experiment) [67] is dedicated to the study of heavy ion collisions and the quark-gluon plasma, placed at Point 2.
- LHCb (Large Hadron Collider beauty) [68] is designed to study of physics of CP violation and b-mesons, located at Point 8.
- LHCf (Large Hadron Collider forward) [69] is designed to measure the neutral collision products in the forward region (close to beam direction) as a source to simulate cosmic rays in laboratory conditions, located near ATLAS.
- MoEDAL (Monopole and Exotics Detector At the LHC) [70] is used to search for a hypothetical particle with magnetic charge: the magnetic monopole, located near LHCb.
- TOTEM (TOTAl, Elastic and diffractive cross-section Measurement) [71], located near CMS, is dedicated to the precise measurements of forward protons.

5.2.3 LHC data taking schedule

The first beams have been injected in the LHC in 2008. Shortly after that, an electrical fault occurred, causing the leakages of liquid helium in the tunnel. Up to 700 meters of magnets were damaged and the LHC roadmap had to be reconsidered. After one year, LHC had recovered from the incident and made first collisions at a center-of-mass energy of 450 GeV on November 2009. In March 2010, LHC Run 1 started and the first 7 TeV center-of-mass collisions were recorded. The LHC delivered an integrated luminosity of 6.14 fb^{-1} of 7 TeV in 2010 and 2011. In 2012, the LHC raised the collision energy to 8 TeV and delivered 23.30 fb^{-1} before the first long shut down (LS1). In 2015, the LHC run 2 started with 13 TeV collision energy. It delivered 4.22 fb^{-1} data in 2015.

During the 2016 data taking period, LHC delivered 40.82 fb^{-1} of data. CMS recorded 37.76 fb^{-1} and 35.9 fb^{-1} were certified as good quality to be used for physics analyses, as those described in this thesis. In 2017, LHC delivered 49.79 fb^{-1} and CMS recorded 44.98 fb^{-1} . In 2018, LHC delivered 68.18 fb^{-1} and CMS recorded 63.98 fb^{-1} .

Now run 2 is finished. The second long shutdown (LS2) is started and will last until 2020. The machine will be upgraded to deliver collisions at center-of-mass energy of 14 TeV. Run 3 will extend until 2023. The high-luminosity LHC (HL-LHC) should start from 2026 and

extend to 2037. At the end of run 3, the LHC should deliver 300 fb^{-1} integrated luminosity in total, while by the end of 2037, 3000 fb^{-1} should be delivered. Figure 5.5 shows the roadmap of LHC.

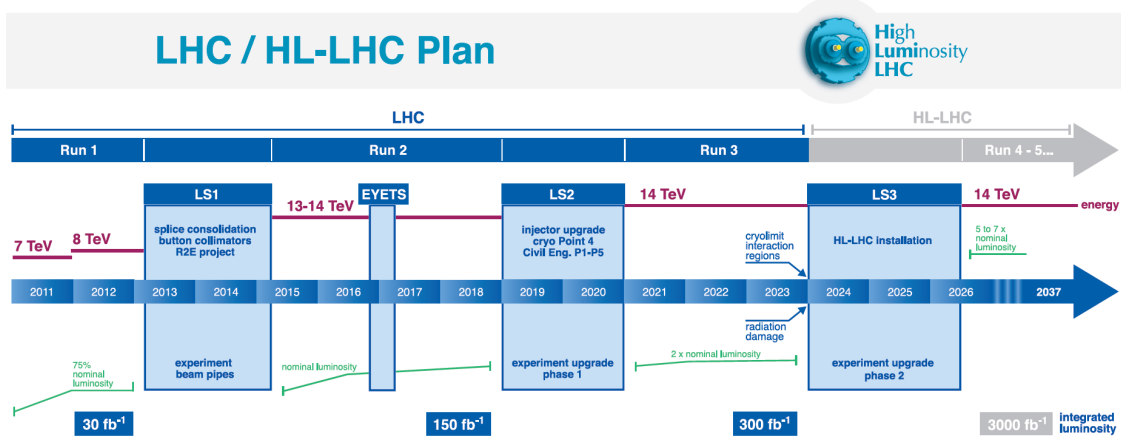


Figure 5.5: Overview of LHC plan until 2037. The LHC run 1 to run 3 period will deliver 300 fb^{-1} integrated luminosity. The high luminosity LHC (HL-LHC) will deliver 3000 fb^{-1} by the end of 2037 [73].

5.3 The CMS Experiment

The Compact Muon Solenoid (CMS) experiment [65] is a general purpose detector that covers a broad range of physics scope, from the precision measurement of the SM parameters and the Higgs boson properties to the search for the BSM physics. The name "CMS" reflects three features of its detector: a limited size considering its complexity, an advance muon sub-detector, and a solenoidal superconducting magnet. The CMS detector is designed to meet the following requirements: a good muon identification and momentum resolution, a good charged-particle momentum resolution and reconstruction efficiency, a good electromagnetic energy resolution and good missing transverse energy and dijet mass resolution [65].

5.3.1 Overview of the CMS detector

The CMS detector is a detector covering 4π solid angle. It has a length of 21.6 m, a diameter of 14.6 m and a total weight of 12500 ton. It has different layers that consist of subdetectors designed to measure different properties of particles created in collisions. It is arranged in a central cylinder, the barrel and closed by two endcaps.

The positions and dimensions of the subdetectors are given with respect to the coordinate systems used within CMS. The cartesian right-handed coordinates are defined with respect to the nominal interaction point of the proton beams, while the x-axis points to the LHC ring center and the y-axis points vertically. The z-axis is orthogonal to the other axes and is directed along the beam direction, towards the Jura mountains from Point 5. The radial distance r is defined in the x - y plane. The azimuthal angle ϕ is measured in the plane defined by the x- and y-axes, from the x-axis. The polar angle θ is defined with respect to the z-axis,

in a plane orthogonal to the xy-plane. In high energy physics, it is convenient to define pseudorapidity η as:

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right). \quad (5.5)$$

In the relativistic limit, where $E \gg m$, this quantity equals to the rapidity y :

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \quad (5.6)$$

The difference of pseudorapidity of two particles $\Delta\eta$ is invariant under a Lorentz transformations in the z direction which makes it advantageous to use in the analysis.

Figure 5.6 shows a schematic view of the CMS detector. The components, which will be discussed in detail in the next sections, are, from the innermost to the outermost part of the detector:

- A silicon based tracking system (tracker), containing both pixel and strip trackers, which measures the trajectory of charged particles and reconstructs the secondary vertices;
- An electromagnetic calorimeter (ECAL), which measures the energy of electrons and photons;
- A hadronic calorimeter (HCAL), which measures the energy of hadrons;
- A superconducting magnet, which provides a uniform 3.8 T magnetic field parallel to the direction of beam and deflects the flight path of charged particles;
- A muon system, which measures the energy and the trajectory of muons.

In addition, due to the high interaction rate of the LHC, trigger and data acquisition systems have been designed to collect only those events which are interesting for physics studies.

5.3.2 Tracker

The tracker is the innermost subdetector layer of the CMS. Its functions are to record the position of charged particles traversing its layers and to measure the curvature of the trajectories bended by the magnetic field. The momentum of charged particles can then be estimated with a great precision, the primary vertex of collision and the secondary vertices from long-lived particle decays can be reconstructed. Because of the very high number of particles produced in overlapping proton-proton collisions, two requirements need to be fulfilled to ensure precision measurements:

- the tracker must have large hit redundancy. This can be accomplished by having a great amounts of tracker layers.
- the tracker must have a high granularity and the amount of material used has to be kept low to limit photon conversion, multiple scattering, bremsstrahlung and nuclear interactions. This can be achieved by having high density of radiation-proof electronics.

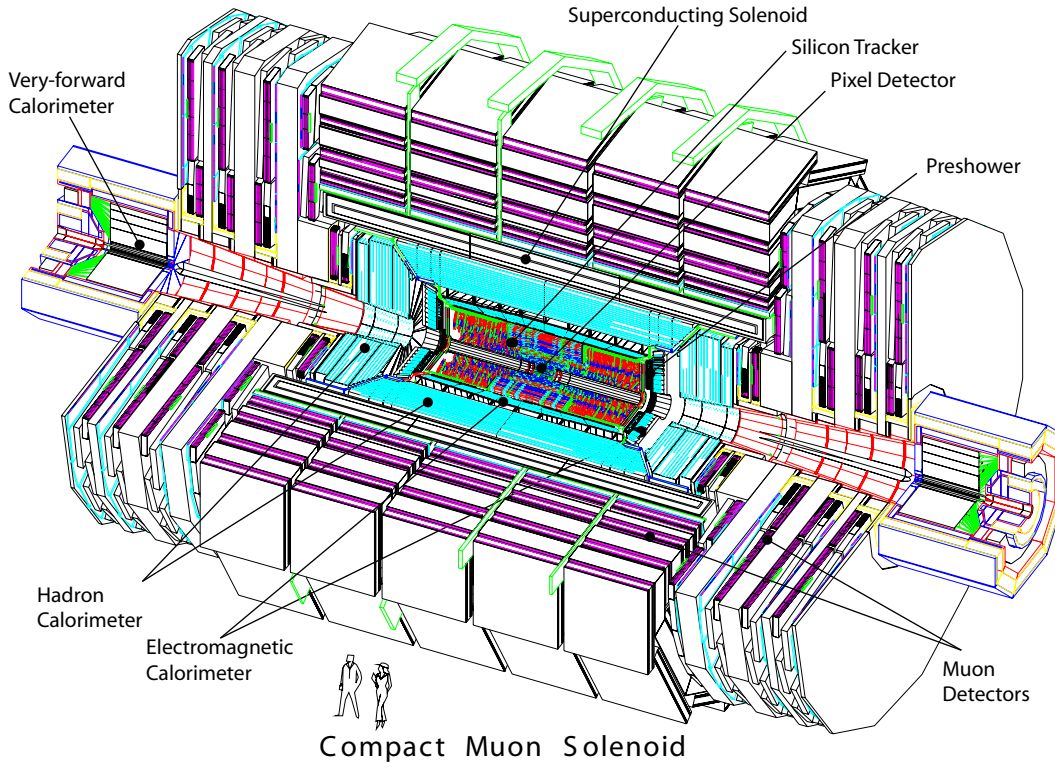


Figure 5.6: Schematic view of the CMS detector showing the different components [74].

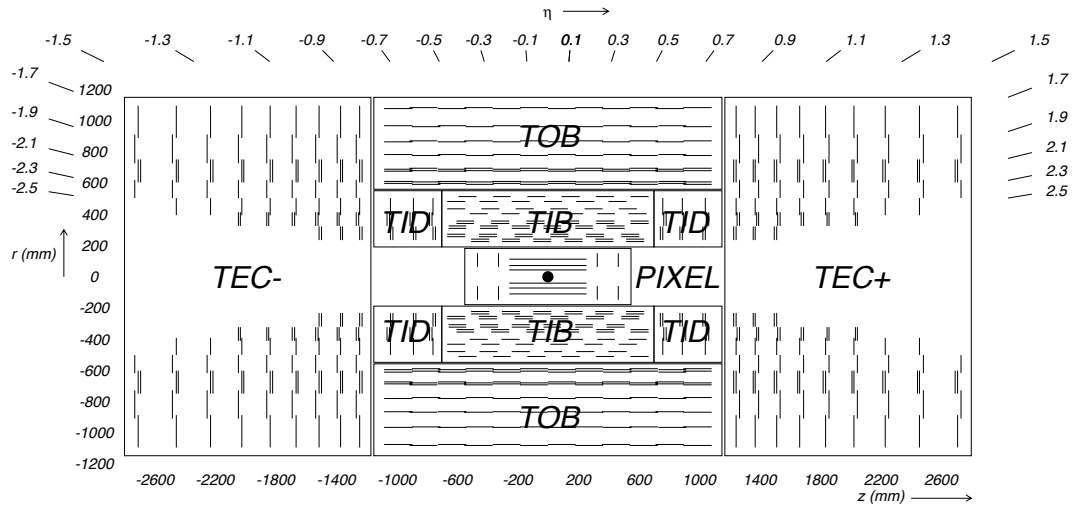


Figure 5.7: Schematic view of the CMS tracker [65]. The line represents a detector module. Double lines represent a double-sided modules.

The tracker has a cylinder shape which has a length of 5.5 m and a diameter of 2.4 m, covering a pseudorapidity region up to $|\eta| < 2.5$. A schematic drawing is shown in Figure 5.7. It consists of a silicon pixel detector part and a silicon strip detector part.

The pixel detector part is placed directly around the beam pipe. It is served as a starting point for reconstruction of charged particle tracks. It is also crucial for the precise measurement of impact parameter. Figure 5.8 shows a schematic view of the structure of pixel detector. It is divided into a barrel with three layers at radii of 4.4, 7.3 and 10.2 cm, and two layers of endcap disk at $|z| = 34.5$ and $|z| = 46.5$ cm with a diameter between 12 and 30 cm. The pixels provide a hit position spatial resolution of 10-40 μm . In total, 66 million pixels are installed on 1440 modules covering an area of approximately 1 m^2 .

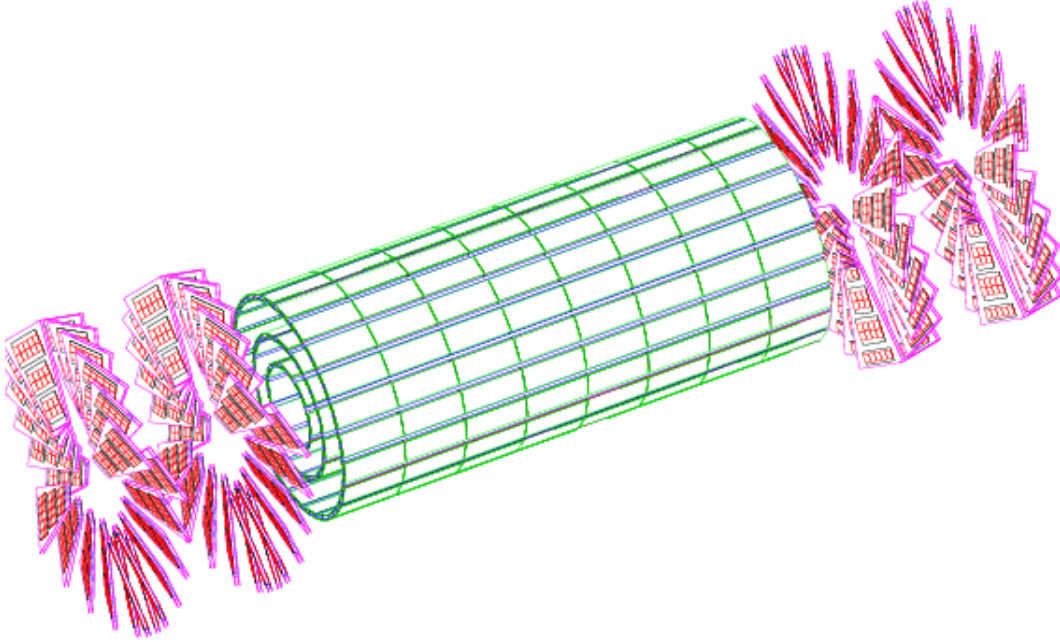


Figure 5.8: A 3D view of the CMS pixel detector part of tracker [65], showing the barrel layers (green) and endcap disks (red).

The strip detector part is placed around the pixel detector. It has 15148 modules, which in total have 9.3 million strips covering an area of approximately 198 m^2 . It is located in the radial region, $20 \text{ cm} < |r| < 116 \text{ cm}$, and consists of four different subsystem. The Tracker Inner Barrel (TIB) has four layers, which cover the radial region with $20 \text{ cm} < |r| < 55 \text{ cm}$. The Tracker Inner Disks (TID) has three disks at each ends of TIB. TIB/TID provide position measurements in $r - \phi$ plane with a resolution of 13-38 μm . The Tracker Outer Barrel (TOB) is placed outside the TIB/TID system and has six layers covering radial region with $55 \text{ cm} < |r| < 116 \text{ cm}$ and z axis between $\pm 118 \text{ cm}$. The Tracker EndCaps (TEC) cover the region with $124 \text{ cm} < |z| < 282 \text{ cm}$ and $22.5 \text{ cm} < |r| < 113.5 \text{ cm}$. Each TEC consists of nine disks. TOB and TEC give position measurement in $r - \phi$ plane with a resolution of 18-47 μm .

5.3.3 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) encloses the tracker and covers pseudorapidity region up to $|\eta| < 3.0$. It is divided into barrel (EB) ($|\eta| < 1.479$) and endcap (EE) ($1.479 < |\eta| < 3.0$). Figure 5.9 illustrates a quadrant of ECAL. The task of the ECAL is to measure the energy of electrons and photons with good resolution and identify them based on the specific shape of their energy deposits. It makes use of scintillating lead tungstate (PbWO_4) crystals. When electrons or photons enter the crystals, electromagnetic showers are produced and excite the crystal atoms. After the atoms de-excite, they emit scintillation light which is collected by silicon avalanche photodiodes (APD) in the barrel or vacuum phototriodes (VPT) in the endcaps. A preshower (ES) detector which consists of planes of silicon sensors and lead is placed in front of endcaps.

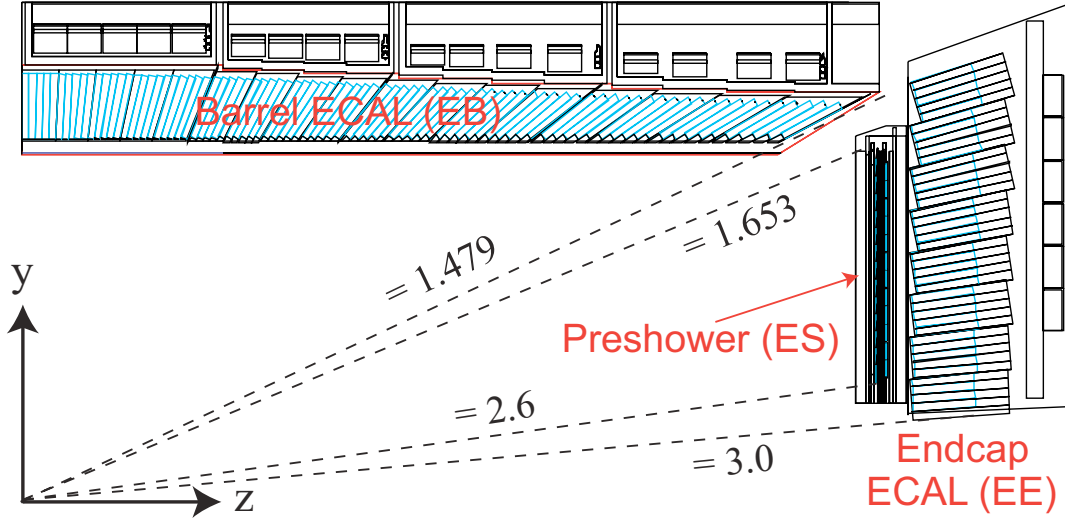


Figure 5.9: A schematic view of one quadrant of the ECAL. The barrel ECAL and the endcap ECAL consist of lead tungstate crystals, silicon avalanche photodiodes or vacuum phototriodes respectively. A preshower is placed in front of the endcap ECAL [74].

ECAL is composed of 75848 lead tungstate crystals in total. The crystal material has a high density ($\rho = 8.28\text{g/cm}^3$), a short radiation length ($X_0 = 0.89\text{cm}$) and a small Moliere radius ($R_m = 2.2\text{cm}$), so that the calorimeter can be compact with fine granularity and high spatial resolution. It also has a fast scintillation latency such that 80% of the light emitted in 25 ns, the LHC design bunch crossing. The length of the crystals is 23 cm, equals to 25 radiation lengths, allowing to measure the whole shower produced by electrons and photons.

The scintillation light is detected by APD in the barrel, two attached to every crystal. In the endcaps, where the radiation rate is high, one VPT is attached to the back of every crystal. APD is chosen because it is more efficient, while VPT is more radiation-resistant. The crystal light yield and the APD gain are temperature dependent, with approximately $-2\%/^{\circ}\text{C}$. Therefore, special care is taken to make the temperature stable within 0.05°C in the barrel and 0.1°C in the endcaps.

The preshower calorimeter detector is used to identify the neutral pions, which decay to a closely spaced pair of photons, in the region $1.653 < |\eta| < 2.6$. It also helps discriminating

electrons from minimum ionizing particles (MIP), i.e., muons and hadrons. It is a sampling calorimeter with the absorber material and sensors that are interleaved. The first layer is a lead plate which initiates electromagnetic showers from the incoming electron and photons. The second layer is made by silicon sensors which measures the deposited energy and shower shape. The identification of neutral pions is achieved by having three radiation lengths for the lead layer, so that about 95% of photons from the neutral pion decay shower before they reach the silicon sensors.

The energy resolution of the ECAL can be parameterized as:

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \quad (5.7)$$

where E is the energy, S is the stochastic term representing the statistical fluctuations in the number of secondary particles produced in shower and collected, N is the noise term coming from electronic device, and C is a constant term arises from energy leakage, non-uniformity of the light collection and calibration uncertainties. In electron test beam studies, the parameters were determined as: $S = 0.028\sqrt{\text{GeV}}$, $N = 0.12 \text{ GeV}$ and $C = 0.003$ in the barrel region, $S = 0.05\sqrt{\text{GeV}}$, $N = 0.5 \text{ GeV}$ and $C = 0.003$ in the endcaps region [75]. The C term becomes dominant at high energy and it can be reduced with stable operating conditions.

5.3.4 Hadronic calorimeter

The ECAL is enclosed by the hadronic calorimeter (HCAL). The purpose of the HCAL is to identify the hadronic jets and to detect the imbalances in the transverse energy. It is a sampling calorimeter with alternating layers of brass absorber and plastic scintillator, except the forward region where steel absorbers and quartz fibers are used as they are more radiation hard.

The architecture of the HCAL is shown in Figure 5.10. The hadron barrel (HB) and endcaps (HE) calorimeter is located inside the magnet coil, covering the $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$ pseudorapidity region, respectively. The HB is divided into towers with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The tower granularity of HE has dimension ranging from $\Delta\eta \times \Delta\phi = 0.087 \times 0.08$ to 0.035×0.08 . The scintillation light is detected by hybrid photodiodes (HPD) that can function in high axial magnetic fields. Outside the magnet, an outer calorimeter (HO) is added to complement the HB to achieve a better sampling in the region $|\eta| < 1.26$. It has also HPD readout. Finally, the forward hadron calorimeter (HF), which covers the very high pseudorapidity region $3.0 < |\eta| < 5.0$ placed 11.5 m away from the collision point, is made of radiation-hard steel absorbers and quartz fibers as active material. The signal is generated via Cherenkov light and collected by photo-multipliers (PMT).

The measured hadron energy resolution of the HCAL and the ECAL combined can be described by [76]:

$$\frac{\sigma(E)}{E} = \frac{100\%}{\sqrt{E(\text{GeV})}} + 5\% \quad (5.8)$$

where the energy E is given in GeV.

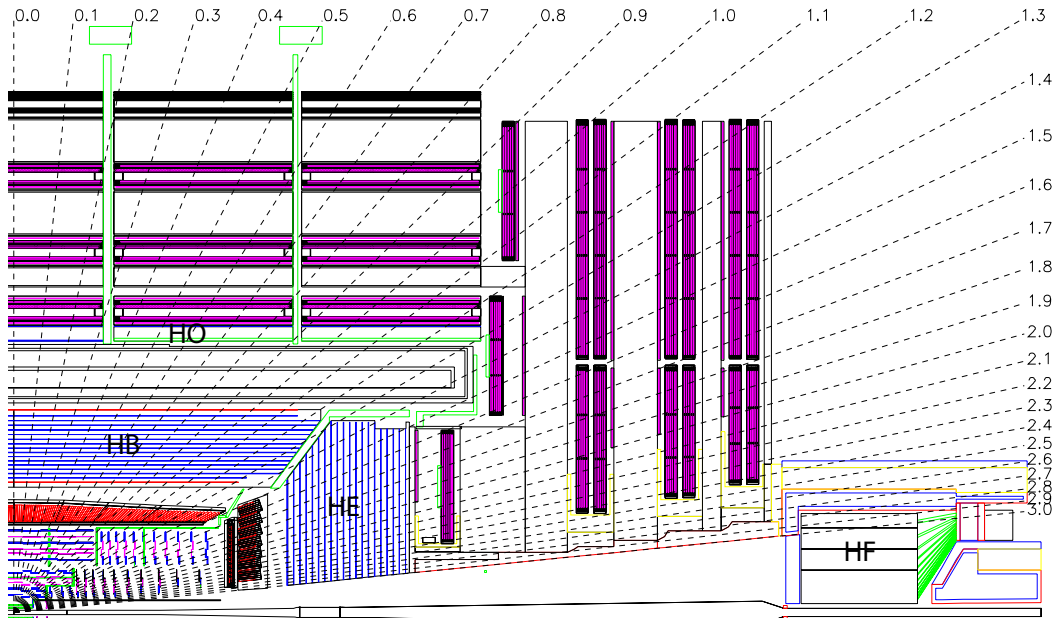


Figure 5.10: A schematic view of one quadrant of the HCAL [65].

5.3.5 Magnet

The CMS magnet is used to bend the tracks of the charged particles. It is the most powerful and the biggest ever built superconducting solenoid. It encloses most of the detector subcomponents (tracker, ECAL, HB and HE) and has a diameter of 6 m and a length of 13 m. The energy stored in the magnet can go up to 2.5 GJ when it is fully functional to provide a 3.8 T magnetic field. The magnet is cooled by using liquid helium. In order to reduce the effects of the intense magnetic field on the devices outside the magnet, a massive iron yoke (10000 tons) is installed outside the magnet to return the magnetic flux.

5.3.6 Muon system

The muon system is the outermost part of the CMS detector. It is positioned outside the magnet and integrated with the magnetic flux-return iron yoke. The coverage region is up to $|\eta| < 2.5$. The whole muon system has a total area of detection planes of 25 000 m² which can be further divided into barrel ($|\eta| < 1.2$) and endcaps ($1.2 < |\eta| < 2.4$) regions.

The working principle for muon system is to use ionized gas chambers. An muon does not interact via the strong force and is too massive to radiate via bremsstrahlung. Due to the long life time of the muon and the relativistic time dilation in the laboratory frame, muon travels through the detector and escapes from it. The muon leave hits when it ionizes the gas of the detector. The path of a muon is reconstructed by using the hits in the muon chambers and the information collected from tracker.

The muon system has a large area of detection planes, thus the technology needs to be inexpensive, reliable and robust. Therefore, a combination of three different types of gas

chambers are in use. Figure 5.11 illustrates the architecture of the muon system.

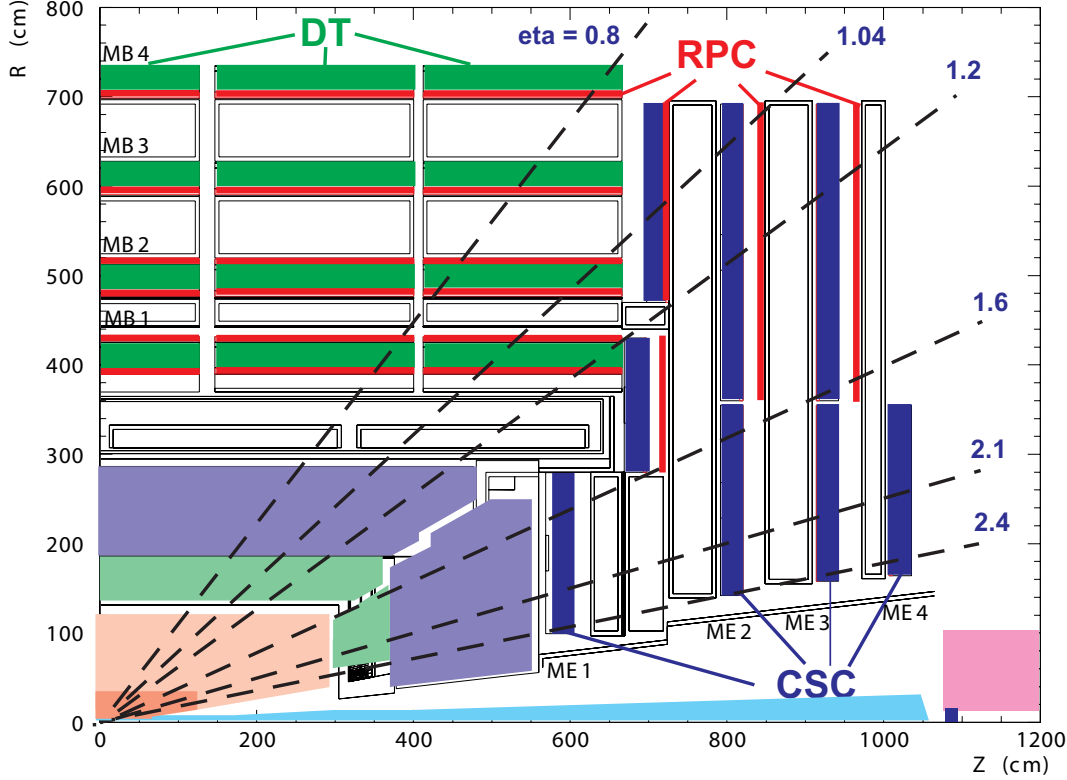


Figure 5.11: A schematic view of one quadrant of the muon system [74].

The drift tubes (DT) are installed in the barrel consisting of four layers. The basic element of DT is the drift cell with a size of $42 \times 33 \text{ mm}^2$. The gas is composed of a mixture of 85% Ar and 15% of CO_2 which provides a saturated drift velocity of about $55 \mu\text{m}$ with a maximum drift time of up to 380 ns. The downside is that this latency makes DT too slow to trigger events in high luminosity.

The cathode strip chambers (CSC) are in the endcaps, covering the region $0.9 < |\eta| < 2.4$. The CSCs are arranged in four disks. They are standard multi-wire proportional chambers, and use a gas mixture of 50% CO_2 , 40% Ar and 10% CF_4 placed in a strong electric field. They provide a precise position measurement in the dimensions $r - \phi$. This detector type has a fast response and a high radiation resistance suited for the high muon flux in the forward region.

Finally, the resistive plate chambers (RPC), installed in both the barrel and the endcaps region, are double-gap chambers filled with a gas mixture of 96.2% R134a ($\text{C}_2\text{H}_2\text{F}_4$), 3.5% iso-butane and 0.3% SF_6 and lined with resistive bakelite plates forming a high voltage capacitor. They are operated in avalanche mode to ensure reliable operation at high rates and serve practically as a triggering system.

5.3.7 Trigger and data acquisition system

The LHC bunch crossing frequency is 40 MHz for a bunch spacing of 25 ns. CMS events have size of 1 MByte, leading to a rate of 40 GByte/s. It is not practical to save this amount of data and a pre-selection of interesting events is necessary. The trigger system is introduced to reduce the rate. In CMS, the event rate is reduced in a two-step approach: the level-1 (L1) trigger, which is hardware-based, filters the rate to 100 kHz; the software-based high-level trigger (HLT) further decreases the rate to 600 Hz for storage. The pre-selection is achieved by deploying a trigger menu, which is modified according to the LHC running condition (e.g. instantaneous luminosity) to reach maximum output. The trigger menu is updated constantly using either tighter selections or pre-scaling, where the trigger only accepts every N th event, where N is being the pre-scale factor.

The L1 Trigger is mainly implemented in Field Programmable Gate Array (FPGA) technology, while Application Specific Integrated Circuits (ASIC) and programmable memory lookup tables (LUT) are also used in special cases to satisfy the need for speed, density or radiation resistance. The L1 trigger decision is based on information from calorimeter and the muon system which can be organized in local, regional and global components. Figure 5.12 illustrates the structure of L1 trigger system. The first step of calorimeter trigger is to use trigger primitive generators (TPG) that are constructed from ECAL and HCAL energy deposits above given threshold as seeds. The regional calorimeter trigger then combines the information from TPG by applying basic quality cuts and define the electron or photon candidates, tau veto bits, transverse energy sums E_T and other interesting information for muons. The third step, the global calorimeter trigger (GCT) determines the total transverse missing energy $\cancel{E}_T$, the jet multiplicity and the scalar sum of energy of jets (H_T). The muon trigger exploits the information from the three muon subdetectors DT, CSC and RPC. Firstly, DT and CSC provide information on the track segments and hit patterns locally. After that, complete tracks are reconstructed regionally. RPC contributes regionally because of its better timing resolution by reconstructing its own track candidates. Finally, the global muon trigger combines all regional information from the three subdetectors.

The software-based HLT trigger is performed by a computing farm. It receives the event from L1 trigger and performs a fast event reconstruction exploiting the full detector information to decide whether the events should be kept for storage.

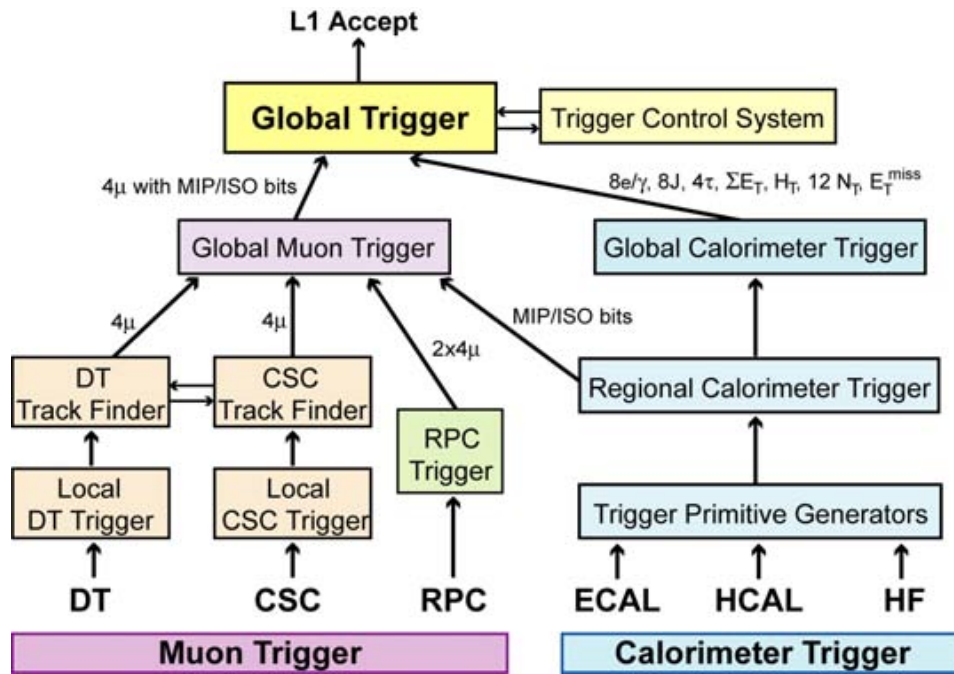


Figure 5.12: Architecture of the CMS level-1 (L1) trigger system [65].

CHAPTER

6

EVENT SIMULATION AND RECONSTRUCTION

Contents

6.1	Event Generation	72
6.1.1	Monte Carlo method	72
6.1.2	Structure of an event	72
6.1.3	Hard scattering	74
6.1.4	Parton showering	75
6.1.5	Hadronization	75
6.1.6	Underlying event	75
6.1.7	Event generators	76
6.1.8	Detector simulation	77
6.2	Object Reconstruction	77
6.2.1	Particle-flow Algorithm	78
6.2.2	Primary vertex	78
6.2.3	Electrons	80
6.2.4	Muons	82
6.2.5	Jets	84
6.2.6	B-tagging	85
6.2.7	Missing transverse energy $\vec{\cancel{E}}_T$	86

In experimental high energy physics, Monte Carlo simulated events are used to model various SM and BSM processes. The generated events will go through the full CMS detector simulation, allowing direct comparisons between data and theoretical models. In this chapter,

how Monte Carlo events are generated is presented in the first part. The second part of this chapter describes the reconstruction the physics objects from the signatures left inside the detector of CMS.

6.1 Event Generation

6.1.1 Monte Carlo method

The nature of quantum mechanics is probabilistic and not deterministic. Calculations can only provide the occurring probability of certain outcomes. High energy proton-proton collisions at the LHC are complex processes due to the substructure of proton and the large amount of final state particles. These characteristics cannot be fully described analytically from first principles. Based on these two facts, a stochastic technique, referred to as Monte Carlo (MC), where pseudo-random numbers are generated according to the probability density functions of physics models, is used for event simulation. Large numbers of events are generated, each event being characterized by unique kinematics encapsulated into its four-momentum of generated particles and their types.

6.1.2 Structure of an event

Based on the theory of factorization [77], simulation of events can be broken down into the following steps: the hard-scattering process, parton showers, the hadronization of quark and gluon jets, the underlying event and the decay to long lived particles. Figure 6.1 illustrates these steps. Firstly, a collision happens between two hadrons. Each hadron is a set of partons. Then one parton from a hadron collides with another parton from the other hadron. This collision constitutes the hard scattering process, illustrated as the big red dot in the middle of Figure 6.1. The momentum of the incoming partons are randomly selected by sampling from the input parton distribution functions (PDFs), which provide the probability that a parton in a proton carries a certain momentum fraction of the proton, at an energy scale Q . The hard scattering calculation can be computed using matrix element formalism in perturbative QCD at leading order (LO) or next-to-LO (NLO) precision depending on the requirements.

Secondly, after the hard scattering process calculation, higher order QCD-effects, shown as blue curly and straight lines in Figure 6.1 are modeled. This is the so-called parton showering process. It describes the initial state radiation (ISR) and final state radiation (FSR) from the partons. Quarks or gluons are radiated and will be further split into gluon pairs or quark-antiquark pairs. The process keeps going until the minimum energy threshold, the hadronization scale about 1 GeV, has been reached, yielding to non-perturbative interactions.

Thirdly, the hadronization of the partons is simulated, illustrated as green spots in Figure 6.1. The process leads to the formation of the final state particles. The perturbative QCD calculation is not valid in this energy scale anymore. Thus, purely phenomenological models that were tuned to previous data are used [77].

Finally, the decay of short-lived hadron and tau lepton is simulated. Once the whole event is generated, it will be the input to the detector simulation with the full information of the detector components, including the geometry, the magnetic field, the interaction of the particles with the detector material and the electronics.

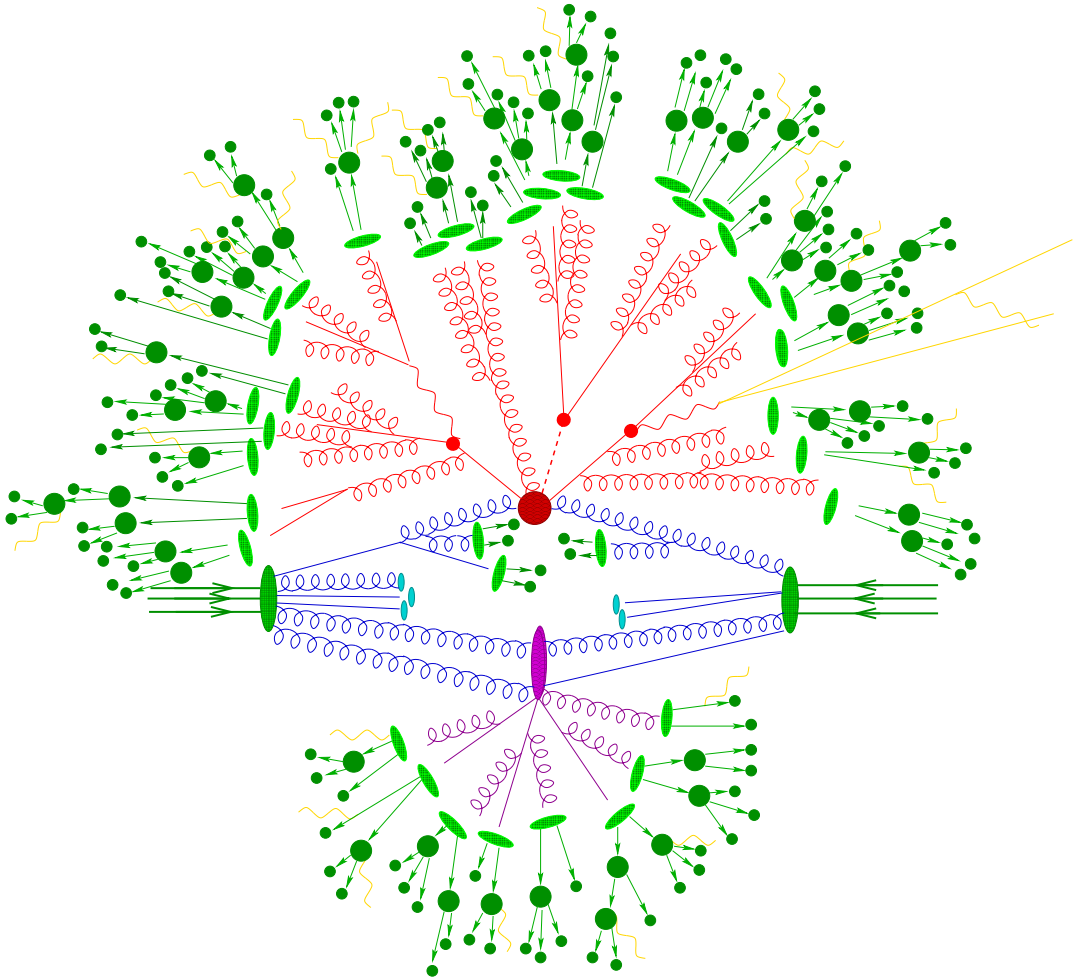


Figure 6.1: Illustration of a proton-proton collision event simulated in a MC event generator [78]. The hard scattering process is shown as red spot, and the purple spot represents the underlying event. Red curly and straight line depict the parton shower. The green spots are the hadronization, while the final hadrons are dark green.

6.1.3 Hard scattering

The cross section of a hard scattering process $p_1 p_2 \rightarrow n$ at a hadron collider, can be written as [79]:

$$\begin{aligned}\sigma &= \sum_{p_1, p_2} \int \int \int dx_1 dx_2 f_{p_1}^{h_1}(x_1, \mu_F^2) f_{p_2}^{h_2}(x_2, \mu_F^2) d\hat{\sigma}_{p_1 p_2 \rightarrow n}(\mu_F, \mu_R) \\ &= \sum_{p_1, p_2} \int \int \int d\Phi_n f_{p_1}^{h_1}(x_1, \mu_F^2) f_{p_2}^{h_2}(x_2, \mu_F^2) \frac{1}{2\hat{s}} |\mathcal{M}_{p_1 p_2 \rightarrow n}|^2(\Phi_n; \mu_F, \mu_R),\end{aligned}\tag{6.1}$$

where:

- μ_F is the factorization energy scale. It is the cut-off value where perturbative QCD fails, thus non-perturbative models are needed to describe the region below that boundary. μ_R is the renormalization scale introduced to save the ultra-violet divergency of the theory. Quite often, the same choice is made for both factorization scale and renormalization scale, meaning $\mu_F^2 = \mu_R^2 = Q^2$, where Q stands for the energy scale of hard scattering.
- $f_p^h(x, \mu)$ are the parton distribution functions (PDFs) as a function of momentum fraction x of parton p with respect to its parent hadron h , and the factorization energy scale of hard scattering μ_F . The evolution of the PDF $f_p(x, Q^2)$ is given by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation [80–82]:

$$Q^2 \frac{\partial f_{p_1}(x_{p_1}, Q^2)}{\partial Q^2} = \sum_j \int_x^1 \frac{dz}{z} \frac{\alpha_s}{2\pi} P_{p_1, p_2}(z, \alpha_s) f_{p_2}(x_{p_1}/z, Q^2),\tag{6.2}$$

where the $P_{p_1, p_2}(z, \alpha_s)$ is the splitting function describing the probability that a parton p_1 radiates a quark or a gluon and become another parton p_2 carrying a momentum fraction z , j is the flavor index (quark or gluon) and α_s is the strong coupling constant. For the CMS Monte Carlo production in LHC run 2, the NNPDF3.0 PDF [83] set is used for most of the generation of the events used for the analysis in this thesis.

- $\hat{\sigma}_{p_1 p_2 \rightarrow n}$ is the parton level cross section for the process $ab \rightarrow n$. It relies on the momentum final phase space Φ_n , on μ_F and renormalization scale μ_R .
- $1/(2\hat{s})$ is the parton flux, where $1/(2\hat{s}) = 1/(2x_1 x_2 s)$, where s is the center-of-mass energy squared.
- $|\mathcal{M}_{p_1 p_2 \rightarrow n}|^2(\Phi_n; \mu_F, \mu_R)$ is the matrix element squared. It could be treated as a sum over all different Feynman diagrams. In the actual case of event simulation, the calculation of matrix element can be achieved by Monte Carlo generator using pre-computed matrix element hardcoded in the program or analytical computation.
- $d\Phi_n$ is the differential phase space elements over the n final state particles,

$$d\Phi_n = \prod_{i=1}^n \frac{d^3 p_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^{(4)}(p_{p_1} + p_{p_2} - \sum_{i=1}^n p_i),\tag{6.3}$$

where p_{p_1} and p_{p_2} are the momentum of initial particles.

This equation is valid for all orders in perturbative theory.

6.1.4 Parton showering

Partons in the initial and final states can produce QCD and QED radiation. The so-called parton showering is the branching of the following processes: $q \rightarrow qg$, $q \rightarrow q\gamma$, $g \rightarrow gg$, $g \rightarrow q\bar{q}$, $\gamma \rightarrow q\bar{q}$, $l \rightarrow l\gamma$ and $\gamma \rightarrow l\bar{l}$. All these processes are generated as long as they are energetically allowed.

The parton showering is the bridge between the hard scattering and the hadronization. It is the higher-order correction to the matrix element calculation. The complete branching process computation using perturbative QCD as the hard scattering is technically impossible to achieve here. Therefore, the parton shower simulation models the probability of additional emissions.

The parton branching in the parton shower is ordered by the evolution scale Λ . In PYTHIA8, the scale is defined as $\Lambda = p_T^2$, where p_T is the radiating parton transverse momentum. The parton showering branching starts from the user-given scale $\Lambda_{\max}$, which usually is the hard scattering scale. The chance that a parton does not branch between the two scales $\Lambda_{\max}$ and Λ_1 , where $\Lambda_{\max} > \Lambda_1$, is described by the Sudakov form factor. By solving the Sudakov form factor [77], the scale Λ_1 of the new branching can be found and the corresponding branching is generated. The parton showering continues until the cut-off scale $\Lambda_{\text{cut-off}}$ is reached. Below this scale, perturbation theory is not valid.

6.1.5 Hadronization

When the parton showering reaches the $\Lambda_{\text{cut-off}}$ scale, approximately 1 GeV, the evolution of the event enters the non-perturbative region. Hadrons start to be formed as a consequence of color confinement. This is the hadronization process and it has to be described by non-perturbative phenomenological model.

PYTHIA8 implements the Lund string fragmentation hadronization model [84]. The idea is that, in QCD, a linear confinement is expected at large distance. Thus, one can define a linear potential between the two partons, $V(r) = \kappa r$, where r is the distance between two partons and the string constant κ is 1 GeV/fm. In left plot of Figure 6.2, a physical picture is shown for the stretching of the string between a quark and an anti-quark. As the distance between the quark and the anti-quark increases, the potential energy increases. The string breaks and a new pair of quark and anti-quark is created, as shown in right plot of Figure 6.2. The hadron production continues as long as the string has enough energy.

Given the fact that it uses many non-perturbative assumptions, the hadronization model needs to be tuned from the previous experiment data.

6.1.6 Underlying event

The underlying event includes the contribution of multiple parton interactions (MPI) and beam-beam remnants (BBR). In MPI, the contribution is coming from the remaining parts of the protons that did not take part in the hard scattering interaction, In BBR, the contribution is emerging from the hadronization of the partonic scattering remnants. The

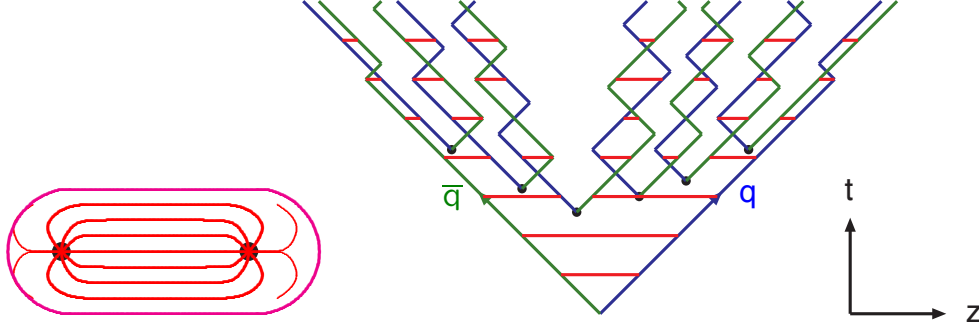


Figure 6.2: **Left:** The potential between a pair of quark and anti-quark. **Right:** The evolution of the string in the coordinate of space and time. The hadrons are produced with an increase in distance. From [77].

existence of underlying event leads to enhancement of the jet multiplicity in the event. Thus, the model of jet activity by the parton shower simulation could also be affected by the underlying event modeling. The model of underlying event has many free phenomenological parameters, therefore, it requires experimental tuning.

6.1.7 Event generators

There are many event generators developed. In the following, those event generators employed to simulate the samples used in the analyses presented in this thesis are discussed:

- **Pythia8**

PYTHIA8 [85] is a leading order multi-purpose event generator, suitable to be used for a wide range of processes, including SM and BSM physics. It can be used for the whole event generation, including all the steps from hard scattering calculation down to parton showering and hadronization simulation. The hard scattering calculation is limited to leading order tree level $2 \rightarrow 1$ and $2 \rightarrow 2$ processes. Jets in final states and higher orders are simulated with parton showering as an approximate approach. The hadronization is based on the Lund string model [86]. The hadronization and underlying event tuning for PYTHIA8 is the CUETP8M1 tune [87]. PYTHIA8 also provide the SLHA interface [88] allowing communication with other programs.

- **MadGraph5_aMC@NLO**

MADGRAPH5 [89] is the most popular event generator in the CMS experiment. It can calculate the matrix elements from tree-level to any given order in perturbation theory. At the same time, generator level cuts can be introduced to constrain the phase space needed by the experimental analysis. Differently from PYTHIA8, the additional parton radiation in ISR and FSR is also calculated on matrix element level. A minimum p_T threshold is set to prevent divergent soft-gluon radiation. The event topology has better modeling in MADGRAPH5, however, the cross section always shows a strong scale

dependence. Usually, the cross section from SM processes is normalized to the higher-order prediction from the MCFM generator. Parton showering and hadronization are not included in MADGRAPH5 and one needs to interface it with PYTHIA8.

AMC@NLO [90] is served as additional package with MADGRAPH5 to combine the NLO matrix element with the parton shower. It gives a more accurate description with reduced uncertainties. The drawback is that there are certain amounts of events with negative weights because of the subtraction for double counting. The MADSPIN package [91] can be used to interface AMC@NLO to model the decay products of resonances and preserve the spin correlation information.

- **PowHegV2**

The Positive Weight Hardest Emission Generation Version 2 (POWHEGV2) [92] can calculate hard scattering at NLO, optimized for heavy quark production. It is used to solve the problem of events generated with negative weight by AMC@NLO. The basic idea in POWHEGV2 is to generate the hardest radiation first, and then feed the event to any shower generator for subsequent softer radiation. It uses the multiplicative scheme to match parton shower with matrix element and only have one additional radiation is allowed. Thus it results in nearly only positive weights.

- **TAUOLA**

TAUOLA [93] is a package designed for the simulation of τ lepton decay. It is interfaced to PYTHIA8 and takes spin information and QED corrections into account.

6.1.8 Detector simulation

After modeling the different stages of an event depicted as in figure 6.1, the interactions of the stable particles with the CMS detector are simulated in a subsequent step. The detector simulation results in the transition from generator level to detector level. A toolkit called GEometry ANd Tracking version 4(GEANT4) [94] is used to perform the full simulation of the interaction of particles with the detector materials, the effects of the magnetic field on the detector response, the conditions of the real detector operation, performance of the electronic readout and non-operating subsystems. Finally, the trigger emulation is added and the output has the same format as the CMS data acquisition quality system, allowing direct comparison between the simulated events and real collision events.

6.2 Object Reconstruction

Physics analyses make use of the identification of particles from hadron collisions. The particles travel through the subdetectors of CMS and leave typical signatures. From these signatures, particle-flow candidates are reconstructed using the particle-flow algorithm. Further high level physics objects, such as electrons, muons, jets or hadronically decaying taus, are reconstructed based on top of particle-flow candidates.

6.2.1 Particle-flow Algorithm

The idea behind the particle-flow (PF) algorithm [95] is to combine the information collected from the different CMS subdetectors in the most efficient way to define the four momentum and the particle types involved in each event. This algorithm is an improved version of energy flow algorithm [96], which was initially used at the ALEPH experiment at LEP.

The algorithm is divided in four major steps. The procedure starts with the reconstruction of charged particle tracks from the set of hits observed in the inner tracker by using an iterative tracking approach based on a Kalman filter [97]. The approach sets a very tight cuts at the beginning and then loosen the cuts in further iterations. In the end, a high efficiency is obtained and the misidentification rate remains on a low level.

The second step involves clustering the energy deposits in the ECAL and HCAL. This method starts with the calorimeter cells with an energy above a threshold acted as seeds. Topological clusters are being merged from adjacent cells also with a minimum energy threshold.

The third step is taking the linking between tracks and clusters of energy in the ECAL and HCAL into account. A linking algorithm is being processed to fully reconstruct the particle candidates while rejecting possible double counting of the signatures in different subdetectors. A schematic view of the linking algorithm is shown in figure 6.3. An energetic π^- is identified by linking a short track (T_1) and a large energy deposits from the ECAL and HCAL (E_1 and H_1). Two photons from a π_0 decay are reconstructed with absence of tracks and two nearby ECAL clusters (E_2 and E_3) linked with a HCAL deposits (H_2). In total, most of the neutral and charged hadrons and photons PF candidates are reconstructed by linking algorithm.

In the final fourth step, muon PF candidates are built from previously reconstructed muons. After that, electron candidates come next, along with the possible bremsstrahlung photons associated to them. The energy cluster of electron are then removed from the set of total clusters. The remaining unassociated ECAL clusters are identified as photons. Figure 6.3 shows an isolated ECAL cluster E_4 , signature of the decay products of K_L^0 ($K_L^0 \rightarrow 3\pi_0 \rightarrow$ photons). The remaining tracks ending in the calorimeters are assumed to be charged hadrons. Again, figure 6.3 shows that a π^+ is identified from its reconstructed track T_2 with the absence of ECAL and HCAL energy deposits. Finally, unassociated HCAL energy clusters are assumed to be neutral hadrons.

To summarize, muon, electron, photon, neutral and charged hadron PF candidates are built by the PF algorithm as shown in 6.4.

6.2.2 Primary vertex

The primary vertex is defined as the position in space where the hard scattering of the most energetic partons happen. It is of high interest in an event. However, at the LHC, due to the large number of protons per bunch, the interested hard interaction point is surrounded by trivial multiple interactions, referred to as pile-up (PU). In other words, we must distinguish the primary vertex point from pile-up interactions. The number of pile-up interactions, shown in Figure 6.5, was on average 23 during the 2016 running conditions.

The primary vertex reconstruction [99] can be divided into two stages. First, the deterministic annealing algorithm [100] is used to cluster a set of selected tracks to primary vertex candidates, where each candidate should be separated by a minimum distance from its closest

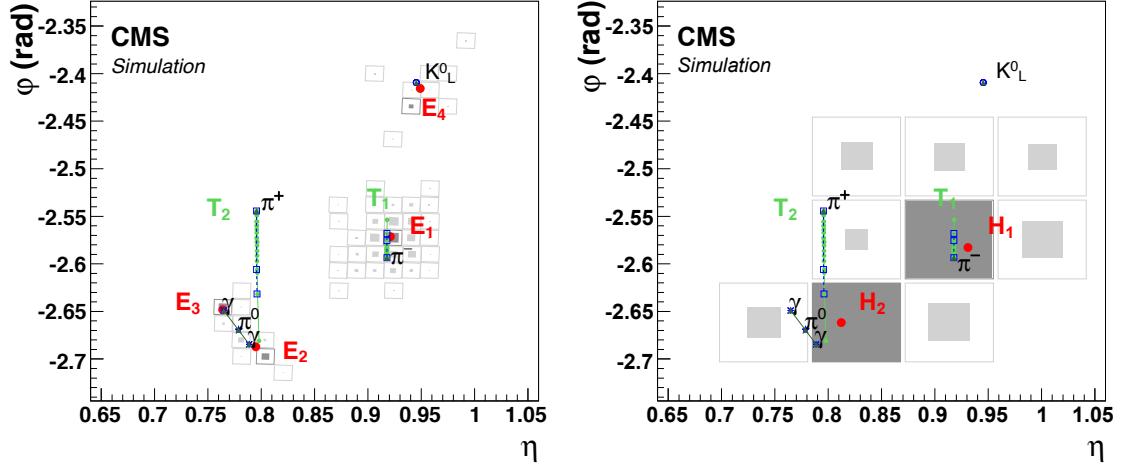


Figure 6.3: A graphical view of an event with five particles in the η - ϕ plane. **Left:** View on the ECAL surface. **Right:** View on the HCAL surface. Tracks (in green), energy deposits (in red) and cells (in grey block) are shown. From [95].

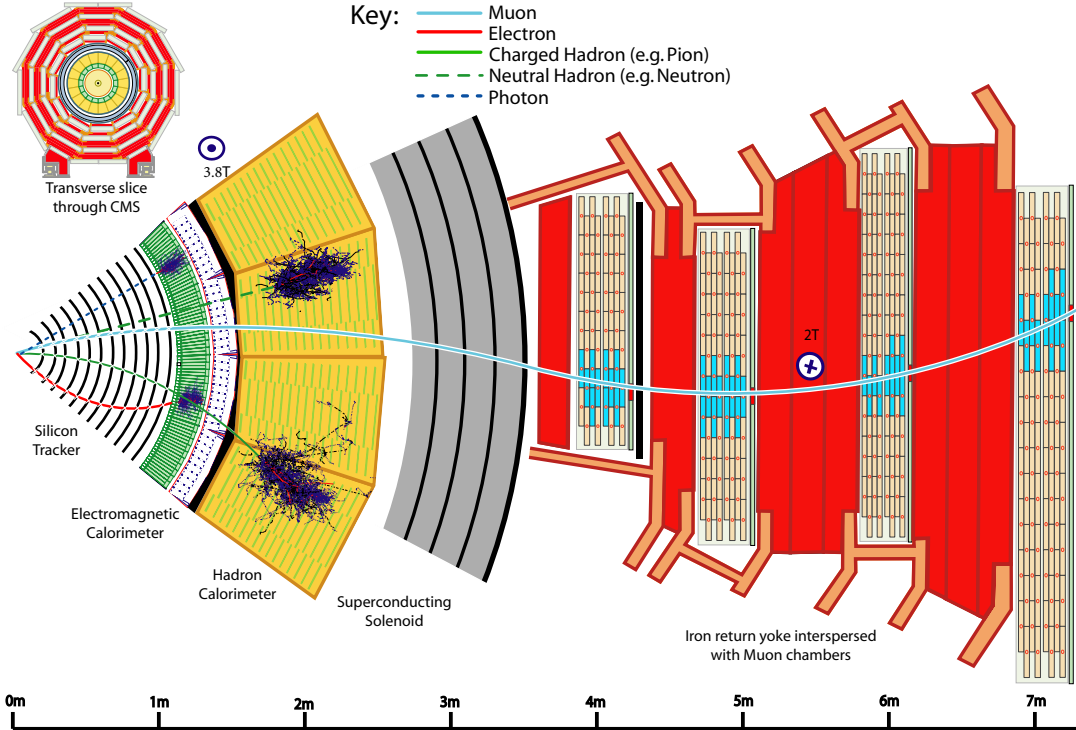


Figure 6.4: Illustrations of signatures from different particle flow candidates [95].

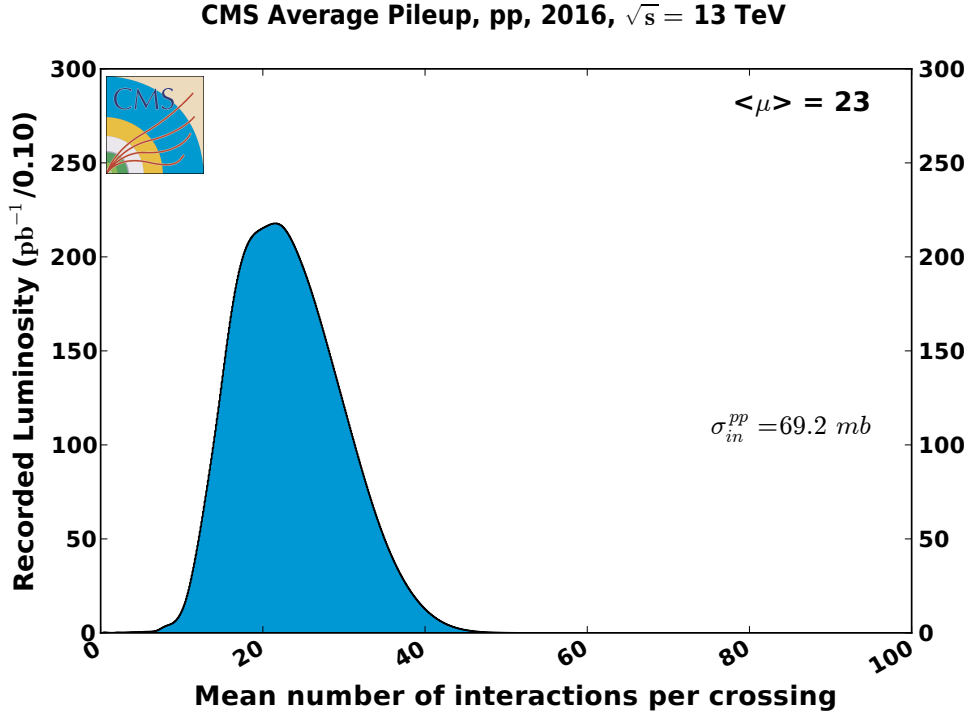


Figure 6.5: The mean number of interactions per bunch crossing in 2016 [98].

neighbor. In the second stage, a weighted least square method adaptive vertex fitting [101] is used to provide robust fit of the vertex position. Quality cuts are applied after the fit to make sure the vertices are close to the detector center and beam spot, and there must be more than four degrees of freedom in the vertex fit. Finally, the primary vertex is assumed to be the vertex with the largest value of the p_T^2 sum of tracks.

6.2.3 Electrons

The signature of an electron in CMS consists of track in the tracker and energy clusters in the ECAL. Due to the relative small mass, electrons suffer from large changes in the direction and losses of energy when scattering against detector materials and radiating bremsstrahlung photons. In particular, 35% of electrons with energy of 10 GeV lose more than 70% of their energy because of bremsstrahlung, and in 10% of the cases the loss reaches more than 95%. In order to have a correct reconstruction of electrons, one has to combine the energy clusters of the electron with the energy clusters of the bremsstrahlung photons in the ECAL. Furthermore, the energy clusters in the ECAL are required to be matched with the electron track in the tracker [102].

Two different algorithms, the hybrid algorithm in the barrel, and the multi- 5×5 algorithm in the endcaps, are used to cluster the energy in the ECAL, exploiting different advantages from different subdetector geometries. The hybrid algorithm makes use of the geometry of the ECAL barrel and properties of the shower shape, collecting the energy in a smaller window in η and larger window in ϕ . It starts from seed crystals with an energy threshold of 1 GeV. Arrays of 5×1 crystals around the seed with an energy threshold of 0.1 GeV are

added to the seed to form a crystal array in the $\eta \times \phi$ plane. A supercluster is then built from the arrays with the seed array with energy of at least 0.35 GeV. In the endcaps, the multi- 5×5 algorithm starts from seed crystals with an energy threshold of 0.18 GeV. Around these seeds and beginning with the crystal with the largest energy, the energy is collected in a cluster of 5×5 crystals. These clusters then form a supercluster not larger than 0.3 in the ϕ -direction and 0.07 in the η -direction. The energy-weighted position of the supercluster are then extrapolated to the planes of the preshower to find the energy deposit in the preshower detector. This energy is further added to the supercluster.

The electron track can be reconstructed by using the Kalman filter method as any charged particle track. However, due to the large energy loss of electron in the tracker detector, a Gaussian Sum Filter (GSF) algorithm [102] is developed to identify the electron track. The electron track reconstruction has two stages: seeding and tracking. The seeding step is using a two-way approach: seeding from the energy cluster in the ECAL and seeding from hits in the tracker. The ECAL-based seeds start from the supercluster and then extrapolate the trajectory to the first layer of the tracker. The tracker-based seeding starts from the hits in the tracker using Kalman filter method and then matches the tracks to the superclusters in the ECAL. The electron seeds built using ECAL-based and tracker-based method are combined. The tracking step make use of the electron seeds. The track building is based on the Kalman filter method. The energy loss of electron caused by the bremsstrahlung in the tracker is not following a Gaussian distribution assumed by the Kalman filter but a Bethe-Heitler distribution, which has a larger tails. A GSF fitting is performed to estimate the track parameters.

An association between ECAL supercluster and electron GSF track is performed. In order to resolve ambiguity, further matching between the electron reconstructed using Kalman filter method and GSF electron is required.

To separate electrons from hadrons that mimic the electron signatures, a MVA approach based on the BDT method, as described in section 4.5, is used. There is no single variable that can separate real electrons from fake electrons. A group of variables are used to build a single MVA classifier. They can be classified into three different sources: the observables that measure the association between ECAL clusters and tracks, such as:

- $\Delta\eta_{\text{out}}$, the distance in η between the super cluster and the track at the ECAL surface;
- $\Delta\eta_{\text{in}}$ and $\Delta\phi_{\text{in}}$, the distances in η and ϕ between the ECAL supercluster and the associated track at the position of the primary vertex;

the observables related to the track reconstruction only, such as:

- The number of valid hits in the track fit, the χ^2 of the Kalman filter fit and the χ^2 of the GSF fit;
- The number of GSF track hits, the number of expected missing inner hits and the result of the conversion vertex fit;

and the observables related to the ECAL information only, such as:

- The cluster shape variables $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$. The two variables are defined:

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_{i \in 5 \times 5} w_i (\eta_i - \langle \eta \rangle_{5 \times 5})^2}{\sum_{i \in 5 \times 5} w_i}} \quad \text{and} \quad \sigma_{i\phi i\phi} = \sqrt{\frac{\sum_{i \in 5 \times 5} w_i (\phi_i - \langle \phi \rangle_{5 \times 5})^2}{\sum_{i \in 5 \times 5} w_i}}, \quad (6.4)$$

where the $\langle\eta\rangle_{5\times 5}$ and $\langle\phi\rangle_{5\times 5}$ are the barycenter in η and ϕ of the supercluster in the 5×5 array, i is the index referred to the crystals of array and the weights w_i is defined as $w_i = \max(0, 4.7 + \ln \frac{E_i}{E_{5\times 5}})$;

- The "R9" variable which is defined as $R_9 = E_{3\times 3}/E_{\text{SC}}$, where $E_{3\times 3}$ is the energy collected in a 3×3 crystal arrays, centered on the most energetic crystal, and E_{SC} is the energy of the supercluster;
- The ratio of the HCAL energy deposits laying behind the ECAL cluster to the ECAL supercluster energy;
- The ratio of the ECAL supercluster energy to the momentum of track associated with the electron;
- The variable exploiting the loss of the electron energy loss, defined as $1/E_{\text{SC}} - 1/P_{\text{GSF}}$, where P_{GSF} is the momentum of the electron track.

One can define the electron relative isolation variable to further separate prompt electrons from electrons produced in jets. The isolation is based on photon, neutral and charged hadron particle flow candidates within a cone size $\Delta R < 0.3$ of the electron direction. It is defined as:

$$I = \frac{\sum_{\text{charged}} p_T + \max\left(0, \sum_{\text{neutral}} p_T - \frac{1}{2} \sum_{\text{charged,PU}} p_T\right)}{p_T}. \quad (6.5)$$

In the expression, $\sum_{\text{charged}} p_T$ is the scalar sum of the transverse momenta of the charged particles originating from the primary vertex. The sum $\sum_{\text{neutral}} p_T$ is a similar quantity for neutral particles. The last term in the numerator is the $\delta\beta$ corrections used to subtract pileup contamination. The neutral contributions from pileup vertices are estimated as a half of the p_T sum of charged particles from pileup vertices inside the isolation cone.

6.2.4 Muons

Muons are much heavier than electrons. Therefore, the energy loss is far smaller than electrons. At the energy relevant to CMS, muons are minimum ionizing particles and can travel through the whole detector. The muon deposits a relatively small amount of energy, approximately 2 GeV [74], in the calorimeters.

The muon reconstruction is performed before running the PF algorithm. Muons can be reconstructed in two ways [103]. In the inside-out approach, reconstruction starts from a track in the tracker. When its trajectory is extended to the outer muon systems, a matching track is found there. In the outside-in approach, it can be reconstructed starting from a track in the muon system and then propagate its trajectory to the inner tracker to locate a matching track. Three types of muon candidates can be reconstructed:

- **Standalone muons** are built using the ionizing signal information from muon systems only, namely DT, CSC and RPC, independently from inner tracker. In a DT cell, hit reconstruction specifies the transverse distance between the wire and the intersection of the muon trajectory with the plane containing the wires in the layer. In a CSC layer, the hit reconstruction measures the position of the traversing muon by combining information from the cathode strips and anode wires. In an RPC, the hit reconstruction

requires clustering of hit strips. Segment can be built from hits in CSC or DT because they are multi-layer subdetector. A standalone muon is reconstructed using a Kalman filter method to gather all information from DT, CSC and RPC. The method starts from the seeding of segments in CSC or DT. Standalone muons can be used in detector studies and served as ingredient to reconstruct global muons.

- **Tracker muons** are built using inside-out approach. The tracker track propagates to the muon system with loose matching to CSC or DT segment. The matching should fulfill one of the following requirements: the absolute value of the difference between their positions in the x coordinate is smaller than 3 cm, or the ratio of this distance to its uncertainty is smaller than 4. The tracker muons have good purity, good momentum resolution and are efficient for low p_T muons.
- **Global muons** are built using outside-in approach. The standalone muon is matched to a tracker track. The matching is done by comparing the parameters of the two tracks propagated onto a common surface. The Kalman filter method is used with information from the tracker track and standalone muon. The global muon reconstruction is efficient for muons leaving hits in several muon chambers and good momentum resolution for muons with $p_T > 200$ GeV.

One can define different muon identification criteria to be used in physics analyses allowing them to tune the balance between purity and efficiency. Differently from the electron identification, the muon identification is cut-based, ie. having cuts on various related variables. The main types of identification include:

- **Loose working point**, defined as a muon selected by PF algorithms, and to be either a global or a tracker muon, aiming to identify prompt muons and muons from decays (such as from the τ lepton) while keeping the misidentification rate low.
- **Medium working point**, is the one used in the analyses presented in later chapters. A medium muon should be a loose muon. If the muon is a tracker muon, the muon segment compatibility probability has to be larger than 0.451. If the muon is both reconstructed as tracker muon and global muon, the segment compatibility has to be greater than 0.303, but the normalized χ^2 per degree of freedom of their global fit is less than 3, the position match between the tracker muon and standalone muon must satisfy $\chi^2 < 12$, and the χ^2 of the kink finding algorithm, which was designed to drop muons from decays in flight, should be less than 20.
- **Tight working point**, having the following requirements: the muon should be a loose muon having six hits in the tracking and at least one hit from pixel detector. The muon should be both a global muon and a tracker muon. Segments matching should be found in two stations for tracker muon. The χ^2 of the global muon fit is required to be less than 10 and have at least one hit from the muon system. The impact parameters of the muon should satisfy $|d_{xy}| < 0.2$ cm and $|d_z| < 0.5$ cm.

The relative isolation for muons to further suppress misidentification rate is defined exactly in the same way as for the electrons in equation 6.5.

6.2.5 Jets

Jets are bundles of collimated hadrons, and are experimental signatures of energetic quarks and gluons produced in high energy collisions. Different jet types can be reconstructed using subdetector information and their combination [104]. In simulated samples, generator level particles can be clustered to form generated jets ("GEN jets"). For reconstructed objects, "Calo jets" make use of calorimeter information only. "Track jets" use the information from inner tracker. "Jet-plus-track jets" are formed by combining "Calo jets" and "Track jets". Finally, the jets with best performance in CMS, "PF jets" are clustered on the PF candidates. The most important jet parameter is the jet energy. Phase space dependent jet energy corrections are further required to be applied after reconstruction to take into account of different effects, such as underlying event and pileup.

The type of jets used in analyses presented in this thesis, is the PF jet. The anti- k_T algorithm [105, 106] implemented by FASTJET package is a sequential clustering trying to combine nearby PF candidate pairs. The goal of this algorithm is to be both infrared-safe and collinear-safe, meaning the reconstruction result should not be different when trying to add another new soft radiated parton or a new splitting of a close-by parton. The algorithm starts the seeding from the most energetic PF candidates, searching all the PF candidates within the cone size $\Delta R < 0.4$ to calculate the following distance variables:

$$d_{ij} = \min\left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2}\right) \frac{\Delta R_{ij}^2}{R^2}, \quad (6.6)$$

$$d_{iB} = \frac{1}{p_{Ti}^2}. \quad (6.7)$$

In the above equations, d_{ij} is the distance between the PF candidates i and j , d_{iB} is the distance between the i PF candidate and the beam axis, $\Delta R_{ij}^2 = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$ where y is the rapidity, ϕ is the azimuthal angle, p_T is the transverse momentum and R is the cone size of 0.4. The algorithm then starts combining the PF candidate with the smallest d_{ij} or d_{iB} variable. The energy and position of the combined product will be recalculated. The process repeats until the smallest is the d_{iB} variable. The combined product i is assumed to be a final state jet and is removed from the list. The algorithm starts the next iteration until all PF candidates are clustered into jets. An example of jet clustering is illustrated in Figure 6.6.

Jet energy correction is following the jet reconstruction [104] step. The goal of this correction is to have reconstructed jet energy matching to the energy of the generated jet. The correction is factorized in several stages as shown in figure 6.7. The correction use the reconstructed jet collection as input. In the first stage, denoted as the pileup offset corrections, charged particles not from the primary vertex are removed. This method, referred to as charged hadron subtraction (CHS), can significantly reduce the effect of pileup. The pileup offset corrections are determined both before and after CHS, with parameterization as a function of offset energy density ρ , jet area A and jet pseudorapidity η and p_T . Corrections for residual differences between data and simulation as a function of η are determined using the random cone method in zero-bias events. The jet energy from PU and electronic noise are then removed. The second stage corrects the response as a function of η and p_T using correction determined from simulation studies. The third and fourth stages correct the residual

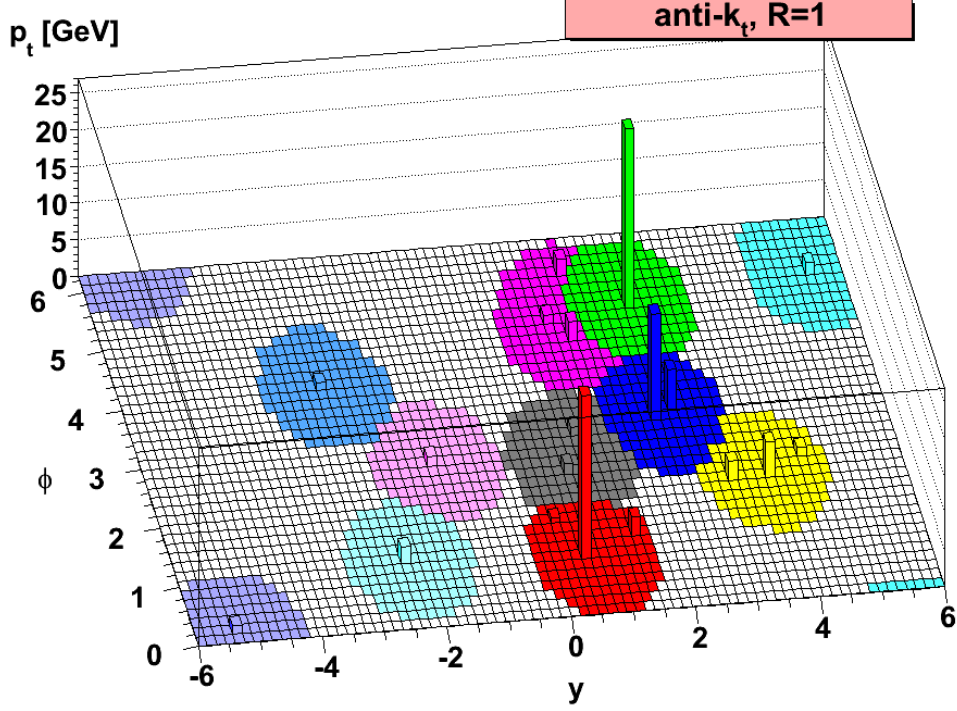


Figure 6.6: A schematic view of how anti- k_T algorithm perform jet clustering [106]. Each color region corresponds to a reconstructed jet.

differences as a function of η and p_T to the data response using correction determined from $Z/\gamma^* + \text{jets}$ events. The final stage corrects the response for different jet flavors.

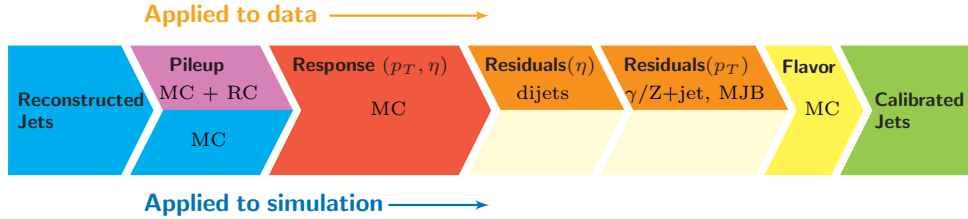


Figure 6.7: Jet energy correction procedure. Stages marked with MC are derived from simulation studies, RC means for random cone, and MJB stands for the analysis of multijet events. From [104].

6.2.6 B-tagging

B quark jets, as a results of the b quark hadronization, can be distinguished from the jets originating from gluons and light quarks (u, d, s) and c quark because b quark have a relatively large mass, long lifetime (about 10^{-12} s) and high momentum decay products. The longer lifetime results in a secondary vertex presented in a event. The high momentum decay

products lead to wider jets compared to the jets from light quarks and gluon. B-tagging algorithms can be built according to these two signatures.

The combined secondary vertex version 2 (CSVv2) algorithm [107] is used in the analyses of this thesis. It starts from reconstructing a secondary vertex and a "pseudo vertex". The reconstruction of the second vertex uses the input of displaced tracks. The following quality requirements of tracks should be fulfilled: the p_T of the track should be above 1 GeV and the χ^2/dof of the Kalman filter fit should be less than 5, the track should have at least 8 hits, where two of them should be from the pixel detector. The most discriminating variable is the impact parameter of the track defined as the distance from the track to the primary vertex at the point of closest approach. Further prevent the contamination from decays from long lifetime particles (such as K meson) and pileup, the tracks should also be displaced from the primary vertex: the absolute transverse impact parameter of the track should be less than 0.2 cm and the longitudinal impact parameter should be less than 17 cm. All the tracks should stay in a cone of jets with a size of $\Delta R < 0.3$.

The next goal is to reconstruct secondary vertices. In CSVv2, the inclusive vertex finder (IVF) is used. It starts from clustering tracks. The track with the largest impact parameter will be served as seed. By using variables of distance and angle between tracks, secondary vertex candidates are formed. An adaptive vertex fitting [101] is then performed on these candidates. After that, if multiple candidates share a large fraction of tracks, these candidates will be merged together. The next step is to balance the primary vertex and secondary vertices by performing continuously vertex fitting and merging to make sure the tracks are correctly associated to. A secondary vertex should be a vertex sharing less than 65% of the tracks with the primary vertex and separated from the primary vertex with a significance greater than 3σ . After the reconstruction of secondary vertex, the relevant variables from the secondary vertex and tracks will be used as inputs to a neural network [108] classifier. If no secondary vertex is found, the IVF algorithm will reconstruct the so-called "pseudo vertex", from tracks with a significance of the impact parameter greater than 2. The "pseudo vertex" allows us to calculate variables that will be input to neural network as if it is a secondary vertex but without using the vertex fit. If no "pseudo vertex" is found, lifetime variables of tracks will be used as inputs. By setting threshold on the MVA output discriminator, a b-tagging working point can be easily defined.

6.2.7 Missing transverse energy $\vec{E}_T$

Neutrinos and other BSM hypothetical neutral weakly interacting particles can not decay nor interact within any subdetector of CMS. If such particles are produced in collision, a momentum imbalance in the transverse plane to the beam axis can be detected.

The reconstruction missing transverse energy heavily relies on other physics objects. The most widely used type of $\vec{E}_T$ in CMS is the PF $\vec{E}_T$ [109], which is defined as the negative vectorial sum of the transverse momentum of all the PF candidates. The jet energy correction is also propagated to the calculation of $\vec{E}_T$:

$$\vec{E}_T = - \sum_{i \in \text{jets}} \vec{p}_{T,i}^{\text{JEC}} - \sum_{j \in \text{others}} \vec{p}_{T,j}. \quad (6.8)$$

CHAPTER

7

PERFORMANCE OF τ_H RECONSTRUCTION AND IDENTIFICATION

Contents

7.1	The τ Lepton Properties	88
7.2	Reconstruction and Identification of τ_h	88
7.2.1	The hadron-plus-strip algorithm	88
7.2.2	Discrimination between τ_h and jets	92
7.2.3	Discrimination against electron	94
7.2.4	Discrimination against muon	96
7.3	Measurement of the $e \rightarrow \tau_h$ Misidentification Rate	97
7.3.1	Datasets and Monte Carlo samples	97
7.3.2	Data-to-Monte Carlo corrections	98
7.3.3	Signals and background modeling	98
7.3.4	Event selection	101
7.3.5	Measurement	101
7.3.6	Systematic uncertainties	102
7.3.7	Results	103
7.4	Measurement of the $\mu \rightarrow \tau_h$ Misidentification Rate	104
7.4.1	Datasets and Monte Carlo samples	105
7.4.2	Data-to-Monte Carlo corrections	105
7.4.3	Background modeling	106
7.4.4	Event selection	106

7.4.5	Measurement	107
7.4.6	Systematic uncertainties	108
7.4.7	Results	109
7.5	Measurement of the Efficiencies of Discriminators against Elec- tron and Muons	111
7.5.1	Strategies	111
7.5.2	Results	111

7.1 The τ Lepton Properties

The τ lepton is a third generation charged lepton, with a lifetime of 2.9×10^{-13} second and a mass of $m_\tau = 1776.86 \pm 0.12$ MeV [14]. It is the only lepton sufficiently heavy to decay hadronically into hadrons and a neutrino. Approximately one third of the time, τ leptons decay leptonically into an electron or a muon, and two neutrinos. The neutrinos are undetectable so they are reconstructed as missing transverse energy, but the electrons and muons are reconstructed and identified by using the common techniques available for such lepton [102,103]. All the remaining decay final states of τ leptons consist of hadrons, typically with a combination of charged and neutral mesons, and a ν_τ . The hadronically decaying τ is denoted as τ_h . To be more specific, 11.5% of the time, τ leptons decay to a charged hadron h^- and a ν_τ , denoted as $h^\pm$ decay mode. In 25.9% of the time, they decay to a h^- , a π^0 and a ν_τ , denoted as $h^\pm\pi^0$ decay mode, via the resonance of $\rho(770)$ meson. In 9.5% and 9.8%, via the resonance of $a_1(1260)$ meson, the τ leptons decay to a h^- , two π^0 , and a ν_τ , denoted as $h^\pm\pi^0\pi^0$ decay mode, and to two h^- , a h^+ and a ν_τ , denoted as $h^\pm h^\mp h^\pm$ decay mode, respectively. In 4.8%, the τ leptons decay to two h^- , a h^+ , a π^0 and a ν_τ , denoted as $h^\pm h^\mp h^\pm\pi^0$ decay mode. The decays branching ratios are also valid for the charge-conjugate processes. To summarize, the final states of τ lepton decay is shown in table 7.1.

7.2 Reconstruction and Identification of τ_h

The τ_h is reconstructed using the Hadron-Plus-Strip (HPS) algorithm [110]. From neutral hadrons, charged hadrons, electron and photon PF candidates, the HPS algorithm checks whether the combined topology is compatible with one of the hadronic decay modes. If it matches to a specific signatures of a τ_h decay mode, the τ_h candidate with the corresponding decay mode is reconstructed. The τ_h reconstruction suffers from the contamination from quark or gluons jets, electrons and muons. τ_h is required to be isolated, by cutting on the isolation variable, to reject quark or gluons jets that might have been misidentified as τ_h . Discriminators built by cut-based or MVA-based method can be further implemented to reject electron and muon faking τ_h .

7.2.1 The hadron-plus-strip algorithm

The HPS algorithm starts from seeding the PF anti- k_T jets with a cone size of $\Delta R = 0.4$. Within the jets, the algorithm searches for charged hadrons, electrons and photons PF candidates. Since many decay modes of τ_h involve π^0 , the next step of the algorithm is to reconstruct π^0 candidates from grouping electron and photon PF candidates within a certain

Decay mode	Resonance	$\mathcal{B}$ (%)
Leptonic decays		35.2
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
Hadronic decays		64.8
$\tau^- \rightarrow h^- \nu_\tau$		11.5
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770)$	25.9
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other		3.3

Table 7.1: Decays of the τ leptons and their branching ratios ($\mathcal{B}$). For simplicity, only τ^- decays are shown. The values are also valid for the charge-conjugate processes.

region of $\Delta\eta \times \Delta\phi$, with the resulting object referred to as a "strip". The size of the strip has a fixed value of 0.05×0.20 in the (η, ϕ) plane in the previous run 1 version of the HPS algorithm. In run 2, the strip reconstruction was improved due to number of reasons:

- Photons from π^0 decays can convert into e^+e^- pairs. When the remaining electrons and photons reach ECAL after multiple scattering and bremsstrahlung, they can spread wider in the ECAL thus end up outside a fixed strip.
- The $\pi^\pm$ from τ_h decays can produce secondary particles from scattering with the tracker material. This leads to low- p_T electrons and photons falling outside the fixed strip.
- For high p_T prompt τ_h , the decay products tend to be Lorentz boosted along the flying direction. A smaller strip size can reduce the background contamination from quark or gluon jets.

Based on these motivations, a dynamic strip reconstruction for HPS algorithm has been developed in run 2: the highest p_T electron or photon e/γ not included in any strip is used to seed a new strip, and the strip position is set to its position. The p_T of the second highest e/γ within

$$\begin{aligned}\Delta\eta &= f(p_T^{e/\gamma}) + f(p_T^{\text{strip}}) \\ \Delta\phi &= g(p_T^{e/\gamma}) + g(p_T^{\text{strip}})\end{aligned}\tag{7.1}$$

of the strip position is grouped into the strip. The functions f and g are

$$\begin{aligned}f(p_T) &= 0.20 \times p_T^{-0.66} \\ g(p_T) &= 0.35 \times p_T^{-0.71},\end{aligned}\tag{7.2}$$

derived from simulation of single τ lepton gun events, such that the strip contains 95% of electrons and photons which are from the τ_h decays. Figure 7.1 shows that the 95% envelope

of points in each bin is fitted by using the analytic fitting form $a/(p_T)^b$. The upper limits on strip size are set to 0.3 in $\Delta\phi$ and 0.15 in $\Delta\eta$. The lower limits are set to 0.05 for both $\Delta\phi$ and $\Delta\eta$. The strip position is recomputed using the p_T weighted average of PF electrons of photons in the strip:

$$\begin{aligned}\eta_{\text{strip}} &= \frac{1}{p_T^{\text{strip}}} \sum p_T^{e/\gamma} \eta_{e/\gamma} \\ \phi_{\text{strip}} &= \frac{1}{p_T^{\text{strip}}} \sum p_T^{e/\gamma} \phi_{e/\gamma}.\end{aligned}\tag{7.3}$$

The reconstruction of the strip will end if there is no other e/γ PF candidates within the $\Delta\eta \times \Delta\phi$ window. Next seed will be selected and the procedure is repeated until all the strips are formed and no more e/γ PF candidates are left.

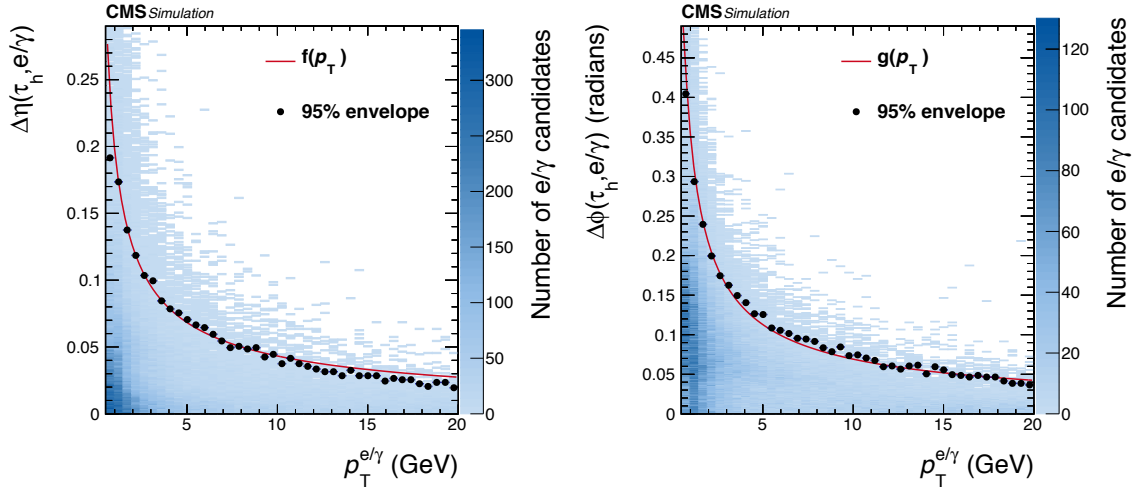


Figure 7.1: Distance in η (left) and in ϕ (right) between the τ_h and e/γ PF candidates from τ_h decay as a function of the p_T of the e/γ PF candidates. The curves (red lines), representing the function f and g given in equation 7.2, are fitted for 95% envelope points [110].

The next step is to reconstruct the relevant decay modes for τ_h . There are four topologies, based on the information of how many charged hadrons and strips are found and whether the reconstructed τ_h mass matches masses of the $\pi^\pm$, $\rho(770)$ or $a_1(1260)$ resonance:

- **One prong $h^\pm$:** The reconstructed visible mass m_{τ_h} is the mass of the reconstructed PF charged hadron. It should be set to the mass of a charged pion.
- **One prong plus one strip $h^\pm\pi^0$:** The charged hadron PF candidates and the strip together should have a reconstructed visible mass falling into the window of

$$0.3 \text{ GeV} - \Delta m_{\tau_h} < m_{\tau_h} < 1.3 \text{ GeV} \sqrt{p_T^{\tau_h}/(100 \text{ GeV})} + \Delta m_{\tau_h}.\tag{7.4}$$

The mass window is enlarged to consider the resolution for high p_T τ_h candidates. The upper limit is fixed to lie between 1.3 and 4.2 GeV.

- **One prong plus two strips** $h^\pm \pi^0 \pi^0$: The reconstructed mass is formed by adding the mass of charged hadron and two strips together. It should fall into the window of

$$0.4 \text{ GeV} - \Delta m_{\tau_h} < m_{\tau_h} < 1.2 \text{ GeV} \sqrt{p_T^{\tau_h} / (100 \text{ GeV})} + \Delta m_{\tau_h}, \quad (7.5)$$

with the upper limit constrained to lie between 1.2 and 4.0 GeV.

- **Three prongs** $h^\mp h^\pm h^\mp$: Three charged hadrons are required to be present. The mass should be in the range $0.8 < m_{\tau_h} < 1.5 \text{ GeV}$. This topology reconstructs two decay modes, $\tau^- \rightarrow h^- h^+ h^- \nu_\tau$ and $\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$.

The variable Δm_{τ_h} is the change in the mass of the τ_h candidates due to the additional number of e/γ PF candidates from size changes using the dynamic strip reconstruction. The exact calculation is:

$$\Delta m_{\tau_h} = \sqrt{\left(\frac{\partial m_{\tau_h}}{\partial \eta_{\text{strip}}} f(p_T^{\text{strip}}) \right)^2 + \left(\frac{\partial m_{\tau_h}}{\partial \phi_{\text{strip}}} g(p_T^{\text{strip}}) \right)^2}, \quad (7.6)$$

with

$$\frac{\partial m_{\tau_h}}{\partial \eta_{\text{strip}}} = \frac{p_z^{\text{strip}} E_{\tau_h} - E_{\text{strip}} p_z^{\tau_h}}{m_{\tau_h}} \quad \text{and} \quad (7.7)$$

$$\frac{\partial m_{\tau_h}}{\partial \phi_{\text{strip}}} = \frac{-(p_y^{\tau_h} - p_y^{\text{strip}}) p_x^{\text{strip}} + (p_x^{\tau_h} - p_x^{\text{strip}}) p_y^{\text{strip}}}{m_{\tau_h}}, \quad (7.8)$$

where $\vec{p}_{\tau_h} = (E_{\tau_h}, p_x^{\tau_h}, p_y^{\tau_h}, p_z^{\tau_h})$ and $\vec{p}_{\text{strip}} = (E_{\text{strip}}, p_x^{\text{strip}}, p_y^{\text{strip}}, p_z^{\text{strip}})$ are the four-momentum of the τ_h candidate and of the strip.

Figure 7.2 shows an illustration of the τ_h decay mode reconstruction process. The reconstructed τ_h visible mass should be compatible with the decay modes. Figure 7.3 shows the successful reconstructed τ_h by comparing data and simulation. The left spike corresponding to the mass of the $\pi^\pm$. The resonances of $\rho(770)$ and $a_1(1260)$ are also shown. The good agreement between data and simulation proves the good performance of the decay mode reconstruction.

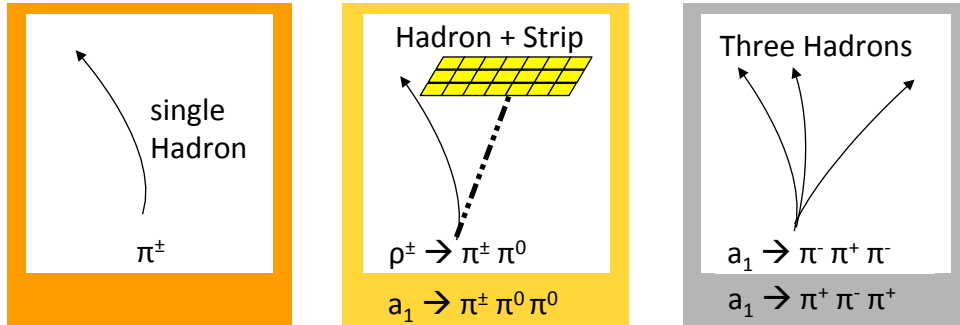
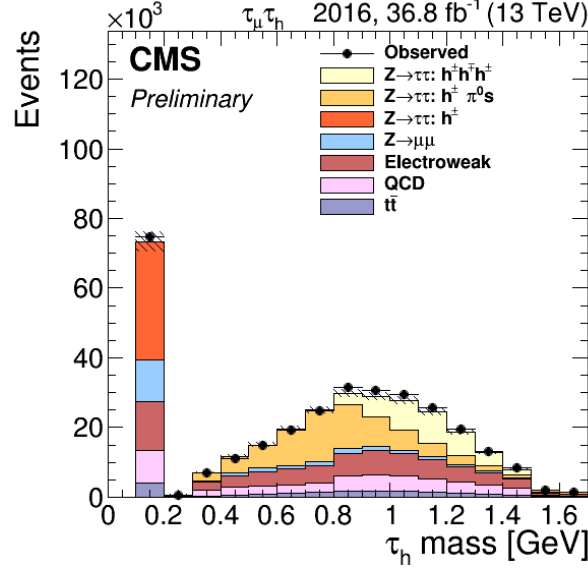


Figure 7.2: The decay mode reconstruction of HPS algorithm. It is capable to reconstruct $h^\pm$ (**left**), $h^\pm \pi^0$ (**middle**), $h^\pm \pi^0 \pi^0$ and $h^\mp h^\pm h^\mp$ (**right**) decay modes of τ_h .


 Figure 7.3: τ_h visible mass distribution in $\mu\tau_h$ events [111].

7.2.2 Discrimination between τ_h and jets

Hadronically decaying taus τ_h are more collimated and isolated than the quark and gluon jets. Isolation requirement can provide handle to reject such jets faking τ_h . Two versions of isolation are developed: cut-based $\Delta\beta$ corrected combined isolation discriminator and MVA-based isolation discriminator.

Cut-based isolation sum discriminator

The cut-based $\Delta\beta$ corrected isolation sum discriminator is computed by scale p_T sums of the charged particles and photons reconstructed using the PF algorithm with a cone centered around the τ_h with a size $\Delta R = 0.5$. The charged particle and photons should have $p_T > 0.5$ GeV. The charged particles should come from the τ_h production vertex, thus fulfilling the requirement of being $d_z < 0.2$ cm along the beamline. The isolation is defined as the following:

$$I_{\tau_h} = \sum p_T^{\text{charged}}(d_z < 0.2 \text{ cm}) + \max(0, \sum p_T^\gamma - \Delta\beta), \quad (7.9)$$

where the $\Delta\beta$ is the correction term for subtracting the contribution of pileup from the photon isolation. It is computed by summing up the p_T of the charged particles within a cone size of $\Delta R < 0.8$ around the τ_h direction and the particles should not be too far from the τ_h production vertex by requiring $d_z < 0.2$ cm. The $\Delta\beta$ is defined as follows:

$$\Delta\beta = 0.20 \times \sum p_T^{\text{charged}}(d_z < 0.2 \text{ cm}), \quad (7.10)$$

where the factor of 0.20 is obtained from fitting $\sum p_T^\gamma$ as a function of $\sum p_T^{\text{charged}}(d_z < 0.2 \text{ cm})$.

In the dynamic strip reconstruction, a photon candidate outside the signal cone can still contribute to the τ_h reconstruction. This increases the quark or gluon jet $\rightarrow \tau_h$ misidentification rates because of the decrease in the value of I_{τ_h} for misidentified τ_h . An additional

handle can be provided to reject quark or gluon jets by exploiting the scalar p_T sum of e/γ candidates included in strips, but outside the signal cone, defined as:

$$p_T^{\text{strip, outer}} = \sum p_T^{e/\gamma} (\Delta R > R_{\text{sig}}). \quad (7.11)$$

where R_{sig} is the signal cone. Requiring $p_T^{\text{strip, outer}} < 0.1 \times p_T^{\tau_h}$ can reduce 20% of $\text{jet} \rightarrow \tau_h$ misidentification rate.

The loose, medium and tight working points of cut-based isolation are obtained by having the I_{τ_h} less than 2.5, 1.5 and 0.8 GeV, respectively. Figure 7.4 shows the misidentification rate as a function of τ_h identification efficiency. The one with dynamic strip reconstruction, $\Delta\beta$ optimized to 0.2 and $p_T^{\text{strip, outer}}$ cut applied has the best performance.

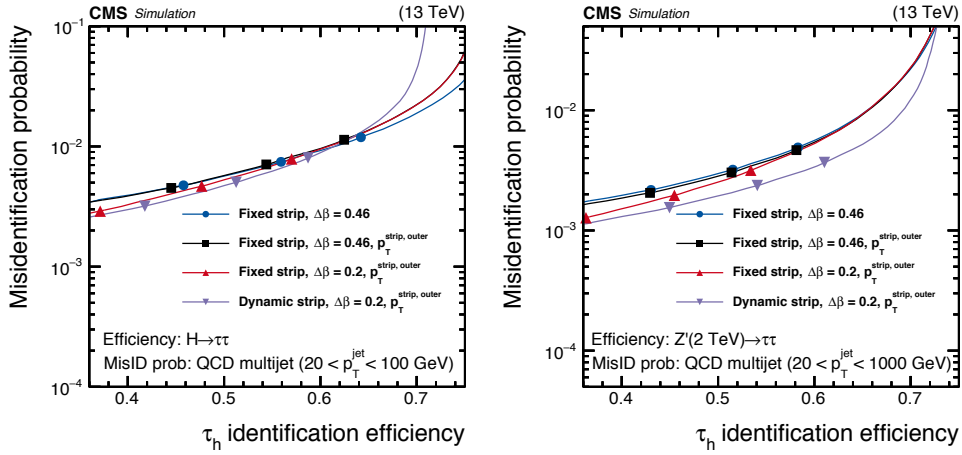


Figure 7.4: The expected misidentification probabilities as a function of the τ_h identification efficiencies, estimated using simulated $H \rightarrow \tau\tau$ (left) and $Z'(2 \text{ TeV}) \rightarrow \tau\tau$ (right), and QCD multijet events. In each plot, from left to right, the three points on each curve represents tight, medium and loose working points for the cut-based isolation sum discriminator [110].

MVA-based isolation discriminator

The MVA-based isolation discriminator is built by using BDT method. The BDT is trained with samples containing genuine τ decays and quark-gluon jets. The input variables have four sources: isolation-like variables, decay mode, τ_h lifetime variables and variables related to dynamic strip reconstruction. The input variables are:

- The scalar p_T sum of charged particles $\sum p_T^{\text{charged}}$ and of photons $\sum p_T^\gamma$.
- The reconstructed tau decay mode.
- The transverse impact parameter d_{xy} of the leading track of the τ_h candidate and its value divided by its uncertainty.
- The distance between the τ production and decay vertices and the value divided by its uncertainty.

- Signed transverse impact parameter (IP) of the leading track and signed 3D impact parameter of leading track.
- The τ flight distance for three prong decay.
- The χ^2 value of the fit for the leading track.
- The multiplicity of photons.
- The fraction of electromagnetic energy in the signal cone.
- The p_T weighted $\Delta\eta$ and $\Delta\phi$ of photons in the strips outside of the signal cone, p_T weighted ΔR of photons in signal cone, and p_T weighted ΔR of photons in the isolation annulus.

The BDT discriminator is trained with samples generated according to LHC run-II scenario, including $Z/\gamma^* \rightarrow \tau\tau$, Higgs with a mass of 125 GeV decaying to $\tau\tau$ (including gluon-fusion, vector boson fusion, and productions associated with vector bosons and $t\bar{t}$), SUSY Higgs decaying to $\tau\tau$ (including gluon-fusion and production associated with $b\bar{b}$), $Z' \rightarrow \tau\tau$ and $W' \rightarrow \tau\nu$ as signal events, and W+jets, QCD multijet and $t\bar{t}$ productions as background events. The events cover a large p_T range from 20 GeV to 2000 GeV. Figure 7.5 shows the BDT output score after training. It shows a good discriminating power between τ_h and quark or gluon jets. The six working points available are very loose, loose, medium, tight, very tight and very very tight working points. The working points are defined based on cuts on the BDT output scores. The cut on the BDT output is applied as a function of the p_T of τ_h to get flat efficiency up to higher p_T . The six working points correspond to 40% to 90% efficiency with incremental steps of 10%, respectively. Figure 7.6 shows the comparisons between cut-based isolation sum discriminator and MVA-based isolation discriminators with different training samples. The MVA-based isolation discriminator has better expected performance than the cut-based. The points of MVA-based discriminator between 2015 and 2016 trainings are shifted but still lying in the same curve. This proves that the performance the MVA-based discriminators is stable.

7.2.3 Discrimination against electron

An electron can also be misidentified as τ_h , mainly in $h^\pm$ and $h^\pm\pi^0$ decay modes. The tracks from electrons can mimic the $\pi^\pm$ from τ_h decay, while the bremsstrahlung photons can be misidentified as π^0 strips. There is no a single variables that can differentiate τ_h from electrons. Thus, BDT method is used to build a MVA-based discriminator against electron.

The chosen input variables can be grouped into four sources: the variables that quantify the distribution in energy depositions in the ECAL, variables sensitive to the amount of bremsstrahlung emitted, the variables to distinguish electromagnetic from hadronic shower and photon related variables. To be more specific, they are:

- The ratio of energy deposits associated to the τ_h in the ECAL to the sum of energy in the ECAL and HCAL, $E_{\text{ECAL}}/(E_{\text{ECAL}} + E_{\text{HCAL}})$.
- The ratios of ECAL and HCAL energies relative to the momentum of the leading charged particle track of τ_h , E_{ECAL}/p and E_{HCAL}/p .

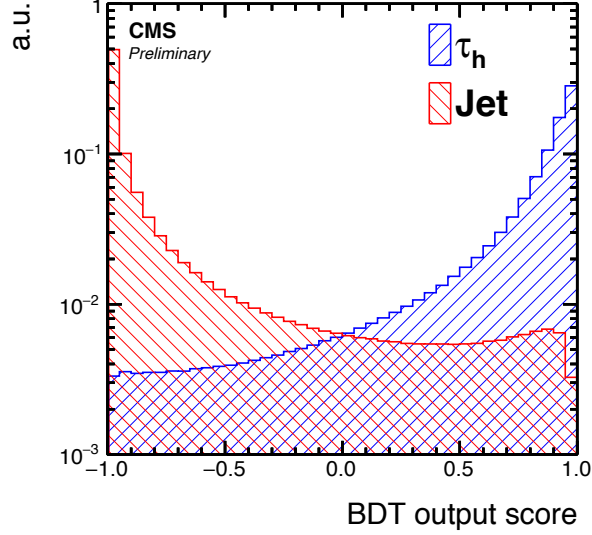


Figure 7.5: The distribution of the BDT output for the MVA-based τ_h isolation discriminator. The blue indicates $Z/\gamma^* \rightarrow \tau\tau$ MC events and red MC W +jets events [110].

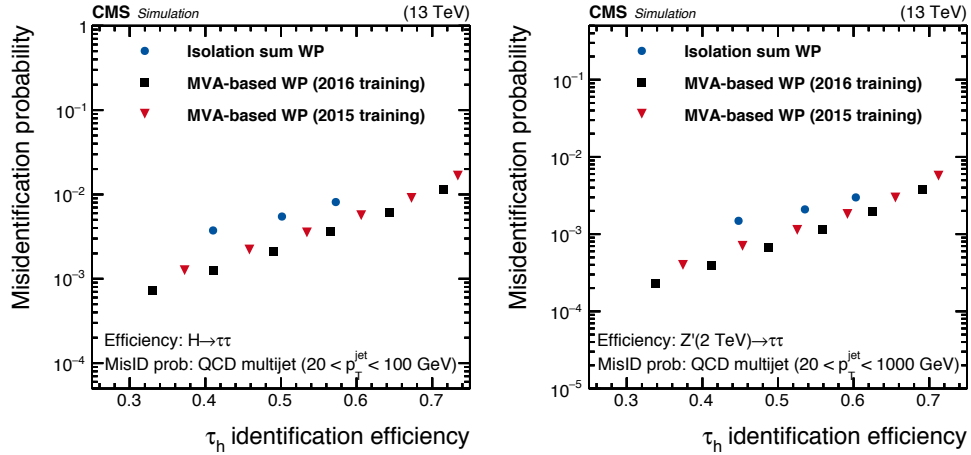


Figure 7.6: The misidentification rates for τ_h as a function of τ_h efficiencies, estimated using simulated $H \rightarrow \tau\tau$ (**left**) and $Z' (2 \text{ TeV}) \rightarrow \tau\tau$ (**right**), and QCD multijet events. MVA-based discriminators trained with different MC events are compared to each other, as well as to the cut-based isolation sum discriminator [110].

- The p_T weighted root-mean-square distances in η and ϕ between the photons in any strip and the leading track, $\sqrt{\sum(\Delta\eta)^2 p_T^\gamma}$ and $\sqrt{\sum(\Delta\phi)^2 p_T^\gamma}$.
- The fraction of τ_h energy carried by photons, $\sum E_\gamma/E_\tau$.
- The fraction of bremsstrahlung, defined by $F_{\text{brem}} = (p_{\text{in}} - p_{\text{out}})/p_{\text{in}}$, where p_{in} and p_{out} is the momentum of the leading track, reconstructed by the GSF algorithm, at the innermost and outermost position of the tracker.
- The ratio between the total ECAL energy and the inner track momentum, $(E_e + \sum E_\gamma)/p_{\text{in}}$, where E_e is the energies of the electron cluster and $\sum E_\gamma$ represents the energies of the bremsstrahlung photons cluster.
- The ratio of energies of the bremsstrahlung photons measured in the ECAL and in the tracker, $\sum E_\gamma/(p_{\text{in}} - p_{\text{out}})$.
- The reconstructed visible mass of τ_h .
- $(N_{\text{hits}}^{\text{GSF}} - N_{\text{hits}}^{\text{KF}})/(N_{\text{hits}}^{\text{GSF}} + N_{\text{hits}}^{\text{KF}})$, where $N_{\text{hits}}^{\text{GSF}}$ and $N_{\text{hits}}^{\text{KF}}$ are the number of hits associated to the track reconstructed using, respectively, the GSF and Kalman filter (KF) track reconstruction algorithms.
- The χ^2 divided by the degree of freedom of the GSF track.

The BDT discriminator is trained with samples generated according to LHC run-II scenario, including $Z/\gamma^* \rightarrow \tau\tau$, $W \rightarrow \tau\nu$, $Z' \rightarrow \tau\tau$ and $W' \rightarrow \tau\nu$, Higgs with a mass of 125 GeV decaying to $\tau\tau$ (including gluon-fusion, vector boson fusion, and productions associated with vector bosons and $t\bar{t}$) and SUSY Higgs decaying to $\tau\tau$ (including gluon-fusion and production associated with $b\bar{b}$) as signals, and $Z/\gamma^* \rightarrow ee$, $W \rightarrow e\nu$, $Z' \rightarrow ee$ and $W' \rightarrow e\nu$ as backgrounds. Different working points are defined on the BDT output, according to the efficiency for real τ_h to pass the discriminator. In total, five working points are available: very loose, loose, medium, tight and very tight. Figure 7.7 shows the expected τ_h efficiencies as a function of transverse momentum of τ_h , $p_T^{\tau_h}$, and expected $e \rightarrow \tau_h$ misidentification rates as a function of generated electron p_T^e .

7.2.4 Discrimination against muon

Muons have a high probability to be misidentified as τ_h in the $h^\pm$ decay mode. A cut-based discriminator is used to reject muons faking τ_h by vetoing τ_h in case muon signals are found in the muon system subdetector near the τ_h direction. Two working points are built according to the following cuts, corresponding to different efficiencies that the τ_h pass the discriminator:

- **anti- μ loose:** τ_h fails the discriminator if track segments are found in at least two muon stations within a cone of size $\Delta R = 0.3$ around the τ_h flying direction or in case the energy in ECAL plus in HCAL to the leading charged hadron of the τ_h amounts to less than 0.2 times the momentum of the leading charged hadron.
- **anti- μ tight:** τ_h fails the discriminator in case it fails the anti- μ loose discriminator and if a hit is present in the CSC, DT or RPC chambers located in the two outermost muon stations within a cone size $\Delta R = 0.3$ around the τ_h flying direction.

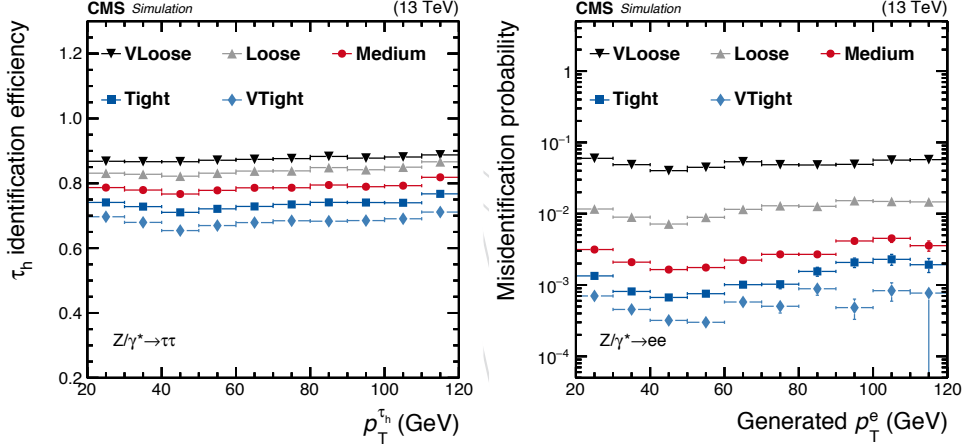


Figure 7.7: The expected efficiency of τ_h as a function of p_T of τ_h via simulated $Z/\gamma^* \rightarrow \tau\tau$ events (**left**), and the $e \rightarrow \tau_h$ misidentification rate as a function of generated electron p_T via simulated $Z/\gamma^* \rightarrow ee$ (**right**) events for the five working points of MVA-based anti-electron discriminator [110].

7.3 Measurement of the $e \rightarrow \tau_h$ Misidentification Rate

In this section, the $e \rightarrow \tau_h$ misidentification rate is measured in data via the tag-and-probe method on $Z/\gamma^* \rightarrow ee$ events. A well-identified and isolated electron is referred to as the "tag", and a τ_h that passed loose preselection criteria corresponds to the "probe". The events are divided into two categories, "pass" and "fail", depending on whether the τ_h probe passes or fails the anti-electron discriminator.

The number of $Z/\gamma^* \rightarrow ee$ signal events in each category is then extracted from a simultaneous maximum likelihood fit of the visible mass m_{vis} of the tag-probe pair in the two regions, with uncertainties modeled by nuisance parameters.

7.3.1 Datasets and Monte Carlo samples

The measurement is performed on 35.9 fb^{-1} of data [112] at 25 ns bunch spacing at center-of-mass energy of 13 TeV collected in 2016 during LHC run 2 by CMS detector. Only those data certified as good proton-proton collisions are used.

The Monte Carlo samples are used to model the $Z/\gamma^* \rightarrow ee$ events as signal and the $Z/\gamma^* \rightarrow \tau\tau$, W +jets, $t\bar{t}$, single top quark and diboson (WW , WZ and ZZ) as background processes. The $Z/\gamma^* \rightarrow ee$, $Z/\gamma^* \rightarrow \tau\tau$ and W +jets events are modeled using LO matrix element calculations from the MADGRAPH5_AMC@NLO event generator. Diboson events are estimated using NLO matrix element calculations also from MADGRAPH5_AMC@NLO, except $WW \rightarrow \ell\nu+2\text{jets}$ ($\ell = e, \mu, \tau$) and $ZZ \rightarrow 4\ell$ which are modeled using the NLO matrix element calculation from POWHEGv2. $t\bar{t}$ and single top quark events are generated using POWHEGv2 in NLO precision. Samples with different numbers of additional jets for W +jets events have been generated in order to increase the sample statistics in regions of high signal purity. PYTHIA8, with tune CUETP8M1, is used to model the parton shower, hadronization processes and underlying events. TAUOLA is implemented to model the τ

lepton decay. The NNPDF3.0 parton distribution functions [113] are used.

7.3.2 Data-to-Monte Carlo corrections

The pileup reweighting is applied to all simulation samples in order to have similar pileup distributions between data and simulation. A minimum bias cross section of 69.2 mb is used. The range of the number of primary vertices used is 0-80. The MC distribution of the number of primary vertices is reweighted to match the number of pileup interactions in data given the recommended minimum bias cross section of 69.2 mb.

The events are triggered by the single electron trigger "HLT_Ele25_eta2p1_WPTight_Gsf". Data-to-MC scale factors are derived from an auxiliary study on $Z/\gamma^* \rightarrow ee$ using a tag-and-probe method selecting a well isolated and identified electron already fired by the trigger as "tag" and another electron as "probe". A probe is defined as pass if an offline trigger object is found within a cone of size $\Delta R < 0.5$ on the probe direction. The efficiency ϵ is defined as

$$\epsilon = \frac{N_{Z/\gamma^* \rightarrow ee}^{\text{pass}}}{N_{Z/\gamma^* \rightarrow ee}^{\text{pass}} + N_{Z/\gamma^* \rightarrow ee}^{\text{fail}}}, \quad (7.12)$$

where $N_{Z/\gamma^* \rightarrow ee}^{\text{pass}}$ is the number of signal events in "pass" category and $N_{Z/\gamma^* \rightarrow ee}^{\text{fail}}$ is the number of signal events in "fail" category. The numbers of signal events in data are determined from a double gaussian fit of the simulation to the data, while the background events are determined from an exponential fit. The observable, used for signal extraction by the fit, is the invariant mass of the two electrons in the mass window of the Z boson mass, $75 < m_{ll} < 105$ GeV. The data-to-MC scale factors are binned in p_T and η of the electron. Figure 7.8 (**left**) shows an example of the single electron trigger efficiencies both in data and in MC, with the scale factors indicated in the lower panel.

The electrons are required to pass the MVA-based tight identification trained with 2016 MC samples, with an efficiency of 80%. The electron should be isolated by requiring the relative isolation of electron $I_e/p_T^e < 0.1$. A set of data-to-MC scale factors are dedicated to correct the difference between data and simulation due to the electron identification and isolation. The scale factors are derived also from an auxiliary study by using "tag-and-probe" approach on $Z/\gamma^* \rightarrow ee$ events as the study to obtain trigger scale factors. The tag is defined as a well identified and isolated reconstructed electron. The probe is a loosely selected reconstructed electron candidate which will be further used to test if it passes or fails the identification and isolation requirements. The efficiency is defined previously as equation 7.12. Figure 7.8 (**right**) shows an example of the electron identification and isolation efficiencies and the corresponding data-to-MC scale factors.

7.3.3 Signals and background modeling

The signal, $Z/\gamma^* \rightarrow ee$ process, with a reconstructed electron matched to a generator level electron, and a reconstructed τ_h matched to a generator level electron, is estimated from MC simulation. Other Drell-Yan events, corresponding to $Z/\gamma^* \rightarrow \tau\tau$, are taken from MC simulation as background events. The $t\bar{t}$, single top quark and diboson processes are modeled directly from the MC simulation. All the simulation events are weighted to certain

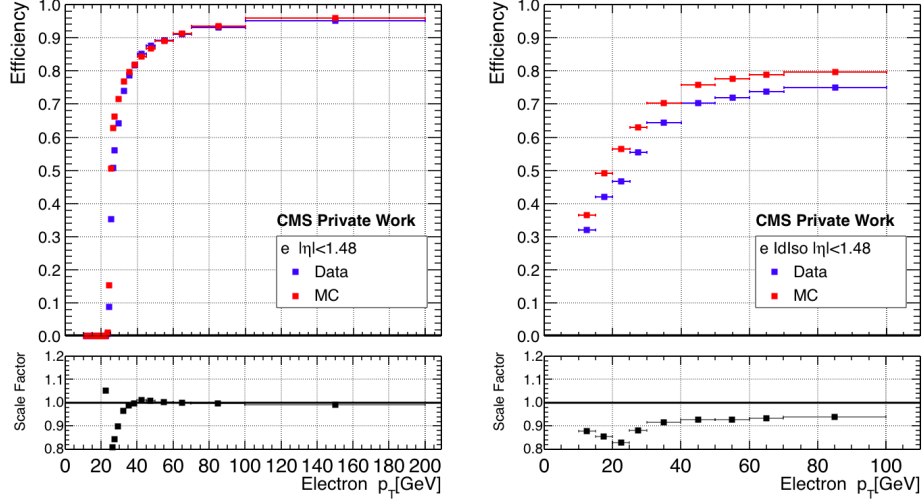


Figure 7.8: The "HLT_Ele25_eta2p1_WPTight_Gsf" (**left**) single electron trigger efficiencies and the electron identification and isolation efficiencies (**right**) for data and for simulation as a function of reconstructed electron p_T . Examples shown are from the ECAL barrel region $|\eta| < 1.48$. The data-to-MC scale factors used for the analysis are shown in the lower panel. The last bin in both plots includes the overflow bin.

values of cross section as the following equation shows:

$$w = \frac{\sigma \times \mathcal{L}}{n} \quad (7.13)$$

where w is the weight, σ is the cross section of a certain process, $\mathcal{L}$ is the integrated luminosity of the data and n is the total number of events for the MC samples. The cross section values for each process are listed in Table 7.2.

W+jets events are modeled using "stitching" technique to increase the statistic. Five MC samples are generated with different numbers of jets: inclusive, W+1 jet, W+2 jets, W+3 jets and W+4jets. The cross sections of these samples are known. The stitching procedure starts from dividing the inclusive W+jets event into five different subsets based on the number of generator level jets. The subsets with 1 jet, 2 jets, 3 jets and 4 jets will be combined with the additional generated W+1 jet, W+2 jets, W+3 jets and W+4 jets samples and will be weighted with the corresponding cross section values, respectively. The remaining subset from inclusive sample with either no jet or larger than 4 jets will be weighted with the inclusive W+jets cross section value. The new formed five subsets are then summed up together.

While the shape of the W+jets background is taken from MC simulation with the stitching method, a normalization factor derived from a W+jets enriched control region is further applied. The control region is defined by requiring events to have transverse mass between the electron and $\vec{\cancel{E}}_T$, the $m_T(e, \vec{\cancel{E}}_T)$, larger than 70 GeV. The $m_T(e, \vec{\cancel{E}}_T)$ is defined by the following equation:

$$m_T(e, \vec{\cancel{E}}_T) = \sqrt{2p_T^e \cancel{E}_T (1 - \cos \Delta\phi)}, \quad (7.14)$$

where the missing transverse energy $\cancel{E}_T$ used as input for the transverse mass computation

Process	Generator	Cross section (pb)
Drell-Yan $Z/\gamma^*(50) + Jets \rightarrow \ell\ell$	MADGRAPH5_AMC@NLO (LO)	1.16×4954.0
$t\bar{t}$	POWHEGv2	831.8
W+jets inclusive	MADGRAPH5_AMC@NLO (LO)	61526.7
W+1jet	MADGRAPH5_AMC@NLO (LO)	1.21×9644.5
W+2jets	MADGRAPH5_AMC@NLO (LO)	1.21×3144.5
W+3jets	MADGRAPH5_AMC@NLO (LO)	1.21×954.8
W+4jets	MADGRAPH5_AMC@NLO (LO)	1.21×485.6
$VV \rightarrow 2\ell 2\nu$	MADGRAPH5_AMC@NLO (NLO)	11.95
$ZZ \rightarrow 2\ell 2jets$	MADGRAPH5_AMC@NLO (NLO)	3.22
$ZZ \rightarrow 4\ell$	POWHEGv2	1.212
$WW \rightarrow \ell\nu 2jets$	POWHEGv2	49.997
$WZ \rightarrow 2\ell 2jets$	MADGRAPH5_AMC@NLO (NLO)	5.595
$WZ \rightarrow \ell 3\nu$	MADGRAPH5_AMC@NLO (NLO)	3.05
$WZ \rightarrow \ell\nu 2jets$	MADGRAPH5_AMC@NLO (NLO)	10.71
single t quark tW channel	POWHEGv2	35.6
single $\bar{t}$ quark tW channel	POWHEGv2	35.6
single t quark t channel	POWHEGv2	136.95
single $\bar{t}$ quark t channel	POWHEGv2	80.95

Table 7.2: Cross section values for all Monte Carlo simulation samples used in $e \rightarrow \tau_h$ misidentification rate measurement. A k-factor of 1.16 is considered for the Z/γ^*+jets samples, and 1.21 for the W+jets samples

is evaluated using the particle flow algorithm and $\Delta\phi$ is the difference in azimuthal angle between the electron momentum and $\cancel{E}_T$.

The QCD multijet events as background is estimated with a fully data-driven method. Its shape is taken from the same signed region (SS region), where "same signed" means the electron and the τ_h candidate are required to carry the same electric charge. In this region, the signals $Z/\gamma^* \rightarrow ee$ can be neglected, and the main contribution is due to QCD multijet and W+jets events. The contribution from other MC simulated processes are subtracted from the data. There is no guarantee that the SS sample has the same yield as the opposite-signed region. A normalization factor, the OS/SS ratio, is used to weight the SS shape. The ratio reflects a charge asymmetry mostly due to the gluon splitting and low mass QCD resonances. The measured value is 1.06 obtained as the ratio between opposite-signed and same-signed events in inverted isolation region where the electron isolation is required to have $I_e/p_T > 0.1$ and the τ_h should fail the isolation discriminator.

7.3.4 Event selection

Events are selected by requiring the presence of a well identified and isolated electron (the tag) plus a τ_h candidate (the probe) passing loose identification criteria. In particular, the selections applied on the tag electron are the following:

- $p_T > 26$ GeV and $|\eta| < 2.1$;
- tight identification and isolation $I_e < 0.1 \cdot p_T$;
- the impact parameters $d_z < 0.2$ cm, $|d_{xy}| < 0.045$ cm;
- matching with the trigger object in a cone of radius $\Delta R < 0.5$.

The probe, τ_h , is required to satisfy the following criteria:

- $p_T > 20$ GeV and $|\eta| < 2.3$;
- be reconstructed in one of the decay modes $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$ or $h^\pm h^\mp h^\pm$;
- pass the loose working point of the cut-based discriminator against $\mu \rightarrow \tau_h$ misidentification;
- pass the tight working point of the MVA-based isolation discriminator;
- be separated from the tag electron by $\Delta R > 0.5$.

The tag and probe pair is required to have opposite electric charge. If an event contains two electrons qualifying as tag, both combinations of tag and probe are considered.

The contribution coming from W +jets and $t\bar{t}$ backgrounds, with one jet misidentified as the τ_h , is reduced by requiring the selected events to pass a cut on the transverse mass of the tag electron plus missing transverse energy of $m_T(e, \vec{E}_T) < 30$ GeV. The motivation of this cut is that, in W +jets, the lepton and the neutrino are supposed to be back-to-back, which leads to a large $m_T(e, \vec{E}_T)$ values.

7.3.5 Measurement

Depending on whether the probe passes or fails the anti-electron discriminator under study, the event enters the "pass" or the "fail" category, respectively. The observed $e \rightarrow \tau_h$ misidentification rate is then measured with a simultaneous maximum likelihood fit of the number of $Z/\gamma^* \rightarrow ee$ events, N_{pass}^{probe} and N_{fail}^{probe} , in both categories. The parameter of interest in the fit is the data-to-MC scale factor for the $e \rightarrow \tau_h$ misidentification rate. The "pass" and "fail" categories are anti-correlated. Supposing a constant total number of signal events N_0 and an $e \rightarrow \tau_h$ misidentification rate ϵ , the number of signal events in "pass" or "fail" category has the following relation:

$$\begin{aligned} N_{pass} &= \epsilon \times N_0, \\ N_{fail} &= (1 - \epsilon) \times N_0. \end{aligned} \tag{7.15}$$

After a maximum likelihood fit, the number of signal events in "pass" or "fail" category will be changed to

$$\begin{aligned} N'_{pass} &= \epsilon' \times N_0 = r\epsilon \times N_0, \\ N'_{fail} &= (1 - \epsilon') \times N_0 = (1 - r\epsilon) \times N_0, \end{aligned} \tag{7.16}$$

where ϵ' is the post-fit misidentification rate and $r = \epsilon'/\epsilon$ is the parameter of interest. So the signal yield in the "pass" or "fail" category must scale as:

$$\begin{aligned}\frac{N'_{\text{pass}}}{N_{\text{pass}}} &= \frac{r\epsilon \times N_0}{\epsilon \times N_0} = r, \\ \frac{N'_{\text{fail}}}{N_{\text{fail}}} &= \frac{(1 - r\epsilon)}{1 - \epsilon}.\end{aligned}\tag{7.17}$$

The visible mass m_{vis} of the tag and probe pair, computed using the energy and the momentum reconstructed by the electron reconstruction algorithm for the tag and the energy and the momentum reconstructed by the HPS algorithm for the probe, is used as discriminant variable for the fit in the pass category. The mass distribution is fitted, in the mass range $60 \text{ GeV} < m_{vis} < 120 \text{ GeV}$ with templates of the $Z/\gamma^* \rightarrow ee$ signal and the $Z/\gamma^* \rightarrow \tau\tau \rightarrow e\tau_h$, $Z/\gamma^* \rightarrow \tau\tau \rightarrow ee$, W +jets, $t\bar{t}$, single-top, di-boson (WW, WZ, ZZ) and QCD multijet backgrounds. The fail category is used to constrain the overall normalization of the $Z/\gamma^* \rightarrow ee$ signal and the fit in this category is performed on the integral distribution of the events, also between 60 GeV and 120 GeV.

7.3.6 Systematic uncertainties

Systematic uncertainties are represented by nuisance parameters in the maximum likelihood fit. The normalization uncertainties following the log-normal distribution (equation 4.7) are:

- **Luminosity:** the uncertainty on the integrated luminosity is taken to be 2.6%, applied on all MC simulated processes.
- **Electron identification/isolation/trigger:** the combination of the three effects is taken to be 2%, applied on all MC simulated processes.
- **Tau identification:** the uncertainty is taken to be 3%, applied on all MC simulated processes.
- **$t\bar{t}$ cross section:** a 10% uncertainty is attributed to the $t\bar{t}$ sample.
- **Di-boson and single-top cross sections:** the uncertainty on the sum of these small backgrounds is considered to be 10%.
- **W+jets normalization:** a value of 20% is associated to it.
- **QCD normalization:** 20% is associated to it.
- **DY normalization:** an uncertainty of 3% is associated to the Drell-Yan process; an additional 3% uncertainty is associated to the $Z/\gamma^* \rightarrow ee$ process only, to disentangle possible differences between the $Z/\gamma^* \rightarrow ee$ and the $Z/\gamma^* \rightarrow \tau\tau$ normalization.

The uncertainties affecting the observable's shape are modeled using equation 4.9. They have the following sources:

- **The tag electron energy scale:** the energy scale of tag electrons matched to a generator level electron has an uncertainty of $\pm 1\%$ in the barrel region, and $\pm 2.5\%$ in endcaps region.

- **The probe τ_h energy scale:** the energy scale of probe τ_h matched to a generator level τ_h candidates has an uncertainty of $\pm 1.5\%$.
- **The $e \rightarrow \tau_h$ energy scale:** the energy scale of probe τ_h matched to electrons has an uncertainty of $\pm 3\%$.
- **The visible mass resolution scale:** the visible mass resolution scale has an uncertainty of $\pm 20\%$, applied on signal $Z/\gamma^* \rightarrow ee$ events. The up and down shapes are defined as:

$$\begin{aligned} m_{\text{vis}}^{\text{up}} &= m_{\text{GEN}}^{\text{peak}} + (1 + \text{scale})(m_{\text{vis}} - m_{\text{GEN}}^{\text{peak}}), \\ m_{\text{vis}}^{\text{down}} &= m_{\text{GEN}}^{\text{peak}} + (1 - \text{scale})(m_{\text{vis}} - m_{\text{GEN}}^{\text{peak}}), \end{aligned} \quad (7.18)$$

where $m_{\text{GEN}}^{\text{peak}}$ is the Z peak value of the generator level dielectron visible mass obtained by performing a fit with a double Gaussian function on the visible mass distribution. Figure 7.9 shows the nominal, up and down shapes of visible mass resolution scale.

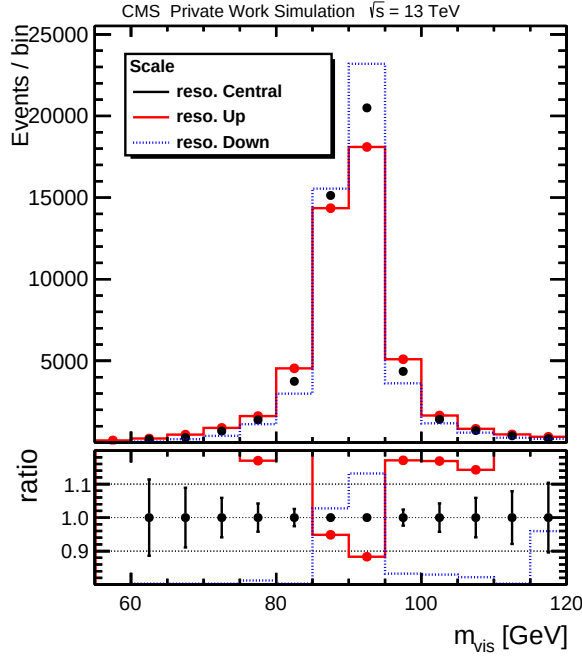


Figure 7.9: The nominal, and the 20% up and down variation shapes of visible mass resolution for $Z/\gamma^* \rightarrow ee$ events.

Finally, the bin-by-bin uncertainties as described in subsection 4.1.2, accounting for the limited number of events in every bin of the distributions, are considered for all MC simulated processes.

7.3.7 Results

Separate fits are used for τ_h in the ECAL barrel ($|\eta| < 1.460$) and in the ECAL endcaps ($|\eta| > 1.558$) region to account for the effects of different ECAL geometries. The fitted visible mass distributions in the "pass" category, for τ_h in barrel region, are shown in figure 7.10 for

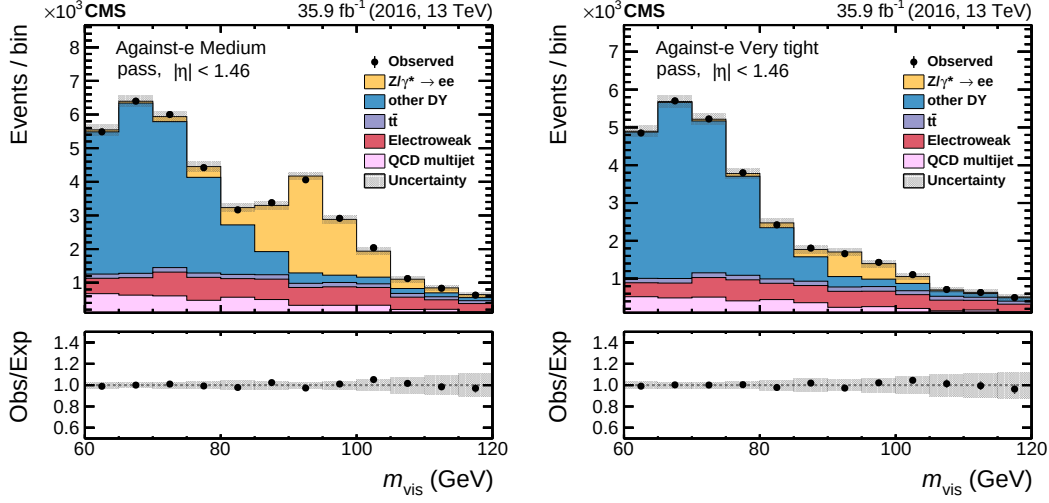


Figure 7.10: The fitted distribution in m_{vis} in the pass category for the medium (**left**) and very tight (**right**) working points of the anti-electron discriminator in the barrel region [110]. The shaped bands represent the statistic uncertainty due to limited events of MC simulation samples and the systematic uncertainties. The vertical bars (not visible) indicate the statistic uncertainty of data. The electroweak background contribution includes the contributions from W +jets, diboson and single top production processes.

medium and very tight working points of the anti-electron discriminator, as examples. On the left of figure 7.10, the distribution shows two peaks. The left peak corresponding to the $Z/\gamma^* \rightarrow \tau\tau \rightarrow e\tau_h$ background. Because hadronic decay products of τ contain neutrinos that do not contribute to the reconstruction of visible mass, the Z peak value is lower than the one of $Z/\gamma^* \rightarrow ee$ process.

The $e \rightarrow \tau_h$ misidentification rates measured for different working points of the anti-electron discriminator are given in table 7.3 and shown in figure 7.11. The data-to-MC scale factors are recommended to be applied on all CMS analyses using 2016 data involving τ_h reconstructed object. For the τ_h in the barrel region, the measured misidentification rates in data exceed the Monte Carlo prediction, with the difference between data and simulation increasing for the tight and very tight working points of the discriminator. A similar trend is observed for the τ_h in the endcaps region.

7.4 Measurement of the $\mu \rightarrow \tau_h$ Misidentification Rate

In this section, the $\mu \rightarrow \tau_h$ misidentification rate is measured in data via a tag-and-probe method on $Z/\gamma^* \rightarrow \mu\mu$ events. The events are divided into two categories, "pass" and "fail", depending on whether the probe reconstructed τ_h passes or fails the cut-based anti- μ discriminator.

The number of $Z/\gamma^* \rightarrow \mu\mu$ signal events in each category is then extracted from a simultaneous maximum likelihood fit of the invariant mass m_{vis} of the tag-probe pair in the two regions, with uncertainties modeled as nuisance parameters of the fit.

ECAL barrel ($\eta < 1.460$)			
WP	Misidentification rate (pre-fit)	Misidentification rate (post-fit)	Scale Factor
Very Loose	$(4.44 \pm 0.66) \cdot 10^{-2}$	$5.39^{+0.22}_{-0.31} \cdot 10^{-2}$	$1.213^{+0.051}_{-0.070}$
Loose	$(8.19 \pm 1.27) \cdot 10^{-3}$	$1.08^{+0.03}_{-0.02} \cdot 10^{-2}$	$1.320^{+0.039}_{-0.025}$
Medium	$(2.06 \pm 0.32) \cdot 10^{-3}$	$2.72^{+0.13}_{-0.13} \cdot 10^{-3}$	$1.323^{+0.066}_{-0.065}$
Tight	$(8.78 \pm 1.38) \cdot 10^{-4}$	$1.23^{+0.10}_{-0.10} \cdot 10^{-3}$	$1.402^{+0.115}_{-0.119}$
Very Tight	$(4.92 \pm 0.77) \cdot 10^{-4}$	$5.95^{+0.87}_{-0.81} \cdot 10^{-4}$	$1.207^{+0.177}_{-0.165}$
ECAL endcap ($\eta > 1.558$)			
WP	Misidentification rate (pre-fit)	Misidentification rate (post-fit)	Scale Factor
Very Loose	$(6.71 \pm 0.98) \cdot 10^{-2}$	$9.23^{+0.32}_{-0.12} \cdot 10^{-2}$	$1.375^{+0.048}_{-0.018}$
Loose	$(1.04 \pm 0.16) \cdot 10^{-2}$	$1.51^{+0.03}_{-0.03} \cdot 10^{-2}$	$1.380^{+0.039}_{-0.036}$
Medium	$(2.63 \pm 0.41) \cdot 10^{-3}$	$4.03^{+0.32}_{-0.35} \cdot 10^{-3}$	$1.527^{+0.122}_{-0.133}$
Tight	$(1.07 \pm 0.16) \cdot 10^{-3}$	$2.03^{+0.31}_{-0.31} \cdot 10^{-3}$	$1.900^{+0.293}_{-0.295}$
Very Tight	$(6.36 \pm 1.00) \cdot 10^{-4}$	$1.25^{+0.27}_{-0.28} \cdot 10^{-3}$	$1.968^{+0.436}_{-0.453}$

Table 7.3: Probability for electrons to be misidentified as τ_h for the different working points of the MVA based anti-electron discriminator, split in barrel and endcaps according to the τ_h pseudorapidity $|\eta|$. For each working point, the $e \rightarrow \tau_h$ misidentification rate is defined as the fraction of τ_h probes passing the given discriminator with respect to the total number of τ_h probes.

7.4.1 Datasets and Monte Carlo samples

The measurement is performed on 35.9 fb^{-1} of data collected at center-of-mass energy of 13 TeV in LHC Run-II by CMS. The same Monte-Carlo samples are used as in the $e \rightarrow \tau_h$ misidentification rate measurement described in section 7.3, except in here, the signal process is the $Z/\gamma^* \rightarrow \mu\mu$.

7.4.2 Data-to-Monte Carlo corrections

The pileup reweighting is applied to all simulation samples, as in the $e \rightarrow \tau_h$ misidentification rate measurement.

The events are triggered by the single muon trigger "HLT_IsoMu24". The data-to-MC scale factors are derived from an auxiliary study on $Z/\gamma^* \rightarrow \mu\mu$ events using the tag-and-probe method selecting a well isolated and identified muon already fired by the trigger as "tag" and another muon as "probe". The definition of the trigger efficiency is the same as equation 7.12, except that the signal is $Z/\gamma^* \rightarrow \mu\mu$ events. The data-to-MC scale factors are binned in p_T and η of the muon. Figure 7.12 (**left**) shows an example of the single muon trigger efficiencies both in data and in MC, with the scale factors indicated in the lower panel.

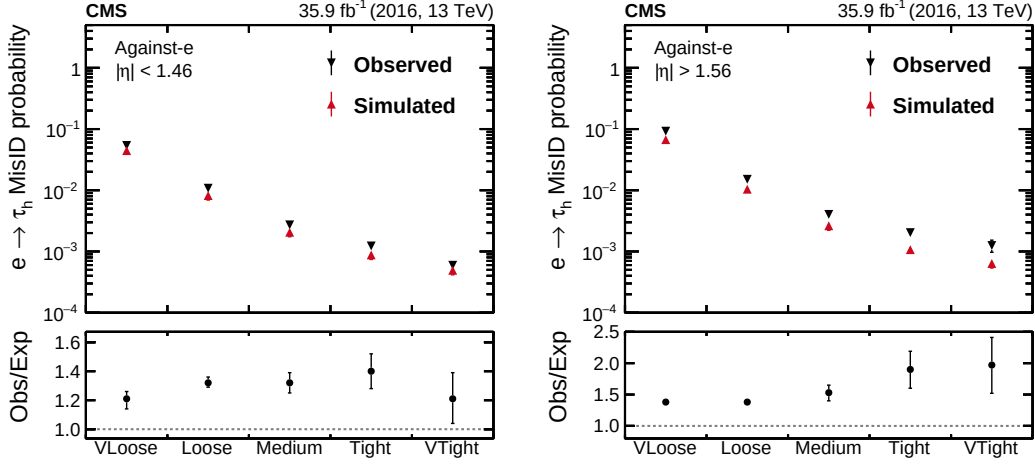


Figure 7.11: Visualization of table 7.3. The misidentification rates are split into the barrel (**left**) and endcaps (**right**) regions [110].

The muons are required to pass the medium identification working point (described in subsection 6.2.4). The muon should be isolated by requiring the relative isolation of muon $I_\mu/p_T^\mu < 0.15$. Data-to-MC scale factors are dedicated to correct the difference between data and simulation due to the muon identification and isolation. The scale factors are derived also from an auxiliary study by using "tag-and-probe" approach on $Z/\gamma^* \rightarrow \mu\mu$ events as the study to obtain the trigger scale factors. The tag is defined as a well identified and isolated reconstructed muon. The probe is a loosely selected reconstructed muon candidate which will be further used to test if it passes or fails the identification and isolation requirements. The efficiency is defined previously as equation 7.12. Figure 7.12 (**right**) shows an example of the muon identification and isolation efficiencies and the corresponding data-to-MC scale factors.

7.4.3 Background modeling

The background modeling method used in the $\mu \rightarrow \tau_h$ misidentification rate measurement is almost identical to the one used in the $e \rightarrow \tau_h$ misidentification rate measurement, except the signal process in here is the $Z\gamma^* \rightarrow \mu\mu$ events and the transverse mass of the tag muon plus missing transverse energy used to define the control region of W+jets is defined as:

$$m_T(\mu, \vec{E}_T) = \sqrt{2p_T^\mu E_T (1 - \cos \Delta\phi)}. \quad (7.19)$$

7.4.4 Event selection

Events are selected by requiring the presence of a well identified and isolated muon (the tag) plus a τ_h candidate (the probe) passing loose identification criteria. In particular, the selections applied on the tag muon are the following:

- $p_T > 26$ GeV and $|\eta| < 2.1$;

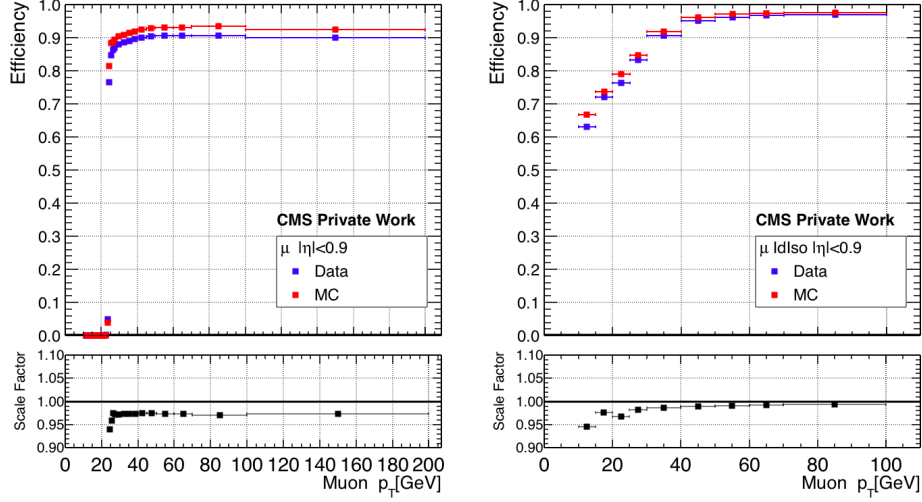


Figure 7.12: The "HLT_IsoMu24" (**left**) single muon trigger efficiencies and the muon identification and isolation efficiencies (**right**) for data and for simulation as a function of reconstructed muon p_T . Examples shown are from region $|\eta| < 0.9$. The data-to-MC scale factors used for the analysis are shown in the lower panel. The last bin in both plots includes the overflowed bin.

- medium identification and isolation $I_\mu < 0.15 \cdot p_T$;
- the impact parameters $d_z < 0.2$ cm, $|d_{xy}| < 0.045$ cm;
- matching with the trigger object in a cone of radius $\Delta R < 0.5$.

The probe, τ_h , is required to satisfy the following criteria:

- $p_T > 20$ GeV and $|\eta| < 2.3$;
- be reconstructed in one of the decay modes $h^\pm$, $h^\pm\pi^0$, $h^\pm\pi^0\pi^0$ or $h^\pm h^\mp h^\pm$;
- pass the tight working point of the MVA-based isolation discriminator;
- be separated from the tag muon by $\Delta R > 0.5$.

The tag and probe pair is required to have opposite electric charge. If an event contains two muons qualifying as tag, both combinations of tag and probe are considered.

The contribution coming from W +jets and $t\bar{t}$ backgrounds, with one jet misidentified as the τ_h , is reduced by requiring the selected events to pass a cut on the transverse mass of the tag muon plus missing transverse energy of $m_T(\mu, \cancel{E}_T) < 30$ GeV.

7.4.5 Measurement

Depending on whether the probe passes or fails the anti-muon discriminator under study, the event enters the "pass" or the "fail" category, respectively. The observed $\mu \rightarrow \tau_h$ misidentification rate is then measured with a simultaneous maximum likelihood fit of the number of $Z/\gamma^* \rightarrow \mu\mu$ events, N_{pass}^{probe} and N_{fail}^{probe} , in both categories. The parameter of interest in the

fit is the data-to-MC scale factor for the $\mu \rightarrow \tau_h$ misidentification rate. The relation between the "pass" and "fail" categories is the same as in $e \rightarrow \tau_h$ misidentification rate measurement.

The visible mass m_{vis} of the tag and probe pair, computed using the energy and the momentum reconstructed by the muon reconstruction algorithm for the tag and the energy and the momentum reconstructed by the HPS algorithm for the probe, is used as discriminant variable for the fit in the pass category. The mass distribution is fitted, in the region $70 \text{ GeV} < m_{vis} < 120 \text{ GeV}$ with templates of the $Z/\gamma^* \rightarrow \mu\mu$ signal and the $Z/\gamma^* \rightarrow \tau\tau \rightarrow \mu\tau_h$, $Z/\gamma^* \rightarrow \tau\tau \rightarrow \mu\mu$, W +jets, $t\bar{t}$, single-top, di-boson (WW, WZ, ZZ) and QCD multijet backgrounds. The fail category is used to constrain the overall normalization of the $Z/\gamma^* \rightarrow \mu\mu$ signal and the fit in this category is performed on the integral distribution of the events, also between 70 GeV and 120 GeV. The mass range is narrower in $\mu \rightarrow \tau_h$ misidentification rate measurement compared to the $e \rightarrow \tau_h$, due to the fact that the resonance of $Z/\gamma^* \rightarrow \mu\mu$ is narrower.

7.4.6 Systematic uncertainties

Systematic uncertainties are represented by nuisance parameters in the maximum likelihood fit. The systematic uncertainties can be grouped into three categories: the uncertainties affecting the yield, the uncertainties affecting the shape of the observable and the Monte-Carlo statistic uncertainty.

The uncertainties affecting the yield are following the log-normal distribution (equation 4.7). They account for the following effects:

- **Luminosity:** the uncertainty on the integrated luminosity is taken to be 2.6%, applied on all MC simulated processes.
- **Muon identification/isolation/trigger:** the combination of the three effects is taken to be 2%, applied on all MC simulated processes.
- **Tau identification:** the uncertainty is taken to be 3%, applied on all MC simulated processes.
- **$t\bar{t}$ cross section:** a 10% uncertainty is attributed to the $t\bar{t}$ sample.
- **Di-boson and single-top cross sections:** the uncertainty on the sum of these small backgrounds is considered to be 10%.
- **W+jets normalization:** a value of 20% is associated to it.
- **QCD normalization:** 20% is associated to it.
- **DY normalization:** an uncertainty of 3% is associated to the Drell-Yan process; an additional 3% uncertainty is associated to the $Z/\gamma^* \rightarrow \mu\mu$ process only, to disentangle possible differences between the $Z/\gamma^* \rightarrow \mu\mu$ and the $Z/\gamma^* \rightarrow \tau\tau$ normalization.

The uncertainties affecting the observable shape are modeled using equation 4.9. They have the following sources:

- **The probe τ_h energy scale:** the energy scale of probe τ_h matched to a generator level τ_h candidates has an uncertainty of $\pm 1.5\%$.

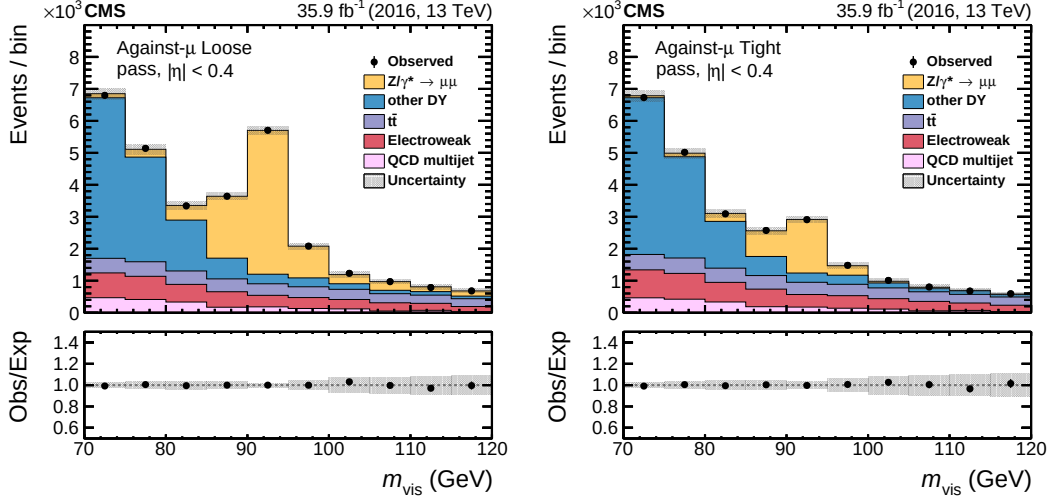


Figure 7.13: The fitted distribution in m_{vis} in the pass category for the loose (**left**) and tight (**right**) working points of the anti-muon discriminator in the $|\eta| < 0.4$ region. [110]. The shaded bands represent the statistic uncertainty due to limited events of MC simulation samples and the systematic uncertainties. The vertical bars (not visible) indicate the statistic uncertainty of data. The electroweak background contribution includes the contributions from W +jets, diboson and single top production processes.

- **The $\mu \rightarrow \tau_h$ energy scale:** the energy scale of probe τ_h matched to muons has an uncertainty of $\pm 3\%$.
- **The visible mass resolution scale:** the visible mass resolution scale has an uncertainty of $\pm 20\%$, applied on signal $Z/\gamma^* \rightarrow \mu\mu$ events. The up and down shapes are defined as equation 7.18.

The bin-by-bin uncertainties as described in subsection 4.1.2, accounting for the limited number of events in every bin of the distributions, are considered for all MC simulated processes.

7.4.7 Results

Separate fits are used for the probe τ_h in the different pseudorapidity region, $|\eta| < 0.4$, $0.4 < |\eta| < 0.8$, $0.8 < |\eta| < 1.2$, $1.2 < |\eta| < 1.7$ and $|\eta| > 1.7$ considering the effects from muons system geometries. The fitted visible mass distributions in the "pass" category, for τ_h in $|\eta| < 0.4$ region, are presented in figure 7.13 for loose and tight working points of the anti-muon discriminator, as examples.

The $\mu \rightarrow \tau_h$ misidentification rate is sensitive to the muon reconstruction quality. In CMS, the "bad" reconstructed muons can be tagged by checking whether there are muon segments that are shared between the inside-out reconstructed muon and the outside-in reconstructed muon. If a inside-out reconstructed muon does not have a shared muon segment with the outside-in reconstructed muon, this PF muon is considered as a "bad" muon. If one filters out the "bad" muons, the $\mu \rightarrow \tau_h$ misidentification rate is lower. Table 7.4 shows

Chapter 7. Performance of τ_h Reconstruction and Identification

the $\mu \rightarrow \tau_h$ misidentification rates after filtering out the "bad" muons and figure 7.14 illustrates the values. The data-to-MC scale factors are recommended to be applied on all CMS analyses using 2016 data involving τ_h reconstructed object and with the "bad" muon filter applied. For τ_h in loose working point, the measured misidentification rates in data exceed the Monte Carlo prediction, with the difference between data and simulation increasing from low pseudorapidity region to high pseudorapidity region. A similar trend is observed for the probes in tight working point.

WP	Misidentification rate (prefit)	Misidentification rate (postfit)	Scale Factor
$ \eta < 0.4$			
Loose	$(3.49 \pm 0.36) \cdot 10^{-3}$	$4.20^{+0.130}_{-0.125} \cdot 10^{-3}$	$1.216^{+0.037}_{-0.036}$
Tight	$(1.02 \pm 0.10) \cdot 10^{-3}$	$1.51^{+0.168}_{-0.107} \cdot 10^{-3}$	$1.474^{+0.163}_{-0.104}$
$0.4 < \eta < 0.8$			
Loose	$(2.85 \pm 0.30) \cdot 10^{-3}$	$3.19^{+0.120}_{-0.123} \cdot 10^{-3}$	$1.120^{+0.042}_{-0.043}$
Tight	$(2.77 \pm 0.29) \cdot 10^{-4}$	$4.30^{+0.828}_{-0.855} \cdot 10^{-4}$	$1.549^{+0.298}_{-0.307}$
$0.8 < \eta < 1.2$			
Loose	$(3.87 \pm 0.40) \cdot 10^{-3}$	$4.86^{+0.143}_{-0.149} \cdot 10^{-3}$	$1.255^{+0.036}_{-0.038}$
Tight	$(2.78 \pm 0.29) \cdot 10^{-3}$	$3.69^{+0.142}_{-0.150} \cdot 10^{-3}$	$1.330^{+0.051}_{-0.054}$
$1.2 < \eta < 1.7$			
Loose	$(9.30 \pm 1.20) \cdot 10^{-4}$	$1.13^{+0.136}_{-0.140} \cdot 10^{-3}$	$1.222^{+0.146}_{-0.151}$
Tight	$(1.70 \pm 0.17) \cdot 10^{-4}$	$2.93^{+0.903}_{-0.916} \cdot 10^{-4}$	$1.721^{+0.530}_{-0.538}$
$ \eta > 1.7$			
Loose	$(4.61 \pm 0.73) \cdot 10^{-4}$	$1.10^{+0.732}_{-0.666} \cdot 10^{-3}$	$2.394^{+0.158}_{-0.144}$
Tight	$(1.66 \pm 0.17) \cdot 10^{-4}$	$4.17^{+1.058}_{-0.963} \cdot 10^{-4}$	$2.501^{+0.634}_{-0.577}$

Table 7.4: Probability for muons to be misidentified as τ_h for the different working points of the cut-based anti- μ discriminator, split in probe pseudorapidity $|\eta|$, applying the "bad" muon filter. For each working point, the $\mu \rightarrow \tau_h$ misidentification rate is defined as the fraction of probes passing the given discriminator with respect to the total number of probes.

The $\mu \rightarrow \tau_h$ misidentification rates and scale factors are also listed in Table 7.5 for the case without using the "bad" muon filter. Compared with the measurement with "bad" muon filter applied, the misidentification rates are higher for all η regions and working points and the data-to-MC scale factors are lower.

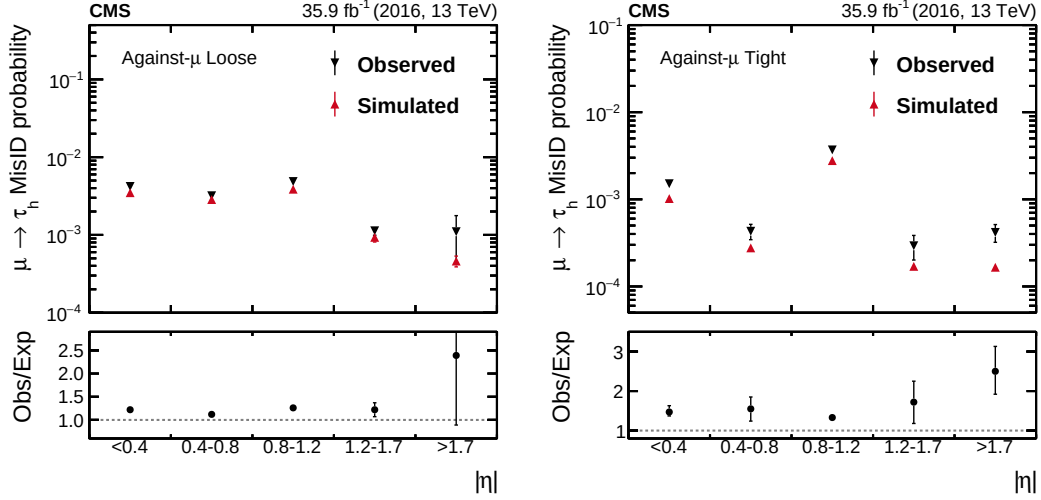


Figure 7.14: Visualization of table 7.4. The misidentification rates are shown for the loose working point (**left**) and the tight working point (**right**) [110].

7.5 Measurement of the Efficiencies of Discriminators against Electron and Muons

The measurements of the efficiencies of the anti-electron discriminator and the anti-muon are described in this section. The tag-and-probe method is used on the $Z/\gamma^* \rightarrow \tau\tau \rightarrow \mu\tau_h$ events, where the events are divided into "pass" and "fail" categories, depending on whether the probe, reconstructed τ_h candidates pass or fail the discriminator under study.

The number of $Z/\gamma^* \rightarrow \tau\tau \rightarrow \mu\tau_h$ signal events in each category is then extracted from a simultaneous maximum likelihood fit of the invariant mass m_{vis} of the tag-probe pair in the two regions, with uncertainties modeled as nuisance parameters of the fit.

7.5.1 Strategies

The datasets and MC samples, data-to-MC corrections, the background modeling techniques, event selections and uncertainties modeling are the same as the $\mu \rightarrow \tau_h$ misidentification rate measurement, except that the signals are $Z/\gamma^* \rightarrow \tau\tau \rightarrow \mu\tau_h$ events, the "bad" muon filter is used, the cut-based anti-muon discriminator tight working point is applied as well as the corresponding data-to-MC scale factors and the mass range 40 – 80 GeV is chosen because of the high signal purity.

7.5.2 Results

The results are given in tables 7.6, 7.7 for the anti-electron and anti-muon discriminator, respectively. The efficiencies measured in data are in good agreement with the MC prediction within the uncertainties of the measurements. The fitted m_{vis} distributions in the "pass" and "fail" categories are presented in figures 7.15, 7.16, 7.17.

Chapter 7. Performance of τ_h Reconstruction and Identification

WP	Misidentification rate (prefit)	Misidentification rate (postfit)	Scale Factor
$ \eta < 0.4$			
Loose	$(3.82 \pm 0.54) \cdot 10^{-3}$	$3.87^{+0.079}_{-0.103} \cdot 10^{-3}$	$1.012^{+0.020}_{-0.026}$
Tight	$(1.12 \pm 0.16) \cdot 10^{-3}$	$1.42^{+0.076}_{-0.082} \cdot 10^{-3}$	$1.263^{+0.068}_{-0.072}$
$0.4 < \eta < 0.8$			
Loose	$(3.03 \pm 0.43) \cdot 10^{-3}$	$3.05^{+0.092}_{-0.095} \cdot 10^{-3}$	$1.007^{+0.030}_{-0.031}$
Tight	$(2.90 \pm 0.41) \cdot 10^{-4}$	$3.96^{+0.765}_{-0.800} \cdot 10^{-4}$	$1.364^{+0.263}_{-0.275}$
$0.8 < \eta < 1.2$			
Loose	$(5.02 \pm 0.72) \cdot 10^{-3}$	$4.36^{+0.086}_{-0.118} \cdot 10^{-3}$	$0.870^{+0.017}_{-0.023}$
Tight	$(3.80 \pm 0.54) \cdot 10^{-3}$	$3.25^{+0.168}_{-0.110} \cdot 10^{-3}$	$0.854^{+0.044}_{-0.028}$
$1.2 < \eta < 1.7$			
Loose	$(9.60 \pm 1.38) \cdot 10^{-4}$	$1.10^{+0.121}_{-0.129} \cdot 10^{-3}$	$1.154^{+0.126}_{-0.135}$
Tight	$(1.69 \pm 0.24) \cdot 10^{-4}$	$2.90^{+0.877}_{-0.902} \cdot 10^{-4}$	$1.712^{+0.516}_{-0.531}$
$ \eta > 1.7$			
Loose	$(4.76 \pm 0.68) \cdot 10^{-4}$	$1.08^{+0.118}_{-0.123} \cdot 10^{-3}$	$2.281^{+0.248}_{-0.258}$
Tight	$(1.66 \pm 0.23) \cdot 10^{-4}$	$3.86^{+0.825}_{-0.843} \cdot 10^{-4}$	$2.324^{+0.496}_{-0.507}$

Table 7.5: Probability for muons to be misidentified as τ_h for the different working points of the cut-based anti- μ discriminator, split in probe pseudorapidity $|\eta|$, without applying the "bad" muon filter. For each working point, the $\mu \rightarrow \tau_h$ misidentification rate is defined as the fraction of probes passing the given discriminator with respect to the total number of probes.

All η regions			
WP	Misidentification rate (pre-fit)	Misidentification rate (post-fit)	Scale Factor
Very Loose	0.914 ± 0.006	$0.932^{+0.003}_{-0.003}$	$1.020^{+0.004}_{-0.004}$
Loose	0.892 ± 0.006	$0.906^{+0.004}_{-0.004}$	$1.016^{+0.005}_{-0.005}$
Medium	0.857 ± 0.009	$0.891^{+0.004}_{-0.004}$	$1.019^{+0.005}_{-0.005}$
Tight	0.814 ± 0.011	$0.832^{+0.004}_{-0.004}$	$1.023^{+0.005}_{-0.005}$
Very Tight	0.771 ± 0.013	$0.788^{+0.004}_{-0.004}$	$1.023^{+0.006}_{-0.006}$

Table 7.6: Probability for τ_h to pass the different working points of the MVA based anti-electron discriminator. For each working point, the efficiency is defined as the fraction of τ_h probes passing the given discriminator with respect to the total number of τ_h probes.

All η regions			
WP	Misidentification rate (pre-fit)	Misidentification rate (post-fit)	Scale Factor
Loose	0.9942 ± 0.0005	$0.952^{+0.077}_{-0.070}$	$0.958^{+0.078}_{-0.071}$
Tight	0.9921 ± 0.0008	$0.953^{+0.077}_{-0.070}$	$0.961^{+0.078}_{-0.071}$

Table 7.7: Probability for τ_h to pass the different working points of the cut-based anti-muon discriminator. For each working point, the efficiency is defined as the fraction of τ_h probes passing the given discriminator with respect to the total number of τ_h probes.

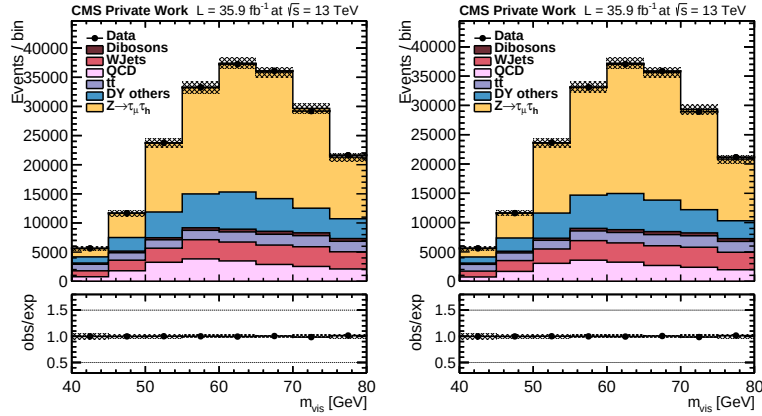


Figure 7.15: The fitted distribution in m_{vis} in the pass category for the loose (**left**) and tight (**right**) working points of the anti-muon discriminator.

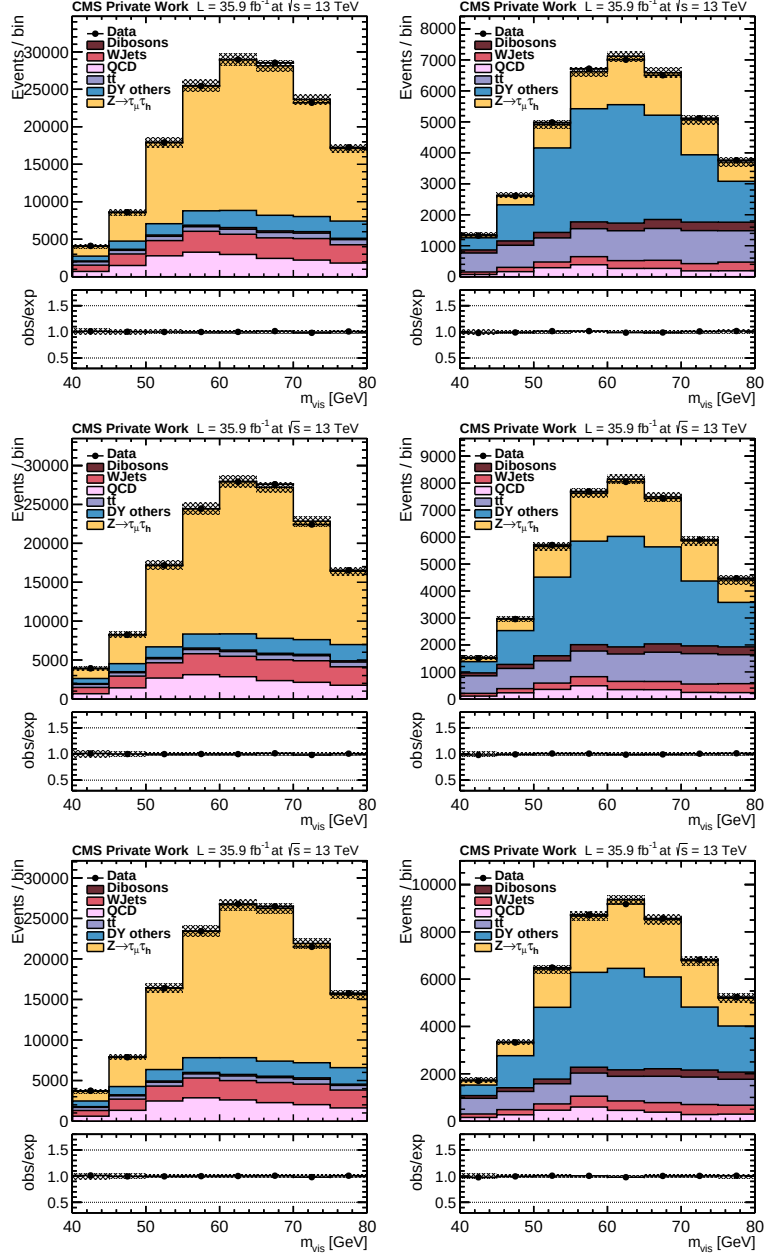


Figure 7.16: Distribution of m_{vis} observed in the pass (left) and fail (right) samples compared to the Monte Carlo expectation, for the very loose (top), loose (middle) and medium (bottom) working points of the MVA-based anti-electron discriminator. The expected m_{vis} distribution is shown for the postfit value to nuisance parameters.

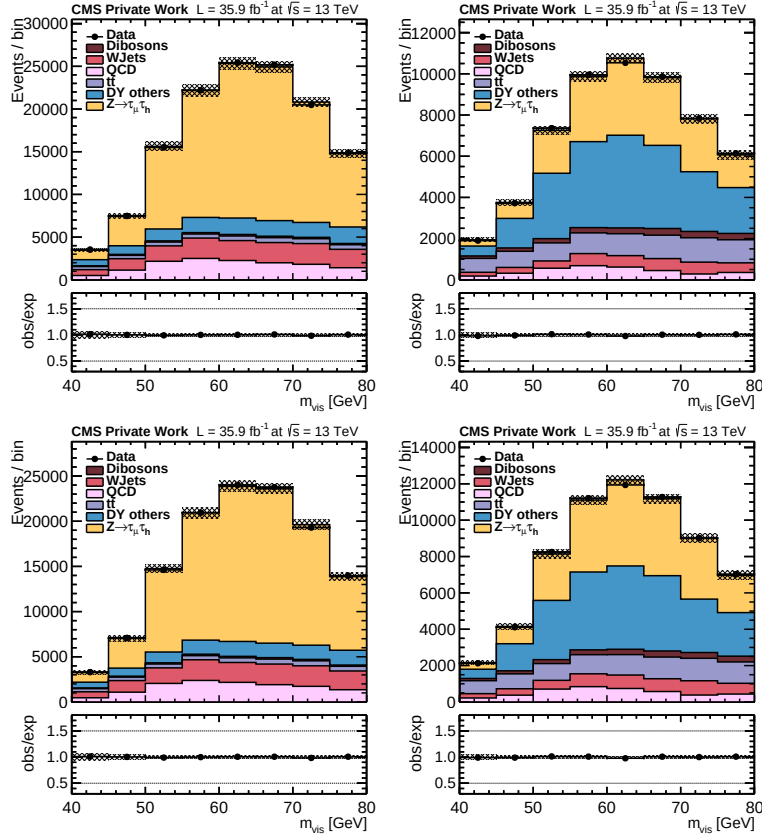


Figure 7.17: Distribution of m_{vis} observed in the pass (**left**) and fail (**right**) samples compared to the Monte Carlo expectation, for the tight (**top**) and very tight (**bottom**) working points of the MVA-based anti-electron discriminator. The expected m_{vis} distribution is shown for the postfit value to nuisance parameters.

CHAPTER

8

ANALYSIS OF SM HIGGS TO $\tau\tau$ IN THE $e\mu$ FINAL STATE

Contents

8.1	Analysis Overview	118
8.2	Datasets and Monte Carlo Samples	118
8.3	Data-to-Monte Carlo Corrections	119
8.3.1	Pileup	119
8.3.2	Triggers	119
8.3.3	Electrons and muon identification and isolation scale factors	122
8.3.4	Recoil corrections	122
8.3.5	Reweighting of LO Madgraph Drell-Yan Z +jets events	123
8.3.6	Top p_T reweighting	123
8.4	Event Selection	124
8.4.1	Baseline selection	124
8.4.2	The D_ζ variable to suppress top pair production background	125
8.4.3	$\tau\tau$ mass reconstruction	126
8.4.4	Event categorization	128
8.5	Background Modeling	129
8.5.1	Drell-Yan Z +jets	129
8.5.2	$t\bar{t}$ events	129
8.5.3	QCD multijet events	130
8.5.4	Other backgrounds	132
8.6	Control Plots	132
8.7	Signal Extraction	132

8.8 Systematic Uncertainties	132
8.8.1 Uncertainties related to object reconstruction and identification . . .	132
8.8.2 Background estimation uncertainties	136
8.8.3 Signal prediction uncertainties	136
8.8.4 Other uncertainties	137
8.9 Results of $e\mu$ Final State	137
8.10 Combination with $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ Final States	137
8.10.1 Brief description of $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states	137
8.10.2 Results	141

8.1 Analysis Overview

The Higgs boson decaying to a τ lepton pair is the most promising channel to study the mass generation mechanism for the fermions, because of the large signal event rate expected in the SM compared to the $\mu\mu$ decay channel, and of the smaller contribution from background events compared to the bb decay channel.

The searches for a Higgs boson decaying to a τ lepton pair have been performed at the LEP [114–117], Tevatron [118, 119] and LHC colliders. At run 1, the CMS experiment reported an evidence for this process with an observed (expected) significance of 3.2 (3.7) σ [17] for a Higgs boson mass of 125 GeV. The ATLAS experiment declared an evidence for the same process with an observed (expected) significance of 4.5 (3.4) σ [120]. The combination of run 1 from both experiments showed an observed (expected) significance of 5.5 (5.0) σ [121].

The $\tau\tau$ pair has six decay channels, $e\mu$, $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, $\mu\mu$ and ee channels. The branching ratios of them are shown in figure 8.1. This thesis is focused on the study of the $e\mu$ channel of the $H \rightarrow \tau\tau$ decay. The smaller branching ratio in the case of the $e\mu$ channels can be counterbalanced by the high efficiency of electron and muon reconstruction, the smaller QCD backgrounds contribution with respect to $\tau_h\tau_h$ channel and absence of the very large $Z/\gamma^* \rightarrow ee$ and $Z/\gamma^* \rightarrow \mu\mu$ background compared to the ee and $\mu\mu$ channels.

The analysis targets both the gluon fusion and the vector boson fusion (VBF) production modes. The results of $e\mu$ channel will be further combined with the results from $\tau_h\tau_h$, $\mu\tau_h$ and $e\tau_h$ channels, covering 94% of all possible $\tau\tau$ decay.

8.2 Datasets and Monte Carlo Samples

The analyzed datasets correspond to an integrated luminosity of 35.9 fb^{-1} , collected in 2016 at a center-of-mass energy of 13 TeV by the CMS detector. Only fully certified data is used.

The MC simulations used to model the significant SM backgrounds in the analysis are listed in table 8.1. The Drell-Yan Z +jets and W +jets samples are generated by MADGRAPH5 in LO accuracy. The single top and $t\bar{t}$ samples are generated by POWHGEV2. The di-boson samples are generated either with MADGRAPH5_AMC@NLO or POWHGEV2 in NLO accuracy. The PYTHIA8, with tune CUETP8M1 is used to model the parton shower,

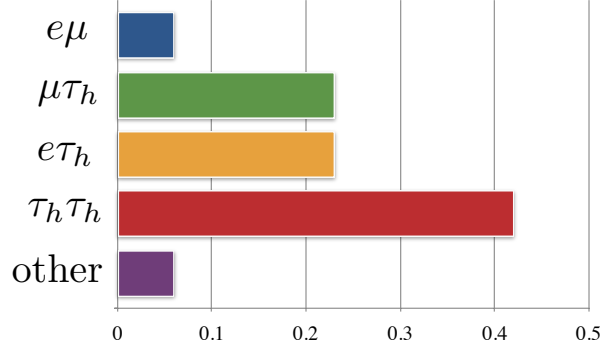


Figure 8.1: The branching ratios of the a pair of τ lepton decay.

hadronization processes and underlying events. The NNPDF3.0 parton distribution functions are used.

MC samples of signals are listed in table 8.2. The signal samples are produced using POWHGEv2. The gluon fusion, vector boson fusion, and W/Z boson associated production modes are considered.

8.3 Data-to-Monte Carlo Corrections

8.3.1 Pileup

Pileup reweighting is applied to the MC simulations in order to bring into consistency distribution in data and in MC, as in $e/\mu \rightarrow \tau_h$ misidentification rate measurements in the previous chapter.

8.3.2 Triggers

The events are required to fire one of the $e + \mu$ cross triggers "HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL*" and "HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL*". Each cross trigger contains one electron and one muon leg. The efficiencies of the legs are determined separately in data and MC using the tag-and-probe technique. The electron legs are estimated from $Z/\gamma^* \rightarrow ee$ events, with the help of a double electron ("HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL") monitoring trigger, sharing the same legs of the trigger used in the analysis. The muon legs are determined from $Z/\gamma^* \rightarrow \mu\mu$ events, with the help of double muon trigger ("HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL"). The exception is the Mu23 leg, whose efficiency is measured from the Mu17 leg placing an additional cut of $p_T > 23$ GeV on the online reconstructed probe muon. The efficiencies are measured in bins of electron or muon p_T and $|\eta|$.

The selections of the tag-and-probe pairs are described as follows. The tag lepton should have the same kinematic, identification and isolation requirements as the analysis itself. It should be matched to a trigger object within a cone of size $\Delta R = 0.5$. The probe lepton should pass the kinematic cuts, identification and isolation requirements. The tag-and-probe pair should be opposite charge and share the same flavor and the invariant mass m_{ll} should be greater than 50 GeV. The matching between the probe and the trigger object, defined

Process	Generator	Cross section (pb)
Drell-Yan $Z/\gamma^*(50) + Jets$	MADGRAPH5_AMC@NLO (LO)	5765.4
Drell-Yan $Z/\gamma^*(50) + 1Jet$	MADGRAPH5_AMC@NLO (LO)	1.16×1012.5
Drell-Yan $Z/\gamma^*(50) + 2Jets$	MADGRAPH5_AMC@NLO (LO)	1.16×332.8
Drell-Yan $Z/\gamma^*(50) + 3Jets$	MADGRAPH5_AMC@NLO (LO)	1.16×101.8
Drell-Yan $Z/\gamma^*(50) + 4Jets$	MADGRAPH5_AMC@NLO (LO)	1.16×54.8
Drell-Yan $Z/\gamma^*(10to50) + Jets$	MADGRAPH5_AMC@NLO (LO)	1.16×18610
Drell-Yan $Z/\gamma^*(150) + Jets$	MADGRAPH5_AMC@NLO (LO)	1.16×6.657
$t\bar{t}$	POWHEGv2	831.8
W +jets inclusive	MADGRAPH5_AMC@NLO (LO)	61526.7
W +1jet	MADGRAPH5_AMC@NLO (LO)	1.221×9644.5
W +2jets	MADGRAPH5_AMC@NLO (LO)	1.221×3144.5
W +3jets	MADGRAPH5_AMC@NLO (LO)	1.221×954.8
W +4jets	MADGRAPH5_AMC@NLO (LO)	1.221×485.6
$VV \rightarrow 2\ell 2\nu$	MADGRAPH5_AMC@NLO (NLO)	11.95
$ZZ \rightarrow 2\ell 2jets$	MADGRAPH5_AMC@NLO (NLO)	3.22
$ZZ \rightarrow 4\ell$	POWHEGv2	1.212
$WW \rightarrow \ell\nu 2jets$	POWHEGv2	49.997
$WZ \rightarrow 2\ell 2jets$	MADGRAPH5_AMC@NLO (NLO)	5.595
$WZ \rightarrow \ell 3\nu$	MADGRAPH5_AMC@NLO (NLO)	3.05
$WZ \rightarrow \ell\nu 2jets$	MADGRAPH5_AMC@NLO (NLO)	10.71
$WZ Jets \rightarrow 3\ell\nu 2jets$	MADGRAPH5_AMC@NLO (NLO)	4.708
single t quark tW channel	POWHEGv2	35.6
single $\bar{t}$ quark tW channel	POWHEGv2	35.6
single t quark t channel	POWHEGv2	136.95
single $\bar{t}$ quark t channel	POWHEGv2	80.95
electroweak $W^- + 2Jets$	MADGRAPH5_AMC@NLO (LO)	20.25
electroweak $W^+ + 2Jets$	MADGRAPH5_AMC@NLO (LO)	25.62
electroweak $Z + 2Jets, Z \rightarrow \ell\ell$	MADGRAPH5_AMC@NLO (LO)	3.987
electroweak $Z + 2Jets, Z \rightarrow \nu\nu$	MADGRAPH5_AMC@NLO (LO)	10.01
gluon fusion $H(125) \rightarrow WW \rightarrow 2\ell 2\nu$	POWHEGv2	1.001
vector boson fusion $H(125) \rightarrow WW \rightarrow 2\ell 2\nu$	POWHEGv2	0.0858
$W\gamma$ production, W decays to $\ell\nu$	MADGRAPH5_AMC@NLO (NLO)	489.0
$W/\gamma^* \rightarrow \ell\nu\mu\mu$	MADGRAPH5_AMC@NLO (LO)	2.793
$W/\gamma^* \rightarrow \ell\nu ee$	MADGRAPH5_AMC@NLO (LO)	3.526

Table 8.1: Monte Carlo background samples included in the analysis. A k-factor of 1.16 is considered for the Z/γ^* +jets samples, and 1.21 for the W +jets samples.

Process	Generator	Production cross section (pb)	$H \rightarrow \tau\tau$ branching ratio
gluon fusion $H(110) \rightarrow \tau\tau$	POWHEGv2	57.90	0.0791
gluon fusion $H(120) \rightarrow \tau\tau$	POWHEGv2	52.22	0.0698
gluon fusion $H(125) \rightarrow \tau\tau$	POWHEGv2	48.58	0.0627
gluon fusion $H(130) \rightarrow \tau\tau$	POWHEGv2	45.31	0.0541
gluon fusion $H(140) \rightarrow \tau\tau$	POWHEGv2	36.00	0.0360
vector boson fusion $H(110) \rightarrow \tau\tau$	POWHEGv2	4.434	0.0791
vector boson fusion $H(120) \rightarrow \tau\tau$	POWHEGv2	4.434	0.0698
vector boson fusion $H(125) \rightarrow \tau\tau$	POWHEGv2	4.434	0.0627
vector boson fusion $H(130) \rightarrow \tau\tau$	POWHEGv2	4.434	0.0541
vector boson fusion $H(140) \rightarrow \tau\tau$	POWHEGv2	4.434	0.0360
$ZH(110) \rightarrow \tau\tau$	POWHEGv2	1.309	0.0791
$ZH(120) \rightarrow \tau\tau$	POWHEGv2	0.994	0.0698
$ZH(125) \rightarrow \tau\tau$	POWHEGv2	0.884	0.0627
$ZH(130) \rightarrow \tau\tau$	POWHEGv2	0.790	0.0541
$ZH(140) \rightarrow \tau\tau$	POWHEGv2	0.651	0.0360
$W^+H(110) \rightarrow \tau\tau$	POWHEGv2	1.335	0.0791
$W^+H(120) \rightarrow \tau\tau$	POWHEGv2	0.9558	0.0698
$W^+H(125) \rightarrow \tau\tau$	POWHEGv2	0.8400	0.0627
$W^+H(130) \rightarrow \tau\tau$	POWHEGv2	0.7414	0.0541
$W^+H(140) \rightarrow \tau\tau$	POWHEGv2	0.6308	0.0360
$W^-H(110) \rightarrow \tau\tau$	POWHEGv2	0.8587	0.0791
$W^-H(120) \rightarrow \tau\tau$	POWHEGv2	0.6092	0.0698
$W^-H(125) \rightarrow \tau\tau$	POWHEGv2	0.5328	0.0627
$W^-H(130) \rightarrow \tau\tau$	POWHEGv2	0.4676	0.0541
$W^-H(140) \rightarrow \tau\tau$	POWHEGv2	0.3940	0.0360

Table 8.2: List of the Monte Carlo signal samples included in the analysis.

by finding a trigger object within a $\Delta R = 0.5$ cone of the probe, is then tested on the probe leptons, that can either pass or fail. The efficiencies are extracted from a fit with an asymmetric gaussian function, modeling the signal, plus an exponential fit, modeling the background, to the invariant mass of the two leptons in the mass window around the Z boson mass, $75 < m_{ll} < 105$ GeV. The same procedure is repeated in bins of $p_T = [10, 15, 20, 25, 30, 40, 60, > 60]$ and $|\eta|$ of the probe lepton. For muons, the binning in $|\eta|$ is $[0.0, 0.9, 1.2, 2.1, 2.4]$ considering the geometries of muon system subdetector, while for electrons the binning is $[0.0, 1.48, 2.1, 2.5]$ exploiting the structures of ECAL.

The scale factor applied to correct the MC samples is $\epsilon_{\text{data}}/\epsilon_{MC}$. In the $e\mu$ channel, the efficiency ϵ is defined by the following combination of efficiencies of the trigger legs:

$$\epsilon = \epsilon(\text{Mu23})\epsilon(\text{Ele12}) + \epsilon(\text{Mu8})\epsilon(\text{Ele23}) - \epsilon(\text{Mu23})\epsilon(\text{Ele23}). \quad (8.1)$$

8.3.3 Electrons and muon identification and isolation scale factors

Electron and muon identification and isolation efficiencies are measured in data and MC using the tag-and-probe method. The scale factor $\epsilon_{\text{data}}/\epsilon_{\text{MC}}$ is applied on an event-by-event basis to correct the simulation. Efficiencies are measured in $Z/\gamma^* \rightarrow ee$ and $Z/\gamma^* \rightarrow \mu\mu$ events, in bins of lepton p_T and $|\eta|$. The efficiency measurement proceeds in a similar way as for the trigger efficiency described in subsection 8.3.2, the main difference is the requirements on the tag and probe leptons. In this case, the tag lepton must satisfy the kinematic, identification and isolation requirements used in the analysis. The probe lepton are only required to pass the kinematic cuts. The requirements on the tag-and-probe pair are unchanged compared to the ones described in subsection 8.3.2. The Identification and isolation requirements used in the analysis are tested on the probe leptons, that can either pass or fail.

8.3.4 Recoil corrections

Recoil corrections are applied to correct for mismodeling of $\vec{E}_T$ in the MC samples of the Drell-Yan Z +jets, W +jets and Higgs signals. The corrections are performed on the variable defined as the vectorial difference of the reconstructed $\vec{E}_T$ and total transverse momentum of neutrinos originating from the decay of Z , W or Higgs boson:

$$\vec{U} = \vec{E}_T - \vec{p}_{T,\nu}. \quad (8.2)$$

In the case of leptonically decaying boson, this variable can be expressed as:

$$\vec{U} = -\vec{p}_{T,B} - \vec{H}_T. \quad (8.3)$$

where $\vec{H}_T$ is the transverse momentum of hadronic recoil and $\vec{p}_{T,B}$ is the transverse momentum of leptonically decaying Z , W or Higgs boson (leptonic recoil).

The vector $\vec{U}$ is projected onto the axes parallel (U_1) and orthogonal (U_2) to the boson p_T and it is measured in $Z/\gamma^* \rightarrow \mu\mu$ events for which the leptonic recoil dose not have neutrinos and the momentum of the Z boson can be measured in good precision. From the measurement, the mean values of the U_1 projection, $\langle U_1 \rangle$, and the resolutions of the U_1 and U_2 projections, $\sigma(U_1)$ and $\sigma(U_2)$, are determined for data and MC events as a function of Z p_T and jet multiplicity.

The corrections to the U_1 and U_2 variables are:

$$U'_1 = \langle U_1 \rangle_{\text{data}} + (U_1 - \langle U_1 \rangle_{\text{MC}}) \frac{\sigma(U_1)_{\text{data}}}{\sigma(U_1)_{\text{MC}}}, \quad (8.4)$$

$$U'_2 = U_2 \frac{\sigma(U_2)_{\text{data}}}{\sigma(U_2)_{\text{MC}}}. \quad (8.5)$$

The corrections are transferred to $\vec{E}_T$ by equation 8.2 and applied to the MC samples of Drell-Yan Z +jets, W +jets and Higgs boson production. Figure 8.2 demonstrates the effect of recoil correction.

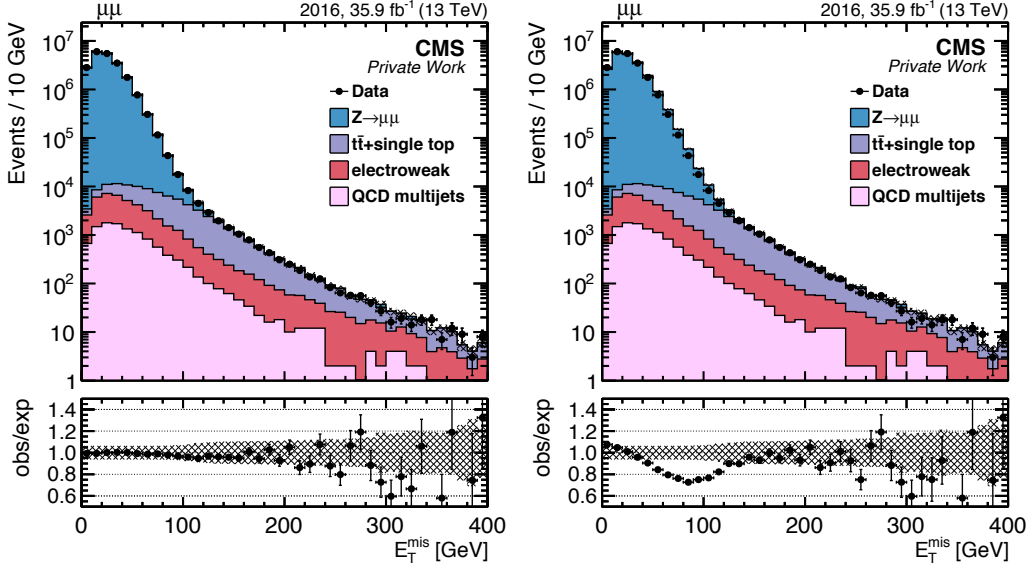


Figure 8.2: The $\cancel{E}_T$ distributions in the $Z \rightarrow \mu\mu$ selection. The one applying the recoil correction (**left**) shows a better data-to-MC agreement than the one without (**right**).

8.3.5 Reweighting of LO Madgraph Drell-Yan Z +jets events

The Drell-Yan Z +jets is estimated with MC simulation. It is the main irreducible background. A reweighting is applied to Drell-Yan Z +jets samples to correct the generator level p_T and generator level invariant mass $m_{\ell\ell}$ distributions in LO MADGRAPH samples. The correction is derived from $Z/\gamma^* \rightarrow \mu\mu$ events, as a function of Z p_T and generator $m_{\ell\ell}$. The generator level Z p_T is defined by summing momentum of generator level muons from $Z/\gamma^* \rightarrow \mu\mu$, electrons and neutrinos either from hard process or from prompt τ lepton decays.

To derive the weights, events are selected in a dimuon control region. In this region, the two opposite sign muons should satisfy the medium muon ID and with a relative isolation less than 0.15. The leading muon is required to have p_T greater than 26 GeV and second leading muon should have p_T greater than 15 GeV. Events with extra identified and isolated electrons and muons are vetoed. The background contributions to the $Z/\gamma^* \rightarrow \mu\mu$ events include $t\bar{t}$, diboson production, single top, W +jets and QCD multijet events. These background events are subtracted from the data. The remaining data events and $Z/\gamma^* \rightarrow \mu\mu$ MC are then normalized and divided in bins of the generator level dimuon invariant mass and generator level p_T to obtain the weights. The weights are further applied in the analysis as a function of generator level Z boson mass and p_T .

8.3.6 Top p_T reweighting

By reweighting the p_T spectrum of the top and anti-top quark, the modeling of the top pair production background is improved in the MC samples. The reweighting method was developed in [122]. Every MC events is assigned a weight:

$$w_{t\bar{t}} = \sqrt{w_t \cdot w_{\bar{t}}}, \quad (8.6)$$

where

$$w_t = \exp(a + bp_T(t)), \quad (8.7)$$

$$w_{\bar{t}} = \exp(a + bp_T(\bar{t})), \quad (8.8)$$

and $p_T(t)$ ($p_T(\bar{t})$) is the transverse momentum of generator level top (anti-top) quarks. The parameter a and b are measured with run 1 data and they are:

$$a = 0.156, \quad (8.9)$$

$$b = -0.00137/\text{GeV}. \quad (8.10)$$

8.4 Event Selection

Particle reconstruction is performed using the particle-flow (PF) algorithm described in subsection 6.2.1 and further high level physics objects are built on top of that. For a given analysis, quality cuts should be placed on relevant physics objects in order to have high yield and purity of signal.

8.4.1 Baseline selection

Electron candidates are reconstructed by the particle flow algorithm. The electron candidates should satisfy the following criteria:

- The electron should pass the tight working point of MVA-based electron identification discussed in subsection 6.2.3. It is defined corresponding to an efficiency of 80%.
- The electron is not from photon conversion.
- The number of missing hits in inner tracker should be less than 1.
- The transverse and longitudinal impact parameters d_{xy} and d_z should be less than 0.45 mm and 2 mm, respectively.
- The $\Delta\beta$ corrected relative isolation I_e , defined as equation 6.5 should be less than 0.15.
- Electron must have $p_T > 13$ GeV and $|\eta| < 2.5$ and matched to the offline electron leg of one of the cross triggers within a cone of $\Delta R < 0.4$.

The muons are reconstructed by the particle flow algorithm. The muon should satisfy the following criteria;

- The muon should pass the medium identification criteria of muon described in subsection 6.2.4.
- The transverse and longitudinal impact parameters d_{xy} and d_z should be less than 0.45 mm and 2 mm, respectively.
- The $\Delta\beta$ corrected relative isolation I_μ , defined as equation 6.5 but with a isolation cone size $\Delta R < 0.4$, should be less than 0.2.

- Muon must have $p_T > 10$ GeV and $|\eta| < 2.4$ and matched to the offline muon leg of one of the cross triggers within a cone of $\Delta R < 0.4$.

The electron should be separated from the muon by requiring $\Delta R(e, \mu) > 0.3$. If more than one $e\mu$ pairs are satisfying the above criteria, the most isolated pair, i.e. with minimum sum of relative isolation variables, $\min(I_e + I_\mu)$, is selected. The electron and muon in the pair should have opposite sign (OS).

The top pair production background events can be further suppressed by vetoing events with at least one b-tagged jet. A b-tagged jet is defined by requiring a jet passing the medium working point of CSVv2 b-tagging algorithm. Figure 8.3 shows the multiplicity of b-tagged jet distributions.

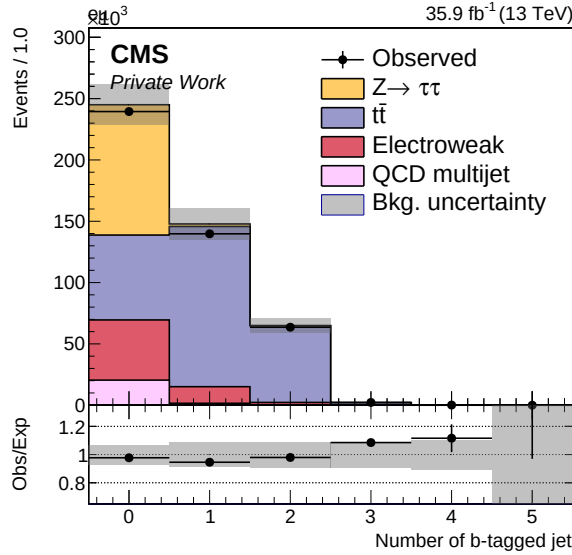


Figure 8.3: The multiplicity of b-tagged jet.

8.4.2 The D_ζ variable to suppress top pair production background

To further suppress the backgrounds from the top pair production and the W+jets events, the D_ζ variable is used. It is calculated as the difference of the projections of visible transverse momentum of τ decay plus $\vec{E}_T$, $\vec{p}_{T,1}^{\text{vis}} + \vec{p}_{T,2}^{\text{vis}} + \vec{p}_T^{\text{mis}}$, and visible transverse momentum of τ decay products on the bisector ζ axis. The bisector between the direction of the electron and that of the muon can be written as

$$\hat{\zeta} = \frac{\vec{\zeta}}{|\vec{\zeta}|} = \frac{\hat{p}_1 + \hat{p}_2}{|\hat{p}_1 + \hat{p}_2|}, \quad (8.11)$$

where $\hat{p}_1 = \vec{p}_{T,1}/|\vec{p}_{T,1}|$ and $\hat{p}_2 = \vec{p}_{T,2}/|\vec{p}_{T,2}|$ are the directions of the electron and muon momentum in the transverse plane of the laboratory frame. The D_ζ variable is defined as

$$D_\zeta = P_\zeta - 1.85 P_\zeta^{\text{vis}} \quad (8.12)$$

with

$$P_\zeta = (\vec{p}_{T,1}^{\text{vis}} + \vec{p}_{T,2}^{\text{vis}} + \vec{p}_T^{\text{mis}}) \frac{\vec{\zeta}}{|\vec{\zeta}|} \quad (8.13)$$

and

$$P_\zeta^{\text{vis}} = (\vec{p}_{T,1}^{\text{vis}} + \vec{p}_{T,2}^{\text{vis}}) \frac{\vec{\zeta}}{|\vec{\zeta}|}. \quad (8.14)$$

The reconstruction of P_ζ variable is shown in figure 8.4. The D_ζ variable quantifies the compatibility of event with the topology, whether the direction of missing neutrinos from τ decays are aligned with the direction of the visible τ decay products. The parameter 1.85 is the optimized value obtained, which gives the best separating power between Higgs signals and top pair production.

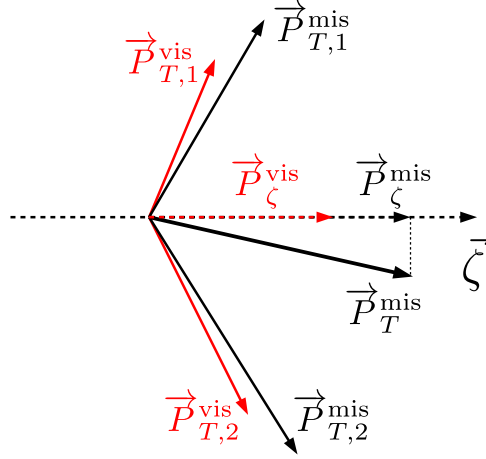


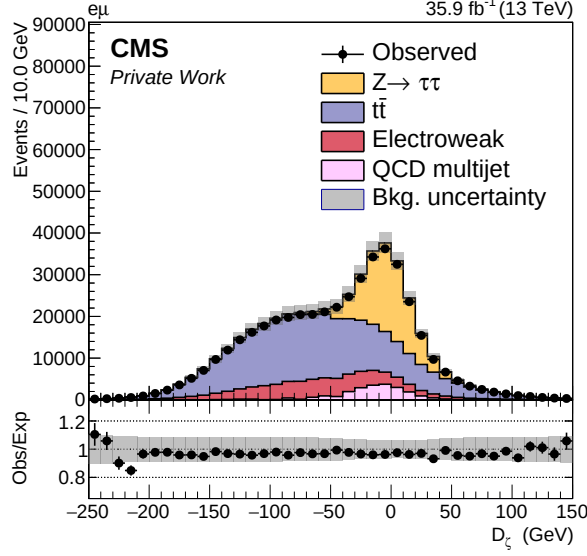
Figure 8.4: The reconstruction of P_ζ variable, from [123].

Figure 8.5 shows the D_ζ variable distributions.

8.4.3 $\tau\tau$ mass reconstruction

The decay of $\tau\tau$ leads to a pair of charged leptons and four neutrinos. The p_T of the four neutrinos cannot be measured individually. Only the sum of p_T of the four neutrinos, the transverse missing energy $\cancel{E}_T$, is accessible. Although the loss of information on each of the τ decay is unrecoverable, it is still possible to find the parent boson of $\tau\tau$ system from a probabilistic approach. The reconstructed mass using such method, denoted as $m_{\tau\tau}$, can provide a better power than the visible mass m_{vis} to separate the peak of the $Z/\gamma^* \rightarrow \tau\tau$ from that of the 125 GeV Higgs $\rightarrow \tau\tau$. In CMS, the SVfit algorithm [17] is used to reconstruct the $\tau\tau$ mass.

The algorithm makes use of a likelihood model that combines the information of visible leptons four-momentum, the components of the $\cancel{E}_T$ and their covariance matrix, in each τ decays. The free parameters for the two τ lepton decays are written as $\vec{a}_1 = (x_1, \phi_1, m_{\nu\nu,1})$ and $\vec{a}_2 = (x_2, \phi_2, m_{\nu\nu,2})$, where x is the fraction of τ lepton energy carried by the visible lepton, ϕ refers to the azimuthal angle of the τ lepton and $m_{\nu\nu}$ represents the invariant

Figure 8.5: The D_ζ variable distributions.

mass of the neutrinos. The visible leptons four-momentum are $\vec{y} = (\vec{p}_1^{\text{vis}}, \vec{p}_2^{\text{vis}})$ and the vector corresponding to the components and the covariance matrix of $\vec{\mathcal{E}}_T$ is $\vec{z} = (\mathcal{E}_x, (\mathcal{E}_y, \mathbf{V}))$. The probability for a given invariant mass hypothesis on the $\tau\tau$ system $m_{\tau\tau}^i$ is

$$P(m_{\tau\tau}^i | \vec{y}, \vec{z}) = \int \delta(m_{\tau\tau}^i - m_{\tau\tau}(\vec{y}, \vec{a}_1, \vec{a}_2)) \cdot f(\vec{z}, \vec{y}, \vec{a}_1, \vec{a}_2) d\vec{a}_1 d\vec{a}_2. \quad (8.15)$$

In the above equation, $f(\vec{z}, \vec{y}, \vec{a}_1, \vec{a}_2)$ is the product of three likelihood functions. Two of them are corresponded to the two τ leptons decay, and relies only on τ mass m_τ and $m_{\nu\nu}$:

$$\mathcal{L}_\tau = \frac{d\Gamma}{dx dm_{\nu\nu} d\phi} \propto \frac{m_{\nu\nu}}{4m_\tau^2} [(m_\tau^2 + m_{\nu\nu}^2)(m_\tau^2 - m_{\nu\nu}^2)]. \quad (8.16)$$

The third likelihood function combines the information of $\mathcal{E}_T$ and its covariance matrix, the momentum of the visible leptons and the free parameter $\vec{a}_1, \vec{a}_2$, which affects the expression from the sums of neutrino momentum $\sum p_x^\nu$ and $\sum p_y^\nu$:

$$\mathcal{L} = \frac{1}{2\pi\sqrt{|\mathbf{V}|}} \exp \left[-\frac{1}{2} \begin{pmatrix} \mathcal{E}_x - \sum p_x^\nu \\ \mathcal{E}_y - \sum p_y^\nu \end{pmatrix}^T \mathbf{V}^{-1} \begin{pmatrix} \mathcal{E}_x - \sum p_x^\nu \\ \mathcal{E}_y - \sum p_y^\nu \end{pmatrix} \right]. \quad (8.17)$$

Figure 8.6 shows the normalized distributions of m_{vis} and $m_{\tau\tau}$ for simulated $Z/\gamma^* \rightarrow \tau\tau$ and $H \rightarrow \tau\tau$, where the mass of the Higgs boson is assumed to be 125 GeV. The mass reconstruction using the SVfit algorithm provides a better separation between signal and background than m_{vis} alone.

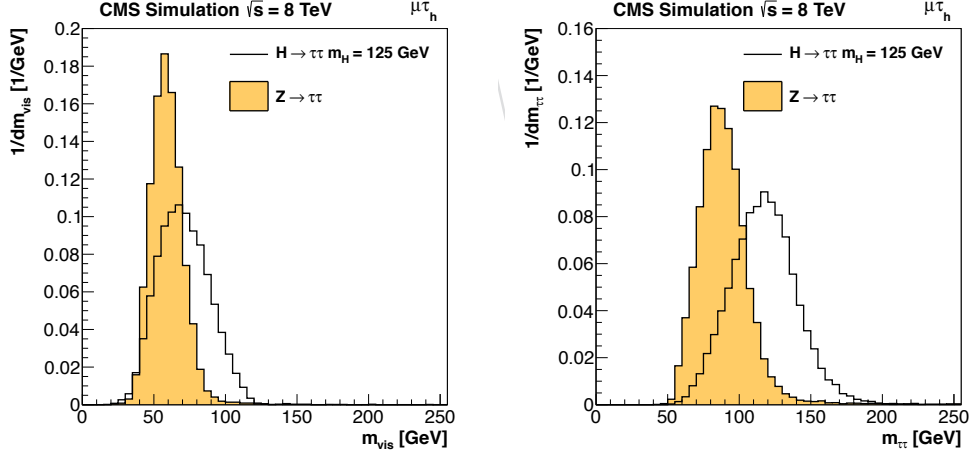


Figure 8.6: Normalized distribution for the invariant mass of the visible decay products of the two τ leptons, m_{vis} (**left**), and the SVFit reconstructed mass, $m_{\tau\tau}$. The distribution from simulated $Z/\gamma^* \rightarrow \tau\tau$ events (yellow histogram) is used to compare the distribution from simulated signal events with a SM Higgs boson with a mass of 125 GeV (open histogram).

8.4.4 Event categorization

The events are split into three mutually exclusive categories. In each category the two variables that maximize the $H \rightarrow \tau\tau$ sensitivity are chosen to build a two dimensional (2D) distribution.

The three categories are defined as:

- **0-jet:** This category targets the events produced via gluon fusion production. The gluon fusion has the largest cross section for the SM Higgs boson at the LHC, but this category contains a large amount of background, including a large contribution of irreducible Drell-Yan Z +jets events. This category is important to constrain the different backgrounds with large statistics. The two variables used to extract the signals are m_{vis} and the p_T of the muon. The m_{vis} variable is used as discriminant in the fit instead of $m_{\tau\tau}$ because it separates the signals from the $Z \rightarrow \ell\ell$ ($\ell = e, \mu$) background, which peaks around the Z boson mass. Signal events typically have leptons with higher p_T than other backgrounds, especially Drell-Yan events, which is why the transverse momentum of the muon are used as second dimension in the fit. The D_ζ variable is required to be greater than -35 GeV to suppress the top pair production background.
- **VBF:** This category focuses the events produced via vector boson fusion production. Events are selected with exactly two jets with $p_T > 30$ GeV. The two leading jets are required to have an invariant mass, m_{jj} , greater than 300 GeV. The D_ζ variable should be greater than -10 GeV. The selection together can suppress the background from SM events composed of QCD multijet events. The two discriminants in the VBF category used for signal extraction are $m_{\tau\tau}$ and m_{jj} .
- **Boosted:** This category has all the events that do not fall into the previous categories, namely events with one jet and events with several jets that fail the requirements of

VBF category. It also contains gluon fusion events produced in association with one or more jets, VBF events where one of the jets has escaped detection or with low m_{jj} , as well as Higgs bosons produced in association with a W or a Z boson decaying hadronically. $m_{\tau\tau}$ is selected as one of the dimensions for signal extraction. $p_T^{\tau\tau}$, defined as the magnitude of the vectorial sum of the $\vec{p}_T$ of the visible decay products of the τ leptons and $\cancel{E}_T$, is used as the second dimension to target Higgs boson events produced in gluon fusion, with a Lorentz-boosted boson recoiling against jets. To be more precise, in this category, events should have $D_\zeta > -35$ GeV, the number of jets should be either greater than two, or equals to one or equals to two but failing the m_{jj} cut in the VBF category.

The categories and the variables used for 2D signal extraction are summarized in Table 8.3.

0-jet	VBF	Boosted
Selection		
No jet and $D_\zeta > -35$ GeV	2 jets, $m_{jj} > 300$ GeV and $D_\zeta > -10$ GeV	Others
Observables used in 2D signal extraction		
$(p_T^\mu, m_{\text{vis}})$	$(m_{jj}, m_{\tau\tau})$	$(p_T^{\tau\tau}, m_{\tau\tau})$

Table 8.3: Categorization and observables used to build 2D kinematic distributions for signal selection. The events neither selected in the 0-jet nor in the VBF categories are included in the boosted category, indicated as "Others".

8.5 Background Modeling

8.5.1 Drell-Yan Z +jets

The Drell-Yan Z +jets background is modeled by the MC simulated samples. A shape correction related to the generator level Z mass and p_T as discussed in 8.3.5 is applied.

Scale factors are applied in order to better model the Drell-Yan background events. The scale factors are obtained from $Z/\gamma^* \rightarrow \mu\mu$ enriched control region. The purity is very high by requiring the visible mass of the muons to be between 70 and 110 GeV. A threshold of 50 GeV on both muons are required. The events are split into the same categorization as signal region. The data-to-MC ratio in these categories are derived as the scale factors. They are inclusively for the 0-jet category, and binned in m_{jj} or $p_T^{\tau\tau}$ for VBF or boosted category, respectively.

8.5.2 $t\bar{t}$ events

The top pair production is estimated from MC simulated samples. Top p_T reweighting is applied to improve the modeling of top pair kinematics as described in subsection 8.3.6. The correction can be validated in the top pair production enriched control region, defined as $D_\zeta < -35$ GeV and $\cancel{E}_T > 80$ GeV. The distribution of m_{vis} before and after applying the top p_T reweighting is shown in figure 8.7. The correction flattens the data-to-MC ratio as

a function of m_{vis} , proving the improvement in estimating the shape of top pair production background.

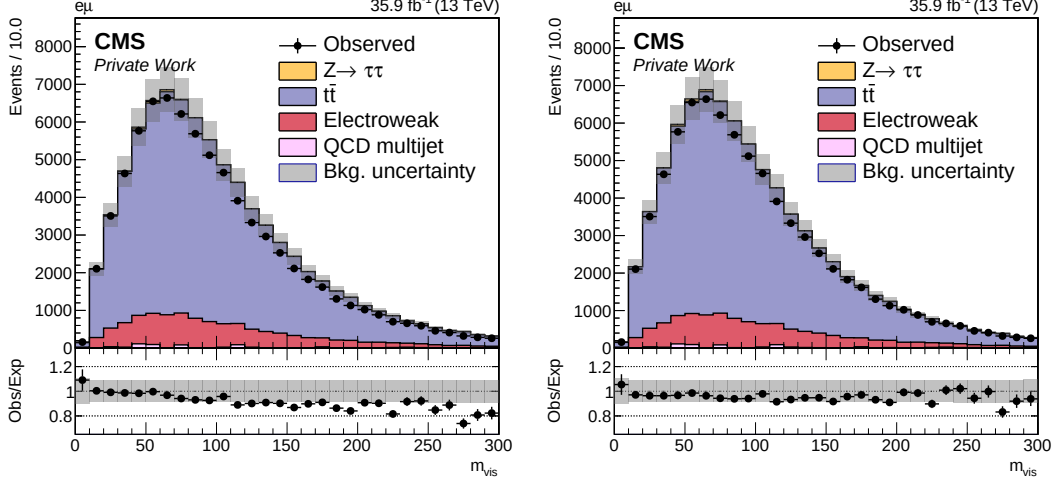


Figure 8.7: Effects of the top quark p_T reweighting on the distribution of m_{vis} in the top pair production enriched region. The reweighting (**right**) shows an improved agreement in the shape of the distribution between data and MC.

8.5.3 QCD multijet events

The contribution of the QCD multijet background in the signal region (SR), where the electron and the muon are required to be of opposite signed (OS) electric charge, is estimated using a data driven technique from a control region, where the electron and the muon are of the same signed (SS) electric charge. The ratio of events number of $e\mu$ pair with OS to the events number of $e\mu$ with SS is measured in sideband (SB) region where the events fail the nominal lepton relative isolation $I_e < 0.15$ and $I_\mu < 0.2$. This ratio is the so-called OS/SS ratio and found to be varied as a function of the isolation criteria that are applied in the isolation sideband (SB). The SB region is defined by requiring that both leptons satisfy a relaxed isolation criterion of $I_{e,\mu} < 0.5$ and at least one lepton fails the nominal isolation requirement, i.e, $I_e > 0.15$ or $I_\mu > 0.2$, to prevent overlapping with the signal region. The ratio measurement is performed in the event sample of selected $e\mu$ pairs with requirements of $D_\zeta > -35$ GeV and the presence of no b-tagged jet to reject a large amount of background contamination from top pair production, W +jets and diboson. The data yields in both OS region and SS region have to be subtracted from the simulated non-QCD contribution. The OS/SS ratio is computed as:

$$\frac{\text{OS}}{\text{SS}} = \frac{N_{\text{data}}^{\text{OS}} - N_{\text{MC}}^{\text{OS}}(\text{Non-QCD})}{N_{\text{data}}^{\text{SS}} - N_{\text{MC}}^{\text{SS}}(\text{Non-QCD})}. \quad (8.18)$$

The OS/SS ratios from the SB region determined with data and with simulation are listed in table 8.4. The OS/SS ratio from the SR region determined with simulation are also listed.

The measurement is performed in the SB individually for each event category presented

Region	Data	MC
SB	2.41 ± 0.02	2.70 ± 0.19
SR	-	2.49 ± 0.30

Table 8.4: The OS/SS ratio measured in isolation sideband (SB) region determined from data and from simulation. The OS/SS ratio measured in signal region (SR) determined from simulation is also presented.

in table 8.3. The ratios measured in SB are extrapolated to SR using extrapolation factors determined from simulation for the inclusive selection:

$$F_{\text{iso}} = \frac{(\text{OS/SS})_{\text{MC}}^{\text{SR}}}{(\text{OS/SS})_{\text{MC}}^{\text{SB}}}, \quad (8.19)$$

where $(\text{OS/SS})_{\text{MC}}^{\text{SR}}$ and $(\text{OS/SS})_{\text{MC}}^{\text{SB}}$ are the corresponding OS/SS ratio obtained from SB and SR, determined from the simulated sample of QCD multijet events as listed in table 8.4. Due to the limited statistics of QCD MC sample, it is not possible to determine the extrapolation factor F_{iso} in the boosted and VBF categories.

The OS/SS ratio in SR for each event category is determined by the following equation:

$$(\text{OS/SS})_{\text{data}}^{\text{SR,cat}} = F_{\text{iso}} \cdot (\text{OS/SS})_{\text{data}}^{\text{SB,cat}}, \quad (8.20)$$

where the $(\text{OS/SS})_{\text{data}}^{\text{SB,cat}}$ is the OS/SS ratio obtained from data in SB according to equation 8.18.

The OS/SS ratio for each category in SB and SR are shown in table 8.5. The yield of QCD events in SR are determined by the following equation:

$$N_{\text{QCD, cat}} = (N_{\text{SS, cat}}^{\text{data}} - N_{\text{SS, cat}}^{\text{non-QCD}}) \cdot (\text{OS/SS})_{\text{data, cat}}. \quad (8.21)$$

Event category	$(\text{OS/SS})_{\text{SB}}$	$(\text{OS/SS})_{\text{SR}}$
0-jet	2.46 ± 0.02	2.26 ± 0.31
boosted	2.45 ± 0.04	2.25 ± 0.31
VBF	3.08 ± 0.47	2.83 ± 0.58

Table 8.5: The OS/SS ratio measured in the isolation sideband (SB) region determined from data and the OS/SS ratio in the SR computed as equation 8.20 for different event categories. The uncertainties include the statistic uncertainties in the SB and the extrapolation uncertainty which also has a statistic source.

The SS signal regions with isolated electron and muons have a large amount of non-QCD events. The backgrounds contamination are from Drell-Yan, W +jets, top pair production and diboson events ranging from 40% in the 0-jet category up to 70% in boosted category. A dedicated control sample is used to model the shape of QCD multijet. By doing this, the contribution of the non-QCD processes can be reduced and the statistics in the QCD

template used in signal extraction can be increased. The control sample is defined by having a relaxed isolation requirements on leptons: $0.2 < I_\mu < 0.5$ and $I_e < 0.5$ in the VBF category and $0.1 < I_\mu < 0.3$ and $I_e < 0.3$ in the 0-jet and boosted categories. The purity of QCD multijet events in these regions ranges from 10% in the 0-jet category to 20% in the boosted category.

The QCD shape templates derived in the control sample in each category, are normalized to the yields computed according to equation 8.21. To prove the shape templates derived from relaxed isolation region are compatible with the shape templates in nominal isolation region, the important variable distributions in the three events categories are checked, as shown in figure 8.8.

8.5.4 Other backgrounds

W +jets, diboson and single top background processes are fully modeled from MC simulation.

8.6 Control Plots

The data-to-MC agreement is checked with a set of control plots. Some representative distributions obtained after cutting $D_\zeta > -35$ GeV are shown in figure 8.9. The visible mass pre-fit distributions are illustrated in figure 8.10 for all three categories. A good pre-fit agreement is achieved. For all the plots, only statistical uncertainty is shown and indicated as the grey shaded area.

8.7 Signal Extraction

Results are extracted by a simultaneous maximum likelihood fit of the three categories. The systematic uncertainties, discussed in the next section, are treated as nuisance parameters in the fit.

Along with the signal regions, a control region is included in the fit as well. The control region is designed to have enriched top pair production events. The control region is defined with the same baseline selection, except that the cut on D_ζ variable is inverted: $D_\zeta < -35$ GeV. To reduce the Drell-Yan background in this region, the visible mass of the electron and the muon is required to be greater than 90 GeV. The diboson background is reduced by having a least one jet in the final state. The purity of top pair production events in this region is 85%. Only one bin is used for the control region distribution. No additional uncertainty is appointed to the extrapolation from this single control region to the three categories, because the top p_T reweighting uncertainty has already covered the uncertainties in this region.

8.8 Systematic Uncertainties

8.8.1 Uncertainties related to object reconstruction and identification

A full description can be found in [126]. A list of these uncertainties in $e\mu$ channel is summarized. The uncertainties related to the physics object reconstruction and identification have the following sources:

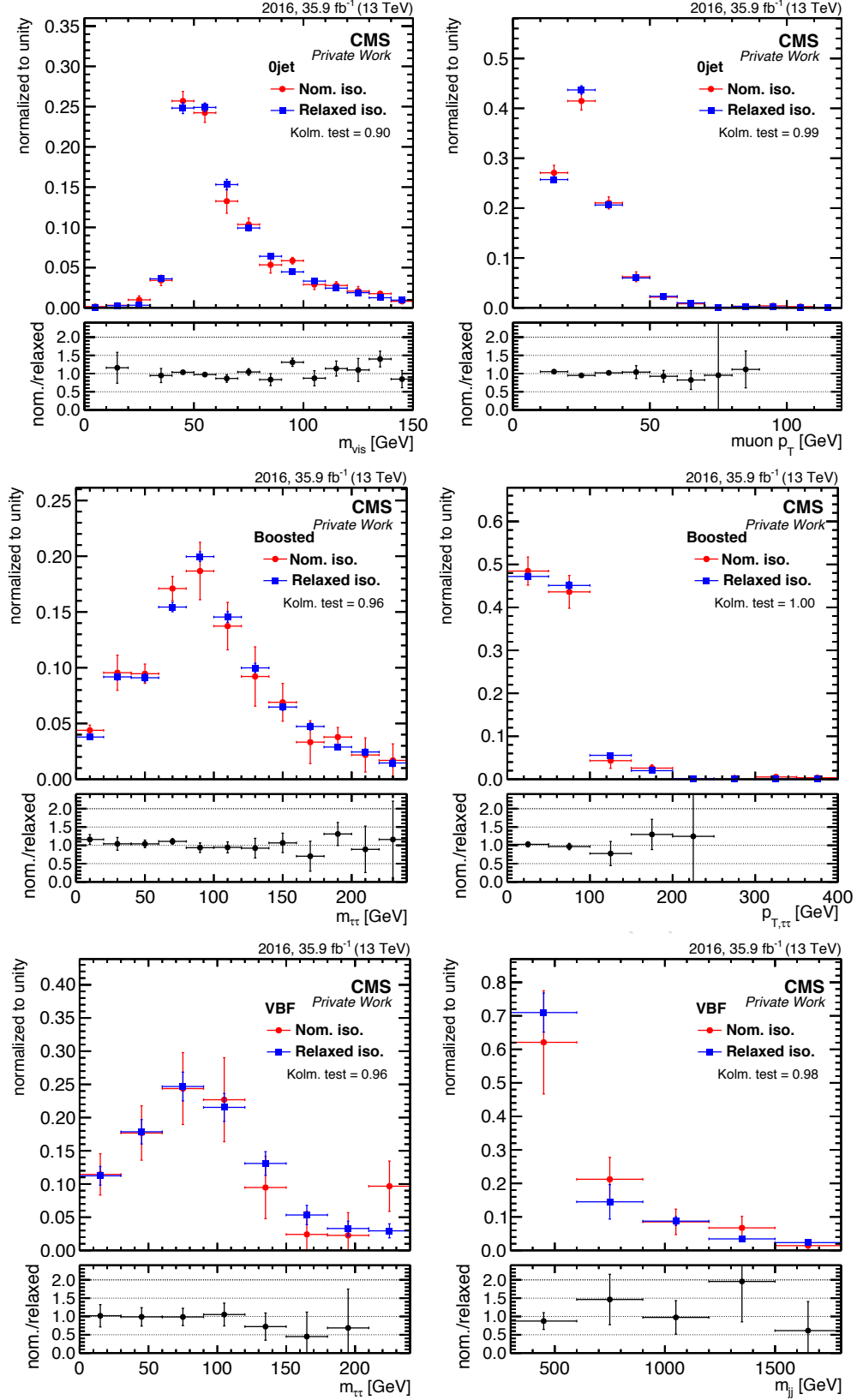


Figure 8.8: The distributions of variables used in the 2D fit signal extraction for QCD multijet events. The distributions show a good compatibility between using templates from nominal isolation region (red) to model QCD events and using templates from relaxed isolation region (blue). Using templates from relaxed isolation region to model QCD events has a smaller statistical uncertainties.

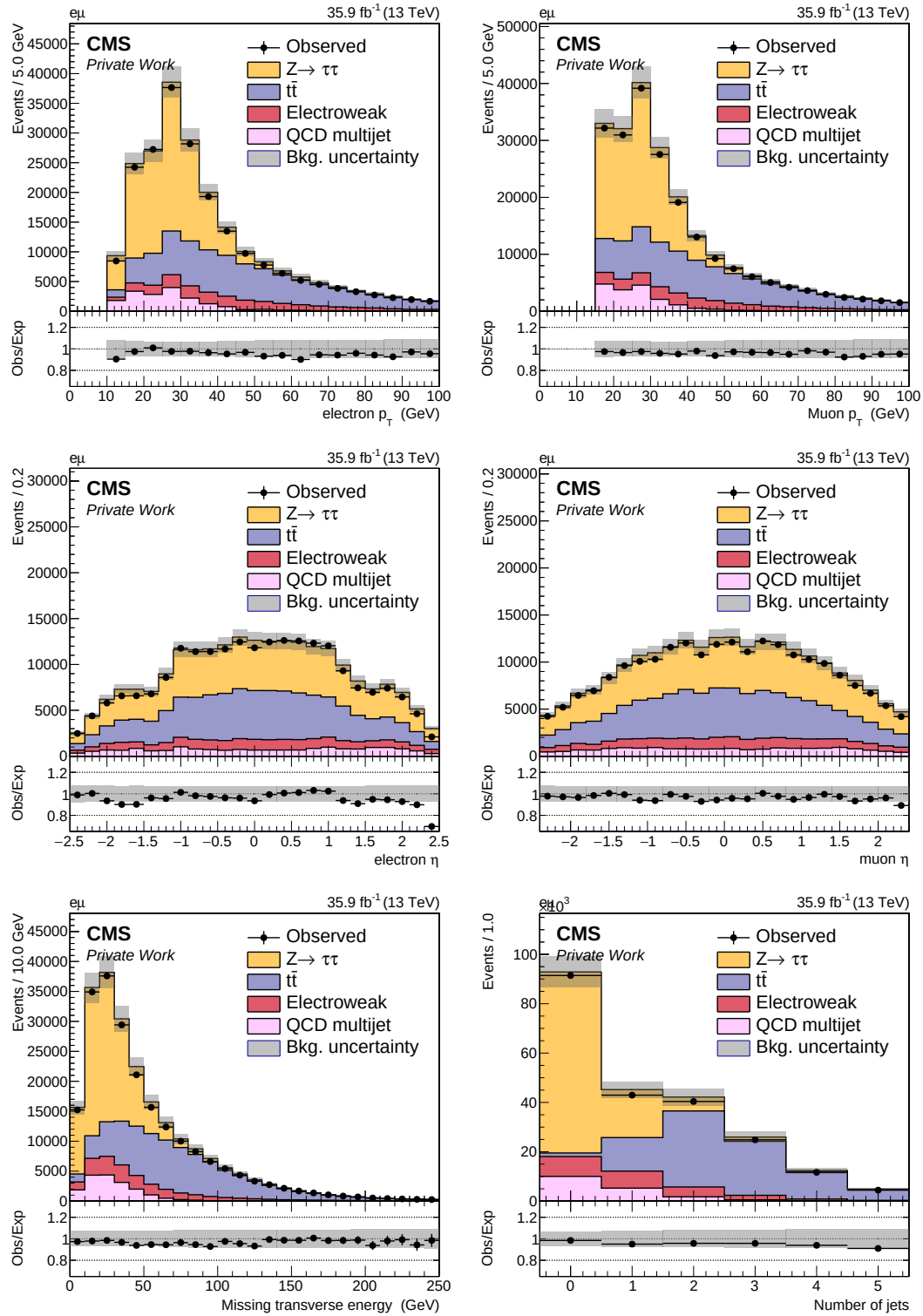


Figure 8.9: Control plots of $e\mu$ channel after applying the baseline selection and the requirement of $D_\zeta > -35$ GeV.

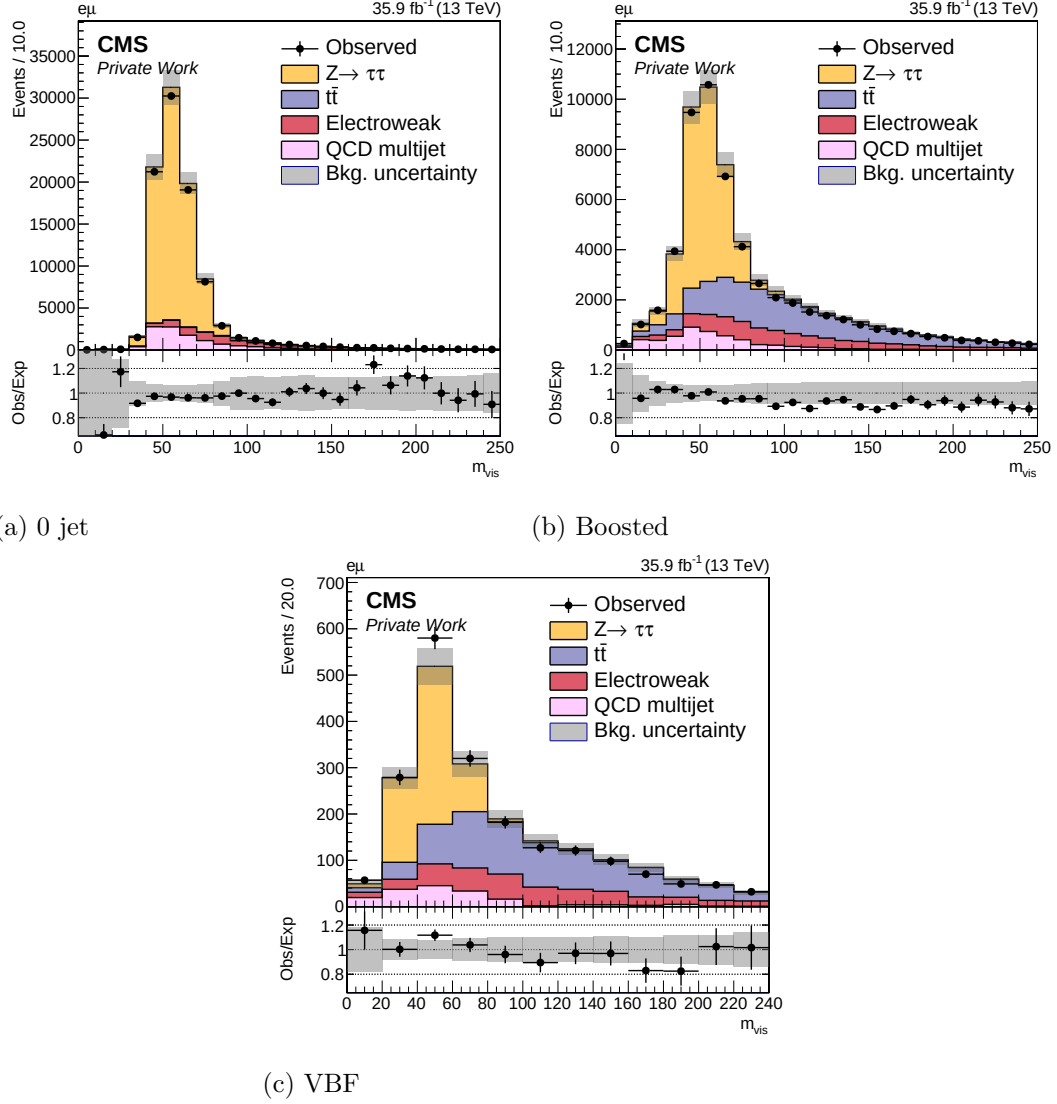


Figure 8.10: Pre-fit visible mass distributions for all three categories.

- **Muon/Electron isolation and identification:** The uncertainty related to the isolation and identification of muons amounts to 2%. It is applied to all MC simulated samples. The same 2% uncertainty goes to the electron isolation and identification efficiency.
- **Electron energy scale:** The electron energy scale is propagated through the SVFit mass calculation. 1% in the barrel and 2.5% in the endcaps are considered.
- **Jet energy scale:** The uncertainties in the jet energy scale depend on the p_T and η of the jet [104]. They are propagated to the computation of the jet multiplicity and to the computation of m_{jj} .
- **Missing transverse energy scale:** The $\cancel{E}_T$ plays an important role via the D_ζ variable in the $e\mu$ channel. The effect of the $\cancel{E}_T$ unclustered energy uncertainty on the

yields of the different processes is evaluated by shifting the $\cancel{E}_T$ within its uncertainties and recomputing the acceptance. The total effect is of the order of 1% for all processes and categories. The uncertainty is considered as flat.

- **B-tagging uncertainty:** A b-tagged jet veto is applied in the $e\mu$ channel. The b-tagging uncertainty for top pair production events in the 0-jet, boosted, VBF categories and top pair production control region, is respectively 3.5, 5, 5 and 5%. The mistagging uncertainty for the diboson background is 1.5% in the boosted and VBF categories. The uncertainty for other processes, including signal, is negligible.

8.8.2 Background estimation uncertainties

The uncertainties related to the background estimation methods have the following sources:

- **W+jets estimation:** The W+jets background is small and it is estimated purely from MC samples. An uncertainty of 4% is assigned to the cross section of the process.
- **QCD estimation:** The uncertainty comes from the OS/SS ratio scale factor, which is between 10 and 20% depending on the category.
- **Top p_T reweighting:** A shape uncertainty, between no and twice the correction on the reweighting, applied to top pair production events.
- **Single top and diboson cross section:** An uncertainty of 5% is assigned to the cross section of single top and diboson background processes, based on the CMS measurements [124, 125].
- **Z p_T reweighting:** A shape uncertainty of Drell-Yan events is computed by applying 0.9 or 1.1 time the correction.
- **Drell-Yan estimation:** Yield uncertainties for Drell-Yan MC sample, based on the data-to-MC agreement observed in the $Z/\gamma^* \rightarrow \mu\mu$ events, and on the variations of the data-to-MC ratio by changing the muon p_T threshold used for the selection in $\mu\mu$ region. It ranges from 7 to 20%, depending on the category.

8.8.3 Signal prediction uncertainties

The rate and acceptance uncertainties for the signal processes related to the theoretical calculations are due to the uncertainties from PDFs, from QCD renormalization and factorization scales, and from parton showering modeling. The following sources of uncertainties are considered:

- **Signal cross section from PDF and α_s :** PDF and α_s uncertainties on signal cross section are 3.2% for gluon fusion production, 2.1% for VBF production, 1.9% for WH production and 1.6% for ZH production.
- **Signal acceptance:** The acceptance uncertainties related to the selection criteria used are less than 1% for the gluon fusion and VBF production modes for the PDF uncertainties. The acceptance uncertainties for the VBF production in the renormalization

and factorization scale uncertainties are also less than 1%. The acceptance uncertainties for gluon fusion production mode from renormalization and factorization scale are treated as shape uncertainties, with a dependence on the $p_T^{\tau\tau}$ and m_{jj} .

- **Parton shower uncertainty:** The changes with variation of the parton shower tune are considered as additional uncertainties, and reach up to 7% in the boosted category.
- **$H \rightarrow \tau\tau$ branching ratio:** The theoretical uncertainty in the branching ratio of the Higgs boson to $\tau\tau$ is equal to 2.1%.

8.8.4 Other uncertainties

The uncertainty in the integrated luminosity is 2.5% [112]. The uncertainty is applied to the signal samples and backgrounds that are estimated using MC simulation.

Uncertainties related to the limited number of simulated events, or to the finite number of events in data control regions, are taken into account. They are considered as bin-by-bin uncertainty which allows each bin to vary within its statistical uncertainty.

8.9 Results of $e\mu$ Final State

The search for an excess of SM Higgs boson events over the expected background makes use of a global maximum likelihood fit based on 2D distributions. The post-fit 2D distributions are shown for all three categories in the figures 8.11, 8.12 and 8.13. The choice of the binning is driven by the statistical precision of the background and data templates, resulting in wider bins in the relatively poor-populated VBF category. The signal prediction for a Higgs boson with $m_H = 125$ GeV is normalized to its best fit signal strength. The backgrounds are also adjusted to the results of the maximum likelihood fit.

The systematic uncertainties are modeled as nuisance parameters that are varied in the fit according to their probability density functions (detailed in 4). A log-normal probability density function is assumed for systematics affecting the event yields. The systematic uncertainties that affect the shape of the distribution are modeled with a linear extrapolation method. In total, the statistical uncertainty in the data is the dominant source of uncertainty.

8.10 Combination with $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ Final States

The results from $e\mu$ final states are further combined with other $\tau\tau$ final states, namely $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states, together with the control regions for top pair production, W +jets and QCD multijet backgrounds.

8.10.1 Brief description of $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states

The kinematic selection criteria for $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states are summarized in table 8.6. In the $e\tau_h$ and $\mu\tau_h$ channels, the large W +jets background is significantly reduced by placing a cut on the transverse mass m_T , to satisfy $m_T < 50$ GeV.

The event categorization used to perform the 2D fit are summarized in table 8.7.

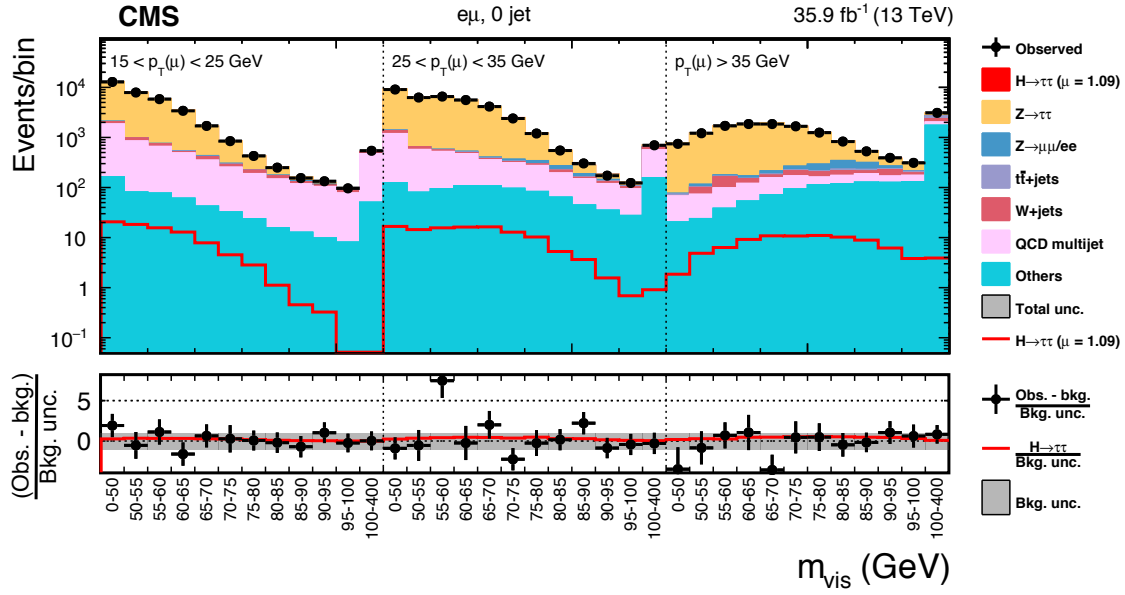


Figure 8.11: Observed and predicted 2D distribution in the 0-jet category of the $e\mu$ final states. The normalization of the predicted background distributions corresponds to the result of the maximum likelihood fit. The signal distribution is scaled to its best fit signal strength value [126].

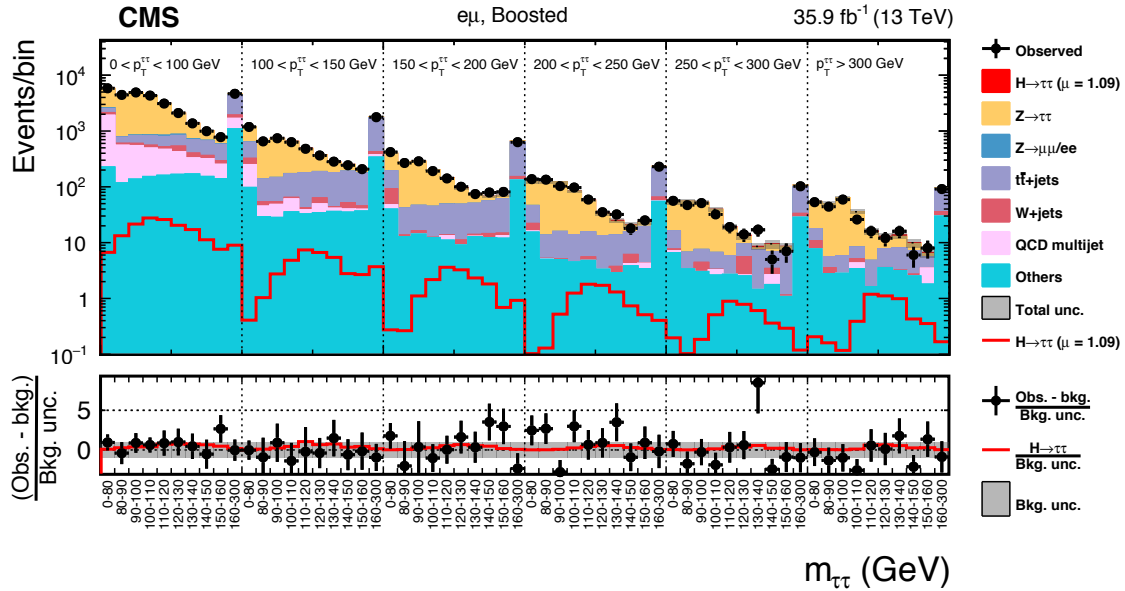


Figure 8.12: Observed and predicted 2D distribution in the boosted category of the $e\mu$ final states [126].

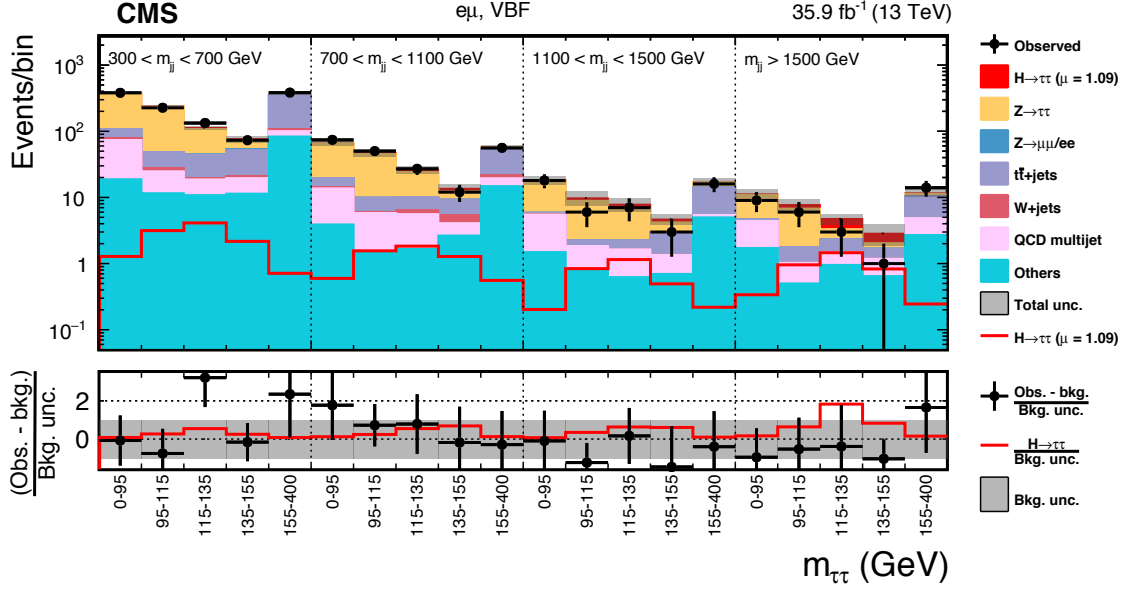


Figure 8.13: Observed and predicted 2D distribution in the VBF category of the $e\mu$ final states [126].

Final states	Trigger requirement	Lepton selection		
		p_T (GeV)	η	Isolation
$\tau_h\tau_h$	$\tau_h(35)\&\tau_h(35)$	$p_T^{\tau_h} > 50\&40$	$ \eta ^{\tau_h} < 2.1$	MVA τ_h tight ID
$\mu\tau_h$	$\mu(22)$	$p_T^\mu > 23$	$ \eta ^\mu < 2.1$	$I_\mu < 0.15$
		$p_T^{\tau_h} > 30$	$ \eta ^{\tau_h} < 2.3$	MVA τ_h tight ID
	$\mu(19)\&\tau_h(21)$	$20 < p_T^\mu < 23$	$ \eta ^\mu < 2.1$	$I_\mu < 0.15$
		$p_T^{\tau_h} > 30$	$ \eta ^{\tau_h} < 2.3$	MVA τ_h tight ID
$e\tau_h$	$e(25)$	$p_T^e > 26$	$ \eta ^e < 2.1$	$I_e < 0.1$
		$p_T^{\tau_h} > 30$	$ \eta ^{\tau_h} < 2.3$	MVA τ_h tight ID

Table 8.6: Trigger requirements and kinematic selections for $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states. The trigger requirement is defined by a combination of trigger candidates with p_T over a given threshold (in GeV), indicated inside the parentheses. The kinematic selections have been optimized to increase the significance of the analysis.

	0-jet	VBF	Boosted
Selection			
$\tau_h\tau_h$	No jet	≥ 2 jets, $p_T^{\tau\tau} > 100$ GeV, $\Delta\eta_{jj} > 2.5$	Others
$\mu\tau_h$	No jet	≥ 2 jets, $m_{jj} > 300$ GeV, $p_T^{\tau\tau} > 50$ GeV, $p_T^{\tau_h} > 40$ GeV	Others
$e\tau_h$	No jet	≥ 2 jets, $m_{jj} > 300$ GeV, $p_T^{\tau\tau} > 50$ GeV	Others
Observables used in 2D signal extraction			
$\tau_h\tau_h$	$m_{\tau\tau}$	$(m_{jj}, m_{\tau\tau})$	$(p_T^{\tau\tau}, m_{\tau\tau})$
$\mu\tau_h$	$(\tau_h \text{ decay mode}, m_{\text{vis}})$	$(m_{jj}, m_{\tau\tau})$	$(p_T^{\tau\tau}, m_{\tau\tau})$
$e\tau_h$	$(\tau_h \text{ decay mode}, m_{\text{vis}})$	$(m_{jj}, m_{\tau\tau})$	$(p_T^{\tau\tau}, m_{\tau\tau})$

Table 8.7: Categorization and observables used to build 2D kinematic distributions for signal selection for $\tau_h\tau_h$, $\mu\tau_h$ and $e\tau_h$ final states. The events neither selected in the 0-jet nor in the VBF categories are included in the boosted category, indicated as "Others".

The background estimation of Drell-Yan Z +jets event in $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channel is the same as in $e\mu$ channel as well as applying scale factors from $Z \rightarrow \mu\mu$ control region and $Z p_T$ mass reweighting.

The W +jets production are estimated using MC simulation in $e\tau_h$ and $\mu\tau_h$ final states, while their yields are normalized in a data-driven way by using W +jets enriched control regions. The control regions are defined in the same way as the signal regions, except that the b-tagged jets are vetoed to reduce top pair production and the transverse mass cut m_T is required to be greater than 80 GeV. The control regions are categorized to 0-jet and boosted category. The constraints obtained in the boosted categories are propagated to the VBF categories. Only one bin per distribution is considered, as shown in figure 8.14. The control regions are fitted simultaneously with the signal regions. In the $\tau_h\tau_h$ final state, the W +jets background is small compared to other backgrounds, so its contribution is estimated from MC simulations.

The QCD multijet background estimation in the $e\tau_h$ and $\mu\tau_h$ final states shares the same idea as the $e\mu$ final state. Shapes are taken from a same sign relaxed lepton isolation region, normalization factors come from same sign reverted lepton isolation region and extrapolation factors are taking into account for the extrapolation from nominal isolation region to relaxed isolation region. The difference is that control regions are added in the fit to better constrain the normalization. The control regions are defined in the same way as the signal regions, except that the muon or electron isolations is inverted. One control region for 0-jet category and one for boosted category per final state. The constraints of boosted categories are propagated to the VBF categories. The QCD control regions are shown in figure 8.15 for the $e\tau_h$ and $\mu\tau_h$ final states.

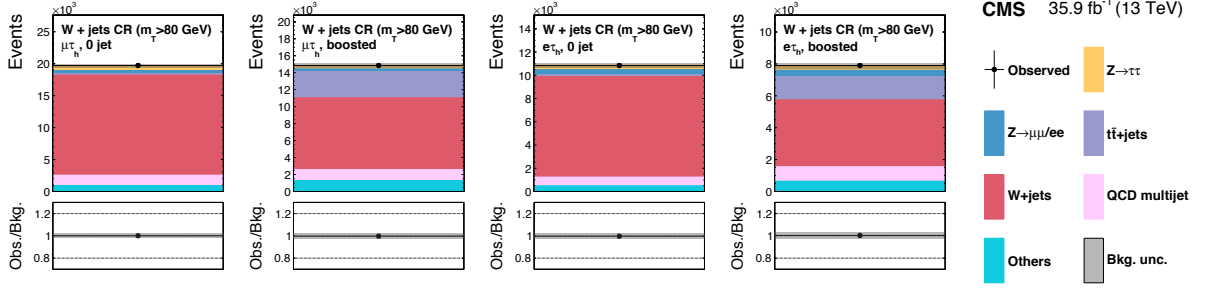


Figure 8.14: W+jets background enriched control regions used in the maximum likelihood fit, together with the signal regions and other control regions, to extract the results. The control regions, defined with $m_T > 80$ GeV, control the yields of the W+jets background. The constraints obtained in the boosted categories are propagated to the VBF categories of the corresponding channels [126].

8.10.2 Results

The 2D distributions of the final observables for each category and each final state in the signal regions, together with the control regions, are fed into the global simultaneously maximum likelihood fit. The expected number of signal events is the predicted SM Higgs boson of mass $m_H = 125.09$ GeV decaying into $\tau\tau$, scaled by the best fit signal strength μ .

A way to visualize an excess for the observed events with respect to the background expectation is to fill events in the signal region by the decimal logarithm of the ratio of the signal (S) to signal-plus-background (S+B) to a binned histogram as shown in figure 8.16. An excess of observed events compared to the SM backgrounds is clearly visible in the most sensitive bins. The expected background and signal contributions, as well as the observed number of events, are listed per final state in table 8.8 for the bins with $\log_{10}(S/(S+B)) > -0.9$.

Another way to visualize the excess of observed events relative to the background expectation is shown in figure 8.17, where the mass distribution is a sum of all mass distributions from all categories and all final states, except that those categories that use m_{vis} are not included. These mass distributions have a weight of $S/(S+B)$ to increase the contribution of the most sensitive distributions.

The excess can be further quantified by calculating the local p -value using a profile likelihood ratio test statistic, as discussed in chapter 4. Figure 8.18 shows the local p -value and the corresponding significance for a SM Higgs boson with a mass of 125.09 GeV. For this particular mass of Higgs boson, the observed significance is 4.9 standard deviations (σ), while the expected significance is 4.7 standard deviations.

The best fit value of signal strength μ is $1.09^{+0.27}_{-0.26}$ at $m_H = 125.09$ GeV. The uncertainty of the best fit signal strength can be decomposed into four sources: the theoretical uncertainties, bin-by-bin Monte Carlo statistic uncertainties, other systematic uncertainties, and the statistical uncertainty:

$$\mu = 1.09^{+0.10}_{-0.08}(\text{theo})^{+0.13}_{-0.12}(\text{bin-by-bin})^{+0.16}_{-0.15}(\text{syst})^{+0.15}_{-0.15}(\text{stat}). \quad (8.22)$$

The best fit signal strengths for each final states and for each category is shown in figure 8.19.

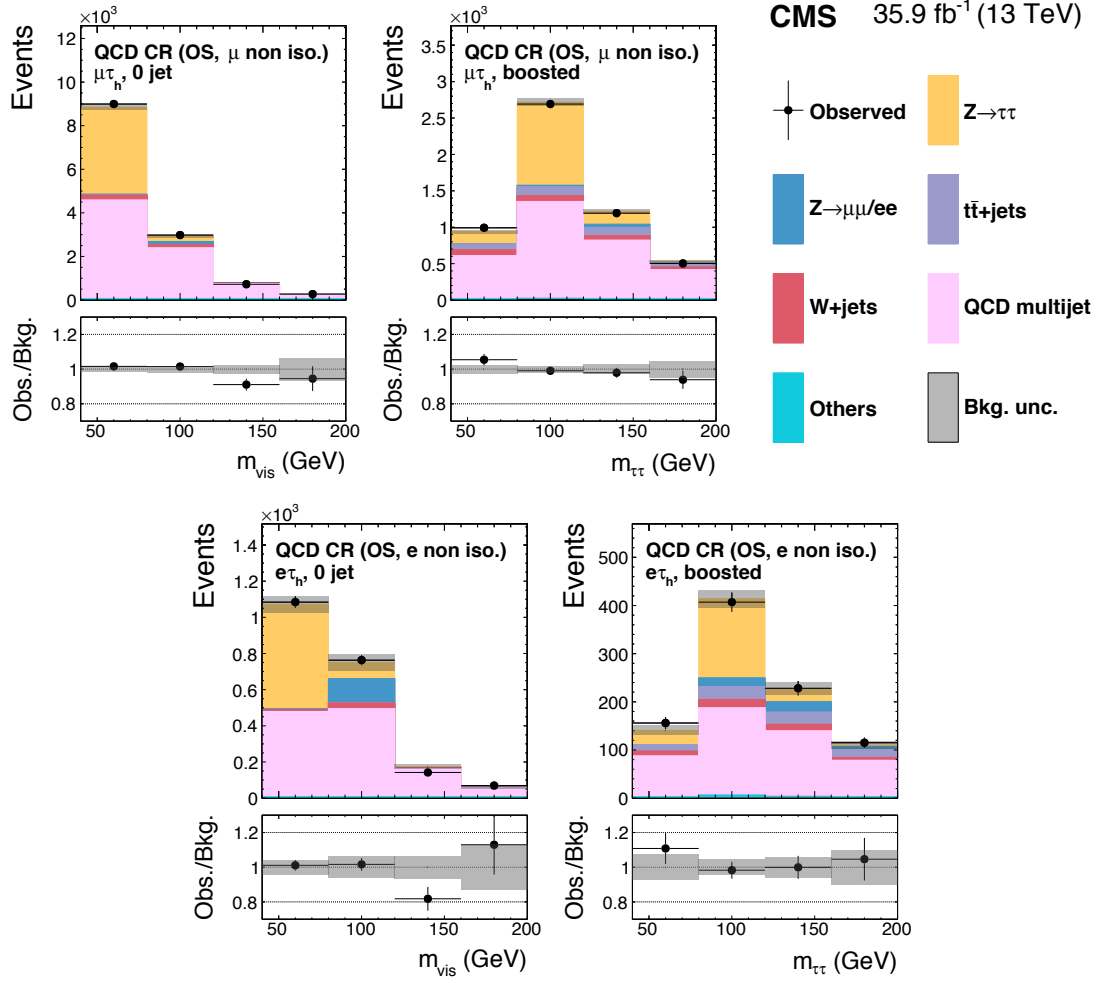


Figure 8.15: QCD control regions used in the maximum likelihood fit, along with other signal regions, to extract the results. The normalization of the background prediction are determined by the fit. The control region are defined by selecting events with inverted isolation of lepton. The yields of QCD are controlled by the fit. The constraints obtained in the boosted categories are propagated to the VBF categories [126].

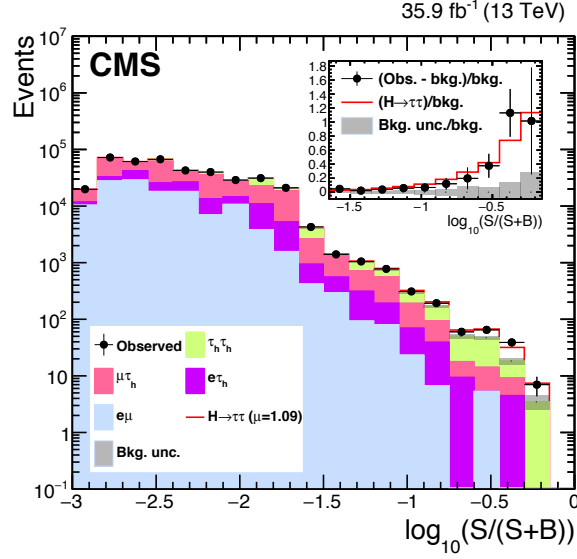


Figure 8.16: Distribution of the ratio between the expected signal and the sum of expected signal and expected background in each bin of the mass distribution used to extract the results, in all four final states. The inset shows the difference between the observed data and background prediction divided by the background prediction. Also shown is the signal prediction divided by the background prediction [126].

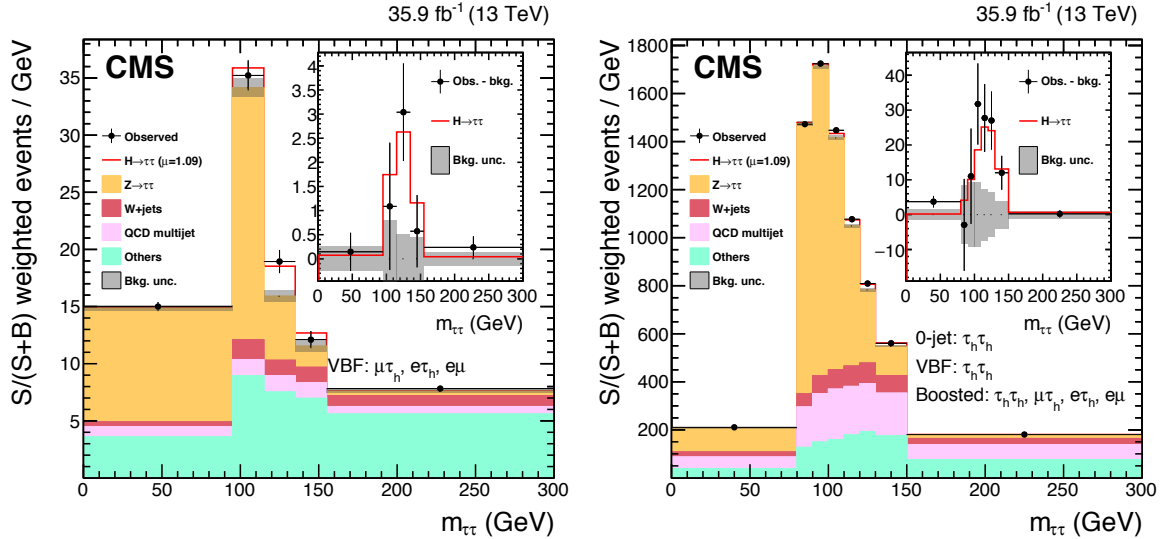


Figure 8.17: The combined observed and predicted $m_{\tau\tau}$ distribution. The left plot includes the VBF category of the $\mu\tau_h$, $e\tau_h$ and $e\mu$ final states, and the right plot includes all other final states that using $m_{\tau\tau}$ as discriminant for the fit. The mass distribution are weighted according to $S/(S+B)$, where S is the signal contribution and B is the background contribution in the mass distribution, excluding the underflow and overflow bins [126].

Process	$e\mu$	$e\tau_h$	$\mu\tau_h$	$\tau_h\tau_h$
$Z/\gamma^* \rightarrow \tau\tau$	5.8 ± 2.2	21.2 ± 3.3	34.6 ± 4.9	89.1 ± 6.9
$Z/\gamma^* \rightarrow ee/\mu\mu$	0.0 ± 0.0	2.9 ± 0.2	3.7 ± 0.2	5.0 ± 0.2
top pair production	1.9 ± 0.1	10.4 ± 0.3	22.2 ± 1.8	13.9 ± 0.5
W +jets	0.8 ± 0.02	4.0 ± 0.3	6.6 ± 1.3	7.6 ± 0.8
QCD multijet	2.1 ± 0.3	3.3 ± 2.5	5.0 ± 1.3	35.5 ± 2.1
Other background	1.4 ± 0.1	5.2 ± 0.2	6.1 ± 0.2	7.3 ± 0.2
gluon fusion, $H \rightarrow \tau\tau$	0.6 ± 0.1	5.0 ± 0.6	6.0 ± 0.6	27.4 ± 2.1
VBF, $H \rightarrow \tau\tau$	2.8 ± 0.3	5.1 ± 0.5	12.55 ± 1.0	17.5 ± 1.0
WH/ZH, $H \rightarrow \tau\tau$	0.0 ± 0.0	0.3 ± 0.0	0.2 ± 0.0	1.3 ± 0.1
Total backgrounds	12.1 ± 2.2	46.5 ± 4.1	77.7 ± 5.5	156.2 ± 7.3
Total signal	3.4 ± 0.4	10.9 ± 0.8	19.2 ± 1.4	48.3 ± 2.6
Observed	11	54	91	207

Table 8.8: Background and signal expectation, along with the number of observed events, for the bins in the signal region for which $\log_{10}(S/(S+B)) > -0.9$, where S and B are, respectively, the number of expected signal events for a Higgs boson with a mass of 125.09 GeV and of expected background events, in those bins.

These signal strengths make use of the constraints obtained from the global simultaneously fit.

A likelihood scan in the (κ_V, κ_f) parameter space is performed for $m_H = 125.09$ GeV, as shown in figure 8.20, where κ_V is the ratio of the measured value for the coupling of the Higgs boson to vector bosons to the SM prediction of that value and κ_f is the ratio of the measured value for coupling of the Higgs boson to fermions to the SM prediction of that value. The method is described in [17]. A compatibility is shown between the observed likelihood contour and the SM expectation value. The best fit value is close to the unity as SM predicted.

The results are combined with the results of the search for Higgs boson decaying to $\tau\tau$ with data collected with the CMS at center-of-mass energies of 7 and 8 TeV during Run 1 [17]. This results in a 5.9 standard deviations observed and expected significance. The uncertainties are treated as uncorrelated between the different center-of-mass energies. The best fit signal strength μ is 0.98 ± 0.18 for a Higgs boson with a mass of $m_H = 125.09$ GeV.

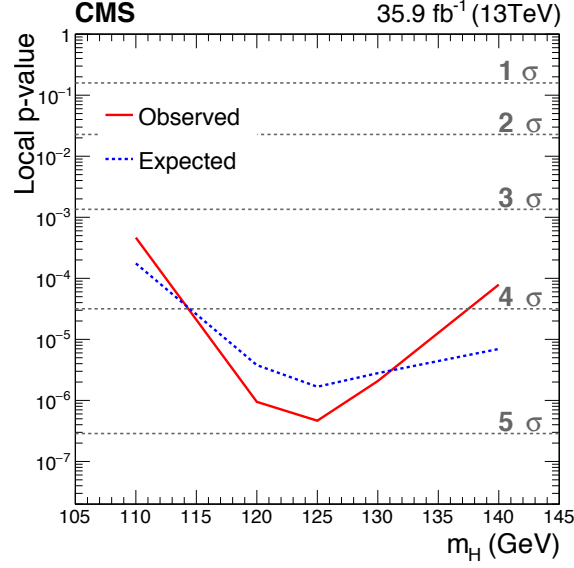


Figure 8.18: Local p -value and the corresponding significance as a function of the SM Higgs boson mass assumption. The red curve represents the observation, while the blue curve is for expectation [126].

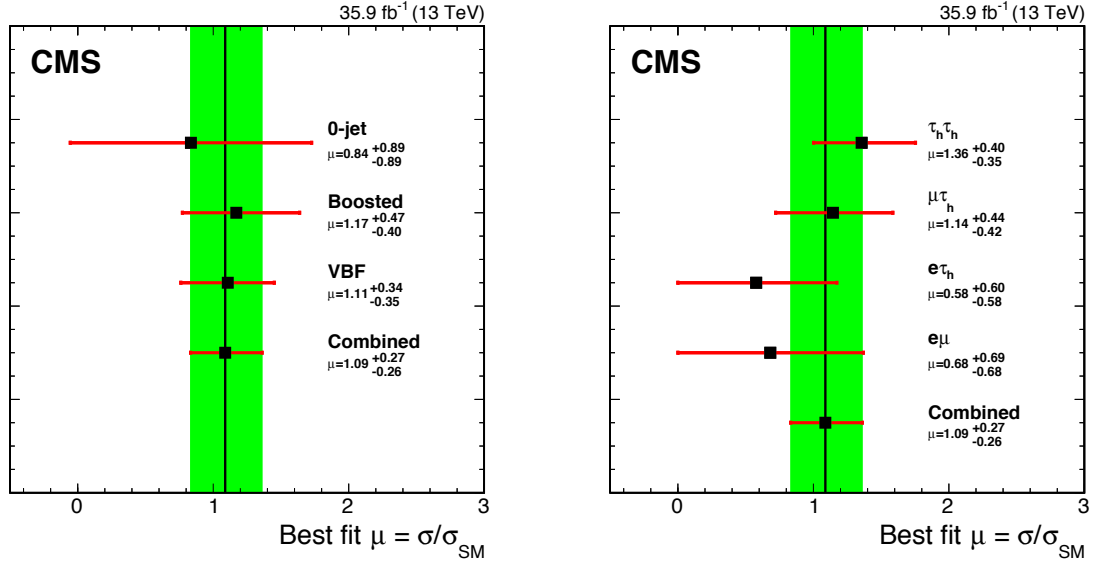


Figure 8.19: Best fit signal strengths for each category (**left**) and for each final state (**right**), for $m_H = 125.09$ GeV. The constraints from the global fit are used to extract the individual signal strengths [126].

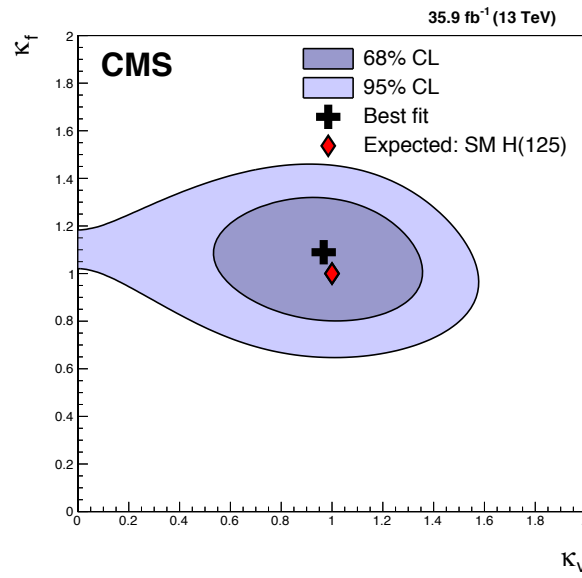


Figure 8.20: A likelihood scan as a function of κ_V and κ_f parameters, for $m_H = 125.09$ GeV [126].

CHAPTER

9

ANALYSIS OF MSSM HIGGS TO $\tau\tau$ IN THE $e\mu$ FINAL STATE

Contents

9.1	Analysis Overview	148
9.2	Datasets and Monte Carlo Samples	148
9.3	Data-to-Monte Carlo Corrections	149
9.3.1	Triggers	149
9.3.2	B-tagging	149
9.4	Event Selection	150
9.4.1	Categorization	150
9.5	Background Modeling	150
9.5.1	Drell-Yan Z +jets	150
9.5.2	Refined modeling of QCD multijet events	151
9.5.3	Modeling of jet $\rightarrow e/\mu$ events	154
9.6	Signal Extraction	156
9.7	Systematic Uncertainties	156
9.7.1	Uncertainties related to object reconstruction and identification	156
9.7.2	Background estimation uncertainties	158
9.7.3	Signal prediction uncertainties	159
9.7.4	Other uncertainties	159
9.8	Results of $e\mu$ Final State	160
9.9	Combination with $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ Final States	160
9.9.1	Brief description of $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states	160
9.9.2	Results	164

9.1 Analysis Overview

After the discovery of a SM-like Higgs boson, the most natural question to ask is whether there are more Higgs bosons. The MSSM introduces two Higgs doublets which can lead to five physical Higgs bosons, as detailed in chapter 3. Three of them are neutral bosons: the light (h) and heavy (H) scalar Higgs, and the pseudoscalar A . Possible extensions to the SM must include the lately discovered 125 GeV boson and its measured properties. Dependencies on additional parameters of the SUSY breaking mechanism enter via higher order corrections in perturbation theory. These parameters can be fixed in benchmark scenarios allowing only the tree level parameters $\tan\beta$ and m_A to freely float. The h takes the role of the observed SM-like Higgs boson at 125 GeV and the H and A are nearly degenerate in mass because, for values of $m_A > 300$ GeV, the MSSM is close to the decoupling limit. The symbol ϕ is used to denote H and A together.

At leading order, the H and A coupling to down-type fermions is enhanced by $\tan\beta$. Due to the relative large mass of the τ lepton, the $\tau\tau$ channel is a promising final state to probe the coupling of neutral Higgs boson to fermions. Though the coupling to $b\bar{b}$ is larger, the b-quark identification is comparably less efficient than the case of τ lepton.

Searches for additional heavy neutral Higgs bosons H/A to $\tau\tau$ final states have been performed at LEP [127] and Tevatron [128, 129]. At the LHC, the CMS and ATLAS experiments have performed such searches [130, 131] as well. So far, no discovery of heavy neutral Higgs boson is observed.

The $\tau\tau$ pair has six decay final states as described in the previous chapter. This thesis is focused on the $e\mu$ channel of the $\tau\tau$ decay and targeted gluon fusion ($gg\phi$) production and production in association with b-quarks ($bb\phi$). The result from this channel is further combined with the results from $\tau_h\tau_h$, $\mu\tau_h$ and $e\tau_h$ channels. As no evidence of signal is found, exclusion upper limits are set on the product of the cross section and the branching ratio for the decay into τ leptons. Exclusion contours in tree level parameters $m_A - \tan\beta$ plane for selected MSSM benchmark scenarios are also presented.

9.2 Datasets and Monte Carlo Samples

The analyzed datasets correspond to an integrated luminosity of 35.9 fb^{-1} , collected in 2016 at a center-of-mass energy of 13 TeV by the CMS detector. Only fully certified data is used.

The MC samples used for modeling SM backgrounds and the SM Higgs boson are the same as the analysis in the previous chapter, listed in tables 8.1 and 8.2.

Additional signal samples have been generated for $gg\phi$ and $bb\phi$, as listed in table 9.1. The samples are generated for ϕ with various mass hypothesis. In total there are 29 mass points corresponding to 80, 90, 100, 110, 120, 130, 140, 160, 180, 200, 250, 300, 350, 400, 450, 600, 700, 800, 900, 1200, 1400, 1500, 1600, 1800, 2000, 2300, 2600, 2900 and 3200 GeV ϕ . The samples of $gg\phi$ production mode are generated by PYTHIA8 at LO precision, while samples of $bb\phi$ production mode are generated using MADGRAPH5_AMC@NLO at NLO precision.

Process	Generator
$gg \rightarrow \phi \rightarrow \tau\tau$	PYTHIA8
$gg \rightarrow bb\phi \rightarrow \tau\tau$	MADGRAPH5_AMC@NLO (NLO)

Table 9.1: List of the Monte Carlo signal samples included in the analysis. The signals are generated for 29 mass points of the neutral scalar $\phi = H/A$.

9.3 Data-to-Monte Carlo Corrections

Several corrections are derived to account for the disagreement between data and MC simulation. Most of the corrections are the same described in section 8.3, including pileup reweighting, electron and muon identification and isolation scale factors, recoil correction, top p_T reweighting and LO Madgraph Z p_T reweighting. Some differences are the different trigger scale factors, because the triggers used in this analysis are different from those used in the SM analysis and b-tagging scale factors are added to the MSSM analysis to better model the multiplicity of b-tagged jets.

9.3.1 Triggers

The 2016 LHC proton-proton collision runs are divided into several periods, denoted as A, B, C, D, E, F, G and H periods. The full 35.9 fb^{-1} data used in this thesis covers the periods from B to G. From B to F, the trigger path used for data are the same cross triggers described in the previous chapter. The data collected from the last two run periods, G and H, exploits the same trigger but added a DZ filter in the sequence to cope with the increased instantaneous luminosity. The DZ filter only trigger the events that have the lepton fulfilling the requirement of longitudinal impact parameter $d_z < 2 \text{ mm}$. The data from G and H is required to fire one of the $e + \mu$ cross triggers "HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ" and "HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ". The DZ filter efficiency for run G to H is measured in data from top pair production events to be $\epsilon_{\text{DZ}}^{\text{G,H}} = 0.953 \pm 0.006(\text{stat}) \pm 0.016(\text{syst})$. As the filter is only applied in run G and H, it can be treated as efficient as the other runs. The DZ filter efficiency for the full 2016 data can be obtained as the luminosity weighted average and estimated to be $\epsilon_{\text{DZ}} = 0.979 \pm 0.003(\text{stat}) \pm 0.008(\text{syst})$. The trigger scale factors to be applied in the analysis are:

$$\text{SF} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}} \cdot \epsilon_{\text{DZ}}. \quad (9.1)$$

where ϵ_{data} and ϵ_{MC} are the trigger efficiencies for data and for MC, respectively, determined by equation 8.1.

9.3.2 B-tagging

The b-tagging data-to-MC scale factors are measured by using "promote-demote" method. Firstly, the b-tagging efficiencies ϵ is measured in top pair production events passing the baseline selection, and combined with the efficiencies measured in Drell-Yan events to get a more accurate modeling at high jet p_T , in bins of jet η and p_T . The scale factors are achieved

by dividing the efficiency measured in data by the efficiency measured in MC, $SF = \epsilon_{\text{data}}/\epsilon_{\text{MC}}$. The scale factors are applied to update the b-tagging status on a jet-by-jet basis, as follows: if $SF < 1$, a fraction of $1 - SF$ of b-tagged jets is demoted to untagged status. If $SF > 1$, a fraction of $1 - SF/(1 - 1/\epsilon_{\text{MC}})$ of untagged jets is promoted to tagged status. The main advantage of this method is that after the b-tagging status has been updated, there is no need to apply any event weights on the simulated events.

9.4 Event Selection

The baseline selection criteria for the $e\mu$ channel are the same as in the previous chapter. The differences with respect to the analysis presented in chapter 8 are: the bad muon filter detailed in subsection 7.4.7 is applied, the D_ζ variable cut is further optimized to $D_\zeta > -50$ GeV, and events are discarded if they have additional muons or electrons, satisfying the following requirements:

- Extra muon veto: no additional muons with $p_T > 10$ GeV, $|\eta| < 2.4$ and $I_\mu < 0.3$ passing the medium identification.
- Extra electron veto: no additional electrons with $p_T > 10$ GeV, $|\eta| < 2.5$ and $I_e < 0.3$ passing the 90% efficiency working point of the electron MVA ID.

9.4.1 Categorization

The events passing the baseline selection are split into two mutually exclusive categories based on the number of b-tagged jets in an event:

- **No b-tag**: this category contains events with no b-tagged jet with $p_T > 20$ GeV and $|\eta| < 2.4$. It is designed to target the $gg\phi$ production mode.
- **b-tag**: this category contains events with at least one b-tagged jet with $p_T > 20$ GeV and $|\eta| < 2.4$. It is designed to target the $bb\phi$ production mode.

Because the signal-to-background ratio has a strong dependence of the D_ζ variable, in each category, three subcategories are introduced to improve the search sensitivity based on the value of the D_ζ variable. Three ranges of D_ζ are taken into account:

- **low** D_ζ : $-50 < D_\zeta \leq -10$ GeV.
- **medium** D_ζ : $-10 < D_\zeta \leq 30$ GeV.
- **high** D_ζ : $D_\zeta > 30$ GeV.

The distributions of b-tagged jet multiplicity and D_ζ variable are shown in figure 9.1.

9.5 Background Modeling

9.5.1 Drell-Yan Z +jets

The Drell-Yan Z +jets background is modeled by the MC simulated samples. A shape correction for the generator level Z boson mass and p_T is applied as in the previous SM analysis chapter.

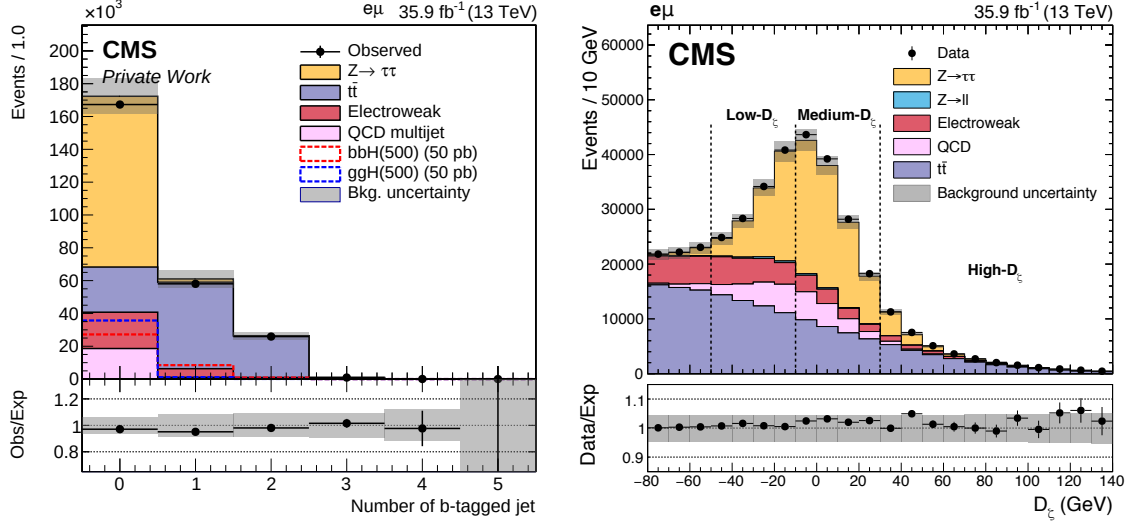


Figure 9.1: Distributions of the b-tagged jet multiplicity (**left**) and the D_ζ variables (**right**) [132].

9.5.2 Refined modeling of QCD multijet events

The same idea is adopted for the search for MSSM Higgs analysis compared to the estimation method in SM Higgs analysis. The contribution of the QCD multijet background in the signal region (SR), where the electric charges of the electron and the muon pair are required to be of opposite sign (OS), is estimated using a data driven technique from a control region, where the electric charges are required to be same signed (SS). The ratio of events number of $e\mu$ pair with OS to the events number of $e\mu$ with SS is measured in a sideband (SB) region where the events fail the nominal lepton relative isolation $I_e < 0.15$ and $I_\mu < 0.2$. The OS/SS ratio is obtained from a relaxed isolation region $I_{e,\mu} < 0.4$ and at least one lepton fails the nominal isolation requirement, i.e. $I_e > 0.15$ or $I_\mu > 0.2$ to avoid overlapping with the SR. The ratio measurement is performed in the events with selected $e\mu$ pair passing the baseline selection and with requirement of $D_\zeta > -50$ GeV. The inclusive OS/SS ratios from the SB region determined with data and with simulation are listed in table 9.2.

Region	Data	MC
SB	2.33 ± 0.05	2.44 ± 0.16
SR	-	2.20 ± 0.32

Table 9.2: The inclusive OS/SS ratio measured in isolation sideband (SB) region determined from data and from simulation. The OS/SS ratio measured in signal region (SR) determined from simulation is also presented.

The OS/SS ratio in the SB region can be measured as a function of p_T of the leading lepton, p_T^{lead} , of the subleading lepton, p_T^{sublead} , and as a function of separation $\Delta R(e, \mu)$ between the leptons in the $\eta - \phi$ plane. Other non-QCD MC background in the SB region are subtracted from data. The OS/SS ratios obtained as a function of lepton kinematics are

listed in table 9.3.

p_T^{lead} [GeV]	p_T^{sublead} [GeV]	$\Delta R(e, \mu)$	OS/SS ratio
[24, 30)	[10, 17)	< 2	2.72 ± 0.07
≥ 30	[10, 17)	< 2	2.60 ± 0.28
[24, 30)	[17, 24)	< 2	2.73 ± 0.09
≥ 30	[17, 24)	< 2	2.91 ± 0.71
[24, 30)	[24, 30)	< 2	2.56 ± 0.18
≥ 30	[24, 30)	< 2	2.46 ± 0.37
≥ 30	≥ 30	< 2	2.44 ± 0.50
[24, 30)	[10, 17)	[2, 4)	2.31 ± 0.04
≥ 30	[10, 17)	[2, 4)	1.97 ± 0.20
[24, 30)	[17, 24)	[2, 4)	2.32 ± 0.04
≥ 30	[17, 24)	[2, 4)	2.00 ± 0.15
[24, 30)	[24, 30)	[2, 4)	2.38 ± 0.06
≥ 30	[24, 30)	[2, 4)	1.95 ± 0.11
≥ 30	≥ 30	[2, 4)	1.75 ± 0.29
[24, 30)	[10, 17)	≥ 4	1.87 ± 0.12
≥ 30	[10, 17)	≥ 4	1.37 ± 0.47
[24, 30)	[17, 24)	≥ 4	1.96 ± 0.10
≥ 30	[17, 24)	≥ 4	1.77 ± 0.68
[24, 30)	[24, 30)	≥ 4	2.02 ± 0.22
≥ 30	[24, 30)	≥ 4	1.57 ± 0.33
≥ 30	≥ 30	≥ 4	1.47 ± 0.55

Table 9.3: The OS/SS ratio measured in isolation sideband (SB) region as a function of the lepton kinematics.

The integrated OS/SS ratios measured for different event categories in SB are shown in table 9.4. The OS/SS ratios in the no btag categories are not so different from the OS/SS ratio in the inclusive category, while, for the btag categories, the OS/SS ratio is much lower than the inclusive. Extrapolation factors are introduced to correct the differences between inclusive category and individual event categories, also shown in table 9.4.

The OS/SS ratios from table 9.3 dependent on the lepton kinematic and the extrapolation factor from table 9.4 are used together as weight applied on an event-by-event basis to the same sign $e\mu$ events. A closure test is performed in the SB region. Figure 9.2 shows the distributions of the m_T^{tot} variable in SB between the SS and OS selected $e\mu$ pairs. The OS/SS

Category	OS/SS ratio	Extrapolation factor F_{cat}
inclusive	2.33 ± 0.05	-
no btag inclusive	2.39 ± 0.06	1.026 ± 0.027
no btag low D_ζ	2.39 ± 0.09	1.026 ± 0.039
no btag medium D_ζ	2.38 ± 0.07	1.021 ± 0.030
no btag high D_ζ	2.65 ± 0.14	1.137 ± 0.060
btag inclusive	1.39 ± 0.11	0.596 ± 0.047
btag low D_ζ	1.37 ± 0.14	0.588 ± 0.060
btag medium D_ζ	1.40 ± 0.12	0.600 ± 0.051
btag high D_ζ	1.60 ± 0.46	0.686 ± 0.197

Table 9.4: The OS/SS ratio measured in isolation sideband (SB) region determined from data for various categories. Also shown are the extrapolation factors from the inclusive category to different event categories.

ratios (listed in table 9.3) and extrapolation factors (listed in table 9.4) are applied on the SS samples. The shapes between OS and SS samples are in good agreement both in btag and no btag categories.

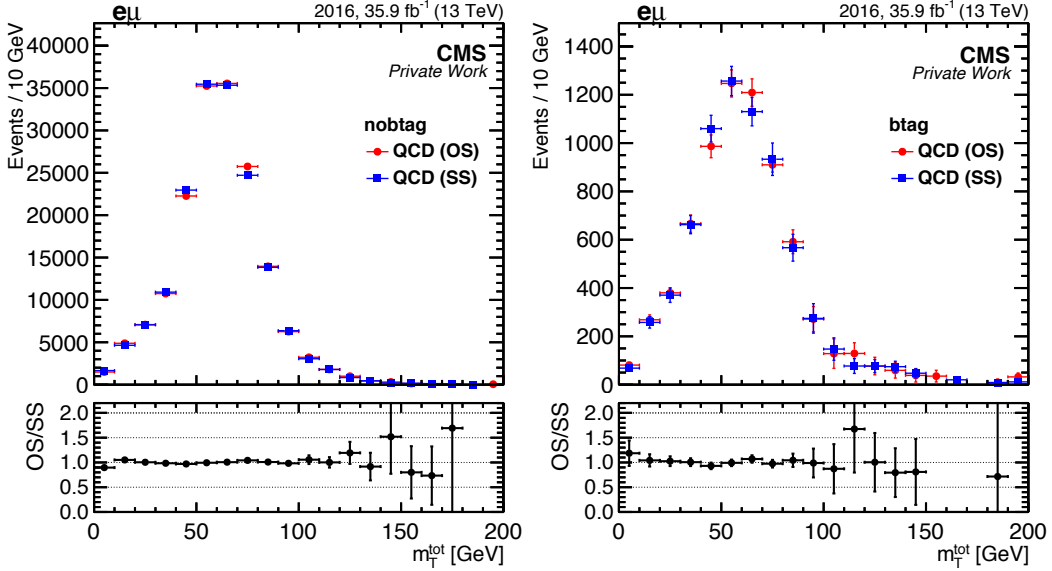


Figure 9.2: A closure test in the SB region is performed. The m_T^{tot} distributions of the SS selected $e\mu$ pairs events (blue) show good agreement to the OS selected (red) both in no btag (left) and btag (right) categories. The OS/SS ratios and the extrapolation factors are applied on the SS selected $e\mu$ pair events.

To account for the difference in the OS/SS ratio between the signal region and the SB region, another extrapolation scale factor is applied to the yield of QCD events. This scale factor, denoted as F_{iso} , is obtained from the MC simulated QCD multijet events as the OS/SS ratio in the signal region divided by that in the sideband:

$$F_{\text{iso}} = \frac{(\text{OS/SS})_{\text{MC}}^{\text{SR}}}{(\text{OS/SS})_{\text{MC}}^{\text{SB}}} = \frac{2.20 \pm 0.32}{2.44 \pm 0.16} = 0.90 \pm 0.14, \quad (9.2)$$

In order to have QCD multijet templates populated with enough events, the shape templates are obtained from the same sign $e\mu$ events selected with relaxed isolation requirements $I_e < 0.5$ and $0.2 < I_\mu < 0.5$. As figure 9.3 shows, the templates are not biased due to the change of isolation requirements. The normalized scale factor applied to the template is calculated by the this formula:

$$w_{\text{QCD}} = (\text{OS/SS})_{\text{kinetic}} \cdot F_{\text{cat}} \cdot F_{\text{iso}}, \quad (9.3)$$

where $(\text{OS/SS})_{\text{kinetic}}$ are the OS/SS values listed in table 9.3.

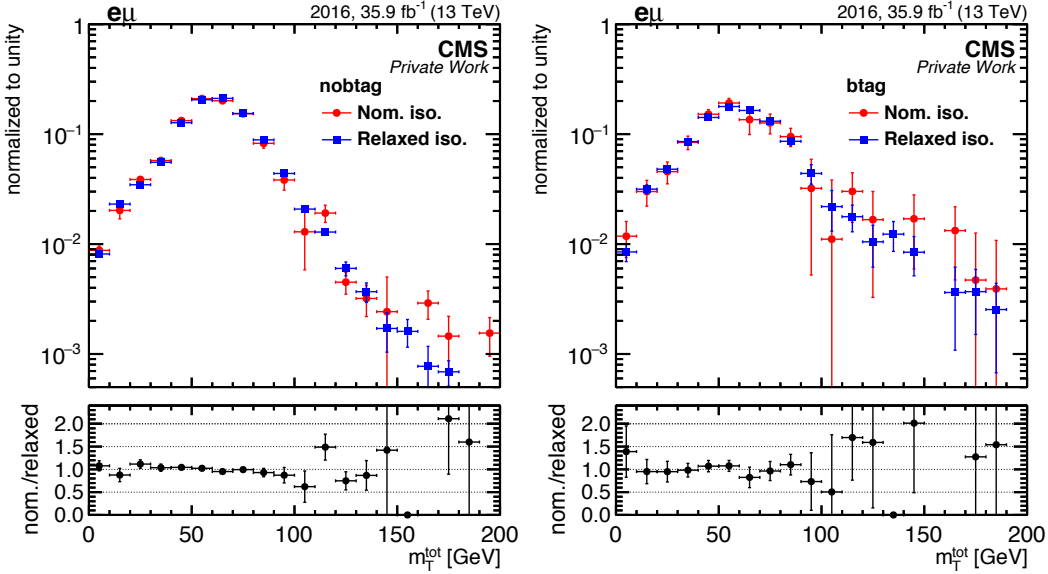


Figure 9.3: The m_T^{tot} variable distributions from relaxed isolation region (blue) and from nominal isolation region (red). Good agreements are shown for both no btag (left) and btag (right) categories.

9.5.3 Modeling of $\text{jet} \rightarrow e/\mu$ events

The non-QCD background can be split into two contributions: the contribution with genuine reconstructed leptons and the contribution from jets misidentified as leptons. The contribution with genuine reconstructed leptons is modeled with pure MC simulation, while the contribution with jets misidentified as leptons is modeled with MC simulation plus a p_T dependence correction.

The correction is accounted for the $\text{jet} \rightarrow e/\mu$ fake rate difference between the data and MC simulation. The $\text{jet} \rightarrow e/\mu$ fake rate is measured with $Z \rightarrow \mu\mu + \text{jets}$ events. The events are triggered by the "HLT_IsoMu24" trigger and selected by requiring two isolated opposite sign muons with invariant mass in the window 70 to 112 GeV. The two muons should have $p_T > 15$ GeV and $|\eta| < 2.4$ with one muon matched to the offline trigger object having $p_T > 25$ GeV. The event is required to have at least one jet separated from the selected muons by $\Delta R(\text{jet}, \mu) > 0.6$. In the case in which more than one jet is found, the probe jet is selected by requiring the largest azimuthal angle with respect to the dimension of the dimuon system. The pass or fail criteria is defined as whether the jet matches within a cone size $\Delta R = 0.5$ to an electron or a muon satisfying the $e\mu$ channel selection criteria. The misidentification rate is measured as function of probe jet p_T . The p_T dependence of data-to-MC scale factors is fitted with a linear function. Figure 9.4 shows the results of the fake rates and the fits of the data-to-MC scale factors. For electrons, the scale factors and the relevant shape systematics corresponding to +1 (up) and -1 (down) standard deviation are:

$$\begin{aligned} F_{\text{jet} \rightarrow e} &= 0.0015 \cdot p_T^e (\text{GeV}^{-1}) + 0.794, \\ F_{\text{jet} \rightarrow e}^{\text{up}} &= 0.0030 \cdot p_T^e (\text{GeV}^{-1}) + 0.883, \\ F_{\text{jet} \rightarrow e}^{\text{down}} &= 0.702. \end{aligned} \quad (9.4)$$

For muons, they are

$$\begin{aligned} F_{\text{jet} \rightarrow \mu} &= 0.0025 \cdot p_T^\mu (\text{GeV}^{-1}) + 0.902, \\ F_{\text{jet} \rightarrow \mu}^{\text{up}} &= 0.00350 \cdot p_T^\mu (\text{GeV}^{-1}) + 0.992, \\ F_{\text{jet} \rightarrow \mu}^{\text{down}} &= 0.812. \end{aligned} \quad (9.5)$$

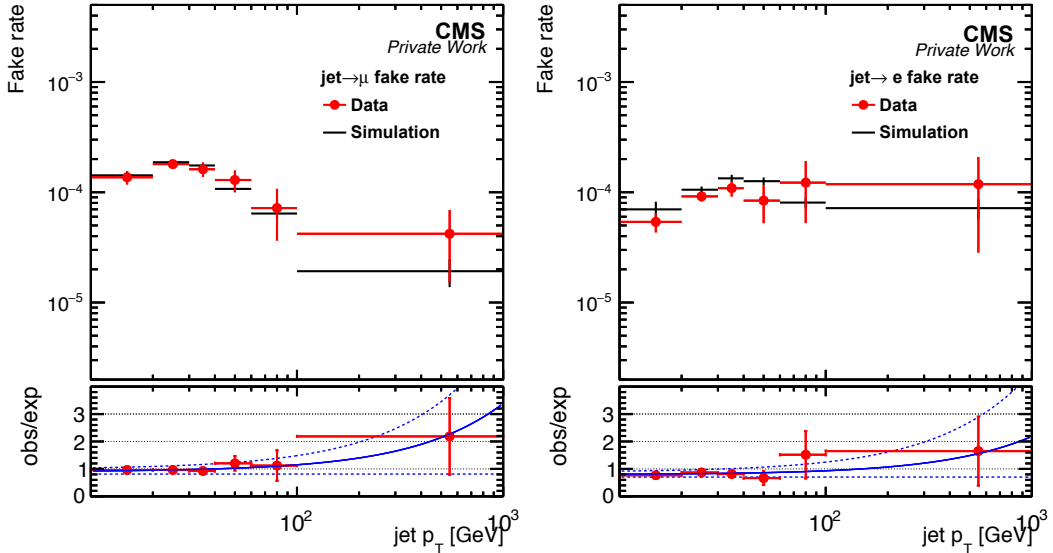


Figure 9.4: The misidentification probabilities of $\text{jet} \rightarrow \mu$ (**left**) and $\text{jet} \rightarrow e$ (**right**) as a function of jet p_T . The lower pads show the fitted function of the data-to-MC scale factors. The function are applied to the corresponding $\text{jet} \rightarrow e\mu$ contribution in the MSSM analysis.

These scale factors are applied to simulated W +jets and diboson events, where a jet is misidentified as electron or muon. The flavor composition in the $Z \rightarrow \mu\mu$ +jets events, where the measurement of jet $\rightarrow$ lepton misidentification rate is performed, is similar to W +jets and diboson events. The scale factors are not applied to $t\bar{t}$ and single top events, where the flavor composition of jets is much different from the $Z \rightarrow \mu\mu$ +jets events.

9.6 Signal Extraction

To extract the signal, the total transverse mass variable is used as the final discriminant:

$$m_T^{\text{tot}} = \sqrt{m_T(\cancel{E}_T, \tau_1^{\text{vis}})^2 + m_T(\cancel{E}_T, \tau_2^{\text{vis}})^2 + m_T(\tau_1^{\text{vis}}, \tau_2^{\text{vis}})^2}, \quad (9.6)$$

where $m_T = \sqrt{2p_T^1 p_T^2 (1 - \cos \Delta\phi)}$ is the transverse mass, τ_1^{vis} is the leading τ visible decay product, τ_2^{vis} is the second leading τ visible decay product and $\Delta\phi$ is the azimuthal angle between the different τ decay visible products. Figure 9.5 shows the pre-fit distributions of the m_T^{tot} variable in the six categories. The medium D_ζ categories have the largest signal-to-background ratio.

By using a maximum likelihood fit of the m_T^{tot} distribution, the upper limits on the MSSM Higgs boson production cross section times the branching ratio for the decay to $\tau\tau$ are determined. The systematic uncertainties are treated as nuisance parameter in the fit and they are discussed in the next section.

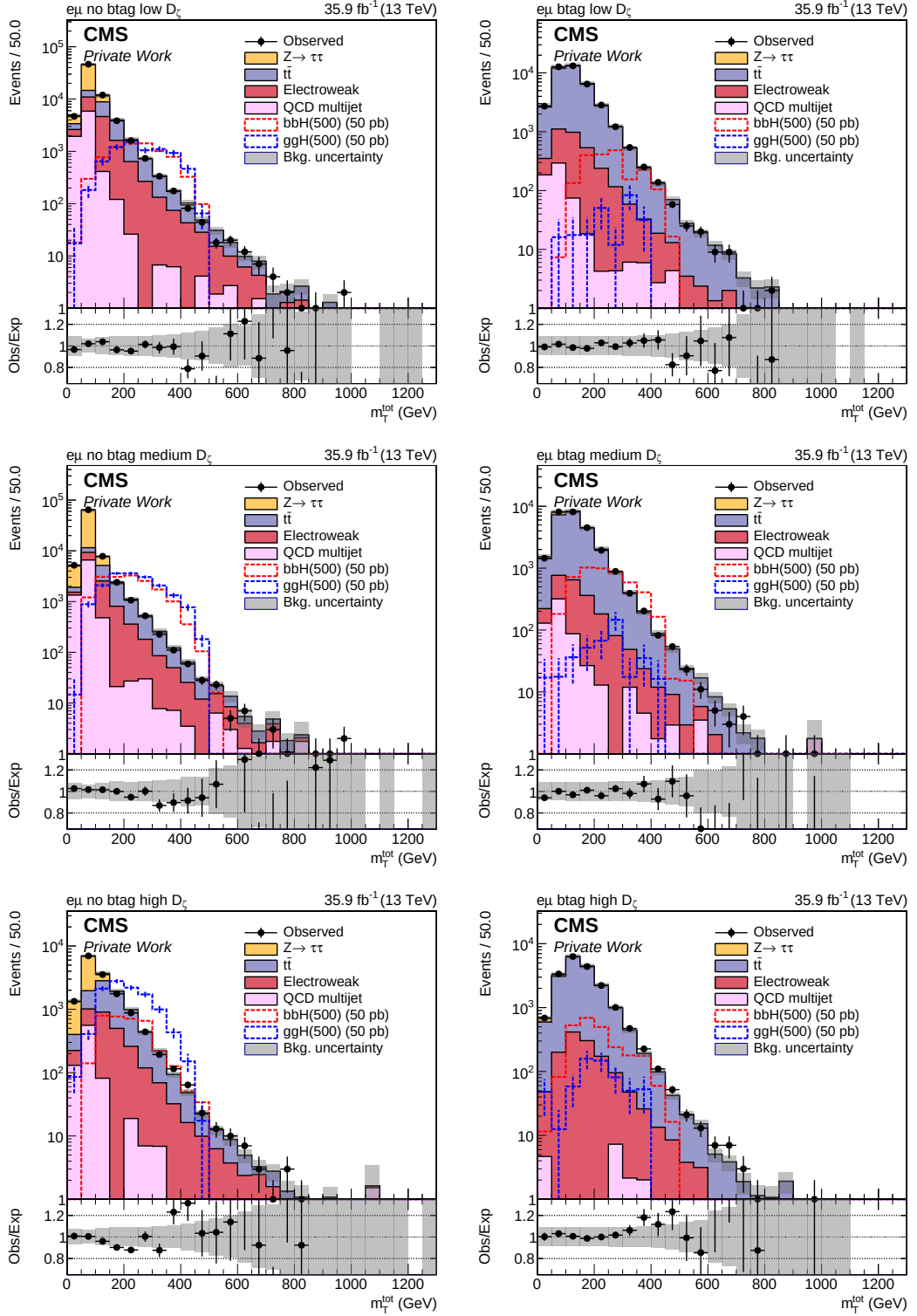
In order to constrain background normalization in the maximum likelihood fit, a $t\bar{t}$ enriched control region is added to the fit as well, together with the signal region. The event sample in this region contains the same $e\mu$ events baseline selection, except the additional selection criteria, $D_\zeta < -50$ GeV and $\cancel{E}_T > 80$ GeV, to enhance the yield of $t\bar{t}$ events. The purity of the top pair events equals to 85% in this region, while the remaining 15% originate from the single top and diboson production events. One bin is used for this control region. The $t\bar{t}$ normalization uncertainty and the top p_T reweighting shape uncertainty in this region are correlated to the same uncertainties in the signal region.

A $Z \rightarrow \mu\mu$ control region dedicated to control the normalization of the $Z \rightarrow \tau\tau$ background is included in the simultaneous fit along with the signal regions. A $\mu\mu$ pair formed by the most isolated muons is selected. The leading muon should have $p_T > 23$ GeV and $|\eta| < 2.1$. The subleading muon should have $p_T > 10$ GeV and $|\eta| < 2.1$. Additionally, the invariant mass of the dimuon system $m_{\mu\mu}$ is required to be between 70 and 110 GeV. Events with extra electrons or extra muons passing the extra lepton veto criteria are rejected. The control region is divided into no-btag and btag categories in order to have a phase space for the $Z \rightarrow \mu\mu$ and the $Z \rightarrow \tau\tau$ processes as similar as possible. Only one bin is used for this control region.

9.7 Systematic Uncertainties

9.7.1 Uncertainties related to object reconstruction and identification

A full description can be found in [132]. A list of these uncertainties in the $e\mu$ channel is summarized. The uncertainties related to the physics object reconstruction and identification have the following sources:

Figure 9.5: Pre-fit distributions of the m_T^{tot} variable in each event category of $e\mu$ channel.

- **Muon/Electron isolation and identification:** The uncertainty related to the isolation and identification of muon/electron amounts to 2%. They are applied to all simulated processes.
- **Electron energy scale:** The uncertainties affecting the shape in energy of electron is 1% in the barrel and 2.5% in the endcaps. It is applied to all simulated processes.
- **Jet energy scale:** The uncertainties in the jet energy scale depend on the p_T and η of the jet [104], ranging from 1 to 6%. They are applied to all simulated processes. The jet energy scale affects the number of events entering each category.
- **B-tagging uncertainty:** The uncertainties in the rate with which both light-flavor jets and true b-jets pass the b-tagging discriminator selection are applied to all processes estimated from simulation, ranging from 1 to 5% as a function of jet p_T and $|\eta|$.
- **Missing transverse energy:** The uncertainties in the resolution and response of the missing transverse energy $\cancel{E}_T$ are derived during the computation of the recoil corrections. This results in uncertainties ranging from 1 to 5% to all processes with recoil correction applied, including all signal processes, Drell-Yan and W +jets events. For those processes without recoil correction applied, the jet energy scale and unclustered energy scale variation are propagated to the $\cancel{E}_T$, also resulting in uncertainties ranging from 1 to 5%.

9.7.2 Background estimation uncertainties

The uncertainties related to the background estimation methods have the following sources:

- **W +jets estimation:** An uncertainty of 4% is assigned to the cross section of the W +jets process.
- **$\text{Jet} \rightarrow e/\mu$:** Shape uncertainties are applied to all processes with jets misidentified as electrons or muons to account for the $\text{Jet} \rightarrow e/\mu$ misidentification rate, as equations 9.4 and 9.5.
- **QCD estimation:** The QCD background is estimated by a data-driven method. The uncertainties should cover the statistical uncertainty of the data and subtracted backgrounds, plus the uncertainties associated with the extrapolation from lepton anti-isolated region, where the estimate is made, to the signal region. A total 30% uncertainty is accounted for the extrapolation method, in addition to category-dependent uncertainties ranging from 4 to 29%, in the measurement of the OS/SS ratios.
- **Top p_T reweighting:** A shape uncertainty to be considered as the uncertainty in the top quark p_T reweighting method is applied to top pair production events. The uncertainty is the difference between not applying the reweighting and applying the nominal reweighting twice.
- **Single top and diboson cross section:** The uncertainties in the cross sections of single top and diboson production amount to 5%, obtained from the CMS measurements [124, 125].

- **Z p_T reweighting:** Five shape uncertainties related to the Z p_T reweighting method which is used to improve the modeling of Drell-Yan events. Two of them are including the 0.2% muon energy scale uncertainty and 6% top pair production cross section uncertainty. The remaining three uncertainties are statistical uncertainties related to the limited statistics in three high mass bins which are found to be non-negligible.
- **MC acceptance extrapolation from $Z \rightarrow \mu\mu$ and top pair production control regions:** These uncertainties are related to the inclusion of the $Z \rightarrow \mu\mu$ and top pair production control regions. Uncertainties in the extrapolation from $Z \rightarrow \mu\mu$ ranges from 1 to 7%, applied to the $Z \rightarrow \tau\tau$ events, depending on the categories. The extrapolation uncertainties applied to the $Z \rightarrow \ell\ell$ events is 4%. The extrapolation uncertainties from the top pair production control region to signal region is smaller than 1%, applied to the top pair production events in signal region. An additional uncertainty of 1% is applied in the control region to account for the fluctuations in the variables used to select the control region events.

9.7.3 Signal prediction uncertainties

The uncertainties related to the signal processes have the following sources:

- **Signal cross section from PDFs:** The MSTW 2008 [133] PDFs are used for the calculation of the signal production cross sections. The uncertainties in the choices for the PDFs are estimated following the recommended procedure given in references [133, 134]. The uncertainties are evaluated for every $m_A - \tan\beta$ point individually. Their uncertainties ranges from 15 to 25%.
- **Signal acceptance:** The theoretical uncertainties in the signal events acceptance in the associated production with b quark are derived from the variations of the renormalization (μ_R) and factorization (μ_F) scales and of the generator matching scale Q_{sh} related to the parton shower. The scales μ_R and μ_F are varied by factors of 0.5 and 2 and the scale uncertainty contains six variations of μ_R and μ_F in total. The scale Q_{sh} is varied by factors of $1/\sqrt{2}$ and $\sqrt{2}$. These are recommended prescription from reference [135].
- **SM Higgs boson production:** The SM Higgs boson production is taken into account in the likelihood ratio as one of the backgrounds. Uncertainties in the choices for PDF and α_s uncertainties on signal cross section are 3.2% for the gluon fusion production, 2.1% for the VBF production, 1.9% for the WH production and 1.6% for the ZH production. Uncertainties due to the choices for renormalization and factorization scales for calculating the SM Higgs boson cross section are 3.9% for the gluon fusion production, 0.4% for the VBF production, 0.5% for the WH production and 2.8% for the ZH production.

9.7.4 Other uncertainties

The uncertainty in the integrated luminosity amounts to 2.5% [112]. The uncertainty is applied to the signal samples and backgrounds that are estimated using MC simulation.

In addition, uncertainties related to the limited number of simulated events are taken into account by allowing each bin of each background template to vary within its statistical uncertainty (bin-by-bin uncertainty).

9.8 Results of $e\mu$ Final State

Figure 9.6 shows the distribution of m_T^{tot} for the no-btag and btag categories in the $e\mu$ channel after the global maximum likelihood fit. The fit is performed for a signal-plus-background model. The expected m_T^{tot} distribution for a signal of three neutral Higgs bosons from gluon fusion and the production in association with b quarks in the MSSM $m_h^{\text{mode+}}$ scenario for $m_A = 700$ GeV and $\tan\beta = 20$ is also shown. The signal distribution has two peaking structures: one related to the signal from the h at about 130 GeV, and the almost mass degenerate H and A at 700 GeV.

9.9 Combination with $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ Final States

The results from the $e\mu$ final states are further combined with three other $\tau\tau$ final states: $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states, together with the control regions for $Z \rightarrow \mu\mu$ and top pair production backgrounds.

9.9.1 Brief description of $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states

The trigger requirement and kinematic selection for $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states are summarized in table 9.5. Data-to-MC scale factors for triggers, electron id and isolation, muon id and isolation and τ_h isolation scale factors [110] are applied. In the $e\tau_h(\mu\tau_h)$ final state, apart from the kinematic selection listed in the table, the τ_h should pass the tight (very loose) working point of the discriminator to suppress electrons and the loose (tight) working point of the discriminator to suppress muons, the relevant data-to-MC scale factors for these discriminators (listed tables 7.3, 7.14) are applied.

Final states	Trigger requirement	Lepton selection		
		p_T (GeV)	η	Isolation
$\tau_h\tau_h$	$\tau_h(35)\&\tau_h(35)$	$p_T^{\tau_h} > 40$	$ \eta ^{\tau_h} < 2.1$	MVA τ_h medium ID
$\mu\tau_h$	$\mu(22)$	$p_T^\mu > 23$	$ \eta ^\mu < 2.1$	$I_\mu < 0.15$
		$p_T^{\tau_h} > 30$	$ \eta ^{\tau_h} < 2.3$	MVA τ_h tight ID
$e\tau_h$	$e(25)$	$p_T^e > 26$	$ \eta ^e < 2.1$	$I_e < 0.1$
		$p_T^{\tau_h} > 30$	$ \eta ^{\tau_h} < 2.3$	MVA τ_h tight ID

Table 9.5: Trigger requirement and kinematic selection of the $\tau_h\tau_h$, $\mu\tau_h$ and $e\tau_h$ final states. The trigger requirement is defined by a combination of trigger candidates with p_T over a given threshold (in GeV), indicated inside the parentheses.

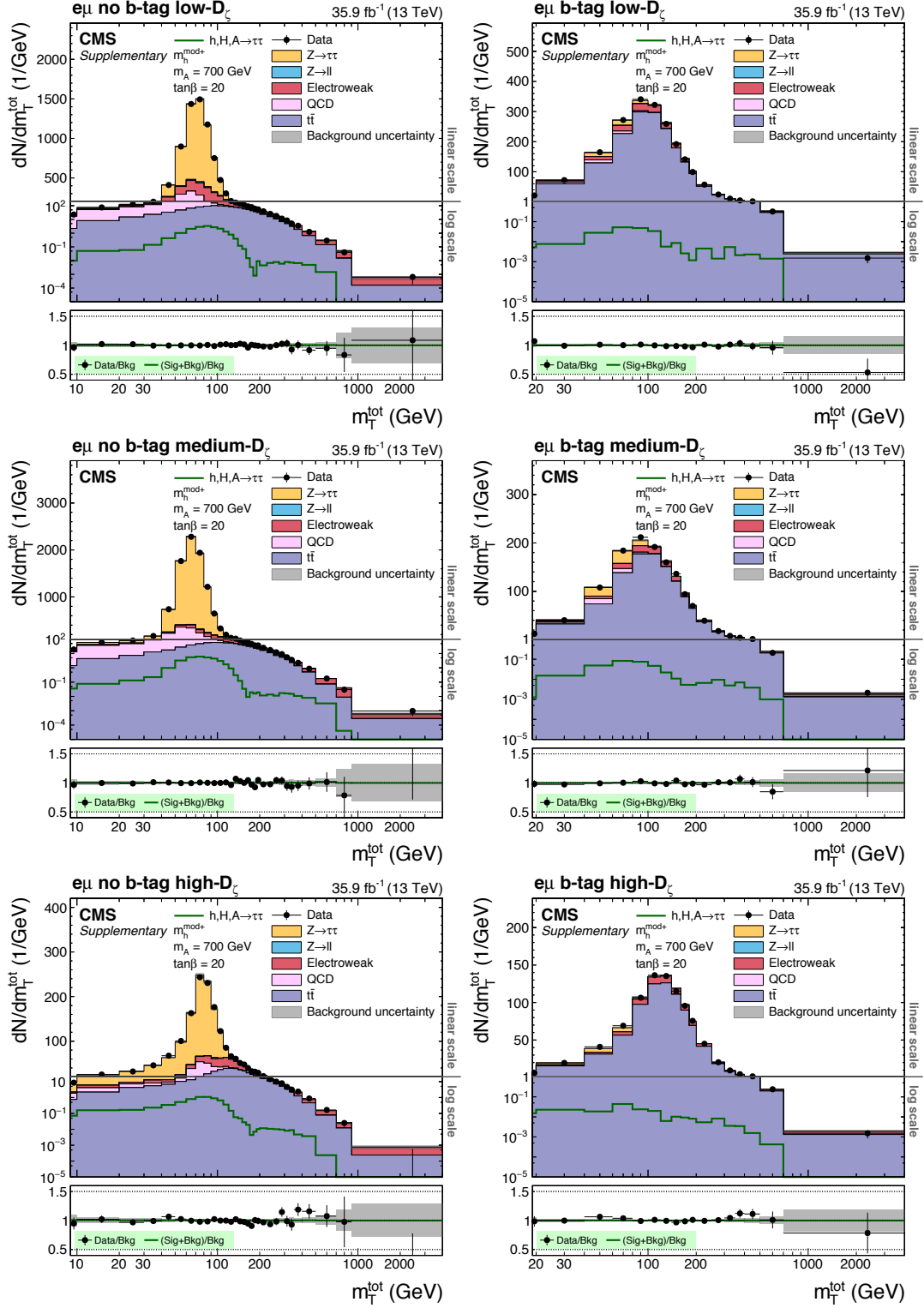


Figure 9.6: Distributions of m_T^{tot} in the no b-tag and b-tag categories in the low D_ζ , medium D_ζ and high D_ζ event subcategory [132].

To increase the sensitivity of the analysis, selected events in $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states are split into b-tag and no b-tag categories using the same definition as in $e\mu$ final state.

In the $e\tau_h(\mu\tau_h)$ final state, each global event category is further split into two subcategories based on the transverse mass of the electron (muon) and $\cancel{E}_T$, m_T , which is used to discriminate between the signal and the backgrounds from W +jets. The subcategories are defined as tight- m_T ($m_T < 40$ GeV) and loose- m_T ($40 < m_T < 70$ GeV). Tight- m_T subcategory is the most sensitive for signal events, especially for the low-mass hypotheses. The loose- m_T subcategory is added to increase the signal acceptance for mass hypotheses of $m_\phi > 700$ GeV. In figure 9.7, the m_T distribution is shown for the $\mu\tau_h$ final state before dividing into two subcategories.

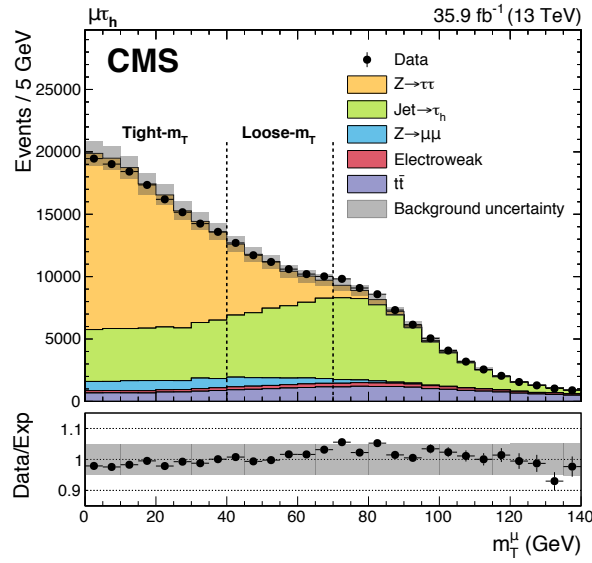


Figure 9.7: The transverse mass m_T distribution in $\mu\tau_h$ final state. The dashed vertical lines represent the definition of the subcategories [132].

In total, sixteen signal categories and three background control regions enter the simultaneous fit to extract the signal. A schematic view of the event categorization is shown in figure 9.8.

The background estimation in $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ channel is split into two parts. The first part contains backgrounds with genuine τ and with ℓ misidentified as τ_h . These contributions are estimated with MC simulation as in the $e\mu$ channel. The second part deals with jet misidentified as τ_h . The "fake factor" method is used to estimate this part of the contribution. A certain background i due to the jet $\rightarrow \tau_h$ misidentification is estimated from a relaxed τ_h isolation region. This region is denoted as application region (AR) and requires the τ_h isolation discriminator to satisfy the very loose but not the tight (medium) working point in the $e\tau_h/\mu\tau_h$ ($\tau_h\tau_h$) final state. This region is dominated by jet $\rightarrow \tau_h$ events, with a few percent of events with genuine τ lepton decays. To estimate the yield for background i due to jet $\rightarrow \tau_h$ misidentification in the signal region, the number in the AR is multiplied by the ratio

$$F_F^i = \frac{N_{\text{pass}}}{N_{\text{fail}}}, \quad (9.7)$$

	No b-tag			b-tag		
$H \rightarrow \tau\tau \rightarrow e\mu$	Low- D_ζ	Medium- D_ζ	High- D_ζ	Low- D_ζ	Medium- D_ζ	High- D_ζ
$H \rightarrow \tau\tau \rightarrow e\tau_h$	Loose- m_T		Tight- m_T	Loose- m_T		Tight- m_T
$H \rightarrow \tau\tau \rightarrow \mu\tau_h$	Loose- m_T		Tight- m_T	Loose- m_T		Tight- m_T
$H \rightarrow \tau\tau \rightarrow \tau_h\tau_h$						
$Z \rightarrow \mu\mu$						
$t\bar{t}(e\mu)$						
	Signal region (SR)			Control region		

Figure 9.8: Overview of all event categories entering the signal extraction fit. There are sixteen signal categories and three background control regions [132].

where N_{pass} represents the number of events satisfying the tight/medium working point and N_{fail} represents the number of events satisfying the very loose but not the tight/medium working point of the τ_h isolation discriminator. The number of events appearing in the above equation are determined from a determination region (DR_i), which is not exclusive to the AR and signal region, and enriched by background i . When estimating the number of events, the contributions from backgrounds other than i are subtracted from the numerator and denominator of equation 9.7. The F_F^i is different for different background processes. A weight w_i is applied on the individually determined F_F^i in order to obtain a weighted factor F_F on an event-by-event basis:

$$F_F = \sum_i w_i F_F^i, \quad w_i = \frac{N_{\text{AR}}^i}{\sum_j N_{\text{AR}}^j}, \quad i, j \in \{\text{QCD}, W + \text{jets}, t\bar{t}\}, \quad (9.8)$$

where N_{AR}^i represents the number of events for background i in AR. The factor F_F will be the factor applied to all events in the AR as the estimation of yield and shape of the sum of QCD multijet, W +jets and $t\bar{t}$ events due to jet $\rightarrow \tau_h$ misidentification. The idea of the fake factor method is shown in figure 9.9.

In the $e\tau_h$ and $\mu\tau_h$ final states, the DR_{QCD} is defined by the same selection as for the signal region, but the electric charge of the $\tau\tau$ pair decay products are required to be same signed. In addition, the transverse mass is required to be $m_T < 40$ GeV to suppress W +jets events, and the isolation of electron (muon) is relaxed to $0.05 < I^{e(\mu)} < 0.15$. The definition of $\text{DR}_{W+\text{jets}}$ has the same selection as for signal region, except that the transverse mass is changed to $m_T > 70$ GeV to have enriched W +jets events and no b-tagged jet is presented to suppress the $t\bar{t}$ events. As for the $\text{DR}_{t\bar{t}}$, because there is no sufficiently pure region for $t\bar{t}$ events, the $F_F^{t\bar{t}}$ are estimated from simulation after the event selection but before the event

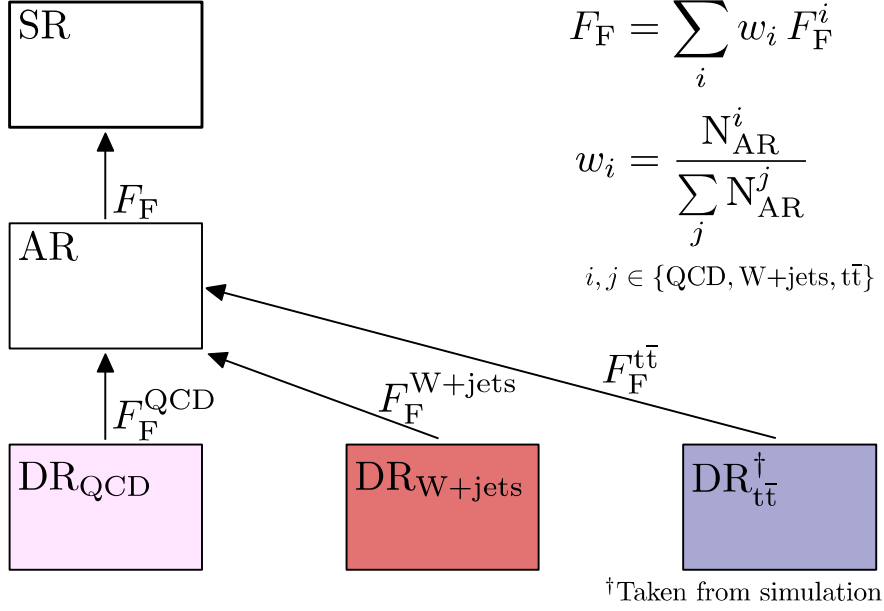


Figure 9.9: Overview of fake factor method used to estimate $\text{jet} \rightarrow \tau_h$ background contribution in $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states [132].

categorization.

In the $\tau_h\tau_h$ final state, only DR_{QCD} is defined from the requirement that electric charges of the $\tau\tau$ decay products should be same signed. The F_F^{QCD} is also used to estimate the $W+\text{jets}$ and $t\bar{t}$ backgrounds since they are far more subdominant. A weight of 0.5 is assumed, accounting for the combinatorial effect arising from the determination of F_F^{QCD} from the fact that each event can enter the AR in one of two mutually exclusive ways, either the leading τ_h candidate satisfying the requirement of nominal τ_h isolation and the subleading τ_h candidate inverted requirement or vice versa. For the backgrounds from $W+\text{jets}$ and $t\bar{t}$ events, these events have a weight of 1 since the fraction of events with two misidentified jets is very small at a few percent.

9.9.2 Results

To extract the signal, a likelihood function given by equation 4.5 is used. The m_T^{tot} distribution from $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$ final states are shown in figures 9.10. They are fitted by using MSSM $m_h^{\text{mod}+}$ signal-plus-background hypothesis, for $m_A = 700$ GeV and $\tan\beta = 20$. No signal is found in the mass region between 90 GeV to 3.2 TeV.

Upper limits on the presence of signal are set for two interpretations of data: background-only and some signal-plus-background hypotheses. The SM Higgs boson contribution is treated as one of the background processes. The upper limit setting is done following the modified frequentist approach as described in subsection 4.4.1.

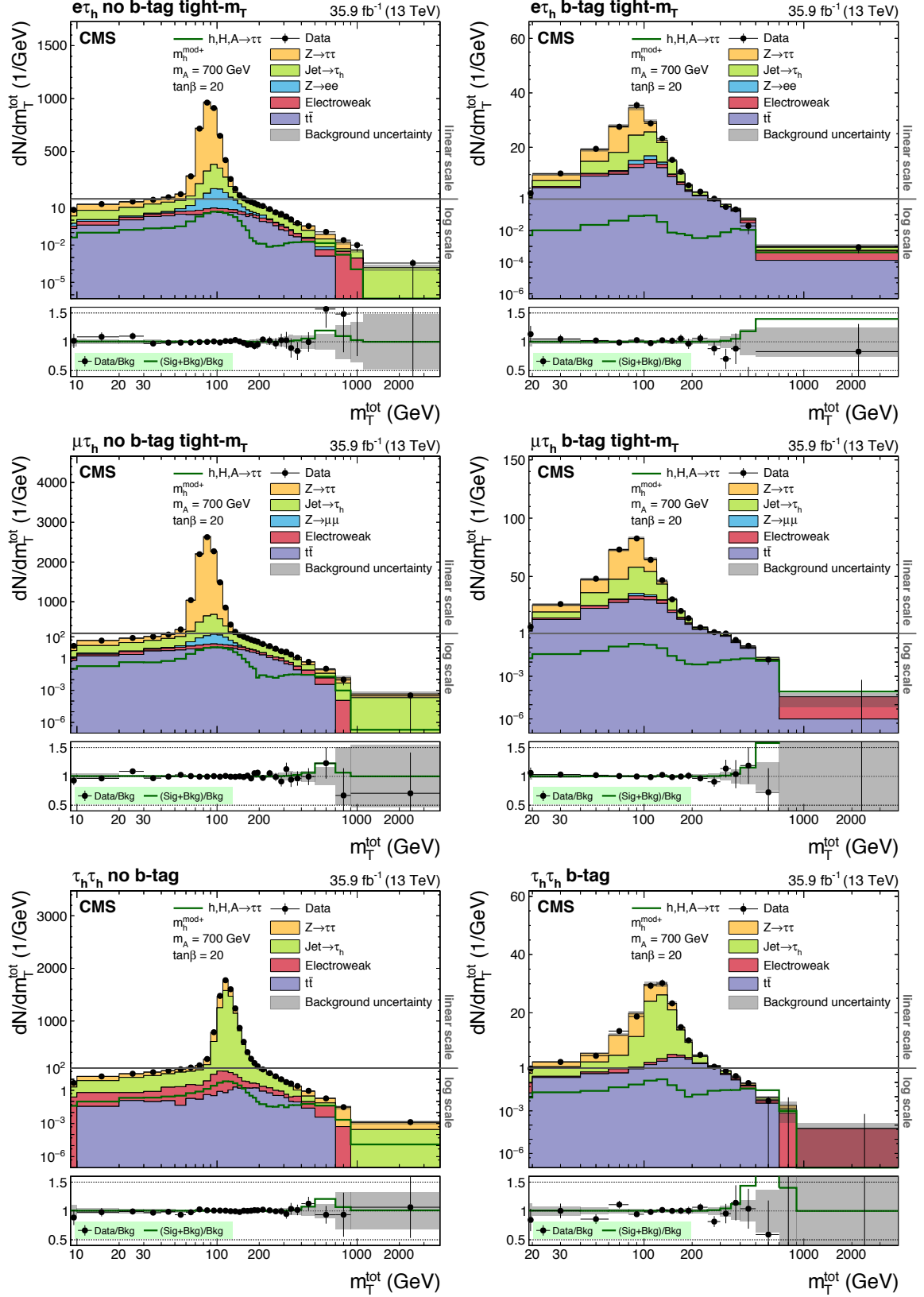


Figure 9.10: Distributions of m_T^{tot} in the no b-tag (left) and b-tag (right) categories for $e\tau_h$ (top), $\mu\tau_h$ (middle) and $\tau_h\tau_h$ (bottom) final states. A signal-plus-background hypothesis is used for the fit, corresponding to $m_A = 700$ GeV and $\tan\beta = 20$ [132].

Exclusion limits of the first interpretation are model-independent. The 95% CL upper limits on the product of cross section times branching ratio for $gg\phi$ and $bb\phi$ production are set. The limits are derived assuming a single narrow resonance, ϕ . The signal is searched for in both production at the same time. Two freely floating parameters of interest, which are the signal strengths for the two production modes, are used in the fit. Expected and observed 95% CL upper limits are shown in figure 9.11. For $gg\phi$ production, expected exclusion limits are shown for the cases where only the b quark and only the t quark are taken into account in the loop contribution, respectively. Figure 9.12 shows the expected 95% CL upper limits divided into four different final states. For $gg\phi$ production, the $\mu\tau_h$ final state is the most sensitive below 300 GeV. $e\mu$ and $e\tau_h$ final states are roughly the same in this mass range. Above 300 GeV, $\tau_h\tau_h$ and $\mu\tau_h$ have roughly the same sensitivity. For $bb\phi$ production, below 200 GeV, $\mu\tau_h$ is the most sensitive final state, while $\tau_h\tau_h$ is the most sensitive for masses $m_\phi > 200$ GeV.

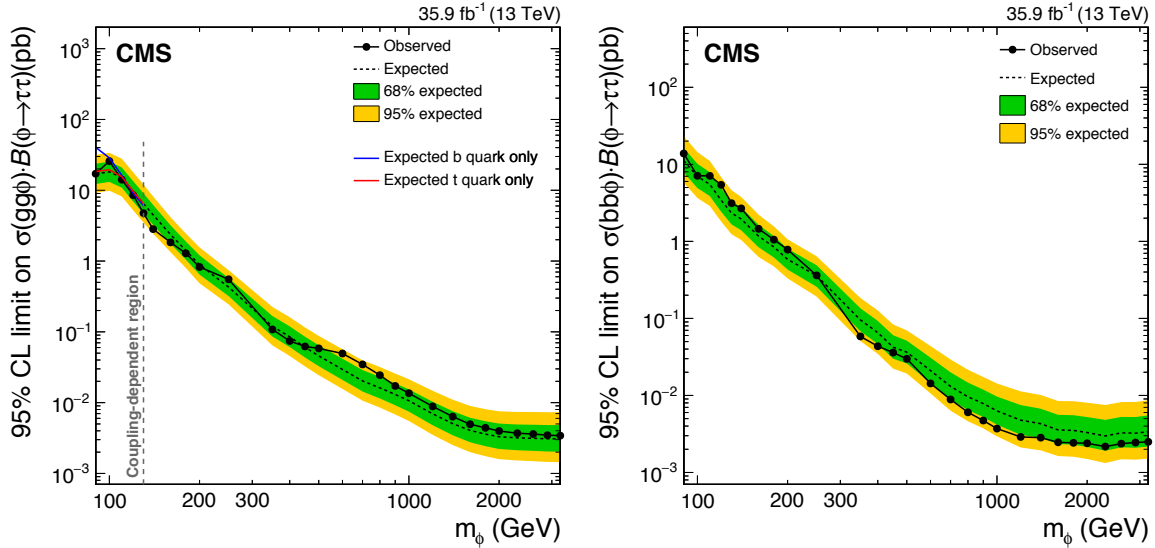


Figure 9.11: Expected and observed 95% CL upper limits for the production of a single narrow resonance ϕ with a mass between 90 GeV and 3.2 TeV in the $\tau\tau$ final state [132]. The SM Higgs boson has been included in the background. Limits are shown for $gg\phi$ (left) and $bb\phi$ (right) production modes.

A likelihood scan is also performed in the $gg\phi - bb\phi$ plane that is spanned for every mass point by the product of cross section time branching ratio ($\sigma \cdot \mathcal{B}$) of $gg\phi$ and $bb\phi$ production. The fit is performed by assuming the SM Higgs boson as one of the backgrounds. The representative results are shown for six mass points in figure 9.13.

The second interpretation is model dependent. The exclusion contours are shown in the $m_A - \tan\beta$ plane, for two MSSM benchmark scenarios: the $m_h^{\text{mod}+}$ and the hMSSM are considered. These scenarios were introduced in section 3.3.2. For each value of m_A and $\tan\beta$, a maximum likelihood fit to the data is performed. The test statistics is slightly different than the one used to obtain model independent limit (equation 4.22). The numerator remains the same, with a fixed value of $\mu = 1$, corresponding to the signal prediction for the given value of m_A and $\tan\beta$. But the denominator is changed to $\mu = 0$, corresponding to the

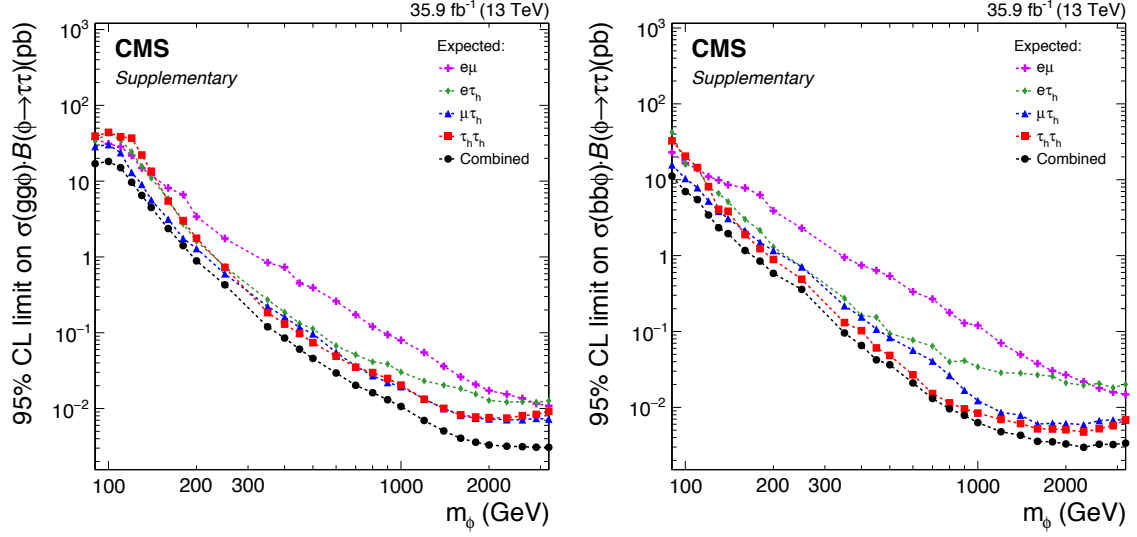


Figure 9.12: Expected 95% CL upper limits by four final states and their combination for $gg\phi$ (**left**) and $bb\phi$ (**right**) production modes [132].

background-only prediction. The SM Higgs is added to the background contribution in the likelihood ratio. In figure 9.15, the observed and expected 95% CL exclusion contours are shown for the MSSM $m_h^{\text{mod}+}$ and the hMSSM scenarios. The exclusion contours reach up to 1.6 TeV, two times broadening the CMS run 1 excluded mass range in m_A [130]. For values $m_A \leq 250$ GeV, the excluded region in both scenarios extend down to values of $\tan\beta \approx 6$. The observed exclusion limits show good agreement with the expected limits for the whole mass range. In the low mass region, the exclusion contour is similar to the previous CMS results [130]. This is due to three main factors: the observable m_T^{tot} used to extract the signal is built to have more sensitivity for the high mass region, but suffers reduced sensitivity for the low mass; the improved signal prediction at NLO precision shows a softer p_T spectrum, which dominates at low $\tan\beta$ region; the choice of single lepton triggers in the $e\tau_h$ and $\mu\tau_h$ final states with higher instantaneous luminosity brings the p_T threshold higher and therefore reduce the signal acceptance.

Figure 9.15 shows the selected CMS results of the MSSM neutral Higgs to $\tau\tau$ searches including the one presented in this thesis [132] and the comparison to the ATLAS run 2 results [131].

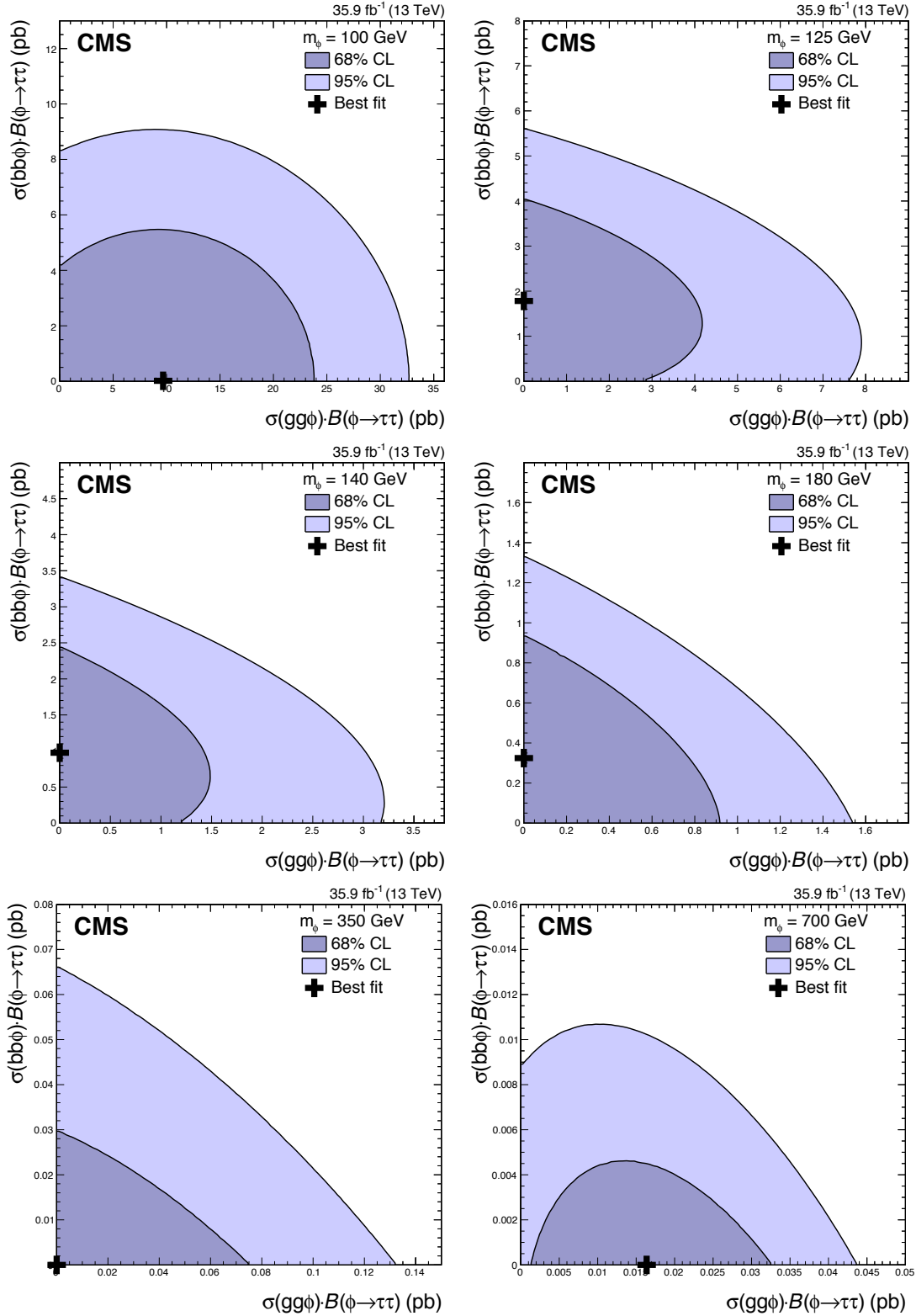


Figure 9.13: Scan of the likelihood function for the search in the $\tau\tau$ final state for a single narrow resonance ϕ produced via $gg\phi$ or $bb\phi$ production mode [132]. A representative subset of the mass points is shown.

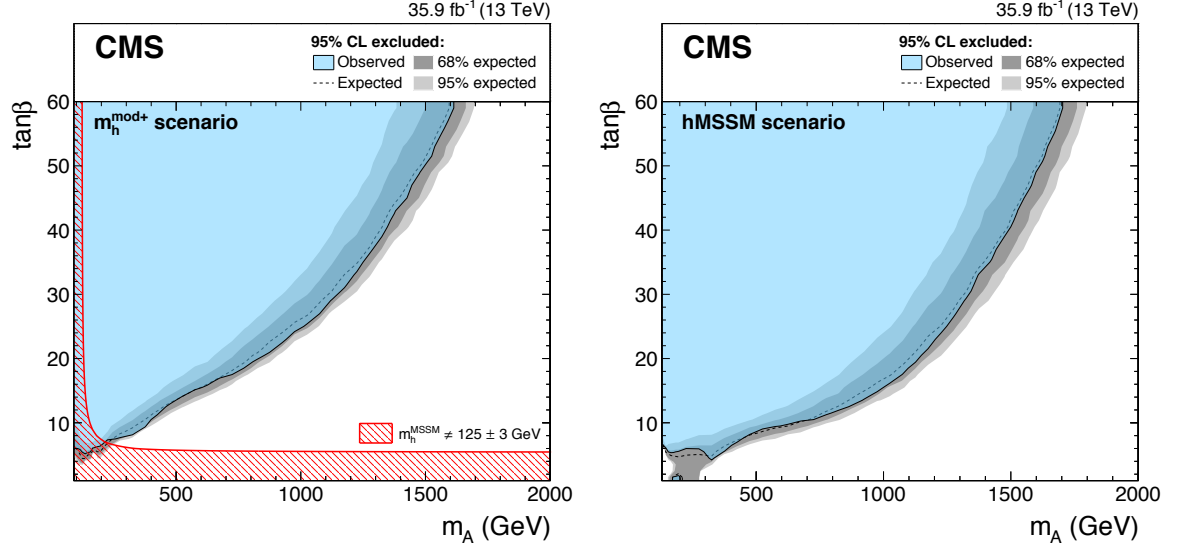


Figure 9.14: Observed and expected 95% CL exclusion contours in the $m_A - \tan\beta$ plane in the MSSM $m_h^{\text{mod}+}$ (left) and in the hMSSM scenario (right) [132].

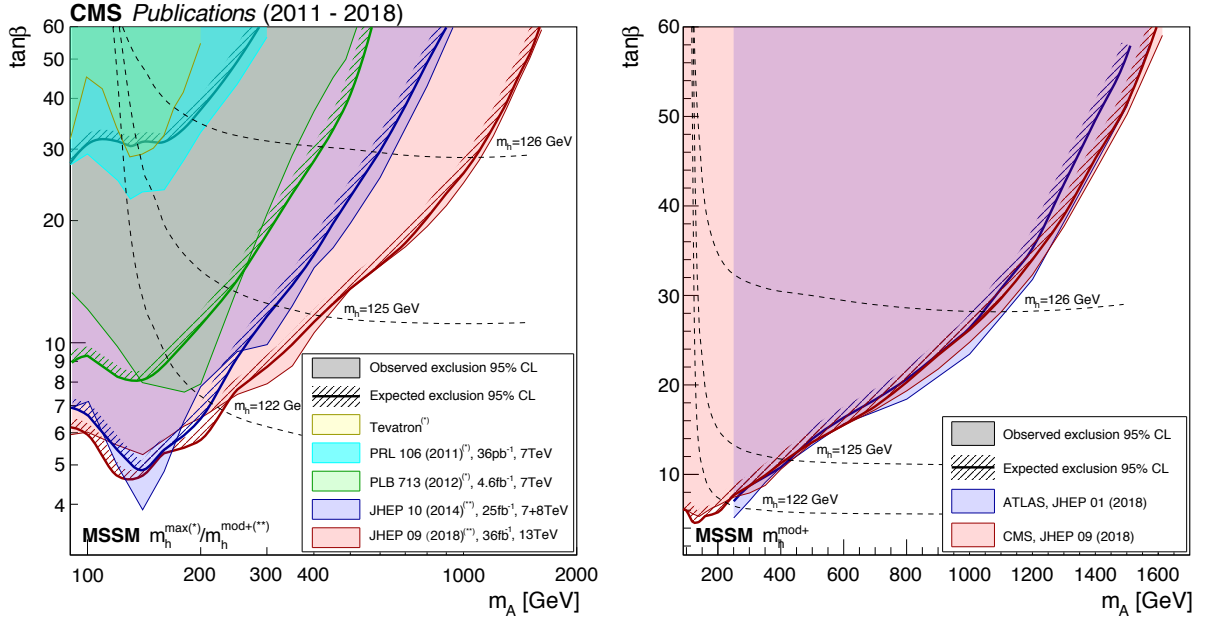


Figure 9.15: Observed and expected 95% CL exclusion contours in the $m_A - \tan\beta$ plane in the MSSM $m_h^{\text{mod}+}$ and $m_h^{\text{max}+}$ scenarios (left) [132]. The exclusion contour comparison between ATLAS and CMS results in the MSSM $m_h^{\text{mod}+}$ scenario. The ATLAS result is taken from [131].

CHAPTER

10

SUMMARY AND CONCLUSIONS

In the Standard Model (SM), the Higgs boson decays approximately in 6% of the cases to a pair of tau leptons if it has a mass of 125 GeV. In the Minimal Supersymmetric Standard Model (MSSM), the heavy neutral Higgs boson can decay to a pair of tau leptons as well. To probe the fermionic coupling of the SM and MSSM Higgs boson, the $\tau\tau$ final state is more promising than the $b\bar{b}$ final state, due to less background contamination, and the $\mu\mu$ final state, due to the larger branching ratio.

In this thesis, the use of tau lepton as a tool to discover the SM Higgs boson and to search for neutral MSSM Higgs boson has been described in detail. These analyses make use of the proton-proton collision data collected by the CMS detector in 2016, corresponding to an integrated luminosity of 35.9 fb^{-1} .

The tau lepton is the only lepton heavy enough to decay hadronically, with a branching ratio of 65%. The reconstruction of hadronic decay taus (τ_h) is achieved in CMS by the Hadron-Plus-Strips (HPS) algorithm. The principle is to relate the τ_h candidates to one of the possible tau hadronic decay modes by counting the number of tracks and deposits in $\eta - \phi$ strips of the electromagnetic calorimeter. Discriminators to reject quark-gluon jets, electron and muons have also been applied. The $e \rightarrow \tau_h$ misidentification rates are kept at percent level or below and $\mu \rightarrow \tau_h$ misidentification rates at permille level or below. The performance of τ identification algorithms is studied with data. It is compatible with the prediction in simulations, but some data-to-Monte Carlo scale factors not equal to unity are derived. These scale factors which were derived in this thesis, are recommended to be used for all CMS analyses involving $e/\mu \rightarrow \tau_h$ misidentification. The uncertainties in the $e \rightarrow \tau_h$ misidentification rates in data ranges from a few percent for very loose selection criteria in the barrel region to 45% for very tight selection criteria in the endcap region. The uncertainties on the $\mu \rightarrow \tau_h$ misidentification rates in data range from a few percent for loose selection criteria in the central region ($|\eta| < 0.4$) to 60% for very tight selection criteria in forward region ($|\eta| > 1.7$). The selection efficiencies for rejecting electrons and muons in data are verified and they are compatible with the simulation.

The decay of the Higgs boson with a mass of 125 GeV compatible with the SM prediction to $\tau\tau$ is discovered with an observed significance of 4.9 standard deviations, to be compared with an expected significance of 4.7 standard deviations. The best-fit signal strength μ is $1.09^{+0.27}_{-0.26}$. The results are obtained by combining four channels: $e\mu$, $\tau_h\tau_h$, $\mu\tau_h$ and $e\tau_h$. The thesis described the analysis strategies, targeting gluon fusion and vector boson fusion production modes, in detail, with an emphasis on the $e\mu$ channel. The missing transverse energy is highly correlated to the sum of momentum of the $e\mu$ pair in the Higgs to $\tau\tau$ decay, while the correlation is weak in the main background $t\bar{t}$. The D_ζ variable is designed to make use of this property and served as one of the variables to categorize events for further signal extraction. The QCD multijet production, as one of the main background in the $e\mu$ channel, is modeled by using a data-driven technique. The shape of QCD is taken from a sideband relaxed lepton isolation region, while the yield is normalized with the ratio of the number of opposite signed $e\mu$ pair events to the number of same signed $e\mu$ pair events (OS/SS ratio) obtained from the lepton reversed isolation sideband region and with the signal region to sideband region extrapolation factor.

In the search for neutral MSSM Higgs boson $\phi = H/A$, a mass range between 90 GeV and 3.2 TeV is probed, making use of four decay channels of $\tau\tau$: $e\mu$, $\tau_h\tau_h$, $\mu\tau_h$ and $e\tau_h$. Model independent upper limits are set on the cross sections of ϕ for the gluon fusion production ($gg\phi$) and b quark associated production ($bb\phi$) times the branching ratio of $\phi \rightarrow \tau\tau$ decay. No evidence of a neutral MSSM Higgs boson is observed. Likelihood scans in the $gg\phi - bb\phi$ plane have been performed, showing compatibility with the background-only hypothesis. The results are further interpreted in two MSSM benchmark scenarios: $m_h^{\text{mod}+}$ and hMSSM. The 2D Exclusion contours are provided for each scenario, as a function of two MSSM parameters: the ratio of the vacuum expectation values of the two Higgs doublets, $\tan\beta$ and the mass of pseudoscalar boson, m_A . Both scenarios exclude a heavy neutral MSSM Higgs boson of $m_A < 250$ GeV for $\tan\beta$ greater than 6. For values of $\tan\beta < 60$, the exclusion contours reach up to 1.6 GeV for m_A . The analysis strategies are described in the thesis, in particular for the $e\mu$ channel. A refined QCD multijet background estimation method is introduced to the MSSM analysis. The OS/SS ratio used to normalize the yield is dependent on the lepton kinematics. The jet faking prompt e/μ contribution is one of the backgrounds. To better estimate this contribution, data-to-Monte Carlo scale factors, derived from $Z \rightarrow \mu\mu + \text{jets}$ events, are applied.

The LHC has collected proton-proton collision data for an integrated luminosity of 150 fb $^{-1}$ so far. No evidence for physics beyond SM (BSM) has been observed. Future operation of the LHC will gather more and more data with an increased center-of-mass energy, which will allow us to explore more regions in the searches for BSM physics with smaller cross sections or heavier particles and to perform more precise measurements of the properties of the observed 125 GeV Higgs boson. Meanwhile, machine learning techniques are becoming more and more widely used in the physics object reconstruction and in physics analyses by providing more discriminating power to separate signals from backgrounds. With new methods and more data, we have reasons to believe that many exciting discoveries most likely lie in front of us.

Appendices

APPENDIX

A

PUBLICATIONS

1. CMS Collaboration, "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV", JINST 13 (2018), P10005.
2. CMS Collaboration, "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector", Phys. Lett. B 779 (2018) 283.
3. CMS Collaboration, "Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", JHEP 01 (2018), 055.

CHAPTER

11

BIBLIOGRAPHY

- [1] D. Griffiths, “Introduction to Elementary Particles”. Wiley-VCH, 2009. 18
- [2] S. Glashow, “Partial-symmetries of weak interactions”, *Nucl. Phys.* **22** (1961) 579, doi:10.1016/0029-5582(61)90469-2. 18, 24
- [3] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264. 18, 24
- [4] A. Salam, “Weak and Electromagnetic Interactions”, in *Proceeding of the eight Nobel symposium*, N. Svartholm, ed., Elementary particle physics: relativistic groups and analyticity. Almqvist and Wiksell, Stockholm, 1968. 18, 24
- [5] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons”, *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321. 18, 25
- [6] P. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964) 132, doi:10.1016/0031-9163(64)91136-9. 18, 25
- [7] P. W. Higgs, “Broken symmetries and the masses of gauge bosons”, *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508. 18, 25
- [8] G. Guralnik, C. Hagen, and T. Kibble, “Global conservation laws and massless particles”, *Phys. Rev. Lett.* **13** (1964) 585, doi:10.1103/PhysRevLett.13.585. 18, 25
- [9] P. Higgs, “Spontaneous symmetry breakdown without massless bosons”, *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156. 18, 25
- [10] T. Kibble, “Symmetry Breaking in Non-Abelian Gauge Theories”, *Phys. Rev.* (155), no. 1554, doi:10.1103/PhysRev.155.1554. 18, 25

- [11] M. Peskin and D. Schroeder, “An Introduction to Quantum Field Theory”. Westview Press, 1995 edition, 1995. 18, 20, 22
- [12] D. J. Gross and F. Wilczek, “Ultraviolet behavior of non-abelian gauge theories”, *Phys. Rev. Lett.* **30** (1973) 1343, doi:10.1103/PhysRevLett.30.1343. 18, 22
- [13] H. Politzer, “Reliable perturbative results for strong interactions?”, *Phys. Rev. Lett.* **30** (1973) 1346, doi:10.1103/PhysRevLett.30.1346. 18, 22
- [14] Particle Data Group, “Review of particle physics”, *Chin. Phys. C* **40** (2016), no. 100001,. 18, 19, 20, 29, 88
- [15] E. Noether, “Invariante variationsprobleme”, *Nachr. D. Ko ĨĹnig. Gesellsch. D. Wiss. Zu G ĨĹttingen, Math-phys. Klasse* **235** (1918), no. 257,. 19, 21
- [16] CMS Collaboration, “Observation of $t\bar{t}H$ production”, *Phys. Rev. Lett.* **120** (2018), no. 231801,. 28
- [17] CMS Collaboration, “Evidence for the 125 GeV Higgs boson decaying to a pair of τ lepton”, *JHEP* **1405** (2014) 104. 28, 118, 126, 144
- [18] “LHC Higgs Cross Section Working Group”. 29, 30
- [19] J. Gunion, H. E. Haber, G. L. Kane, and S. Dawson, “The Higgs hunter’s guide”. Perseus Publishing, 1990. 29
- [20] A. Djouadi., “The anatomy of electro-weak symmetry breaking- Tome I: The Higgs boson in the standard model.”, *arXiv:0503.172* (2005). 29
- [21] J. Wess and B. Zumino, “Supergauge transformations in four-dimensions”, *Nucl. Phys., B* **70** (1974) 39. 34
- [22] Y. A. Golfand and E. P. Likhtman, “Extension of the algebra of Poincare group generators and violation of p invariance”, *JETP Lett.* **13** (1971) 323. 34
- [23] S. Martin, “A supersymmetry primer”, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1. 34, 36
- [24] F. Zwicky, “Die Rotverschiebung von extragalaktischen Nebeln”, *Helv. Phys. Acta* **6** (1933) 110. 35
- [25] “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation”, *Astrophys. J. Suppl* **192** (2011) 18. 35
- [26] “Planck 2015 results. XIII. Cosmological parameters”, *arXiv:1502.01589* (2015). 35
- [27] G. R. Farrar and P. Fayet, “Phenomenology of the Production, Decay, and Detection of New Hadronic States Associated with Supersymmetry”, *Phys.Lett. B* **76** (1978) 575. 35
- [28] P. Fayet, “Relations Between the Masses of the Superpartners of Leptons and Quarks, the Goldstino Couplings and the Neutral Currents”, *Phys.Lett. B* **84** (1979) 416. 35

-
- [29] S. Dimopoulos and H. Georgi, “Softly Broken Supersymmetry and SU(5)”, *Nucl.Phys. B* **193** (1981) 150. 35
 - [30] P. Fayet, “Spontaneously Broken Supersymmetric Theories of Weak, Electromagnetic and Strong Interactions”, *Phys.Lett. B* **69** (1977) 489. 35
 - [31] P. Fayet, “Supersymmetry and Weak, Electromagnetic and Strong Interactions”, *Phys.Lett. B* **64** (1976) 159. 35
 - [32] G. C. Branco et al., “Theory and phenomenology of two-Higgs-doublet models”, *Phys. Rept.* **516** (2012) 1. 37
 - [33] “The Higgs particles in the MSSM”. (2001). 38
 - [34] A. Djouadi, “The anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model”, *Phys. Rept.* (2008), no. 459, 1. 38
 - [35] M. Carena et al., “MSSM Higgs boson searches at the LHC: benchmark scenarios after the discovery of a Higgs-like particle”, *Eur. Phys. J. C* **73** (2013), no. 9, 2552. 39
 - [36] L. Maiani, A. D. Polosa, and V. Riquer, “Bounds to the Higgs Sector Masses in Minimal Supersymmetry from LHC Data”, *Phys. Lett. B* (2013) 274. 39
 - [37] A. Djouadi et al., “The post-Higgs MSSM scenario: habemus MSSM?”, *Eur. Phys. J. C* **73** (2013) 2650. 39
 - [38] A. Djouadi et al., “Fully covering the MSSM Higgs sector at the LHC”, *JHEP* **06** (2013) 168. 39
 - [39] E. Bagnaschi and et al., “Benchmark scenarios for low $\tan\beta$ in the MSSM”, *Tech. rep. LHCHSWG-2015-002*, CERN (2015). 39
 - [40] the CMS Collaboration, the ATLAS Collaboration, and The LHC Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, technical report, CERN, 2011. 42, 47, 48, 49
 - [41] A. L. Read, “Linear interpolation of histograms.”, *Nucl. Instrum. Meth.* **A425** (1999) 357. 44
 - [42] R. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples”, *Comput. Phys. Commun.* **77** (1993) 219. 44
 - [43] E. Gross and O. Vitells, “Trial factors or the look elsewhere effect in high energy physics”, *Eur. Phys. J.* **C70** (2010) 525. 46
 - [44] R.B. Davies, “Hypothesis testing when a nuisance parameter is present only under the alternative”, *Biometrika* **74** (1987) 33. 46
 - [45] A. L. Read, “Presentation of search results: The CL(s) technique”, *J. Phys.* **G28** (2002) 2693. 47
 - [46] A. L. Read, “Modified frequentist analysis of search results (the CLs method)”, (2000). CERN-OPEN-2000-205. 47

- [47] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth.* **A434** (1999) 435. 47
- [48] G. Cowan et al, “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J.* (2011), no. C71, 1554. 48, 50
- [49] A. Hocker et al., “TMVA - Toolkit for multivariate data analysis”, *PoS ACAT* (2007) 40. 49, 51
- [50] B. Roe et al, “Boosted decision trees, an alternative to artificial neural networks”, *Nucl. Instrum. Meth.* **A543** (2005) 577. 49
- [51] C. Lefèvre, “The CERN accelerator complex. Complexe des accélérateurs du CERN”, (Dec, 2008). <https://cds.cern.ch/record/1260465>. 56
- [52] Gargamelle Neutrino Collaboration, “Search for elastic muon-neutrino electron scattering”, *Phys. Lett. B* **46** (1973) 121. 55
- [53] Gargamelle Neutrino Collaboration, “Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment”, *Phys. Lett. B* **46** (1973) 138. 55
- [54] UA1 Collaboration, “Experimental observation of isolated large transverse energy electrons with associated missing energy at $s=540$ GeV”, *Phys. Lett. B* **122** (1983) 103. 55
- [55] U. 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. 55
- [56] UA1 Collaboration, “Experimental Observation of Lepton Pairs of Invariant Mass Around $95\text{-GeV}/c^2$ at the CERN SPS Collider”, *Phys. Lett. B* **126** (1983) 398. 55
- [57] UA2 Collaboration, “Evidence for $Z^0 \rightarrow e^+ e^-$ at the CERN anti-p p Collider”,. 55
- [58] NA31 Collaboration, “First Evidence for Direct CP Violation”, *Phys. Lett. B* **206** (1988) 169. 55
- [59] NA31 Collaboration, “A new measurement of direct CP violation in the neutral kaon system”, *Phys. Lett. B* **317** (1993) 233. 55
- [60] NA48 Collaboration, “A new measurement of direct CP violation in two pion decays of the neutral kaon”, *Phys. Lett. B* **465** (1999) 335. 55
- [61] 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. 55
- [62] 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. 55
- [63] L. Evans and P. Bryant, “LHC machine”, *JINST* **3** (2008), no. S08001,. 56, 58, 59

-
- [64] ATLAS Collaboration. 57
 - [65] the CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008), no. S08004,. 60, 61, 63, 64, 67, 70
 - [66] the ATLAS Collaboration, “The ATLAS experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008), no. S08003,. 60
 - [67] the ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008), no. S08002,. 60
 - [68] the LHCb Collaboration, “The LHCb detector at the LHC”, *JINST* **3** (2008), no. S08005,. 60
 - [69] LHCf collaboration, “The LHCf detector at the CERN Large Hadron Collider”, *JINST* **3** (2008), no. S08006,. 60
 - [70] MoEDAL collabration, “Technical design report of the MoEDAL experiment”, technical report, CERN, 2009. 60
 - [71] TOTEM collaboration, “The TOTEM Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008), no. S08007,. 60
 - [72] CMS Collaboration.
 - [73] J. Butler et al., “CMS Phase II upgrade scope document”, Technical Report CERN-LHCC-2015-019, CERN, 2015. 61
 - [74] CMS Collaboration, “CMS Physics: Technical Design Report”, Technical Report CERN-LHCC-2006-001, 2006. 63, 65, 68, 82
 - [75] P. Adzic et al., “Energy resolution of the barrel of the CMS electromagnetic calorimeter”, *JINST* **2** (2007), no. P04004,. 66
 - [76] CMS Collaboration, “Missing transverse energy performance of the CMS detector”, *JINST* **6** (2001), no. P09001,. 66
 - [77] A. Buckley et al., “General-purpose event generators for LHC physics”, *Phys. Rept* **504** (2011) 145. 72, 75, 76
 - [78] T. Gleisberg et al., “Event generation with SHERPA 1.1”, *JHEP* **0902** (2009) 7. 73
 - [79] R. K. Ellis, et al., “QCD and collider physics”, *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol* **8** (1996) 1. 74
 - [80] V. N. Gribov, L. N. Lipatov, “Deep inelastic e-p scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15** (1972) 438. 74
 - [81] Y. L. Dokshitzer, “Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics”, *Sov. Phys. JETP* **46** (1977) 641. 74

- [82] G. Altarelli, G. Parisi, “Asymptotic freedom in parton language”, *Nucl. Phys. B* **126** (1977) 298. 74
- [83] The NNPDF collaboration, “Parton distributions for the LHC run II”, *JHEP* (2015), no. 40,. 74
- [84] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton fragmentation and string dynamics”, *Phys. Rept.* **97** (1983) 31. 75
- [85] T. Sjöstrand, et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159. 76
- [86] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, “Parton fragmentation and string dynamics”, *Phys. Rept.* **97** (1983) 31. 76
- [87] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2015) 155. 76
- [88] P. Z. Skands et al., “SUSY Les Houches accord : Interfacing SUSY spectrum calculators, decay packages, and event generators”, *JHEP* **0407** (2004) 36. 76
- [89] J. Alwall et al., “MadGraph 5: Going Beyond”, *JHEP* **1106** (2011) 128. 76
- [90] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079. 77
- [91] P. Artoisenet et al., “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”, *JHEP* **2013:15** (2013). 77
- [92] P. Nason and B. Webber, “Next-to-Leading-Order Event Generators”, *Annual Review of Nuclear and Particle Science* **62** (2012) 187. 77
- [93] Z. Was, “TAUOLA for simulation of tau decay and production: perspectives for precision low energy and LHC applications”, *Nucl. Phys. B* **218** (2011) 249. 77
- [94] S. Agostinelli et al., “GEANT4: A simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250. 77
- [95] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017), no. P10003,. 78, 79
- [96] D. Buskulic et al., “Performance of the ALEPH detector at LEP”, *Nucl. Instrum. Meth. A* **360** (1995) 481. 78
- [97] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. P10009,. 78
- [98] CMS Collaboration, “Average pileup in 2016”, 2018. 80
- [99] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. P10009,. 78

-
- [100] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems”, *Proceedings of the IEEE* (1998) 2210. 78
 - [101] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, “Adaptive vertex fitting”, *J. Phys.* **G34** (2007) N343. 80, 86
 - [102] CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. P06005,. 80, 81, 88
 - [103] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018), no. P06015,. 82, 88
 - [104] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017), no. P02014,. 84, 85, 135, 158
 - [105] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *Eur. Phys. J. C.* **72** (2012) 1896. 84
 - [106] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04** (2008) 063. 84, 85
 - [107] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018), no. P05011,. 86
 - [108] W. S. Sarle, ed., “Neural networks and statistical models”. (1994). 86
 - [109] CMS Collaboration, “Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. P02006,. 86
 - [110] CMS Collaboration, “Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018), no. P10005,. 88, 90, 93, 95, 97, 104, 106, 109, 111, 160
 - [111] CMS Collaboration, “Tau-Id performance with full 2016 dataset using $Z \rightarrow \tau_\mu \tau_h$ events”, Technical Report CMS DP-2017/002, CMS Collaboration, 2017. 92
 - [112] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, (2017). CMS-PAS-LUM-17-001. 97, 137, 159
 - [113] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015), no. 040,. 98
 - [114] OPAL Collaboration, “Search for the Standard Model Higgs boson in e^+e^- collisions at $\sqrt{s} = 192 - 209$ GeV”, *Phys. Lett. B* **499** (2001) 38. 118
 - [115] L3 Collaboration, “Standard model Higgs boson with the L3 experiment at LEP”, *Phys. Lett. B* **517** (2001) 319. 118

- [116] DELPHI Collaboration, “Final results from DELPHI on the searches for SM and MSSM neutral Higgs bosons”, *Eur. Phys. J. C* **32** (2004) 145. 118
- [117] ALEPH Collaboration, “Observation of an excess in the search for the Standard Model Higgs boson at ALEPH”, *Phys. Lett. B* **495** (2000) 1. 118
- [118] CDF Collaboration, “Search for a low-mass standard model Higgs boson in the $\tau\tau$ decay channel in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV”, *Phys. Rev. Lett.* **108** (2012), no. 181804,. 118
- [119] D0 Collaboration, “Search for the standard model Higgs boson in tau lepton final states”, *Phys. Lett. B* **714** (2012) 237. 118
- [120] ATLAS Collaboration, “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”, *JHEP* **04** (2015) 117. 118
- [121] ATLAS and CMS Collaboration, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **08** (2016), no. 045,. 118
- [122] CMS Collaboration, “Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J.* **C75** (2015), no. 11, 542. 123
- [123] D. Jang, “Search for MSSM Higgs decaying to tau pairs in $p\bar{p}$ collision at $\sqrt{s} = 1.96$ TeV at CDF”. PhD thesis, Rutgers University, 2006. 126
- [124] CMS Collaboration, “Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **772** (2017) 752. 136, 158
- [125] CMS Collaboration, “Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 13$ TeV”, *Phys. Lett. B* **766** (2017) 268. 136, 158
- [126] CMS Collaboration, “Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector”, *Phys. Lett. B* **779** (2018) 283. 132, 138, 139, 141, 142, 143, 145, 146
- [127] DELPHI, OPAL, ALEPH, L3 and LEP Working Group for Higgs Boson Searches Collaboration, “Search for neutral MSSM Higgs bosons at LEP”, *Eur. Phys. J. C* **47** (2006), no. 547,. 148
- [128] CDF Collaboration, “Search for Higgs bosons predicted in Two-Higgs-Doublet models via decays to tau lepton pairs in 1.96 TeV $p\bar{p}$ collisions”, *Phys. Rev. Lett.* **103** (2009), no. 201801,. 148
- [129] D0 Collaboration, “Search for Higgs bosons decaying to $\tau^+\tau^-$ pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV”, *Phys. Lett. B* **707** (2017), no. 323,. 148
- [130] CMS Collaboration, “Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions”, *JHEP* **10** (2014), no. 160,. 148, 167

-
- [131] ATLAS Collaboration, “Search for additional heavy neutral Higgs and gauge bosons in the ditau final state produced in 36 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector”, *JHEP* **01** (2018), no. 055,. 148, 167, 169
- [132] CMS Collaboration, “Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ”, *JHEP* **09** (2018), no. 007,. 151, 156, 161, 162, 163, 164, 165, 166, 167, 168, 169
- [133] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the LHC”, *Eur. Phys. J. C* **63** (2009) 189. 159
- [134] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Uncertainties on α_s in global PDF analyses and implications for predicted hadronic cross sections”, *Eur. Phys. J. C* **64** (2009) 653. 159
- [135] D. de Florian et al., “Handbook of LHC Higgs cross sections 4. deciphering the nature of the Higgs sector”, Technical Report CERN Report CERN-2017-002-M, CERN, 2016. 159

ACKNOWLEDGEMENTS

First of all I would like to thank Elisabetta Gallo. At every moment she has been present, supportive, and has given me the best advice. This thesis would not have been the same without her, and I want to thank her for the careful review of the previous versions.

I also want to thank Alexei Raspereza for being my supervisor and a good friend. He has taught me so many things, and has always tried to give me visibility and responsibilities, from the very beginning of my Ph.D. It has always been a pleasure to work with him.

I am grateful to the jury members, Elisabetta Gallo, Alexei Raspereza, Dieter Horns, Gudrid Moortgat-Pick and Christian Sander for taking part in my disputation and evaluating my work.

I would like to thank Qiang Li of Peking University for encouraging me to applied DESY Ph.D program. Without his guidance in my master period, I would not have walked this far.

I am also grateful to the support from the DESY CMS Higgs group. In particular, Dr. Rainer Mankel for many supports and fruitful discussions, Dr. Mareike Meyer for helping me with the German version of the abstract, Dr. Teresa Lenz for many useful discussions.

I have a wonderful time working in DESY. In particular, I would like to thank my colleagues, Chayanit Asawatangtrakuldee, Andrej Saibel, Rostyslav Shevchenko, Patrick Conner, Sandra Consuega Rodriguez, Danyer Perez Adan, Qun Wang, Isabell Melzer-Pellmann.

I want to express my gratitude to my colleagues in the CMS TAU community, Fabio Colombo, Cécile Caillol, Yuta Takahashi, Isobel Ojalvo, Jan Steggemann, Michal Bluj, Camilla Galloni, Christian Veelken, Roger Wolf, Aruna Nayak, Riccardo Mazoni, Hale Sert, Olena Hlushchenko, Martin Flechl, Artur Gottmann, and many others. They made up such a nice atmosphere and I am very happy to work with them.

At this point, I would like to thank my office mates, Valeria Botta, Ilya Bobovnikov, Andrea Cardini, Antonio Vagnerini. I very much enjoyed the discussion of physics and everything.

I would like to thank my landlord Mrs. and Mr. Ruchhöft for helping me living at Germany, also as friendly neighbor.

And the last word goes to my parents and my brother, who always support me both ma-

Chapter 11. Bibliography

terialy and mentally. My deepest gratitude goes to Ruoxi Zhang who brightens my everyday, thank you!
