Shiraz University

Faculty of Science

Ph.D. Thesis in Physics (Fundamental Particles and High Energy Physics)

Tau at CMS

(From Tau Lepton and Tau in SM measurements to Tau Role in Discovery)

By

Abdollah Mohammadi

Shiraz University and Institute for Fundamental Science(IPM), Iran

Supervised by

Alexandre Nikitenko (Imperial College, UK)

Azizollah Azizi (Shiraz University, Iran)

October 2011

Abstract

Large Hadron Collider (LHC) is the world's largest particles accelerators which enables the exploration of physics at the TeV scale. Compact Muon Solenoid (CMS) as one of the multi-purpose detectors aims to test the Standard Model (SM) measurements and search for the missing part of it (Higgs boson) as well as new physics. Tau lepton plays a significant role in CMS analysis. Exploiting the advanced methods for tau reconstruction and identification is a crucial tool to test the standard model measurements (such as $Z \to \tau\tau$ and $W \to \tau\nu$ cross-section) as a test for lepton universality. In addition, tau lepton incorporates in several analyses involving search for standard model Higgs boson and beyond standard model. In this thesis, after investigating the properties of hadronic decay of tau, we proceed with the measurement of cross section production of $W \to \tau\nu$ and finally search for standard model Higgs boson decaying to a pair of Z boson with final state of 2 light leptons and two taus has been performed.

Acknowledgments

First and foremost I offer my sincerest gratitude to my supervisor, Alexandre(Sasha) Nikitenko, who has supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. It has been an honor to be his Ph.D. student.

I would like to thank the Institute for Fundamental Science, school of particles and accelerators (IPM), Farhad Ardalan, Hesammedin Arfaei and my advisor Mohsen Khakzad for all their support and help. I warmly thank all my professors from department of Physics at Shiraz University which helped me to do my Ph.D research at CERN especially, Azizollah Azizi (my supervisor at Shiraz university), Mohammad Zebarjad, Nader Ghahramany, Nematollah Riazi and Hamid Nadgaran.

I owe my deepest gratitude to Daniel Denegri for all his guidance and motivation and especial thanks to Giuseppe Bagliesi, Chiara Mariotti, Michail Bachtis, Monica Vazquez Acosta, Liis Rebana, and Nicola De Philips which I had lots of fruitful discussion with them. I am indebted to Christian Veelken, who never let me stuck with my questions.

I must also acknowledge my friends and colleagues from IPM, Majid Hashemi, Saeid Paktinat and those who I have been with during my PhD at CERN, Ali Fahim, Hamed Bakhshian, Maryam Zeinali and Nadjie Jafari. I had a very fantastic experience being with them and learn from them.

And last but no the least I would like to thank my family for the endless support they provided me through my entire life and in particular, I must acknowledge my wife Afrooz, without whose love, encouragement and assistance, I would not have finished this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The purpose of Standard Model(SM) is to describe the behavior of all known subatomic particles. All parts of Standard Model have been well tested experimentally and the only unobserved particle is the Higgs boson which give mass to quarks, leptons and gauge bosons. Large Hadron Collider (LHC) is the world's largest and the highest-energy particle accelerator and enables the exploration of physics at the TeV scale. Compact Muon Solenoid (CMS) is one of the two multi-purpose detectors at the LHC. One of the primary motivations of this experiment is to evaluate the consistency of the Standard Model at the new energy frontier opened up by the LHC. The other goal of CMS experiment is to investigate for the last piece of SM and search for new phenomena such as Super Symmetric particles (SUSY), or new exotic particles.

Tau leptons play a significant role in physics analyses performed by the CMS experiment. Studies covering diverse topics ranging from electroweak measurements to searches for Higgs bosons, supersymmetric particles and other new physics phenomena benefit from efficient tau reconstruction and powerful background rejection. In about two thirds of cases taus decay hadronically, typically into either one or three charged mesons (predominantly π^+,π^-) in the presence of up to two neutral pions, decaying via $\pi^0 \to \gamma\gamma$, and a tau neutrino. This leads to an experimental signature that is similar to that expected for quark and gluon jet production. Due to the large jet production cross section, the experimental challenge in reconstructing and identifying hadronic tau decays is to discriminate efficiently between them and jets misreconstructed as tau candidates.

In the framework of the standard model, tau leptons are produced in decays of electroweak vector bosons: $Z \to \tau^+\tau^-$ and $W^\pm \to \tau^\pm\nu$. These processes have relatively large cross sections and are the largest sources of tau leptons at LHC.

The $W \to \tau^{\pm}\nu$ channel benefits from a large production cross section, exceeding the production rate of $Z \to \tau^{+}\tau^{-}$ by nearly an order of magnitude. However, the experimental signature of a single tau-jet and undetected neutrino is challenging, requiring a good understanding of the tau identification and missing transverse energy (MET).

The study of $W \to \tau\nu$ production in the $\tau_{had}\nu$ final state is an important contribution to tau-physics studies at LHC. Also, $W^{\pm} \to \tau^{\pm}\nu$ production has to be well understood as a test of the standard model and as a measure of important background process in several searches for new physics. In particular, it is the major background in the search for the charged Higgs boson in the $\tau\nu$ final state.

The search for a signal compatible with the production of the standard model (SM) Higgs boson was the most important guiding criterion to define the performance requirements of the CMS detector. The discovery of that signal would, in particular, shed light on the mechanism of spontaneous breaking of the electro-weak symmetry. Direct searches for the SM Higgs particle at the LEP $e^{+}e^{-}$ collider have led to a lower mass bound of $m_{\mathrm{H}} > 114.4\,\mathrm{GeV/c^2}$ (95% C.L.) [1]. On-going direct searches at the Tevatron $\mathrm{p\bar{p}}$ collider by the D0 and CDF experiments set exclusion limits for the SM Higgs boson in a mass range between 156 and $177\,\mathrm{GeV/c^2}$ at 95% C.L. [2]. The recent measurements at the LHC excluded SM Higgs at 95% C.L. in the mass ranges 155–190 and 245–450 for the ATLAS and 145–216, 226–288 and 310-400 for the CMS experiment [3, 4].

Tau can play an important role in the SM Higgs boson discovery mainly with $H \to \tau\tau$ for light Higgs and $H \to ZZ \to 2l2\tau$ for heavy Higgs. Studying the $H \to ZZ \to 2l2\tau$ can complement the search in the $H \to ZZ^{(*)} \to 4l$ channel [5] at Higgs masses above the kinematical threshold for ZZ production. With four leptons in the final state, it provides a clean signature with small background contributions. Adding all final states of the Higgs, leads to increasing the sensitivity of the Higgs boson search.

This thesis starts with the description of standard model and production/decay of Higgs boson at the colliders and the most recent results from different experiments. In chapter 2, the LHC machine and several parts of the CMS detector is described. Chapter 3 is devoted to the Generation, Simulation and Reconstruction

of the Monte–Carlo samples and an extensive description of reconstruction of particles. Chapter 4 aims to describe the hadronic decays of tau particles and present the results of tau identification and fake rate. Chapter 5 proceeds with the measurement of the cross–section of W boson in the final state of hadronic tau and finally chapter 6, is the search for the Higgs boson in the process of $H \to ZZ \to 2l2\tau$ with 2.1 fb^{-1} data.

Chapter 2

Standard Model and Higgs Boson

Elementary particle physics addresses the questions, 'What is matter made of?' and 'How do they interact?' Matter at the subatomic level, consist of microscopic particles with vast empty space in between and are categorized in small number of different types: protons, neutrons, electrons, pions, neutrinos and etc. In addition there are forces which causes these particles to interact each other. A model to include these particles and their interactions is so called standard model of elementary particles and by now all of the aspects of this model is known and the last piece of this puzzle that has yet to be observed is Higgs boson which is expected to be the origin of mass as a direct physical manifestation of Higgs mechanism.

2.1 The Standard Model of Elementary Particles

2.1.1 Interactions

As far as we know, there are four fundamental forces in nature. Table 2.1 exhibits these forces together with their coupling constants, mediator and ranges.

Force	Strength	Mediator	Range
Strong	10	Gluon	10^{-15}
Electromagnetic	10^{-2}	Photon	∞
Weak	10^{-13}	W and Z Boson	10^{-18}
Gravitational	10^{-42}	Graviton	∞

Table 2.1: Table of four fundamental forces in the nature [6]

To each of these force there belongs a physical theory. For gravity there exists a classical theory (which is the Newton's law of universal gravitation) and for its relativistic generalization, it is Einstein theory for general relativity. At the moment there in not a satisfactory quantum theory of gravity. This force which acts as the

most important force in large scale (i.e. galaxies), is too weak to play a significant role in elementary particle physics.

The physical theory that describes electromagnetic force is called *electrodynamics*. The classical formulation of this force (which was also consistent with special relativity) was formulated by Maxwell. The quantum theory of electrodynamics(QED) was perfected by Tomonaga, Feynman and Schwinger in the 1940s. QED is one of the most accurate physical theories constructed so far. Based on the comparison of the electron anomalous magnetic dipole moment and the Rydberg constant from atom recoil measurements, there has been found an agreement between theory and experiment within ten parts in a billion.

The weak force was originally described by Fermi in 1933 as the theory of a contact four-fermion interaction; which is to say, a force with no range (i.e. entirely dependent on physical contact). However, it is now best described as a field, having range, albeit a very short range. In 1968 the electromagnetic force and the weak interaction were unified, when they were shown to be two aspects of a single force, known as electro-weak force.

The strong interaction is the interaction between quarks and gluons and since these particles have color, the theory which is a representative of this force is called Quantum ChromoDynamics (QCD). Although QCD is not tested to the same precision as QED, it is nevertheless in impressive agreement with extensive experimental data.

Each of these forces is mediated by the exchanged particles which are called gauge bosons because they have integral spins and their existence and physical behavior are predicted and studied by gauge theories. The gravitational forces are mediated by graviton, electromagnetic forces are mediated by the photons, weak forces by W and Z vector bosons and strong forces are mediated by gluon.

2.1.2 Particles

In the most general classification, there exist two main types of particles: the *matter constituents*, which include quarks and leptons, and the *interaction quanta* which incorporate photons and other gauge bosons particles [7].

Leptons

Leptons are indivisible spin 1/2 particles, without any internal structure. There are six leptons, they occur in pairs, called generations. Each generation has a charged lepton with electric charge $-e$ and a neutral neutrino. The three charges leptons (e^-, μ^-, τ^-) are the familiar electron, together with two heavier particles, which are called muon and tau. Ignoring gravity, the charged leptons participate in only electromagnetic and weak interactions, while for the neutrinos, only weak interactions have been observed and therefore due to their very tiny mass they can be detected with difficult procedures.

Name	Mass[MeV]	Charge	Lifetime [s]	Major decays
e^-	0.511	-1	stable	None
ν_e	< 2 eV	0	stable	None
μ	106	-1	2.2×10^{-6}	$e\bar{\nu}_e\nu_\mu$
ν_μ	< 0.19	0	stable	None
τ	1777	-1	2.9×10^{-13}	leptonic (35%) hadronic (65%)
ν_τ	< 18.2	0	stable	None

Table 2.2: Properties of leptons [7]

The lifetimes and mass of the leptons are summarized in Table 2.2.

As the lightest charged particles and due to the two conservation laws (charged and leptons number), electrons can not decay to other particles and therefore it is stable. Muons decay by the weak interaction (mostly to electrons and neutrinos) with the life time of 2×10^{-6} s.

Among the leptons, taus are interesting particles. Due to their high mass they can decay either leptonically(with Branching Ratio (BR) of about 35%) to electrons and muons or decay hadronically(with BR of about 65%) to charged (and neutral) hadrons. The Feynman diagram for decay of tau is shown in figure 2.1.

Each generation of leptons is associated with a quantum number, called lepton numbers. The first of these lepton numbers is the *electron number*, defines by:

$$L_e = N(e^-) - N(e^+) + N(\nu_e) - N(\bar{\nu}_e) \tag{2.1}$$

where $N(e^-)$ and $N(e^+)$ are the number of electrons and positrons, respectively. For single particles states, $L_e = 1$ for e^- and ν_e, $L_e = -1$ for e^+ and ν_e and it is

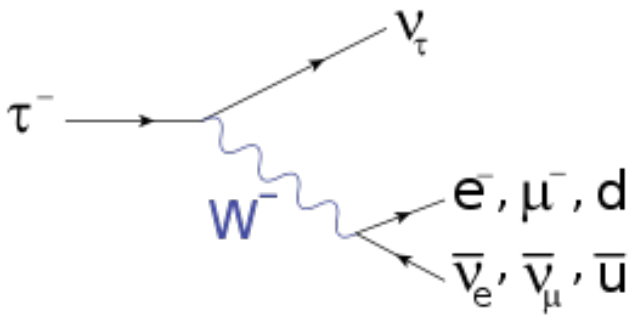

Figure 2.1: Dominant Feynman diagram for the decay of τ

zero for all other particles. Similar way is applied to muon and tau numbers. In the standard model, the lepton number is postulated to be separately conserved in any reaction.

Quarks

Similar to leptons, quarks are indivisible spin 1/2 particles, without any structure. Due to the color confinement, quarks are never directly observed or found in isolation; they can only be found within bound states (hadrons).

Six distinct types, or flavors, of quarks are known to exist. Like the leptons, they occur in pairs. Each type, or generation, consist with a quark with charge $+\frac{2}{3}$, (u, c, t), together with a quark of charge $-\frac{1}{3}$,(d, s, b), in units of e. They are called down (d), up (u), strange (s), charmed (c), bottom (b) and top (t) quarks. The quantum number associated with the s,c,b and t quarks are called strangeness, charm, bottom and top, respectively.

Some properties of quarks are given in Table 2.3. Except for the top quark, these masses are inferred indirectly from the observed masses of their hadron bound states, together with models of quark binding [7].

Associated to any quark, a quark number is defined which is 1 for the quark and and -1 for the anti-quark. Quarks number needs to be conserved in strong and electromagnetic interactions while in the weak interaction, only the total quark number is conserved.

Since quarks can not be found in isolation, they occur in different bound states

Name	Mass[GeV]	Charge	Lifetime [s]	Major decays
down	0.3	-1/3	-	-
up	0.3	2/3	-	-
strange	0.5	-1/3	$10^{-8} - 10^{-10}$	$s \to u + X$
charmed	1.5	2/3	$10^{-2} - 10^{-13}$	$c \to (d/s) + X$
bottom	4.5	-1/3	$10^{-2} - 10^{-13}$	$b \to c + X$
top	171	2/3	10^{-25}	$t \to b + X$

Table 2.3: Properties of quarks [7]

which are called hadrons. Hadrons are mostly classified in two different categories based on the number of constituent quarks:

1. **Mesons** consist of one quark and one anti-quarks which have integer spin. The simple example of mesons are charged pions as a bound state of $u\bar{u}$ and $d\bar{d}$ quarks.

2. **Baryons(anti-baryons)** are the bound states of 3 quarks(anti-quarks) which have half integer spin. The most well-known examples of baryons are protons and neutrons which consist of (uud) and (udd) quarks, respectively.

2.2 Electroweak theory

The electroweak theory was originally proposed to solve problems associated with the Feynman diagrams in which more than one W was exchanged. Figure 2.2 shows a reaction of $e^+\mu^- \to e^+\mu^-$ from exchange of two W bosons.

Such contributions are expected to be small because they are higher order in the weak interactions; however calculation of these higher order contributions leads to divergent integrals. In the unified theory, this problem is solved by involving of diagrams with the exchange of Z^0 and photons and when all the diagrams of a given order are added together the divergences cancel.

This cancellation is not accidental but follows provided that two relations are hold. One of this relation is called 'unification conditions' and relates the weak and electromagnetic coupling constant:

$$\frac{e}{2(2\varepsilon_0)^{1/2}} = g_W \sin\theta_W = g_Z \cos\theta_W \tag{2.2}$$

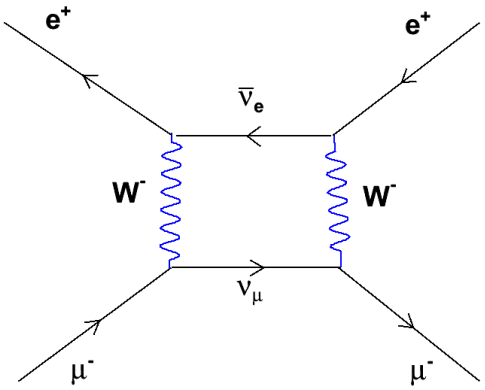

Figure 2.2: Higher order contribution to the reaction $e^+\mu^- \to e^+\mu^-$ from exchange of two W bosons.

where mixing angle $\theta_W = \cos^{-1}(M_W/M_Z)$ is in the range of $(0, \pi/2)$ and g_W and g_Z are coupling constants characterizing the strength of the charged and neutral current vertices, respectively.

The other one is called 'anomaly condition' which relates the electric charges Q_l and Q_a of the leptons l and quarks a:

$$\Sigma_l Q_l + 3\Sigma_a Q_a = 0. \tag{2.3}$$

Where the sum extends over l (all leptons) and a (all quark flavors) and the factor 3 arises because each quark comes in three color states.

To find out the origin of equation(2.2), we make a flash back to the history of electroweak.

For the first time Glashow tried to unify the weak and electromagnetic interaction to combine them as a single theory and as a manifestation of the one fundamental electroweak interaction. The main obstacle in the path through unification of electromagnetic and weak interactions was the huge discrepancy between their coupling constant. It was Weinberg and Salam that finally solve this problem by exploiting the 'Higgs Mechanism'.

One of the structural differences between the electromagnetic and weak vertex factor is that former is purely vectorial (γ^μ), whereas the latter contains vector

and axial vector part $\gamma^\mu(1-\gamma^5)$. A brilliant idea to overcome this problem was to decompose the particle spinor into left-handed and right-handed components and absorb the matrix γ^5 into spinor. So they would be :

$$u_L(p) = \frac{(1-\gamma^5)}{2}u(p) \qquad \text{and} \qquad u_R(p) = \frac{(1+\gamma^5)}{2}u(p)$$

The weak vertex factor (for example in the vertex of electron, neutrino and W boson) will be re-configured from

$$j_\mu^- = \bar{\nu}\gamma_\mu(\frac{1-\gamma^5}{2})e$$

to

$$j_\mu^- = \bar{\nu}_L\gamma_\mu e_L \tag{2.4}$$

which is now purely vectorial, but it couples left-handed electrons to left-handed neutrinos. One may write the electromagnetic analog of the above equation by using chiral spinors ($u = u_L + u_R$) which the fundamental vertex in QED would be:

$$j_\mu^{em} = -\bar{e}\gamma_\mu e = -(\bar{e}_L + \bar{e}_R)\gamma_\mu(e_L + e_R) = -\bar{e}_L\gamma_\mu e_L - \bar{e}_R\gamma_\mu e_R \tag{2.5}$$

Now the equations(2.4) and (2.5) look similar to each-other and this is the first step toward building a unified theory.

The next step is writing the negatively and positively charged weak current in a more compact notation by introducing the left-handed doublet as $\chi_L = \binom{\nu_e}{e}$

as following:

$$j_\mu^\pm = \bar{\chi}_L\gamma_\mu\tau^\pm\chi_L \tag{2.6}$$

where $\tau^\pm$ are the linear combination of the first two Pauli spin matrices. ($\tau^\pm = \frac{1}{2}(\tau^1 \pm i\tau^2)$)

Exploiting the third Pauli spin matrix into the weak current, the third weak current would appear:

$$j_\mu^3 = \bar{\chi}_L \gamma_\mu \frac{1}{2}\tau^3 \chi_L = \frac{1}{2}\bar{\nu}_L \gamma_\mu \nu_L - \frac{1}{2}\bar{e}_L \gamma_\mu e_L \tag{2.7}$$

which is the neutral weak current and couple left-handed particles (neutrino to neutrinos and electron to electrons).

Based on the weak analog of the hypercharge,(Y), which is related to the electric charge(Q) and the third component of the iso-spin (I^3) by ($Q = I^3 + \frac{1}{2}Y$), the 'weak hypercharge' current is written as:

$$j_\mu^Y = 2j_\mu - 2j_\mu^3 = -2\bar{e}_R \gamma_\mu e_R - \bar{e}_L \gamma_\mu e_L - \bar{\nu}_L \gamma_\mu \nu_L \tag{2.8}$$

which is invariant under $SU(2)_L \otimes U(1)_Y$ where $SU(2)_L$ refers to weak iso-spin(involving only left-handed states) and $U(1)_Y$ refers to weak hypercharge (involving both chiralities).

Based on GWS model, the three weak iso-spin currents couple, with strength g_w, to a weak isotriplet of vector boson W, whereas the weak hypercharge current couples with strength $g'/2$ to an isosinglet B:

$$-i[g_w \mathbf{j}_\mu . \mathbf{W}^\mu + \frac{g'}{2} j_\mu^Y B^\mu] \tag{2.9}$$

Now, the last step is to extract the EWK bosons from the above current.

A linear combination of $W^{1,2}$ would corresponds to the charged bosons, $W^\pm$:

$$W_\mu^\pm = (1/\sqrt{2})(W_\mu^1 \mp i W_\mu^2) \tag{2.10}$$

while the combination of the W^3 and B leads to the neutral bosons Z and γ:

$$\begin{aligned} Z_\mu &= -B_\mu \sin\theta_w + W_\mu^3 \cos\theta_w \\ A_\mu &= B_\mu \cos\theta_w + W_\mu^3 \sin\theta_w \end{aligned} \tag{2.11}$$

θ_w is called the 'weak mixing angle'.

Consistency of unified electroweak theory with ordinary QED necessitate the following relation:

$$g_w \sin\theta_w = g' \cos\theta_w = g_e \tag{2.12}$$

which is very similar to the unification condition mentioned earlier.

Now the main question that is addressed here is 'by which mechanism is the $SU(2)_L \otimes U(1)_Y$ symmetry of electroweak interaction broken?'

The answer to this question would be in next chapter.

2.3 Higgs Boson

Higgs boson is the last missing piece of the Standard Model theory that so far has not been observed. The field associated to the Higgs Boson could be responsible for the mass of all other particles and keep the massless particles intact.

Higgs field is a field that fills the universe like a water fills a pool. As particles move through the universe they acquire mass by interacting with the Higgs field. One way to imagine the Higgs field is to imagine trying to walk through a pool. The water pushes against you making you feel heavier, and making it harder for you to move. This effectively generates inertia or mass. Of course, one can come out of the pool and walk normally, but particles can never escape the Higgs field since it is everywhere, including the vacuum of space.

2.3.1 Spontaneous Symmetry Breaking

Given a Lagrangian $\mathcal{L}$ possessing a particular symmetry(i.e. invariant under transformation), two situations can occur in classifying the lowest energy level of the system [8]:

- non-degenerate $\rightarrow$ ground state is unique and possess the symmetries of $\mathcal{L}$
- degenerate $\rightarrow$ no unique eigenstate to represent the ground state

In the second case, if one of the degenerate states (of the minimum) is selected, then the ground state would not share the symmetries of $\mathcal{L}$ which in the other word, spontaneous asymmetry breaking has occurred. The asymmetry is not due to adding a non-invariant asymmetry term to $\mathcal{L}$, but to the arbitrary choice of one of the degenerate states.

Goldstone model is a simple example of such a phenomena. The following Lagrangian:

$$\mathcal{L}(x) = [\partial^{\mu}\phi^{*}(x)][\partial_{\mu}\phi(x)] - \mu^{2}|\phi(x)|^{2} - \lambda|\phi(x)|^{4} \tag{2.13}$$

with

$$\phi(x) = \frac{1}{\sqrt{2}}[\phi_1(x) + i\phi_2(x)] \tag{2.14}$$

as a complex scalar field is invariant under the global $U(1)$ phase transformation ($\phi(x) \rightarrow \phi'(x) = \phi(x)e^{i\alpha}$) and μ^2 and $\lambda(>0)$ are arbitrary real parameters. The potential energy density of the field is written as:

$$\mathcal{V}(\phi) = \mu^{2}|\phi(x)|^{2} + \lambda|\phi(x)|^{4} \tag{2.15}$$

and its minimum depends on the sign of the μ^2.

1. if $\mu^2 > 0$: $\mathcal{V}(\phi)$ has an absolute minimum at $\phi(x) = 0$ and therefore, spontaneous symmetry breaking can not occur.

2. if $\mu^2 < 0$: $\mathcal{V}(\phi)$ has a local maximum at $\phi(x) = 0$ following by a minimum at $\phi_0 = \left(\frac{-\mu^2}{2\lambda}\right)^{1/2} = \frac{1}{\sqrt{2}}v$.

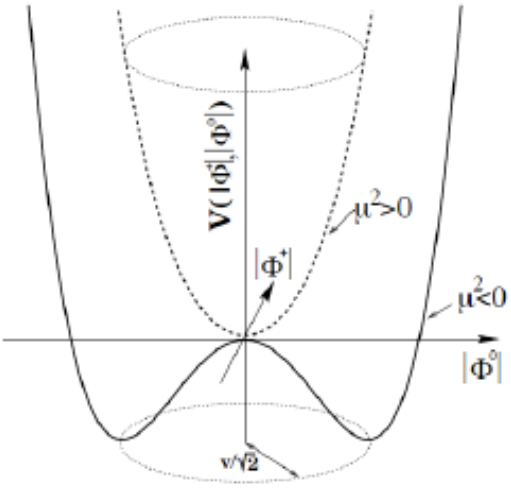

Figure 2.3: The potential energy density $\mathcal{V}(\phi) = \mu^{2}|\phi(x)|^{2} + \lambda|\phi(x)|^{4}$ with $\lambda > 0$ and for two different sign of μ^2. While for the case of $\mu^2 > 0$ the $\phi(x) = 0$ is a stable minimum for the other case it is unstable maximum.

Figure 2.3 illustrates the potential energy for these two different modes.

Rewriting the Lagrangian with :

$$\phi(x) = \frac{1}{\sqrt{2}}[v + \sigma(x) + i\eta(x)] \tag{2.16}$$

where $\sigma(x)$ and $\eta(x)$ are two Klein-Gordon fields corresponding to two neutral spin 0 particles, one of the fields would have real, positive mass of $\sqrt{2\lambda v^2}$ while the other is massless and is known as Goldstone boson.
Since by definition, there is no particles present in the vacuum, the expectation value of the field in vacuum would be as

$$< 0|\phi(x)|0 >= \phi_0. \tag{2.17}$$

Which is the condition of spontaneous symmetry breaking in quantized theory.

2.3.2 Higgs Mechanism

In order the electroweak theory to be invariant under gauge transformation and guarantee the renormalizablity of the theory, all the particles in the theory should be massless. Explicit mass term for fermions, would not violate the gauge invariance, however for GWS model to be invariant under chirality transformation, fermions are required to be massless. This picture is against our observation from experiments which require leptons and three out of four vector bosons to be massive. The process that leptons and $W^{\pm}$ and Z^0 acquire mass in the EWK theory is called Higgs mechanism. Such mechanism is accomplished by means of a doublet of complex scalar field [8].

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix} \tag{2.18}$$

which is introduced in the EWK Lagrangian:

$$\mathcal{L}_{EWSB} = [D^\mu \phi]^\dagger [D_\mu \phi] + \mathcal{V}(\phi^\dagger \phi) \tag{2.19}$$

where the the $D_\mu = \partial_\mu - igt_a W_\mu^a + \frac{i}{2} g' Y B_\mu$ is the covariant derivatives. The above Lagrangian is invariant under the $SU(2)_L \otimes U(1)_Y$. The Potential term $\mathcal{V}$ is similar to the equation (2.15) where for a choice of $\lambda > 0$ and $\mu^2 < 0$, there would

be a stable minimum for

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}. \tag{2.20}$$

Choosing a particular value ϕ for the ground state, leads to spontaneous symmetry breaking. Without loss of generality we can choose

$$< \phi >= \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix}, \qquad v^2 = -\frac{\mu^2}{\lambda} \tag{2.21}$$

The ϕ filed can thus be rewritten in a generic gauge, in terms of its vacuum expectation value:

$$\phi = \frac{1}{\sqrt{2}}e^{\frac{i}{v}\phi^a\tau_a}\begin{pmatrix}0\\H+v\end{pmatrix}, \qquad a = 1,2,3 \tag{2.22}$$

where the three fields ϕ^a and the fourth $\phi^4 = H + v$ are called the Goldstone fields. The unitary gauge is then fixed by the transformation

$$\phi' = e^{-\frac{i}{v}\phi^a\tau_a}\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\\phi^4\end{pmatrix} \tag{2.23}$$

The remaining field, the Higgs field, has now a zero expectation value. Rewriting the Lagrangian with the ϕ field in the unitary gauge, the $\mathcal{L}_{EWSB}$ can be written as:

$$\mathcal{L}_{EWSB} = \mathcal{L}_{HH} + \mathcal{L}_{HW} + \mathcal{L}_{HZ} \tag{2.24}$$

where the first term would be the Higgs term and the other two are vector boson mass and their couplings to Higgs field.

$$\begin{aligned}\mathcal{L}_H &= \frac{1}{2}\partial_\mu H\partial^\mu H + \mu^2H^2\\ \mathcal{L}_{HW} &= \frac{1}{4}v^2g^2W^\mu W_\mu^\dagger + \frac{1}{2}vg^2HW^\mu W_\mu^\dagger\\ \mathcal{L}_{HZ} &= \frac{1}{8}v^2(g^2+g'^2)Z_\mu Z^\mu + \frac{1}{4}v(g^2+g'^2)HZ_\mu Z^\mu\end{aligned} \tag{2.25}$$

From the above equation, we can find the mass of bosons and their couplings to the Higgs field.

Vector Boson Mass

$$\begin{aligned} m_W &= \frac{1}{2}vg \\ m_Z &= \frac{1}{2}v\sqrt{g^2+g'^2} \\ \frac{m_W}{m_Z} &= \frac{g}{\sqrt{g^2+g'^2}} = \cos\theta_W \end{aligned} \tag{2.26}$$

Vector Boson Coupling

$$\begin{aligned} g_{WH} &= \frac{1}{2}vg^2 = \frac{2}{v}m_W^2 \\ g_{HZ} &= \frac{1}{2}v(g^2+g'^2) = \frac{2}{v}m_Z^2 \end{aligned} \tag{2.27}$$

Based on the above relation, the ratio of decay of $H \to WW$ to $H \to ZZ$ would be:

$$\frac{BR(H \to W^+W^-)}{BR(H \to ZZ)} = \Big(\frac{g_{HW}}{1/2 g_{ZH}}\Big)^2 = 4\Big(\frac{m_W^2}{m_Z^2}\Big)^2 \simeq 2.4 \tag{2.28}$$

2.3.3 Higgs Production

There are several mechanisms in the proton-proton collider that Higgs can be produced. These are shown in Figure 2.4.

1. **Gluon-Gluon Fusion**

   Among all Higgs boson production mechanisms at the LHC, the gluon-gluon fusion is the dominant one over whole mass range, due to the high luminosity of gluons in p-p collisions. This process is performed via a top quark triangle procedure. The large cross section production is mainly due to the large coupling constant of Higgs and top quark. There is about 15 to 20% uncertainty on the cross section of Higgs from gluon-gluon fusion mostly due to the parton density function.

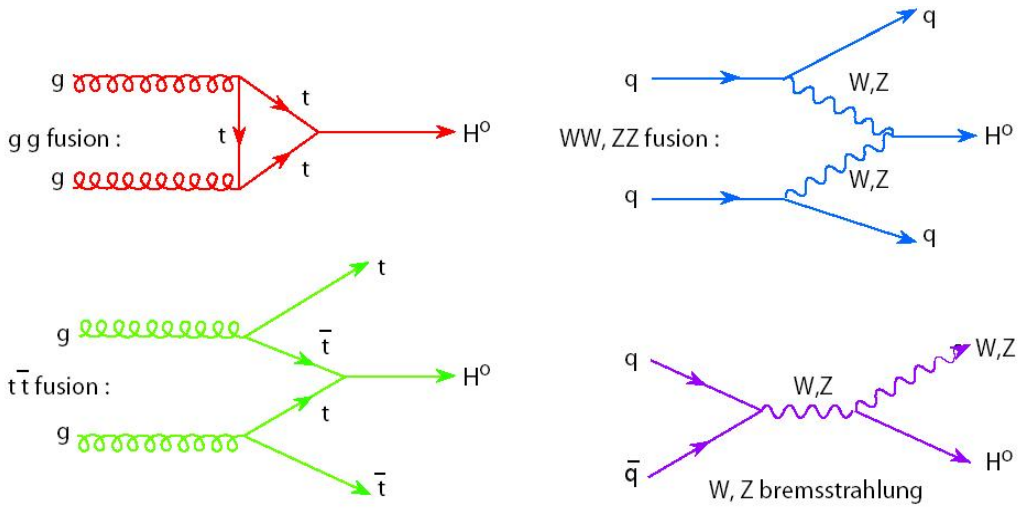

Figure 2.4: Different mechanisms of the Higgs boson production at the pp collisions.

2. **Vector Boson Fusion**

   The vector boson fusion (VBF) is the second origin of the Higgs boson production which its cross section is approximately, one order of magnitude less than gluon-gluon fusion in most of the Higgs mass regions; however the cross section of this mechanism become comparable to Gluon-Gluon fusion for the high masses of Higgs (about 1 TeV). Due to the presences of two forward jets in the large psudorapidity regions, this channel has an advantage in terms of suppressing more backgrounds and increasing signal over background ratio. The uncertainty in this process is at the order of 10%.

3. **Associated Production**

   The third possible production of Higgs boson in the Standard Model is the Higgsstrahlung where a vector boson (either $W^{\pm}$ or Z^0) is produced in association with the higgs. The cross section for this processes is even less than VBF and there is about 20-25% uncertainty for this process of production.

   The last possible process is the associated production in the presences of a pair of $t\bar{t}$, albeit with lower cross section but cleaner signature in the presence of two heavy quarks.

Figure 2.5 compares the cross section of different processes for two different energy of center of Mass.

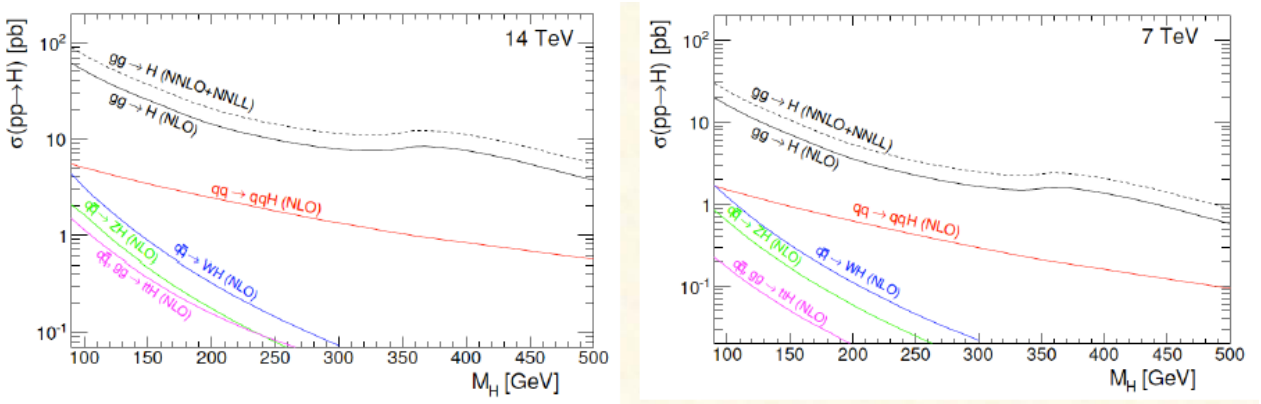

Figure 2.5: Comparison of the cross section of different processes for two different energy of center of Mass, 14 TeV(left) and 7 TeV(right) .

2.3.4 Higgs Decay

The Higgs boson has several decay modes with their branching ratios depending on the mass. Regarding the mass, the Higgs decay modes is divided into three complementary low, intermediate and high mass regions. Fig. 2.6 and 2.7 show the branching fractions of the Higgs boson as a function of mass for the low and high mass decay, respectively.

1. **Low Mass Regions** ($115 < M_H < 140$)
   In the low mass range, dominated by fermions, there are two important decay modes:

   (a) $b\bar{b}$ channel, has the highest branching ratio with very huge QCD di-jet background. Looking for the Higgs in this decay mode is only feasible for the Higgs which are produced in association with W/Z bosons.

   (b) $\gamma\gamma$ channel, has a much lower branching ratio but resemble as a very clean channel with only very few backgrounds from $q\bar{q} \to \gamma\gamma$ (irreducible) and $Z \to e^+e^-$(reducible). Assuming that Higgs exist in the region, the mass can be measured with high precision.

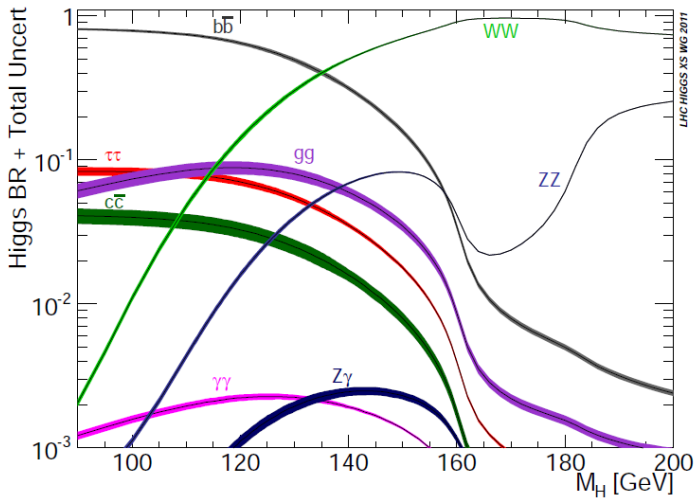

Figure 2.6: Different decay modes of the Higgs for the mass less than 200 GeV. As it is shown the branching ratio of each mode depends on the Higgs mass [9]

.

2. **Intermediate Mass Regions** ($140 < M_H < 180$)

   In the intermediate mass range the main channels are the decay of Higgs into pairs of W and Z bosons.

   (a) Branching ratio of $H \to WW^*$ is always higher than $H \to ZZ^*$ particularly in the mass regions between $2M_W$ and $2M_Z$ where the Higgs can decay to two on-shell W and can not decay to on-shell Z. The branching ratio in this region is very close to 1. Amongst all decay of WW^* the decay into $2l2\nu$ is the most promising channel for the discovery based on its low backgrounds; however due to the presence of neutrinos in the final states, it is almost very difficult to measure the Higgs mass with the precision.

   (b) $H \to ZZ^* \to 4l$ is also important in this region and in comparison to WW^* channel, despite the lower branching ratio, it can measure the Mass with high precision due to the presence of the 4 leptons.

3. **High Mass Regions**($M_H > 180$)

   And finally for the Higgs in high mass range, decay to pairs of W and Z bosons are dominant.

(a) $H \rightarrow ZZ$ is the most important one, even though its branching ratio is less than $H \rightarrow WW$, it provides a clean signature. Not only $4l$ but also the other decay modes of ZZ, like $2l2q$, $2l2\nu$ and even $2l2\tau$ become important in this range. With the high precision measurement for the Higgs mass in $4l$ final state, this channel is usually called the "golden channel".

(b) In the high mass regions, in addition to the $H \rightarrow WW \rightarrow 2l2\nu$, the $H \rightarrow WW \rightarrow 2ql\nu$ mode also become important especially for the boosted Higgs where 2 high energetic jets in association with one lepton can suppress the background and leads to reasonable signal over background ratio.

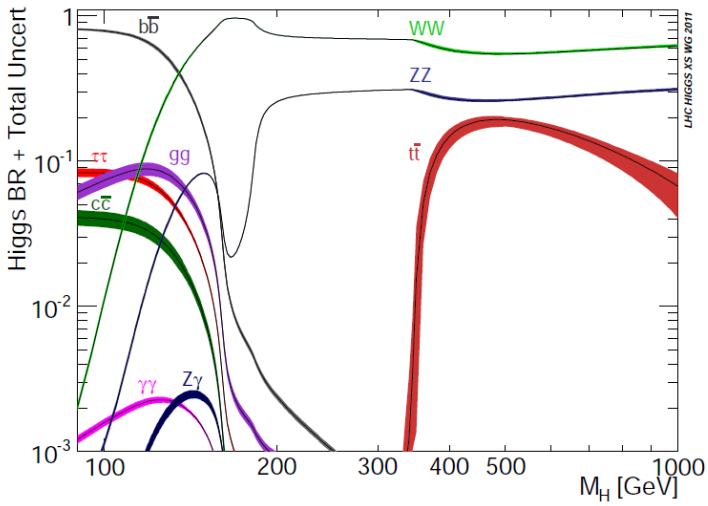

Figure 2.7: Different decay modes of the Higgs for all masses. For the Higgs at high masses mainly decay to vector bosons are dominant [9]

2.4 LHC is a Higgs Discovery Machine!?

The implication of the Higgs mechanism(which was discussed in section 2.3.2) is the existence of a single neutral scalar particle, the Higgs boson.

The mass of Higgs particle is not specified, but indirect experimental limits are obtained from precision measurements of the electroweak parameters which depend

logarithmically on the Higgs boson mass through radiative corrections. In addition to indirect search, there have been several direct searches for Higgs boson in the past, in the LEP and Tevatron accelerators. While LEP has terminated its mission in 2000, Tevatron was still doing search and is going to stop the data taking by the end of the September 2011. However the more promising experiment that is eligible to unreveal the questions from existence of the Higgs boson is LHC.

2.4.1 Indirect Search of the Higgs

Although the Higgs boson mass is not predicted by the Standard Model it may be constrained by fitting the model to the observed electroweak data. In electroweak fits, measured parameters are allowed to vary within their errors and hence indirectly place limits on the range that the mass of the Higgs boson could take to remain compatible with the measurements[10, 11].

Figure 2.8 shows the result of this fit for m_H, the Higgs boson mass. The preferred value for its mass, corresponding to the minimum of the curve, is at 92 GeV, with an experimental uncertainty of +34 and -26 GeV (at 68 percent confidence level derived from $\Delta\chi^2 = 1$ for the black line, thus not taking the theoretical uncertainty shown as the blue band into account). The upper mass limit at 95% confidence level is 185 GeV/c^2.

2.4.2 Higgs Search at the LEP

The Large Electron Positron (LEP) collider was CERN's flagship accelerator from 1989 to 2000. The four LEP experiments carried out extensive searches for the Higgs boson predicted in the context of the Standard Model (SM). Figure 2.9 shows different possible processes that Higgs could be produced at LEP. The prominent Higgs production channel is in association with the Z boson through the Higgsstrahlung ($e^+e^- \rightarrow HZ$) which has a kinematic threshold at $m_h = \sqrt{s} - m_Z$ [14]. Small additional contributions can be added mostly at the end of the kinematic range of the Higgsstrahlung process from W and Z boson fusion, which produce a Higgs boson and a pair of neutrinos or electrons, respectively, in the final state. Figure 2.10 compares the cross section of different procedure of Higgs mass production at $\sqrt{s} = 200$ GeV. The main search topologies are therefore dictated by the dominant

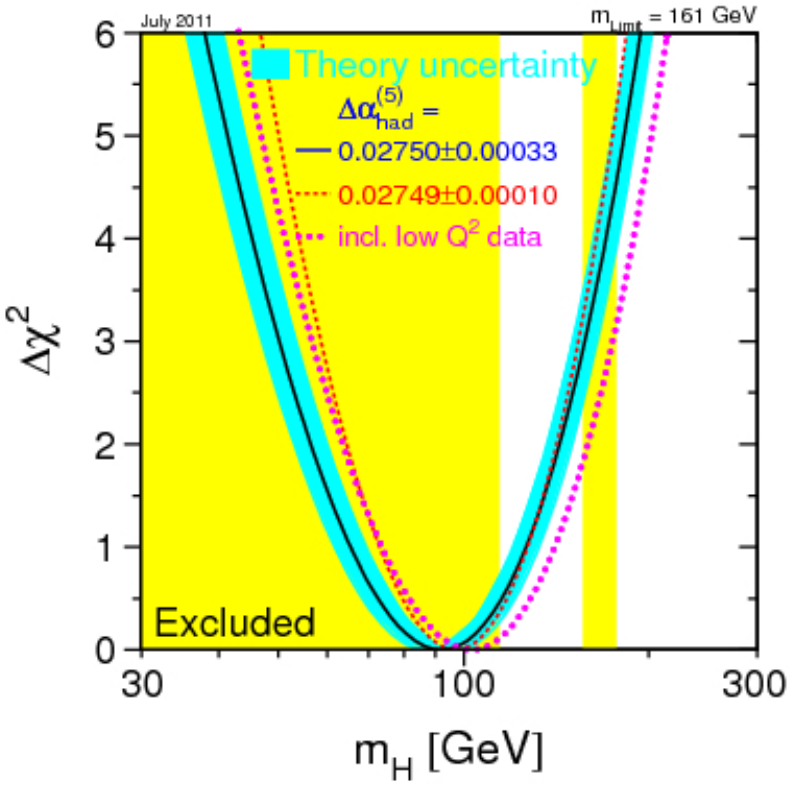

Figure 2.8: Electroweak fit for the mass of the Standard Model Higgs boson. The shaded region labeled "Excluded" corresponds to the lower limit set on the Higgs boson mass by direct searches. The fit alone gives an upper mass limit of 185 GeV/c^2 at 95% confidence level on the Higgs boson mass[10].

Higgs decay modes (mostly $b\bar{b}$ and $\tau^+\tau^-$) and the Z decay modes[1].

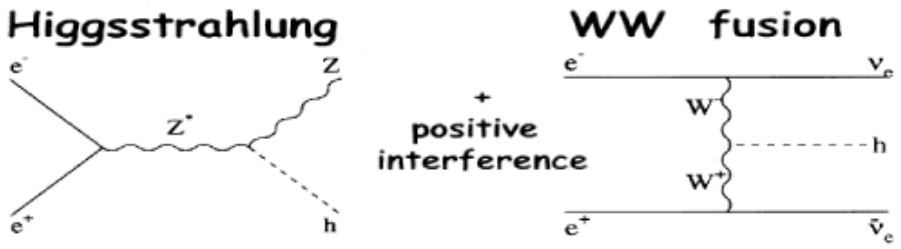

Figure 2.9: The Most important Higgs production modes at LEP

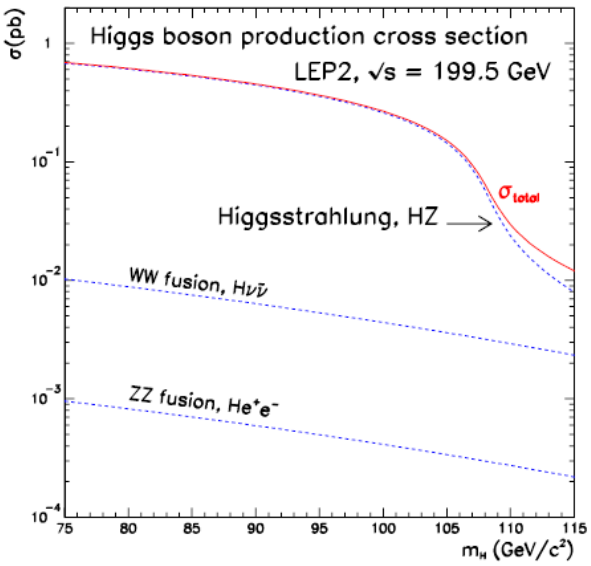

Figure 2.10: Cross section for the e^+e^- production of the Higgs boson of the Standard Model v.s. Higgs boson mass at 199.5 GeV for three different mechanism. The cross section for Higgsstrahlung drops rapidly with m_H once the kinematic threshold for HZ production at $m_H = \sqrt{s} - m_Z$ is crossed[12].

Combining the final results from the four LEP experiments, ALEPH, DELPHI, L3 and OPAL, a lower bound of 114.4 GeV/c^2 is set on the mass of the Standard Model Higgs boson at the 95% confidence level.

2.4.3 Higgs Search at the Tevatron

At Tevatron, which is a proton-antiproton collider, Higgs can be produced quit similarly to LHC. However due to the lower energy of collision, the cross section for Higgs production is lower. Figure 2.11 shows the cross section of Higgs production at the Tevatron with the energy of center of mass about 1.96 TeV.

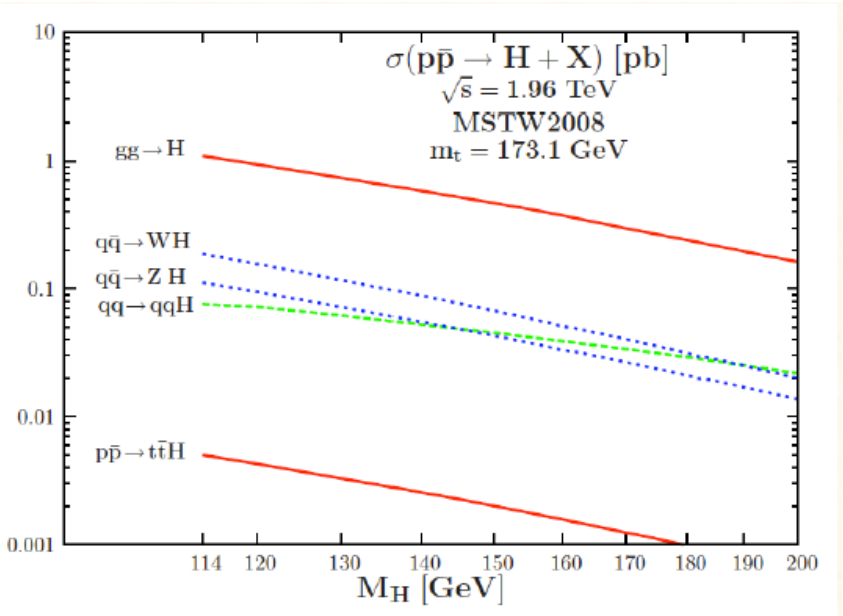

Figure 2.11: Cross section of different processes of Higgs production at the Tevatron with the energy of center of Mass about 1.96 TeV.

All production channels and the decay modes of the Higgs to which the Tevatron experiments are even marginally sensitive are included in the Higgs search. Then all search results are combined together and a further combination occurs between results of CDF and D0 (the two experiment of Tevatron) which leads producing combined search results with the highest possible sensitivity. Figure 2.12 shows the latest combination of CDF and D0 with more than $8fb^{-1}$ data.

This 'Brazil band plot' shows the possibility of the exclusion of the Higgs mass as a function of the mass. The Y axis is normalized such that the expected standard model production rate is at unity (such that what is effectively drawn is the rate in units of 'times the SM predicted rate'); this means that when the curve goes below 1.0 the corresponding mass values are excluded, at 95% confidence level.

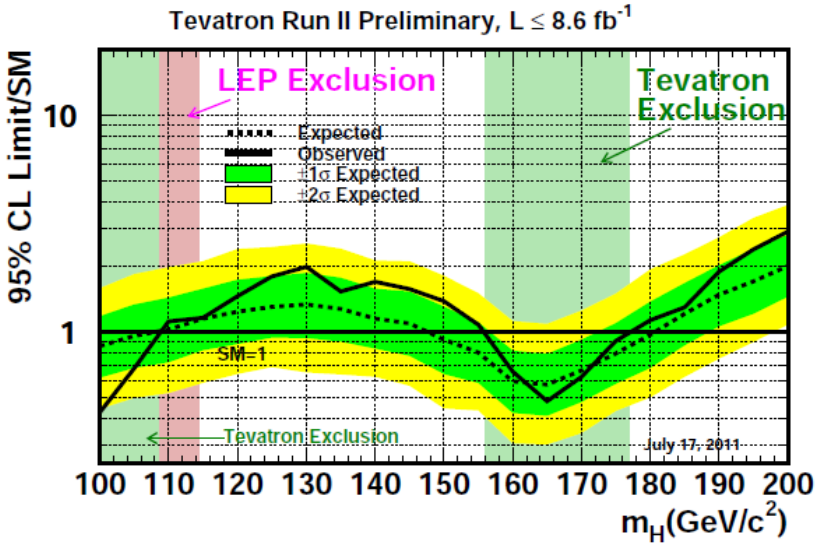

Figure 2.12: Tevatron's exclusion limit for SM Higgs boson

Based on the plot, Higgs with mass between $156 < m_H < 177\ GeV/c^2$ has been excluded at the 95% C.L. at Tevatron [2].

2.4.4 Higgs Search at the LHC

Going back to the title of this chapter, the most prominent goal of the LHC is seeking the standard model Higgs boson. Although so far both Tevatron and LEP could exclude some parts of the mass phase space, yet it seems that SM Higgs would be either excluded(in the full mass range) or discovered by CMS and ATLAS experiments at LHC. Figure 2.13 shows a MC study on the Higgs Boson with CMS detector with $5fb^{-1}$ at 7 TeV. From this plots one can conclude that within this amount of data CMS can excluded the CM Higgs boson in a whole possible mass range (from 114 to 600 GeV), provided that no excess of signal is observed.

Latest Results from ATLAS

Search for Higgs boson in ATLAS currently cover nine different decay modes. Figure 2.14 shows the expected and observed sensitivity of the search for the Higgs that arises from combining the results of searches in all decay modes studied to date. The black undulating dashed line shows ATLAS' predicted sensitivity to the Higgs

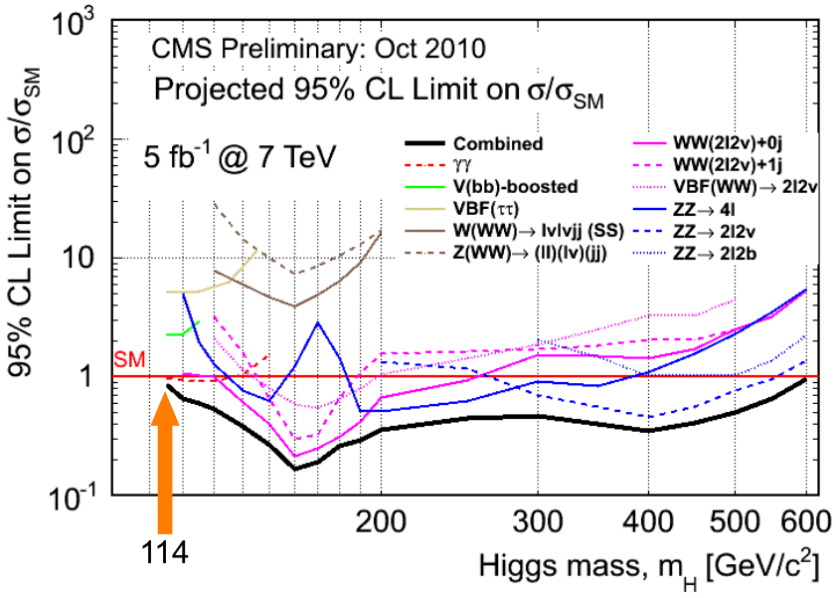

Figure 2.13: The CMS Collaboration projected sensitivity to excluding the higgs boson with $5fb^{-1}$ at $\sqrt{s} = 7$ TeV; the black line gives combined (total) sensitivity.

boson in the mass range 100-600 GeV, based on the simulations. The green and yellow bands correspond to the uncertainty in these predictions. The solid black line shows ATLAS' limit on Higgs production based on actual data collected. The plot shows that, with 1-2 fb^{-1} of data collected, ATLAS can exclude with 95% confidence the existence of a Higgs wherever the solid line dips below the horizontal dashed line at 1 [4].

from ATLAS results, many mass regions are excluded. In some other regions, there are small excesses above expectations. These regions where the Higgs discovery is possible are : 115-146 GeV, 232-256 GeV, 282-296 GeV plus any mass above 466 GeV [4].

Latest Results from CMS

CMS has combined 9 channels with the following final states:

- $H \to WW$ (0,1,2 jets)
- $H \to ZZ$, $(4l, 2l2\nu, 2l2q, 2l2\tau)$
- $H \to \gamma\gamma$

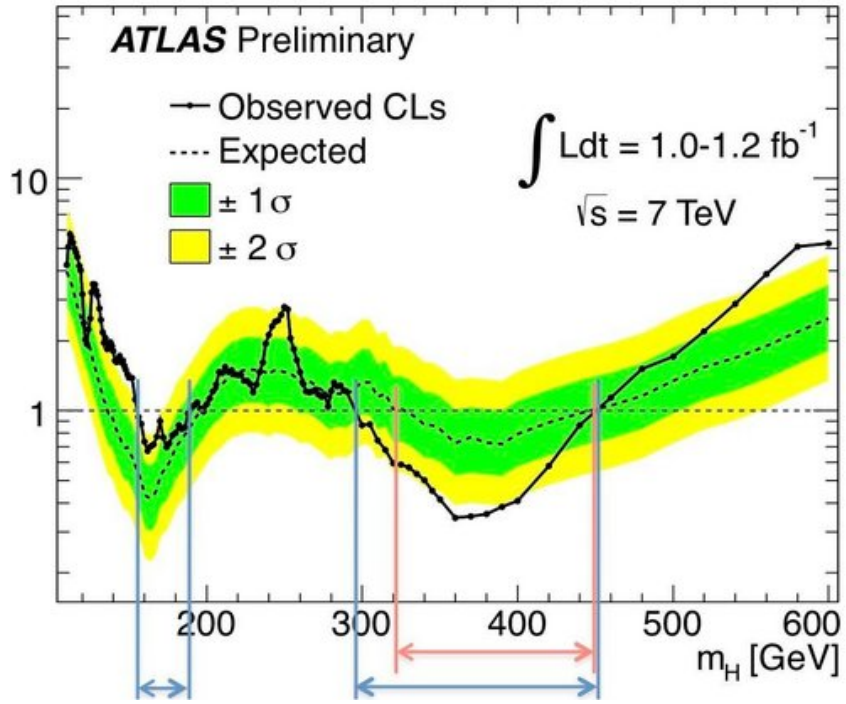

Figure 2.14: ATLAS's exclusion limit for SM Higgs boson

- $H \to \tau\tau$
- $VH \to b\bar{b}$

With most recent of the LHC data, 1.1–1.7 fb^{-1}, CMS could exclude the SM Higgs boson at 95% C.L. in three mass ranges 145-216, 226-288, and 310-400 GeV/c^2 [3].

A summary of the exclusion of the Higgs mass 95% C.L. made by all experiments[by the end of Summer 2011] is summarized in Table 2.4.

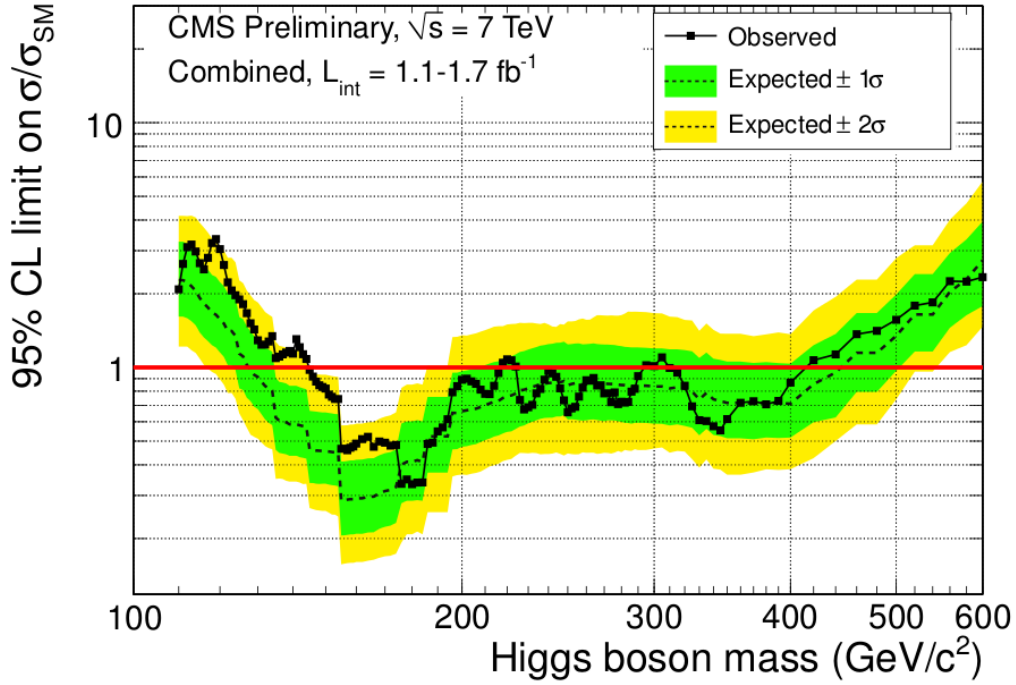

Figure 2.15: The combined 95% C.L. upper limits on the signal strength modifier $m = s/s_{SM}$, as a function of the SM Higgs boson mass in the range 110-600 GeV/c^2. The observed limits are shown by the solid symbols and the black line. The dashed line indicates the median expected limit on m for the background-only hypothesis, while the green/yellow bands indicate the ranges that are expected to contain 68%/95% of all observed limit excursions from the median.

Experiment	Mass Limit Exclusion [GeV] @ 95% C.L.
LEP	< 114.4
Tevatron (CDF + D0)	[156-177]
ATLAS	[155-190], [295-450]
CMS	[145-216], [226-288], [310-400]

Table 2.4: Higgs mass exclusion limit made by different experiment. Results based on the latest report until summer 2011

Chapter 3

Large Hadron Collider and CMS Detector

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) with 27 km in circumference, is the largest and highest energy particle accelerator in the world. It is located between 50 to 175 m under the surface of Switzerland and France. It has been built at the European Organization for Nuclear Research(CERN) laboratory with the aim to study physical interactions at the TeV energy scale.

In order to reach its final center of mass energy a chain of several accelerators is needed. Figure 3.1 shows a schematic view of the accelerator chain which provides protons for the LHC. The proton beams are obtained by stripping the electrons from hydrogen atom and are injected into the LINear particle ACcelerator LINAC2. LINAC2 accelerates the protons to 50 MeV, injecting them into the Proton Synchrotron Booster (PSB) where they are accelerated to 1.4 GeV. In the next step, the Proton Synchrotron (PS), the protons reach 26 GeV and obtain their final bunch structure. Before the bunches enter the LHC ring, they are accelerated to 450 GeV in the Super Proton Synchrotron (SPS).

In this ring two counter-rotating proton beams will be accelerated to a final center of mass energy of 7 TeV per proton(in the nominal case) resulting in a total energy of 362 MJ stored in each beam. 1232 dipole magnets with a magnetic field of 8 T are used to keep the protons within the ring. Furthermore 858 quadra-pole magnets are used as magnetic lenses for focusing and correcting the beams. At the designed value, each beam will consist of 2808 bunches with gaps of 25 ns between them. Each bunch will consist on average of 1.15×10^{11} protons and will be 7.48 cm long and 1 mm in diameter.

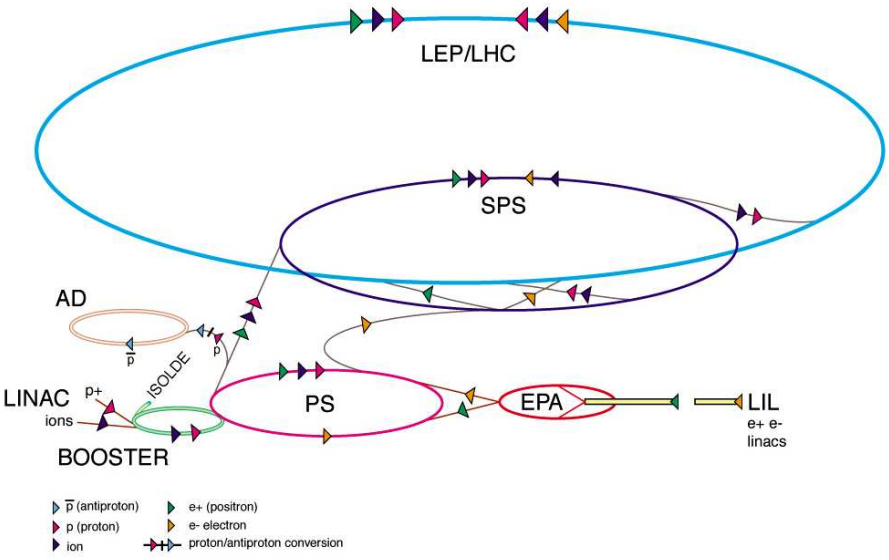

Figure 3.1: Schematic view of the accelerator chain for the LHC

At four collision sites the bunches will be squeezed down to 16 $\mu m \times$ 16 μm and focused at the interaction points where the experiments CMS, ATLAS, LHCb and ALICE are located. Their locations are indicated as yellow points in Figure 3.2. ATLAS and CMS are multi-purpose detectors, while LHCb and ALICE will study b-meson and heavy ion collisions, respectively.

The nominal energy and instantaneous luminosity for the LHC are 14 TeV and $10^{34} cm^{-2} s^{-1}$, respectively. Figure 3.3 shows the maximum Instantaneous luminosity and amount of integrated luminosity delivered to/recorded by CMS experiment in 7 TeV collisions, so far.

For a Gaussian beam distribution, as it will be at the LHC, the machine luminosity is defined as:

$$\mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi \varepsilon_n \beta^*} F \tag{3.1}$$

where γ is the relativistic gamma (Lorentz) factor, f is the revolution frequency, k_B the number of bunches per beam, N_p the number of protons/bunch, ε_n the normalized transverse beam emittance, β^* is the β function at the collision point and F is the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP).

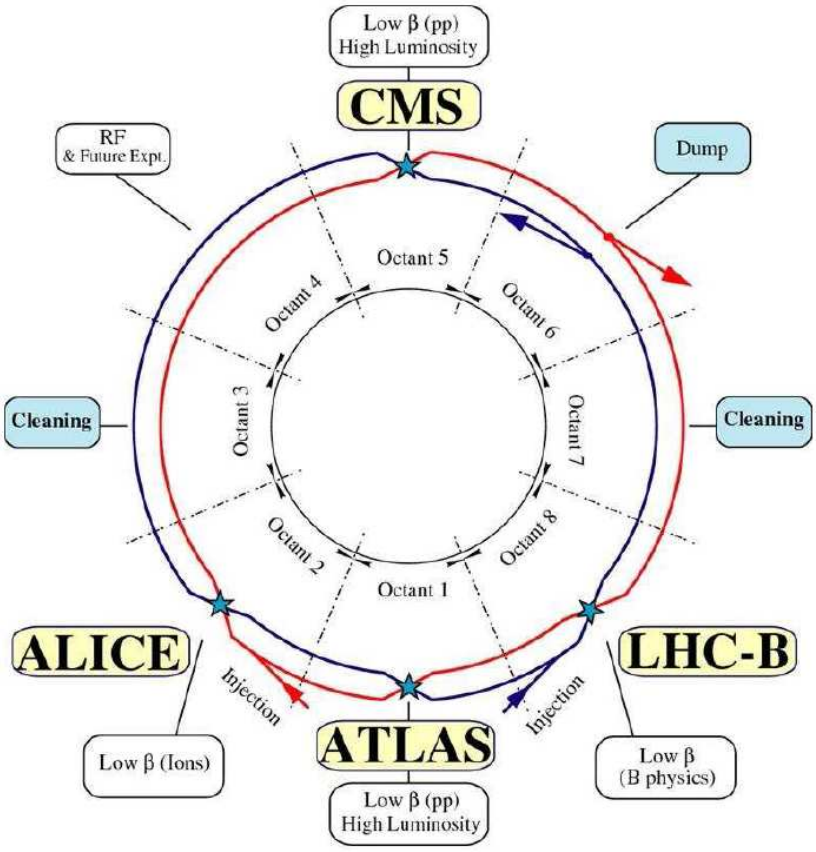

Figure 3.2: Schematic layout of the LHC complex and the experiments location

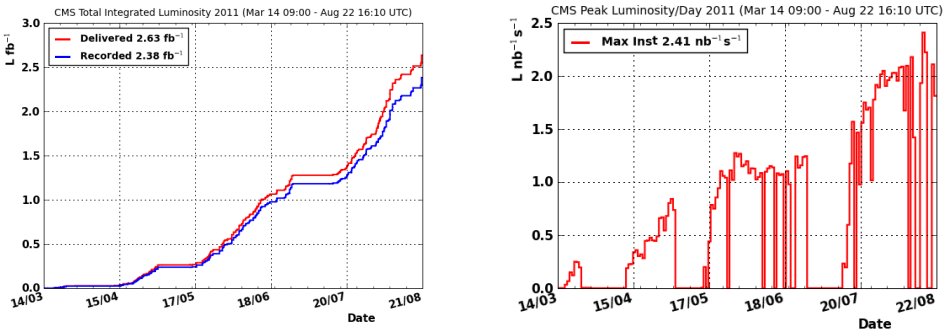

Figure 3.3: Integrated luminosity versus time delivered to (red), and recorded by CMS (blue) (left) and maximum Instantaneous luminosity per day delivered to CMS during stable beams at 7 TeV center-of-mass energy(right).

Table 3.1 shows the nominal proton beam parameters.

		Nominal
Energy(per beam)	TeV	7
Luminosity	$cm^{-2}s^{-1}$	10^{34}
Bunch spacing	[ns]	24.95
Number of bunches		2808
N_b intensity per bunch	[p/b]	1.15×10^{11}
Beam current	[A]	0.58
ε_n (transverse emittance, rms, normalised)	[μ m]	3.75
Bunch length, total (4σ)	[ns]	1.0

Table 3.1: LHC nominal proton beam parameters

3.2 CMS Detector

The Compact Muon Solenoid (CMS) [13, 14] is a multi-purpose detector for precise tests of the Standard Model(SM) and physics beyond SM. A schematic of the detector layout is shown in Figure 3.4. CMS has a length of 21.5 m, a diameter of 15 m and a weight of 12500 t. As shown in Figure 3.5 it consists several substructures in order to identify and measure the quantities of different particle types produced in the collisions. It is composed of the tracking system which builds the innermost part and measures the momentum of charged particles. Surrounding it is a scintillating crystal electromagnetic calorimeter, which is itself surrounded with a sampling calorimeter for hadrons which are designed for energy measurement. Both tracking system and calorimeters are confined by the solenoid which generates a powerful magnetic field of 3.8 T. The outer most part of the detector, the muon system, measures the momentum of muons and identifies them.

3.2.1 Coordination System

At the CMS detector, the interaction point is chosen as the center of the coordinate system, the z-axis along the tunnel, the y-axis is slightly toward vertical and the x-axis pointing to the center of the tunnel. It is described by the radial distance to the beam-line r, the polar angle θ and the azimuthal angle ϕ. The polar angle is measured with respect to the z-axis; $\theta = 0$ corresponds to the positive z-direction and $\theta = \pi$ to the negative z-direction. The azimuthal angle is measured from

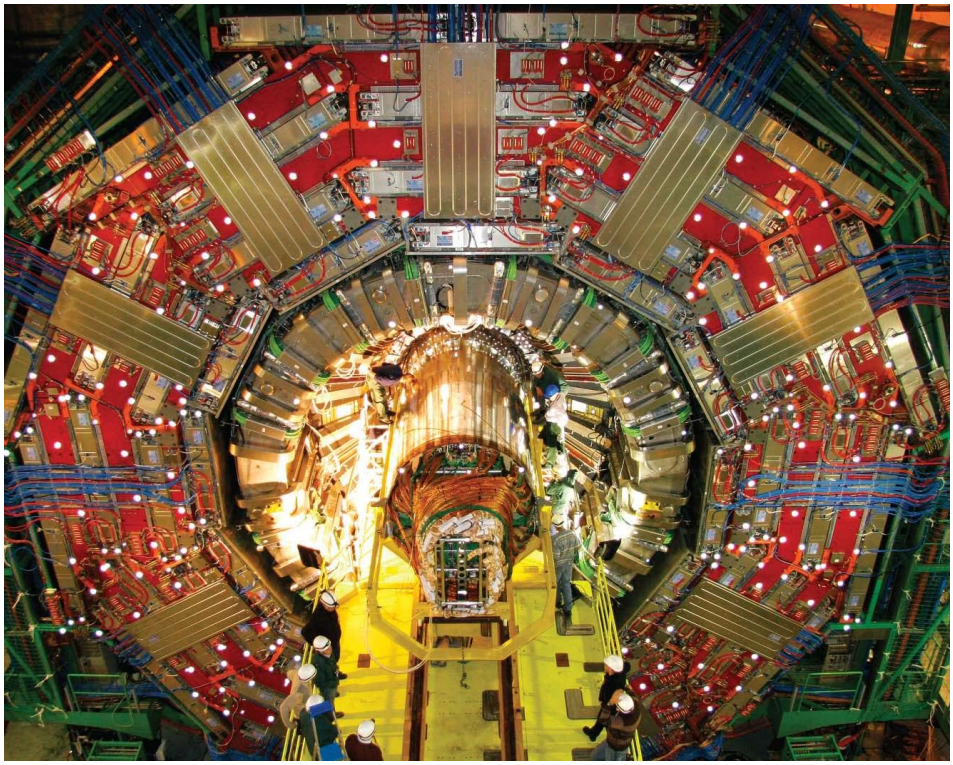

Figure 3.4: An over view of the CMS

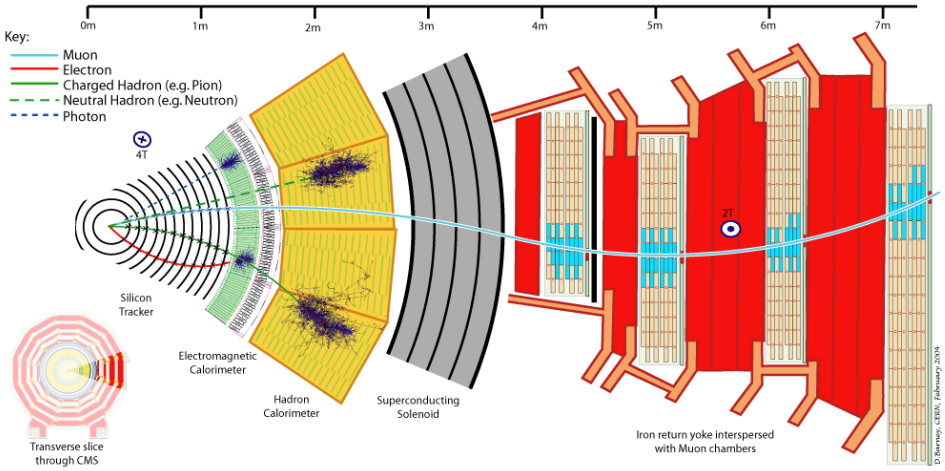

Figure 3.5: Several layer of the CMS Detector

the x-axis in the xy-plane. $\phi = 0$ points to the $+x$-direction and $\phi = \pi/2$ to the $+y$-direction. The pseudo-rapidity η is defined as:

$$\eta = -\ln\tan(\theta/2) \tag{3.2}$$

Zero value of η corresponds to a moving particles perpendicular to to the z and $\eta = \pm\infty$ is in the $\pm z$ direction. The region of $|\eta| < 1.4$ is referred to as the central region. Both $\Delta\eta$ and $\Delta\phi$ of two particles are independent of Lorentz boosts, therefore the distance between two particle(ΔR) can be measured in a third Lorentz invariant variable:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.3}$$

3.2.2 Magnet System

A strong magnetic field is required to bend the high energy particles enough to obtain charge identification and momentum measurement. Equation(3.4) shows the relation between momentum p, magnetic field B and the radius of the curvature R for single charged particles.

$$R = \frac{p}{0.3B} \tag{3.4}$$

where R is measured in m, p is measured in GeV and B is measured in T. The necessary magnetic field is provided by the solenoid magnet. It is about 13 m long and has an inner diameter of 5.9 m. The superconducting is cooled by liquid helium. It was designed to produce a magnetic field of up to 4 T in the inner region. In order to maximize its lifetime, the magnet will run at 3.8, though which is still 100,000 times stronger than the Earth's magnetic field. The current required for this strong magnetic field is 19 500 A resulting in a stored energy of 2.7 GJ.

3.2.3 Tracking System

Since the particle flux within the detector decreases as $1/r^2$, the region closest to the interaction point has the highest requirements on the detector. With a high occupancy, reconstruction of the tracks is impossible, because too many possibilities exist to combine hits to a track. For this reason a very strong and powerful tracker

has been designed for the CMS[15]. It has a length of 5.8 m and a diameter of 2.5 m. It is immersed in a coaxial magnetic field of 3.8 T provided by the CMS solenoid. A schematic drawing of the CMS tracker is shown in Figure 3.6.

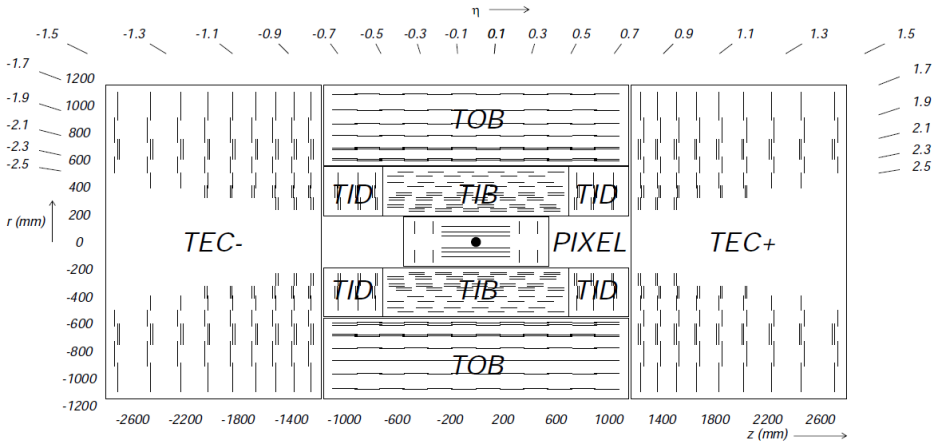

Figure 3.6: A schematic r-z view of the CMS tracking system. All the tracking subsystems are denoted and their coverage in z and η is also shown. Each line represents a detector module.

CMS Tracker consist pixel tracker and silicon tracker which are as follows:

Pixel Detector

The pixel detector, though about the size of a shoe box, contains 65 million pixels, allowing it to track the paths of particles emerging from the collision with extreme accuracy. It is also the closest detector to the beam pipe, with cylindrical layers at 4 cm, 7 cm and 11 cm and disks at either end, and so will be vital in reconstructing the tracks of very short-lived particles. The pixel sensors are 100 μm (in $r\phi$) $\times$ 150 μm^2 (in z) in size. This leads to a high granularity and therefore decreases the occupancy below 1% [16]. The pixel detector delivers high precision space-points with resolutions of 15-20 μm .

Silicon Detector

After the pixels and on their way out of the tracker, particles pass through ten layers of silicon strip detectors, reaching out to a radius of 130 cm. The tracker

silicon strip detector has a total of 9.3 million strips and 198 m^2 of active silicon area consists of four inner barrel (TIB) layers assembled in shells with two inner endcaps (TID), each composed of three small discs. The outer barrel (TOB) consists of six concentric layers. Finally two endcaps (TEC) close off the tracker. Each has silicon modules designed differently for its place within the detector. Resolution for different parts of the silicon detector are different which depend on their position with respect to the interaction point and are within the range of 23-35 μm for TIB/TID, 35-53 μm for TOB and 230-530 μm for TEC.

The tracking system provides a transverse momentum resolution for charged particles, which is given by:

$$\left(\frac{\Delta P}{P}\right)^2 = (0.15.p_T)^2 + (0.005)^2 \quad \text{in}[TeV^2] \tag{3.5}$$

It is expected to improve the resolution to 10 % after short-term alignment and down to 0.025 % after the long-term alignment [10].

3.2.4 Electromagnetic Calorimeter

The next detector layer around the tracking system is the electromagnetic calorimeter (ECAL)[17] which has a total coverage of $|\eta| < 3$. This subdetector is especially designed to measure the energy of electron and photons.

ECAL comprises three different components which schematically is shown in Figure 3.7. They are:

ECAL Barrel (EB)

ECAL has a cylindrical barrel which consist of 61200 lead tungstate ($PbWO_4$) scintillating crystals. Lead tungstate crystal has such a high density($8.28\frac{g}{cm^3}$) which produce scintillation light in a small, fast and well-defined photon showers. Consequently measurement from calorimeter would be very precise. EB coverage is to $\eta = 1.479$. Furthermore it allows a compact detector design.

ECAL Endcap (EE)

The two flat ECAL Endcap of the CMS, each consisting 7324 crystal of the same

type as the EB crystals, seal off the barrel at either end and complemented the coverage of the EB to the $\eta = 3.0$.

ECAL pre-Shower (ES)

ECAL pre-showers in the CMS detector are located in front of the (EE) and are made of silicon-strip. They covers $1.653 < \eta < 2.6$. This part of the ECAL is particularly used to enhance the spatial resolution of those particles in the above η range by lowering the required detector depth. That would lead to distinguish between energetic photons (which often are signs of exciting physics) and the less interesting close pair of low-energy photons.

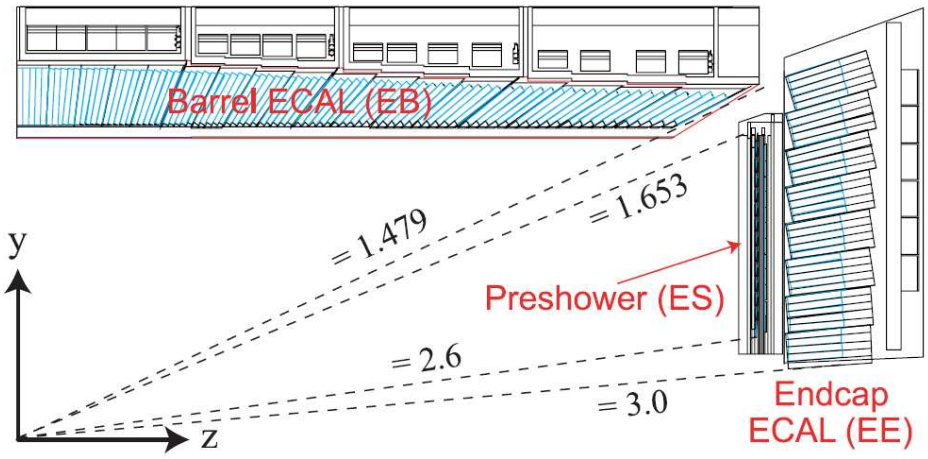

Figure 3.7: Transverse section through the ECAL, showing geometrical configuration

When electrons, positrons or photons hit the crystals they create electromagnetic showers. The number of produced particles in this shower is proportional to the energy of the incoming particle. Energy deposits within the scintillating crystals create a flash of light which is detected by silicon avalanche photo-diodes in the EB and vacuum photo-triodes in the EE. Electrons and unconverted photons deposit 94% of their energy in an array of 3×3 crystals(which is called 'cluster') and 97% in an array of 5×5 crystals (which is called 'super–cluster'). The energy of the electron or photon can be obtained by summing the energy measured around the hit positions while the hit position itself is defined as the center of the energy deposit. In order to retrieve this center each position in a crystal is weighted by the fraction

of the total energy deposit:

$$x = (\frac{\Sigma x_i.W_i}{\Sigma W_i}), \qquad W_i = W_0 + \log(\frac{E_i}{\Sigma E_j}) \tag{3.6}$$

where x is the hit position, x_i the centers of the crystals and W_i the weights. In case of photon conversion or bremsstrahlung radiation in the tracker (which some times happens due to the interaction of photon and electron with the tracker material) the energy reaching the calorimeter is spread in ϕ direction because of the magnetic field. In this case other algorithms are necessary to calculate the hit position of the initial electron or photon.

The energy resolution for the calorimeter $\frac{\sigma_E}{E}$ can be written as [18]:

$$\frac{\sigma(E)}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{3.7}$$

‘

where the three contributions correspond to the stochastic, noise, and constant terms, respectively.

3.2.5 Hadronic Calorimeter

The Hadron Calorimeter (HCAL)[19] which consists of layers of dense material (brass or steel), is supposed to measures the energy of ‘hadrons’. It is a ‘sampling’ calorimeter, which means that measuring the position and energy of the particles by exploiting the alternating layers of ‘absorber’ and fluorescent scintillator materials that produce a rapid light pulse when the particle passes through. This light is collected and send into the readout boxes. The energy of a particle is then measures as the sum of the amount of light over the layer of tile in the depth (which is also called as tower)

Figure 3.8 is an elevation view of the CMS detector showing the HCAL components with lines of constant pseudo rapidity overlayed. They are as follows:

HCAL Barrel (HB)

The HCAL barrel covers the region $|\eta| < 1.4$ and consists of about 2300 towers, resulting in a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Each tower consists of

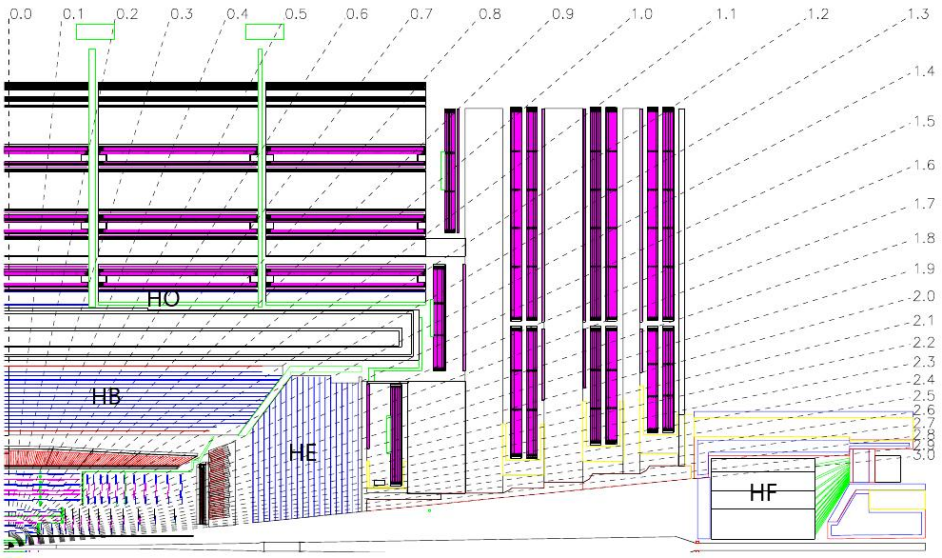

Figure 3.8: Longitudinal view of one quarter of the detector in the r-π plane, showing the positions of the HCAL parts: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF)

alternating layers of non-magnetic brass absorber and plastic scintillator material. The absorber material has to be non-magnetic to not distort the magnetic field.

HCAL Outer barrel (HO)

HCAL Outer barrel are located just behind the HB and outside of the solenoid and act as 'tail catchers' effectively increasing the thickness of the calorimeter in the central pseudo rapidity region and ensure that no energy leaks out the back of the HB undetected. HO has the same granularity and η range as HB.

HCAL Endcap (HE)

The Hadronic endcap calorimeters (HE) cover the pseudo rapidity range $1.3 < |\eta| < 3$ and are located in the end parts of the CMS detector and thus are allowed to contain magnetic material. Here iron is used as the absorber material. The granularity ramps from $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ at $\eta < 1.6$ up to $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ at $1.6 < \eta < 3$.

HCAL Forward (HF)

The hadronic forward calorimeter (HF) extends the coverage $3.0 < |\eta| < 5.0$, which is not covered by any other detector part. It is a Cerenkov light sub-detector that made up of quartz fibers embedded within a long steel absorber.

Together, the components of the HCAL cover a range of $|\eta| < 5$ and only a small range of $< 0.7°$ around the beam direction remains uncovered. This almost total enclosure of the interaction point is very important for the CMS detector's ability of determining possible missing transverse energy, caused by neutrinos or a signature of a new physics.

The expected energy resolution of the full HCAL (including the HO) is

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{120\%}{\sqrt{E}}\right)^2 + (5\%)^2 \tag{3.8}$$

Combining the energy measurement of ECAL and HCAL, the CMS calorimeter is expected to achieve a resolution of

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{100\%}{E}\right)^2 + (4.5\%)^2 \tag{3.9}$$

for energies of 30 GeV $< E <$ 1 TeV.

3.2.6 Muon System

Muon detection and reconstruction is a powerful tool for the discovery of new physics and precision measurements of standard model physics. This requires the robust detection of muons over the full acceptance of the CMS detector and over the very high background rate expected at the LHC. Since muons' life time is large and they can penetrate several meters of iron without interacting, unlike most particles they are not stopped by any of CMS's calorimeters. Therefore, muon sub-detectors are placed at the very edge of the detector[20].

detection of the muons are handled via three muon sub-detectors(Layout of one quarter of the CMS muon systems is shown in Figure 3.9) arranged in a cylindrical barrel section and 2 planar endcap regions, with a total area of 25,000 m^2 and are as follows:

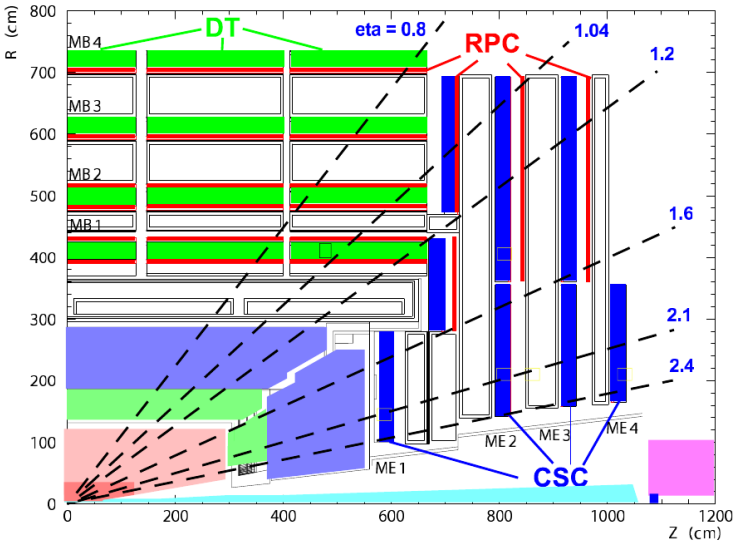

Figure 3.9: Layout of one quarter of the CMS muon system for initial low luminosity running. The RPC system is limited to $|\eta| < 1.6$ in the endcap, and for the CSC system only the inner ring of the ME4 chambers have been deployed.

Drift Tubes (DT)

The drift tube system aims to measures the muon positions in the barrel part of the detector. There are 250 tube with 4 cm wide and each contains a stretched wire within a gas volume. When a muon or any charged particle passes through DT, they ionize the gas inside the cell and the electrons and ions will accelerate to the anodes and cathodes, respectively. The drift time, which is a function of the random diffusing motion and the drift velocity due to an electric field, provides the hit spatial resolution for the particle. Figure 3.10 illustrate a cross section of the CMS drift tube with the anode wire.

Cathode Strip Chamber (CSC)

Cathode strip chambers are used in the endcap disks where the magnetic field is uneven and particle rates are high.

CSCs consist of arrays of positively-charged 'anode' wires crossed with negatively-charged copper 'cathode' strips within a gas volume. When muons pass through, they knock electrons off the gas atoms, and electron accelerate to the anode wires and creates an avalanche of electrons. The same procedure happens for ion

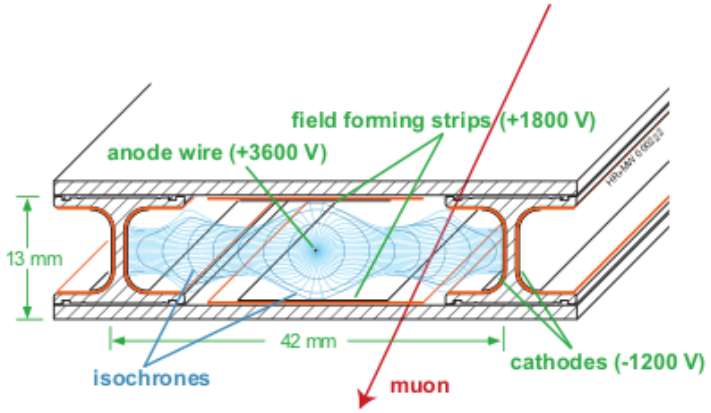

Figure 3.10: Cross section of a CMS drift tube with the anode wire, which is spanned in the middle of the tube, and field lines of drift field.

toward the copper cathode. Because the strips and the wires are perpendicular, for each passing particles two position coordination is dedicated.

Resistive Plate Chambers (RPC)

Resistive Plate Chambers are embedded in both barrel (six layers) and endcap (three layer). They consist of two thin gas-filled gaps between Bakelite plates. Two of these Bakelite chambers are combined via aluminum strips, which are used for the readout. High voltage is attached to the Bakelite plates, leading to an avalanche effect of free electrons created by incoming ionizing radiation. Due to the small width of the gas gap, the avalanche reaches the Bakelite plates and is read out within $\sim$ 5 ns which is particularly highly efficient for the L1 Trigger.

All three muon systems(DT, CSC and RPC), take part in the trigger. The barrel DT chambers provide local trigger information in the form of track segments in the ϕ -projection and hit patterns in the η -projection. The endcap CSCs deliver 3-dimensional track segments. All chamber types also identify the bunch crossing from which an event originated. The Regional Muon Trigger consists of the DT and CSC Track Finders, which join segments to complete tracks and assign physical parameters to them. In addition, the RPC trigger chambers, which have excellent timing resolution, deliver their own track candidates based on regional hit patterns. The Global Muon Trigger then combines the information from the three

sub-detectors, achieving an improved momentum resolution and efficiency compared to the stand-alone systems.

3.2.7 Trigger and Data Acquisition

For every bunch crossing at a nominal luminosity of $10^{34} cm^{-2} s^{-1}$, 25 collisions are expected to occur on average, producing soft scattering events with low transverse momentum particles. Only in rare cases hard scattering processes take place. In addition it can occur that the same proton takes part in two collisions (pile-up). Every event produces 1-2 MB of raw data. Taking a bunch crossing rate of 40 MHz leads to a total data production of 40 TB/s. Since only the hard collisions are of interest the trigger system was developed to reduce the amount of data with minor effect on interesting events.

The CMS trigger system consists of two independent levels: Level 1 (L1) and the High Level Trigger (HLT)[21].

L1

In the nominal case, L1 trigger needs to process events every 25 ns. The total time allowed for making a trigger decision is limited to 3.2 μs, limiting therefore the time to reconstruct trigger primitive objects to about 1 μs. Therefore, L1 should be a pure hardware trigger which only exploits the calorimeter and muon system information on its decision. The tracker system is not considered since the reconstruction of tracks would exceed the time limit of L1 decision. The L1 subsystems are distributed on the respective detector parts where trigger primitives with reduced granularity and resolution are generated as candidates for the global L1. The global L1 is located 90 m outside the detector. The latency between the bunch crossing and the L1 accept signal is 3.2 μs. The signal information is buffered for this period of time. The L1 reduces the event rate from 40 MHz to 100 kHz, which corresponds to a data flow of 100 GB/s. The accepted events are passed to the HLT.

HLT

The HLT is a software trigger and thus very flexible. It can analyze the entire high-resolution data with highly complex and customized algorithms if necessary. The HLT decision is based on the output of all subsystems including the tracker, but only regions of interest in the tracker, where a L1 object was found before, are reconstructed. The full reconstructions take place after the HLT at where calibration data can be used for re-reconstruction. Both triggers reduce the initial input of 40 MHz to 100 Hz.

Figure 3.11 shows an illustration of the L1 and HLT process in the CMS detector

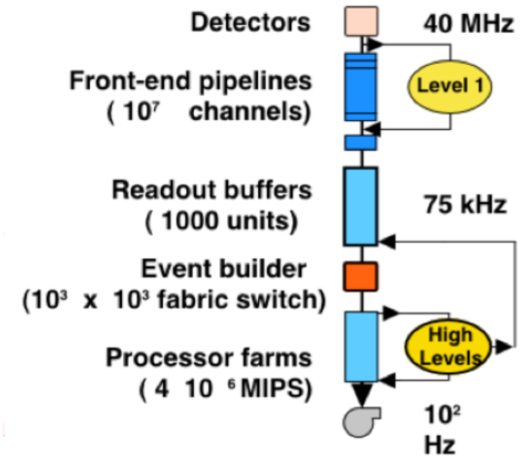

Figure 3.11: illustration of the L1 and HLT process in the CMS detector

Chapter 4

Generation, Simulation and Reconstruction

4.1 Monte Carlo Generation

4.1.1 Event Generation

To understand and interpret data, sets of Monte Carlo (MC) models are required. Monte Carlo simulations should describe the real data as realistic as possible based on theoretical models. The underlying theoretical framework for the description of interactions between particles is quantum mechanics, where many properties of the particles, like their momenta, are stochastically distributed. Thus, also the output of a MC event generator should have some fluctuations, which turn to be the basic distinguishable feature of every MC event generator. Random numbers are thrown to obtain candidate events, while all significant variables with the predicted probability distributions are taken into account. But as pp collisions have a complex nature, the event generator subdivides the process into several components as below[21] :

Parton Distribution

When extended particles like protons collide, the kind of interaction and consequently the simulation parameters depend on the momentum fraction of the partons.

Hard Subprocess

With the term 'hard subprocess' one refers to the description of the events based on the perturbation theory. This can be applied in the case of collisions because the generation of heavy particles (short-lived resonances) from interactions and the

subsequent decay to partons happen in a shorter time than the beginning of the parton shower.

Parton Shower

Higher order QCD effects are taken into account in the 'parton shower' approach. For example, MC generators, based on parton branching or showering approach, describe the evolution of gluon initial-state or final-state radiation from partons until a certain cutoff is reached. Due to this showering, jets of quarks and gluons are produced in the direction of the primary parton and can be easily detected.

Hadronization

Colored quarks and gluons, produced in the showers, cannot exist as free particles due to the 'color confinement'. This leads to the formation of quark-antiquark pairs if two colored particles separate by more than a certain distance. Because of this, partons in jets have to be grouped into colorless objects, i.e. hadrons which can be detected. This phenomenon, the 'hadronization', proceeds at the scale of large strong coupling constant, i.e. at the scale of low momentum transfer.

4.1.2 Generators

In order to generate a full and complete set of events, several Monte Carlo generators need to be adopted. Some of them are only describing the LO interactions while the others using Matrix Elements, can take care of NLO correction.
The mostly used generator in the CMS is Pythia[22]. It is a program for the generation of high-energy physics events, i.e. for the description of collisions at high energies between elementary particles such as e^+, e^-, p and $\bar{p}$ in various combinations. It contains theory and models for a number of physics aspects, including hard and soft interactions, parton distributions, initial–state and final–state parton showers, multiple interactions, fragmentation and decay.

The Pythia picture of the event and the physics which lies below is schematized as follows:

Initially two beam particles which are characterized by their partonic substructure come into each other and start showering. One parton from each beam starts

the shower, i.e. the sequence of branchings (for example $q \rightarrow qg$). One parton from each of the shower enters the hard process and the out-coming partons are produced. The short-lived particles like Z and W are produced which immediately decay to other particles. The out-coming partons undergo a branching process to build up final- state showers.

In the Pythia program, a large variety of hard processes are modelized and available such as Hard QCD processes(like $qg \rightarrow qg$), Soft QCD processes (like diffractive and elastic scattering), minimum-bias events and Heavy-flavor production.

Another important generator is MadGraph [23] which has the ability to identify the relevant sub-processes of a given process. It generates both the amplitudes and the mappings needed for an efficient integration over the phase space.

MadGraph is specially preferable with respect to Pythia in the events with many jets and therefore many particles. It can measure the kinematics of the jets more precisely.

Other used generators at CMS are POWHEG [24](which is used for NLO correction), Herwig (which is a Monte Carlo package for simulating Hadron Emission Reactions With Interfering Gluons) and MC@NLO (which is a package for combining a Monte Carlo event generator with Next-to-Leading-Order calculations of rates for QCD processes)

4.2 Detector Simulation

After an event is generated and a full picture of events is ready, the event with all particles then undergo a detailed simulation of the detector which is called detector simulation. This simulation is based on the Geant4 [25]. The detector simulation step takes as input the generated particles, propagates them through the matter and models physics processes that happen during this passage. The resulting data are stored in form of 'simulated hits'. Subsequently, there is the digitization step to model the response of the detector readout electronics; signal collection and electronic effects are computed, noise is added and pile-up events are superimposed. The following step in the analysis of data is called reconstruction and can be applied independently from the origin of input data (simulation or real data).

4.3 Object Reconstruction

In the reconstruction phase, the high-level objects, such as jets, tracks, vertices, leptons are reconstructed and the relative collections are created.

The reconstructions algorithms collect and link informations from different subdetectors taking into account also differences between the subdetectors themselves, such as for example calibrations, noise thresholds, readout geometries. To do this, before the application of any high-level algorithm, a preliminary step is added to create the so-called 'RecHits'. RecHits are objects containing informations such as energy releases, 3D-position and collected charge of the particle which is interacting with the subdetector matter. Different RecHits collections are created for each different part of the detector but the type of the associated information is the same. The main advantage of RecHits objects is that different high level reconstruction algorithms can share the same code; It is just the input RecHits collection that should be changed accordingly to the desired output high-level object type.

In the following, reconstruction algorithms of some of the high-level objects are prescribed.

4.3.1 Tracks

The reconstructed tracks of charged particles are one of the most fundamental objects in the reconstruction of pp collisions, since the reconstruction of leptons, charged hadrons, jets, and etc would highly depend on track reconstruction [26, 27]. In addition, Particle Flow algorithm, which is an algorithm to reconstruct Jets, MET and Taus with high resolution are based on a perfect Track reconstruction.

The default track reconstruction at CMS is performed by the combinatorial track finder (CTF). Triplets of hits in the tracker or pairs of hits with an additional constraint from the beamspot or a vertex are used as initial estimates, or seeds of tracks. The seeds are then propagated outward in a search for compatible hits. As hits are found, they are added to the seed trajectory and the search continues until either the limit of the tracker is reached or no more compatible hits can be found. In the final step, this collection of hits is fit to obtain the best estimate of the track parameters.

Tracking system of CMS is able to to reconstruct tracks with p_T as low as 0.7 GeV/c in the central region ($\eta \sim 0$) to the tracks with p_T near to 1 TeV/c.

4.3.2 Vertices

Two different concepts are usually discussed under vertices reconstruction: primary vertex and beamspot[26].

Primary Vertex

The Primary Vertex (PV) indicates the real point where particles collide, which can be different from the nominal interaction point (IP). The actual CMS algorithm for the PV finding uses fully reconstructed particles. First there is a track preselection, based on the distance of closest approach to the beam which is quantified considering the transverse impact parameter significance, i.e. the ratio of the transverse impact parameter $d0$ divided by its uncertainty, which is required to be < 3. Then tracks are further selected requiring their p_T to be > 1.5 GeV/c. Then, clusters of the selected tracks are formed, based on the z-coordinate of their point of closest approach with respect to the beam line. For each of these clusters, a fit of the PV candidate is performed, discarding incompatible tracks. Finally, from the list of the candidates, those with poor quality fits (with χ^2 probability $< 1\%$) are excluded.

Secondary Vertex finding

Finding a secondary vertex is a crucial issues especially for channels including b-quark such as light Higgs in decay of $b\bar{b}$. A good discriminator is the distance of the vertex to the beam line but in general a filter is applied to the output of Vertex Finder to select only SV based on the following cuts: the distance in the transverse plane L_T between the PV and the SV should be between 100μ and 2 cm.

Beamspot

The beamspot represents the profile of the luminous region where the LHC beams collide at CMS. The beamspot is determined in an average over many events, in contrast to the event-by-event primary vertex which gives the precise position of a single collision. A precise measurement of the position and slope of the beamspot

is an important component of the event reconstruction. The beam position can be used, especially in the High Level Trigger, as a precise estimate of the primary interaction point prior to the reconstruction of the primary vertex and even as the primary interaction point in low multiplicity data.

4.3.3 Muons

In the standard CMS reconstruction for pp collisions, tracks are first reconstructed independently in the silicon tracker (tracker track) and in the muon spectrometer (standalone-muon track). Based on these, two reconstruction approaches for muons are used[28]:

1. **Global Muon reconstruction (outside-in)**
   Starting from a standalone muon in the muon system, a matching tracker track is found and a global-muon track is fitted combining hits from the tracker track and standalone-muon track. At large transverse momenta (p_T 200 GeV/c), the global-muon fit can improve the momentum resolution compared to the tracker-only fit.

2. **Tracker Muon reconstruction (inside-out)**
   In this approach, all tracker tracks with $p_T > 0.5$ GeV/c and $p > 2.5$ GeV/c are considered as possible muon candidates and are extrapolated to the muon system, taking into account the expected energy loss and the uncertainty due to multiple scattering. If at least one muon segment (i.e. a short track stub made of DT or CSC hits) matches the extrapolated track, the corresponding tracker track qualifies as a tracker-muon track. At low momentum (roughly p $<$ 5 GeV/c) this approach is more efficient than the global muon reconstruction, since it requires only a single muon segment in the muon system, whereas global muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station.

   The majority of muons from collisions (with sufficient momentum) are reconstructed either as a Global Muon or a Tracker Muon, or very often as both.

For a small fraction of the cases where both approaches fail, the track reconstructed at the muon chamber is called 'stand-alone muon'.

Figure 4.1 shows some identification variables for muons with $p_T > 20$ GeV/c.

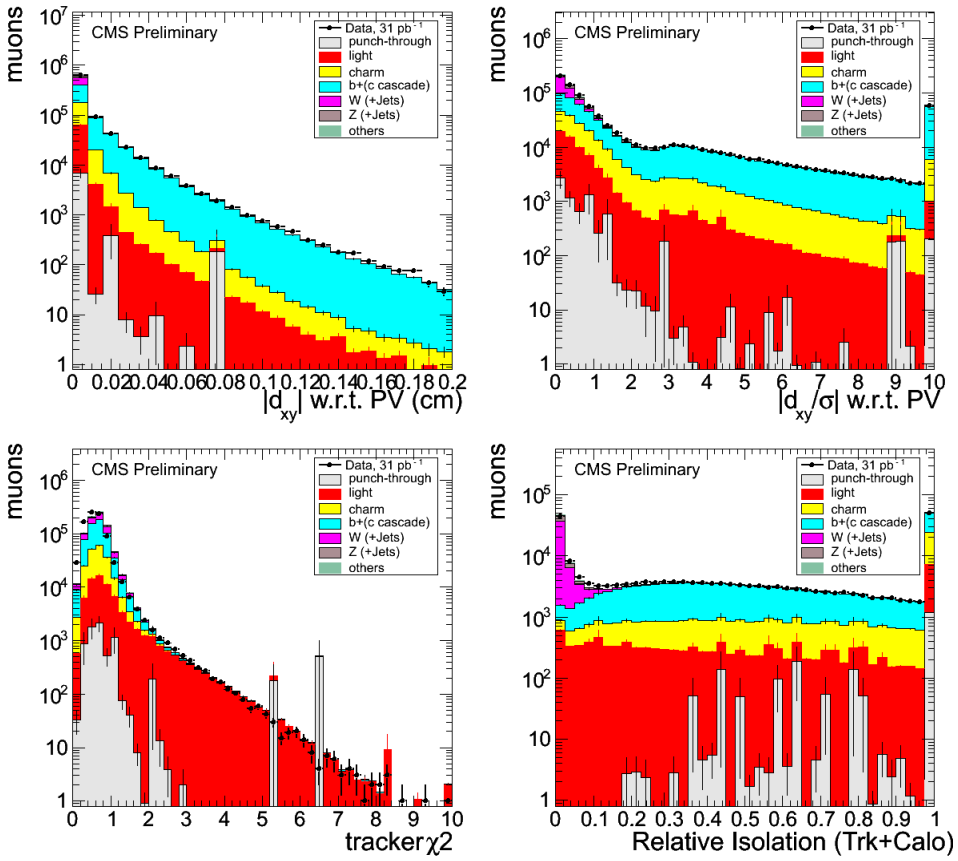

Figure 4.1: Identification variables for Tight Muons with $pt > 20$ GeV/c: (upper left) transverse impact parameter; (upper right) significance of the transverse impact parameter; (lower left) normalized χ^2 of the fit of the track in the silicon tracker; (lower right) relative combined isolation (tracker+calorimeters), with a cone size $\Delta R = 0.3$, for events with one good primary vertex. The error bars indicate the statistical uncertainties [28]

4.3.4 Electrons

Two different and complementary algorithms are used to reconstruct electrons which can cover the reconstruction of both high p_T and low p_T electrons[29].

1. 'Tracker driven' which is based on tracker seeding and is useful for low energy

electrons specially those inside jets

2. 'ECAL driven' which starts by reconstruction of ECAL 'supercluster' and is optimized for isolated electrons mostly from decay of W and Z bosons with $p_T > 5$ GeV

'Supercluster' is a group of one or more associated clusters of energy deposits in the ECAL constructed using an algorithm which takes account their characteristic narrow width in the η coordinate and their characteristic spread in ϕ due to the bending in the magnetic field of electrons radiating in the tracker material. As a first filtering step, superclusters are matched to track seeds (pairs or triplets of hits) in the inner tracker layers, and electron tracks are built from these track seeds. Trajectories are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian Sum Filter (GSF). The filtering performed at the seeding step is complemented by a preselection. For candidates found only by the 'tracker driven' seeding algorithm, the preselection is based on a multivariate analysis. For candidates found by the 'ECAL driven' seeding algorithm, the preselection is based on the matching between the GSF track and the supercluster in η and ϕ. The few 'ECAL driven' electron candidates ($\sim$ 1% for isolated electrons) not accepted by these matching cuts but passing the multivariate preselection are also kept.

Figure 4.2 presents the kinematical p_T and η distributions of the reconstructed electron candidates. The data corresponds to an integrated luminosity of $3\ \text{nb}^{-1}$. The Monte Carlo distribution is normalized to the total number of electron candidates in the data. No other selection than the reconstruction preselection is applied to electron candidates.

4.3.5 Jets

Four types of jets are reconstructed at CMS, which differently combine individual contributions from subdetectors to form the inputs to the jet clustering algorithm: Calorimeter Jets, Jet-Plus- Track (JPT Jets), Track Jets and Particle-Flow (PFlow or PF) Jets[30]. Since PFlow jets are the mostly used jets in CMS, we only concentrate on the this type:

The Particle Flow algorithm combines the information from all CMS sub-

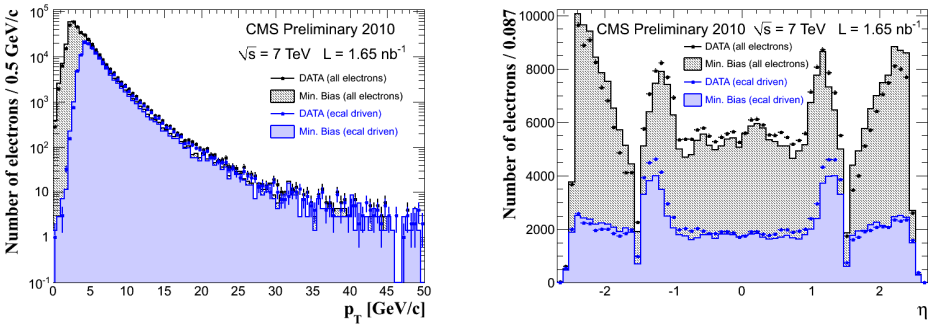

Figure 4.2: Kinematical distributions of electron candidates in minimum bias events (dots) compared with Monte Carlo (histograms) for all reconstructed candidates and for reconstructed candidates from the ECAL driven seeding algorithm: (a) transverse momentum distribution and (b) pseudorapidity [29].

detectors to identify and reconstruct all particles in the event, namely muons, electrons, photons, charged hadrons and neutral hadrons. The detailed description of the algorithm and its commissioning can be found in [31]. Charged hadrons, in particular, are reconstructed from tracks in the central tracker. Photons and neutral hadrons are reconstructed from energy clusters in the electromagnetic and hadron calorimeters. Clusters separated from the extrapolated position of tracks in the calorimeters constitute a clear signature of these neutral particles. A neutral particle overlapping with charged particles in the calorimeters can be detected as a calorimeter energy excess with respect to the sum of the associated track momenta. PFlow jets are then reconstructed from the resulting list of particles. The jet momentum and spatial resolutions are expected to be improved with respect to calorimeter jets as the use of the tracking detectors and of the excellent granularity of the ECAL allows to resolve and precisely measure charged hadrons and photons inside jets, which constitute ~ 90 % of the jet energy.

Jet energy measured in the detector is typically different from the corresponding particle jet energy [32]. The latter is obtained in the simulation by clustering, with the same jet algorithm, the stable particles produced during the hadronization process that follows the hard interaction. The main cause for this energy mismatch is the non-uniform and non-linear response of the CMS calorimeters. Furthermore,

electronics noise and additional pp interactions in the same bunch crossing (event pile-up) can lead to extra unwanted energy. The purpose of the jet energy correction is to relate, on average, the energy measured in the detector to the energy of the corresponding particle jet.

CMS has developed a factorized multi-step procedure for the jet energy calibration (JEC). The following three subsequent (sub-)corrections are devised to correct calorimeter, PFlow and JPT jets to the corresponding particle jet level; offset, relative and absolute corrections.

- The offset correction aims to correct the jet energy for the excess unwanted energy due to electronics noise and pile-up.
- The relative correction removes variations in jet response versus jet η relative to a central control region chosen as a reference because of the uniformity of the detector.
- The absolute correction removes variations in jet response versus jet pt.

The default sequence for the jet energy corrections is expressed mathematically as:

$$E_{\text{corrected}} = (E_{\text{uncorrected}} - E_{\text{offset}}).C_{\text{Rel}}(\eta, p_T^{''}).C_{\text{Abs}}(p_T^{'}) \tag{4.1}$$

where $p_T^{''}$ is the transverse momentum of the jet corrected for offset and $p_T^{'} = p_T^{''}.C_{\text{Rel}}(\eta, p_T^{''})$ is the transverse momentum of the jet corrected for offset and pseudorapidity dependence.

Figure 4.3 shows the absolute jet energy correction factors C_{Abs} derived from simulation for calorimeter, JPT, and PFlow jets as a function of corrected jet transverse momentum.

4.3.6 MET

In general, E_T^{miss} or MET is the negative of the vector sum of the transverse momenta of all final-state particles reconstructed in the detector. CMS has developed three distinct algorithms to reconstruct E_T^{miss} [33].

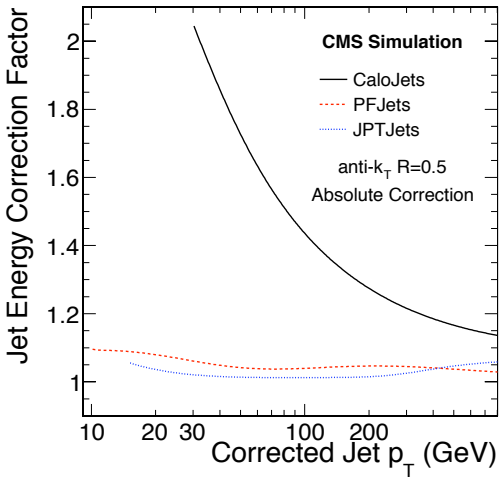

Figure 4.3: Absolute jet energy correction factors C_{Abs} derived from simulation for calorimeter, JPT, and PFlow jets at $\sqrt{s} = 7$ TeV as a function of corrected jet transverse momentum.

1. **PFMET**

   is based on the complete particle-flow technique, is calculated from the reconstructed PF particles. PFET is the associated scalar sum of the transverse energies of the PF particles.

2. **CaloMET**

   is based on calorimeter energies and the calorimeter tower geometry, is calculated using the energies contained in and directions of calorimeter towers to define massless pseudo-particles. The sum excludes energy deposits which are below noise thresholds. Since a muon deposits only a few GeV on average in the calorimeter, independent of its momentum, the muon p_T is included in the Calo E_T while simultaneously excluding the small calorimetric energy deposit associated to the muon track. CaloET is the associated scalar sum of the transverse energies of the calorimeter towers and muons.

3. **TCMET**

   is based on CaloMET but also includes the p_T of tracks which have been reconstructed in the inner tracker, while removing the expected calorimetric energy deposit of each track. All tracks that are not identified as electrons or

muons use the predicted energy deposition for charged pions. The calorimetric energy deposit is estimated from simulations of single pions, in bins of p_T and η, and an extrapolation of the track in the CMS magnetic field is used to determine its expected position. No correction is applied for very high-p_T tracks ($p_T > 100$ GeV), whose energy is already well measured by the calorimeters. Low-p_T tracks ($p_T < 2$ GeV), are fully compensated for, assuming no response from the calorimeter.

The particle-based SumEt is seen in Figure 4.4 to be close to the true generated SumEt in comparison to the calo-based one. Three reasons govern this observation.

- First, charged hadrons (measured by the tracker) and photons (measured by the ECAL) are reconstructed at the correct energy scale and represent about 80% of the event energy.
- Second, the particle-flow algorithm is able to reconstruct very low-energy particles, down to a p_T of 100 MeV/c for charged hadrons, and to an energy of 200 MeV for photons.
- Third, the hadronic-cluster calibration brings the neutral hadron energy, which accounts for the remaining 20% of the event energy, to the proper scale as well.

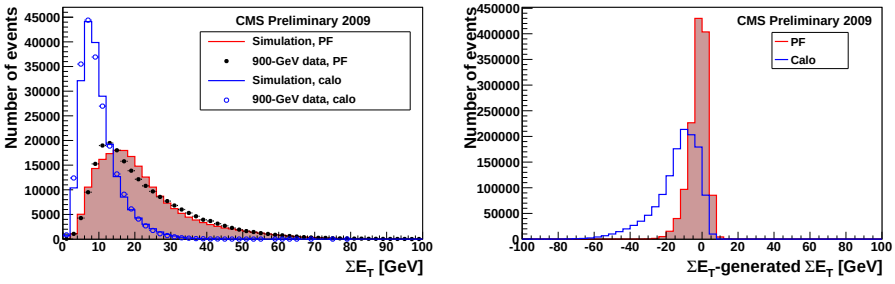

Figure 4.4: (a) Distribution of the calorimeter-based (hollow) and particle-based (solid) SumEt in the data (dots) and in the simulation (histogram). (b) Difference between the calorimeter-based (hollow histogram) or the particle-based (solid histogram) SumEt and the generated SumEt, in the simulation. [33]

Chapter 5

τ-lepton Reconstruction and Identification

5.1 Tau Reconstruction Algorithms

Different algorithms are used to reconstruct tau at CMS. They are either based on calorimeters and tracker information or objects reconstructed with Particle Flow algorithm. Taus which are based on PFlow can further be categorized with respect to their exploitation of finding tau decay mode in their reconstruction. Figure 5.1 shows a complete picture of different tau reconstructing algorithm. In the following all methods are described. More details for each algorithm can be found in [34]

5.1.1 TCTAU

The Track Corrected Tau (TCTau) algorithm is to identify tau leptons decaying hadronically considering jets reconstructed using tracks and energy deposits in the electromagnetic and hadronic calorimeters.

All reconstructed tracks with the cone size of $\Delta R = 0.1$ and minimum p_T of 0.5 GeV/c are sorted and the track with highest p_T track (at least $p_T > 5.0$ GeV/c) is selected as a 'leading' track. All other tracks within the signal cone of size $\Delta R = 0.07$ around the leading track direction are taken as tau lepton decay products. Tracks outside the signal cone, but within an isolation annulus of outer radius $\Delta R = 0.5$ are counted for the track isolation requirement.

One of the main differences in the topology of the QCD jet and tau is the narrowness of tau jets. The track isolation criteria is satisfied if no tracks of transverse momenta $p_T > 1.0$ GeV/c are found within the isolation annulus and the ECAL isolation requirement based on a maximum value of 1.5 GeV for summing the energy deposits in the ECAL.

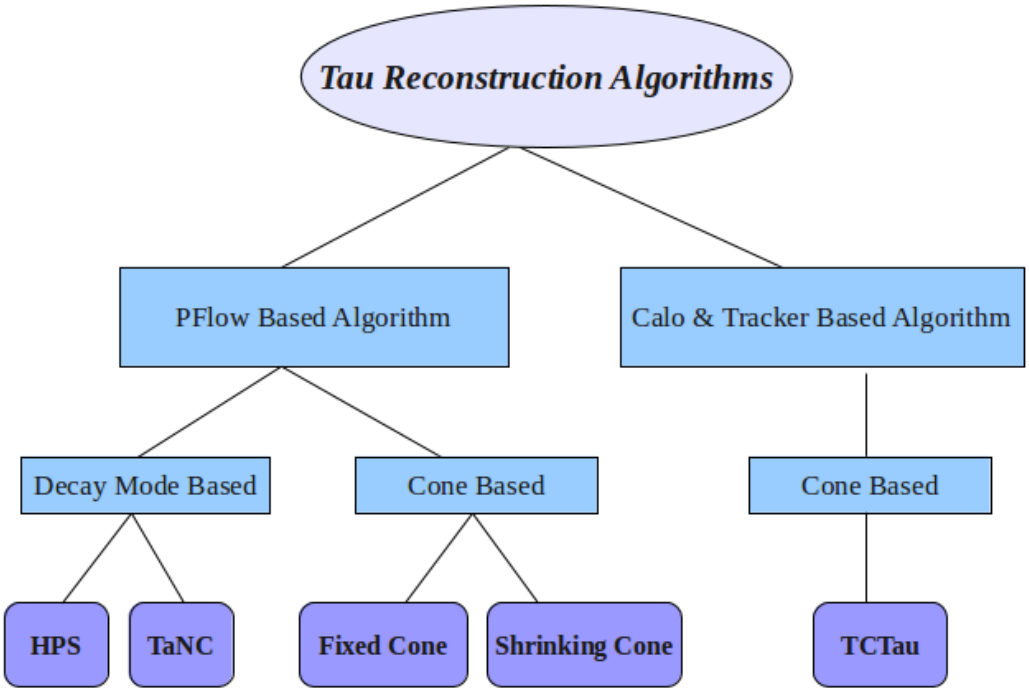

Figure 5.1: Different tau reconstruction algorithms at CMS

The energy and momentum of the tau jet is computed by the TCTau algorithm, which as an input uses jets which energy and direction have been corrected using tracker information [35].

5.1.2 PFTAU

Fixed and Shrinking Cone

In the particle–flow approach, the tau four-momentum is reconstructed as a sum of the four–momenta of all particles with p_T above 0.5 GeV/c in a signal cone of radius $\Delta R = 0.15$ around the direction of the highest p_T ('leading') particle in the jet. The leading particle should be within a cone of radius $\Delta R = 0.1$ (matching cone) around jet direction and is required to have p_T above 5 GeV/c. The tau isolation requirement is based on particle counting in an isolation annulus of outer radius $\Delta R = 0.5$ around the leading particle. The isolation annulus is defined by a narrow signal cone surrounded by a wider isolation cone (the same as TCTau) and is required to contain no charged hadrons with $p_T \geq 1$ GeV/c and no photons with $p_T \geq 1.5$ GeV/c. There are two different types of particle flow tau identification algorithms which their isolation cone are the same $\Delta R = 0.5$, while they differ in signal cone as follows:

- the **fixed signal cone** algorithm uses a cone of size $\Delta R = 0.15$ for photons and $\Delta R = 0.07$ for charged hadrons.

- the **shrinking signal cone** algorithm uses a cone of fixed size $\Delta R = 0.15$ for photons and a shrinking cone, $\Delta R = 5.0/E_{\mathrm{T}}$ for charged hadrons, where E_{T} is transverse energy of the particle–flow tau jet.

  A further increase of the signal-cone size was shown to have no visible effect on the isolation efficiency, while noticeably increasing the background contamination [36].

More detailed information on the cut based particle–flow isolation for tau identification algorithm can be found in [36, 37].

Tau Neural Classifier

The Tau Neural Classifier (TaNC) algorithm was designed to improve the performance of the isolation–based tau identifications algorithms by considering the different hadronic decay modes of the tau individually.

Since the majority of the tau decay modes(a full picture is given in Table 5.1) are via decaying an intermediate resonance, the problem of hadronic tau identification re-framed from global search for collimated hadrons satisfying the tau mass constraint into an ensemble of searches for single production of the different hadronic tau decay resonances.

Decay Mode	Resonance	Mass (MeV/c^2)	Branching ratio(%)
$\tau^- \to h^-\nu_\tau$			11.6 %
$\tau^- \to h^-\pi^0\nu_\tau$	ρ	770	26.0 %
$\tau^- \to h^-\pi^0\pi^0\nu_\tau$	$a1$	1200	10.8 %
$\tau^- \to h^-h^+h^-\nu_\tau$	$a1$	1200	9.8 %
$\tau^- \to h^-h^+h^-\pi^0\nu_\tau$			4.8 %
Total			63.0%
Other hadronic modes			1.7%

Table 5.1: Resonances and branching ratios of the dominant hadronic decays of the τ-lepton [38]. The decay products listed correspond to a negatively charged τ-lepton; the table is identical under charge conjugation.

TaNC algorithm implements this approach using two complimentary techniques: a method to reconstruct the decay mode and an ensemble of neural network classifiers to suppress quark and gluon jets which can mimic the same final state topology.

The major task in reconstructing the decay mode of the tau is determining the number of π^0 mesons which decays almost instantaneously to a pair of photons.(BR of $\pi^0 \to \gamma\gamma \geq 99\%$). The photon pairs that are in the shrinking signal cone and fulfills the condition of π^0 mass(less than 0.2 GeV/c^2) are combined as π^0 candidate. There is also a possibility for unpaired photons to be considered as π^0 candidates, provided that their transverse momenta exceed 10% of all tracks plus tagged π^0 candidates within the signal cone.

After the decay mode of the tau candidate has been reconstructed, the TaNC algorithm feeds the tau candidate to a neural network in order to compute the classification decision. By adjusting the thresholds of cuts on the neural network

output three "working points": 'loose','medium' and 'tight' were defined.

Hadron Plus Strips

The Hadron Plus Strips(HPS) algorithm [39] starts from a particle–flow jet and searches for tau lepton decay products produced by any of the hadronic decay modes enumerated in Table 5.1.

As neutral pions are produced very often in hadronic tau decays, one focus of the HPS tau identification algorithm is an optimized π^0 reconstruction. The possible broadening of calorimeter signatures by photon conversions is accounted for in the HPS algorithm by reconstructing photons in 'strips', objects which are built out of electromagnetic particles.

The strip reconstruction starts by centering one strip on the most energetic electromagnetic particle reconstructed by the particle–flow algorithm. The algorithm then searches for other electromagnetic particles within the window of size $\Delta\eta = 0.05$, $\Delta\phi = 0.20$ until no further particles are found which can be associated to the strip. After that search for a new strip continues. Strips are finally combined with the charged hadrons to reconstruct individual hadronic tau lepton decay modes.

The decay modes which are considered by the HPS tau identification algorithm are:

- **Single hadron**: This signature reconstructs $\tau^- \to h^-\nu_\tau$ decays and $\tau^- \to h^-\pi^0\nu_\tau$ decays in which the neutral pions have too little energy to be reconstructed as strips.

- **One hadron + one strip**: This signature aims to reconstruct the decay mode $\tau^- \to h^-\pi^0\nu_\tau$ in events in which the impact position of the photons from π^0 decays are close together on the calorimeter surface.

- **One hadron + two strips**: This signature aims to reconstruct the decay mode $\tau^- \to h^-\pi^0\nu_\tau$ in events in which the impact positions of photons from π^0 decays are well separated on the calorimeter surface.

- **Three hadrons**: This signature aims to reconstruct the decay mode $\tau^- \to$

$h^-h^+h^-\nu_\tau$ and the three charged hadrons are required to have charge sum $|q| = 1$.

All charged hadrons and strips are required to be contained within a narrow cone of size $\Delta R = 2.8/p_T^\tau$,

The four–vectors of hadrons and strips are reconstructed according to the respective tau decay mode hypothesis and are required to be compatible with the masses of intermediate meson resonances listed in Table 5.1. (for example in case of hadron + two strips, a massless four–vector is reconstructed for each strip and the invariant mass of the two strips is required to be within 50 MeV/c^2 and 200 MeV/c^2. The invariant mass of hadron plus strips is required to be in the range $0.4-1.2$ GeV/c^2.)

Finally, reconstructed candidates are required to satisfy isolation criteria which are based on counting the number of charged hadrons and photons not associated to the tau decay signature within an isolation cone of size $\Delta R = 0.5$. Three sets of tau candidate selection criteria ('working–points')are defined:

1. **Loose isolation**: requires no charged hadrons with $p_T > 1.0$ GeV/c and no photons with $E_T > 1.5$ GeV within the isolation cone

2. **Medium isolation**: requires no charged hadrons with $p_T > 0.8$ GeV/c and no photons with $E_T > 0.8$ GeV within the isolation cone

3. **Tight isolation**: requires no charged hadrons with $p_T > 0.5$ GeV/c and no photons of $E_T > 0.5$ GeV within the isolation cone.

Recently another WP, 'VLoose' has been added to the HPS algorithm which has higher fake rate and efficiency. It requires no charged hadrons with $p_T > 1.5$ GeV/c and no photons with $E_T > 2$ GeV within the isolation cone.(ΔR cone of 0.3)

A correlation between generated and reconstructed tau decay modes was studied using a sample of simulated $Z \to \tau_{had}\tau_{had}$ MC events, where τ_{had} denotes the taus decaying hadronically. The results are presented in Fig. 5.2. Each column reproduces one generated decay mode normalized to unity. The y-axis represents the fraction of generated taus of a given type reconstructed in a certain decay mode.

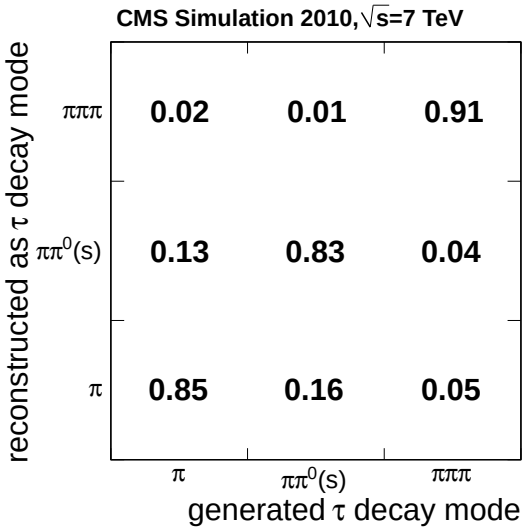

Figure 5.2: The plot shows a fraction of generated taus of a given type reconstructed in a certain decay mode for the HPS 'loose' working point and a sample of $Z \to \tau^+\tau^-$ simulated MC events [40].

The correlation between generated and reconstructed decay modes exceeds 80% even for decays including neutral pions and reaches 90% for three charged pion decay mode.

Figure 5.3 shows a first measurement of tau fake rate with early data and compares the probability of jet to be misidentified as tau in each of the algorithms. The efficiency of tau reconstruction is also shown using MC simulation. As it is shown, the two advanced algorithms which exploit the decay mode finding(although have a bit lower efficiency) performs better in term of QCD jet rejection and due to the large cross section of QCD jet or V+jet in all analysis, these two methods have been considered in most of the CMS analysis.

5.2 Efficiency of tau reconstruction and identification

To find the tau reconstruction and identification efficiency from data [40], a sample of $Z \to \tau\tau \to \mu\tau_{had}$ which is triggered by single muon has been selected. Then we profit from 'tag and probe' method as following:

A well identified and isolated muon with $p_T > 15$ GeV/c has been considered

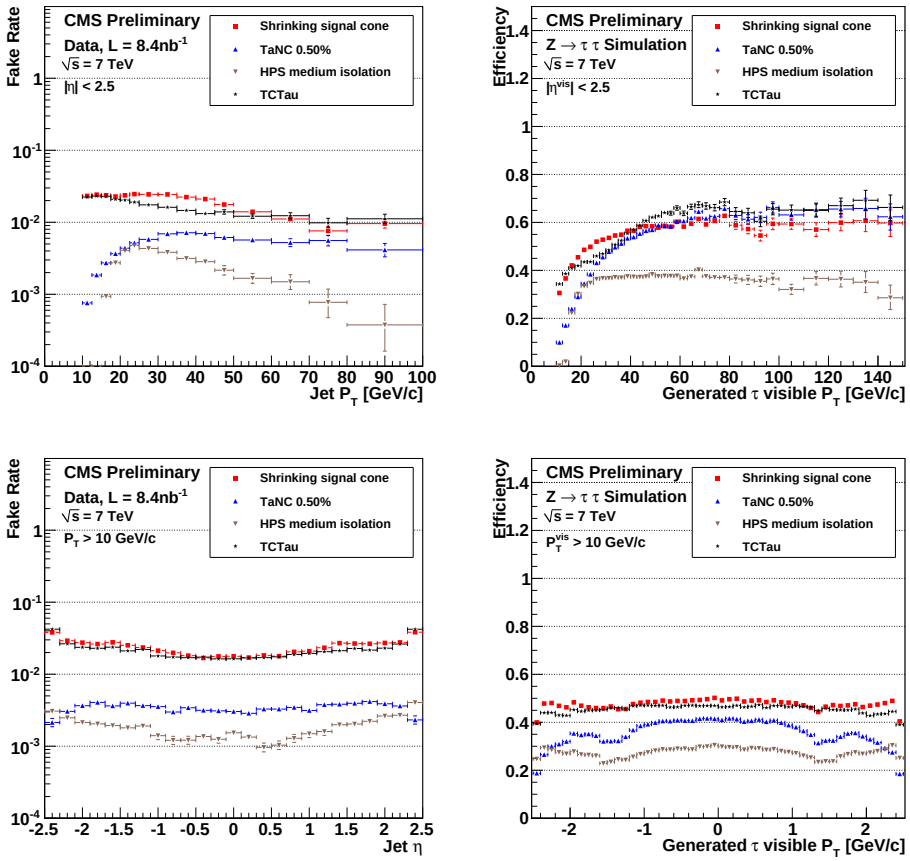

Figure 5.3: Left: Measured probabilities of quark/gluon jets to pass the tau candidate selection criteria of the TCTau (black stars), shrinking cone (red squares), HPS (brown downward facing triangles) and TaNC (blue upward facing triangles) tau identification algorithms as function of jet p_T and η Right: Efficiencies of the algorithms to identify genuine tau lepton hadronic decays as function of transverse momentum and pseudo–rapidity of the visible tau decay products on generator level, estimated using a sample of simulated $Z \to \tau^+\tau^-$ events [34].

as 'tag' and is required to have opposite charge to tau–jet candidate of $P_T^{\tau_{had}} > 20$ GeV/c, which would be the 'probe'. While no cut on tau identification is applied, several cuts are required to suppress the contribution of background like QCD , W and $Z \to \mu\mu$.

The sample of tau–jet candidates passing the 'leading' track P_T and loose isolation requirements in the selected $Z/\gamma^* \to \tau^+\tau^- \to \mu\tau_{had}$ candidate events is divided into two subsamples: the sample of tau–jet candidates which pass the TaNC/HPS discriminators and those which fail. In each subsample, the number of $Z/\gamma^* \to \tau^+\tau^-$ signal and $Z \to \mu^+\mu^-$, W + jets, $t\bar{t}$ and QCD background events is determined by a fit of the $\mu + \tau_{had}$ visible invariant mass distribution.

The efficiency of each TaNC/HPS discriminator is obtained by the normalization factor obtained by the fit for the $Z/\gamma^* \to \tau^+\tau^-$ signal contribution to the 'pass' and 'fail' subsamples:

$$\varepsilon = \frac{N_{pass}^{Z\to\tau\tau}}{N_{pass}^{Z\to\tau\tau} + N_{fail}^{Z\to\tau\tau}}.$$

The template shapes used in the fit for signal and background processes are taken from the Monte Carlo simulation.

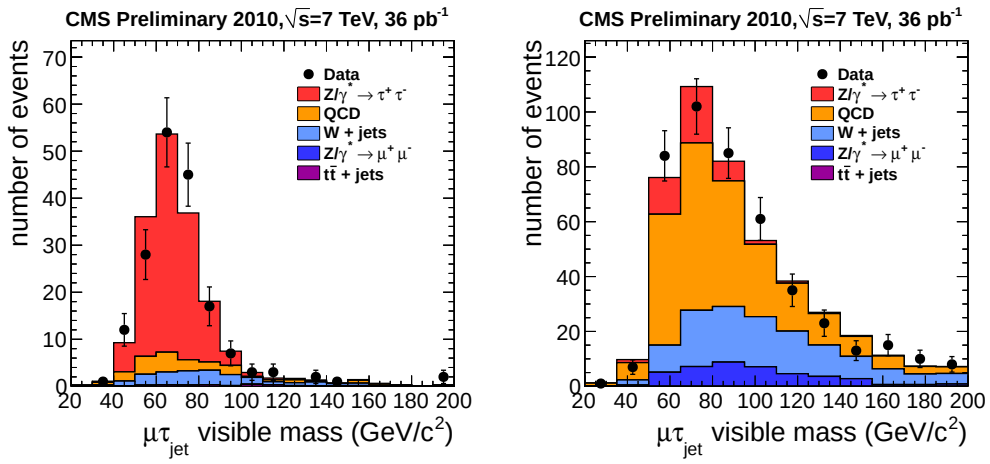

Figure 5.4: The visible invariant mass of the $\mu\tau_{jet}$ system for preselected events which passed (left) and failed (right) the HPS 'loose' tau identification requirements compared to predictions of the MC simulation [40]

Figure 5.4 shows visible invariant mass of the $\mu\tau_{jet}$ system for preselected events which passed (left) and failed (right) the 'loose' tau identification requirements.

Since in the 'failed' sample there is no tau reconstructed, to make the plots consistent the visible mass is always computed using the jet four–vector for taus and not the tau four–vector as reconstructed by identification algorithms. The MC predictions for signal and background events are also shown. The MC describes data well in the 'pass' region which has a well defined shape dominated by Z events and a small background contribution. The sample of 'failed' events is dominated by background contributions. The MC predictions describe data reasonably well.

Algorithm	Fit data	Expected MC	DATA/MC
TaNC "loose"	0.76 ± 0.20	0.72	1.06 ± 0.30
TaNC "medium"	0.63 ± 0.17	0.66	0.96 ± 0.27
TaNC "tight"	0.55 ± 0.15	0.55	1.00 ± 0.28
HPS "loose"	0.70 ± 0.15	0.70	1.00 ± 0.24
HPS "medium"	0.53 ± 0.13	0.53	1.01 ± 0.26
HPS "tight"	0.33 ± 0.08	0.36	0.93 ± 0.25

Table 5.2: Efficiency for hadronic tau decays to pass TaNC and HPS tau identification criteria measured by fitting the $Z \to \tau^+\tau^-$ signal contribution in the samples of the 'passed' and 'failed' preselected events. The errors of the fit represent statistical uncertainties. The last column represents the data to MC correction factors and their full uncertainties including statistical and systematic components.

Results of the fits are summarized in Table 5.2. The values measured in data, 'Fit data' are compared with the expected values, 'Expected MC', which is found by performing the fitting on simulated events.

Efficiency of tau reconstruction and identification depends on the p_T and η. Table 5.3 shows the expected efficiency values for hadronic taus from $Z \to \tau\tau$ process for two different selections,(one with p_T of hadronic tau to be greater than 15 GeV/c and the other to be greater than 20 GeV/c).

Algorithm	TaNC			HPS		
	"loose"	"medium"	"tight"	"loose"	"medium"	"tight"
Efficiency ($p_T^{\tau_{had}} > 15$ GeV)	53.6%	43.1%	30.4%	45.9%	33.8%	22.9%
Efficiency ($p_T^{\tau_{had}} > 20$ GeV)	57.8%	47.9%	35.6%	49.9%	36.5%	24.6%

Table 5.3: The expected efficiency for hadronic tau decays to pass TaNC and HPS tau identification criteria estimated using $Z \to \tau\tau$ events simulated using MC for two different cuts on the visible p_T of the taus.

5.3 Tau scale energy and the uncertainty

Since hadronic taus consist of charged hadrons and photons reconstructed with a high precision using the PF techniques the reconstructed tau energy is expected to be close to the true energy of the visible tau decay products. According to simulation, the ratio of the reconstructed to the true visible tau energy for the HPS algorithm is constant and within 2% from unity, while for TaNC it decreases by about 2% towards $p_T^\tau = 60$ GeV/c.

The η dependence is more pronounced. For both algorithms the reconstructed tau energy degrades with respect to the true one by moving from the central, barrel, to endcap region by about 5%.

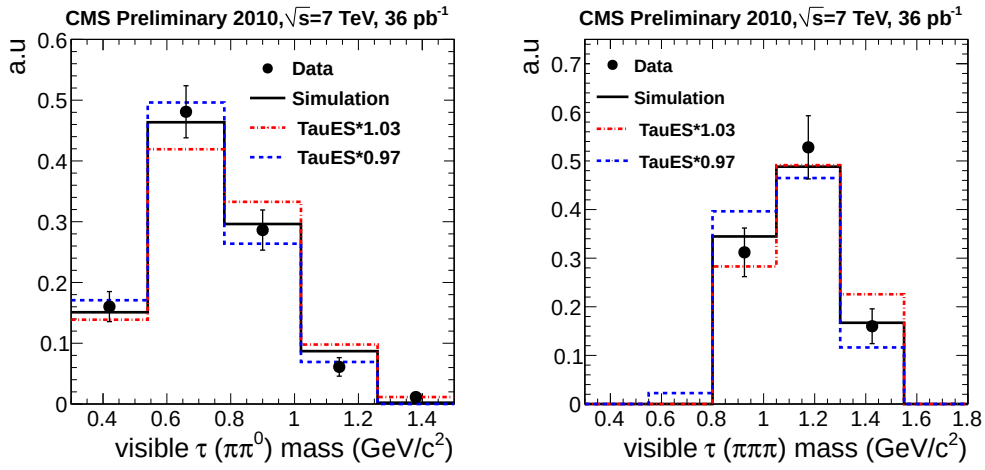

Figure 5.5: The reconstructed invariant mass of taus decaying into one charged and one neutral pion (left) and into three charged pions (right) compared to predictions of the simulation. The black lines represent results of the fit and the dashed lines represent the predictions with tau energy scale, TauES, varied up and down by 3%.

It is expected that the p_T and η-dependent behavior of the tau response is well reproduced by the MC simulation, and that the ratio of the tau response in data to the response in MC is very close to unity. The quality of the tau energy scale simulation can be examined by analyzing the $Z \to \tau^+\tau^- \to \mu\tau_{had}$ data sample, assuming that the $\tau\tau$ mass spectrum is not altered significantly by non-SM sources of $\tau\tau$ events. The visible invariant mass of the $\mu\tau_{had}$ system is very sensitive to the energy scale of the τ, since the muon four–momenta are measured with a high

precision. By varying the tau energy scale simultaneously in the signal and background MC samples a set of templates was produced. The resulting templates were fit to the data and the best agreement was achieved by scaling the tau energy in simulation by a factor 0.97 ± 0.03, where the uncertainty of the tau energy scale is averaged over the pseudorapidity range of the data sample.

A complementary procedure, which does not assume knowledge of the $\tau\tau$ invariant mass spectrum, was based on the invariant mass of reconstructed tau decays and is shown in Fig. 5.5. The method uses taus as standalone objects but still relies on good understanding of underlying background events which contribute to the signal sample. The fit was performed separately for $\pi\pi^0$ and $\pi\pi\pi$ decay channels, since the major source of the uncertainty is expected to come from reconstruction of the electromagnetic energy. The simulation describes both decay channels well. The data to MC ratio was measured to be 0.97 ± 0.03 for the $\pi\pi^0$ decay mode and 1.01 ± 0.02 for the $\pi\pi\pi$ decay mode. The effect of the energy scale uncertainty on the shape of the tau invariant mass is also illustrated in Fig. 5.5. Varying the energy scale in simulation by the uncertainty delivered from the $\mu\tau_{had}$ invariant mass fit, 3%, leads to a significant deviation in the predicted mass shapes of taus.

5.4 Tau Fake Rate

5.4.1 Fake Rate of Jets

Since the signature of jets are very similar to taus , they can easily fake taus. In this section the probabilities with which generic, quark and gluon jets pass the selection criteria of different tau identification algorithms is described. The main sources of these jets are jets produced in QCD processes and jets produced in association with Z and W bosons. For each sample an special selection has been used as follow:

QCD **dijet**

Selecting dijet events requires at least one jet of transverse momentum $p_T^{jet} >$ 15 GeV and a jet of $p_T^{jet} > 20$ GeV within $|\eta| < 2.3$. The jet that we want to find fake rate should not match to the trigger object.

QCD enriched Muon

To select those QCD events which are enriched by muon(in these events muons mainly are coming from heavy quarks like b–quark), a muon is required to be identified with $p_T^{jet} > 15$ GeV, but not isolated and to discriminate these events from W, we require that transverse mass of muon and MET to be less than 40 GeV. Furthermore one jet of transverse momentum $P_T^{jet} > 20$ GeV within $|\eta| < 2.3$ is also required.

W + jet

To select an event sample dominated by $W \to \mu\nu$ events, one muon of transverse momentum $P_T^{jet} > 15$ GeV and $|\eta| < 2.1$ and relative isolation $relIso < 0.1$ is required in addition to one jet of transverse momentum $P_T^{jet} > 20$ GeV within $|\eta| < 2.3$.

Z + jet

And finally for jet associated Z, events were selected with a good Z bosons(which is explained at [41] with decays to $\mu\mu$ and an associated jet is required to pass the 20 GeV cuts on its p_T and 2.3 on η.

For each of these samples additional cuts were also applied to suppress the background contribution from events with jets of other types.

Fake–rates values are determined separately for each working points of TaNC and HPS algorithm by counting the fraction of quark and gluon jets passing the identification criteria of one particular algorithm in a given p_T^{jet} or η_{jet} bin [34]:

$$P_{fr}\,(\text{bin}) := \frac{N_{jets}\,(\text{bin})_{\text{passed tau identification}}}{N_{jets}\,(\text{bin})}, \tag{5.1}$$

where N_{jets} denotes the number of jets with $p_T^{jet} > 20$ GeV/c and $|\eta_{jet}| < 2.3$ reconstructed in that bin.

Figure 5.6 shows the fake–rate as a function of the jet p_T for the 'loose' working points of the TaNC and HPS algorithms, where the measured values are compared with the MC predictions for the different type of jets.

The values of the fake–rates expected from simulation and the measured data to

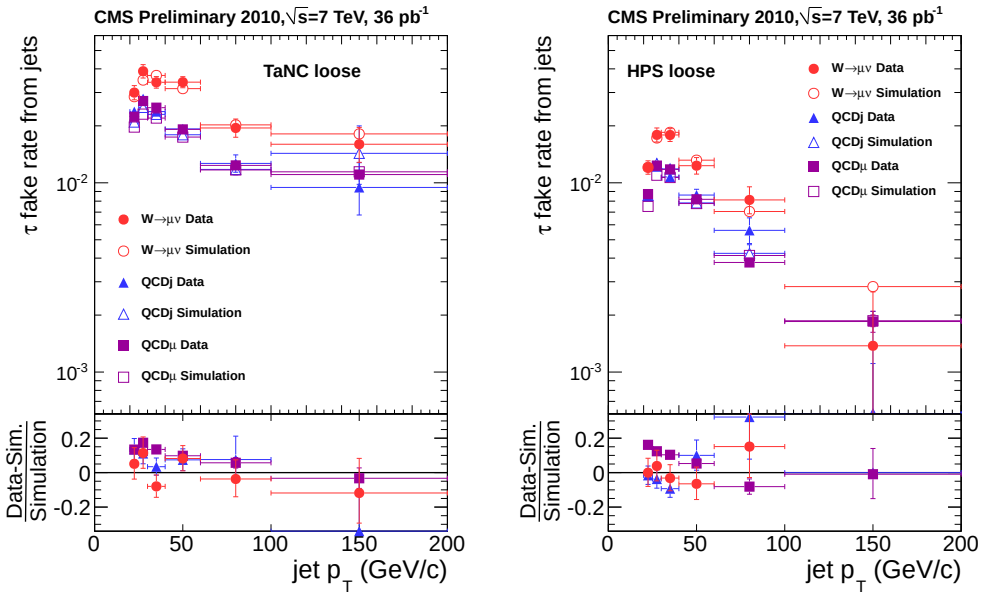

Figure 5.6: Probabilities of quark and gluon jets to pass "loose" working points of the TaNC (left) and HPS (right) algorithms as a function of jet p_T for QCD, QCD μ-enriched and W type events. Fake-rate measured in data are represented by solid symbols and compared to MC prediction represented by open symbols.

MC ratios integrated over the p_T and η phase space used in the $Z \to \tau\tau$ analysis, $p_T^{jet} > 20$ GeV and $|\eta| < 2.3$, are summarized in Table 5.4 for the three working points of both reconstruction algorithms.

Algorithm	QCDj		QCDμ		W + jets	
	MC	DATA/MC	MC	DATA/MC	MC	DATA/MC
TaNC "loose"	2.1%	1.05 ± 0.04	1.9%	1.12 ± 0.01	3.0%	1.02 ± 0.05
TaNC "medium"	1.3%	1.05 ± 0.05	0.9%	1.08 ± 0.02	1.6%	0.98 ± 0.07
TaNC "tight"	0.5%	0.98 ± 0.07	0.4%	1.06 ± 0.02	0.8%	0.95 ± 0.09
HPS "loose"	1.0%	1.00 ± 0.04	1.0%	1.07 ± 0.01	1.5%	0.99 ± 0.04
HPS "medium"	0.4%	1.02 ± 0.06	0.4%	1.05 ± 0.02	0.6%	1.04 ± 0.06
HPS "tight"	0.2%	0.94 ± 0.09	0.2%	1.06 ± 0.02	0.3%	1.08 ± 0.09

Table 5.4: The MC expected fake–rate values and the measured data to MC ratios integrated over the p_T and η phase space typical for the $Z \to \tau\tau$ analysis.

Since More data is needed to measure the fake rate in Z+Jets events, the same approach has been done with 200 pb^{-1} and the fake rates has been measured and shown in Figure 5.7;

The fake–rate as a function of efficiency for all working points for both (HPS

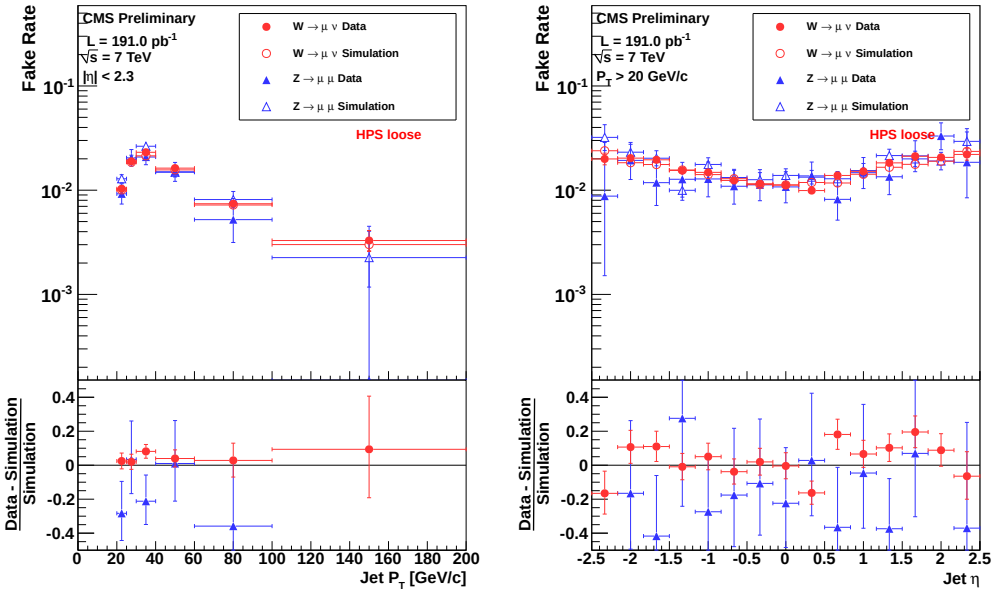

Figure 5.7: Probabilities of quark and gluon jets to pass "loose" working points of the HPS algorithms as a function of jet p_T for (left) and as a function of jet η(right) for W type and Z type events. Fake-rate measured in data are represented by solid symbols and compared to MC prediction represented by open symbols.

and TaNC) algorithms is shown in Fig. 5.8, which summarizes the MC estimated efficiency and the measured fake–rate values presented in Tables 5.6 and 5.3. Since the QCD and QCD μ-enriched fake–rates values were observed to be similar only one set of QCD points is shown. Open symbols represent results obtained by running a fixed cone algorithm based on the PF taus on simulated events only. The algorithm is effectively the same which was used in the CMS physics technical design report (PTDR) [13] and is shown for comparison.

5.4.2 Fake Rate of Electron

Isolated electrons passing the identification and isolation criteria of the tau algorithms can also be misidentified as hadronic taus and represent an important source of background to many analyses with taus in the final state. Both electron and charged pion(which is the main constitute of hadronic tau) deposit their energy in calorimeter and while the latter deposit most of its energies in HCAL, the former does in ECAL. However it is very probable that electron is misidentified as

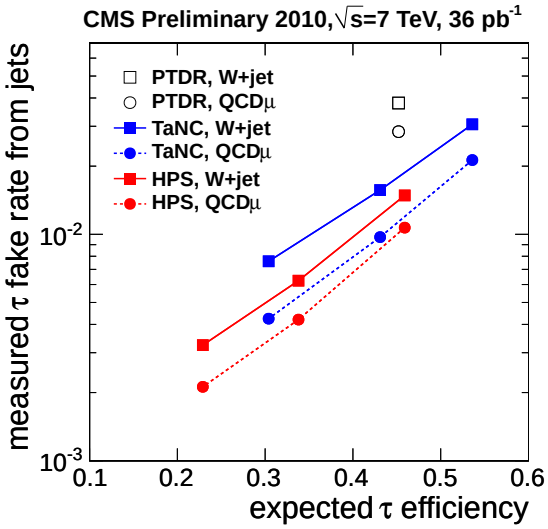

Figure 5.8: The measured fake–rate as a function of MC estimated efficiency for all working points for QCD μ-enriched and W data samples. The PTDR points represent results of the fixed cone algorithm based on the PF taus.

a pion. A multivariate discriminator is used to reduce this background improving the pion/electron separation. The same discriminator, implemented as a Boosted Decision Tree is used in the PF algorithm and its output is denoted by ξ.

Two working points are defined which the looser is optimized for a low fake–rate, $O(2\%)$, at the price of about 4% losses of real hadronic taus and the tighter corresponds to a larger fake-rates ($O(20\%)$) but higher efficiency exceeding 99.5%

To estimate the $e \to \tau_{had}$ fake-rate a sample of isolated electrons coming from the decay of the $Z \to ee$ resonance was selected and tag and probe method is performed. Tag would be a well identified and isolated electron while the probe is the electron which passes all tau-identification and isolation discriminators without any cut to veto electron.

The ratio between the number of probes passing the anti-electron discriminator to the overall number of selected probes is an estimator of the $e \to \tau_{had}$ fake–rate.

Table 5.5 shows the ratio between fake-rates as measured in the data and those obtained using MC for two $|\eta|$ bins. The results from a closure test on MC is also shown in the table.

bin	discriminator by $\xi < -0.1$			discriminator by $\xi < 0.6$		
$\lvert\eta\rvert$	MC(exp),%	MC(t&p),%	DATA/MC	MC(exp),%	MC(t&p),%	DATA/MC
< 1.5	2.22 ± 0.03	2.21 ± 0.05	1.13 ± 0.17	13.19 ± 0.06	13.10 ± 0.08	1.14 ± 0.04
> 1.5	3.90 ± 0.06	3.96 ± 0.09	0.82 ± 0.18	26.90 ± 0.13	26.80 ± 0.16	0.90 ± 0.04

Table 5.5: The expected MC values of the fake–rates of electron and measured ratio of the data to MC response.

Chapter 6

$W \to \tau\nu$ cross–section measurement

6.1 Motivation of the Analysis

The study of $W \to \tau^{\pm}\nu$ production in the $\tau^{\pm}_{had}\nu$ final state is an important contribution to tau–physics studies at the LHC. Besides testing the Standard Model, the analysis of $W \to \tau^{\pm}\nu$ production is important, as it is an important background process to several searches for new physics [13, 14]. In particular it is the main irreducible background to the search for charged Higgs bosons in the $\tau^{\pm}\nu$ final state.

In the following we describe the measurement of the $W^{\pm} \to \tau^{\pm}\nu$ cross–section[42]. The tau lepton is identified by its hadronic decay mode. The $W \to \tau^{\pm}\nu$ channel benefits from the large production cross-section, exceeding the production rate of $Z \to \tau^{+}\tau^{-}$ events(another dominant source of tau leptons in SM [43]) by nearly an order of magnitude. However, the experimental signature of a single tau-jet plus undetected neutrinos is more challenging, requiring a good understanding of hadronic tau identification and missing transverse energy.

6.2 Data and Monte Carlo samples

The measurement of the $W \to \tau_{had}\nu$ cross–section is based on analyzing the b–tau secondary datasets for 7 TeV proton–proton collisions recorded by CMS in 2010. The data has been re–reconstructed with CMSSW (a CMS Software) and corresponds to an integrated luminosity of 32.4 pb^{-1}.

Monte Carlo simulated events are used to estimate acceptance and efficiency for the $W \to \tau\nu$ signal and expected contributions of $W \to e\nu$, $W \to \mu\nu$ and $Z \to \tau^{+}\tau^{-}$

Trigger path	L [pb^{-1}]
HLT_SingleIsoTau20_Trk15_MET20	14.6
HLT_SingleIsoTau20_Trk15_MET25	17.8

Table 6.1: Triggers used for $W \to \tau\nu$ cross–section measurement.

background processes. The QCD background contribution is determined directly from Data. All samples have been generated by PYTHIA [22] and POWHEG [24]. The generated events are passed through the full GEANT [25] based simulation of the CMS apparatus and are reconstructed using CMSSW.

The simulated event samples are normalized to next–to–next–to–leading order (NNLO) cross–sections [44, 45] and are reweighted to match the the number of pile–up interactions in 2010 data.

6.3 Trigger

$W \to \tau\nu$ candidate events are triggered using a 'cross–channel' trigger requiring the presence of a single tau–jet plus missing transverse energy (E_T^{miss}). Two versions of this 'cross–channel' trigger were used in most of the 2010 data–taking period. As the instantaneous luminosity delivered by the LHC increased at the end of 2010, the E_T^{miss} threshold of the trigger was raised from 20 to 25 GeV, to keep the trigger rate within bandwidth restrictions. 17.8 pb^{-1} (14.6 pb^{-1}) of data was collected by the higher (lower) E_T^{miss} threshold version of the $\tau_{had} + E_T^{miss}$ trigger, *cf.* Tab. 6.1.

Requirement at L1 and HLT are as follow:

L1

Both versions of the trigger are seeded by a Level 1 (L1) tau object reconstructed solely based on calorimeter information and the E_T of L1 tau is required to exceed 20 GeV. In order to increase the trigger efficiency, jets not passing the tau object selection at L1, but passing a higher E_T threshold (30 GeV), are also considered as seeds. Events passing the L1 seed conditions are then processed by the High Level Trigger (HLT).

HLT

At the HLT, event is required to pass different criteria at different level:

- **L2 τ:** Reconstructed tau–jet candidates are required to have $E_T > 20$ GeV and $|\eta| < 3.0$ and satisfy ECAL isolation:($\sum E_T < 5$GeV, computed by summing energy deposits in the ECAL within an annulus of outer (inner) radius $\Delta R = 0.50$ (0.15) around the tau direction).

- **L2 $\mathbf{E_T^{miss}}$:** $CaloE_T^{miss}$ is required to exceed 20(25) for two different range of runs.

- **L2.5 Leading Track:** Track of $P_T > 15$GeV matching the direction of the tau–jet candidate within $\Delta R = 0.20$ is required.

- **L3 Tracker Isolation:** No tracks of $P_T > 1$ GeV within an annulus of outer (inner) radius $\Delta R = 0.50$ (0.15) around the direction of the tau.

6.3.1 Trigger efficiency

The precision with which the efficiency of the $\tau_{had} + E_T^{miss}$ trigger is modeled by the Monte Carlo simulation is verified by comparison to data by decoupling the trigger to two parts. The efficiency of τ_{had} and E_T^{miss} conditions are checked independently, using different event samples. The efficiency for hadronic tau decays to pass the τ_{had} requirements is checked using a sample of $Z \to \tau^+\tau^- \to \mu\tau_{had}$ events triggered by single muon triggers(according to the event yields measured in [43]). The efficiency to pass the E_T^{miss} conditions is measured in samples of $W \to e\nu$ events triggered by single electron triggers and in QCD events triggered by jet triggers.

Using Monte Carlo simulated events it has been verified that the efficiency for $W \to \tau\nu$ events to pass the $\tau_{had} + E_T^{miss}$ trigger factorizes into independent terms for the τ_{had} and E_T^{miss} legs.

For the τ_{had} and E_T^{miss} trigger efficiencies to factorize, it is necessary to parametrize the efficiency of the τ_{had} (E_T^{miss}) leg by the P_T of the offline reconstructed calorimeter jet corresponding to the tau–jet candidate (by the offline reconstructed calo–E_T^{miss}).

The values of P_T^τ and E_T^{miss} computed by the $\tau_{had} + E_T^{miss}$ trigger at HLT level are solely based on calorimeter information. These parameters are hence highly correlated with offline reconstructed calo–P_T^τ and calo–E_T^{miss} values, while being quite different from PF–P_T^τ and PF–E_T^{miss}. The correlations between P_T^τ and E_T^{miss} computed by the $\tau_{had} + E_T^{miss}$ trigger at HLT level and offline reconstructed values are illustrated in Fig. 6.1.

As a consequence of these correlations, the trigger efficiency turn–one curves of P_T^τ leg as well as of E_T^{miss} leg are significantly 'sharper' when parametrized by offline reconstructed calorimeter observables compared to the parametrization by corresponding observables reconstructed by the PF algorithm. This effect is illustrated in Fig. 6.2.

The trigger efficiencies of τ_{had} and E_T^{miss} legs measured in data are compared to Monte Carlo expectations in Fig. 6.3. The efficiency of the E_T^{miss} leg is very similar in simulated $W \to \tau\nu$, $W \to e\nu$ and QCD events and in good agreement with the trigger efficiencies measured in data. There is also agreement between data and MC for the efficiency of the τ_{had}.

As there is no indication of a difference between Data and Monte Carlo simulation in $Z \to \tau^+\tau^- \to \mu\tau_{had}$, $W \to e\nu$ and QCD events, we take the trigger efficiency for $W \to \tau\nu$ events from the Monte Carlo simulation.

We assign an uncertainty of 15% to the efficiency of the $\tau_{had}+E_T^{miss}$ trigger. Two components are included in this uncertainty. The dominant component is due to statistical uncertainties, in particular of the $Z \to \tau^+\tau^- \to \mu\tau_{had}$ sample selected in data. We propagate the statistical uncertainties of individual calo–jet P_T bins into an uncertainty on the $W \to \tau\nu$ trigger efficiency by fitting the efficiency measured in data by a Gaussian error function. We use the resulting fit parameter values plus uncertainties to compute upper and lower bounds for the the τ_{had} trigger efficiency as function of calo–jet P_T. The uncertainty on the $W \to \tau\nu$ trigger efficiency is then obtained by taking the difference between upper and lower bounds to the central value of the fit, averaged over the calo–jet P_T distribution expected for $W \to \tau\nu$ signal events after all analysis cuts are applied. The value obtained amounts to 15%. The same procedure is applied to estimate the uncertainty on trigger efficiency of the E_T^{miss} leg. The event statistics of the $W \to e\nu$ and QCD samples is sufficient

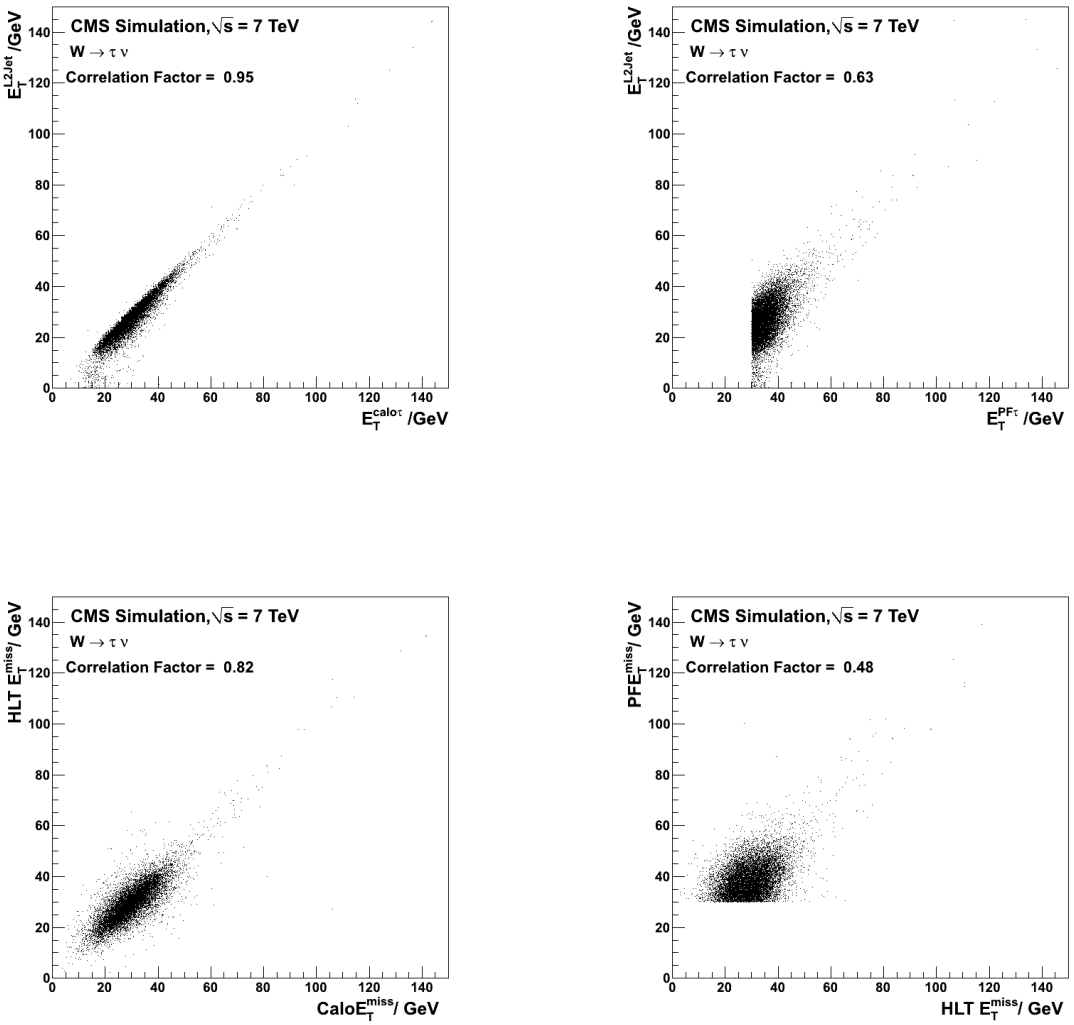

Figure 6.1: Correlation between E_T^{τ} (top) and E_T^{miss} (bottom) computed at HLT level with offline reconstructed calorimeter (left) and PF quantities (right) in simulated $W \to \tau_{had}\nu$ signal events.

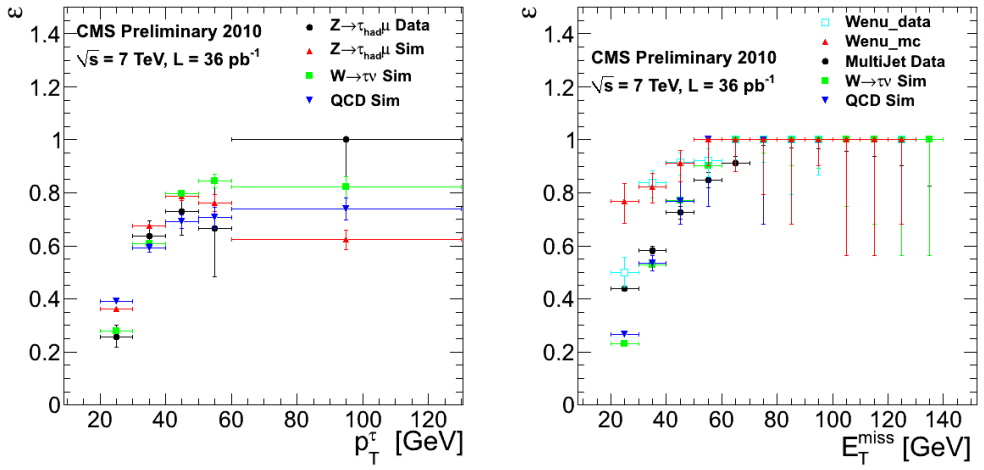

Figure 6.2: Efficiency of the τ_{had} leg of the $\tau_{had} + E_T^{miss}$ 'cross–channel' trigger as function of offline reconstructed PF–P_T^{τ} (left) and efficiency of the E_T^{miss} leg as function of offline reconstructed PF–E_T^{miss} (right).

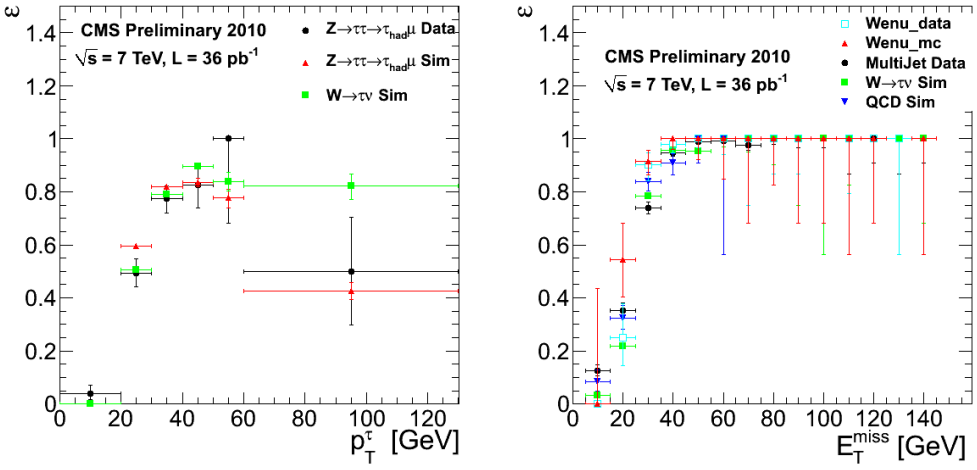

Figure 6.3: Efficiencies of τ_{had} (left) and E_T^{miss} (right) legs of the $\tau_{had} + E_T^{miss}$ 'cross–channel' trigger. The trigger efficiency of the τ_{had} (E_T^{miss}) leg is parametrized by the P_T of the offline reconstructed calorimeter jet corresponding to the tau–jet candidate (offline reconstructed calo–E_T^{miss}).

to yield small uncertainties, amounting to 2% (2%) in case of $W \to e\nu$ (QCD). The two fits are illustrated in Fig. 6.4. The second component of the uncertainty on the $\tau_{had}+E_T^{miss}$ trigger efficiency is related to the choice of binning plus fitting function. The corresponding uncertainties are estimated by repeating the fit using a different binning and replacing the Gaussian error function by the integral of the Crystal ball function [46]. The effect of using a different binning (fitting function) amounts to 2% (2%). We add these differences in quadrature, as well as the statistical uncertainties on the efficiencies of the τ_{had} and E_T^{miss} legs and obtain a relative uncertainty on the $\tau_{had} + E_T^{miss}$ trigger efficiency of 15%.

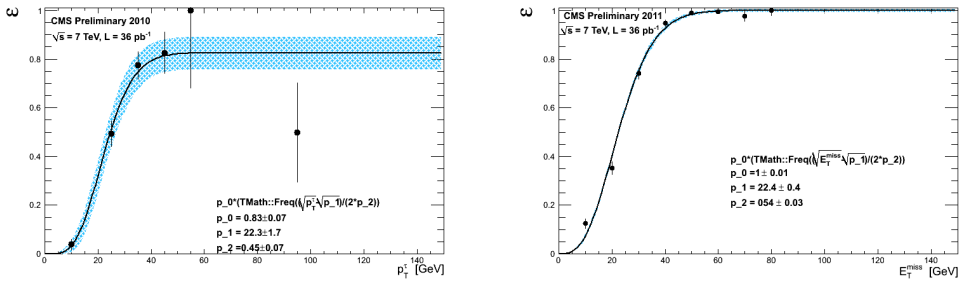

Figure 6.4: Efficiency plots for data, fitted with the Gaussian error function for $Z \to \tau^+\tau^- \to \mu\tau_{had}$(left) and Multijet (right). The colorful band corresponds to the uncertainty on the trigger efficiency.

6.4 Particle reconstruction and identification

Hadronic tau decays are reconstructed by the HPS tau identification algorithm described in 5.1. All other particles that have been used in the analysis(Jets, MET, muons and electrons) are described in 5.1. Isolated electrons and muons are used in this analysis only for the purpose of vetoing $W \to e\nu$, $W \to \mu\nu$ and $Z \to \tau^+\tau^-$ background events. In order to preserve efficiency for the $W \to \tau\nu$ signal, muons are required to satisfy $P_T^{\mu} > 15$ GeV, $|\eta_{\mu}| < 2.4$ and to pass 'VBTF' muon identification criteria [41]. Electrons are required to satisfy $P_T^{e} > 15$ GeV, $|\eta_{e}| < 2.4$ and to pass identification criteria of the nominal 85% efficiency working–point (WP) [47].

6.5 Event selection

$W \to \tau_{had}\nu$ candidate events are selected by requiring:

- The event to pass the $\tau + E_T^{miss}$ trigger path described in section 6.3.
- An event vertex with ≥ 4 DoF, reconstructed within $-24 < z_{vtx} < +24$ cm and $|\rho| < 2$ cm, where ρ denotes the distance between reconstructed vertex and beam–spot position in the transverse plane.
- A τ_{had} candidate with $P_T^{\tau} > 30$ GeV and $|\eta_{\tau}| < 2.3$, reconstructed by the HPS tau identification algorithm, and passing the HPS 'medium' isolation discriminator(as discussed before). Additionally, the τ_{had} candidate is required to contain at least one charged hadron of $P_T > 15$GeV and to pass the 'loose' discriminator against muons and 'tight' discriminator against electrons [40].
- The missing transverse energy reconstructed in the event to satisfy PF–$E_T^{miss} > 30$ GeV and Calo–$E_T^{miss} > 25$ GeV.
- The event not to contain any muon or electron of $P_T^{\ell} > 15$ GeV and $|\eta_{\ell}| < 2.4$, which passes the identification criteria detailed in section 6.4 and passing a 'relative' isolation criterion: $Iso < 0.10 \cdot P_T^{\ell}$ which by Iso mean the sum of energy deposit in tracker, ECAL and HCAL within a cone of size $\delta R = 0.3$ around the lepton direction.
- The ratio of P_T of the τ_{had} candidate to the sum of transverse momenta of PF–jets with $P_T > 15$ GeV and $|\eta| < 3$ reconstructed in the event (R_{HT}) exceed 0.65.
- The transverse mass, M_T, of τ_{had} candidate plus missing transverse energy:

$$M_T = \sqrt{(P_T^{\tau} + E_T^{miss})^2 - \left((P_x^{\tau} + E_x^{miss})^2 + (P_y^{\tau} + E_y^{miss})^2\right)}$$

  to satisfy $M_T > 40$ GeV.

The aim of the cuts on R_{HT} and M_T is to reduce QCD background.

6.6 Background estimation

Background contributions to the sample of $W \to \tau_{had}\nu$ candidate events passing the selection criteria described in section 6.5 are due to QCD multi–jet events and electroweak (EWK) backgrounds: $W \to e\nu$, $W \to \mu\nu$ and $Z \to \tau^+\tau^-$. All background contributions are estimated directly from data, via the 'template method'. The estimate obtained for the dominant background, QCD, which is also the background most difficult to model by the Monte Carlo simulation, is cross–checked using a complementary method, the 'ABCD' method.

6.6.1 Template method

The idea of the template method is to determine the contribution of signal plus background processes to the sample of $W \to \tau_{had}\nu$ candidate events selected in data via a fit of distributions observed in data with a set of 'template' histograms. Each template histogram is normalized to unit area and represents the shape of the distribution expected for one particular signal or background processes. The contributions of signal plus background processes to the event sample selected in data is then obtained by the normalization factors obtained by the best fit of all template shapes to the distribution observed in data.

The choice of observable which gets used in the template fit is made based on the level of discrimination between the signal and background processes provided by the observable: Good separation between the template shapes causes the normalization factors to be well constrained by the fit, yielding small uncertainties on the signal and background contributions to the event sample selected in data.

We have chosen *metTopology* to be the template fit observable used in the $W \to \tau_{had}\nu$ analysis. *metTopology* is defined as the ratio between the P_T sum of PF–candidates projected in the *met* direction to the P_T sum of PF–candidates projected opposite to the direction of *met*. For signal events we expect very little activity in direction of *met* and thus small values of *metTopology*. QCD multi–jet background events have typically higher values of *metTopology*, even after passing a PF–$E_T^{miss} >$ 30 GeV cut, as QCD events are intrinsically balanced and the reconstructed E_T^{miss} typically results from a mismeasurement of one (or more) of the jets in the event.

The shape templates for the $W \to \tau_{had}\nu$ signal and EWK background processes are taken from the Monte Carlo simulation. The template for the QCD background is obtained from data, using a control region. The control region is chosen such that the a high purity QCD sample is acquired, while the shape of the *metTopology* distribution in the control region represents the distribution of QCD events passing the selection criteria described in section 6.5.

With respect to the selection of $W \to \tau_{had}\nu$ candidate events, the QCD enriched control region is defined by inverting the isolation requirement of the τ_{had} candidate. The vetos against additional electrons or muons in the event are tightened by lowering the P_T^{ℓ} threshold to 8 GeV. The purity of QCD events selected in the QCD enriched control region is expected to be $\mathcal{O}(75\%)$.

The signal and background yields obtained by fitting the the *metTopology* shapes for $W \to \tau_{had}\nu$, QCD and EWK backgrounds to the *metTopology* distribution observed in data are given in Table 6.2.

Process	Estimate
QCD	158.8
$W \to e\nu$	45.2
$W \to \mu\nu$	4.9
$Z \to \tau\tau$	53.2
$\sum$ backgrounds	261.3

Table 6.2: Normalization factors for background processes obtained by the template fit.

The sum of signal and background templates scaled by the normalization factors obtained by the fit is compared to the *metTopology* distribution observed in the sample of $W \to \tau_{had}\nu$ candidate events selected in data in Fig. 6.5. The distribution observed in data is fitted well by the sum of templates.

It is also needed to correct the contamination from EWK and signal to the control region. The corrected value for the QCD background contribution to the sample of $W \to \tau_{had}\nu$ candidate events passing the selection criteria described in section 6.5 amounts to 159 ± 30 events.

The EWK background contributions determined by the template fit are in agree-

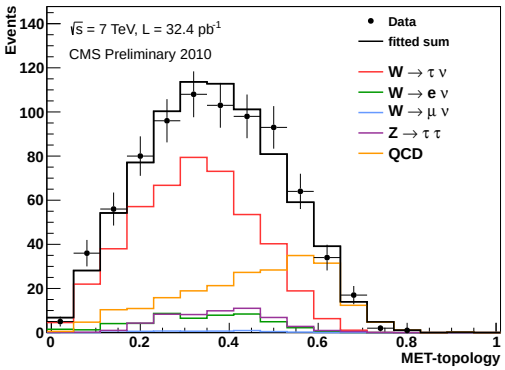

Figure 6.5: *metTopology* distribution observed in data compared to sum of shape templates scaled by normalization factors determined by the fit.

ment with the Monte Carlo expectation.

6.6.2 ABCD method

In the ABCD method, the contribution of QCD multi–jet background to the sample of $W \to \tau_{had}\nu$ candidate events (signal region A) is estimated via extrapolation of event yields measured in three sidebands (B, C and D). The method utilizes two observables, which need to be uncorrelated in order for the method to work.

We use *met* and R_{HT} and define the four regions by:

- signal region A: $R_{\mathrm{HT}} > 0.65$, $met > 30$ GeV.
- sideband B: $R_{\mathrm{HT}} > 0.65$, $met < 30$ GeV
- sideband C: $R_{\mathrm{HT}} < 0.65$, $met < 30$ GeV
- sideband D: $R_{\mathrm{HT}} < 0.65$, $met > 30$ GeV.

Provided that there is no correlation between *met* and R_{HT}, the number of QCD events contributing to the signal region A is related to the number of QCD events in other regions by:

$$N_A = N_B \times \frac{N_D}{N_C}. \tag{6.1}$$

Fig. 6.6 demonstrates that the correlation between *met* and R_{HT} is indeed low.

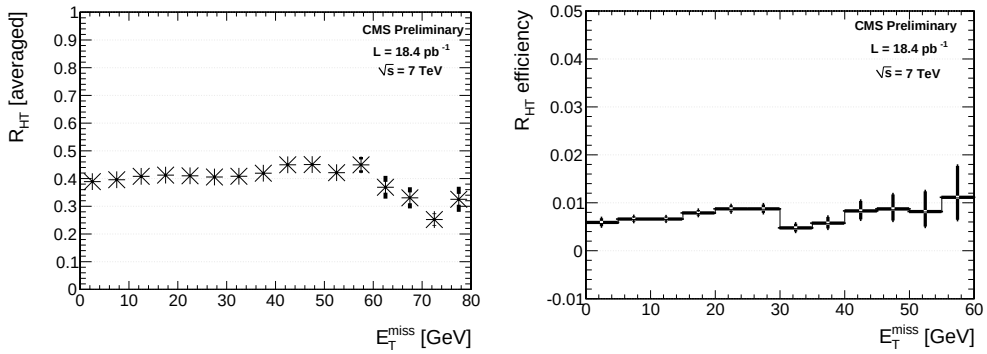

Figure 6.6: Average value of R_{HT} (left) and fraction of events passing $R_{\mathrm{HT}} > 0.65$ cut (right) in different bins of *met*. The events shown in the plots are selected in data and are required to pass all selection criteria described in section 6.5, except *met* and R_{HT}, which are both not applied [48].

Equation 6.1 can be used directly to estimate the QCD background contribution in signal region A, in case the contributions from EWK backgrounds and from the $W \to \tau_{had}\nu$ signal in the sidebands B, C and D is zero. Fig. 6.7 illustrates that the contributions to regions B, C and D from processes other than QCD are low. Monte Carlo simulation predicts that the sum of contributions from EWK backgrounds plus $W \to \tau_{had}\nu$ signal amounts to about 1% in region C, 6% in region B and 15% in region D, respectively. The expected contributions of processes other than QCD are subtracted from the number of events selected in the regions B, C and D before equation 6.1 is applied. The resulting estimate for the QCD event yield in signal region A is $N_A = 209 \pm 6$ events, compatible with the estimate obtained via the template method. The quoted uncertainty represents only the statistical uncertainties on the number of events observed in the sidebands B, C and D.

6.7 Systematic uncertainties

Systematic uncertainties affecting the measurement of the $W \to \tau\nu$ cross–section are due to uncertainties on the efficiency of the $\tau_{had} + E_T^{miss}$ trigger, the τ_{had} identification efficiency, uncertainties on the energy scale of tau–jets and other jets in the event, theoretical uncertainties and uncertainties on the integrated luminosity of the analyzed dataset.

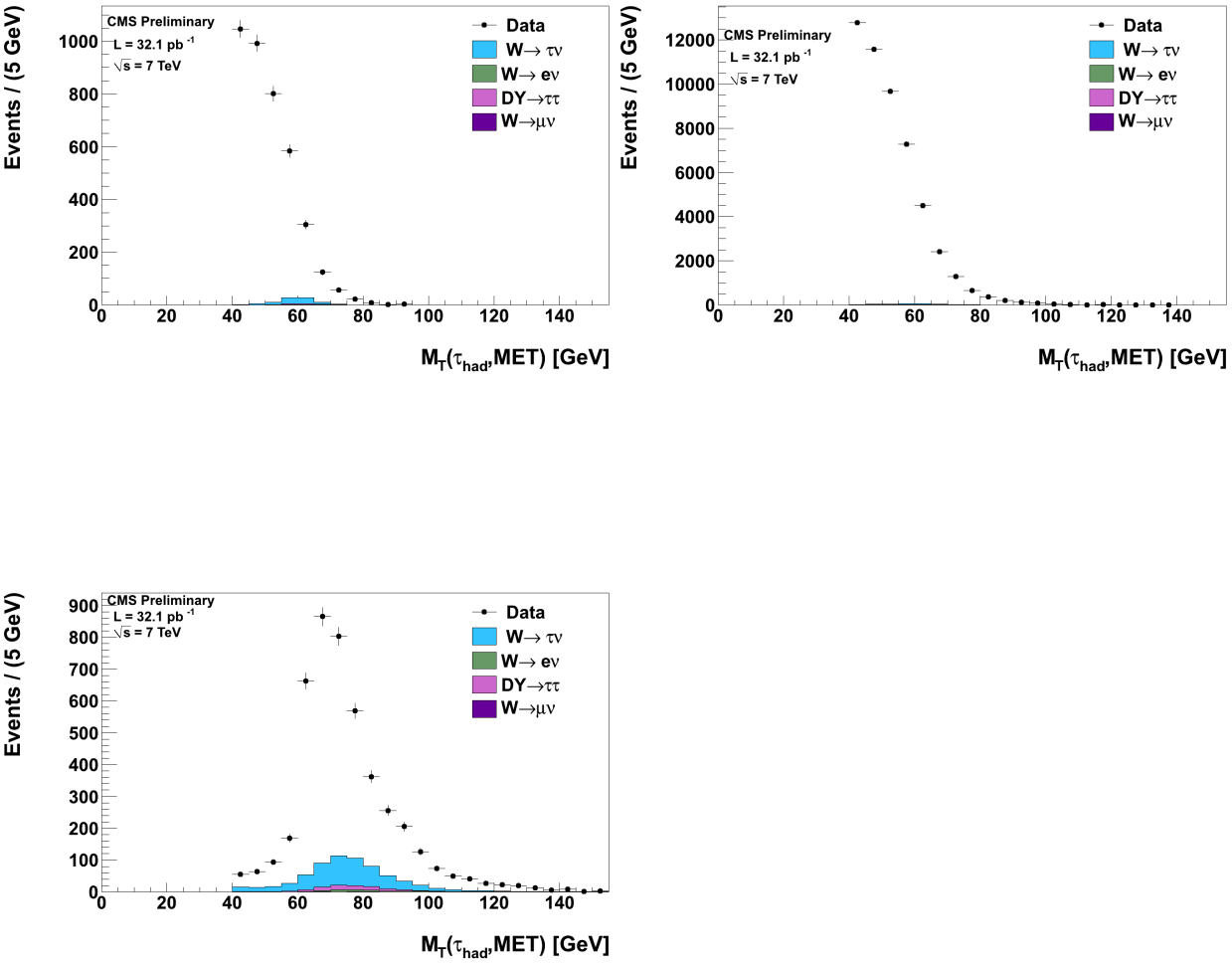

Figure 6.7: Transverse mass distributions of the τ_{had} candidate and met, for the four designated regions in phase space: Region D (bottom left) where $R_{\text{HT}} < 0.65$ and $met > 30$ GeV, Region C (upper right) where $R_{\text{HT}} < 0.65$ and $met < 30$ GeV, and Region B (upper left) where $R_{\text{HT}} > 0.65$ and $met < 30$ GeV. The points represent the data. Simulated signal and electroweak backgrounds are represented by the filled histograms.

The trigger efficiency is measured with an uncertainty of 15%, as described in section 6.3.1.

The efficiency for identifying hadronic tau decays is measured by the ratio of $Z \to \tau^+\tau^-/Z \to \mu^+\mu^-$ event yields with an uncertainty of 7% [40].

The energy scale of tau–jets is known with an uncertainty of 3% [40]. The energy scale of quark and gluon jets is determined from an analysis of the P_T balance between photons and recoil jets in γ + jets events [32]. Jet energy scale uncertainties depend on jet P_T and η and are applied to jets of $P_T > 10$ GeV.

The effect of energy scale uncertainties on the efficiency to pass the $P_T^\tau > 30$ GeV, PF–$E_T^{miss} > 30$ GeV, $R_{\mathrm{HT}} < 0.65$ and $M_T > 40$ GeV cuts applied in the event selection is estimated by varying the energy of tau–jets, quark and gluon jets of $P_T > 10$ GeV plus unclustered energy within their respective uncertainties, re–reconstructing the event after each variation and reevaluating all event selection criteria. The event yield is compared to the nominal value and the difference is taken as systematic uncertainty. The effect of the 3% tau–jet energy scale uncertainty on the $W \to \tau\nu$ event yield amounts to $^{+15.1}_{-14.2}$%. The effect is rather large, as an upward (downward) variation of the tau–jet energy scale simultaneously increases (decreases) the efficiencies to pass P_T, PF–E_T^{miss}, R_{HT} and M_T cuts. Variation of the jet energy scale has the opposite effect: an upward (downward) variation decreases the efficiencies to pass the PF–E_T^{miss} and R_{HT} cuts. We conservatively assume that tau–jet energy scale and jet energy scale are not correlated and obtain an uncertainty on the $W \to \tau\nu$ event yield of $^{+9.5}_{-10.6}$% due to the jet energy scale uncertainty.

Uncertainties due to parton–distribution functions (PDFs) are estimated following the PDF4LHC recommendations [49]. The total theoretical uncertainties amount to 2%.

The integrated luminosity of the analyzed dataset is measured with an uncertainty of 4.5% [50].

The systematic uncertainties considered in the $W \to \tau\nu$ cross–section measurement are summarized in Tab. 6.3

Source	Uncertainty
Trigger efficiency	15%
Tau–jet energy scale	$^{+15.1}_{-14.2}$%
Jet energy scale	$^{+9.5}_{-10.6}$ %
ISR + FSR + PDF	2%
Luminosity	4.5%
Background contributions	15.1%

Table 6.3: Effect of systematic uncertainties on the $W \to \tau\nu$ cross–section measurement. The uncertainty arising from background contributions to the sample of $W \to \tau_{had}\nu$ candidate events selected in data is described in section 6.8.

6.8 Cross–section extraction

The cross–section for $W \to \tau_{had}\nu$ production is determined by the number N_{sig} of signal events, the integrated luminosity $\mathcal{L}$ of the analyzed dataset, plus signal acceptance A and signal selection efficiency ε using the relation:

$$\sigma = \frac{N_{sig}}{\mathcal{L} \cdot A \cdot \varepsilon}. \tag{6.2}$$

The number of signal events N_{sig} is determined by subtracting from the number of $W \to \tau_{had}\nu$ candidate events passing the selection criteria described in section 6.5 the expected contribution of EWK plus QCD backgrounds.

The signal acceptance is defined to be the fraction of $W \to \tau_{had}\nu$ signal events in which the W boson decays into a tau–jet with $P_T^{\tau_{had}} > 30$ GeV and $|\eta_{\tau_{had}}| < 2.3$ plus neutrino with $P_T^{\nu} > 30$ GeV on generator level. The signal acceptance is obtained from Monte Carlo simulation and is quoted relative to the branching fraction for tau leptons to decay hadronically [38]:

$$A/BR = 0.0790 \pm 0.0002(\text{stat.}) \tag{6.3}$$

$$BR(\tau \to \tau_{had}) = 0.648. \tag{6.4}$$

The signal selection efficiency ε is defined as the fraction of $W \to \tau_{had}\nu$ events within signal acceptance that pass the selection criteria described in section 6.5. Its value is obtained from the Monte Carlo simulation and amounts to:

$$\varepsilon = 0.0334 \pm 0.0006(\text{stat.}). \tag{6.5}$$

6.9 Results

In the analyzed dataset, corresponding to an integrated luminosity of 32.4 pb^{-1}, 793 $W \to \tau_{had}\nu$ candidate events pass the event selection criteria.

Distributions of various kinematic observables reconstructed in the selected events are displayed in Figure 6.8.

The estimated QCD background contribution to the selected event sample, determined by combining the data–driven background estimates obtained via template and ABCD method (*cf.* section 6.6), amounts to 182 ± 19 events.

The contribution of EWK backgrounds is taken from Monte Carlo simulation. The sum of $W \to e\nu$, $W \to \mu\nu$ and $Z \to \tau^+\tau^-$ background contributions is estimated to be 109 ± 44 events. We attribute 40% relative uncertainty to the EWK background contribution. The attributed uncertainty includes uncertainties on $e \to \tau_{had}$ and $\mu \to \tau_{had}$ fake–rates, which amount to about 20% [40, 51], plus energy scale uncertainties, which are estimated to be about 14% for $W \to e\nu$, 12% for $W \to \mu\nu$ and 25% for $Z \to \tau^+\tau^-$, respectively.

The number of $W \to \tau_{had}\nu$ signal events in the selected event sample is estimated to be 502 ± 48 events.

Using equation 6.2 to compute the cross–section with the values 6.4 and 6.5 for signal acceptance and selection efficiency, we measure the $W \to \tau_{had}\nu$ production cross–section in pp collisions at $\sqrt{s} = 7$ TeV center–of–mass energy of

$$\sigma(pp \to WX) \times \mathcal{B}(W \to \tau\nu) = 9.08 \pm 0.51(\text{stat.})^{+2.44}_{-2.38}(\text{sys.}) \pm 0.36(\text{lumi.})\ \text{nb}.$$

The measured cross–section is in agreement with the theory prediction of 10.44 nb, computed at next–to–next–to–leading (NNLO) order, and with the $W \to e\nu$ and $W \to \mu\nu$ cross–sections measured by CMS: $\sigma(pp \to WX) \times \mathcal{B}(W \to e\nu, W \to \mu\nu,) = 9.95 \pm 0.07(\text{stat.}) \pm 0.28(\text{sys.}) \pm 1.09(\text{lumi.})$ nb [41].

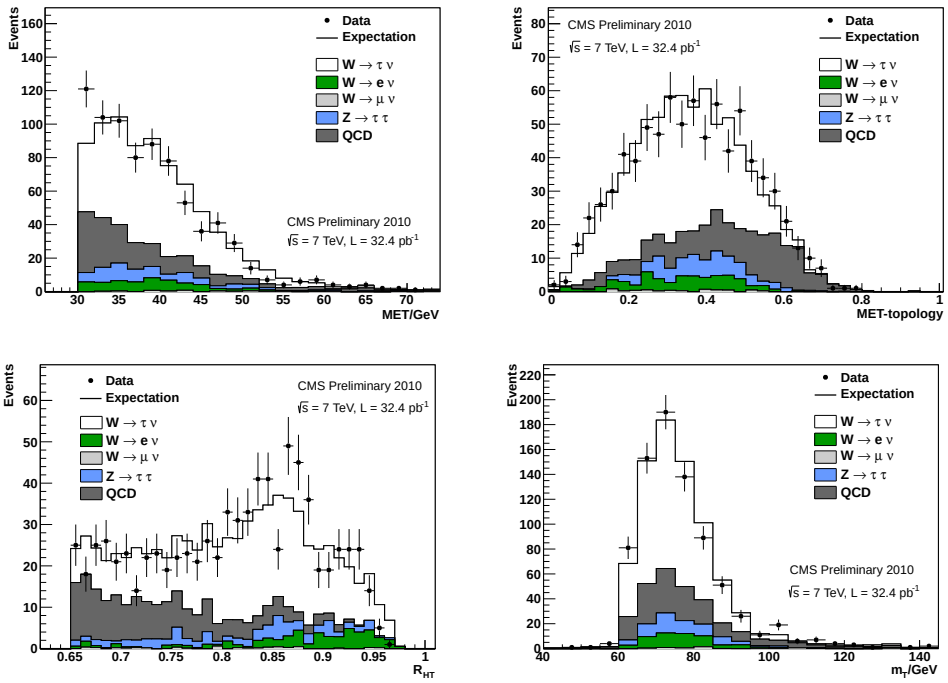

Figure 6.8: Distributions of PF–E_T^{miss} (a), *metTopology* (b) R_{HT} (c) and M_T (d) in the sample $W \to \tau_{had}\nu$ candidate events compared to the expected sum of signal plus background processes. The $W \to \tau_{had}\nu$ signal is scaled according to the measured cross–section, EWK backgrounds according to the Monte Carlo expectation and QCD background to the yield determined from data. The expected shapes are taken from the Monte Carlo simulation, except for the QCD background, which is obtained from a control region in data, as described in section 6.6.

Chapter 7

Search for Higgs in $H \to ZZ \to 2l2\tau$ final state

7.1 Analysis Motivation

The search for the SM Higgs in the decay $H \to ZZ \to 2l2\tau$ [52] complements the search for $H \to ZZ^{(*)} \to 4l(l = e, \mu)$ [5] at large Higgs mass above the kinematical threshold of the ZZ production. Presence of four leptons in the final state provides a clean signature with small background contributions. The major sources of backgrounds are expected to be: irreducible SM ZZ production and reducible SM Z and WZ production in association with jets, and $t\bar{t}$.

The analysis has been done in two steps. At the step 1 the electron, muon and tau algorithms were selected. The CiC has been chosen for the electrons and VBTF Muon for muons, mainly due to these algorithms are used in [5] and demonstrated a very good performance. For the taus the HPS was chosen, which performance was studied in the $Z \to \tau\tau$ analysis [53, 54]. The values were adjusted by looking at the MC signal sample and mixture of all possible backgrounds as explained later on in the text. The cuts on $Z \to \tau\tau$ leg have been set not as tight as for the inclusive Z production, since the background in ZZ case is expected to be much smaller. An absolute minima of the cuts were chosen to ensure robust performance of the analysis and data–MC comparison.

At the step 2 the tuning was performed using data. The signal efficiency was taken from the ZZ MC as in step 1, but the background contribution was estimated using data–driven method. The cuts were tuned finally to get the best signal to background rejection as explained later on in the text.

Study on $H \to ZZ \to 2l2\tau$ has been done by 1.1 fb^{-1} and updated with 2.1 fb^{-1}. The MC samples, triggers, selections and systematical errors are the same in

two analysis while the method for reducible background estimations and extracting the exclusion limit differs between two analyses.

7.2 Data and MC samples and Trigger Selection

A set of MC samples is used to reproduce the mixture of signal and background events.

The Drell–Yan background is simulated with the next-to-leading order (NLO) Monte Carlo generator POWHEG [55, 56, 57]. The diboson backgrounds are simulated with PYTHIA [58]. The top samples are simulated with Madgraph [59]. The τ–lepton decays are generated with Tauola [60]. The Higgs boson samples are generated using POWHEG, which incorporates NLO gluon fusion ($gg \to H$) and vector-boson fusion ($q\bar{q} \to q\bar{q}H$). The MC samples are mixed based on the corresponding cross-sections corrected to NLO. The pileup conditions in the MC are re-weighted to the pile-up observed in the data.

The primary data set which are trigger by double muon or double electron has been used in this analysis. MC are also required to pass the doubleMuon or doubleElectron triggers. A summary of the triggers is presented in Table 7.1.

HLT path	L1 seed
$\mu\mu$ channels	
DoubleMu7 Mu13_Mu8	L1_DoubleMu3 L1_DoubleMu_5_Open
ee channels	
Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL	L1_SingleEG12

Table 7.1: HLT paths used to select the final sample, based on the decay products of the leading Z leg

7.3 Event Selections

Eight different final states are investigated in this analysis. Each signal consists of two to four light leptons (e, μ) and up to two hadronic taus.

The 'leading' Z consists of pair of well identified leptons(μ or e) which are also expected to trigger the events. The first leg in common between both the signal and

background contributions with the exception of the $t\bar{t}$ background which is largely rejected by requiring a narrow mass around the leading Z mass.

The strategy for the leading leptons is to be as loose as possible for leading lepton ID and isolation, only requiring the cuts to be tighter than those applied on the trigger. The background rejection is then achieved by requiring two more leptons in the event, distinguishing between jets and light leptons or hadronic taus.

Details of the event selection is defined as follows.

7.3.1 Leading Z Selection

The leading Z boson candidate is required to have leptons of opposite charge that are isolated and well-identified using criteria discussed below.

$\mathbf{Z \to \mu + \mu}$

- Two opposite charged, global or tracker muons with $p_T > 20$ GeV (leading), 10 GeV (subleading) and $|\eta| < 2.4$
- Muon Combined Relative Isolation less than 0.25 with fast jet corrections.

  The fast jet correction uses the Fastjet algorithm to calculate the energy density(ρ) due to PU and UE. Therefore since the isolation cone is known, this energy can be subtracted from the isolation sum as following:

$$I_{rel}(\rho) = \frac{\Sigma\left(P_T^{Tracker} + max(E_T^{ECAL} + E_T^{HCAL} - \rho\pi\Delta R^2, 0.0)\right)}{P_T^l}$$

  where ΔR is the isolation cone size(0.3) and $P_T^{Tracker}$, E_T^{ECAL} and E_T^{HCAL} correspond to energy deposit in tracker, ECAL and HCAL, respectively.
- Invariant mass of two muons to be between 60 and 120 GeV

$\mathbf{Z \to e + e}$

- Two opposite electrons(CiC tight ID) with $p_T > 20$ GeV (leading), 10 GeV (subleading) and $|\eta| < 2.5$
- No more than 1 missing inner tracker hits for the electron.

- Electron combined relative isolation less than 0.2 with fast jet corrections.
- Invariant mass of two electrons to be between 60 and 120 GeV

Data to MC comparison for the leading Z decaying into $\mu\mu$ and ee with 2.1 fb^{-1} is shown in Fig. 7.1.

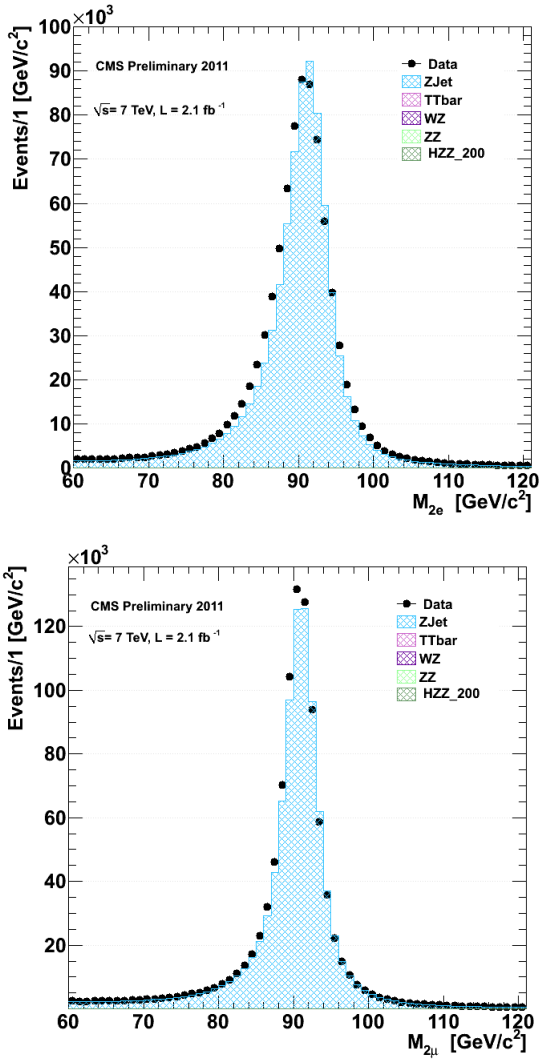

Figure 7.1: Reconstructed invariant mass of the leading Z to ee and $\mu\mu$ candidates.

7.3.2 Second Z Selection

$Z \to \mu + \tau_{had}$

In this mode the second Z decays into a pair of taus with one τ decaying to a μ

and two neutrinos and the second τ decays hadronically with the addition of another neutrino. The presence of the muon makes this channel relatively clean but small backgrounds remain from Z+Jets or WZ, where a real Z decay to a pair of electrons or muons and a real muon from a b jet or W and a jet faking the hadronic τ. In order to reduce these backgrounds the following selection is applied:

- Identified(VBTF) and isolated(relIso < 0.15) muon with $p_T > 10$ GeV and $|\eta| < 2.4$
- HPS tau with transverse momentum greater than 20 GeV and $|\eta| < 2.3$
- Tau is required to pass electron loose discriminator and muon tight discriminator and Loose Isolation
- Charge of the muon and tau must be opposite, $(Q_{\tau_{had}} + Q_{\mu} = 0)$.
- $30 \leq$ Visible Mass $(\mu\tau) \leq 80$

$\mathbf{Z \rightarrow e + \tau_{had}}$

In this mode second Z decays to a pair of taus where one τ decays to an electron and two neutrinos and the second τ decaying hadronically with the addition of another neutrino. This channel is expected to have more backgrounds than the $\mu - \tau$ channels from Z+Jets or WZ. There is also a chance of a charged pion or a photon faking the electron. In order to reduce these backgrounds the following selection is applied:

- Identified (CIC tight) and isolated(relIso < 0.05) electron with $p_T > 10$ GeV and $|\eta| < 2.5$
- No missing inner tracker hits for the electron.
- HPS tau with transverse momentum greater than 20 GeV and $|\eta| < 2.3$
- tau is required to pass electron loose discriminator and muon loose discriminator and Loose Isolation
- Charge of the electron and tau must be opposite, $(Q_{\tau_{had}} + Q_{e} = 0)$.
- $30 \leq$ Visible Mass $(e\tau) \leq 80$

$\mathbf{Z \to \tau_{had} + \tau_{had}}$

This mode where the taus from second Z decay fully hadronically, has the largest background due to the fake–taus being reconstructed out of hadronic jets, the main background source comes from a Z production in association with two or more jets. This background can be controlled by requiring a stringent tau isolation requirement summarized as follows:

- Two opposite charge HPS τ with minimum p_T 20 GeV and $|\eta|$ of less than 2.3.
- tau is required to pass electron loose discriminator and muon loose discriminator and Medium Isolation
- $30 \leq$ Visible Mass $(\tau\tau) \leq 80$

$\mathbf{Z \to e + \mu}$

In this mode, Z decays into a pair of taus with the one tau decaying to a muon and a pair of neutrinos and the second tau decaying to an electron and a pair of neutrinos. This channel is very clean but also has the lowest branching ratio. The remaining small backgrounds come from Z+Jets and WZ, these backgrounds are reduced with the following cuts:

- Identified(VBTF) and isolated(relIso < 0.25) muon with $p_T > 10$ GeV and $|\eta| < 2.4$
- Identified (CIC tight) and isolated(relIso < 0.2) electron with $p_T > 10$ GeV and $|\eta| < 2.5$
- No more than 1 missing inner tracker hits for the electron.
- The charge of the electron and muon must be opposite ($Q_e + Q_\mu = 0$).
- $30 \leq$ Visible Mass $(e\mu) \leq 80$

7.4 Analysis I (1.1 fb^{-1}) data

The two main sources of reducible backgrounds are represented by $Z+jets$ including heavy and light flavor jets and standard model WZ backgrounds. In both cases a

real Z decaying to a pair of electrons or muons is detected and another lepton is produced from the W, bottom quark, or charm quark decay. In addition to a real lepton there is also a possibility for a charged pion or photon to fake an electron or an electron, muon, quark, or gluon jet to fake a hadronic tau. To provide a data-driven estimate of the reducible background contributions, a fake–rate method is used. The fake–rate is estimated in a similar way separately for each lepton flavor.

7.4.1 Background estimation via fake rate method

The probability for jets to fake taus is measured using events with a leading Z, after all selections cuts are applied, and a second leading Z decaying into $\tau\tau$, where no requirements on the τ isolation are applied and only decay mode is required. Additionally, we require that the taus have the same charge to ensure a background dominated sample. After this selection a ratio of number of τ's which pass HPS Loose or Medium discriminator to the initial number of τ's is measured as a function of the tau transverse momenta for each tau individually. This region is dominated by Z+Jets background with signal contamination less than 0.1%. The $2l2\tau$ visible invariant mass distributions measured in data are compared to predictions of the MC simulation in Fig. 7.2 for the control regions. The MC simulation describes the data well.

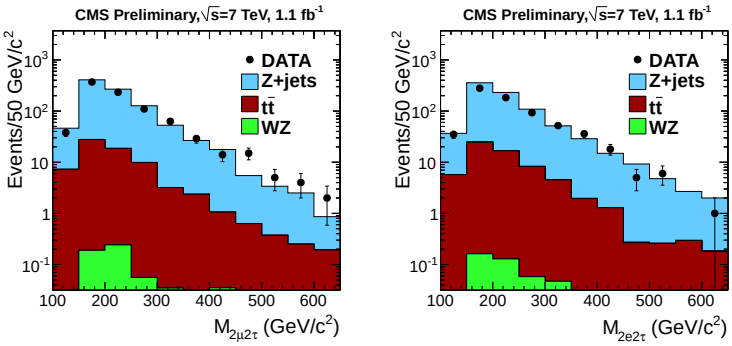

Figure 7.2: Comparison between data and simulation for the background estimation in same-sign, anti-isolated control region for $\mu\mu\tau\tau$ (left plot) and $ee\tau\tau$ (right plot) events.

The measured fake–rates are then fit as

$$F(p_T(\tau)) = C_0 + C_1 e^{C_2 p_T(\tau)} \tag{7.1}$$

Results of the fit can be seen in Fig. 7.3 for the Loose and Medium HPS working points. To perform the background estimate the fit results are applied to the signal sample (N_S), which is selected using all final cuts and an anti-isolation requirement on the tau isolation (to avoid contribution of real events). The background contribution (N_B) is estimated as follows.

For semi-leptonic channels:

$$N_B = \frac{N_S F(p_T(\tau))}{1 - F(p_T(\tau))} \tag{7.2}$$

For fully hadronic channels:

$$N_B = \frac{N_S F(p_T(\tau_1)) F(p_T(\tau_2))}{1 - F(p_T(\tau_1)) F(p_T(\tau_2))} \tag{7.3}$$

Data derived estimation values are shown in Table 7.2

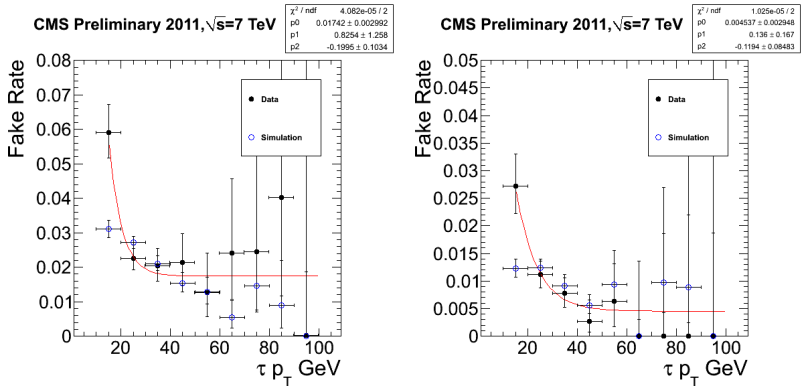

Figure 7.3: Fake rate as a function of τp_T for the Loose (Left) and Medium (right) τ working points with the resulting fit overlaid. Results are from 1.1 fb^{-1} of data and compared with simulated background.

In the channels where no hadronic tau are involved a similar fake–rate method is used. A background dominated sample is selected by applying all cuts except the isolation on inspected leptons and again requiring the leptons from the second Z to have the same charge. After selecting the background dominated sample, a p_T independent fake rate (F) is calculated separately for electrons and muons. This

fake rate is then applied to the signal like region (N_S), selected by applying all cuts except the leptons coming from the second leading Z are anti-isolated. The number of expected background events (N_B) is then calculated as follows.

$$N_B = \frac{N_S F_1 F_2}{1 - F_1 F_2} \tag{7.4}$$

Where F_1 and F_2 correspond to the fake rate for 1st and 2nd lepton from the second leading Z respectively. Table 7.2 shows the results of reducible background estimation via fake rate method.

Channel	Estimated Events
$ee\tau\tau$	0.084 ± 0.004
$\mu\mu\tau\tau$	0.066 ± 0.004
$eee\tau$	0.24 ± 0.07
$\mu\mu e\tau$	0.12 ± 0.05
$ee\mu\tau$	0.07 ± 0.04
$\mu\mu\mu\tau$	0.05 ± 0.03
$\mu\mu e\mu$	0.12 ± 0.09
$eee\mu$	0.06 ± 0.05

Table 7.2: Results from data driven background estimate with 1.1 fb^{-1}

The estimation of the ZZ contribution is based on a comparison to the well–measured inclusive Z production cross-section. The number of expected ZZ events can be written as

$$N_{ZZ}^{\text{estimated}} = N_Z^{\text{obs}} \cdot \frac{\sigma_{ZZ}^{\text{SM}} \cdot A_{ZZ}}{\sigma_Z^{\text{SM}} \cdot A_Z}, \tag{7.5}$$

where N_Z^{obs} is the number of observed events via inclusive Z production, A_Z is the analysis acceptance estimated using a MC simulation and scaled with measured data/MC correction factors, and A_{ZZ} is the analysis acceptance for ZZ events.

Table 7.3 compares the estimated number of ZZ events in the signal region in 1.1 fb^{-1} of data with the expected numbers from the MC simulation.

7.4.2 Systematic uncertainties

The main sources of systematic uncertainties which are common for all final states are summarized in Table 7.4. The pp integrated luminosity has an uncertainty of 4.5%. Additional systematics affecting the reducible backgrounds come from the

	N_{ZZ}^{exp}	N_{ZZ}^{est}
$\mu\mu\tau\tau$	0.142 ± 0.013	0.145 ± 0.013
$\mu\mu\mu\tau$	0.281 ± 0.017	0.287 ± 0.018
$\mu\mu e\tau$	0.316 ± 0.019	0.323 ± 0.019
$\mu\mu e\mu$	0.179 ± 0.014	0.183 ± 0.014
$ee\tau\tau$	0.120 ± 0.012	0.125 ± 0.012
$ee\mu\tau$	0.240 ± 0.016	0.250 ± 0.017
$eee\tau$	0.311 ± 0.018	0.323 ± 0.019
$eee\mu$	0.157 ± 0.013	0.163 ± 0.014

Table 7.3: The ZZ background yield expected from the MC simulation compared to the yield obtained using the data–driven method.

limited statistics in the background control regions which propagates to the background evaluation in the signal region, and from the uncertainty on the extrapolation factor from the background control to the signal control regions and from the MC based closure tests.

In addition, in Table 7.5, systematic for e and μ reconstruction, identification, isolation and energy scales are quoted.

Source	Uncertainty
Luminosity	1.045
Higgs cross section	1.17–1.20
ZZ Background Estimate	1.1
Z+Jets/WZ Background Estimate	1.3

Table 7.4: Systematic uncertainties common to all channels.

Channel	e Reco / ID/ Iso	μ Reco / ID / Iso	τ ID	τ Energy Scale	Trigger
$\mu\mu\tau\tau$	-	1.01	1.10	1.04	1.01
$\mu\mu\mu\tau$	-	1.02	1.06	1.03	1.01
$\mu\mu e\tau$	1.06	1.01	1.06	1.03	1.01
$\mu\mu e\mu$	1.03	1.02	-	-	1.01
$ee\tau\tau$	1.02	-	1.10	1.04	1.01
$ee\mu\tau$	1.02	1.01	1.06	1.03	1.01
$eee\tau$	1.06	-	1.06	1.03	1.01
$eee\mu$	1.04	1.02	-	-	1.01

Table 7.5: Channel specific systematic uncertainties

7.4.3 Results

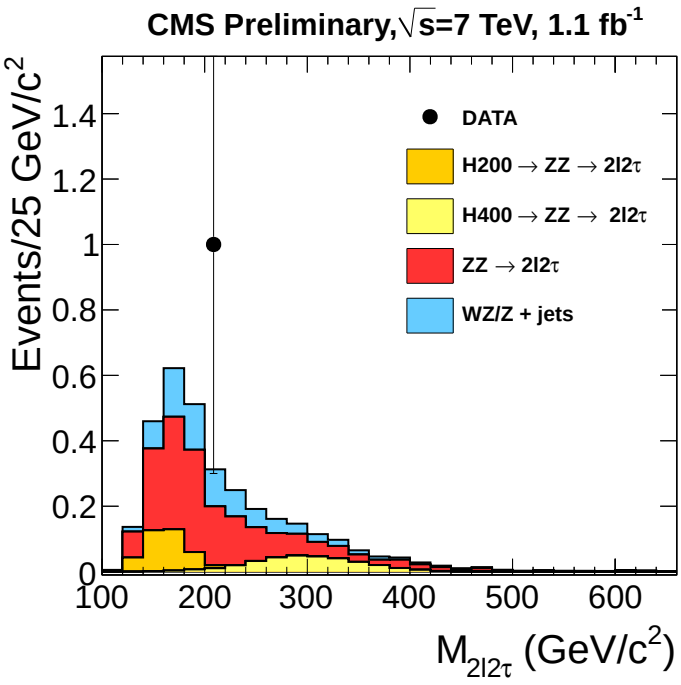

Figure 7.4: The four-lepton reconstructed mass summed for all $2l2\tau$ final states. The data corresponds to an integrated luminosity of 1.1 fb^{-1}. The stacked background shapes are the result of a MC simulation and normalized to the values obtained using the data–driven techniques. The expected mass distributions for the SM Higgs boson with a mass 200 or 400 GeV are also shown.

In 1.1 fb^{-1} of data one event is observed in the 8 search channels, in the $eee\tau$ state, as compared to 2.5 background events expected. The reconstructed $2l2\tau$ invariant mass for all decay channels is presented in Fig. 7.4. The stacked background shapes are simulated using MC and normalized to the values obtained using the method outlined above. The expected mass distributions for the SM Higgs with masses $m_H = 200$ and $m_H = 400 GeV/c^2$ are also shown. No evidence is found for a significant deviation from standard model expectations anywhere in the ZZ mass range considered in this analysis.

Expected and observed 95% C.L. (Confidence Level) upper limits are set on the cross-section ratio to the nominal SM Higgs cross-section using the modified frequentist construction CL$_s$ [61, 62]. They are presented as a function of the Higgs

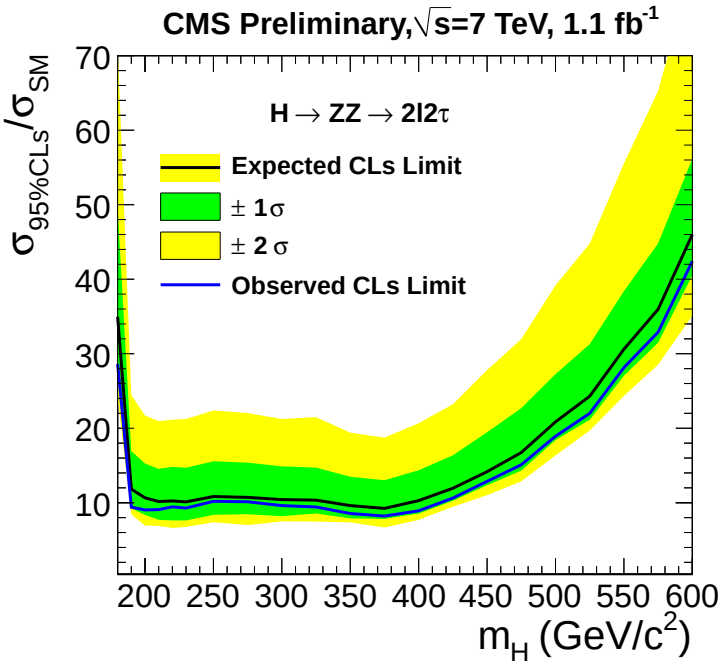

Figure 7.5: The expected and observed 95% C.L. upper limits on the product of the production cross-section and decay branching ratio for a Higgs boson normalized to the SM expectation as a function of m_{H}. At $m_{\mathrm{H}} = 180\,\mathrm{GeV/c^2}$ the less restrictive limit is due to the rapidly decreasing branching ratio of $H \to ZZ$ in the SM. The one- and two-standard-deviation ranges are also shown.

mass in Fig. 7.5. The bands represent the 1σ and 2σ probability intervals around the expected limit. The upper limit on the cross-section is approximately ten times the SM cross-section in the range of $200 < m_{\mathrm{H}} < 400\,\mathrm{GeV/c^2}$.

7.5 Analysis II (2.1 fb^{-1}) data

In this part, the updated results of this analysis with 2.1 fb^{-1} is presented with the same selection as discussed in 7.3 and same MC with added more data. The systematics is the same as pointed out in section 7.4.2. The main difference in this update is usage of a different method for estimating the reducible background. At the end expected and observed 95% C.L. (confidence level) upper limits are set on the cross-section ratio to the nominal SM Higgs cross-section using the MarkovChainMC method(which computes a Bayesian limit).

7.5.1 Reducible background estimation via ABCD method

This method aims to estimate the amount of reducible backgrounds (ZJet, WZ and TTbar) by dividing the phase space of the isolation of the two objects from decaying subleading Z into four regions, after applying all other cuts. One region which is called "signal region" is where two objects pass the isolation where HZZ and ZZ samples dominates, while in the 3 other regions at least one of the objects fail the isolation and are called "background regions" which dominates by backgrounds.

For events with two hadronic tau in the final states the definition of the regions would be as following:

- region A where both taus are isolated. This region is dominated by signal and irreducible backgrounds; we want to account reducible background, such as Zjets,WZ and TTbar here.
- In region B and C one of the taus are isolated while the other is anti-isolated.
- In Region D, both taus are anti-isolated.

Figure 7.6 illustrates the visible mass distributions of $\mu\mu\tau\tau$ in regions B, C and D. One sees that indeed these regions are dominated by background.

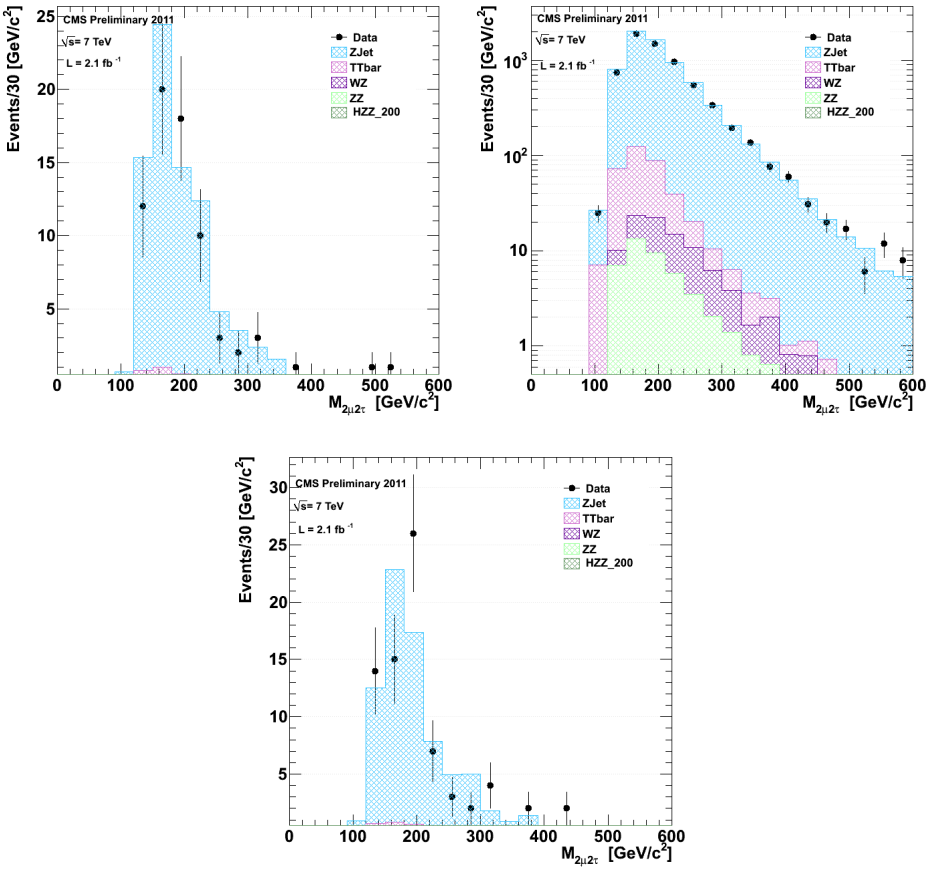

Figure 7.6: Visible mass distributions of $\mu\mu\tau\tau$ final states for the selections that all cuts has been applied and one of the taus are anti-isolated(top-left and bottom) and for the case where all cuts are applied and both taus are anti-isolated(top-right).

Provided that the isolation of the first tau does not correlate with the isolation of the second tau, the amount of reducible background events in the signal region A from the numbers of events observed in the other regions would be: $N_A = \frac{N_B \times N_C}{N_D}$.

Assuming that the charge of two leptons are not correlated with isolation, one can relax the cut on the charge of the subleading Z. Therefor regions B,C and D include both same-sign and opposite-sign ditaus and the estimated reducible formula changes to:

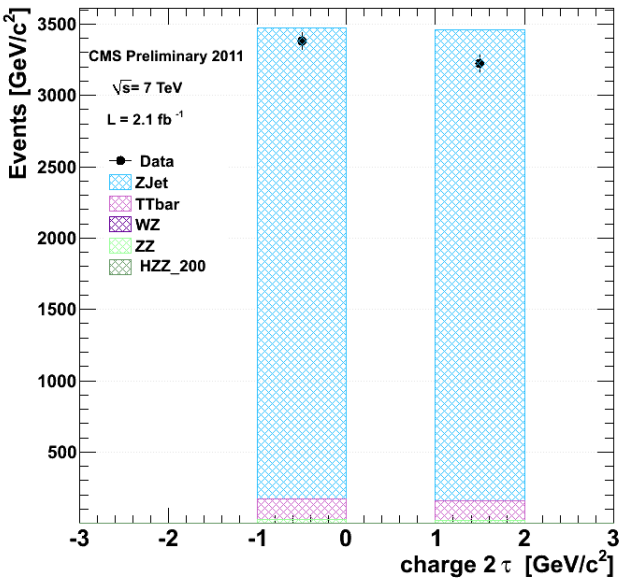

Figure 7.7: ditau charge in $\mu\mu\tau\tau$ channels where both taus are anti-isolated

$$N_A = (\frac{N_B \times N_C}{N_D}) \times (\frac{N_{OS}}{N_{OS} + N_{SS}})$$

where $N_{OS}(N_{SS})$ is the amount of opposite–sign(same–sign) for two taus in region D where statistics is higher than other regions. Relaxing the cut on ditau charge leads to more statistics in control regions B,C and D and consequently, decreases the statistical error on background estimation. Figure 7.7 shows the sign of ditau in the $\mu\mu\tau\tau$ final states where both taus fail the isolation. The charge ratio is about 0.50±0.01.

The same strategy is applied for $3l1\tau$ and $4l$ final state. In the former the four

regions are based on the isolation of the μ/e and τ_{had} while for the latter, the four regions are based on the isolation of the e and μ.

Channel	Expected events
$\mu\mu\tau\tau$	0.41 ± 0.07
$\mu\mu e\tau$	0.25 ± 0.08
$\mu\mu\mu\tau$	0.22 ± 0.09
$\mu\mu\mu e$	0.13 ± 0.13
$ee\tau\tau$	0.38 ± 0.07
$eee\tau$	0.16 ± 0.08
$ee\mu\tau$	0.16 ± 0.08
$eee\mu$	0.17 ± 0.17

Table 7.6: Results from data driven background estimate with 2.1 fb^{-1} using ABCD method

Table 7.6 shows the estimated number of reducible backgrounds for all 8 channels. Agreement between data and MC in shown in Figure 7.6.

7.5.2 Results

Five events(2 $\mu\mu\mu\tau$, 2 $eee\tau$ and 1 $\mu\mu\tau\tau$ final state) have been observed in 2.1 fb^{-1}. Comparing to 4.5 events in the MC shows no evidence for the significant deviation from standard model expectations anywhere in the ZZ mass range considered in this analysis. Table 7.7 shows the final yield of Higgs at Mass 200, irreducible ZZ background and the estimated irreducible background which were measured via data-driven method.

Channel	H200→ZZ	irreducible ZZ	reducible WZ/Z + jet BG
$\mu\mu\tau\tau$	0.080 ± 0.003	0.324 ± 0.031	0.41 ± 0.07
$\mu\mu e\tau$	0.137 ±0.004	0.431 ±0.036	0.25 ± 0.08
$\mu\mu\mu\tau$	0.115 ±0.004	0.493 ± 0.036	0.22 ± 0.09
$\mu\mu\mu e$	0.053 ± 0.003	0.202 ±0.025	0.13 ± 0.13
$ee\tau\tau$	0.065 ± 0.003	0.241 ± 0.027	0.38 ± 0.07
$eee\tau$	0.106 ± 0.004	0.398 ± 0.035	0.16 ± 0.08
$ee\mu\tau$	0.115 ± 0.004	0.388 ±0.034	0.16 ± 0.08
$eee\mu$	0.046 ± 0.002	0.182 ± 0.023	0.17 ± 0.17

Table 7.7: Final yield of H200, ZZ and WZ/Z+jet backgrounds. Value of signal and ZZ background are from MC simulation and the value of reducible backgrounds are from data using ABCD method

The reconstructed $2l2\tau$ visible mass for all decay channels is presented in Fig. 7.8.

The shape and yield of Signal(200 GeV) and ZZ background has been obtained from the MC while for the reducible background the shape of all channel is from data(where in the subleading Z both objects are anti-isolated and shapes are normalized to the estimated value from data-driven method, summarized in table 7.6).

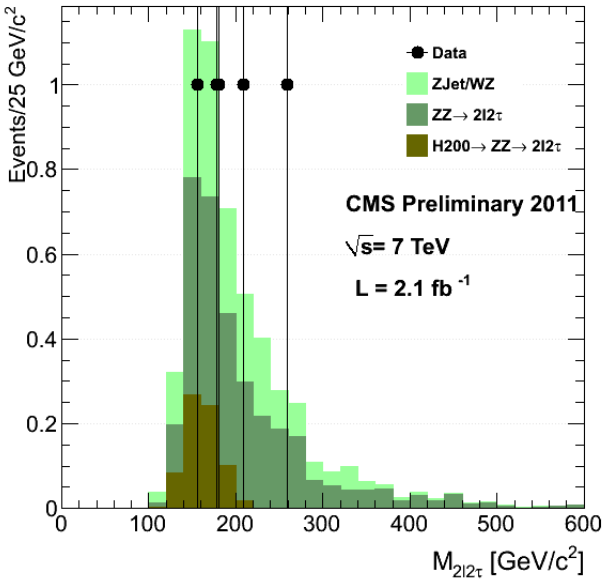

Figure 7.8: The four-lepton reconstructed mass summed for all $2l2\tau$ final states. The data corresponds to an integrated luminosity of 2.1 fb^{-1}. The shape and yield of Signal(200 GeV) and ZZ background has been obtained from the MC while for the reducible ZJet/WZ background the shape of all channel is from data (where in the subleading Z both objects are anti-isolated and shapes are normalized to the estimated value from data-driven method, summarized in table 7.6). Indeed there is a correlation between visible mass and isolation of the taus since fake rate depends on the tau p_T. However it has been shown that this effect on the final visible mass for all final states is small.

The final step before making the exclusion plot is to define a mass windows for each final states and each Higgs mass. This cut has a significant impact on reducing the ZZ irreducible background. Fig. 7.9 compares the shape of the HZZ at Mass 200 GeV with ZZ in four different final states. Cutting on the invariant mass aims to select HZZ events as significant as possible and suppressing ZZ background.

Figure 7.10 shows the mass window for all eight final states and for 10 different Higgs masses. As the Higgs with higher mass has a broader shape, their visible mass windows become larger. Approximately the window for each mass is within

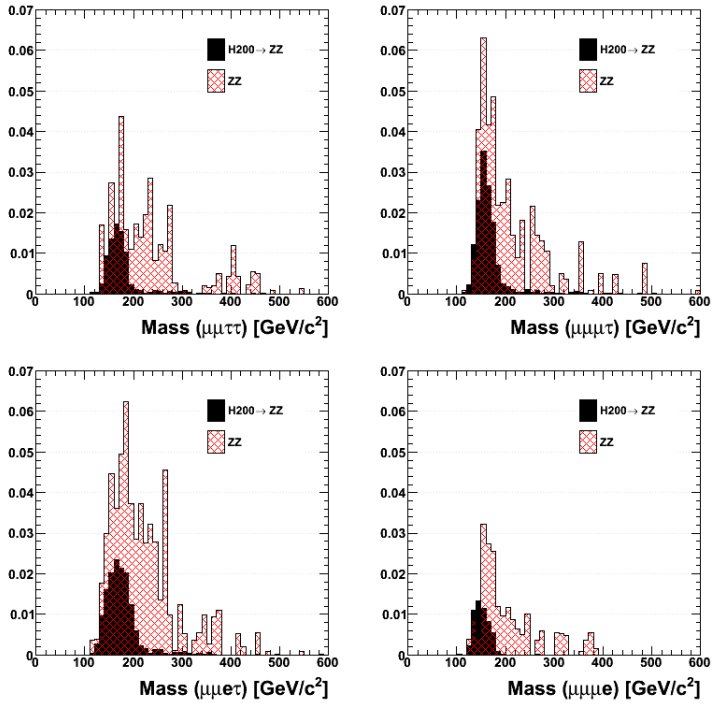

Figure 7.9: Shape of signal(H200) and ZZ background for four final states. The plots are normalized to 2.1 fb^{-1}

$(0.60 \pm 0.05, 1.00 \pm 0.05)$ times the mass of the Higgs.

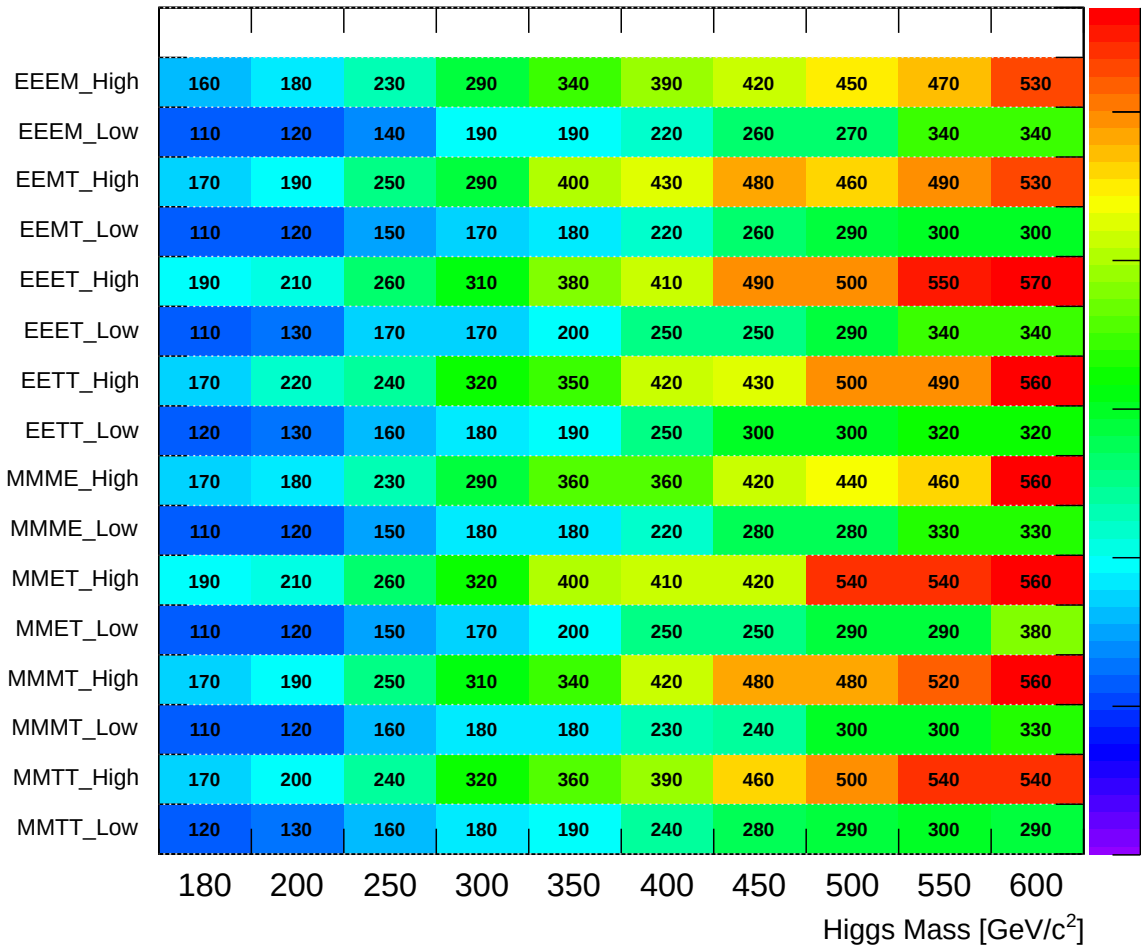

Figure 7.10: Different value of the mass windows for 8 final states and 10 Higgs mass points.

Expected and observed 95% C.L. (confidence level) upper limits are set on the cross-section ratio to the nominal SM Higgs cross-section using Markov-Chain MC method. They are presented as a function of the Higgs mass in Fig. 7.11. Similar to the analysis done with 1.1 fb^{-1}, the upper limit on the cross-section is approximately ten times the SM cross-section in the range of $200 < m_{\mathrm{H}} < 400\,\mathrm{GeV/c^2}$.

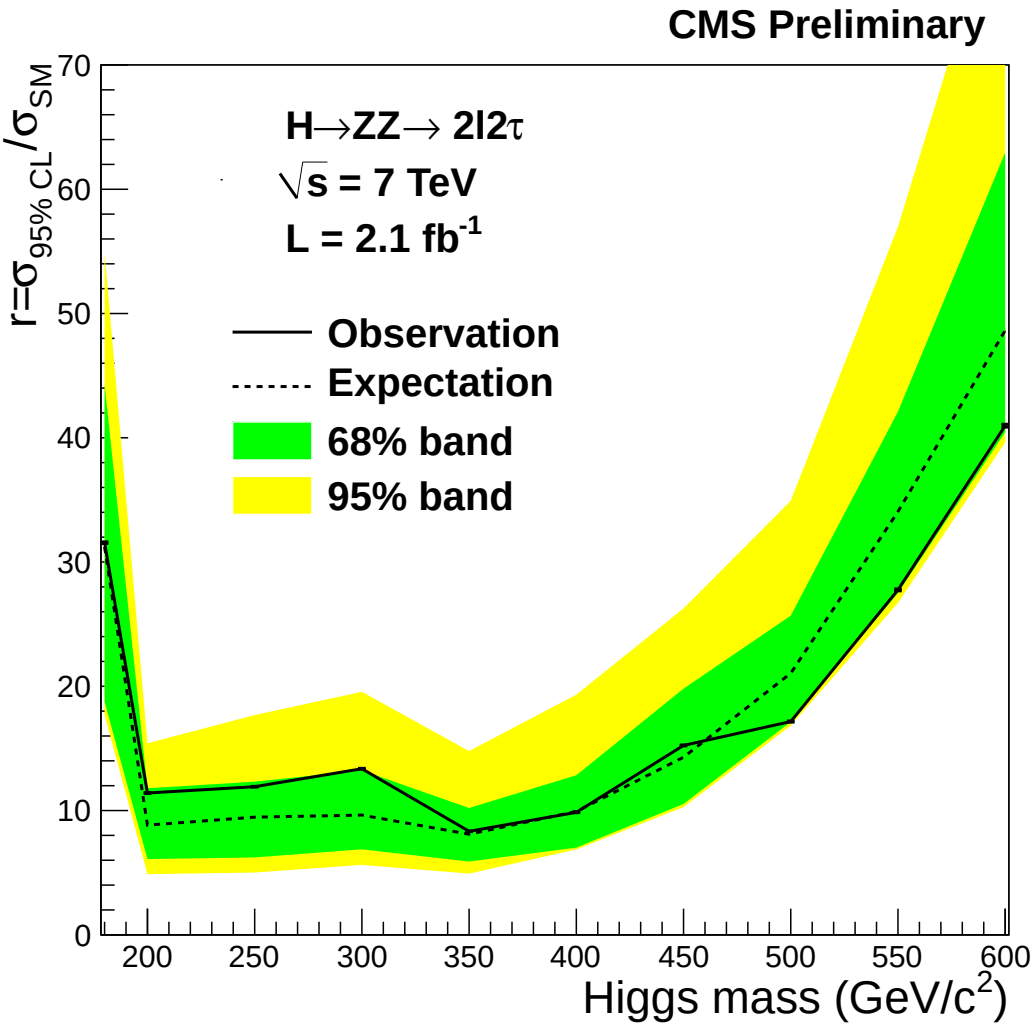

Figure 7.11: The expected and observed 95% C.L. upper limits on the product of the production cross-section and decay branching ratio for a Higgs boson normalized to the SM expectation as a function of m_{H}. The one- and two-standard-deviation ranges are also shown.

7.6 Conclusions

A search for the Standard Model Higgs in the decay mode $H \to ZZ \to 2l2\tau$ has been performed using CMS data corresponding to an integrated luminosity of 1.1 fb^{-1} and updated with 2.1 fb^{-1}. No evidence is found for a significant deviation from Standard Model expectations anywhere in the ZZ mass range considered in this analysis. An upper limit is placed at 95% C.L. on the product of the cross-section and decay branching ratio for a Higgs boson decaying with SM couplings, which excludes cross-sections of about ten times the expected value for Higgs masses in the range $200 < m_{\mathrm{H}} < 400\,\mathrm{GeV/c^2}$.

Bibliography

[1] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys.Lett.*, B565:61–75, 2003. (Cited on pages 2 and 24.)

[2] CDF and D0 Collaborations. Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.2 fb-1 of Data. 2011. (Cited on pages 2 and 26.)

[3] CMS Collaboration. Search for standard model higgs boson in pp collisions at $\sqrt{s} = 7$ tev. *CMS Physics Analysis Summary*, CMS-PAS-HIG-11-011, 2011. (Cited on pages 2 and 28.)

[4] ATLAS Collaboration. Combined standard model higgs boson searches in pp collisions at $\sqrt{s} = 7$ tev with the atlas experiment at the lhc. *ATLAS Conference Note*, ATLAS-CONF-2011-112, 2011. (Cited on pages 2 and 27.)

[5] CMS Collaboration. Search for a standard model higgs boson produced in the decay channel 4l. *CMS Physics Analysis Summary*, CMS-PAS-HIG-11-004, 2011. (Cited on pages 2 and 95.)

[6] David Griffiths. *Introduction to Elementary Particles*. WILEY-VCH, 2010. (Cited on page 5.)

[7] B.R.Martin. *Nuclear and Particle Physics*. John Wiley & Sons Ltd, 2009. (Cited on pages 6, 7, 8 and 9.)

[8] F.Mandle and G.Shaw. *Quantum Field Theory*. John Wiley & Sons, 1988. (Cited on pages 13 and 15.)

[9] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard Model Higgs-Boson Branching Ratios with Uncertainties. 2011. (Cited on pages 20 and 21.)

[10] Scott Willenbrock John D. Hobbs, Mark S. Neubauer. Tests of the standard electroweak model at the energy frontier. *arXiv:1003.5733v2*, 2010. (Cited on pages 22 and 23.)

[11] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, and the SLD Electroweak and Heavy Flavour Groups. Precision electroweak measurements and constraints on the Standard Model. 2010. (Cited on page 22.)

[12] David Smith. Higgs boson searches at lep ii. *JHEP*, Third Latin American Symposium on High Energy Physics. (Cited on page 24.)

[13] CMS Collaboration. CMS technical design report, volume I: Physics performance. *J. Phys.*, G34:995–1579, 2007. (Cited on pages 34, 73 and 77.)

[14] CMS Collaboration. CMS technical design report, volume II: Physics performance. *J. Phys.*, G34:995–1579, 2007. (Cited on pages 34 and 77.)

[15] Description and performance of the cms track reconstruction. *CMS AN*, , 2011. (Cited on page 37.)

[16] J.L.Agram. The silicon sensors for the compact muon solenoid tracker-design and qualification procedure. *Technical Report CMS-NOTE-2003-015,CERN,Geneva*, June 2003. (Cited on page 37.)

[17] Electromagnetic calorimeter commissioning and performance with 7 tev data. *CMS PAS*, EGM-10-002, 2010. (Cited on page 38.)

[18] Electromagnetic calorimeter commissioning and performance with 7 tev data. *CMS PAS*, EGM-10-003, 2010. (Cited on page 40.)

[19] Commissioning and performance of the cms hadronic calorimeters in proton collisions at a center of mass energy of 7 tev at the large hadron collider. *CMS AN*, , 2010. (Cited on page 40.)

[20] Muon reconstruction in the cms detector. *CMS AN*, , 2008. (Cited on page 42.)

[21] Letizia Lusito. Reconstruction and selection of $z \to \tau\tau \to \tau_h\tau_\mu$ decays at the cms experiment. *CMS Thesis Notes, UNIVERSITA DEGLI STUDI DI BARI*, 2010. (Cited on pages 45 and 47.)

[22] T. Sjostrand, S. Mrenna, and P.Z. Skands. Pythia 6.4 physics and manual. *JHEP*, 05, 2006. (Cited on pages 48 and 78.)

[23] Tim Stelzer Fabio Maltoni. Madevent: Automatic event generation with madgraph. *JHEP0302:027,2003*, 2003, 2003. (Cited on page 49.)

[24] S. Frixione, P. Nason, and C. Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *JHEP*, 0302:070, 2007. (Cited on pages 49 and 78.)

[25] S. Agostinelli et al. G4–a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. (Cited on pages 49 and 78.)

[26] Tracking and vertexing results from first collisions. *CMS PAS*, TRK-10-001, 2010. (Cited on pages 50 and 51.)

[27] Measurement of tracking efficiency. *CMS PAS*, TRK-10-002, 2010. (Cited on page 50.)

[28] Performance of muon identification in *pp* collisions at $\sqrt{s} = 7$ tev. *CMS PAS*, MUO-10-002, 2010. (Cited on pages 52 and 53.)

[29] Electron reconstruction and identification at $\sqrt{s} = 7$ tev. *CMS PAS*, EGM-10-004, 2010. (Cited on pages 53 and 55.)

[30] Jet performance in pp collisions at $\sqrt(s) = 7$ tev. *CMS PAS*, JME-10-003, 2010. (Cited on page 54.)

[31] CMS Collaboration. Commissioning of the particle-flow event reconstruction with the first lhc collisions recorded in the cms detector. *CMS Physics Analysis Summary*, CMS-PAS-PFT-10-001, 2010. (Cited on page 55.)

[32] Jet energy corrections determination at $\sqrt{s} = 7$ tev. *CMS PAS*, JME-10-010, 2010. (Cited on pages 55 and 90.)

[33] Met performance in pp collisions at $\sqrt(s) = 7$ tev. *CMS PAS*, JME-10-009, 2010. (Cited on pages 56 and 58.)

[34] Study of tau reconstruction algorithms using pp collision data collected at $\sqrt(s) = 7$ TeV. *CMS PAS*, PFT-10-004, 2010. (Cited on pages 59, 66 and 71.)

[35] The jet plus tracks algorithm. *CMS PAS*, JME-09-002, 2009. (Cited on page 61.)

[36] Tau reconstruction using the particle flow techniques. *CMS PAS*, PFT-08-001, 2008. (Cited on page 61.)

[37] Particle flow events reconstruction in cms and performance for jets, taus and met. *CMS PAS*, PFT-09-001, 2009. (Cited on page 61.)

[38] Particle Data Group Collaboration. Review of particle physisc. *Phys. Lett, B*, 667:1, 2008. (Cited on pages 62 and 91.)

[39] M. Bachtis, S. Dasu, and A. Savin. Prospects for measurement of $\sigma(pp \to z) \cdot b(z \to \tau^+\tau^-)$ with cms in pp collisions at $\sqrt(s) = 7$ tev. *CMS Note*, 2010/082, 2010. (Cited on page 63.)

[40] Performance of tau reconstruction algorithms in 2010 data collected with cms. *CMS Note*, TAU-11-001, 2011. (Cited on pages 65, 67, 84, 90 and 92.)

[41] CMS Collaboration. Measurements of inclusive w and z cross sections in pp collisions at $\sqrt{s} = 7$ tev. *CMS Note*, EWK-10-002, 2010. (Cited on pages 71, 83 and 92.)

[42] et all G.Bagleissi. Measurement of the $w \to \tau\nu$ cross–section in pp collisions at $\sqrt{s} = 7$ tev. *CMS Note, EWK-11-019*, 2011. (Cited on page 77.)

[43] CMS Collaboration. Measurement of the inclusive $z \to \tau^+\tau^-$ cross section in pp collisions at $\sqrt{s} = 7$ tev. *CMS Note*, EWK-10-013, 2010. (Cited on pages 77 and 79.)

[44] K Melinkov and F Petriello. Electroweak gauge boson production at hadron colliders through o(α^2). *Phys.Rev. D*, 74:114017, 2006. (Cited on page 78.)

[45] K Melinkov and F Petriello. The w boson production cross section at the lhc through o(α^2). *Phys.Rev.Lett.*, 96:231803, 2006. (Cited on page 78.)

[46] Crystal Ball collaboration. (Cited on page 83.)

[47] C. Seez and N. Rompotis. . (Cited on page 83.)

[48] Abdollah Mohammadi. Observation of $W \to \tau\nu$ production in PP Collisions with CMS detector. *arXiv:1108.1089*, 2003. (Cited on page 88.)

[49] PDF4LHC Working Group. http://www.hep.ucl.ac.uk/pdf4lhc/PDF4LHCrecom.pdf. (Cited on page 90.)

[50] CMS Collaboration. Absolute luminosity normalization. *CMS Detector Performance Summary*, DPS-2011-002, 2011. (Cited on page 90.)

[51] M. Bachtis et al. Performance of tau reconstruction algorithms in 2011 data collected with cms. *CMS Note*, 2011/239, 2011. (Cited on page 92.)

[52] Search for a standard model higgs boson produced in the decay channel $h \to zz \to 2l2\tau$ with cms detector at $\sqrt{s} = 7$ tev. *CMS PAS*, HIG-11-013, 2011. (Cited on page 95.)

[53] CMS Collaboration. Measurement of inclusive $z \to \tau\tau$ cross section in pp collisions at $\sqrt{s}$=7 tev. *submitted to JHEP*, EWK-10-013, 2011. (Cited on page 95.)

[54] CMS Collaboration. Performance of tau reconstruction algorithms in 2010 data collected with CMS. *CMS Physics Analysis Summary*, CMS-PAS-TAU-11-001, 2011. (Cited on page 95.)

[55] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO vector-boson production matched with shower in POWHEG. *JHEP*, 07:060, 2008. (Cited on page 96.)

[56] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. (Cited on page 96.)

[57] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. (Cited on page 96.)

[58] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006. (Cited on page 96.)

[59] Fabio Maltoni and Tim Stelzer. MadEvent: Automatic event generation with MadGraph. *JHEP*, 02:027, 2003. (Cited on page 96.)

[60] S. Jadach, Z. Wąs, R. Decker, and Johann H. Kuhn. The tau decay library TAUOLA: Version 2.4. *Comput. Phys. Commun.*, 76:361, 1993. (Cited on page 96.)

[61] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl.Instrum.Meth.*, A434:435–443, 1999. (Cited on page 105.)

[62] A. L. Read. Modified frequentist analysis of search results (the CLs method). *CERN Yellow Report*, CERN-2000-005:81, 2000. (Cited on page 105.)