A man ought to read
just as his inclination leads him;
for what he reads as a task
will do him little good.

SAMUEL JOHNSON

THESIS ENTITLED

“Search for the SM Higgs boson in the $H \rightarrow WW \rightarrow l\nu jj$ & Probe of WW Production in Vector Boson Fusion Topology in the CMS Experiment at the LHC”

Submitted to

DEPARTMENT OF PHYSICS AND ASTROPHYSICS
UNIVERSITY OF DELHI, DELHI-110007
INDIA

IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
BY

AJAY KUMAR

UNDER THE SUPERVISION OF
Dr. Kirti Ranjan

April 2015

Abstract

The Large Hadron Collider (LHC) at CERN is the world's largest and highest energy particle accelerator. It was built with the aim of allowing physicists to test the predictions of the Standard Model (SM) of particle physics, and particularly for the existence of the hypothesized Higgs boson; and to search for new particles predicted by theories beyond the SM. The Compact Muon Solenoid (CMS) is one of the main experiments at the LHC. The main objective of the CMS detector is to perform research in the experimental frontier of High Energy Physics, which is made possible because of the excellent running of LHC. This thesis aims to work on two fundamental problems of the SM. First, a search for the SM Higgs boson (H) is carried out in H decaying to WW in the semileptonic final state (H $\rightarrow$ WW $\rightarrow$ lνjj, with l being e or μ) with the CMS detector at the LHC. Further, a measurement of WW production in vector boson fusion (VBF) topology is also carried out with the CMS detector at the LHC. Both these analyses are performed using data recorded at $\sqrt{s} = 8$ TeV in 2012 corresponding to an integrated luminosity of 19.3 fb^{-1}. The decay channel, H $\rightarrow$ WW $\rightarrow$ lνjj, is sensitive to higher Higgs boson masses ($\geq$ 200 GeV) and is one of the leading channels in the CMS experiment for search of the SM Higgs boson in the higher mass range. The analysis selects events with one well-identified lepton, large missing energy and at least two high transverse momentum jets. A multivariate discriminant is designed in order to control the background while preserving, as much as possible, the signal component. The main background contaminating the signal region is $W + jets$, which has much higher production cross-section than that of the Higgs boson, and hence it is a major experimental challenge. This analysis imposes limits on the SM Higgs cross section in the high mass region of the Higgs boson, as well as on the heavy Higgs boson coupling and branching ratios in the BSM framework. This thesis also reports a first measurement on vector boson scattering in the semileptonic decay mode of the $WWWW$ interaction. The final states consists of a lepton, missing energy, two central jets and two forward jets. A special jet selection technique is developed with emphasis on having higher signal efficiency. We extract the VBF signal yield from an unbinned maximum likelihood fit to the dijet invariant mass distribution in the data. This work can serve as a baseline for future WW scattering analysis.

Dedication

Dedicated to my loving mother
and to everyone who supported,
encouraged and laughed with
me along the way...

THANKS

Declaration

This thesis presents a search for the SM Higgs boson in the $H \to WW \to \ell\nu$jj and probe of WW Production in Vector Boson Fusion Topology in the CMS Experiment at the LHC.

The work presented in this thesis has been published. This work has been done under the supervision of Dr. Kirti Ranjan. This work has been done by the candidate himself and to the best of his knowledge, no part of this work has earlier been submitted for any degree or diploma of this or any other university.

(AJAY KUMAR)
Candidate

(Dr. KIRTI RANJAN)
Supervisor

Prof. AMITABHA MUKHERJEE
Head

Department of Physics & Astrophysics

University of Delhi

Delhi 110 007

Department of Physics & Astrophysics
University of Delhi
Delhi - 110 007, India

Certficate of Originality

The research work embodied in this thesis entitled **"Search for the SM Higgs boson in the $H \rightarrow WW \rightarrow \ell\nu$jj and probe of WW Production in Vector Boson Fusion Topology in the CMS Experiment at the LHC"** has been carried out by me at the Department of Physics & Astrophysics, University of Delhi, Delhi, India. The manuscript has been subjected to plagiarism check by Turnitin.com software. The work submitted for consideration of award of Ph.D. is original.

AJAY KUMAR

Name and Signature of the Candidate

Acknowledgements

At the end of a journey as graduate student which has been at once romantic, tragic and heroic, I want to thank a lot of people who not only influenced me as person but have given direction to my work and my life. I feel very indebted and obliged to a number of people who start from my Supervisor Dr. Kirti Ranajn. Sir, It's been great great privilege to work under you. I feel that world has changed since i joined as a Ph.D student and so has I. And I feel extremely happy that I have people like you around me who shaped my path, guided my thoughts, gave a vision and most importantly introduced to real world -The Particle Physics. I really admire your dedication, enthusiasm and interest towards physics.

Next, I want to thank Prof. R. K. Shivpuri for being so kind, helpful and motivating. Thanks a ton sir, I owe you gratitude for all the support which was always there whenever i needed.

My special thanks goes to Dr. Ashutosh Bhardwaj for being so friendly and cooperating. He is so calm all the time and have been always available for me whenever needed. My sincere thanks goes to Prof. Brajesh Choudhary, Dr. Mohd. Naimuddin, Dr. Ashok Kumar for being wonderful people they are.

I have been fortunate to have an international exposure which allowed me to collaborate with people from all around the world. I visited CERN and Fermilab many times and interacted and worked with many people, many of them became my friend and always will be.

Many thanks to my Ph.D. advisory committee, Prof. D. S. Kulshrestha and Dr. Mohd. Naimuddin for giving me valuable suggestions time to time. I am also grateful to Prof. Amitabha Mukharjee, Head, Physics Department for providing me adequate facilities to work in the department and helping me at difficult times. I know him since my M.Sc. times, and he is truly a gentleman and helpful.

I owe sincere and earnest thankfulness to Kalanand Mishra, Dan Green,Jeffrey Berryhill, Ilya Osipenkov, Phil Dudero, Jack Anderson, Jordon Damgov, Nhan Tran, ,Pratima Jindal, Fan, Wei, Andy, John, Ashish kumar,Patricia Teles and list goes on. They have been always helpful and approachable during my stay at Fermialb. I specially want to mention

Kalanand mishra for being wonderful human being and a leader he is. It was real pleasure working with you. Your door was always open for my academic or non-academic discussion. Ilya and Phil you both also have been really great. Thanks a lot for all discussion, I really enjoyed working with you guys.

I want to recollect and thank Meenakshi, Boaz and Rick for helping me out throughout whenever i needed at Fermilab. I thank to collaborator from other part of the world, people from Milano, CERN, Brazil and China. I specially want to thank Pietro Govoni for all the help and being available for me. Thanks you all for all the suggestions, discussion through out my PhD. I also want to thanks Marco Musich for helping me out through tracking and APE codes of CMSSW.

I acknowledgement my seniors and fellow colleagues at our research lab (CDRST Lab) at University of Delhi consisting of Arun saini, Ruchika N. Sharma, Sudha Ahuja, Arun Kumar, Shilpi Jain, Sandhya Jain, Pooja, Saxena, Shivali Malhotra, Varun Sharma, Daljeet Kaur, Rocky Bala Garg, Raman Khurana, Bhawna Gomber, Ramkrishna Sharma , Sumit Kesri, Geetika Jain, , Ankit Gaur, Aashiq Hussain, Purnendu Kumar, Mohd. Hasibuddin, Mohit Gola, Aman Phogat, Prajesh Sharma, Prabhjot, Swati Mishra,Vijayender Narwal for helping me out with their suggestions and creating a friendly working environment in the lab. I would like to specially thank Kavita Lalvani, Priyanka Phogat for all the last minute corrections. It has not been possible without your help.

My special thanks to my friends Shruti, Sudha, Vibhuti, Savita, Ayushi, Punit, Manoj, Atul, Ajay, Varun and never ending list. Sorry guys for not mentioning you all. I also thank friends at Fermilab Sonam, Richa, Pavan, Sadia, Titas, Sudha (again), Sumit, Amandeep, Navaneet, Sid, Shiva, Bibhuti and friends from PU, TIFR, SINP and many more. I value each of your friendship and always will be in your life that's my promise. I also would like to thank the "chor-sipahi" gang formed during last visit of CERN, Pooja, Ankita, Sonia, Jitu, Harsh and Ram, guys it was real pleasure meeting you all. Hope our friendship will grow for forever.

My thanks to the staff of CDRST lab, Yunus Jee, Sanjay Singh, Aamir and Puran Singh for their innumerable help. A special word of gratefulness to Mohd. Yunus who has helped me with several matters and he is always there whenever needed.

And at the end of this i would like to thanks my mother, my sisters, brothers, cousins and all children. I am really sorry for not making to many gathering but i love you all unconditionally. And I will always be like that, that's my promise.

I would also like to acknowledge and thank DAE-DST, UGC, Fermilab for providing me funding for various travels.

Ajay Kumar

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Mankind has often wondered about how universe occurred, how life on earth occurred, what are we made up of and so on. Early human wondered about various things related with Nature, like why sun rises everyday? Why there is day and night? etc. Ancient indians believed that everything on earth are made from five basic elements (pancha mahabhutas): earth, water, fire, air, and space. The early ideas were founded through more abstract, philosophical reasoning rather than experimentation and empirical observation and were represented only in a line of thought among many. However as time passed by, humans approach of looking at Nature changed, they turned more thoughtful and argumentative.

This idea that all matter is composed of elementary particles were put forth in the 6th century BC. In the 19th century, John Dalton, through his work on stoichiometry, inferred that each element of nature was composed of a single, unique type of particle, which were thought to be fundamental and were called as atoms. However, close to the end of 19th century, it was found that fundamental particles of nature were conglomerates of even smaller particles. Electron was the first one to make the splash; in 1896, J. J. Thomson through his cathode ray experiment claimed that electron was a negatively charged particle. It's charge was carefully measured by Robert Andrews Millikan and Harvey Fletcher in the oil drop experiment in 1909. Physicists theorized that negatively charged electrons are constituent part of "atoms", along with some positively-charged substance, which were confirmed later. Electron became the first elementary, truly fundamental particle discovered. In 1911, Ernest Rutherford discovered nucleus in his famous gold foil experiment. It was a ground breaking experiment, envisaging the structure of an atom which was supposed to made up of nucleus (having positively charged protons) and electrons outside of the nucleus. But, soon it was found that mass of heavy nucleus was roughly twice as that of protons. This was an anomaly which led to many studies and culminated in the discovery of neutron in 1932 by James Chadwick. By the mid-1960's, physicists were cognizant of the fact that - 'all matter are composed of the fundamental protons, neutrons, and electron' - was insufficient to explain the myriad of new particles being discovered. This gave birth to what we call as modern particle physics.

A brief overview of the standard model (SM) is given in this Chapter. Since this thesis deals with the search for the SM Higgs boson, hence, an emphasis is given on electroweak symmetry breaking and Higgs mechanism. Constraints on the Higgs boson mass arising from the theory is then discussed. Searches for the Higgs boson in various colliding beam experiments are also reported. Finally, Higgs boson production and decay at the LHC are described, followed by thesis organization.

1.1 The Standard Model: An Overview

The theories and discoveries by many physicists since 1900s led to a remarkable insight into the elementary structure of matter. Our best understanding of how these particles and forces are related to each other is encapsulated in the SM of particle physics. Within the SM, everything in the universe is made from a few basic building blocks called fundamental particles, governed by fundamental forces. The SM explains many seemingly unrelated and, in some cases, even puzzling phenomena. It has successfully explained almost all experimental results and precisely predicted a wide variety of phenomena. Through many experiments conducted in past 50 years, the SM has become a well-tested physics theory.

Within the SM, all matter around us is made of elementary particles (belonging to broader group called "fermions"). These particles occur in two basic types called quarks and leptons. Each group consists of six particles, which are related in pairs, or generations. The lightest and most stable particles make up the first generation, whereas the heavier and less stable particles belong to the second and third generations. All stable matter in the universe is made from particles that belong to the first generation; any heavier particles quickly decay to the next most stable level. Quarks also come in three different "colors" [1] and all observed particles are colorless combinations of quarks. "Color" charge plays a vital role in strong interaction, very similar to the role played by electric charge in electromagnetic interaction. Quarks, and three of the leptons, i.e. electron, muon and tau carry electric charge and have a sizeable mass ($\sim$ MeV to GeV), whereas the neutrinos are electrically neutral and have very little mass (<KeV). There are four fundamental forces at work in the universe: the strong force, the weak force, the electromagnetic force, and the gravitational force. They work over different ranges and have different strengths, however, the most familiar force in our everyday lives, gravity, is not part of the SM.

1.1.1 The SM as Gauge Theory

The mathematical discription of the SM is based on Quantum Filed Theory (QFT), which uses idea of symmetry structure to have space-time invariant theory. The words "gauge theory" refer to the most successful framework for the description of Nature, i.e. of matter

[1] A Quantum Number

and its interactions. A gauge theory is a kind of field theory in which the Lagrangian is invariant under a continuous group of local transformations. Two matter particles interact with each other through an exchange of a virtual gauge boson. Thus, each of the interaction in the SM has its associated gauge bosons.

The transformations between possible gauges described as gauge transformations form the symmetry group or the gauge group of the theory. Each such group is associated with a group of generators, and for each such generator there must be a vector field. These gauge fields are introduced in the Lagrangian to preserve 'gauge invariance' under local group transformation. When such a theory is quantized, the quanta associated with the gauge fields are called gauge bosons. If the symmetry group is non-commutative, the gauge theory is termed as non-abelian, the usual example is the Yang–Mills theory. The gauge symmetry group $SU(3)_C \times SU(2)_L \times U(1)_Y$ [2], [3] represents the SM. $SU(3)_C$ is symmetry group of strong interaction which corresponds to Quantum Chromodynamics (QCD) part of the SM, whereas $SU(2)_L \times U(1)_Y$ represents unified electromagnetic and weak interactions part of the SM, called as Electroweak (EW) theory [1–3]. The theory is commonly viewed as containing the fundamental set of particles – the leptons, quarks, gauge bosons and the Higgs particle. A Lagrangian controls the dynamics and kinematics of the SM theory. The SM Lagrangian ($\mathcal{L}_{SM}$) can be written as [4] :

$$\mathcal{L}_{SM} = \mathcal{L}_{YM} + \mathcal{L}_{fermi} + \mathcal{L}_H + \mathcal{L}_{Yukawa} \tag{1.1}$$

Where, $\mathcal{L}_{YM}$ is the Yang-Mills component, and it describes the dynamics of the gauge fields any their their self-interactions,

$$\mathcal{L}_{YM} = -\frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{1.2}$$

where

$$W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g_2\epsilon^{ijk}W^j_\mu W^k_\nu, \qquad i,j,k = 1,2,3, \tag{1.3}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.4}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc}G^b_\mu G^c_\nu, \qquad a,b,c = 1,..,8 \tag{1.5}$$

W^i_μ for the $SU(2)_L$ group, B_μ for $U(1)_Y$ of weak hypercharge and G^a_μ for the $SU(3)_C$ of color are the field strength tensors for the associated gauge fields. g_2 and g_s are gauge couplings associated with fields corresponding to weak and strong interactions respectively. ϵ^{abc} and f^{abc} are the structure constants of the groups $SU(2)$ and $SU(3)$ respectively.
The term $\mathcal{L}_{fermi}$[4] in equation (1.2) represents the interaction of the gauge fields with the fermions.

$$\mathcal{L}_{fermi} = \bar{\psi}_L iD_\mu\gamma^\mu\psi_L + \bar{\psi}_{\ell_R} iD_\mu\gamma^\mu\psi_{\ell_R} + \bar{\psi}_Q iD_\mu\gamma^\mu\psi_Q + \bar{\psi}_{u_R} iD_\mu\gamma^\mu\psi_{u_R} + \bar{\psi}_{d_R} iD_\mu\gamma^\mu\psi_{d_R} \tag{1.6}$$

[2] electromagnetic and weak interactions are represented by a single lie group.

[3] C represents color quantum number, L represents left-handedness, Y is weak hyper-charge.

[4] ψ and $\bar{\psi}$ are Dirac-fields (matter fields) and γ^μ are Dirac matrices for relativistic fermions.

where L $= (\nu_{\ell_L}, \ell_L)^T$ are left-handed $SU(2)$ doublets of charged leptons $\ell = e,\ \mu,\ \tau$ and neutrinos $\nu_\ell = \nu_e,\ \nu_\mu,\ \nu_\tau$; Q $= (u_L, d_L)^T$ are the left-handed $SU(2)$ doublets of up-type quarks $u = u,\ c,\ t$ and down-type quarks $d = d,\ s,\ b$; and ℓ_R, u_R, d_R are the respective right-handed $SU(2)$ singlets.
The $\mathcal{L}_{YM} + \mathcal{L}_{fermi}$ part of equation (1.2) is invariant under local $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge transformations for fermion and gauge fields. The coupling of matter fields ψ to the gauge fields is given by the covariant derivative.

$$D_\mu \psi = (\partial_\mu - i g_s T_a G^a_\mu - i g_2 T_i W^i_\mu - i g_1 \frac{Y}{2} B_\mu)\psi \tag{1.7}$$

The couplings between the fermion and gauge fields are given by $-g_i \bar{\psi} V_\mu \gamma^\mu \psi$. T_i, Y and T_a are the generators of the respective gauge groups, i.e. $T_i = \sigma_i/2$ (σ_i are the Pauli matrices) for the left-handed $SU(2)$ doublets and $T_i = 0$ for the right-handed singlets. $T_a = \lambda^a/2$ (λ^a are the Gell-Mann matrices) for $SU(3)$ quark triplets and $T_a = 0$ for the leptons. Y is the weak hypercharge. g_1 is gauge coupling associated with electromagnetic interaction.

The above Lagrangian does not include any mass term, i.e. the gauge and fermion fields are kept massless. However, if one tries to add the mass term for the gauge bosons, like $M_W^2 W_\mu W^\mu$, it violates the local $SU(2) \times U(1)$ gauge invariance. This is also true for fermions if one includes standard mass term like $-m_f \bar{\psi}_f \psi_f$ for the SM fermion f.
Thus, the mass terms for gauge bosons and fermions destroy the local $SU(2) \times U(1)$ gauge invariance. The unbroken symmetry implies masslessness for all fundamental particles. However, it is experimentally observed that fundamental particles (fermions), and W and Z bosons having finite masses. To address this problem, Peter Higgs, in 1964, coined the idea of mass generation through Higgs field. Robert Brout and Francois Englert independently explained the working of mass-generation mechanism. This was also performed independently by Dick Hagen, Gerald Guralnik and Tom Kibble. This mechanism is commonly referred to as "Higgs Mechanism" or "BEH Mecahnism".

1.2 The Higgs or the BEH mechanism

The choice of Higgs field is very important. In the SM, the Higgs field is chosen to be four-component scalar field that forms a complex doublet of the weak isospin SU(2) symmetry. This choice of field Φ is motivated by the fact there are four gauge bosons in EW theory and hence at least four degree of freedoms are required.

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix} \tag{1.8}$$

where, ϕ^i are four real scalar fields. Now, The corresponding Higgs field Lagrangian would be

$$\mathcal{L}_H = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \tag{1.9}$$

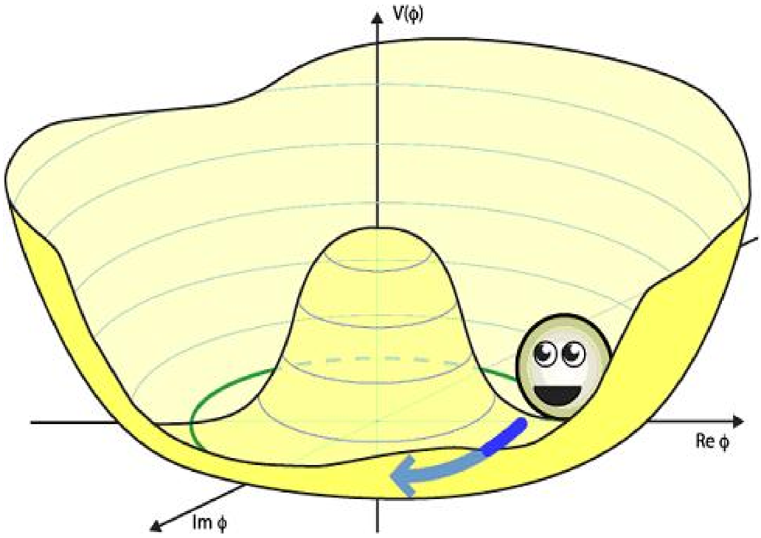


Figure 1.1: The Higgs potential $V(\Phi)$ for $\lambda > 0$ and $\mu^2 < 0$ condition.

The potential V(Φ) gives the self-interaction of the field.

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2 \tag{1.10}$$

Where λ is the self-coupling constant and μ is the mass related parameter. In equation (1.10), both the products $\Phi^\dagger\Phi$ and the potential $V(\Phi)$ are invariant under the local gauge transformations

$$\Phi(x) \rightarrow \Phi^{'}(x) = e^{i\alpha^i(x)\sigma^i/2}\Phi(x), \tag{1.11}$$

where σ^i are Pauli matrices and $\alpha^i(x)$ are transformation parameters. When $\lambda > 0$ and $\mu^2 > 0$, one gets a minima at zero whose symmetry can not be violated. That is not an interesting case for symmetry breaking. However, When $\lambda > 0$ and $\mu^2 < 0$ one obtains a continuous set of minima. The fields satisfy the condition:

$$\Phi^\dagger\Phi = \frac{1}{2}\Big((\phi^1)^2 + (\phi^2)^2 + (\phi^3)^2 + (\phi^4)^2\Big) = -\frac{\mu^2}{2}\lambda = v^2 \tag{1.12}$$

The potential for this case looks like a Mexican hat as shown in Figure 1.1. Once we choose the ground state to be a particular Φ among the many satisfying the equation of the minima, and then expand the potential in the vicinity of that minimum, the Lagrangian is transformed such that the original gauge symmetry is hidden or spontaneously broken. A minimum can be chosen to be: $\phi^1 = 0$, $\phi^2 = 0$, $\phi^4 = 0$, and $(\phi^3)^2$ is given by, $(\phi^3)^2 = -\frac{\mu^2}{\lambda} = v^2$. Expanding $\Phi(x)$ about this particular vacuum:

$$\Phi^0 \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.13}$$

Using the gauge symmetry pattern of Equation (1.11) one can write the field Φ in the exponential form via four fields $\theta^{1,2,3}(x)$ and $h(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2}} e^{i\theta^a(x)\tau^a(x)/v} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.14}$$

Using appropriate gauge transformation of the field, one "gauges away" three θ^a fields, chooses only one direction, and leaves only one invariant in the Lagrangian.

In the SM Lagrangian, the mass terms appears like $M^2\Phi^\dagger\Phi$, by inserting the vacuum expectation value, i.e. by choosing a minimum out of many possible states, one is able to study the structure of mass term. The covariant derivative chosen to preserve gauge invariance for the EW part is defined by:

$$D_\mu\Phi = \left(\partial_\mu - ig_2 W_\mu^i \frac{T^i}{2} - ig_1 \frac{Y}{2} B_\mu\right)\Phi \tag{1.15}$$

Using vacuum expectation value and covariant chosen above (Equation (1.15)) in Equation (1.9), the kinetic term can be written as:

$$|D_\mu\Phi|^2 = \left|\left(\partial_\mu - ig_2 W_\mu^a \frac{\tau^a}{2} - ig_1 \frac{Y_H}{2} B_\mu\right)\Phi\right|^2 \tag{1.16}$$

$$= \frac{1}{2}(\partial_\mu h)^2 + \frac{g_2^2}{8}(v+h)^2(W_1^\mu + iW_2^\mu)(W_\mu^1 - iW_\mu^2) + \frac{1}{8}(v+h)^2(g_2 W_3^\mu - g_1 Y_H B^\mu)^2$$

$$= \frac{g_2^2 v^2}{8}(W_1^\mu + iW_2^\mu)(W_\mu^1 - iW_\mu^2) + \frac{v^2}{8}(g_2 W_3^\mu - g_1 Y_H B^\mu)^2 + \frac{1}{2}(\partial_\mu h)^2 + ... \tag{1.17}$$

The first term is proportional to the mass term $M_W^2 W_\mu^+ W^{-\mu}$ for any charged gauge boson field

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad with \quad M_W = \frac{1}{2} v g_2 \tag{1.18}$$

Essentially, the last relation fixes the vacuum expectation value v in terms of the W boson mass M_W and their coupling g_2. Using Fermi constant ($G_F = 1.667 \times 10^{-5}\text{GeV}^{-2}$) M_W becomes:

$$M_W = \frac{g_2 v}{2} = \left(\frac{\sqrt{2}g_2^2}{8G_F}\right)^2 \quad and \quad v = 2\frac{M_W}{g_2} = \frac{1}{(\sqrt{2}G_F)^{1/2}} \approx 246 \text{ GeV}. \tag{1.19}$$

Using Equation (1.17), the physical states or mass eigenstates become:

$$Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, \qquad A_\mu = \frac{g_2 W_\mu^3 + g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}} \tag{1.20}$$

This mixing can also be represented as:

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \tag{1.21}$$

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \tag{1.22}$$

Where θ_W is the Weinberg or weak mixing angle. Its value is observed to be $\sin^2\theta_W = 0.231$. So, the mass term $\frac{1}{2}M_Z^2 Z_\mu^+ Z^{-\mu}$ can be interpreted as:

$$M_Z = \frac{1}{2}v\sqrt{g_2^2 + g_1^2} \tag{1.23}$$

Also, it can be shown that the masses of W and Z bosons are connected through the following relation:

$$\cos\theta_W = \frac{M_W}{M_Z} \tag{1.24}$$

As desired, no mass term appears for neutral field A_μ, and thus the photon field (A_μ) remains massless.

The arrangement of the SM scalar Higgs field Φ in the complex SU(2) doublet (Equation (1.8)) allows one to construct $SU(2) \times U(1)$ invariant interaction of the Higgs field with fermions, being only the SU(2) doublets or singlets. The most general form of these interactions are expressed as:

$$\mathcal{L}_{Yukawa} = -\lambda_e \bar{\Psi}_L \psi_{\ell_R} \Phi - \lambda_u \bar{\Psi}_Q \psi_{u_R} \tilde{\Phi} - \lambda_d \bar{\Psi}_Q \psi_{d_R} \Phi + h.c. \tag{1.25}$$

where "h.c." means hermitian conjugate; and $\tilde{\Phi} = i\sigma^2\Phi^* = \left((\phi^0)^*, -\phi^-\right)^T$ denotes the charge-conjugate Higgs doublet with quantum numbers opposite to Φ and $\lambda' s$ are the coupling strengths of fermions with Higgs field. The terms are only shown for electron, u quark and d quark. Other leptons have same kind of terms as first term, and u-type and d-type quarks have similar terms as second and third terms respectively. Therefore, after the Electroweak symmetry breaking (EWSB) [5–10] when the Higgs field has obtained the non zero vacuum expectation value one can generate masses for all fermions of the SM through the interaction Lagrangian shown in Equation (1.25). If one considers the case of electron, i.e. the first term and its complex conjugate in Equation (1.25) then with the Higgs field one gets:

$$\begin{aligned}\mathcal{L}_e &= -\lambda_e(\bar{e}_L e_{\ell_R}\Phi + \Phi^\dagger e_{\ell_R} e_L) \\ &= -\frac{\lambda_e v}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) - \frac{\lambda_e}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L)h\end{aligned} \tag{1.26}$$

The first term in Equation (1.26) looks exactly as a mass term for fermions $-m_f\bar{\psi}\psi$, with the electron mass given as:

$$m_e = \frac{\lambda_e v}{\sqrt{2}}, \tag{1.27}$$

The kinetic part of the Higgs field, $\frac{1}{2}(\partial_\mu h)^2$, comes from the covariant derivative $|D_\mu\Phi|^2$ (the last term in Equation (1.17)), while the Higgs mass and Higgs self-interaction contributions parts, come from the scalar potential $V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$. Finally, with the equation $\mu^2 = -v^2\lambda$ the pure SM Higgs Lagrangian is given by

$$\mathcal{L}_H = \frac{1}{2}(\partial^\mu h)^2 - \frac{2\lambda v^2}{2}h^2 - \lambda v h^3 - \frac{\lambda}{4}h^4 + \frac{\lambda v^4}{4} \tag{1.28}$$

From this relation we can see that

$$M_H^2 = 2\lambda v^2 \tag{1.29}$$

The mass of the Higgs boson is generated by the self-coupling of Higgs boson and is defined by the parameter λ. As λ is a free parameter so does the Higgs boson mass. There are constraints on the Higgs boson mass from theory and experimental searches which are described in the next section.

1.2.1 Theoretical Constraints on Higgs Mass

The EWSB of the SM does not predict any particular value of mass for Higgs boson. It remains a free parameter in the SM but at the same time it can not assume an arbitrary value. The mass of the Higgs boson is dependent upon v and on λ. The latter cannot be determined other than measuring the Higgs mass itself. However, there are both theoretical and experimental constraints, including those from direct searches at colliders. Theoretical constraints include: unitarity, triviality and vacuum stability [11–21].

The scattering amplitudes for longitudinal gauge bosons swell as the square of the Higgs boson mass ($V_L V_L \to V_L V_L$) where V is W/Z boson. This can be calculated using the EW equivalence theorem, which is valid in the high energy limit (at very high energies ($s = Q^2 >> M_V^2$)); where M_V is the mass of vector boson and Q^2 is the square of four-momentum transfer in a process. Using the optical theorem from QFT and applying the unitarity requirements, it is given that:

$$|a_l|^2 = Re(a_l)^2 + Im(a_l)^2 = Im(a_l) \to |Re(a_l)| \leq \frac{1}{2} \tag{1.30}$$

Where, a_l is the partial wave amplitude for orbital angular momentum quantum number ℓ. By considering coupled channels like $W_L^+ W_L^- \to Z_L Z_L$ the bound can be found to be 710 GeV.

Further, taking totally different point of view, if it is assumed that there is no Higgs boson at all, in other words if $M_H^2 \gg s$, then Equation (1.30) gives indications on the critical scale $\sqrt{s_c}$ above which new physics should be expected. And, considering $W_L W_L \to W_L W_L$ scattering, it can be shown that:

$$a_0(\omega^+\omega^-) \to -\frac{s}{32\pi\nu^2} \tag{1.31}$$

from which, using Equation (1.30), we get:

$$\sqrt{s_c} < 1.8 \quad \text{TeV}, \tag{1.32}$$

Using more constraining channels the bound can be further reduced to $\sqrt{s_c} < 1.2$ TeV, This is very suggestive: it tells us that new physics must to be found around 1.2 TeV, i.e.

exactly in the range of energies that has been explored by the Tevatron and now by the LHC.

According to $\lambda\phi^4$ theory triviality argument, the dependence of the quartic coupling λ on the energy scale Q is regulated by the renormalization group equation which suggests that λ is directly proportional to energy. In the high energy regime, the theory becomes non-perturbative, since λ grows and it can remain perturbative only if λ is set to zero, i.e. only if the theory is assumed to be trivial. In the SM, λ goes as the M_H^2. Thus in the case when M_H becomes large, the equation for running of $\lambda(Q)$ can be written as:

$$\frac{d\lambda}{dt} = \frac{3\lambda^2}{4\pi^2} \tag{1.33}$$

where $t = \ln(Q^2/Q_0^2)$ and Q_0 is reference energy scale. $\lambda(Q)$ becomes infinite for $\lambda \to \infty$, the Landau pole. This is independent of the reference scale $\lambda(Q_0)$. Alternatively $\lambda(Q_0) \to 0$ for $Q \to 0$ with $\lambda(Q) > 0$. So, to assure the stability of the vacuum state of the theory it is required:

$$\frac{1}{\lambda(\Lambda)} > 0 \tag{1.34}$$

This is expressed as a bound on the Higgs boson mass as:

$$M_H^2 = \frac{8\pi^2 v^2}{3\log\left(\Lambda^2/v^2\right)} \tag{1.35}$$

The behaviour of M_H and Λ is shown in figure 1.2. By assuming the SM to be valid up to the scale of grand unified theories, $\Lambda \sim 10^{16}$ GeV, the approximate upper bound on the Higgs boson mass is [22]: $M_H < 160$ GeV. As the scale Λ becomes smaller the limit on the M_H becomes progressively weaker. Another bound is derived from the top quark Yukawa coupling which causes the scalar self-coupling to decrease, eventually driving it to negative if the Higgs boson mass is too low. In such situation, the value of the effective potential becomes negative, dropping beyond the SM vacuum. This is equivalent to the argument that λ remains positive at all scales of Λ, $\lambda(\Lambda) > 0$, because if λ becomes negative the potential has no state of minimum energy and the vacuum is not stable. It can be derived that if the SM is valid up to the scales of $\sim 10^{16}$ GeV one obtains:

$$M_H(GeV) = 130 + 2(m_t(GeV) - 170) \tag{1.36}$$

where m_t is the mass of top-quark. Which means that the mass of the Higgs boson is restricted between 126 to 160 GeV. It should be noted that this is exactly the mass range preferred by the EW precision measurements as discussed in section 1.3.2.

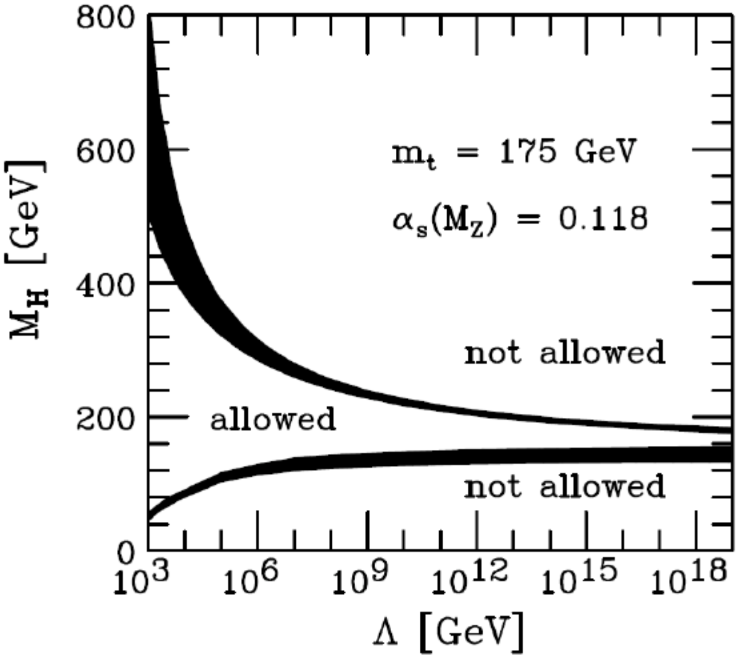


Figure 1.2: Upper and lower bounds on the SM Higgs boson mass from theoretical arguments as a function of the scale of new physics, Λ, The allowed region is between the curves [22].

1.3 Hunt for the SM Higgs boson

1.3.1 Direct searches

LEP1 and LEP2

The first ever extensive search for the SM Higgs boson was conducted at the Large Electron-Positron Collider (LEP1 and LEP2) at CERN in the 1990s. At the end of its Run in 2000, LEP had found no evidence for the SM Higgs boson and it gave a lower bound of 114.4 GeV on SM Higgs boson mass at 95 % C.L. (confidence level) [23]. The production of Higgs boson is dependent on centre of mass energy, type of beam and Higgs boson mass itself. LEP was a e^+e^- collider with energy reach of 209 GeV. At this energy the SM Higgs boson is expected to be produced mainly in association with the Z boson through the Higgsstrahlung process $e^+e^- \to HZ$. Small additional contributions are expected at the end of the kinematic range of the Higgsstrahlung process from W and Z boson fusion, which produces a Higgs boson and a pair of neutrinos or electrons or combinations, in the final state. For the Higgs boson masses which are relevant at LEP, the SM Higgs boson is expected to decay mainly into $b\bar{b}$ quark pairs (the branching ratio is 74% for a mass of 115 GeV) while decays to $\tau^+\tau^-$, WW^*, gg ($\approx$ 7% each), and $c\bar{c}$($\approx$ 4%) constitute the rest of the Higgs boson decay width (H_L^W).

Tevatron experiments

The search continued at Fermilab in the United States, where the Tevatron collider that discovered the top quark in 1995 had been upgraded for this purpose. Tevatron was a $p - \bar{p}$ collider with centre of mass energy reach of 1.96 TeV, and the dominant Higgs boson production modes were gluon-gluon fusion and associated productions, while decay modes with significant branching ratio were $b\bar{b} \sim 74\%$, $\tau^+\tau- \sim 10\%$ with WW and ZZ were at the tail of its high mass reach. The Tevatron was only able to exclude further ranges for the Higgs boson mass, and was shut down on 30th September 2011. The final analysis of the data excluded the possibility of a Higgs boson with a mass between 100 GeV to 103 GeV, and 147 GeV to 180 GeV at 95% C.L. [24]. In addition, there was a small (but not significant) excess of events possibly indicating a Higgs boson with a mass between 115 GeV and 140 GeV. The ratios of the 95% C.L. expected and observed limit to the SM cross section are shown in figure 1.3 for the combined CDF and DO analyses.

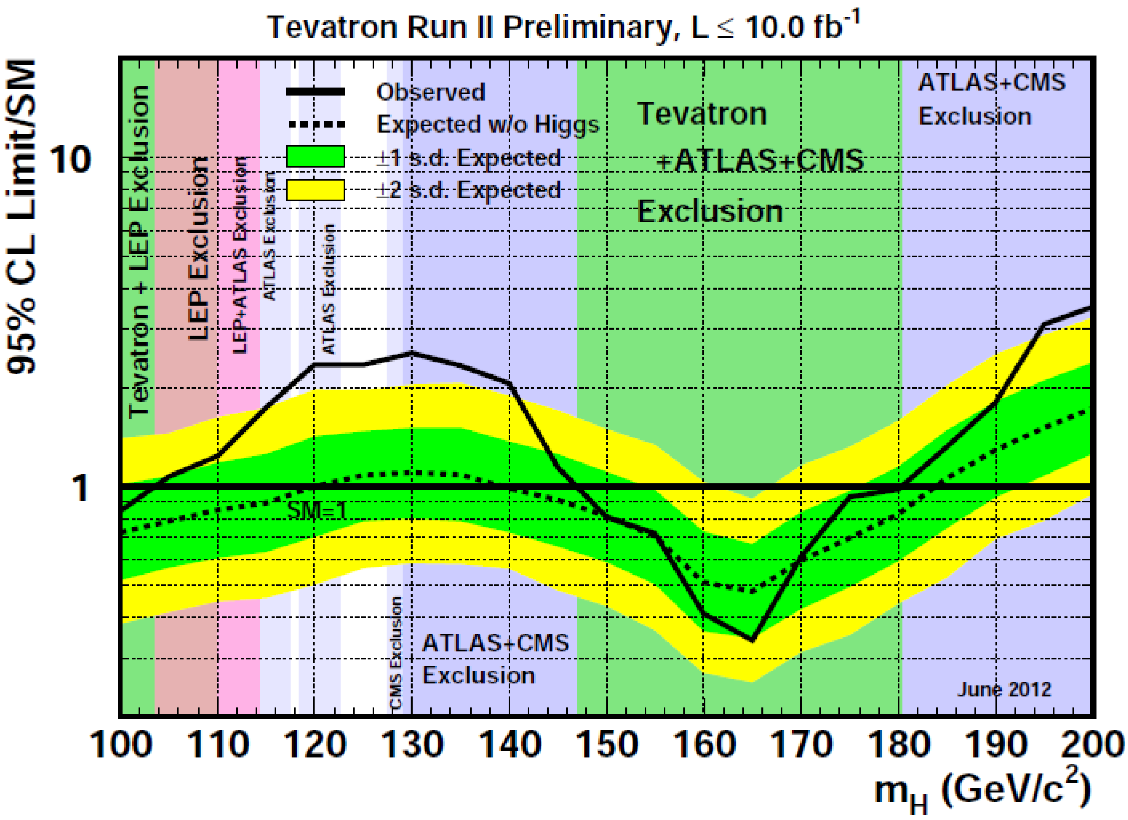


Figure 1.3: Observed as well as expected (median, for the background-only hypothesis) at 95% C.L. upper limits on the ratios to the SM cross-section, as a function of the Higgs boson mass for the combined CDF and DO analyses are shown. The limits are expressed as a multiple of the SM prediction for various test mass points at interval of 5 GeV for which both experiments have performed dedicated searches in different channels. The 68% and 95% bands are drawn where the limits can fluctuate, in the absence of signal. In the absence of signal, also shown are the regions excluded by LEP, ATLAS, and CMS. Plot is taken from Ref. [24].

The Large Hadron Collider (LHC) at CERN in Switzerland, is designed specifically to be able to either confirm or exclude the existence of the Higgs boson. It is described in detail in Chapter 2.

1.3.2 Indirect searches

At LEP collider, it was possible to measure masses of particles, in particular W boson and t (top quark) to a very high degree of precision. Thus, assuming that the SM Higgs boson exists, it can contribute to EW observables (like mass of $W^{\pm}$ boson via virtual corrections, particularly through loop diagrams shown in figure 1.4). Using, precision measurements of

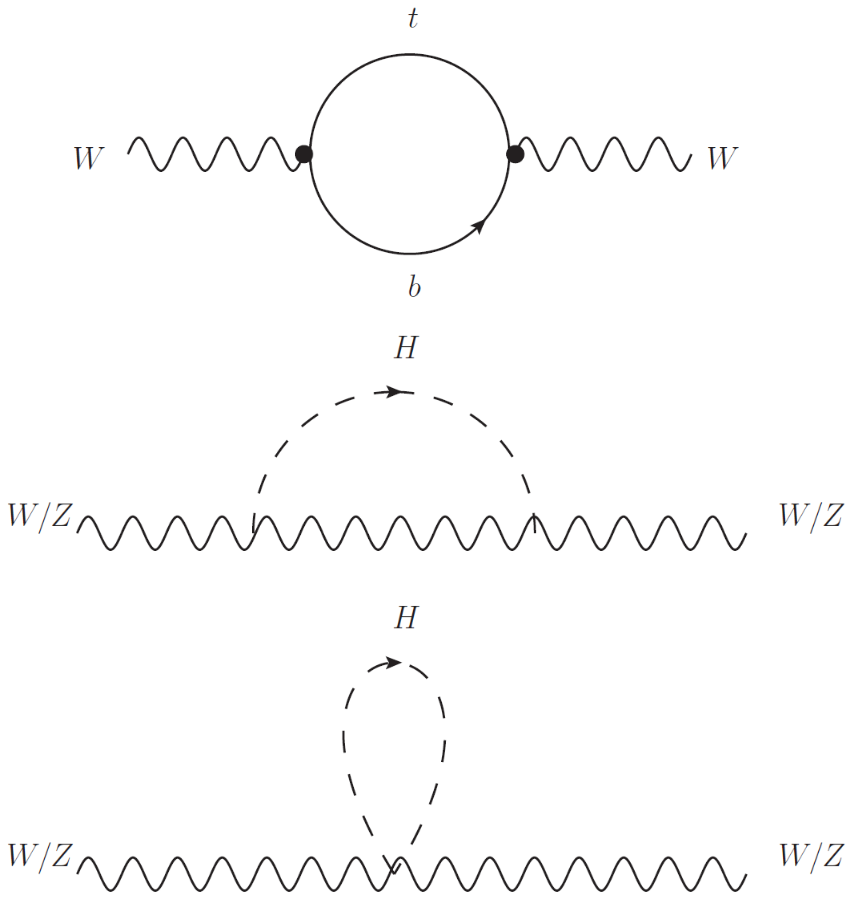


Figure 1.4: The masses of top quark, W and Higgs bosons are related via radiative loop corrections.

the W boson and top quark masses (shown in figure 1.5), and also using other observables measured at LEP and the Tevatron, LEP EW Working Group [25] tested the most prefered Higgs boson mass. Figure 1.6 shows the result of the combined Tevatron and LEP results in 2012, giving the following best fit value.

$$M_H = 94^{+29}_{-24}\ \text{GeV} \tag{1.37}$$

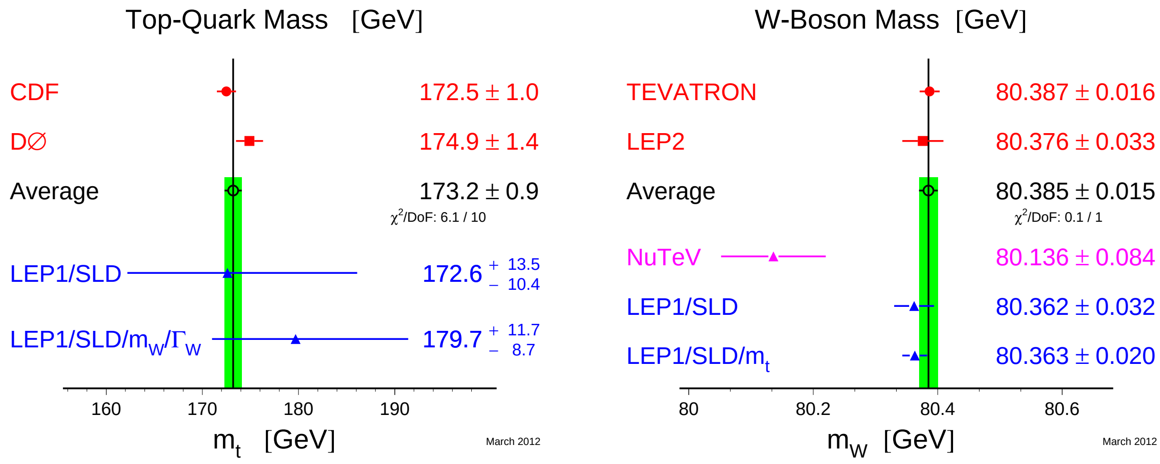


Figure 1.5: World average and individual measurements of the top quark m_t (left) and W boson mass m_w (right) including both direct measurements of the top quark mass as well as indirect measurements via higher order corrections to the W boson mass.

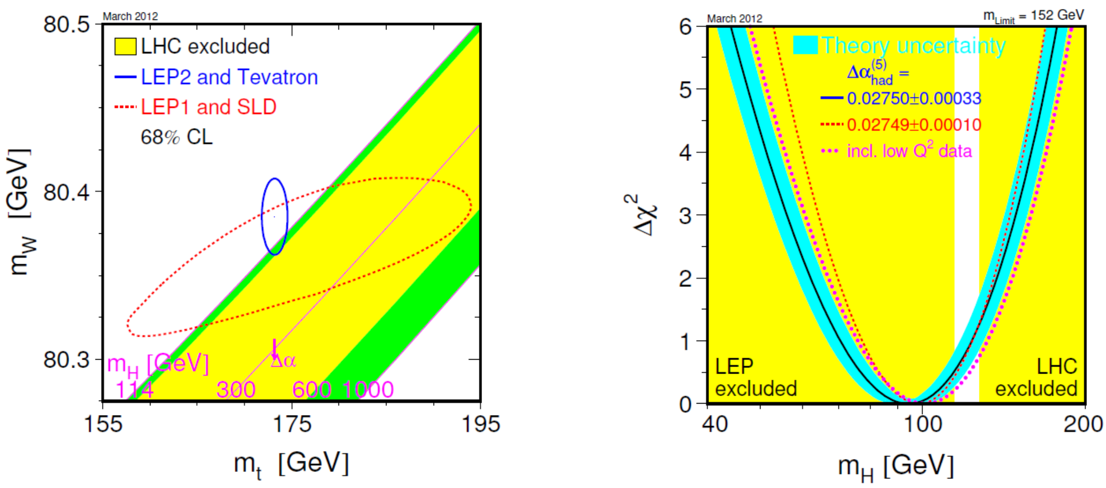


Figure 1.6: Combined results between m_t and M_W. The left plots shows the 68% C.L. for the indirect measurements of LEP1 and SLD (dashed ellipses) and direct observations from LEP2 and Tevatron (solid ellipses). The right plot shows the constraints from the EW fits on a potential SM Higgs boson mass as $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ vs. M_H curve.

1.4 Higgs Production at the LHC

The LHC is a $p-p$ collider where gluon density is very huge. Due to this the SM Higgs boson production through gluon-gluon fusion is most dominant. However, there are other SM Higgs boson production modes which are significant. There is vector boson fusion production which is an order in production cross-section lower that gluon-gluon fusion. Besides that Higgs boson can also be produced in association with W/Z and $t\bar{t}$ whose contributions are small. The variation of production cross-section as a function of the SM Higgs boson mass are shown for each process in figure 4.1 (Chapter 4) on left.

1.4.1 Gluon-Gluon Fusion

The leading order (LO) Feynman diagram for gluon-gluon fusion involves a quark loop shown in figure 1.7. Its interaction cross-section is proportional to the square of the number of heavy quarks. In the SM, the amplitude is dominated by the contribution from the top quark, the only one with mass scale comparable to the Higgs boson mass. The production cross-section is naturally sensitive to contributions from hypothetical particles with QCD interactions beyond the SM ones at higher energy scales. Higher order QCD

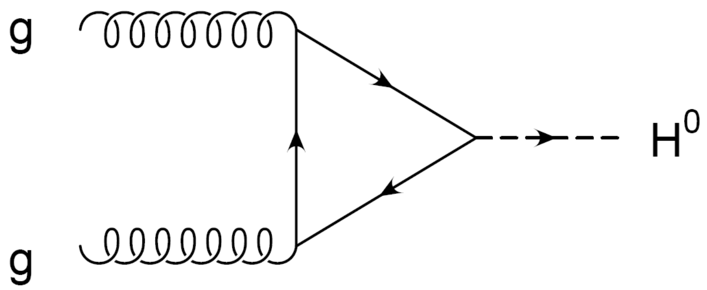


Figure 1.7: Leading order Feynman diagram for Higgs boson production through gluon-gluon fusion in QCD.

corrections to the process are finitely large, increasing the cross-section by about 100% at the next-to-leading order (NLO), 25% at next-to-next-to-leading order (NNLO), and a further 8% when leading logarithmic corrections are resummed (NNLL). The state of the art theoretical predictions also include NLO corrections from EW and mixed QCD-EW terms, which do not contribute significant. Because of the slow convergence of the QCD radiative corrections, the theoretical uncertainty on the cross-section from neglected higher order terms is 8% - 12% depending on the Higgs boson mass point even for the most precise NNLO + NNLL result. The theoretical uncertainties from the knowledge of the gluon parton density function are likewise significant, $\sim$ 8% up to Higgs boson masses below 250 GeV and it grows up to 10% for $M_H = 600$ GeV. As a natural consequence of the large QCD corrections, final states with one or more hadronic jets are not uncommon, introducing complications in the analyses performed in exclusive jet multiplicity final states.

1.4.2 Vector Boson Fusion

The Higgs boson production via vector boson fusion is important for high Higgs boson masses where the coupling to longitudinal polarised vector bosons would be strong. At higher energies where heavy Higgs boson particle can be produced the vector bosons act essentially as massless particles and can be treated as particles present inside the colliding protons. From experimental point of view, since there are two quarks coming directly from initial state quarks, these will carry very high energy and will be produced at large rapidity gap. This is a distinct signature and provides a handle to separate this process by tagging such quark initiated jets. The LO Feynman diagram is shown in figure 1.8.

For this process, QCD radiative corrections are significantly smaller, $O(10\%)$ at NLO [26]. As a consequence, the theoretical uncertainty from higher order terms on the most accurate prediction available, at NNLO accuracy for QCD effects and NLO for EW ones, is also smaller: below 0.5% for Higgs boson masses up to 200 GeV, and below 2% up to 600 GeV. As the parton density function for quarks is also known to higher accuracy than the corresponding one for gluons, the theoretical uncertainties from this source are also smaller, and it's range is from about 2.5% at low M_H (200 GeV) to about 7.5% at 600 GeV.

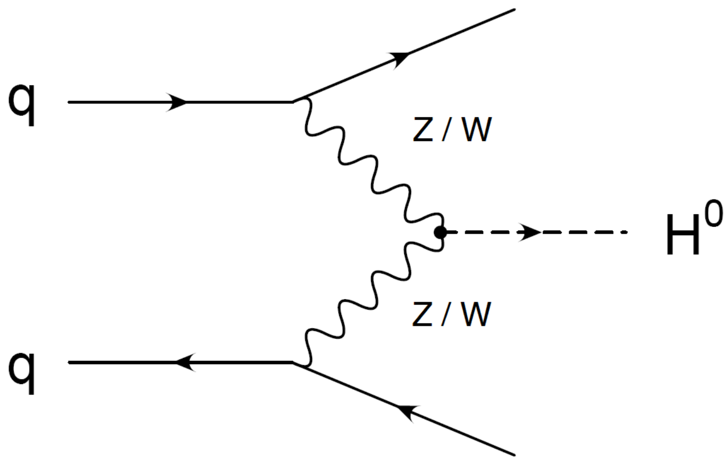


Figure 1.8: Leading order Feynman diagram for Higgs boson production through Vector Boson Fusion.

1.4.3 Associated Production

The LO Feynman diagrams for these modes are shown in figures 1.9 and 1.10. Although these processes have lower cross-section but still they have a very distinctive signature. The Higgs boson produced in association with either a top quark pair or a vector boson can have a very effective trigger for the Higgs boson events by requiring a high energy lepton from one of the top quarks or the vector boson decay. The next handle for the decay is to identify the jets with b-quarks. While the HW mode can in general have two b-quarks in the final state, the $Ht\bar{t}$ can have four b-quarks. The use of multiple b-tagging can provide a larger rejection of jets and thus counteract the lower cross section for $Ht\bar{t}$ production in the interesting mass range.

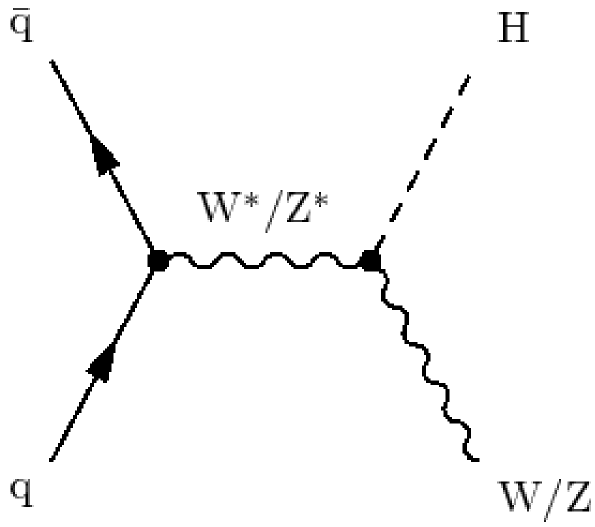


Figure 1.9: LO Feynman diagram for WH and ZH associated production.

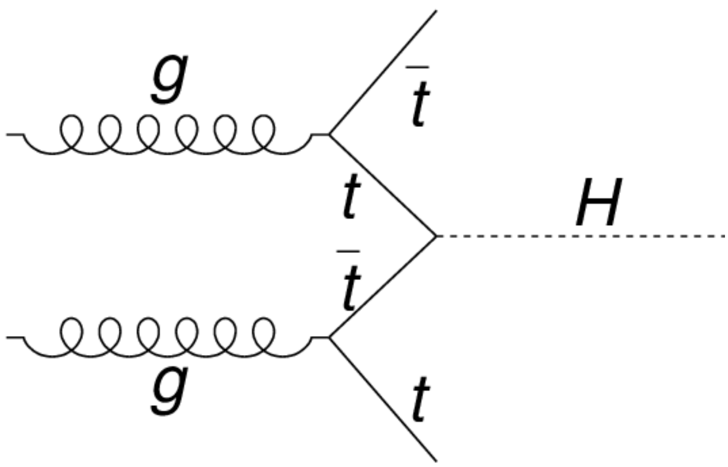


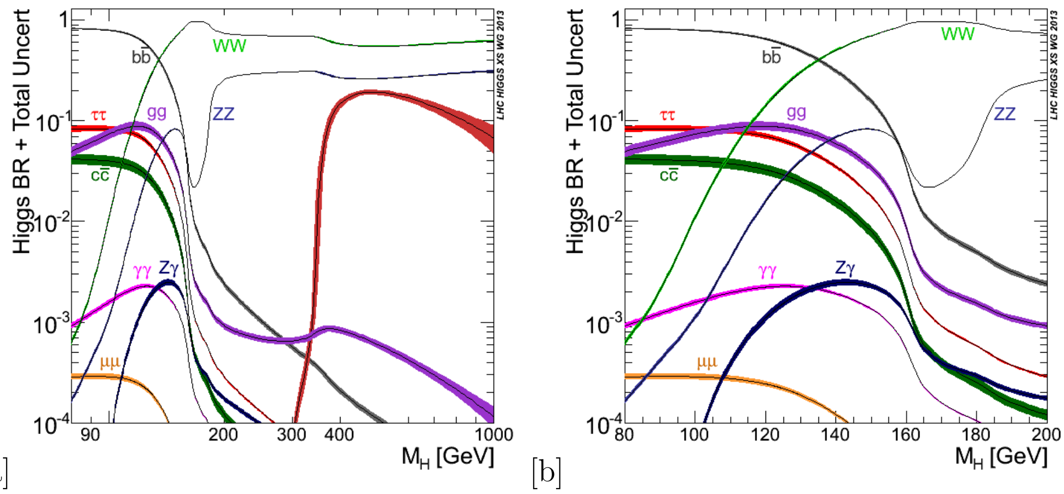
Figure 1.10: LO Feynman diagram for $t\bar{t}$H associated production.

1.5 Higgs Decays at the LHC

Since, the SM Higgs boson can decay directly into pairs of all massive particles, and via loop diagrams even into pairs of massless gluons and photons, the field of Higgs physics is vast. After a discovery of a Higgs particle it is important to detect the Higgs decay into several decay channels to check whether the coupling strength of the Higgs boson is in proportion to the mass of all fermions, as the SM predicts. The Higgs boson branching ratios and total width are shown in figures 1.11, 1.12 [27–29]. It is clear from the figures that the branching ratios change dramatically across the possible range of the Higgs boson mass making it necessary to have different strategies for the Higgs boson identification depending on its mass. One can categorize the Higgs boson searches being carried out at LHC in three classes based on mass range they cover which are discussed below:

1.5.1 Low Mass Region

The $M_H < 130$ GeV region is termed as low mass region. In this region the channel with the highest branching ratio (BR) is $H \to b\bar{b}$. However, the search for the SM Higgs boson in this channel is quite challenging due to the overwhelming QCD background contribution since this is also a low traverse momentum (p_T) region. In this region, there are decay modes which have very distinctive signature such as $H \to \gamma\gamma$. It is termed as golden channel because Higgs boson mass peak can be reconstructed with very high resolution as γ is very well reconstructed in the detector. Further, the SM Higgs boson produced through associated production modes have very distinctive final state due to the presence of lepton (provide good trigger) and b-quarks, for which there are well established techniques to separate it from light quarks.



[a] [b]

Figure 1.11: The BRs for the different Higgs boson decay channels as a function of M_H (a) in the entire mass range (b) zoomed at low mass region.

1.5.2 Intermediate Mass Region

The mass starting from 130 GeV to 180 GeV can be classified as intermediate mass range, where Higgs boson can decay predominantly in a pair of W/Z bosons, with atleast one of the two on mass shell. In fact this is most interesting decay mode as both of these have large production cross-section throughout. Fully leptonic and then semileptonic decay modes of the W and Z bosons are important. $H \to ZZ^* \to 4\ell$ is golden channel because leptons (e, μ) are very well reconstructed in the detector.

1.5.3 High Mass Region

This region corresponds to the Higgs mass values above the $2M_Z$ threshold, $M_H > 180$ GeV, where the Higgs boson can decay into a pair of on-shell W/Z bosons. Since

Higgs boson production cross-section decreases very fast at higher masses, channels with high BR become more important. Hence, $H \to WW \to \ell\nu qq$, $H \to WW \to 2\ell 2\nu$, $H \to ZZ \to 2\ell 2\nu$ and $H \to ZZ \to 2\ell 2q$ are most sensitive final states for High mass SM Higgs boson searches. Figure 4.1 (in Chapter 4) shows the cross-section times BR of various high mass channels studied at the LHC.

1.5.4 Higgs Decay Width

The total decay width, Γ, of the Higgs boson which is given by the sum over all the possible decay channels, is shown in Figure 1.12. The total decay width is very narrow in the low mass range, $\Gamma < 10$ MeV, but the width becomes rapidly wider for masses larger than 130 GeV, and reaches to about 1 GeV slightly above the ZZ threshold. The low mass range is therefore the most challenging region for the experiments, because the total decay width is dominated by the experimental resolution. In the high mass region, the total decay width

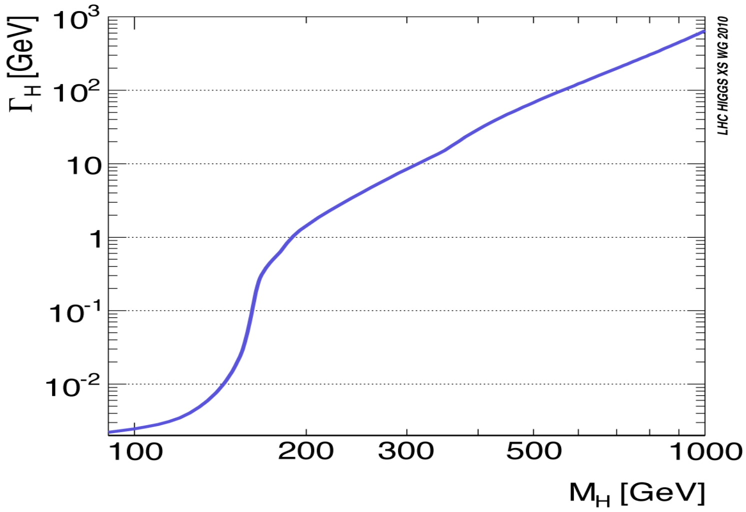


Figure 1.12: The SM Higgs boson total decay width as a function of M_H (GeV).

comprises of the ZZ and the WW partial widths. For the still higher Higgs boson masses, decay width is comparable to its mass. For $M_H \sim 1\ TeV$, the total decay width of Higgs boson is $\Gamma_H \sim 700\ GeV$, that results in a very broad resonant structure.

1.5.5 Electroweak Singlet Model

In the Beyond the SM (BSM) interpretation we interpret our result in electroweak singlet model [30–43] , in which a search for the EW singlet scalar is performed where a heavy Higgs boson can mix with the recently discovered Higgs boson candidate. Phenomenologically the couplings of the two gauge eigenstates (SM and singlet) can be related by unitarity and the original coupling strength of the light Higgs boson is therefore reduced with respect to the SM case. if C (C') are defined as the scale factor of the couplings of the low (high) mass with respect to the SM, one can write $C^2 + C^{'2} = 1$ for preserving unitarity. And, indirectly one can set an upper limit at 95% C.L. on $C^{'2} < 0.446$ using the signal strength fits to the Higgs candidate as obtained in Ref. [44]. The heavy Higgs boson cross-section is modified by a factor, μ', and the modified width is Γ'. They are defined as:

$$\mu' = C'^2(1 - \mathcal{BR}_{\text{new}}) \tag{1.38}$$

$$\Gamma' = \Gamma_{\text{SM}} \times \frac{C'^2}{(1 - \mathcal{BR}_{\text{new}})} \tag{1.39}$$

where $\mathcal{BR}_{\text{new}}$ is the branching ratio of the heavy Higgs boson to non-SM-like decay modes. The BSM heavy Higgs line shape is evaluated by reweighting the SM POWHEG samples following the procedure outlined in Ref. [45]. The model is implemented as a rescaling of the SM at NLO in QCD and LO in EWK. This analysis intends to interpret the results in another model called EW Singlet Model besides the SM Higgs boson search, for that there are no specific signal samples as such available. However, one can do that by following a procedure in which the Higgs line-shape is reweighted as per expected cross-section for EW Singlet model resonance. Benchmark models for BSM heavy Higgs boson searches are outlined in Ref. [46]. Consider a heavy Higgs called h_2 where h_1 is the boson discovered at $\sim$125 GeV. Couplings of h_1 to vector bosons and fermions are denoted as C_V and C_f respectively. Couplings of h_2 to vector bosons and fermions are denoted as $C_V^{'}$ and $C_f^{'}$ respectively. Unitarity requires the following conditions:

$$C_V^2 + C_V^{'2} = 1 \tag{1.40}$$

$$C_V C_f + C_V^{'} C_f^{'} = 1 \tag{1.41}$$

Benchmark models are proposed where:

Model 1:$C \equiv C_V = C_f$ such that the h_2 coupling is given as $C'^2 = \sqrt{1 - C^2}$
Model 2: $C_V \neq C_f$ such that the h_2 couplings are given as $C_V' = \sqrt{1 - C_V^2}$ and $C_f' = (1 - C_V C_f)/\sqrt{1 - C_V^2}$

An additional parameter is needed in both cases for a heavy Higgs decaying into non-SM-like modes which we denote as BR_{new}. In Model 1, we are then left with the observable defined in Equation (1.38) and (1.39). We consider now the reweighting necessary for Model 1. Given the new benchmark parameters of the BSM model, it is necessary to reweight the line shape of the heavy Higgs boson to arbitrary width. Therefore, we propose to modify the reweighting scheme for the SM case given in the previous section. First, we reweight the complex pole scheme (CPS) lineshape, before the inclusion of interference reweighting with $gg \to WW$, to arbitrary width. Then we apply the interference reweighting, which has been modified to include effects from a narrower width resonance with modified couplings. Currently we lack a numerical solution to reweight the CPS from the SM width to an arbitrary width, so we take an analytical solution. We fit the SM CPS lineshape with a Breit-Wigner function with a running width. The shape is given in Eq. 4.1.

$$F_{BW}(x) = \frac{s\Gamma/m_H}{(s - m_H^2)^2 + (s\Gamma/m_H)^2} \tag{1.42}$$

where $s = m_{WW}^2$, m_H is Higgs boson mass at virtuality, Γ is Higgs boson width at virtuality, The fitted width is then scaled by a factor $(C'^2)/(1 - BR_{\text{new}})$ according to Model 1.

1.6 Thesis Organization

This thesis aims to work on two fundamental problems of the SM. First, a search for the SM Higgs boson (H) is carried out in H decaying to WW in the semileptonic final state ($H \to WW \to ł\nu jj$) with the Compact Muon Solenoid (CMS) detector at the LHC. Further, a measurement of WW production in vector boson fusion topology is carried out with the CMS detector at the LHC. Both these analyses are performed using data recorded at $\sqrt{s} = 8\ TeV$ in 2012 corresponding to an integrated luminosity of 19.3 fb^{-1}. This thesis is organized in **six** chapters. A brief description of each chapter is given below:

Chapter 1: This chapter introduce the SM of particle physics. Starting from the gauge invariance of the SM, it discusses Higgs mechanism and generation of mass for vector bosons and also for fermions by Yukawa interactions with Higgs field. The experimental and theoretical bounds are reported. It gives a historical journey of Higgs searches at various colliders and differentiates direct and indirect searches of the Higgs boson.

Chapter 2 describes, in detail, about the experimental set up used for the collection of data, which has been analysed in this thesis work. These are the LHC accelerator and the CMS detector based at CERN, Geneva, Switzerland. This chapter starts with an introduction of the LHC accelerator and its operating parameters. Then, it discusses the CMS detector with an explanation of all its components, purpose and design. The analysis discussed in this thesis has leptons (electrons and muons), missing transverse energy and Jets in the final state. So, every sub-detector of the CMS plays a critical role. The CMS trigger system and data acquisition system are briefly discussed along with the physics requirements of the trigger system. In the end, computing tools are explained which are useful in carrying out physics analysis. A brief overview of the CMS computing structure is also given.

Chapter 3: The data collected from collision of proton beams in every few miliseconds are in the form of digitized electronic signal. These signals are recorded from electronic channels of tracker, calorimeters and muon system and stored as RAW (straight from detector) data. Starting from raw data produced from the online system, successive degrees of processing (event reconstruction) refines this data, applies calibrations and creates higher-level physics objects [5]. CMS uses a number of data formats with varying degrees of detail, size, and refinement to write this data in its various stages. In turn, the data formats get grouped within an 'Event' file into multiple Event formats, according to the data origin or content. An 'Event' is reconstructed from the signals from all particles of an interaction, or possibly even several interactions, joined together. After sorting out which bits of information are related to the same particle (this process is called pattern recognition) the kinematical properties of each particle have to be reconstructed to reveal the physical nature of the whole event. This last process is called reconstruction.

This chapter describes the details of the objects associated with physics analyses such as electrons, photons, muons, missing transverse energy and jets; heavy-quarks and taus. These objects are available in the standard AOD (Analysis Object Data) format. They are produced using a wide variety of sophisticated reconstruction algorithms.

Chapter 4: A search for the SM Higgs boson is described in $H \to WW \to ł\nu jj$ decay channel [6]. This decay channel is sensitive to higher Higgs boson masses ($\geq$ 200 GeV). $H \to WW \to ł\nu jj$ is one of the leading channels in the CMS experiment for search of the SM Higgs boson in the high mass region. The observation of a Higgs boson with mass of about 125 GeV is consistent with the theoretical constraint coming from the unitarization of WW scattering at high energies. However, there is still a possibility that the newly discovered particle has no connection to the EW symmetry breaking mechanism; thus, it is important to continue with the search for the SM Higgs boson in the high mass regime to further confirm the properties of the new boson at 125 GeV. In addition, several popular scenarios, such as general two-Higgs-doublet models or models in which the SM Higgs boson mixes with a heavy EW singlet, predict the existence of additional resonances at high mass, with couplings similar to the SM Higgs boson. The search for Higgs-like boson decaying into WW bosons, using the SM Higgs as a benchmark, is therefore a way to exclude such models or restrict their parameter space. For Higgs boson masses above or near the threshold for decay into two W/Z bosons, the decay modes of choice are dominated by those decays in bosons because of their large branching fractions. The events where one W decays leptonically, which provides the main trigger elements, while the other decays hadronically, have the second highest branching fraction and have a reconstructable Higgs boson mass peak.

The chapter starts with an introduction of signal and various background processes. The full selection procedure is then described, starting with the selection of physics objects

[5] Reconstructed particles

[6] decay channel corresponds to a particular decay mode of the Higgs boson

and then the kinematical selection of W boson candidates. The analysis selects events with one well-identified and isolated lepton, large missing energy and at least two high transverse momentum jets. A MVA discriminant [7] is designed in order to control the background while preserving, as much as possible, the signal component. The main background contaminating the signal region is $W + jets$ having production cross-section much higher than that of Higgs boson and hence it is a major experimental challenge. Its contribution is determined from data. Minor backgrounds include $t\bar{t}$, single top quark, and diboson, and are determined using Monte Carlo. This analysis obtains limits on the SM Higgs cross section in the high mass region of the Higgs boson, as well as on the heavy Higgs boson coupling and branching ratios in the BSM framework.
Once the analysis is optimized, a simultaneous fit using unbinned maximum likelihood fit is performed for background only and signal plus background only hypotheses. Further, limits extraction is performed for each Higgs boson mass point.

Chapter 5: This chapter discusses a first measurement on vector boson scattering. This analysis will serve as a baseline for future WW scattering analysis. The SM of particle physics has been extremely successful in predicting a vast variety of phenomena, that it is easy to forget that some of its predictions have not yet been verified. One such extremely important, which is related intimately to EW symmetry breaking, is that the gauge bosons (γ, W and Z) can interact with each other through quartic interactions. Four such interactions are allowed in the SM: $WWWW$, $WWZZ$, $WWZ\gamma$ and $WW\gamma\gamma$. The other boson couplings are forbidden on the basis of symmetry grounds. The CMS and A ToroidaL LHC Apparatus (ATLAS) collaboration has recently found evidence for a process involving the first of these four, the $WWWW$ interaction in fully leptonic channel. The $WWWW$ and $WWZZ$ interactions, in particular, are of great theoretical interest. If there were no Higgs boson, the rate of these processes would become unphysically large. With the discovery of a Higgs boson, these processes provide useful probe to study EW symmetry breaking and even to probe for new heavier Higgs bosons. At the LHC, the $WWWW$ interactions can be studied through the radiation of W bosons from quarks, which is a rare process. The final state has two central jets and two forward jets. A special jet selection technique is developed with emphasis on having higher signal efficiency. This work is performed for the first time in the CMS. An understanding of signal in this unexplored phase space is mandatory and to achieve this goal, a generator level study using various signal event generators is performed. We extract the vector boson fusion signal yield from an unbinned maximum likelihood fit to the dijet invariant mass distribution in the data.

[7]MVA analysis is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time.

Chapter 6 summarizes the results presented in this thesis and recent results on the search for the SM Higgs boson and vector boson scattering is also discussed.

Chapter 2

The Large Hadron Collider and the CMS Detector

This chapter describes the details of the tools used for this thesis, i.e. the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector. It also discusses the goals and sub-detectors of the CMS experiment at the LHC.

2.1 The Large Hadron Collider

The LHC (as shown in figure 2.1) is the world's most powerful particle accelerator. It is

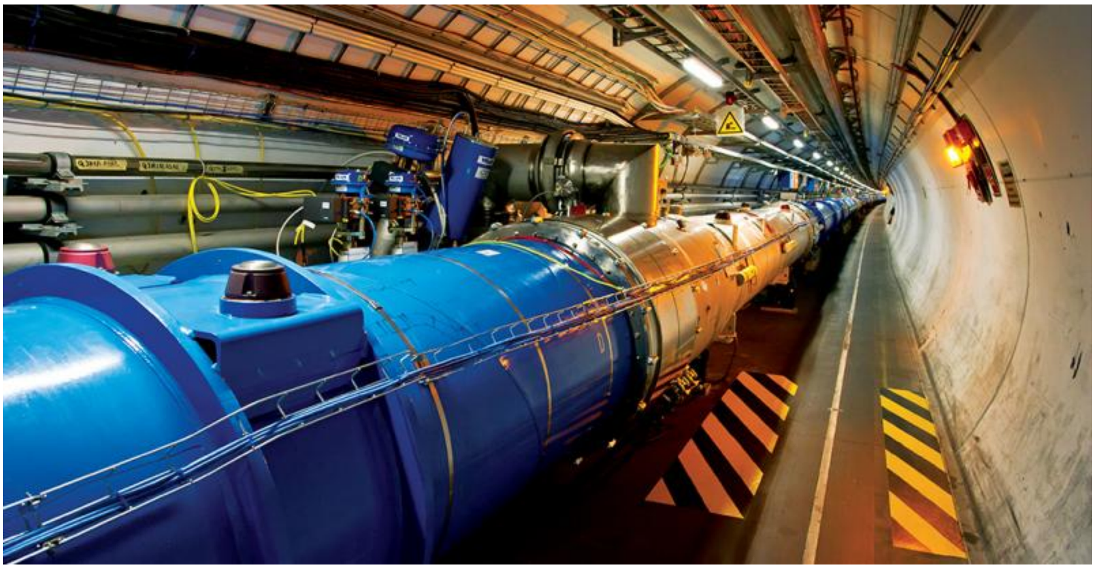


Figure 2.1: The LHC Tunnel.

situated at the border of France and Switzerland at the CERN (European Organization for Nuclear Research) laboratory, about 100 meter underneath the earth. The layout of the LHC in shown in figure 2.2.

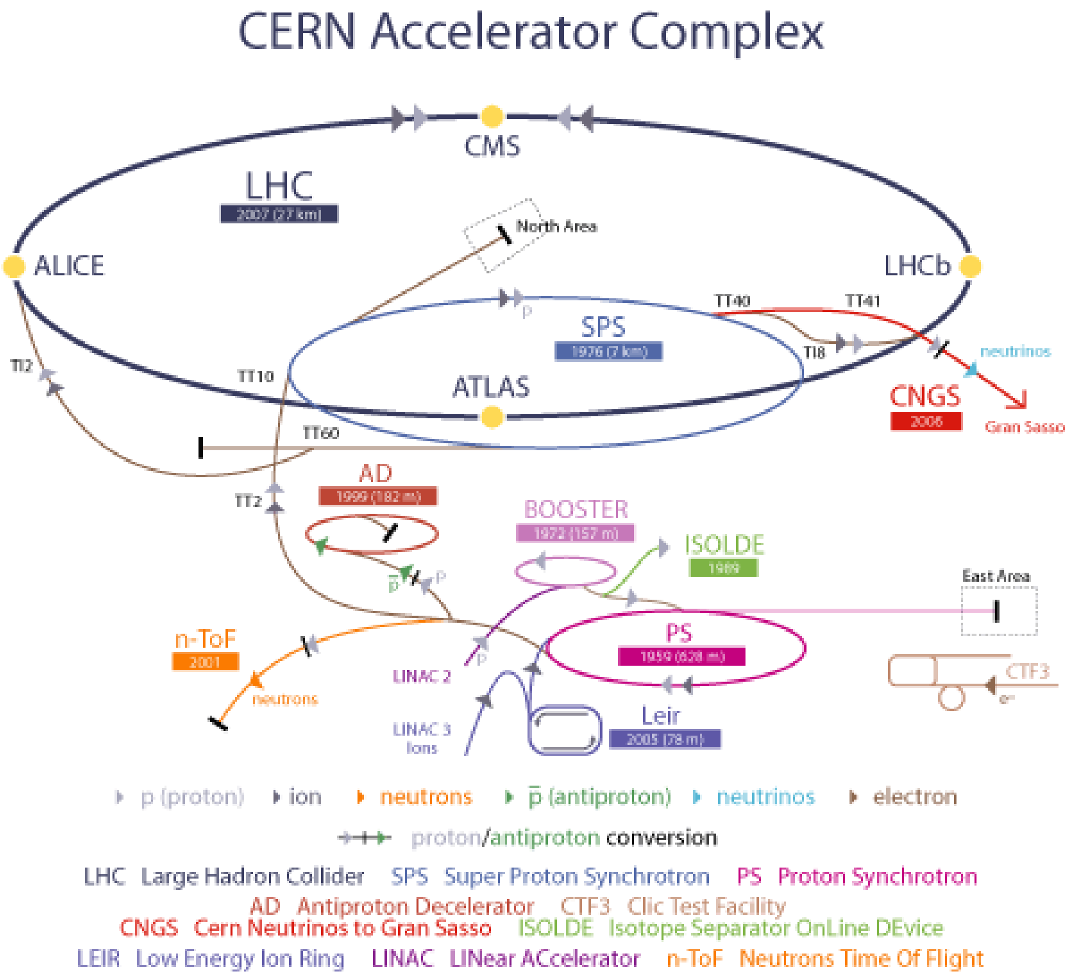


Figure 2.2: The LHC layout and accelerator chain.

The LHC consists of an almost circular tunnel, 27 km in circumference, comprising of superconducting magnets with a number of accelerating structures to boost the energy of the charged particles along the way. Inside the accelerator, two high-energy particle beams travel at close to the speed of light before they are made to collide. The beams travel in opposite directions in separate beam pipes in two tubes kept at ultra-high vacuum. They are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets. The electromagnets are built from coils of special electric cable that operates in a superconducting state, efficiently conducting electricity almost without resistance or loss of energy. This requires cooling the magnets to -271.3^0C, a temperature colder than outer space. For this reason, much of the accelerator is connected to a distribution system of liquid helium, which cools the magnets, as well as the other supply services. Thousands of magnets of different varieties and sizes are used to direct the beams around the accelerator. These include 1232 dipole magnets, 15 m in length which bend the beams, and 392 quadrupole magnets, each 5-7 m long, which focus the beams. Just prior to collision, another type of magnet is used to "squeeze" the particles together to increase the chances of collisions. The particles are so tiny that the task of making them collide is akin to firing two needles 10 kilometres apart with such precision that they meet halfway.

The LHC is designed to collide proton beams at a centre-of-mass energy ($\sqrt{s}$) upto 14 TeV with an unprecedented peak instantaneous luminosity of about 10^{34} cm^{-2}s^{-1}. It can also collide heavy ions (Pb) at an energy of 2.8 TeV per nucleon with the peak luminosity of 10^{27}cm^{-2}s^{-1}. The LHC comprises of four active interaction points where beams travelling in opposite directions are brought to collision. The interaction points correspond to four currently operational major experiments: ATLAS (A Toroidal LHC ApparatuS), CMS, ALICE (A Large Ion Collider Experiment) and LHCb (Large Hadron Collider-beauty). The main focus of the ATLAS and CMS experiments, among other things, have been the search for the Higgs boson. ALICE is mainly used to study quark-gluon plasma under conditions theorized to be very close to those just after the Big Bang. LHCb is a specialized bottom quark-physics ("b-physics") experiment, measuring the parameters of CP-violation in the interactions of B-hadrons. The main aim of the LHCb experiment is to establish a better understanding of the matter-antimatter asymmetry of the universe. In addition, three smaller experiments share interaction points with CMS, ATLAS and LHCb: TOTEM (TOTal Elastic and diffractive cross section Measurement), LHCf ("Large Hadron Collider forward", studying particles in the very high pseudorapidity region with an aim to explain the origin of ultra-high-energy cosmic rays) and MoEDAL (Monopole and Exotics Detector at the LHC, searching for the theoretical magnetic monopole).

The LHC is expected to answer some of the fundamental open questions in physics, concerning the basic laws governing the interactions among the elementary objects, the deep structure of space and time, and in particular the interrelation between quantum mechanics and general relativity, where current theories and knowledge are unclear or break down altogether. Many theorists expect new physics beyond the SM to emerge at the TeV energy level, as the SM appears to be unsatisfactory. Major issues that will be explored during the next run of the LHC collisions includes: Is the newly discovered SM Higgs like boson is actually the SM Higgs boson? Are the masses of elementary particles actually generated by the Higgs mechanism via electroweak symmetry breaking? Is there any evidence for WW scattering in accordance with the Unitarity conservation? Does supersymmetry (SUSY), an extension of the SM, having all known particles have supersymmetric partners exist? Are there extra dimensions, as predicted by various models based on string theory, and can we detect them? What is the nature of the dark matter that appears to account for 27% of the mass-energy of the universe? and so on.

2.1.1 LHC accelerator Chain

The LHC accelerator aspires to achieve an energy so high that it's impossible to achieve it with simple singly designed accelerator due to technical reasons, for example limitation from highest achievable magnetic field. So, it does that by assembling a chain of small and large accelerators, though each one of them has its own complex design structure and facility. It consists of one linear and four synchrotron accelerators, each pushes energy up to certain level before it feeds the beam to next accelerator. The free proton source is obtained by simply stripping off proton from hydrogen atoms using an electric field. Those hydrogen atoms are accelerated to an energy of 50 MeV and gained in mass by 5 % by first and only linear accelerator called Linear accelerator 2 (LINAC 2) in the chain. LINAC 2 uses radio-frequency cavities to accelerate protons. Next in the chain is first synchrotron accelerator named Proton Synchrotron Booster (PSB), which accelerates the protons to an energy of 1.4 MeV. The PSP is made up of four superimposed synchrotron rings to help in maximizing the intensity of protons beam. A synchrotron works on the principle that a charged particle can be accelerated by alternating electric field while keeping it confined in a constant circular orbit by magnetic field. The PSP feeds up the protons to next synchrotron accelerator called Proton Synchrotron (PS) which accelerates protons to an energy of 25 GeV. PS also receives beam for heavy ion collisions from Low Energy Ion Ring (LEIR). Next, PS delivers the protons to Super Proton Synchrotron (SPS) which accelerates the protons to an energy ot 450 GeV. Protons in the form of packets are now ready to feed in main LHC collider. The LHC has two rings for proton beams to come from opposite direction and collide at defined collision points where detectors like CMS, ATLAS etc are stationed. Figure 2.2 shows complete accelerator chain of the LHC.

2.1.2 The LHC parameters

Figure 2.3 lists the values for the main luminosity-related parameters at peak performance of the LHC from 2010 to 2012 and at the design values. It shows that, even though the beam size is naturally larger at lower energy, the LHC has achieved 77% of design luminosity at four-sevenths of the design energy with a β^* of 0.6 m (compared with the design value of 0.55 m) with half of the nominal number of bunches. For further details about LHC parameters, one can refer to Ref. [47].

Parameter	2010	2011	2012	design value
Beam energy	3.5	3.5	4	7
β^* in IP 1 and 5 (m)	2.0/3.5	1.5/1.0	0.6	0.55
Bunch spacing (ns)	150	75/50	50	25
Max. number of bunches	368	1380	1380	2808
Max. bunch intensity (protons per bunch)	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}	1.15×10^{11}
Normalized emittance at start of fill (mm mrad)	≈2.0	≈2.4	≈2.5	3.75
Peak luminosity ($cm^{-2}s^{-1}$)	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}	1×10^{34}
Max. mean number of events per bunch crossing	4	17	37	19
Stored beam energy (MJ)	≈28	≈110	≈140	362

Figure 2.3: An overview of performance-related parameters during LHC operations in 2010–2013.

2.1.3 Operational History

The first beam in the LHC was circulated through the collider on the morning of 10th September 2008. CERN successfully fired the protons around the tunnel in stages, three km at a time. The particles were fired in a clockwise direction into the accelerator and successfully steered around it. The LHC successfully completed its major test: after a series of trial runs, two white dots flashed on a computer screen showing the protons travelled the full length of the collider.

On 19th September 2008, a magnet quench occurred in about 100 bending magnets in sectors 3 and 4, where an electrical fault led to a loss of approximately six tonnes of liquid helium (the magnets cryogenic coolant), which was vented into the tunnel. The escaping vapour expanded with explosive force, damaging over 50 superconducting magnets and their mountings, and contaminating the vacuum pipe, which also lost vacuum conditions.

On 20th November 2009, low-energy beams circulated in the tunnel for the first time since the incident, and shortly after, on 30th November, the LHC achieved 1.18 TeV per beam to become the world's highest-energy particle accelerator, beating the Tevatron's (at Fermilab, U.S) previous record of 0.98 TeV per beam held for eight years. The early part of 2010 saw the continued ramp-up of beam in energies and early physics experiments towards 3.5 TeV per beam and on 30th March 2010, LHC set a new record for high-energy collisions by colliding proton beams at $\sqrt{s}$ of 7 TeV. This also marked the start of its main research program.

Since the start of LHC it has worked beyond expectations and integrated luminosity delivered in 2011 is $6.13 fb^{-1}$, as shown in figure 2.4 on left. In 2012, $\sqrt{s}$ has been incremented to 8 TeV, with an integrated luminosity of $23.3 fb^{-1}$ as shown in figure 2.4 on right.

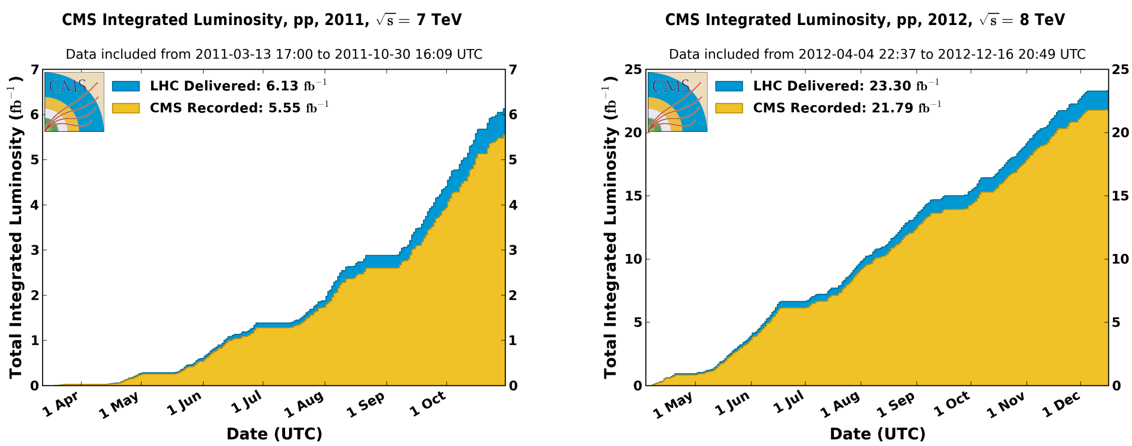


Figure 2.4: Cumulative luminosity versus day delivered by LHC (blue), and recorded by CMS (orange) during stable beams and for $p-p$ collisions at $\sqrt{s} = 8$ TeV in 2011 and 2012 on left and right respectively.

2.2 The CMS Detector

The CMS, together with the ATLAS, is one of the two main experiments at the LHC collider. The main objective of the CMS detector is to make research in the experimental frontier of High Energy Physics (HEP). This includes, for instance, Higgs boson search, SUSY, extra dimensions search etc. In order to fulfill physics aims, the CMS is designed as multi-purpose detector with several sub-detector layers. Each layer is designed to stop, track or measure different type of particles emerging from $p-p$ collisions and heavy ion collisions. The main components of the CMS are (listed from the inside to the outside) a full silicon tracker, a scintillating electromagnetic calorimeter (ECAL), a hadron calorimeter (HCAL), a superconducting solenoid magnet, and a muon detection system. More than 3000 scientists are working or collaborating in the CMS, coming from 155 institutes in 37 countries.

2.2.1 The Co-ordinate system

The CMS experiment uses a right-handed coordinate system, with the origin at the nominal collision point, the $x-axis$ pointing to the centre of the LHC ring, the $y-axis$ pointing up (perperpendicular to the LHC plane), and the $z-axis$ along the anticlockwise beam direction. The polar angle (θ) is measured from the positive $z-axis$ and the azimuthal angle ϕ is measured from the positive $x-axis$ in the $x-y$ plane. The radius (r) denotes the distance from the $z-axis$ and the pseudo-rapidity (η) is defined as $\eta = -ln[\tan(\frac{\theta}{2})]$.

2.2.2 Detector layout

In present-day HEP experiments, any detector is made up of a number of sub-detectors, each providing some level of particle identification and kinetic measurements. The CMS is no exception and it has a number of components explained below and a perspective view is shown in figure 2.5.

2.2.3 The Superconducting Magnet

The CMS magnet is the central device around which the experiment is built, with a 4 Tesla magnetic field that is 100,000 times stronger than the earth's magnetic field. In figure 2.6, LHC magnet is shown at the time of assembly. The job of the magnet is to bend the paths of charged particles emerging from high-energy collisions at the LHC. The more the momentum a charged particle has, the less its path is curved by the magnetic field. Thus, tracing its path gives a measure of momentum. The CMS started its operation with the aim of having the strongest magnet possible because a higher strength field bends paths more and, combined with high-precision position measurements in the tracker and

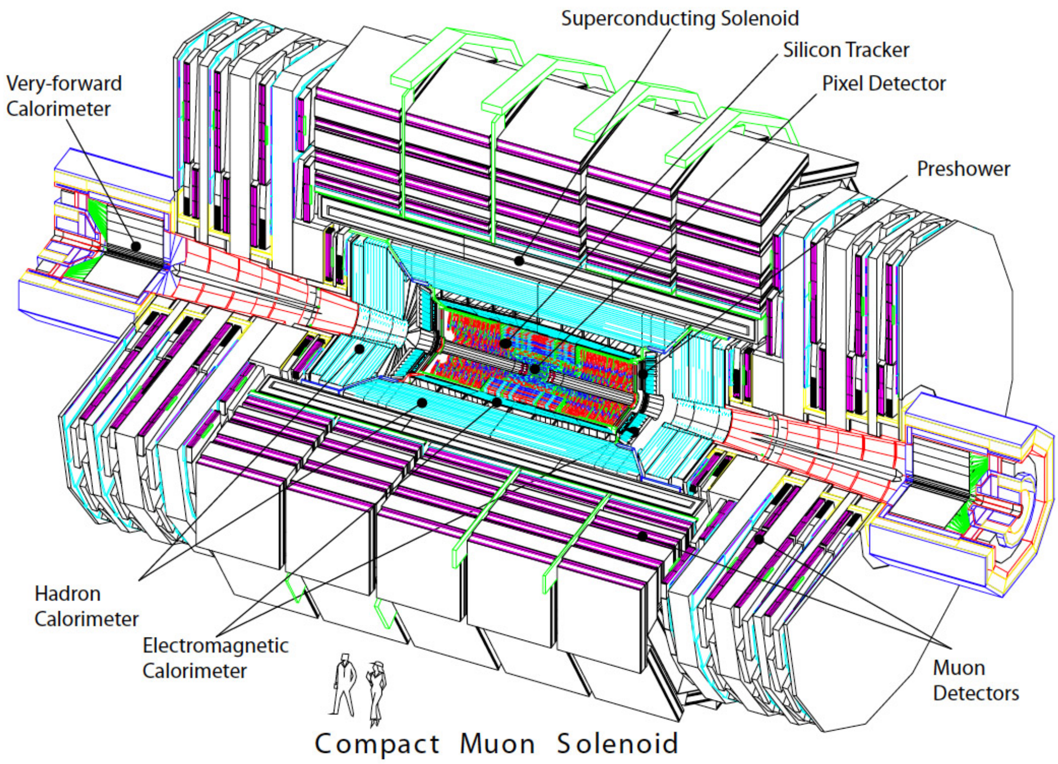


Figure 2.5: A perspective view of the CMS detector.

muon detectors, this allows accurate measurement of the momentum of even high-energy particles.

The CMS magnet [48] is a “solenoid” - a magnet made of coils of wire that produces a uniform magnetic field when current flows through them. The CMS magnet is “superconducting”, allowing current to flow almost without resistance and creating a powerful magnetic field. In fact, at ordinary temperatures, the strongest possible magnet has only half the strength of the CMS solenoid.

The tracker and calorimeter detectors (ECAL and HCAL) fit snugly inside the magnet coil whilst the muon detectors are interleaved with a 12-sided iron structure that surrounds the magnet coils and contains and guides the field. The “return yoke” made up of three layers reaches out to 14 m in diameter and also acts as a filter, allowing through only muons and weakly interacting particles such as neutrinos. The enormous magnet also provides most of the experiments structural support, and is very strong itself to withstand the forces of its own magnetic field.

2.2.4 The Tracker

Momentum of particle is crucial in helping us to build up a picture of events at the heart of the collision. One method to calculate the momentum of a charged particle is to track its path through a magnetic field; the more curved the path, the less momentum the

Figure 2.6: LHC Magnet.

particle has. The CMS tracker records the paths taken by charged particles by finding their positions at a number of key points. The tracker can reconstruct the paths of high-energy muons, electrons and hadrons (particles made up of quarks) as well as see tracks coming from the decay of very short-lived particles such as beauty or “b quarks”. The tracker needs to record particle paths accurately and yet be lightweight so as to disturb the particle as little as possible. It does this by taking position measurements so accurate that tracks can be reliably reconstructed using just a few measurement points. Each measurement is accurate to 10 μm, a fraction of the width of a human hair. It is also the inner most layer of the detector and so receives the highest volume of particles, the construction materials were therefore carefully chosen to resist radiation. The final design consists of a tracker made entirely of silicon: the pixels, at the very core of the detector and dealing with the highest intensity of particles, and the silicon microstrip detectors that surround it. As particles travel through the tracker the pixels and microstrips produce tiny electric signals that are amplified and detected. The tracker employs sensors covering an area of about the size of a tennis court, with 75 million separate electronic read-out channels.

The Silicon Pixel detector

The pixel detector shown in figure 2.7 is about the size of a shoebox and 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 about 4 cm, 7 cm and 11 cm and disks at either end, and so will be vital in reconstructing

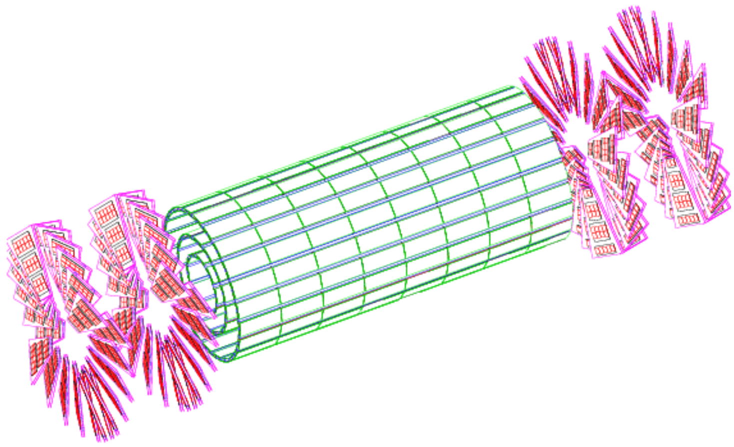


Figure 2.7: CMS Silicon Pixel Detector.

the tracks of very short-lived particles. However, being so close to the collision means that the number of particles passing through this detector system is huge, the rate of particles received at 8 cm from the beam line will be around 10 million particles per square cm per second. The pixel detector is able to disentangle and reconstruct all the tracks they leave behind, and withstand such a pummelling over the ten-year duration of the experiment. Each layer is spilt into segments like tiny kitchen tiles, each little silicon sensor is of dimension 100 μm by 150 μm. When a charged particle passes through, it gives enough energy to silicon atoms to create electron-hole pairs. Each pixel uses an electric current to collect these charges on the surface as a small electric signal. An electronic silicon chip, one for each tile is attached, using an almost microscopic spot of solder using the so-called bump bonding technique, amplifies the signal.

Knowing which pixels have been hit (producing signal) allows one to deduce the particle's trajectory. And because the detector is made of 2-dimensional tiles, rather than strips, and has a number of layers, we can create a three-dimensional picture. Because there are 65 million channels, the power for each pixel must be kept to a minimum. Even with each only generating around 50 μW, the total power output is around the same as the energy produced by a hot plate. So as not to overheat the detector, the pixels are mounted on cooling tubes.

The Silicon Strip detector

After the pixels and on their way out of the tracker, particles pass through ten layers of silicon strip detectors as shown in figure 2.8, reaching out to a radius of 130 cm. The tracker silicon strip detector 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. This part of

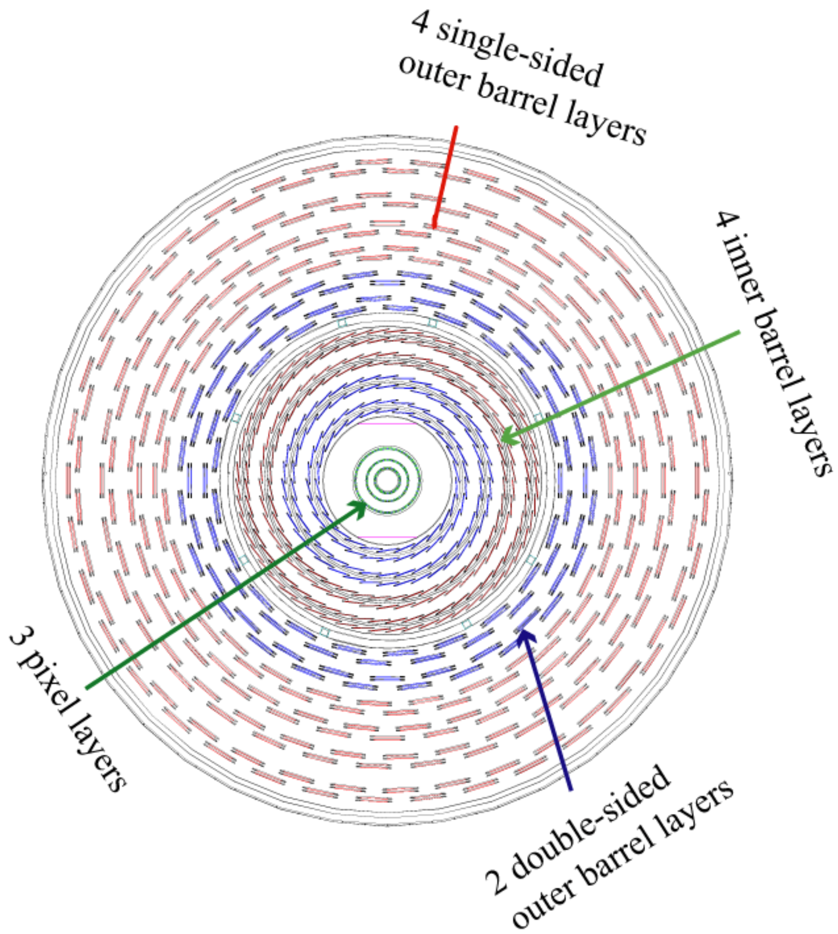


Figure 2.8: CMS Tracker layers shown in the plane perpendicular to the beam.

the tracker contains 15,200 highly sensitive modules with a total of 10 million detector strips read by 80,000 microelectronic chips. Each module consists of three elements: a set of sensors, its mechanical support structure and readout electronics. Silicon sensors are highly suited to receive many particles in a small space due to their fast response and good spatial resolution. The silicon strip detectors work in much the same way as the pixels: as a charged particle crosses the material it knocks electron from atoms and within the applied electric field, these carriers move giving a very small pulse of current lasting a few nanoseconds, giving us "hits" when a particle passes, allowing us to reconstruct its path. Due to the proximity to collision point, tracker electronics are struck repeatedly by radiation but they are designed to withstand it. To minimise disorder in the silicon this section of the detector is kept at -20^0C, to "freeze" any damage and prevent it from perpetuating. The charge on each microstrip is read out and amplified by an Analogue Pipeline Voltage (APV25) chip. Four or six APV25 chips are housed within a "hybrid", which also contains electronics to monitor key sensor information, such as temperature, and provide timing information in order to match 'hits' with collisions. The APV25 stores the signals in a memory for several microseconds and then processes them before sending to a laser to be converted into infrared pulses. These are then transmitted over a 100 m fibre optic cable for analysis in a radiation-free environment. The tracker uses 40,000 such fibre optic links providing a low power, lightweight way of transporting the signal. Much of the technology behind the tracker electronics came from innovation in collaboration with industry.

2.2.5 Electromagnetic Calorimeter (ECAL)

In order to build up a picture of events occurring in the LHC, CMS must measure the energies of emerging particles. Of particular interest are electrons and photons, because of their use in finding the Higgs boson and other new physics. The energies of these particles are measured using an electromagnetic calorimeter (ECAL) as shown in figure 2.9. But

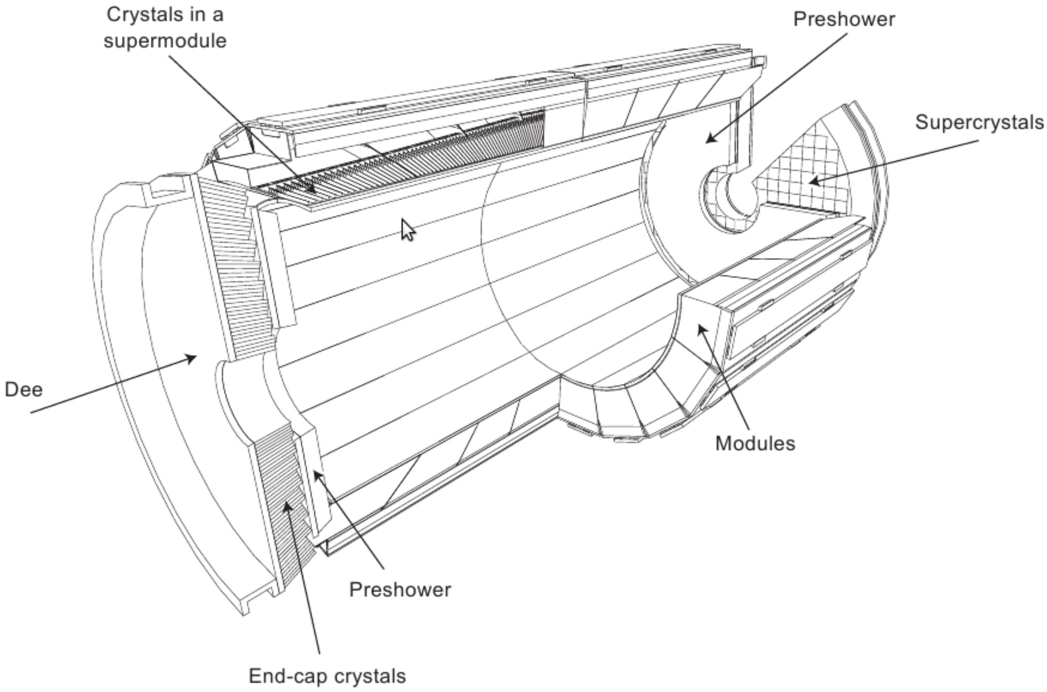


Figure 2.9: Layout of the CMS ECAL, showing the barrel supermodules, the two end-caps and the preshower detectors.

due to the very difficult conditions of the LHC, i.e. having high magnetic field, high levels of radiation and only 50 ns duration (or time gap) between collisions, energy measurement with necessary precision requires particular detector materials. One such choice is Lead tungstate crystal, which made primarily of metal and is heavier than stainless steel, but with a touch of oxygen in this crystalline form it is highly transparent and "scintillates" when electrons and photons pass through it. It produces light in proportion to the particles energy. These high-density crystals produce light in fast, short, well-defined photon bursts and thus allows a precise, fast and fairly compact detector. Photo-detectors that have been especially designed to work within the high magnetic field, are also glued onto the back of each of the crystals in order to detect the scintillation light and then convert it to an electrical signal that is amplified. The ECAL, made up of a barrel section and two “endcaps”, forms a layer between the tracker and the Hadron Calorimeter. The cylindrical “barrel” consists of 61,200 crystals formed into 36 supermodules, each weighing $\sim$ 3 tonnes and containing 1700 crystals. The flat ECAL endcaps seal off the barrel at either end and are made up of almost 15,000 further crystals. For extra spatial precision, the ECAL also contains Preshower detectors that sit in front of the endcaps. These allow CMS to distinguish between single high-energy photons (often signs of exciting physics) and the less interesting close pairs of low-energy photons (say, for example coming from π).

The energy resolution (σ) of ECAL has been parametrized as:

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

where ‘a’ is the stochastic term, ‘b’ is the the noise, and ‘c’ the constant term. The stochastic term includes fluctuations in the shower containment as well as a contribution from photostatistics. The noise term contains the contributions from electronic noise and pileup energy. The former is quite important at low energy and the latter is negligible at low luminosity. The measured resolution for the ECAL has the following values: $\sqrt{a} = 2.8\%$ of E, $b = 0.0415$ GeV and $c = 0.3\%$ [49]. The contribution changes at different pseudorapidities and with the luminosity of the machine.

2.2.6 The Hadronic Calorimeter

The Hadron Calorimeter (HCAL), shown in figure 2.10, measures the energy of “hadrons” and the particles made of quarks and gluons. Additionally, it provides indirect measurement of the presence of non-interacting, uncharged particles such as neutrinos. Measuring these particles is important as they can tell us if new particles such as the Higgs boson have been formed or not. As these particles decay, they may produce new particles that do not leave any signaturerecord of their presence in any part of the CMS detector. To spot them the HCAL must be “hermetic”, i.e. it should capture, to the extent possible, every particle emerging from the collisions. This way if we see particles shoot out on one side of the detector, but not on the other, with an imbalance in the momentum and energy (measured in the sideways “transverse” direction relative to the beam line), we can deduce that the “invisible” particles are produced. To ensure that we are seeing something new, rather than just letting familiar particles escape undetected, layers of the HCAL were built in a staggered fashion so that there are no gaps in direct lines that a familiar particle might escape through. The HCAL consists of layers of brass or steel densely interleaved between the tiles of plastic scintillators. These are read out via

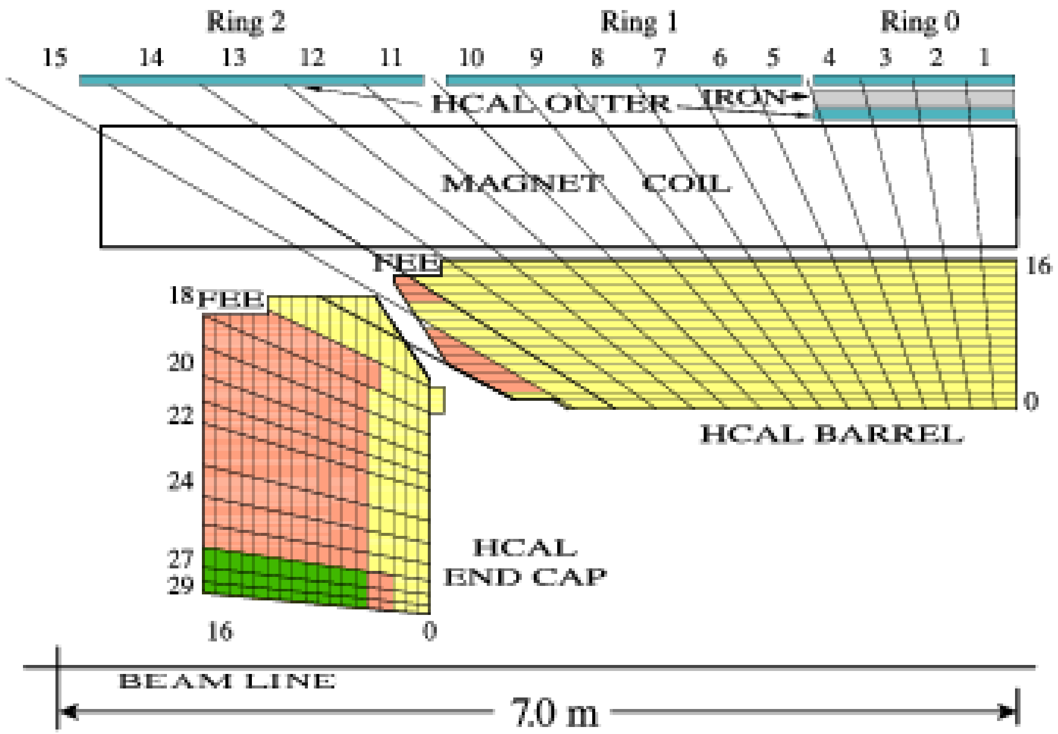


Figure 2.10: A quarter slice of the CMS HCAL detector. The right end of the beam line is the interaction point. FEE denotes the location of the Front End Electronics for the barrel and the endcap. In the diagram, the numbers on the top and left refer to segments, and the numbers on the right and the bottom refer to scintillator layers. Colors/shades indicate the combinations of layers that form the different depth segments, which are numbered sequentially starting at 1, moving outward from the interaction point. The outer calorimeter is assigned depth 4.

wavelength-shifting fibres using hybrid photodiodes. The HCAL is a sampling calorimeter as it measures the particle position, energy and arrival time. It has alternating layers of “absorber” and fluorescent “scintillator” materials which produces a rapid light pulse when the particle passes through it. Special optic fibres collects this light and feed it into readout boxes where photodetectors amplify the signal. The amount of light in a given region is then summed up over many layers of tiles in depth, called a “tower”, and this total amount of light is a measure of a particle’s energy. The HCAL is massive and thick, as the cascades of particles produced (known as showers), when a hadron hits the dense absorber material, are large, and even the minimum amount of material needed to contain and measure them is about one metre. Therefore, the HCAL is organised into barrel (HB and HO), endcap (HE) and forward (HF) sections. There are 36 barrel “wedges”, each weighing 26 tonnes. These form the last layer of detector inside the magnet coil whilst a few additional layers, the outer barrel (HO), sit outside the coil, ensuring no energy leaks out the back of the HB undetected. Similarly, 36 endcap wedges measure particle energies as they emerge through the ends of the solenoid magnet. Lastly, the two hadronic forward calorimeters (HF) are positioned at either end of the CMS, to pick up the myriad particles coming out of the collision region at shallow angles relative to the beam line. These receives the bulk of the particle energies contained in the collision, and hence are resistant to radiation. The energy resolutions (σ) for different parts of HCAL are parametrized by eq. (2.2) [50] :

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus b \tag{2.2}$$

where 'a' corresponds to stochastic term and 'b' to a constant term. The values of the coefficients are measured to be 'a' = 0.847 $GeV^{\frac{1}{2}}$ and 'b' = 0.074, with E measured in GeV.
The energy resolution in the endcaps is similar to that in the barrel. The corresponding values for HF are $a = 1.98$ $GeV^{\frac{1}{2}}$ and $b = 0.09$. Since the forward jets typically have very high energies, the stochastic term is higher for HF than for the other calorimeters.

2.2.7 The Muon System

To identify muons and measure their momenta, CMS uses three specialized detectors: drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). The DTs are used for precise position measurements, i.e. for track reconstruction in the central barrel region, while the CSCs are used in the end caps. The RPCs provide a fast signal when a muon passes through the muon detector. It works as the trigger system, and is installed in both the barrel and the end caps.

In figure 2.11, a muon track is shown. A muon in the plane perpendicular to the LHC beams, leads a curved trajectory in four layers of muon stations due to lorentz force acting upon it.

As the name "Compact Muon Solenoid" suggests, detecting muons is one of CMS's most important tasks. Muons are charged particles just like electrons and positrons, but are 200 times heavier. We expect them to be produced in the decay of a number of potential new particles. For instance, one of the clearest "signatures" of the Higgs Boson is its decay into four muons. Because muons can penetrate several metres of iron almost without interacting (as minimum ionizing particle), unlike most of the other particles, they are not stopped by any of CMS's calorimeters. Therefore, chambers to detect muons are placed at the very end of the experimental setup where these are the only particles that are likely to register a signal. A particle is measured by fitting a curve to hits among the four muon stations, which sit outside the magnet coil and are interleaved with iron "return yoke" plates (shown in red for the barrel region in Figure 2.11). By tracking its position through the multiple layers of each station along with tracker measurements the detectors precisely trace the particle's path. This gives a measurement of its momentum. Therefore, the CMS

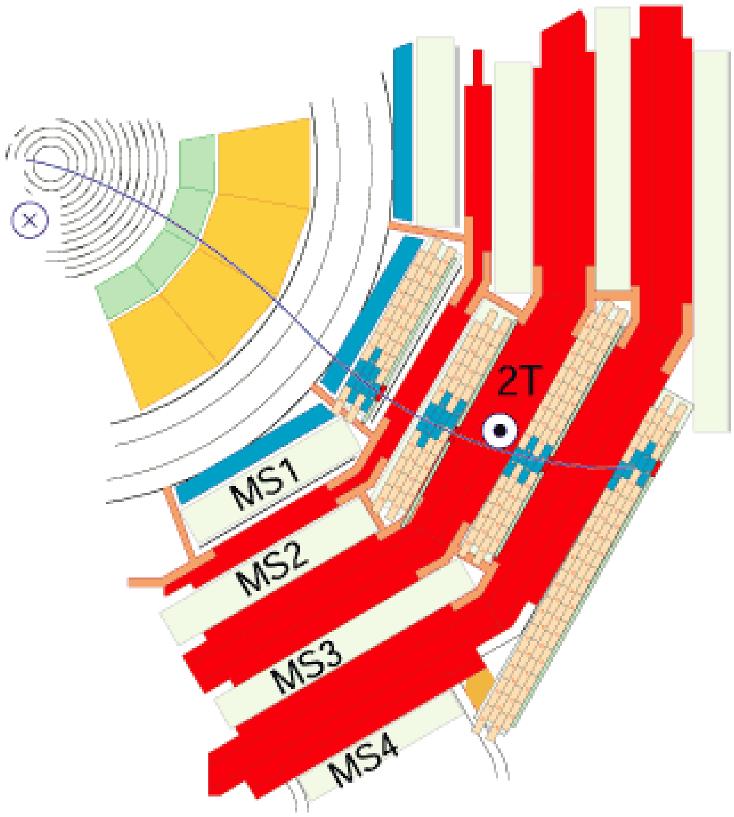


Figure 2.11: A muon, in the plane perpendicular to the LHC beams, leaves a curved trajectory in four layers of muon detectors.

magnet is very powerful so that we can bend even the paths of the very high-energy muons and calculate their momenta. In total, there are 1400 muon chambers: 250 DTs and 540 CSCs which tracks the particles, positions and provides a trigger, while the 610 RPCs forms a redundant trigger system, which quickly decides whether to keep the acquired muon data or not. Because of the many layers of detector and different specialities of each type, the system is naturally robust and able to filter out background noise. DTs and RPCs are arranged in concentric cylinders around the beam line ("the barrel region") whilst CSCs and RPCs, make up the "endcaps" disks that cover the ends of the barrel.

2.3 Triggering and Data Acquisition

In particle physics, a trigger is a system that rapidly decides which events in a particle detector should be kept when only a small fraction of the total events can be recorded. Trigger systems are necessary due to real-world limitations in data storage capacity and rates. Since experiments are typically searching for "interesting" events (such as decays of rare particles) that occur at a relatively low rate, trigger systems are used to identify the events that should be recorded for later analysis. When CMS is performing at its peak, during the peak performance of CMS about one billion proton-proton interactions takes place every second inside the detector, the Higgs boson is expected to be produced there at a rate of at least 1/100 events per second. There is no way that the data from all these events could be read out, and even if it could, most of it would less likely reveal interesting phenomena. For example there might be low-energy glancing collisions rather than energetic, head-on interactions of interest. We therefore need a “trigger” that can select the potentially interesting events, such as those which, for examples, can produce the Higgs particle, and reduce the rate to just a few hundred “events” per second, which can be read out and stored on computer disk for subsequent analysis. However, in the designed system, with groups of protons colliding 40 million times per second there are only ever 25 nanoseconds (25 billionths of a second) before the next lot arrive. Thus, the new waves of particles are being generated even before those from the last event have left the detector. The solution is to store the data in pipelines that can retain and process the information from many interactions at the same time. In order to discriminate the events, the detectors must have very good time resolution and the signals from the millions of electronic channels must be synchronised so that they can all be correlated with the original event. Triggers usually make heavy use of a parallelized design, exploiting the symmetry of the detector. However, on a global scale they are essentially serial devices and in fact, they are usually divided in “levels”. The idea is that each level selects the data which becomes an input for the subsequent level and thus it has more time and information available to take a better decision. Custom hardware processors make an initial decision to keep an event for a few μs using coarsely segmented data from a subset of the detectors, while holding all the high-resolution data in pipelined memories. Commodity processors make subsequent decisions using more detailed information from all of the detectors in more sophisticated algorithms that eventually approach the final reconstruction. In CMS, two step trigger system is used. The first step trigger is called Level 1 trigger, which is basically hardware based while second step trigger system is called High Level Trigger (HLT), which is completely software based.

Level 1 Triggers

The Level-1 trigger (as shown in figure 2.12) is hardware trigger built of custom-made electronics dedicated to analyzing the detector information in a fairly coarsely segmented data from the calorimeter and the muon detectors, while holding all the high-resolution data in pipeline memories in the front-end electronics. It stores all data for $3.2\mu s$ and after that about only 100 KHz of the stored events are forwarded to the High Level Triggers. This must be done for all channels without dead time. This step of trigger is an extremely fast and wholly automatic process that looks for simple signs of interesting physics, for example particles with a large amount of energy or in unusual combinations. This is like a reader simply scanning the headlines of a newspaper to see if anything catches their eye. This way we select the best 100,000 events or “issues” each second from the billion events

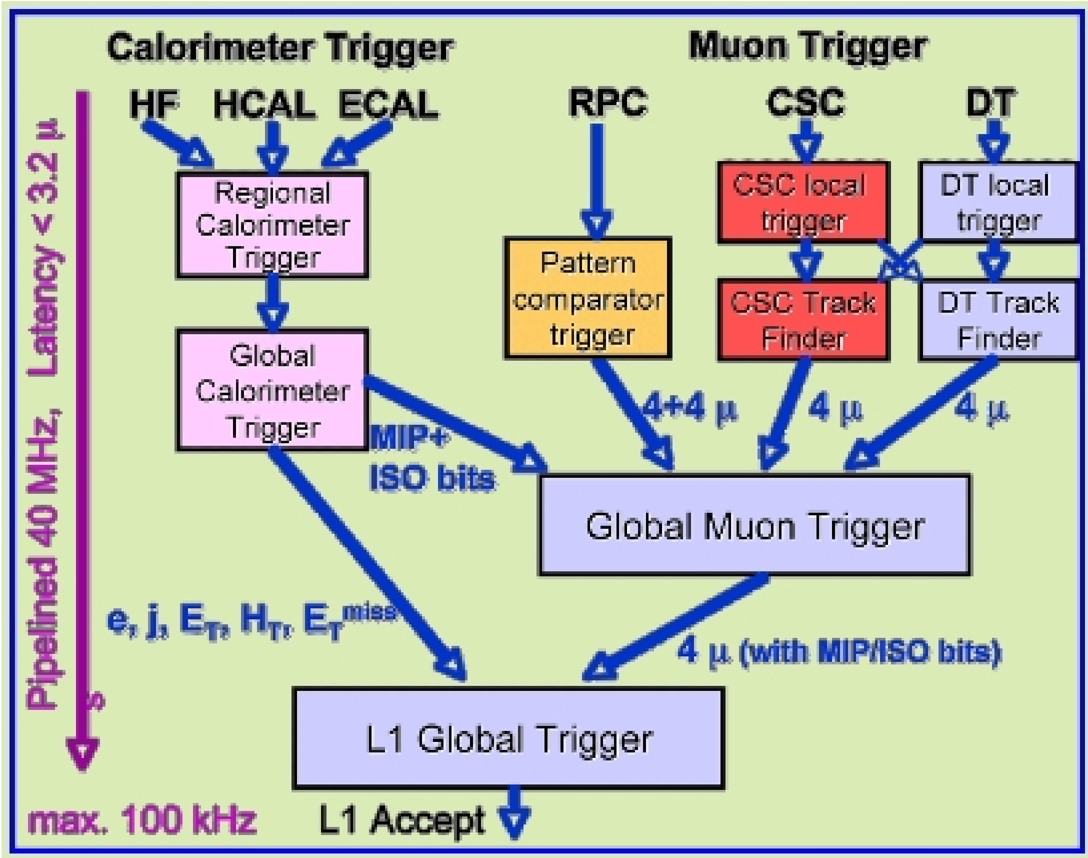


Figure 2.12: A Block Diagram of CMS L1 Trigger System.

available.

2.3.1 The High Level Trigger

The second step (High-Level Trigger or "HLT") is designed to reduce this maximum Level-1 accept rate of 100 kHz to a final output rate of $O(100\text{ Hz})$. The HLT system relies upon commercial processors. It is implemented in a single processor farm (referred to as the "Event Filter Farm") as shown in figure 2.13. As a result, in inspecting the events that have passed the Level-1 trigger the full event information is available to a fully programmable processor. We assimilate and synchronise information from different parts of the detector to recreate the entire event - like collating the different pages to form the full newspaper - and send it to a farm of more than 1000 standard computers. Here the PCs are like speed readers, who, with more detailed information, review the information for longer, less than a tenth of a second. They run very complex physics tests to look for specific signatures, for instance matching tracks to hits in the muon chambers, or spotting photons through their high energy but lack of charge. Overall they select $O(100)$ events per second and the remaining 99,900 are thrown out. We are left with only the collision events that might give indication of something new about physics. Despite the trigger system, CMS still records and analyses several petabytes of data, that's millions of gigabytes, around the

Figure 2.13: A computer farm, cluster of CPU used to process data.

same amount of information as in 10,000 Encyclopaedia Britannica every second, although not all these data are stored.

2.4 CMS Computing Model

Even after significant reduction of data by the two level trigger system, CMS still produces a huge amount of data that must be analysed. It amounted to more than five petabytes per year when running at peak performance. To meet this challenge, the LHC developed a novel computing system, a distributed computing and data storage infrastructure called the Worldwide LHC Computing Grid (WLCG). In 'The Grid', tens of thousands of standard PCs collaborate worldwide to have much more processing capacity than could be achieved by a single supercomputer, giving access to data to thousands of scientists world-wide.

The WLCG is composed of four levels, or "Tiers", called 0, 1, 2 and 3. Each Tier is made up of several computer centres and provides a specific set of services. The Tiers process, store and analyse all the data from the LHC. Tier 0 is the CERN Data Centre. All of the data from the LHC passes through this central hub, but it provides less than 20% of the Grid's total computing capacity. The "Tier 0" centre reconstructs the full collision events and analysts start to look for patterns. CERN is responsible for the safe keeping of the raw data (millions of digital readings from across the detectors), and performs the first pass at reconstructing the raw data into meaningful information. Tier 0 distributes the raw data and the reconstructed output to Tier 1s, and reprocesses data when the LHC is not running.

Tier 1 consists of 13 computer centres, large enough to store LHC data. They provide round-the-clock support for the Grid, and are responsible for storing a proportional share of raw and reconstructed data, as well as for performing large-scale reprocessing and storing the corresponding output; distributing data to Tier 2s and storing a share of the simulated data that the Tier 2s produce. Optical-fibre links working at 10 gigabits per second connect CERN to each of the 13 major Tier 1 centres around the world. This dedicated high-bandwidth network is called the LHC Optical Private Network (LHCOPN). In this way information braches out from each Tier across the world so that, on a local level, physicists and students can study CMS data from their own computer, updated on a regular basis by the LHC Computing Grid.

2.5 CMS Software Components

The overall design of the CMS Software Architecture is motivated by the following underlying principles:

- Multiple Environments: Various software modules must be able to run in a variety of environments as different computing tasks are performed (examples of these environments are level 3 triggering, production reconstruction, program development, and individual analysis);
- Migration between environments: a particular software module may start out being developed for one environment, and then later used in other unforeseen environments as well;
- Dispersed code development: the software will be developed by organisationally and geographically dispersed groups of part-time non-professional programmers – only some portions of the code will be written by computing professionals;
- Flexibility: not all software requirements will be fully known in advance, therefore the software systems must be adaptable without requiring total rewrites;
- Ease of use: the software systems must be easily usable by collaboration physicists who are not computing experts and cannot devote large amounts of time to learning computing techniques.

These requirements on the software architecture result in a customisable application framework for each of the computing environments, which is given below

- physics software modules with clearly defined interfaces that can be plugged into the framework;
- a service and utility toolkit that can be used by any of the physics modules. The framework will do the program control, module scheduling, and input/output. It will be tailored to the task in hand and will most likely be written by computing

professionals (or physicists acting in that role) using disciplined software engineering methods. The tasks that the framework should be able to handle are:

- program development: emphasis on interactivity and access to small numbers of events;
- calibration and monitoring: moderate interactivity, high performance serial access to full runs of calibration or special data samples;
- event analysis: reasonable interactivity, rapid access to random selections of data, and excellent data presentation tools;
- event reconstruction: batch processing, high performance serial I/O for entire runs, well controlled versioning and software engineering;
- trigger level 3 reconstruction: real time processing, high performance I/O, rigidly controlled versioning and software engineering;
- simulation: limited interactivity, highly compute bound jobs.

The physics and utility modules are written by detector group and physicists. The modules can be plugged in to any of the application frameworks at run time. One can easily choose between different versions of various modules. The physics modules will not communicate with each other directly but only through the data access protocols that are part of the service layer below. These are publicly accessible to the collaboration, but may also include private code, not yet released to others. The level of software engineering required varies depending on the environments in which the modules are to be certified for use. The service and utility toolkit will consist of two major categories of services: physics type services (histogrammers, fitters, physics calculation routines) and computer services (data access, inter module communication, event input and output, etc).

2.6 CMS Remote Analysis Builder (CRAB)

CRAB [51] is a utility to submit CMSSW jobs to distributed computing resources. The data analysis on a distributed environment is a complex computing task because the used data have a dimension of hundred Megabytes, that makes the data transfer not convenient for just some small analysis. Also the other heterogeneous instruments are not local, but distributed all around the world. CRAB [3] is the official CMS analysis software that easily interfaces the user with the computing grid environment hiding the system complexities. Following the analysis model it allows an easy access to the data distributed over the grid in a very transparent way and the user is not required to have any deep knowledge about the computing grid. In fact CRAB simplifies the process of CMS analysis allowing to process officially published data and hiding as much as possible the grid complexity to the final user, so that remote data can be accessed with the same facility of local data: CMS user sends through this tool his analysis code to the site where the selected data are. CRAB needs to be installed on the User Interface (UI) -which is the user acces point to the grid, it supports any CMSSW (the CMS software framework) based executable, with any modules/libraries, including the user provided ones, and finally it deals with the output produced by the executable. From a user point of view, the basic steps of the workflow of this software are:

- Job Creation : interaction with data discovery services and with the user environment; the task is splitted into smaller jobs and the input sandbox is prepared.
- Job Submission : interaction with Resource Broker, Workload Management System and proxy services to submit jobs to sites matching the user requirements.
- Job Status : check the status of the jobs.
- Job Output : retrieval of the job output from the grid (output sandbox); also the output can be transferred to a nearby "Storage Element" or to another one specified by the user. Other functionalities are: the possibility to get information about aborted jobs; to stop the running jobs and to resubmit failed jobs

Chapter 3

Physics Object Reconstruction

The data collected from collision of proton beams every few seconds are in the form of digitized electronic signals. These signals are recorded from electronic channels of tracker, calorimeters and muon system and stored as RAW (straight from detector) data. Starting from raw data produced from the online system, successive degrees of processing (event reconstruction) refines this data, apply calibrations and create higher level physics objects [1]. CMS uses a number of data formats with varying degrees of detail, size, and refinement to write this data in its various stages. In turn, the data formats get grouped within an 'Event' file into multiple Event formats, according to the data's origin or content. An 'Event' is reconstructed from the signals from all particles of an interaction, or possibly even several interactions, joined together. After sorting out which bits of information are related to the same particle (this process is called pattern recognition), the kinematical properties of each particle have to be reconstructed to reveal the physical nature of the whole event. This last process is called reconstruction. This chapter describes the details of the objects associated with physics analyses such as electrons, photons, muons, jets and missing transverse energy (E_T^{miss}); heavy-quarks and taus. These objects are available in the standard AOD (Analysis Object Data) data samples. They are produced using a wide variety of sophisticated reconstruction algorithms.

3.1 Determination of Beam Spot

The beam spot is the luminous region produced by the collisions of proton beams. The position and spread of the beam spot needs to be measured precisely because of a number of reasons listed below:

- Important input for physics, for example b-tagging, lifetime, etc.

[1]A reconstructed particle

- For pattern recognition (input for High Level Trigger (HLT) and offline reconstruction).

- Extraction of the tracking impact parameter resolution.

- Quick check of global alignment.

- Provide feedback to accelerator groups.

- Beam monitor.

  The algorithm used to calculate the transverse beam position is the so called $d0 - \phi$ algorithm [52]. This is a robust and fast χ^2 fit which needs 1000 good tracks to reach micron precision. Using a Log-Likelihood fit, the beam width can be extracted using 20000 good tracks.

3.2 Track Reconstruction

Track reconstruction primarily uses signals in the tracking detectors to measure tracks of elementary particles. The reconstruction process is divided into an initial phase where starting points for tracks are found ("Seed Finding") followed by the collection of all measurements corresponding to a track ("Pattern Recognition"). The final step consists of a fit of the collected measurements ("Final Fit") which produces tracks.

Currently, two different ideas have been realized in CMS software framework (CMSSW). The first uses both the Pixel and the Strip detectors for Seed Finding and an iterative process going from layer to layer of the tracker detectors taking into account multiple scattering for the Pattern Recognition step using Kalman Filter techniques such as Combinatorial Track Finder (CTF) [53]. This is the default Pattern Recognition algorithm in CMSSW. The second algorithm uses only the Strip Tracker for Seed Finding and opens up 'Roads' around the Seed Trajectories to collect measurements. Following both implementations, a final fit using Kalman Filter technique [54] produces the tracks. The modular structure followed during the implementation of the tracking in CMSSW allows the exchange of different components of the two implementations and opens up the possibility to easily integrate new ideas for reconstructing tracks. Recently, the default CTF reconstruction has been modified employing an Iterative Tracking approach.

'Local' Reconstruction

Local reconstruction transforms the digitized hits from the tracker into reconstructed hits the local coordinate system of the silicon sensors. This is performed in two stages:

- Clustering: forming clusters of adjacent pixels or strips above threshold

- RecHit conversion: giving the reconstructed hit position with the errors

The reconstructed hits are needed as input for the track reconstruction.

Tracking

The track reconstruction is CMSSW is done in four processes:

- **The seed finding:** 'Seeds' define the starting trajectory parameters and associated uncertainties of potential possible tracks. In the quasi-uniform magnetic field of the tracker, charged particles follow helical paths and five parameters are needed to define a trajectory. Extraction of these five parameters requires either three hits (pixel triplet), or two hits (pixel pairs, strip pairs and mixed pairs) and a constraint on the origin of the trajectory based on the assumption that the particle is originated near the beam spot. To limit the number of hit combinations, seeds are required to satisfy certain weak restrictions, for example, on their minimum transverse momentum p_T and their consistency with originating from the $p - p$ interaction region.

- **The pattern recognition:** The pattern recognition is based on a combinatorial Kalman filter method. The iteration proceeds iteratively from the seed layer, starting from a coarse estimate of the track parameters provided by the seed, and including the information of the successive detection layers one by one. On each layer, i.e. with every new measurement, the track parameters are known with a better precision, up to the last point, where they include the full tracker information.

- **Ambiguities resolution:** Ambiguities in track finding arise because a given track may be reconstructed starting from different seeds, or because a given seed may result in more than one trajectory candidate. These ambiguities, or mutually exclusive track candidates are resolved in order to avoid double counting of tracks. The ambiguity resolution is based on the fraction of hits which are shared between two trajectories. If this fraction exceeds a value of 0.5, the track with the least number of hits is discarded, or, if both tracks have the same number of hits, the track with the highest value of χ^2 is discarded.

- **The final fit:** For each trajectory, the building stage results in a collection of hits and in an estimate of the track parameters. However, the full information is only available at the last hit of the trajectory and the estimate can be biased by constraints applied during the seeding stage. Therefore the trajectory is refitted using a least-squares approach, implemented as a combination of a standard Kalman filter and smoother.

Refitting Tracks

Some cases require performing a new final fit over already reconstructed tracks. For example, in alignment studies, we refit Tracks after having modified the geometry of detectors. Refitting tracks is also useful if one needs access to their trajectory.

3.3 Vertex Reconstruction

Vertex reconstruction [55,56] usually involves two steps, vertex finding and vertex fitting [57]. Vertex finding involves grouping of tracks into vertex candidates. The vertex-finding algorithms differs in approach depending on the physics case such as primary or secondary vertex finding, reconstruction of exclusive decays, etc. Vertex fitting involves determining the best estimate of the vertex parameters such as position, covariance matrix, and track parameters constrained by the vertex position and their covariances for a given set of tracks, as well as indicators of the fit quality like total χ^2 , number of degrees of freedom, or track weights.

Pixel standalone track and vertex reconstruction is fast enough to be used at HLT. Fitting a single vertex from reconstructed tracks is a common step in the reconstruction of high-level physics objects such as tau's and B-mesons. Offline primary vertex reconstruction yields ultimate vertex position resolution as well as the best possible estimation of the primary vertex position error matrix.

3.3.1 Primary vertex reconstruction

The goal of primary vertex reconstruction [58] is the precise determination of the collision points in $p-p$ interactions. With the exception of early 2010 data, multiple collisions in the same bunch-crossing (pile-up) are non-negligible. Therefore, it is necessary to identify primary vertex among many associated with interesting physics, which we want to persue. In CMS the offline primary vertex reconstruction proceeds as follows

- reconstructed tracks are selected based on their compatibility with the beam spot, number of hits and fit quality,

- the tracks are clustered into several primary vertex candidates, according to the $z-$coordinate of the point of closest approach of the tracks to the $z-$axis,

- a 3-dimensional vertex fit is performed with the tracks of each primary vertex candidate, and

- primary vertex candidates compatible with the beam line are retained.

Although a large number of inelastic collisions can be present in each bunch crossing, there is typically only one signal vertex containing the event of interest. The primary vertex collection is sorted according to the sum of the p_T squared of the tracks associated to each vertex, such that the vertex with largest sum, likely to be the "signal" vertex, is selected.

3.4 Electron Reconstruction

Reconstruction of electrons are very important to the success of CMS physics program. Several SM signatures (W, Z, top quark etc. are good examples) and evidence for New Physics (SUSY, Z, etc.) can be observed by reconstructing decays that involve electrons and/or muons. In many cases, a decay involving electrons is rare. Taking Z boson decay as an example, only about 3% of Z bosons decay to a pair of electrons; the dominant decay of the Z boson is to hadrons. LHC collisions are dominated by QCD processes, so the bulk of the data is full of hadronic activity (i.e. jets). If we exploit some of the features of true electrons, we can isolate a sample of interesting events from the bulk of the data.

3.4.1 Supercluster reconstruction

Electron and photon showers deposit their energy in several crystals in the ECAL. Approximately 94% of the incident energy of a single electron or photon is contained in 3×3 crystals, and 97% in 5×5 crystals. Summing the energy measured in such fixed arrays gives the best performance for unconverted photons, or for electrons. The presence of material in front of the calorimeter results in bremsstrahlung and photon conversions. Because of the strong magnetic field the energy reaching the calorimeter is spread in ϕ. The spread energy is clustered by building a cluster of clusters, a "supercluster", which is extended in ϕ.

The reconstruction of electrons [59] in CMS starts by the reconstruction of clusters in the ECAL. The hybrid algorithm is used in the barrel and the multi 5x5 (Modified Island algorithm) is used in the endcaps. The second one is complemented by the recovery of clusters in a ϕ window around a seed cluster so that the superclusters formed efficiently collect the energy of radiated bremsstrahlung photons. Currently a ϕ window around the seed crystal of cone of radius 0.3 radian in both directions is used. The hybrid clustering algorithm internally performs the bremsstrahlung recovery, using a window of similar ϕ length.

3.4.2 ECAL Clustering

Hybrid superclusters

The $\eta - \phi$ geometry of the barrel crystals allows the Hybrid algorithm to exploit our knowledge of the lateral shower shape in the η direction (by taking a fixed bar of 3 or 5 crystals in η), while searching dynamically for separated energy in the ϕ direction. Unlike the Island algorithm (described next), the Hybrid algorithm cannot be used without making superclusters. The algorithm makes superclusters which are then decomposed into sub-clusters ("basic clusters").

Island basic clusters and superclusters

The Island algorithm starts with a seed crystal having a local energy maximum above a defined threshold. Starting from the seed position, adjacent crystals are examined, scanning first in ϕ and then in η. Along each scan line, crystals are added to the cluster until a rise in energy (meaning that the area with energies at the noise level or a crystal belonging to another cluster is reached) or crystal that has not been read out (i.e. without recorded energy) is encountered. For a crystal to be added to a cluster the following must be true:

- The crystal contains a rechit (reconstructed hit) with positive energy.
- The crystal has not been assigned to another cluster already.
- The previous crystal added (in the same direction) has higher energy.

In much the same way as energy is clustered at the level of the crystals, non-overlapping Island clusters can, in turn, be clustered into superclusters. The procedure is seeded by searching for the most energetic cluster and then collecting all the other nearby clusters in a very narrow η-window, and much wider ϕ-window.

3.4.3 Electron seeding

Superclusters are then used to select trajectory seeds built by the combination of tracker hits in the innermost tracker layers. The seeding algorithm make use of pixel layers and also tracker endcap (TEC) layers, which covers the forward region where the pixels start to be inefficient. The selection is made by matching supercluster's position with trajectory seeds using seed match windows in ϕ and z to match the two hits of the trajectory seeds (in case of triplets, at least two out of the three hits are required to be matched). This matching takes advantage of the fact that the energy-weighted average impact point of the electron and its bremsstrahlung photons is precisely where a non-radiating electron would have impacted the ECAL. The position measured by the supercluster can thus be used to predict the position of hits in the innermost tracker layers before which most of the electrons have not radiated significantly. This allows for a very efficient filtering and a similar procedure is used at HLT. The match windows are defined in electron seeding configuration. The seeds are further preselected using an H/E [2] threshold (0.1) and a threshold on traverse energy ($E_t > 4$ GeV). Such ECAL-driven electron seeding [60] strategy is very efficient for $p_T > 5$ GeV and for isolated electrons. At very low p_T, the ϕ recovery window starts to be too small (i.e. some electrons which radiate leads to photon clusters more separated from the electron cluster than 0.3 radian in the CMS magnetic field). Also, for the cases of electrons in jets, the energy collected in the supercluster might include some neutral contribution from the jets and therefore is not appropriate to seed electron tracks. For these reasons, the above seeding strategy is complemented by a tracker driven approach following electron pre-identification strategy developed for non-isolated electrons. In this approach, seeds are built from the general tracks which are followed as close as possible to the ECAL front face and matched with clusters. These two procedures are finally merged into a single collection of electron seeds, keeping track of the provenance of which algorithm has found the corresponding seed.

3.4.4 Electron tracking

All electron seeds are then used to initiate a dedicated electron tracking. It uses a modeling of energy loss for electrons in the search for compatible hits and for the final fit of the track parameters. The building parameters are similar to the ones used in the general tracking, apart from the use of a dedicated electron propagator, a very loose χ^2 threshold $<$ (200). and a higher p_T threshold $>$ (2 GeV). Trajectories are required to have at least 5 hits and at most 1 lost hit. Following Ref. [61], a gaussian sum filter (GSF) fit is applied to estimate the electron track parameters at each measurement point, modeling the Bethe Heitler probability distribution function for energy loss by a superposition of gaussian distributions. This procedure leads to multi-component trajectory states for each measurement point, with weights for each component describing the associated probability. Although more information is available, we usually make use of only two combinations for the final track parameters evaluation: the weighted mean of the components (so called "mean") and the component with highest weight (so called "mode"). It has been shown in Ref. [59] that the mode estimate is equivalent in terms of resolution to a Kalman filter fit on the part of the trajectory that do not radiate, while allowing to include all the points of the electron trajectory in the fit. Finally, the difference between the momentum magnitude at the outermost track position and the momentum magnitude at the innermost track position is an estimate of the true fraction of energy radiated by the electron. This variable, called f_{brem} (bremsstralung fraction), is used in the electron classification for

[2]ratio of energy deposition in HCAL to ECAL

the final determination of the electron momentum and is a key ingredient of electron ID (electron identification method). From the reconstruction of the GSF tracks, the GSF electron objects are built. The reconstruction is split into two steps: in a first step an object called ‘gsfElectronCore’ is built which encapsulates the track-cluster association, i.e. contains a reference to the GSF track and the associated supercluster. In the case of ECAL-driven seeds, the supercluster is directly obtained from the track seed. For the tracker driven electrons, the supercluster is built by matching the GSF track with all ECAL clusters, using tangents to identify corresponding bremsstrahlung clusters. When the electron seed has been found by the tracker driven approach, the supercluster found by this method is also attached to the electron object.

3.4.5 Electron Preselection

In the preselection step, loose geometrical and energy-momentum matching is imposed between the reconstructed electron track and the corresponding supercluster. Good energy-momentum matching is obtained by matching the corrected supercluster energy E_{corr} with the track momentum p_{in} taken at the closest position from the nominal vertex. The geometrical matching is performed by taking the track parameters at interaction vertex, η_{in} and ϕ_{in}, extrapolating to the ECAL assuming a perfect helix, and matching the resulting $\eta_{in}^{extrap.}$ and $\phi_{in}^{extrap.}$ to energy-weighted position of the supercluster, η_{sc} and ϕ_{sc}. The compatibility of the matched object with the expectations from an electron is reinforced by setting an upper threshold on the fraction of the supercluster energy collected in the hadron calorimeter (HCAL). Initially the preselection is made very loose in order to very efficiently reconstruct electrons and satisfy a large number of possible analyses. For the case of tracker driven only electrons, the multivariate discriminant provided by the Particle-Flow [62] is used to preselect electrons, with a threshold greater than 0.4. The selection of primary electrons is based on a further requirement on the impact parameter with respect to the nominal vertex. As the beam spot is much better defined in the transverse plane, the normalized transverse impact parameter is used to select primary electrons.

3.4.6 Ambiguous electron resolution

Once preselected, a further step is applied to remove ambiguous electrons that arise from the reconstruction of several conversion legs of radiated photon(s) from the primary electrons. With an up to $\sim 2X_0$ [3] (X_0 is radiation length) integrated amount of material budget in the tracker, electrons encounter in many cases several X_0 eventually followed by conversions, so that such situation is not marginal, i.e to say energy loss is significant. In

[3] the mean distance over which a high-energy electron loses all but $\frac{1}{e}$ of its energy by bremsstrahlung.

very rare cases two different superclusters are attached to the same track as having pointed to the same seed. All these cases have electrons sharing the same superclusters (or in few rare cases the same track) and are resolved. The ambiguity solving algorithms identify cases of common superclusters or tracks, classify the electrons according to innermost track hit position to favour primary electrons and keep the best combination according to this criteria. In case both ambiguous combinations have a track with first hit on the same tracker layer, the best E/p (ratio of energy deposited in ECAL to the momentum measured in tracker) combination is kept. Finally, in order to resolve the cases where the supercluster overlaps because one of the candidate is tracker driven only and therefore the Particle-Flow supercluster is used, a threshold in the energy shared between the two superclusters is used.

3.4.7 Electron Isolation and Identification

In the CMS experiment, most recent and widely used isolation method is Particle flow-based relative isolation, which is defined as

$$RelIso_{PF} = \frac{I_{\mathrm{CH}} + max(0, I_{\mathrm{NH}} + I_{\mathrm{PHOTON}} - (\mathrm{EA} \cdot \rho))}{E_{\mathrm{T}}},$$

where I_{CH}, I_{NH} and I_{PHOTON} are the charged hadron, neutral hadron and photon isolation variables (using an isolation cone of 0.3). The isolation is corrected for contributions from pile-up using charged hadron subtraction in the isolation cone using fastjet algorithm [63] and for neutral particles using the effective area correction, $(\mathrm{EA} \cdot \rho)$ where EA is the cone effective area and ρ is the average neutral particle density of the event. The PF charged particle isolation is obtained by summing the momenta of all particles inside the cone that are charged, imposing an extra requirement that the z coordinate of the track trajectory at the point of closest approach to the beamline is within 0.1 cm of the event primary vertex, which is intended to reject particles produced from pile-up events. The PF photon isolation is computed by summing the momenta of all photons inside the cone, but vetoing those PF photons which lie within a narrow strip in the coordinate defined by $|\delta\eta| < 0.025$.

In physics analysis, high identification efficiency is an advantage as it enhances signal selection, in particular at low transverse energy E_T where the background increases and the fake rate (or misidentification rate) is much higher. The following variables are used to discriminate between real and mis-identified electrons in the CMS experiment:

- supercluster energy / track momentum at vertex
- $\Delta\eta$ between supercluster position and track direction at vertex extrapolated to ECAL assuming no radiation

- $\Delta\Phi$ between supercluster position and track direction at vertex extrapolated to ECAL assuming no radiation

- Ratio of energy in HCAL behind supercluster to supercluster energy

- Energy in 3×3 crystals / energy in 5×5 crystals

- Energy of closest basic cluster to track impact point at ECAL / outermost track momentum

- Energy of closest basic cluster to track impact point at ECAL / innermost track momentum

- $\Delta\Phi$ between track impact point at ECAL and closest basic cluster

- 1/E(supercluster) - 1/p(track at vertex)

- Brem fraction = (track momentum at vertex - track momentum at ECAL)/(track momentum at vertex)

A selected subset of the above variables are combined to perform electron identification, using certain thresholds, or multivariate techniques such as a likelihood, or neural network.

In this thesis, electrons are reconstructed using GSF algorithm [64] described above, and are required to pass electron Identification thresholds according to a multivariate identification technique [65]. We also require that selected electron candidates are well isolated, and the technique which we use for that is particle-flow based isolation described in detail in section 3.4.7.

- Electron trasverse energy $E_{\mathrm{T}} > 30\,\mathrm{GeV}$.

- Pseudorapidity $|\eta| < 2.5$. There is an exclusion region due to the ECAL barrel-endcap transition region, defined by $1.444 < |\eta_{\mathrm{sc}}| < 1.566$, where η_{sc} is the pseudorapidity of the ECAL supercluster.

- Impact parameter: We impose a threshold on the absolute value of the impact parameter calculated with respect to the average primary vertex $\left(d_0^{PV}\right)$ to be less than 0.02 cm.

- In order to make sure that the selected electron and the selected jets come from the same hard interaction and not from pile up events, we require that the z coordinate of the PV of the event and the z coordinate of the electron's vertex lie within a distance of less than 0.1 cm.

- In order to reject events in which the electron candidate actually originates from a conversion of a photon into an e^+e^- pair, we use an approach using the vertex fit probability of fully reconstructed conversions combined with the requirement that the number of missed inner tracker layers of the electron track must be exactly zero (i.e. there are no missed layers before the first hit of the electron track from the beam line).

Lepton η	$\lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 1.479$	$1.479 < \lvert\eta\rvert < 2.5$
MVA threshold value (tight lepton)	0.913	0.964	0.899
Isolation threshold value (tight lepton)	0.105	0.178	0.150
MVA threshold value (loose lepton)	0.877	0.811	0.707
Isolation threshold value (loose lepton)	0.426	0.481	0.390

Table 3.1: Threshold values for electron identification multivariate discriminant.

Depending on the thresholds, electrons can be categorized into "tight" and "loose" ones. We consider electrons which have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 2.5$, and which satisfy electron $RelIso_{PF}$ and multivariate (MVA) identification thresholds. The threshold values for the electron identification and isolation used in the analysis can be found in table 3.1.

3.5 Photon Reconstruction

Photon reconstruction is very much similar to electron reconstruction except that it does not have tracks. Photons are built from ECAL super clusters, constructed using the corrected Hybrid super clusters algorithm in the ECAL barrel and the corrected Multi 5x5 algorithum in endcaps. Each supercluster is potentially a Photon candidate (threshold on super cluster $E_T > 5$ GeV ad H/E < 0.2), so that one photon object is created for every super cluster. Each photon object carries information about photon energy, position and information related to isolation in the Tracker, ECAL, and HCAL.

3.5.1 Converted Photon

Conversion signal mostly comes from low p_T π^0s and constitutes background to all analyses involving isolated photons. The relatively massive CMS Tracker ($2X_0$) leads to a large probability of photon conversion and electron bremsstrahlung radiation in the tracker volume. The constituent basic clusters are used as starting points for the inward conversion track finding, using the E_T of the basic cluster as an initial guess for the trajectory of the track. If a pair of matching hits are found in the last three layers of the tracker, they are used to re-evaluate the track parameters. Track finding then proceeds inward, considering the mean energy loss experienced by electrons in the tracker material. Once the inward ECAL-seeded tracks are built, their innermost hits are used as starting points for the other arm of the conversion. Inside out tracks are then built outwards, again considering the mean energy loss for an electron. The possible combinations of oppositely charged tracks are selected and combined to form conversion candidates. Photon conversions are characterized by a pair of oppositely charged tracks with small opening angle and consistent with zero invariant mass. The variables used to separate the conversion candidates from long-lived particles and vertices from mis-reconstructed tracks are the angular separation measured at the production vertex between the conversion electrons in the transverse plane ($\Delta\phi$), in the longitudinal plane ($\Delta \cot\theta$), and the χ^2 of the vertex fit. For signatures with photons and missing energy, another significant source of background are the bremsstrahlung photons emitted by high energy muons from cosmic rays or beam halo. The timing distribution of cosmic rays is uniform, while the beam halo is tightly correlated with the interaction time, making it more challenging to suppress once the instantaneous luminosity is high enough to produce multiple interactions per beam-crossing.

3.5.2 Photon Identification and Isolation

The variables which are used to identify photons are:

- *R9* : The ratio of energy contained in a 3×3 array of crystals centred on the crystal with the highest deposited energy to the total super cluster energy.
- *H/E*: Ratio of the energy deposited in the single closest HCAL tower to the supercluster inside a cone of radius 0.15 in the $\eta - \phi$ plane centered on the photon direction and the energy deposited in ECAL to that supercluster.
- $\sigma_{i\eta i\eta}$: The energy weighted standard deviation of single crystal η within the 5 crystals centered at the crystal with maximum energy.
- $\sigma_{i\phi i\phi}$: The energy weighted standard deviation of single crystal ϕ within the 5×5 crystals centered at the crystal with maximum energy.
- *Pixel seed veto*: There should not be any hit present in the pixel detector. This is mainly to distinguish between electron and photon.

3.6 Muon Reconstruction

The muon reconstruction chain starts with the "local reconstruction". First, hits in DTs, CSCs and RPCs are reconstructed from digitized electronics signals. Hits within each DT and CSC chamber are then matched to form "segments" (track stubs).

3.6.1 Standalone Muons

In the offline reconstruction, the segments reconstructed in the muon chambers are used to generate "seeds" consisting of position and direction vectors and an estimate of the muon transverse momentum. These initial estimates are used as seeds for the track fits in the muon system, which are performed using segments and hits from DTs, CSCs and RPCs and are based on the Kalman filter technique. The result is a collection of Track objects reconstructed in the muon spectrometer, which are referred to as “standalone muons”. To improve the momentum resolution, a beam-spot constraint can be applied in the fit.

3.6.2 Global muons

For each standalone muon track, a search for tracks matching it among those reconstructed in the inner tracking system (referred to as “tracker tracks”, “inner tracks” or “silicon tracks”) is performed, and the best-matching tracker track is selected. For each “tracker track” - “standalone muon” pair, the track fit using all hits in both tracks is performed, again based on the Kalman filter technique. The result is a collection of Track objects referred to as “global muons”.

3.6.3 Tracker Muons

An approach complementary to the global-muon reconstruction consists in considering all tracker tracks to be potential muon candidates and in checking this hypothesis by looking for compatible signatures in the calorimeters and in the muon system. Tracker tracks identified as muons by this method are referred to as “tracker muons”. The tracker-muon algorithm is particularly useful for the identification of low-pT muons (with p_T of the order of several GeV), which may not leave enough hits in the muon stations for a standalone muon to be reconstructed.

3.6.4 Muon Identification and Isolation

There are several categories of muon identification algorithms that have been developed:

- Cut-based identification for global muons, which consists of a set of track-quality requirements.
- Likelihood-based identification for tracker muons, which uses compatibility of the calorimeter response with the muon hypothesis and the presence of matched segments in the muon system.
- Cut-based identification for tracker muons, which selects muons on the basis of the track-penetration depth in the detector.

Isolation of a muon is done relative to measurements in the detector (tracks, calorimeter towers and hits) that can be viewed as a swarm of other objects. Isolation of a muon relative to a list of other physics candidates objects is not a trivial task in general. Highly flexible methods are provided in the CMSSW together with the configurations which are run by the standard reconstruction. The outputs of these methods are intended to be flexible enough to suite most analysis purposes.

In this thesis, muon candidates are identified by two of these above algorithms: one proceeds from the inner tracker outwards, the other one starts from tracks measured in the muon chambers and matches and combines them with tracks reconstructed in the inner tracker. These selection criteria are summarized below:

- The muon candidate is reconstructed both as a global muon and as a tracker muon.
- Number of pixel hits of the Tracker track ≥ 1,
- Number of muon system hits of the Global track ≥ 1,
- Normalized χ^2 of the Global track < 10.0,
- Muon $p_{\mathrm{T}} > 25\,\mathrm{GeV}$,
- Pseudorapidity $|\eta| < 2.1$,
- Impact parameter: We impose a threshold on the absolute value of the impact parameter calculated with respect to the primary vertex, i.e. we require $d_0(\mathrm{PV})$ to be less than 0.02 cm.,
- In order to make sure that the selected muon come from the same hard interaction and not from pile up events, we require that the z coordinate of the PV of the event and the z coordinate of the muon's inner track vertex lie within a distance of less than 0.5 cm,
- The number of tracker layers with hits from the muon track ($N_{\mathrm{layers}} > 5$) has to be,

The selected muon candidates also have to be isolated. Particle flow-based relative isolation for muons is defined as

$$RelIso_{PF} = \frac{I_{\mathrm{CH}} + max(0, I_{\mathrm{NH}} + I_{\mathrm{PHOTON}} - (0.5\; p_T^{sumPU}))}{p_T},$$

where I_{CH}, I_{NH} and I_{PHOTON} are the charged hadron, neutral hadron and photon isolation variables (using an isolation cone of 0.4). The charged hadron isolation variable uses tracks from the primary vertices only. The neutral hadron and photon isolation variables are corrected for contributions from pile-up. We require muon to have $RelIso_{PF} < 0.12$ in order to be considered isolated.

Additionally, to reject events with more than one lepton, we define a 'loose' muon, which has looser thresholds. We consider all global muons which have $p_{\mathrm{T}} > 10\,\mathrm{GeV}/c$, $|\eta| < 2.5$, and $\mathrm{RelIso_{PF}} < 0.2$ to be 'loose' muons.

3.7 Jets Reconstruction

Jets arise from the hadronization of colored (quarks and gluons)[4] particles. Colored quantum state particle does not exist as independent particle that is to say they always appear in an experiment as group of hadrons. Thus, a jet of hadrons can be considered as the final state signature of quarks or gluons produced in the hard scattering. Jets can be made of leptons, hadrons and even bosons (specifically the photon). These particles are usually collimated in a given direction, and one can draw a cone around them (like a shotgun blast!). So, to reconstruct jet object we can collect energy deposition in a material around a cone of any defined radius (dependent on particular algorithm). There are a number of algorithms in particle physics motivated by kind of jets we expect, such as $anti - k_t$ clustering algorithm, Cambridge-Aachen jet algorithm etc. Historically there are two types of mainstream jet clustering algorithm named as:

- Cone-Type jet clustering algorithm
- Sequential clustering algorithm

Jet Algorithums: The majority of jet algorithms at CMS use a so-called "clustering sequence". This is essentially a pairwise examination of the input four-momentum. If the pair satisfy some criteria, they are merged. The process is repeated until the entire list of constituents is exhausted. In addition, there are several ways to determine the "area" of the jet over which the input constituents lay. This is very important in correcting for pileup, because some algorithms tend to "consume" more constituents than others and hence are more susceptible to pileup. Furthermore, the amount of energy that is inside of a jet due to pileup is proportional to the area, so to correct for this effect it is very important to know the jet area.

On the basis of input information jet clustering algorithm uses, jets are further classified as Particle-flow jets (PFJets), MC generator jets (GenJets) etc. The jet algorithms takes input in the form of a set of four momentum. At CMS, the most popular jet type is the "Particle Flow Jet", which attempts to use the entire detector at once and derive single four-vectors representing specific particles. For this reason it is very comparable (ideally) to clustering generator-level four-vectors also.

Particle Flow Jets (PFJets): Particle Flow candidates (potential jet) combine information from various detectors to make the best combined estimation of particle properties. PFJet is made from PF Candidates and contains information about contributions of every particle class: Electromagnetic/hadronic, Charged/neutral etc.Jet response is high. These jets p_T resolution is quite good overall: It varies from 15–20% at low p_T ($\sim$ 20 GeV) and asymptotically reaching 5% at high pT.

[4]color a quantum number

Monte Carlo Generator-level Jets (GenJets): GenJets are pure Monte Carlo simulated jets. Useful for analysis with MC samples. Generated by clustering energies of Monte Carlo particles. May include "invisible" particles (muons, neutrinos, ..) Jet response is 1, i.e., the jet energy scale is 1. Resolution is perfect, by definition. GenJets include information about the 4-vectors of the constituent particles i.e. hadronic and electromagnetic components of the energy etc.

3.7.1 Jet Identification

Pile-Up Jet Identification

The LHC is running at such high intensities that multiple $p - p$ collisions per bunch interaction occur with high probability. In such case, one is concerned about identifying and reconstructing a single primary collision where a physics event of interest occurs amongst the background of the additional $p - p$ collisions. Such backgrounds are due to processes that occur with high probability like low-p_T jets production. These additional collisions are known as pileup (PU). The rate of pileup is quoted in units of the number of additional collisions. The 2012, with the p-p collision at $\sqrt{s} = 8$ TeV, it recorded an average pileup rate of 23 additional collisions, with some events exhibiting well over 40 pileup collisions. In the CMS detector, some of its sub-detectors reads data in an extended window from the time of the current collision. This allows for pileup from both previous and following proton bunches to affect the reconstructed event. This effect is termed as out-of-time pileup (as opposed to in-time-pileup).

Due to the fact that pileup jets primarily come from overlapping jets incurred during pileup interactions, pileup jets show two characteristic features: first, they are both diffused and, wherever charged particle identification is possible, some fraction of the charged particles does not point to the primary vertex. These characteristics allow the identification of pileup jets in both regions, where charged particle tracking is available and regions where jet shape identification is possible. Both vertex and shape information can be combined through a multivariate analysis technique, to give a single discriminator targeting the identification of pileup jets. This technique is known as the pileup jet identification (PF Jet ID). The PFJet ID rejects mis-identified, badly reconstructed and noise jets while retaining 98-99% of real jets. Taking into consideration the correlations and the discriminating power of all the PF jet variables, as well as the importance of variables for possible HCAL and ECAL noise rejection, the following three sets of criteria were reconstructed named as Loose, Medium and Tight Ids. The identification of jets from pile-up relies on three types of propeties of the jets:

- inside the tracker acceptance, the trajectories of tracks associated to the jets can be used to establish the compatibility of the jet with the primary interaction vertex.

- the topology of the jet shape can be used in order to disentangle jets arising from the overlap of multiple interactions from truly hard jets.
- the object multiplicity can be used as an additional criteria.

3.7.2 Jet Energy Resolution

Jets are stochastic objects. The content of jets fluctuates quite a lot, and the content also depends on what actually caused the jet (light quarks (u,d or s), gluons, etc). In addition, there are also experimental limitations to the measurement of jet energies. Both of these aspects limit our accuracy to which we can measure the four-momentum of a jet.

3.7.3 Jet Energy Correction

The calorimeter response to particles is not linear and therefore it is not straight-forward to translate the measured jet energy to the true particle or parton energy. The jet corrections are a set of corrections that allows the proper mapping of the measured jet energy deposition to the desired level (particle or parton) energy, also called as true energy. In the following, the CMS strategy for jet energy corrections is described:

- **Factorized Approach:** CMS has adopted a factorized solution to the problem of jet energy corrections, where each level of correction takes care of a different effect. Each level of correction is essentially a scaling of the jet four momentum with a scale factor (correction) which depends on various jet related quantities (p_T, η, flavor, etc.). The levels of correction are applied sequentially (the output of each step is the input to the next) and with fixed order.
- **L1 Pile Up:** The goal of the L1 correction is to remove the energy coming from pile-up events. In principle this will remove any data dependence on luminosity so that the following corrections are applied upon a luminosity independent sample data.
- **L2 Relative Jet Correction:** The goal of the L2 Relative correction is to achieve uniform jet response as a function of η. Essentially, the uniformity in η is achieved by correcting a jet in arbitrary η relative to a jet in the central region ($|\eta| < 1.3$). The Relative correction is derived either using MC truth (parton information) or by employing a data driven method (dijet balance).
- **L3 Absolute Jet Correction:** The goal of the L3 Absolute correction is to achieve uniform jet response as a function of p_T. Once a jet has been corrected for η dependence (L2 relative correction), it is corrected back to particle level (this means that the p_T of the corrected jets reconstructed from calorimeter is equal on average,

to the MC generator level p_T). The derivation of the Absolute correction is done either by using MC truth information or by employing data driven techniques (Z/γ+jet balance).

The L2+L3 corrections scale the energy of an “average QCD jet” back to the energy of the corresponding generator level particle jet called true jet. However, an analysis of the individual jet flavors (u, d, s, c, b, gluon) shows that different corrections are needed for different jet flavors. This leads to an over or under correction if the L2+L3 corrections are applied to jets with a flavor composition different than that of QCD jets. For example, jets from the hadronic decay of a W boson typically consist of u, d, s and c quarks, which have a harder energy response than b and gluon jets. Consequently, applying L2+L3 corrections to these jets will result in an overcorrection. A first pass of flavor-specific corrections is provided as a a tool to minimize the flavor-dependence of the L2+L3 corrections.

- **L2L3Residuals: The Residual Calibration of Data:** After the first collision data at $\sqrt{s} = 7$ TeV, it was realized that the CMS jet energy response simulation is very successful. However, the comparison between data and MC is not perfect but there are small differences, up to 10%, depending on η. Therefore, the following strategy was adopted: first, the MC truth L2L3 JEC calibration is applied, which takes care of the bulk of the energy response. Second, a small residual calibration (η and p_t dependent) is applied which fixes the small differences between data and MC.

By definition, the residual calibration is applied to data only. The MC truth calibration is good enough for the MC. The need for a residual calibration will remain until CMS develops a perfectly tuned simulation that will reproduce the data features out of the box. The approach of a residual calibration, instead of a full-blown data driven one was found to be more effective in reducing the systematic uncertainties. Currently there is a need for a residual calibration only for the relative energy scale. The absolute energy scale seems to be modeled very well in the simulation.

In this thesis, jets are reconstructed with the $anti-K_T$ algorithm [66]. The algorithm start from the set of objects reconstructed by the particle flow [67–69]. Jets are corrected such that the measured energy of the jet correctly reproduces the energy of the initial particle. In this analysis we use jets with measured (corrected) p_T greater than 30. We require $|\eta| < 2.4$ so that the jets fall within the tracker acceptance. Jets from pile-up are identified and removed with Pileup Jet ID, as described in section 3.7.1.

To reduce the contribution of jets from top quark decays, we require that each jet that falls within the tracker acceptance should not be b-tagged as defined by the combined secondary vertex (CSV) ‘loose’ b-tagging working point (described later).

In order to account for electron and muon objects that have been reconstructed as jets, we remove from the jet collection any jet that falls within a cone of radius $R = 0.5$ of a "loose" electron or a "loose" muon. This "cleaning" procedure is applied in order to ensure that the same particle is not double counted as two different physics objects.

3.8 Missing Traverse Energy Reconstruction

Missing transverse momentum (E_T^{miss}) is the imbalance in the transverse momentum of all visible particles, i.e. particles which interact with the electromagnetic or strong forces, in the final state of collisions. Because momentum is conserved in each direction, E_T^{miss} is the transverse momentum that must have been carried by something invisible which is basically neutrinos in the CMS detector. Therefore, E_T^{miss} is an estimate of transverse momentum of neutrinos. We use E_T^{miss} in measurements of W bosons, top quarks, tau leptons etc. as these particles can decay into neutrinos. Further, many models of physics beyond the SM predict the existence of particles or something else which are invisible and can carry momentum; for example, Dark Matter models, super-symmetric models, unparticle models, and models with large extra dimensions. For this reason, we use E_T^{miss} to test such models. Accurate reconstruction of E_T^{miss} is demanding because it entails reconstruction of all visible particles in an event with precision. This requires a hermetic detector which can detect all particles having electromagnet or strong interaction with matter. With its highly granular ECAL, hermetic HCAL, redundant muon systems, and all silicon trackers in a strong magnetic field, the CMS detector meets the requirement.

In this thesis, a threshold on $E_T^{miss} > 30$ GeV is chosen.

3.9 b-Tagging

Jets reconstructed from bottom-quark (b-quark) hadronization (b-jet) are present in many physics processes, such as in top quark and the SM Higgs boson decay. Various new particles predicted by supersymmetric models and other theories can also decay into b-quark. b-tagging is a multivariate technique based algorithm which associates a single, real number as a discriminator - with each jet. The b-quark initiated jets will always tend to show higher value of the discriminator compared to light quarks (u,d,s). The methods for b-tagging are based on the unique features of b-jets. These are explained below:

1. Hadrons containing b-quarks have sufficient lifetime that they travel some distance before decaying. On the other hand, their lifetimes are not so high as those of light quark hadrons, so they still decay inside the detector. The advent of precision silicon detectors within particle detectors has made it possible to identify particles that originate from a different place where the b-quark was formed (for example the beam–beam collision point in a particle accelerator), and thus indicating the likely presence of a b-jet.

2. The b-quark is much more massive than anything it decays into. Thus its decay products tend to have higher transverse momentum (momentum perpendicular to the original direction of the bottom quark, and therefore of the b-jet). This causes b-jets to be wider, have higher multiplicities (numbers of constituent particles) and invariant masses, and also contain low-energy leptons with momentum perpendicular to the jet. These features can be measured, and jets that have them are more likely to be b-jets.

The b-tagging algorithms and their study are based on the measurement of three main variables:

1. The impact parameter significance of the tracks,

2. the position of the secondary vertex, and

3. the transverse momentum of the muon relative to the jet direction.

The impact parameter (IP) is defined as the distance between the track and the primary interaction vertex (PV) at the point of closest approach. The IP is positive (negative) if the track is produced downstream (upstream) with respect to the PV along the jet direction. The IP is calculated in 3-dimensions due to excellent 3-dimensional resolution provided by the silicon pixel detector. [An important feature of the IP is that it is Lorentz invariant and due to the b-hadron lifetime the typical IP scale is set by $c\tau \sim 480\mu m$]. In practice, the impact parameter significance $IP/\sigma_{(IP)}$ is used in order to take into account resolution effects. Due to the long lifetime of the b-hadrons, the IP from b-jets is expected to be mainly positive, while for the light jets it is almost symmetric with respect to zero. The output of each algorithm is a discriminator value on which the user can choose a threshold to select different regions in the efficiency versus purity phase space. Various b-tagging algorithm employed in the CMS detector are:

- **Track Counting algorithm:** This is a very simple 'tag', exploiting the long lifetime of B-hadrons. It calculates the signed impact parameter significance of all good tracks, and orders them in order of decreasing significance. Its b-tag discriminator is defined as the significance of the n'th track. It comes in two variations for N = 2 (high efficiency) or N = 3 (high purity).

- **Jet Probability algorithm:** This is a more sophisticated algorithm, also exploiting the long lifetime of B hadrons. Its b-tag discriminator is equal to the negative logarithm of the confidence level that all the tracks in the jet are consistent with originating from the primary vertex. This confidence level is calculated from the signed impact parameter significances of all good tracks.

- **Soft Muon and Soft Electron:** These two algorithms tag b-jets by searching for the lepton from a semi-leptonic B-hadron decay, which typically has a large $P_{t_{rel}}$ [5] with respect to the jet axis. Their b tag discriminators are the output of neural nets based on the leptons $P_{t_{rel}}$, impact parameter significance and a few other variables. For each of these taggers, we want to have a simple one (cutting basically only on the presence and P_t of the lepton), and a complex one using also jet quantities to compute a multivariate discriminate. This is already the case for muons: Electron taggers need refinement and are not really usable.

- **Simple Secondary Vertex algorithms:** These class of algorithms reconstruct the B hadron decay vertex using an adaptive vertex finder, and then uses variables related to it, such as decay length significance to calculate its b tag discriminator. It has been found to be more robust to tracker misalignment than the other lifetime-based tags.

- **Combined Secondary Vertex algorithm:** This sophisticated and complex tag exploits all known variables, which can distinguish b-jets from all non-b jets. Its goal is to provide optimal b-tag performance, by combining information about impact parameter significance, the secondary vertex and jet kinematics. The variables are combined using a likelihood ratio technique to compute the b-tag discriminator. A variant of this tagger combines the variables using the multivariant discriminator.

Varying the threshold on the discriminator, we obtain different efficiencies of the taggers. We establish standard operating points as, loose (L), medium (M), and tight (T), being the values at which the tagging of u, d, s and gluon initiated jets are estimated from monte-calro (MC) to be 10%, 1%, or 0.1%, respectively, for jet transverse momentum of about 80 GeV.

3.10 Tau Tagging

Because the τ lepton is the heaviest of the three leptons, specific final states involving τs are expected to show up in the SM, and would appear abundantly in many processes arising from new physics beyond the SM. In about two thirds of cases τ leptons decay hadronically, typically into either one or three charged mesons (predominantly π^+, π^-) in presence of up to two neutral pions, decaying via π^0 to $\gamma\gamma$. Similar signature is expected for generic QCD jets production. Since the cross–section of jet production exceeds the cross–section of τ lepton production by several orders of magnitude, the experimental challenge in reconstructing and identifying hadronic τ decays is to discriminate efficiently between genuine τ hadronic decays and quark/gluon initiated jets misreconstructed as τ candidates.

[5] lepton transverse momentum p relative to the charged jet axis

- **The Track corrected Tau referred as "TCTau":** This algorithm aims to identify τ lepton hadronic decays considering jets reconstructed using tracks and energy deposits in the ECAL and HCAL.
- **Particle-Flow τ tagging:** 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 in a signal cone of radius ΔR = 0.15 around the direction of the highest P_T (leading) particle in the jet, reconstructed using the anti–kT algorithm in a cone of size 0.5. The leading particle should be within a cone of radius ΔR =0.1 around jet direction and is required to have p_T above 5 GeV. The tau isolation requirement is based on particle counting in an isolation annulus of outer radius ΔR = 0.5 around the leading particle. The isolation annulus is defined by a narrow signal cone surrounded by a wider isolation cone and is required to contain no charged hadrons with $p_T \geq 1.0$ GeV and no photons with $p_T \geq 1.5$ GeV.
- **Hadron plus Strips:** The most popular algorithm in the CMS is Hadron plus Strips (HPS) tau identification algorithm, which is based on charged hadrons and neutral electromagnetic objects ("photons"), in a way very similar to the cut based particle flow isolation tau identification algorithm. The HPS algorithm starts from a particle–flow jet and searches for tau lepton decay products produced by various hadronic decay modes. As π^0s are produced very often in hadronic τ decays, one focus of the HPS τ identification algorithm is an optimized π^0 reconstruction. Special attention is given to photon conversions in the CMS tracker material, which typically result in broader calorimeter signatures for π^0s in the azimuthal direction, due to the bending of electron/positron tracks in the 3.8 Tesla solenoidal magnetic field of the CMS detector.

3.11 The data flow, from detector to analysis

To enable the most effective access to CMS data for end user, the data are first split into Physics Datasets (PDs) and then the events are filtered. The division into the PDs is done based on the trigger decision. The PDs are structured and placed to minimize the need of an average user to execute analysis on very large amounts of data. However, the PDs are still too large to make direct access by users reasonable or even feasible. The main strategy in dealing with such a large number of events is to filter them, and do that in layers of ever-tighter event selection. The process of selecting events and saving them in output is called 'skimming'. The intended *modus operandi* of CMS analysis groups is the following:

The PDs and skims are produced; they are defined using the trigger information and produced centrally on Tier 1 systems. The secondary skims are produced by the physics groups (say Higgs group) by running on the primary skims; the secondary skims are usually produced by group members running on the Tier 2 clusters assigned to the given group optionally. The user then skims once again, applying an even tighter event selection.

The primary skims (step 1 above) reduce the size of the PDs in order to reduce the time of subsequent layers of skimming. The target of the primary skims is a reduction of about a factor of 10 in size with respect to the PDs.

The secondary skimming (step 2 above) must be tight enough to make the secondary skims feasible in terms of size. And yet it must not be too tight since otherwise some information may be lost for certain analyses. However, in this case what is 'tight' is analysis-dependent, so it is vital for the group members to be involved in the definition of their group's secondary skims.

The user selection (step 3) is made on the Tier 2 by the user, and it's the main opportunity to reduce the size of the samples the user will need to deal with. In many cases, this is where the preliminary event selection is done, and thus it is the foundation of the analysis. It is expected that the user may need to re-run this step (for example, in case of finding out that the thresholds were too tight), but this is not a problem since the tertiary skims are being run on the secondary skims which are already reduced in size.

That being said, it is important to tune the user's skim to be as close to 'just right' as possible: the event selection should be looser than they are expected to be after the final threshold optimization, but not too loose – otherwise the skimming would not serve its purpose.

If someone wants a
sheep,
then that means that he
exists.

- ANTOINE DE SAINT-EXUPERY

Chapter 4

Search for the Standard Model Higgs Boson

The Standard Model (SM) of electroweak interactions [1–3] relies on the existence of the SM Higgs boson, H, a scalar particle associated with the field responsible for spontaneous electroweak symmetry breaking [5–10]. The observation of a Higgs boson with a mass of 125 GeV [70–72] is consistent with the theoretical constraint coming from the unitarization of WW scattering at high energies [12–21]. However, there is still a possibility that the newly discovered particle has no connection to the electroweak symmetry breaking mechanism [73, 74]; thus, it is important to continue searching for the SM Higgs boson in the high mass regime to further confirm the properties of the new boson. In addition, several attractive extensions to the SM, such as general two-Higgs-doublet models [75] or models in which the SM Higgs boson mixes with a heavy electroweak singlet [30–43], predict the existence of additional resonances at high mass, with couplings similar to the SM Higgs boson.

For Higgs boson masses above or near the threshold for decay into W/Z bosons, the decay modes of choice are dominated by those decays because of their large branching fractions. It is clear that the events where one W boson decays leptonically, which provides the main trigger elements, while the other decays hadronically, have the second highest branching fraction and have a reconstructable Higgs boson mass peak [76].

This chapter contains the analysis that sets a limit on the Higgs boson cross-section based on this decay mode, performed on data acquired by CMS at $\sqrt{s}$ = 8 TeV during the year 2012 in the mass range 170-600 GeV. The analysis selects events with one well-identified and isolated lepton, large missing transverse energy and at least two high p_T jets. Therefore, the main experimental challenge is to control the large W+jets background that the advantages of using this final state are realized.

A discussion about the data samples used in the analysis and the trigger selections are presented in sections 4.1 and 4.2. The physics objects reconstruction and all pre-selections details are already discussed in chapter 3.

After the pre-selections, the signal-over-background ratio is enhanced by means of a selection on a multivariate discriminant (MVA), designed in order to control the background while preserving as much as possible the difference in shape with respect to the signal (section 4.8). The input variables of the MVA exploit the decay angles of the four-body mass system (e, μ, ν, two-jets), and the kinematics of the entire four-body system. The MVA variable definition is optimized with dedicated trainings for 12 Higgs mass hypothesis (170, 180, 190,200, 250, 300, 350, 400, 450, 500, 550 and 600 GeV) case and for each lepton flavour (e, μ). In this way, 24 different configurations are obtained.

The main background process contaminating the signal region is W+jets; its contribution is determined from data. Minor background processes include $t\bar{t}$, single top, and diboson, and are determined using Monte Carlo(MC). The procedures for extracting the di-jet invariant mass m_{jj} and WW system invariant mass $m_{\ell\nu jj}$ shapes and normalizations for these MC backgrounds are described in sections 4.11 and 4.12.

The systematic uncertainties present in the signal description are discussed in section 4.13. section 4.14 describes the obtained limits on the SM Higgs cross section, as well as limits on the heavy Higgs coupling and branching ratios in the Beyond Standard Model (BSM) framework, and section 4.15 summaries this work.

4.1 Signal and backgrounds

4.1.1 The signal

The signal is characterized by the presence of a well isolated lepton (e or μ), large missing energy E_T^{miss} and a pair of jets with high traverse momentum (p_T). For the search of the SM Higgs boson the most dominant mode of the Higgs boson production at LHC, gluon-gluon fusion is considered. The analysis is optimized to search for the Higgs produced through this mode. However, the next dominant mode of the SM Higgs boson production at LHC, vector-boson fusion production contribution is also added up in analysis though it is not optimized separately for that topology. This analysis search for the SM Higgs boson in the mass range 170-600 GeV. In the mass range below 170 GeV, its is not sensitive and is limited by huge QCD background, so much so that we do not trust the reconstructed jets as genuine jets and we need better techniques to deal with it. The SM Higgs production cross-section for various production modes are shown in figure 4.1 (on left), and branching ratio of Higgs boson into the four fermions final states are shown in figure 4.1 (on right) as a function of Higgs boson mass. As it is visible the cross-section times branching ratio in the final state $\ell\nu jj$ is second highest throughout in the higher mass region. The inclusive cross-sections used for the Higgs boson signal at center-of-mass energy of 8 TeV have been calculated by the Higgs Cross Section Working Group [77], [28] for the gluon-gluon fusion (ggF) and vector boson fusion (VBF) processes. All cross-section values are taken from

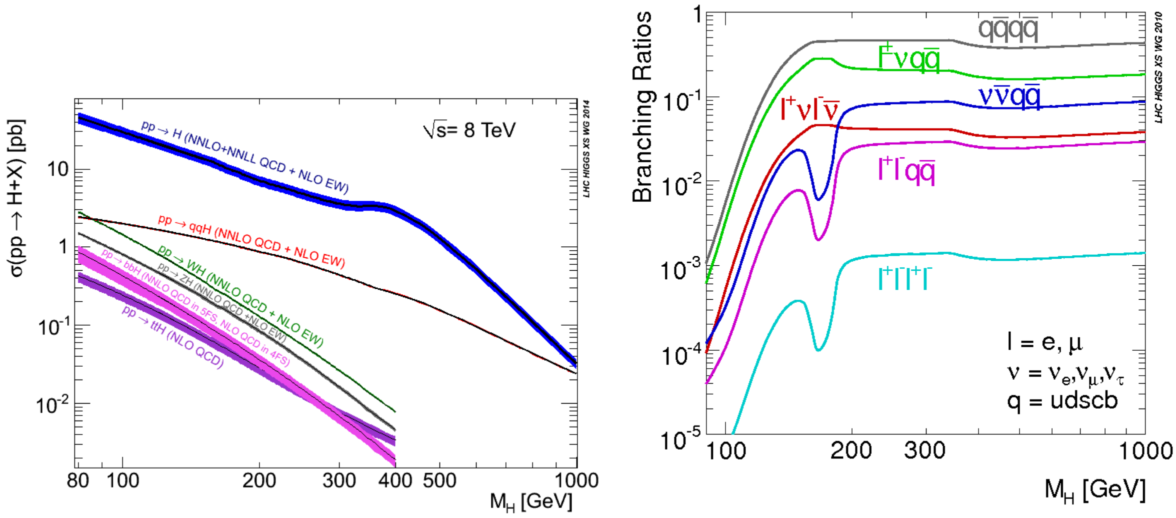


Figure 4.1: SM Higgs boson production cross sections at center-of-mass energy of 8 TeV (left) and branching ratio of four fermion final states (right).

Ref. [78].

4.1.2 The backgrounds

All the SM processes with final state having one lepton, two jets and E_T^{miss} are considered as possible sources of backgrounds, which mimic our signal. The most relevant ones are:

- W+jets: this is the production of single W vector boson in association with quarks or gluons. The leading order (LO) Feynman diagram is shown is figure 4.2. Because of its huge cross-section compared to our signal, this is by far the most important background to the analysis.

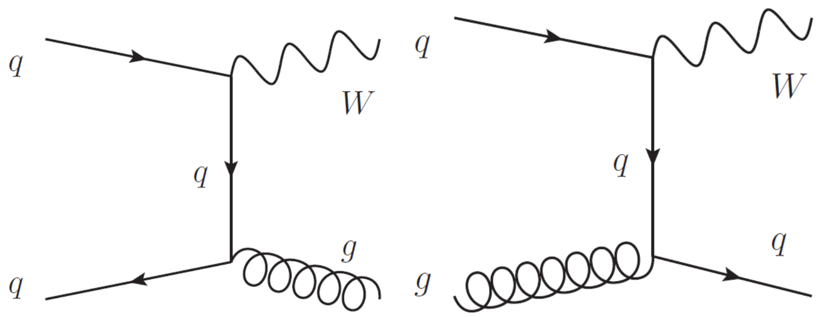


Figure 4.2: W+jets production Feynman diagram at LO.

- Drell-Yan (Z/γ^*+jets): this is the production of single Z/γ^* bosons in association with quarks or gluons, where one lepton from the decay of Z/γ^* boson gets undetected because of acceptance or inefficiency effects. The Feynman diagram, at LO is shown in figure 4.3.

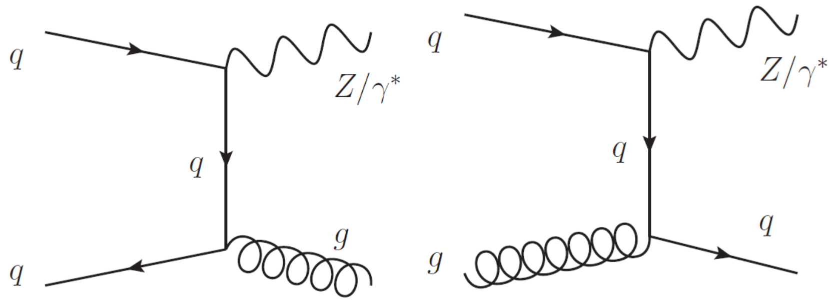


Figure 4.3: Z+jets production diagrams at LO.

- WW: this non-resonant production is an irreducible background for the analysis.
- WZ: in case the Z boson decays hadronically, or otherwise the W boson decays hadronically and one of the Z boson lepton is not identified by the detector, this sample also contributes to the backgrounds.
- ZZ: in case one Z boson decays hadronically, and one lepton is not identified by the detector, this sample contributes to the backgrounds. Feynman diagrams for production of these processes are shown in figure 4.4.

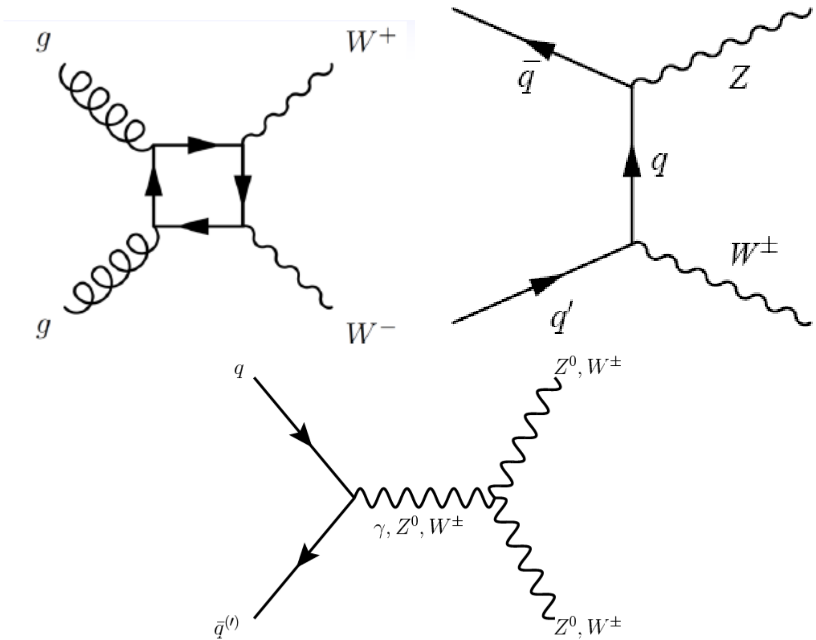


Figure 4.4: WW, WZ and ZZ production Feynman diagrams at LO.

- $t\bar{t}$: top quark pairs are produced at the LHC via the gluon fusion process $gg \to t\bar{t}$ or via QCD quark annihilation $qq \to t\bar{t}$. The semi-leptonic sample, where one W boson decays hadronically, can be reduced by applying a threshold on the number of jets, while the fully leptonic one is reduced by requiring only one good lepton in the event. In any case, because of acceptance and inefficiencies, this background still contaminates the signal phase space. The Feynman diagram showing their production at LO is shown in figure 4.5.

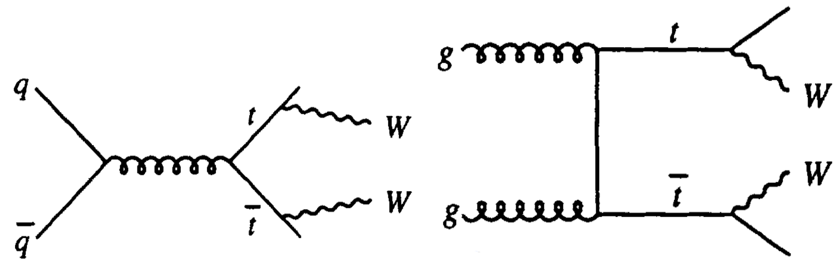


Figure 4.5: $t\bar{t}$ pair production Feynman diagrams at LO.

- Single top quark production: it proceeds through three separate channels:

  1. t-channel: top quark is produced after a quark-gluon interaction with the exchange of a virtual W boson.
  2. s-channel: top quark is produced in association with an anti-bottom quark, after the annihilation of a pair of quarks in a weak vertex.
  3. tW-channel: top quark is produced in association with a charged vector boson in a weak process, from a gluon-bottom quark pair in the initial state. The Feynman diagram representing their LO production at LHC are shown in figure 4.6.

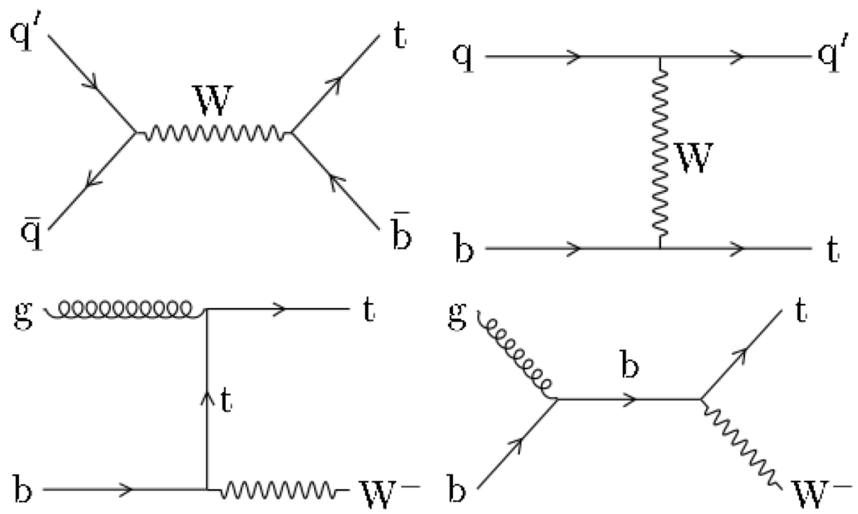


Figure 4.6: Single top quark production Feynman diagrams at LO.

- QCD multi-jet events generate a background because of the non-negligible probability of jets to be reconstructed as leptons.

The cross section for the backgrounds, multiplied by the branching ratio, are reported in Table 4.1. Each different sample notation includes all the decays taken into account in the effective cross section calculation.

Channel	Cross-section (pb)
W+jets	36257
Z+jets	3503
WW	57.1
WZ	32.3
ZZ	8.3
$t\bar{t}$+jets	225.2
$t/\bar{t}$+jets (t-channel)	85.5
$t/\bar{t}$+jets (s-channel)	5.65
$t/\bar{t}$+jets (tW-channel)	22.4

Table 4.1: The cross section for the backgrounds, multiplied by the branching ratio when meaningful. Each different sample notation includes all the decays taken into account in the calculation.

4.2 Datasets

4.2.1 $p-p$ Collision Data

The data sample used in this analysis was recorded by the CMS experiment in 2012. Only certified good runs and luminosity sections are considered, which means that a good functioning of all the CMS sub-detectors is required. The total statistics analyzed correspond to an integrated luminosity of $19.3 fb^{-1}$ for muons and $19.2 fb^{-1}$ for electrons.

Data were collected using high-level [1] single lepton triggers: a single muon trigger with a p_T threshold of 24 GeV and a single electron trigger with a p_T threshold of 27 GeV.

4.2.2 Simulated Data samples

SM Higgs boson samples, as well as samples for a large variety of electroweak and QCD-induced background sources, have been generated and showered using different MC generators. To better reproduce the actual data-taking conditions, where there is a significant probability that more than two protons interact in the same bunch crossing, pile-up (PU) events are added on top of the hard scattering. Particle interactions with the detector were reproduced through a detailed description of CMS.

[1] A High level trigger is the trigger applied offline at the analysis level.

The POWHEG generator [79–82] has been used to produce signal events, and the showering has been performed with PYTHIA6 [83]. For this analysis, samples with Higgs mass hypotheses ranging from 170 to 600 GeV have been used.

4.3 Reweightings

4.3.1 Pile Up reweighting

Multiple $p-p$ collisions in a single bunch crossing is termed as pile-up. When MC samples gets generated, they are simulated with expected number of pile-up interactions in intended data taking scenario. However, when actual data get recorded, it has been found that the number of pile-up interactions are not the same as expected. There could be a number of reasons behind this, such as data taking conditions change than what is planned. In such case one needs to correct the simulated sample to what is observed in the data. The procedure to do that is based on calculating a weight factor event by event, which is basically ratio of number of pile-up interactions observed in data to the number of pile up interaction with which MC samples were simulated. Then all (signal and backgrounds) MC samples are scaled by this weight factor.

4.3.2 Higgs Line-shape reweighting

The SM Higgs signal samples generated with POWHEG generator assume Higgs production as on shell but sampled with Breit-Wigner distribution with fixed width or running width approximation. This naturally needs correction. To correct the Higgs signal from this inefficiency a complex pole reweighting scheme (CPS) [84–86] is employed. This procedure gives each signal event a new weight to correct for the deficiencies in the POWHEG generation. These weights are cross section neutral and computed on an event by event basis for the signal MC samples only.

4.3.3 Reweighting due to Heavy Higgs Interference effect

The SM process $gg \to WW$, which is a background for this analysis interfere with our signal $gg \to H \to WW$. With increasing Higgs boson mass, the interference effect becomes more and more substantial between the signal and the continuum background. As shown in [87], the effect of this on LO total cross section is already over 30% for M_H 600 GeV or above. Further, it changes the M_{WW} spectrum substantially, with constructive behavior for M_{WW} less and close to M_H and destructive for M_{WW} greater and close to M_H. So, one needs to correct that by reweighting the signal's M_{WW} distributions. For more Appendix B.1 can be refereed which describes in further detail the reweighting procedure for this effect.

4.3.4 Beyond the SM reweighting

This analysis intends to interpret the results in another model called EW Singlet Model (described in section 1.5.5) besides the SM Higgs boson, for that there are no specific signal samples as such available. However, one can do that by following a procedure in which the Higgs line-shape is reweighted as per expected cross-section for EW Singlet model resonance. This is worked out on the same line as was done in previous high mass (600 - 1000 GeV) Higgs boson search analysis [88] at $\sqrt{s} = 8\ TeV$. We consider now the reweighting necessary for Model 1 as discussed in section 1.5.5. Given the new benchmark parameters of the BSM model, it is necessary to reweight the line shape of the heavy Higgs to arbitrary width. Therefore, we propose to modify the reweighting scheme for the SM case given in the previous section. First we reweight the CPS line-shape, before the inclusion of interference reweighting with $gg \to WW$, to arbitrary width. Then we apply the interference reweighting, which has been modified to include effects from a narrower width resonance with modified couplings.

Currently we lack a numerical solution to reweight the CPS from the SM width to an arbitrary width, so we take an analytical solution. We fit the SM CPS line-shape with a Breit-Wigner function with a running width. The shape is given by Eq.(4.1).

$$F_{BW}(x) = \frac{s\Gamma/m_H}{(s - m_H^2)^2 + (s\Gamma/m_H)^2} \tag{4.1}$$

where $s = m_{WW}^2$. The fitted width is then scaled by a factor $(C'^2)/(1 - BR_{\text{new}})$.

4.4 Selection Criteria

To perform physics analysis on experimental data, we always need to choose loosely selected sample of data, which in all probability, should contain the signal (if it exist in nature). Such details of all selections along-with their reconstruction, identification and isolations are described in great detail in Chapter 3. We term selections at this level as "pre-selection" since we always devise few more selections applied on top of these before final result is prepared and a conclusion is drawn. Pre-selections on physics objects are given in Chapter 3.

4.4.1 Event selection

On top of those pre-selections on each physics object [2], we apply event level selections. Each passing event should have a good primary vertex (PV). This means selecting the primary vertex with the highest sum of p_T^2 of the tracks associated with it and requiring it to have a number of degrees of freedom (ndof) ≥ 4, where ndof corresponds to the weighted sum of the number of tracks used for the construction of the PV. In addition, the PV must lie in the central detector region of $abs(z) \leq 24$ cm and $\rho \leq 2$ cm around the nominal interaction point where z and ρ are the longitudinal and transverse distance of PV from the beam spot, respectively.

4.4.2 Event categorization

Two exclusive samples are chosen on the basis of lepton flavour. They are e+jets category and μ+jets category.

In the e+jets category, we select events which contain exactly one tight electron candidate fulfilling the criteria described in electron isolation section of Chapter 3, and reject events which contain a loose electron in addition to the tight electron. In this category we are only interested in the decay to electron and jets, and we therefore reject events containing a loose muon.

In the μ+jets category, we select events which contain exactly one tight muon candidate whose criteria are described in muon identification and isolation section of Chapter 3, and reject events which contain an additional loose muon. In an analogous way to the e+jets category, we reject events containing a loose electron.

Finally, we require an event with E_T^{miss} in excess of 25(30) GeV for muons (electrons) and also with W transverse mass of at least 30 GeV. These criteria are designed to reduce the background contributions from QCD multijet production.

4.5 Data - MC comparisons after selection

To assess the quality of the modeling provided by the MC simulation, we make comparisons between the MC shape, normalized to the number of collected events in the data, so that the background is compared to the overall data yield. After applying all selection criteria, the distributions of kinematic variables are shown in figure 4.7 for the μ+jets category and in figure 4.8 for the e+jets category. The color code for each of the distributions under this section is same. The MC has been been corrected for lepton reconstruction efficiency and the trigger efficiency (described in next section). We also Data-MC comparison plots for variables used to reconstruct MVA discriminant. They are shown in figures 4.9–4.11

[2] A reconstructed physics candidate

for the μ+jets category and figures 4.12–4.14 for the e+jets category. Seeing reasonable agreement gives us confidence in the qualitative aspects of the MC modeling.

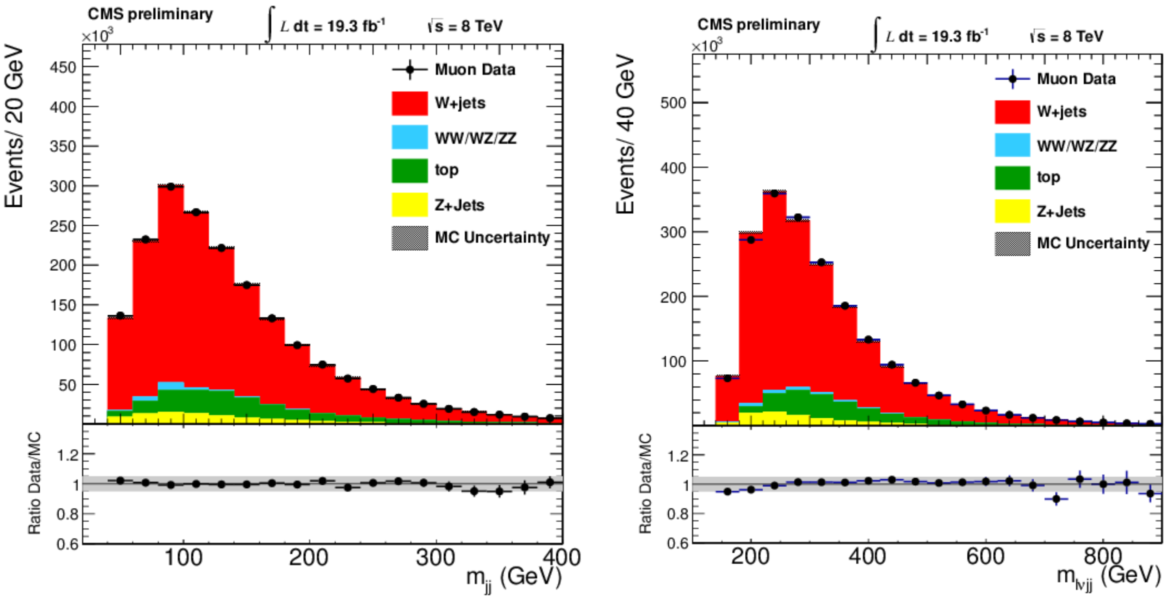


Figure 4.7: Comparison of the dijet mass (m_{jj}) distribution (left) and WW invariant mass (m_{lvjj}) (right) from data and MC for the μ+jets category.

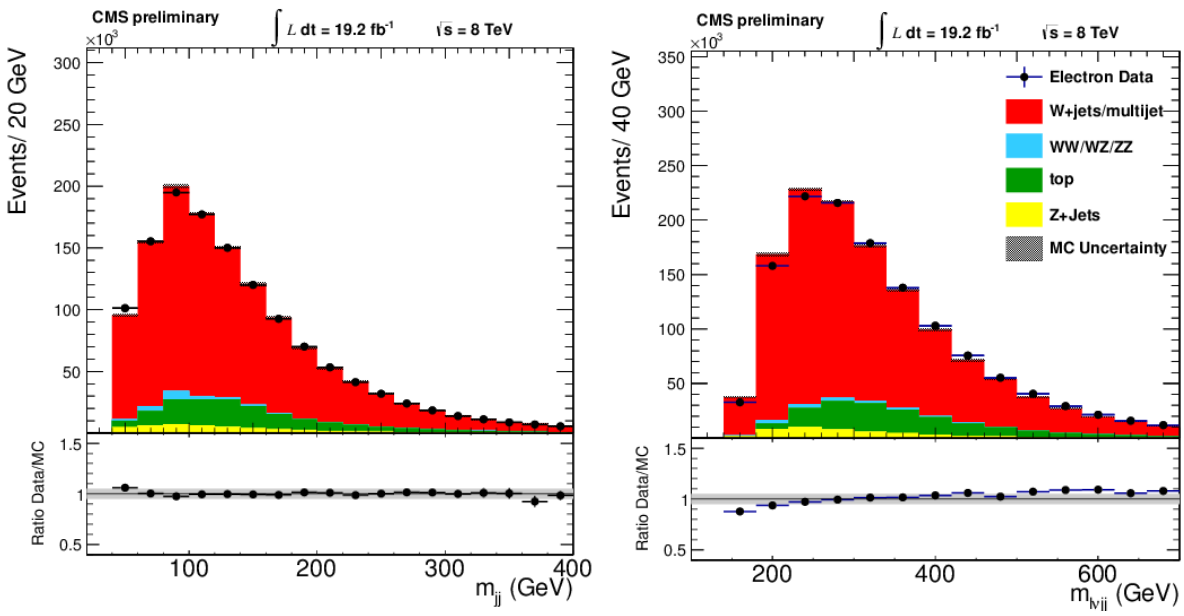


Figure 4.8: Comparison of the dijet mass (m_{jj}) distribution (left) and WW invariant mass (m_{lvjj}) (right) from data and MC for the e+jets category.

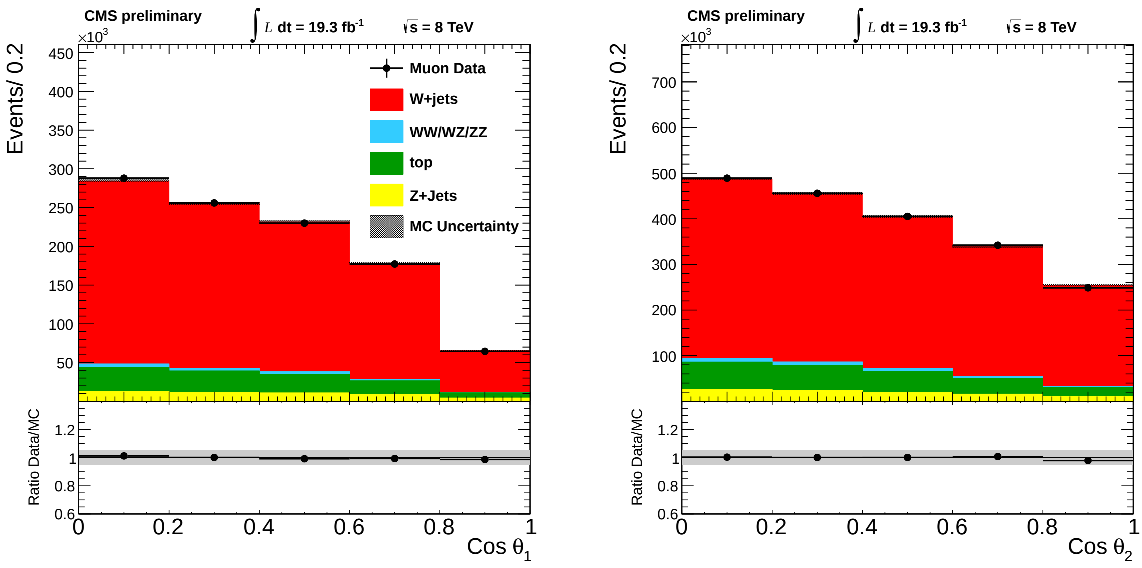


Figure 4.9: Comparison of the angular variables (described later) distribution for $\cos\theta_1$ (left) and $\cos\theta_2$ (right) from data and MC for the μ+jets category.

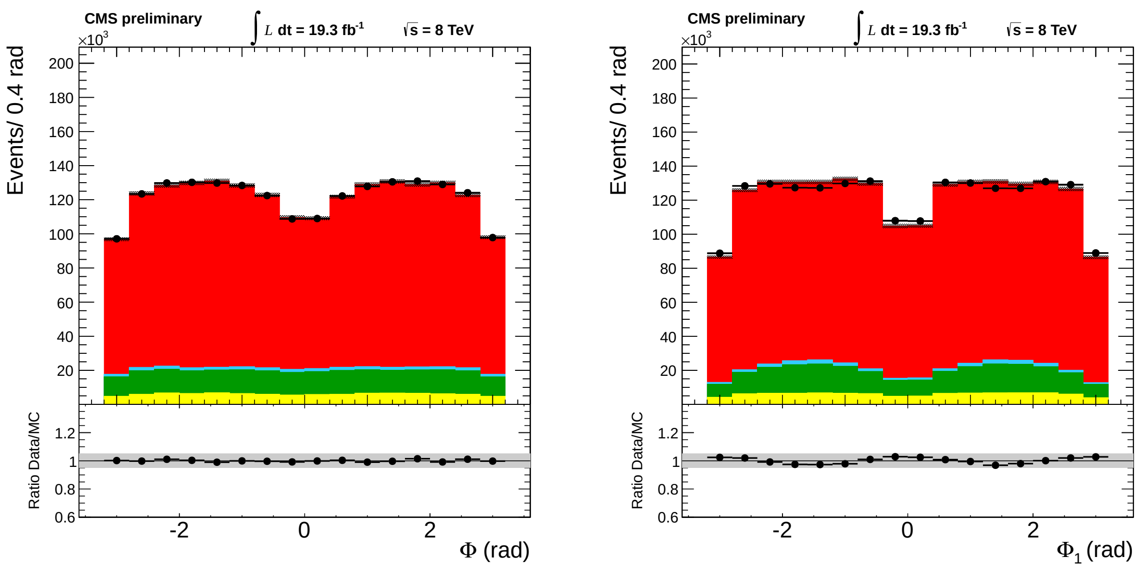


Figure 4.10: Comparison of the angular variables (described later) distribution for Φ (left) and Φ_1 (right) from data and MC for the μ+jets category.

4.6 Lepton reconstruction, identification and trigger efficiencies

Since the lepton reconstruction, selection and trigger efficiencies can be slightly different between data and simulation, correction factors have to be applied to the MC to account for these differences. The efficiencies are calculated using a Tag and Probe technique exploiting Z boson decays to a pair of electrons or muons, respectively. One of the leptons is used as tag and has to pass a tight selection criteria, while the second one is used as

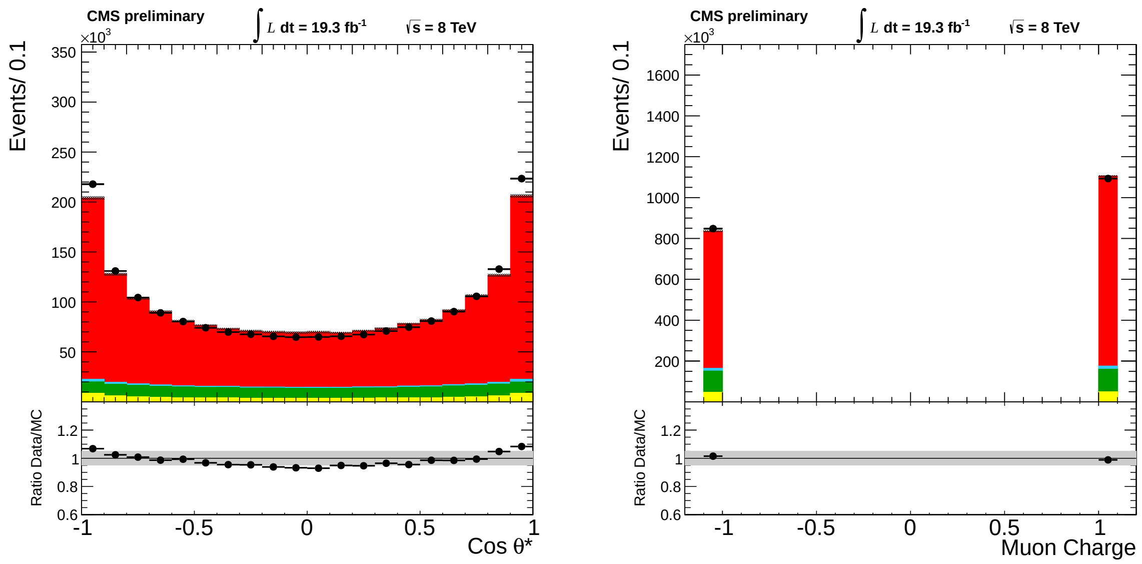


Figure 4.11: Comparison of the angular variable (described later) distribution for $\cos\theta^*$ (left) and charge distribution (right) from data and MC for the μ+jets category.

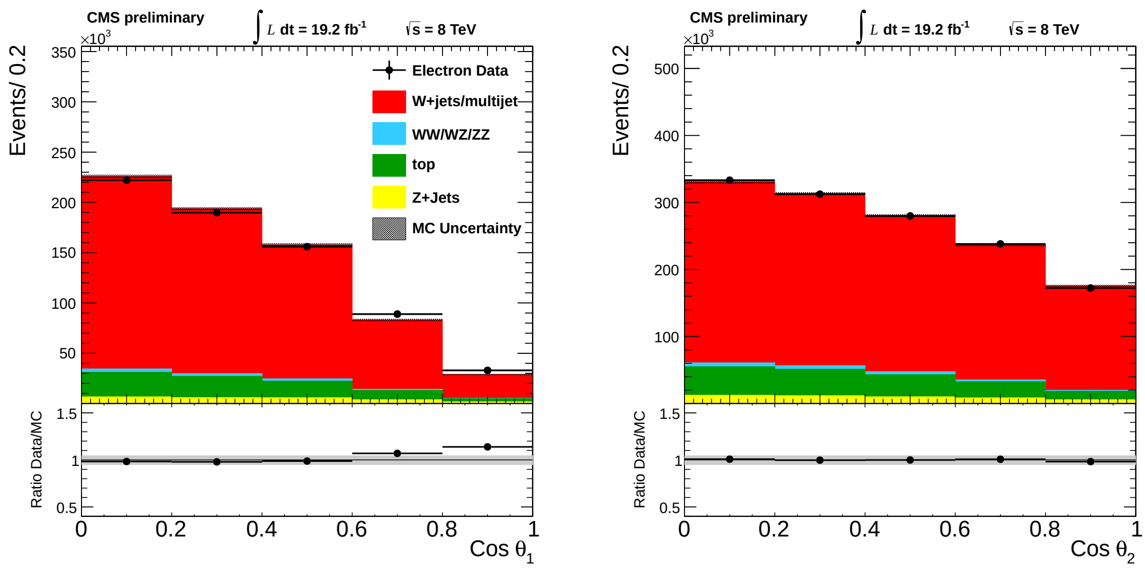


Figure 4.12: Comparison of the angular variables (described later) distribution for $\cos\theta_1$ (left) and $\cos\theta_2$ (right) from data and MC for the e+jets category.

probe if the tag-probe pair combines to the Z boson mass. The total lepton efficiency can be factorized into three components; reconstruction efficiency (ϵ_{Reco}), identification efficiency (ϵ_{Id}) and HLT (trigger) efficiency (ϵ_{HLT}). And hence ϵ_{total}, which is product of all three is called total efficiency.

$$\epsilon_{\text{total}} = \epsilon_{\text{Reco}} \cdot \epsilon_{\text{Id}} \cdot \epsilon_{\text{HLT}} \tag{4.2}$$

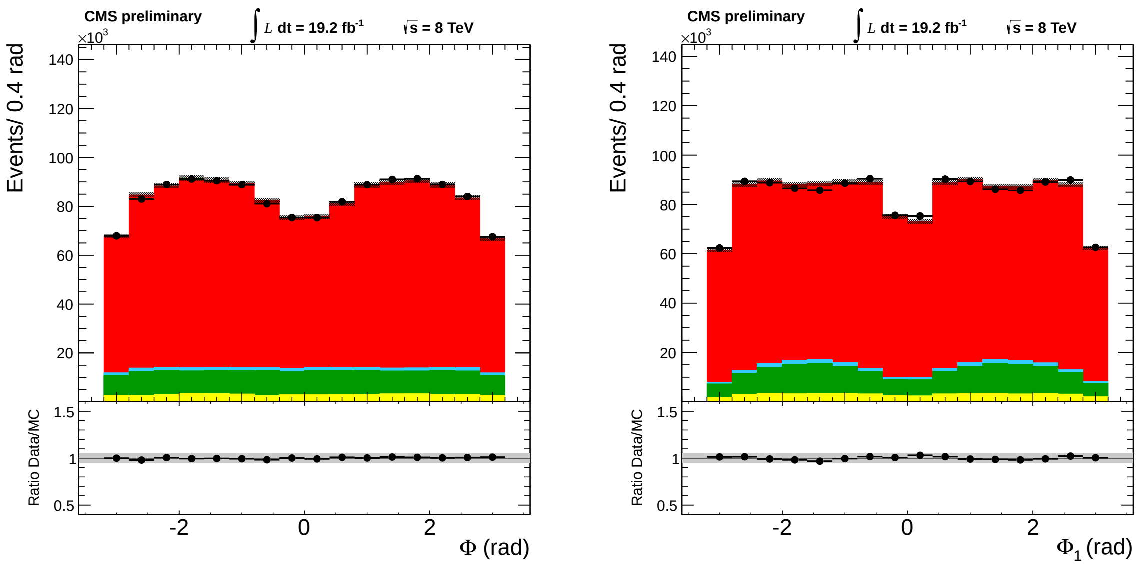


Figure 4.13: Comparison of the angular variables (described later) distribution for Φ (left) and Φ_1 (right) from data and MC for the e+jets category.

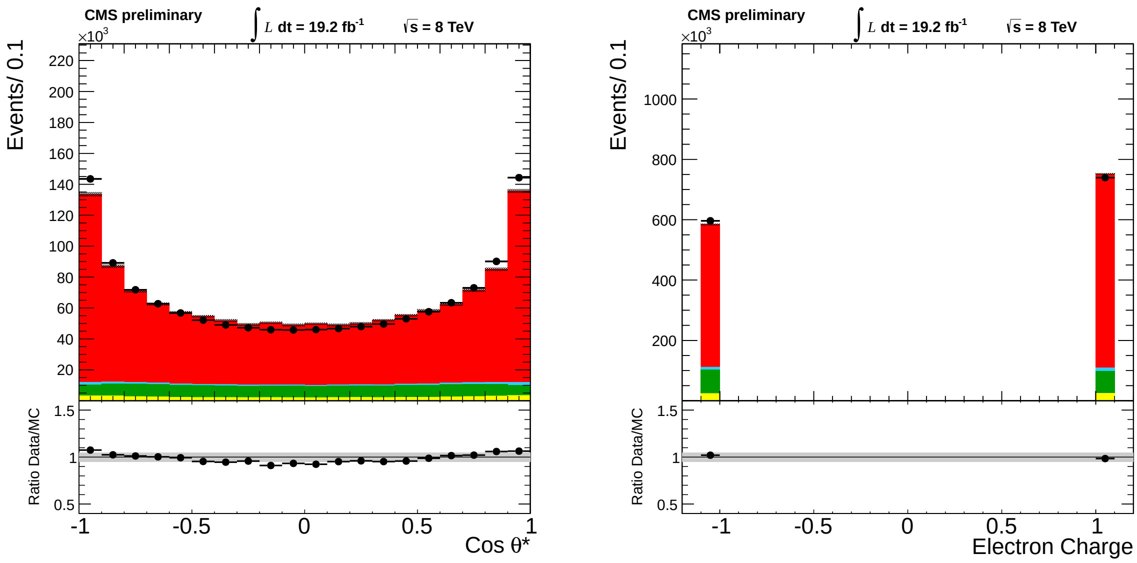


Figure 4.14: Comparison of the angular variable (described later) distribution for $\cos\theta^*$ (on left) and charge of e (on right) from data and MC for the e+jets category.

4.6.1 Electron efficiencies

In the electron case, the reconstruction efficiency ϵ_{Reco} characterizes the transition from a supercluster in the electromagnetic calorimeter to a reconstructed Particle Flow electron. The ability of a reconstructed electron to pass the offline selection consisting of several isolation and identification criteria is given by the identification efficiency ϵ_{Id}. Finally, the selected electron has a certain probability to fire the high level trigger (HLT) and the efficiency to fulfill the HLT requirements is parametrized as ϵ_{HLT}. In data, a single electron trigger is used at HLT level, while in MC the HLT requirements are dropped.

Since the HLT efficiency in MC is equal to 100%, the HLT efficiency measured on data is applied directly in the analysis of MC samples, while the other two efficiency components are calculated both for data and MC, so that a rato of data to MC scale factor is applied in the other cases.
In general, since the efficiency depends both on p_T and η of the electron, the measurement is done in p_T intervals as (30, 35, 40, 45, 50, 200 GeV) and in η intervals as (-2.5, -1.5, 0.0, 1.5, 2.5) of the probe electron. The resulting efficiencies and scale factors are summarized in table 4.2 and shown in figure 4.15.

$p_{\mathrm{T,min}}$ $[GeV]$	$p_{\mathrm{T,max}}$ $[GeV]$	η_{min}	η_{max}	$\epsilon_{\mathrm{Reco,data}}/\epsilon_{\mathrm{Reco,mc}}$	$\epsilon_{\mathrm{ID,data}}/\epsilon_{\mathrm{ID,mc}}$	$\epsilon_{\mathrm{HLT,data}}$
30	35	-2.5	-1.5	1.000 ± 0.002	0.973 ± 0.004	0.639 ± 0.003
30	35	-1.5	0	0.996 ± 0.001	0.981 ± 0.003	0.874 ± 0.001
30	35	0	1.5	0.996 ± 0.001	0.980 ± 0.003	0.874 ± 0.001
30	35	1.5	2.5	1.002 ± 0.001	0.999 ± 0.004	0.650 ± 0.003
35	40	-2.5	-1.5	1.001 ± 0.001	1.005 ± 0.003	0.686 ± 0.002
35	40	-1.5	0	0.999 ± 0.001	0.978 ± 0.002	0.896 ± 0.001
35	40	0	1.5	0.998 ± 0.001	0.978 ± 0.002	0.891 ± 0.002
35	40	1.5	2.5	1.001 ± 0.001	1.003 ± 0.085	0.690 ± 0.002
40	45	-2.5	-1.5	1.001 ± 0.001	1.005 ± 0.003	0.708 ± 0.002
40	45	-1.5	0	0.999 ± 0.001	0.985 ± 0.001	0.909 ± 0.001
40	45	0	1.5	0.999 ± 0.001	0.983 ± 0.001	0.906 ± 0.001
40	45	1.5	2.5	1.000 ± 0.001	1.014 ± 0.003	0.720 ± 0.002
45	50	-2.5	-1.5	1.001 ± 0.001	1.017 ± 0.003	0.724 ± 0.002
45	50	-1.5	0	1.000 ± 0.001	0.984 ± 0.002	0.917 ± 0.001
45	50	0	1.5	0.999 ± 0.001	0.985 ± 0.002	0.911 ± 0.001
45	50	1.5	2.5	1.001 ± 0.001	1.021 ± 0.003	0.733 ± 0.002
50	200	-2.5	-1.5	0.999 ± 0.001	1.023 ± 0.003	0.733 ± 0.003
50	200	-1.5	0	0.999 ± 0.001	0.990 ± 0.002	0.925 ± 0.001
50	200	0	1.5	0.999 ± 0.001	0.991 ± 0.003	0.920 ± 0.001
50	200	1.5	2.5	1.000 ± 0.001	1.019 ± 0.003	0.745 ± 0.003

Table 4.2: Electron efficiency and data/MC scale factors for super-cluster to reconstructed electrons (ϵ_{Reco}), reconstructed to selected electrons (ϵ_{ID}) and selected to HLT electrons (ϵ_{HLT}). The errors are statistical only.

4.6.2 Muon efficiencies

In the muon case, the reconstruction efficiency ϵ_{Reco} describes the ability to reconstruct a "Particle Flow" muon starting with a particle track and can be assumed to be one [89]. The identification efficiency ϵ_{Id} gives an estimate for a reconstructed muon to pass the offline selection criteria. It can be computed for both data and simulation and thus a scale factor being the ratio of the two efficiencies is derived.
The trigger efficiency ϵ_{HLT} is the fraction of selected muons fulfilling the HLT requirements and, since the HLT requirement is dropped on the MC analysis, the efficiency computed on data is used directly to correct the MC event expectation.
The efficiency measurement is divided in intervals both in P_T and η of the probe muon covering the relevant intervals (25, 30, 35, 40, 45, 50, 200 GeV) in P_T and (-2.1, -1.5, -1.0, -0.5, 0.0, 0.5, 1.0, 1.5, 2.1) in η. The resulting selection and trigger efficiencies and scale factors are summarized in table 4.3 and figure 4.16.

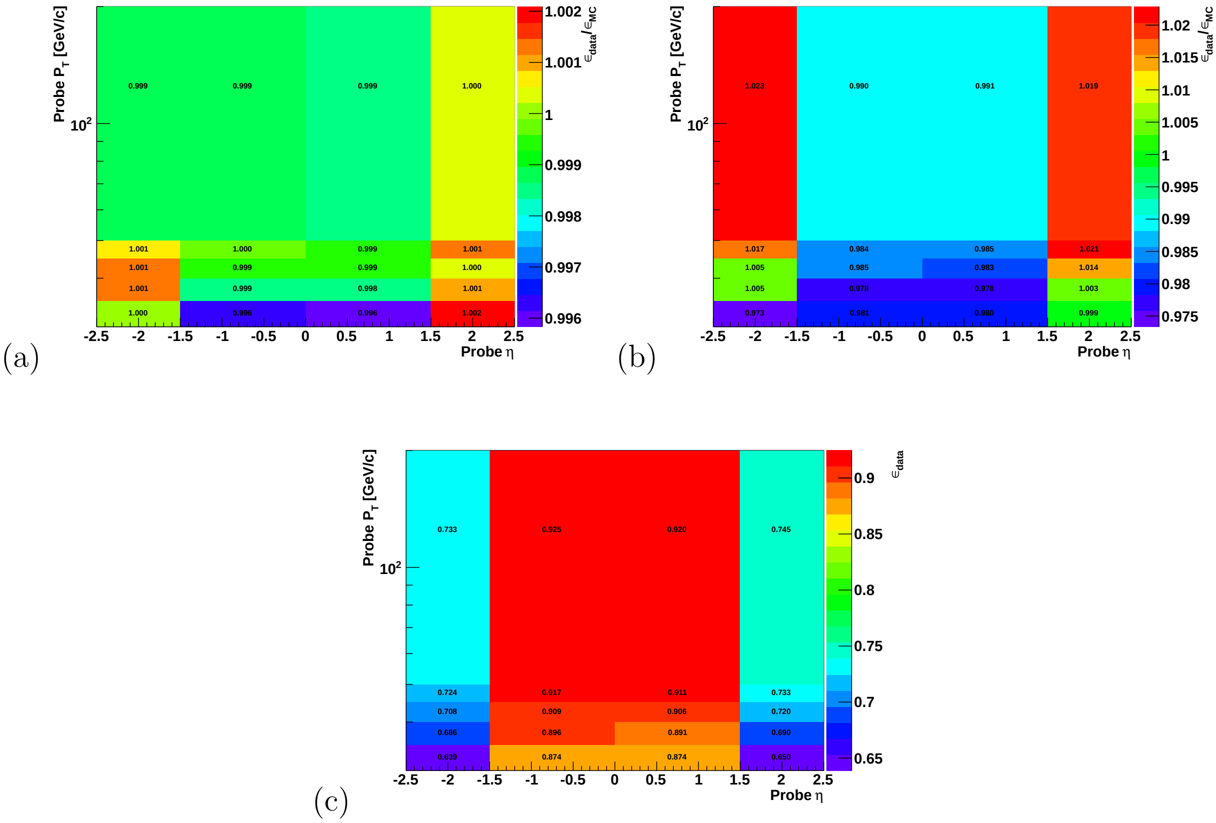


Figure 4.15: Electron efficiency and data/MC scale factors for super-cluster to reconstructed electrons ϵ_{Reco} (a), reconstructed to selected electrons ϵ_{Id} (b) and selected to HLT electrons ϵ_{HLT} (c).

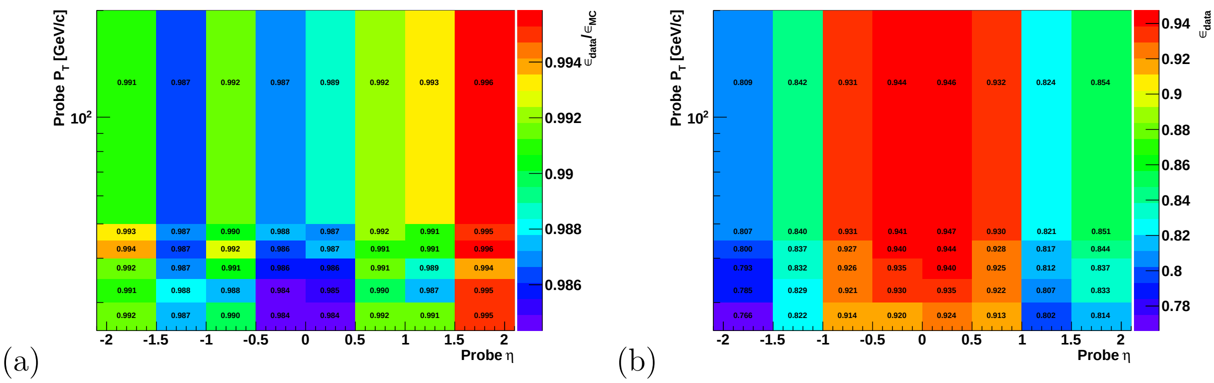


Figure 4.16: Muon scale factors for reconstructed to selected muons $\epsilon_{\mathrm{ID,data}}/\epsilon_{\mathrm{ID,mc}}$ (a) and efficiency for selected to HLT muons $\epsilon_{\mathrm{HLT,data}}$ (b).

4.7 Kinematic fit

In a given event hypothesis, measured and unmeasured quantities have to match certain kinematic constraints such as energy and momentum conservation or invariant masses of

$p_{\mathrm{T,min}}$ $[GeV]$	$p_{\mathrm{T,max}}$ $[GeV]$	η_{min}	η_{max}	$\epsilon_{\mathrm{ID,data}}/\epsilon_{\mathrm{ID,mc}}$	$\epsilon_{\mathrm{HLT,data}}$
25	30	-2.1	-1.5	0.992 ± 0.003	0.766 ± 0.003
25	30	-1.5	-1	0.987 ± 0.003	0.822 ± 0.003
25	30	-1	-0.5	0.990 ± 0.003	0.914 ± 0.002
25	30	-0.5	0	0.984 ± 0.003	0.920 ± 0.002
25	30	0	0.5	0.985 ± 0.003	0.924 ± 0.002
25	30	0.5	1	0.992 ± 0.003	0.913 ± 0.002
25	30	1	1.5	0.991 ± 0.003	0.802 ± 0.003
25	30	1.5	2.1	0.995 ± 0.002	0.814 ± 0.003
30	35	-2.1	-1.5	0.991 ± 0.002	0.785 ± 0.002
30	35	-1.5	-1	0.988 ± 0.002	0.829 ± 0.002
30	35	-1	-0.5	0.988 ± 0.002	0.921 ± 0.002
30	35	-0.5	0	0.984 ± 0.002	0.930 ± 0.001
30	35	0	0.5	0.985 ± 0.002	0.935 ± 0.001
30	35	0.5	1	0.990 ± 0.002	0.922 ± 0.002
30	35	1	1.5	0.987 ± 0.002	0.807 ± 0.002
30	35	1.5	2.1	0.995 ± 0.002	0.833 ± 0.002
35	40	-2.1	-1.5	0.992 ± 0.002	0.793 ± 0.002
35	40	-1.5	-1	0.987 ± 0.002	0.832 ± 0.002
35	40	-1	-0.5	0.991 ± 0.002	0.926 ± 0.001
35	40	-0.5	0	0.986 ± 0.002	0.935 ± 0.001
35	40	0	0.5	0.986 ± 0.002	0.940 ± 0.001
35	40	0.5	1	0.991 ± 0.002	0.925 ± 0.001
35	40	1	1.5	0.989 ± 0.002	0.812 ± 0.002
35	40	1.5	2.1	0.994 ± 0.002	0.837 ± 0.002
40	45	-2.1	-1.5	0.994 ± 0.002	0.800 ± 0.002
40	45	-1.5	-1	0.987 ± 0.001	0.837 ± 0.002
40	45	-1	-0.5	0.992 ± 0.001	0.927 ± 0.001
40	45	-0.5	0	0.986 ± 0.001	0.940 ± 0.001
40	45	0	0.5	0.987 ± 0.001	0.944 ± 0.001
40	45	0.5	1	0.991 ± 0.001	0.928 ± 0.001
40	45	1	1.5	0.991 ± 0.001	0.817 ± 0.002
40	45	1.5	2.1	0.996 ± 0.001	0.844 ± 0.002
45	50	-2.1	-1.5	0.993 ± 0.002	0.807 ± 0.002
45	50	-1.5	-1	0.987 ± 0.002	0.840 ± 0.002
45	50	-1	-0.5	0.990 ± 0.001	0.931 ± 0.001
45	50	-0.5	0	0.988 ± 0.002	0.941 ± 0.001
45	50	0	0.5	0.987 ± 0.002	0.947 ± 0.001
45	50	0.5	1	0.992 ± 0.001	0.930 ± 0.001
45	50	1	1.5	0.991 ± 0.002	0.821 ± 0.002
45	50	1.5	2.1	0.995 ± 0.002	0.851 ± 0.002
50	200	-2.1	-1.5	0.991 ± 0.002	0.809 ± 0.002
50	200	-1.5	-1	0.987 ± 0.002	0.842 ± 0.002
50	200	-1	-0.5	0.992 ± 0.002	0.931 ± 0.001
50	200	-0.5	0	0.987 ± 0.002	0.944 ± 0.001
50	200	0	0.5	0.989 ± 0.002	0.946 ± 0.001
50	200	0.5	1	0.992 ± 0.002	0.932 ± 0.001
50	200	1	1.5	0.993 ± 0.002	0.824 ± 0.002
50	200	1.5	2.1	0.996 ± 0.002	0.854 ± 0.002

Table 4.3: Muon selection scale factors and HLT efficiencies. The errors are statistical only.

particles. Due to the presence of uncertainties in the measured quantities, these constraints are not exactly matched. The constraints then can be used to slightly change the measured values within their uncertainties. The concept of an event-by-event least square fitting along-with the application of Langrange Multipliers can be employed to ensure that the measured and unmeasured quantities match the kinematic constraints obtained from the event hypothesis. This procedure is referred to as a kinematic fit. It not only results in a χ^2 test of the event hypothesis but also in improved estimators of the underlying kinematics for a given event.

For this analysis, in order to improve the mass resolution in the $m_{l\nu jj}$ measurement, a two constraint (2C) fit with one unknown, the neutrino z momentum component p_Z, was made. The constraints were that the lepton-neutrino pair and the di-jet system both separately form an invariant mass equal to the W boson mass, within its known width. The four particles are part of the fit, with a covariance matrix supplied in the fit for the jets and the lepton. The starting value input to the fit for the x and y components of the neutrino is the measured E_T^{miss} along the x and y axes, while the z component is chosen as the solution with the smallest absolute value (or the one that produces the neutrino closest to the lepton). The errors for the x and y transverse components come likewise from the E_T^{miss} measurement, whereas z component is assigned a small error. In case no real solution is found, the real part of the solution is used.

The improved resolution in the $m_{l\nu jj}$ is a crucial ingredient in the Higgs search. Without the improved resolution, the rise and fall of the low mass spectrum of electroweak WW continuum production is not well resolved from a resonant Higgs mass spectrum, were it to occur near the WW threshold. The kinematic fit also improves our ability to model the dominant W+Jets background and reduce its shape uncertainty, which is one of our dominant systematic uncertainty. Since in the limit setting $m_{l\nu jj}$ shapes is used to distinguish signal from background, the kinematic fit has a direct impact on the final limit.

Figure 4.17, 4.18 and 4.19 shows the effect of the kinematic fit on the $m_{l\nu jj}$ distribution for the signal at the four lowest mass points, as well as for WW and W+Jets. The distributions obtained before and after the kinematic fit are reported. The effect of the kinematic fit becomes less pronounced for increased signal mass.

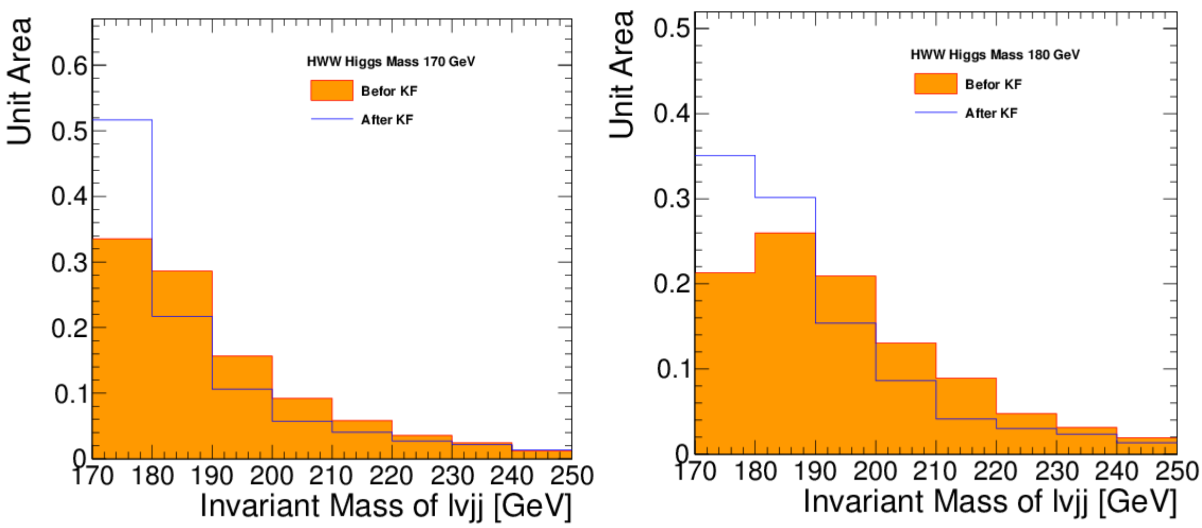


Figure 4.17: The $m_{l\nu jj}$ distributions obtained before and after the kinematic fit are shown for the Higgs mass point 170 GeV and 180 GeV.

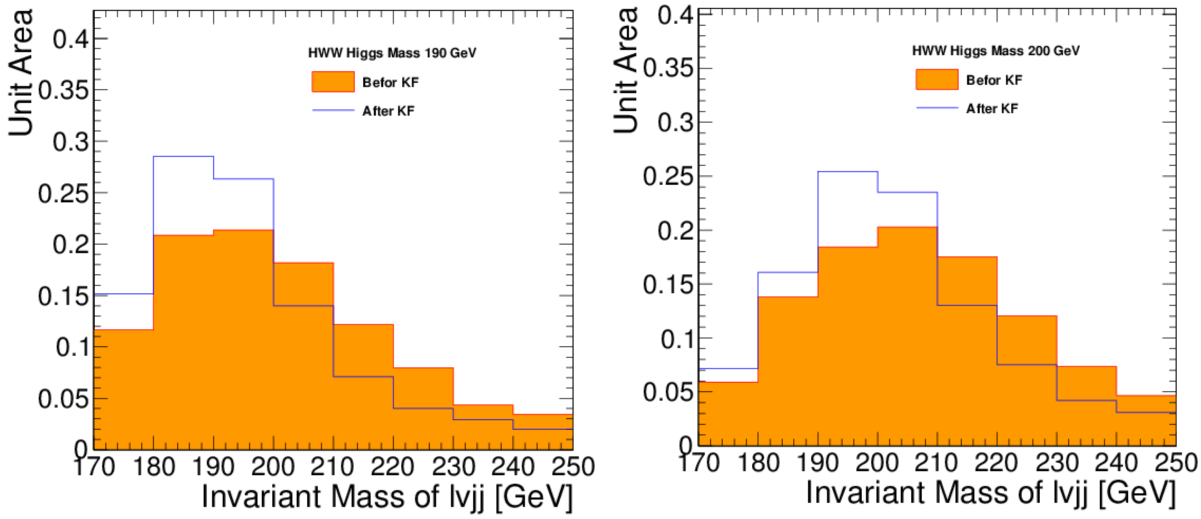


Figure 4.18: The $m_{l\nu jj}$ distributions obtained before and after the kinematic fit are shown for the Higgs mass point 190 GeV and 200 GeV.

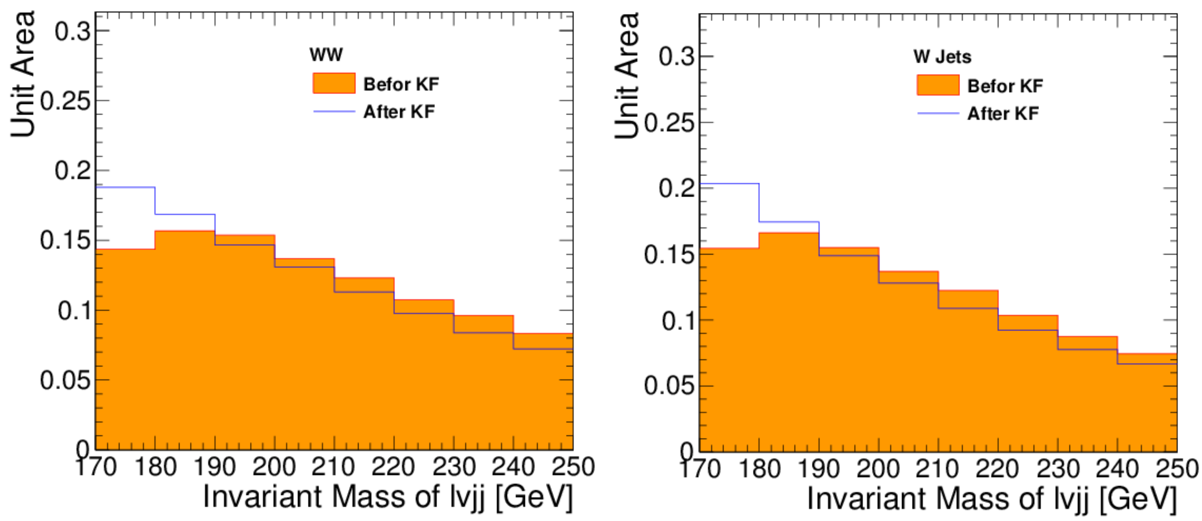


Figure 4.19: The $m_{l\nu jj}$ distributions obtained before and after the kinematic fit are shown for WW (on left) and W+Jets (on right) backgrounds.

4.8 Multivariate optimization

In experimental high energy physics, often we have to extract signal from huge background. And, more often than not, we do not find a single variable whose nature of distribution is such that, it can significantly discriminate between the signal and background. To deal with such situations a method has been devised using which, by taking in account any possible discrimination between signal and background in a set of input variables, a new single variable termed as "Discriminator" [3] is developed. In practice this is implemented in ROOT [90] [4] and is called as TMVA [91] (Toolkit for Multivariate Data Analysis). There

[3]Popularly known as MVA output
[4]An object oriented data analysis framework

are a number of methods implemented in it, we found likelihood based discriminator as better choice compared to others. We also call this section as MVA likelihood analysis. In addition to the selections mentioned in Section. 4.4, the MVA likelihood analysis applies the following selections additional selections, these selection is devised to choose jets well separated from E_T^{miss} in the event, the threshold in electron channel is higher because it has huge multi-jet background:

- $|\Delta\phi_{\text{leading jet},E_T^{miss}}| > 0.4$ for muons
- $|\Delta\phi_{\text{leading jet},E_T^{miss}}| > 0.8$ for electrons

4.9 Inputs to Multivariate Analysis

The kinematics of the Higgs boson decay into WW in semileptonic final state at leading order are described by the following observables [92, 93]:

$$\{m_{l\nu jj}, m_{jj}, \cos\theta_1, \cos\theta_2, \Phi, \cos\theta^*, \Phi_1\} \tag{4.3}$$

The invariant mass of the leptonic W, $m_{l\nu}$, is constrained in the kinematic fit to compute the longitudinal momentum of the neutrino. The angular variables are correlated and defined in Fig. 4.20. The angle is the polar angle between the parton collision axis z and

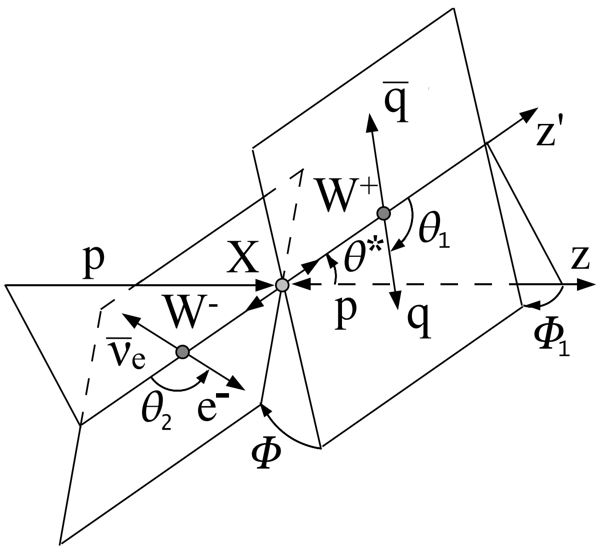


Figure 4.20: Angular definition for the $WW \to l\nu jj$ process

the X decay axis z', both defined in the rest frame of X. The angle Φ_1 is the azimuthal angle between the zz' plane and the decay plane of hadronic W. The angles and Φ_1 are denoted as production angles because they depend on the production mechanism, gg or $q\bar{q}$. For the SM Higgs which is a spin-zero particle, the production angles are flat (before acceptance). The angle Φ is the angle between the decay planes of the two W systems in the rest frame of X. The angle θ_i is the angle between the direction of the fermion f from $W \to f\bar{f}$ and the direction opposite to the X in the W_i rest frame, where index i = 1, 2 refers to the first or second W boson. In the case of the, angle from the hadronic W, it is ambiguous as to which jet is originating from the fermion or anti-fermion, so the angles is defined from 0 to π for the leading p_T jet. Finally, The angles Φ, do not depend on the production mechanism and are denoted as the helicity angles.

The observables $m_{l\nu jj}$, m_{jj} and Φ_1 are excluded from the MVA inputs. The observable, $m_{l\nu jj}$, is not used because it is the distribution of this observable which is used in the extraction of the upper limit. The observable, m_{jj}, is not used since the m_{jj} distribution is used to extract from the sideband to the signal region to extract a data-driven shape for the background in the $m_{l\nu jj}$ spectrum. The observable, Φ_1, is not used because of its under performance. That leaves four angular variables as inputs to the multivariate discriminant. As an example, the related search for the Higgs in the $ZZ \to 2l2q$ final state at CMS [94] uses these observables in an angular discriminant. Since the invariant masses are not used in the multivariate discriminant, the angular variables are defined in a loose window in $m_{l\nu jj}$ in order to take into account the correlations between the angles and the masses. In addition, to the four angular variables, the lepton charge is also included to give some discrimination power since the W + jets background is asymmetric with respect to charge while the SM Higgs signal will be symmetric in lepton charge. Combining all the inputs together the final set of inputs to the multivariate discriminant is:

$$\{\cos\theta_1, \cos\theta_2, \Phi, \cos\theta^*, \text{lepton charge}\} \tag{4.4}$$

The distributions of input variables are shown in figures 4.21–4.32 for various Higgs boson mass hypothesis in μ+jets category. The blue color represent signal while red shows backgrounds shapes.

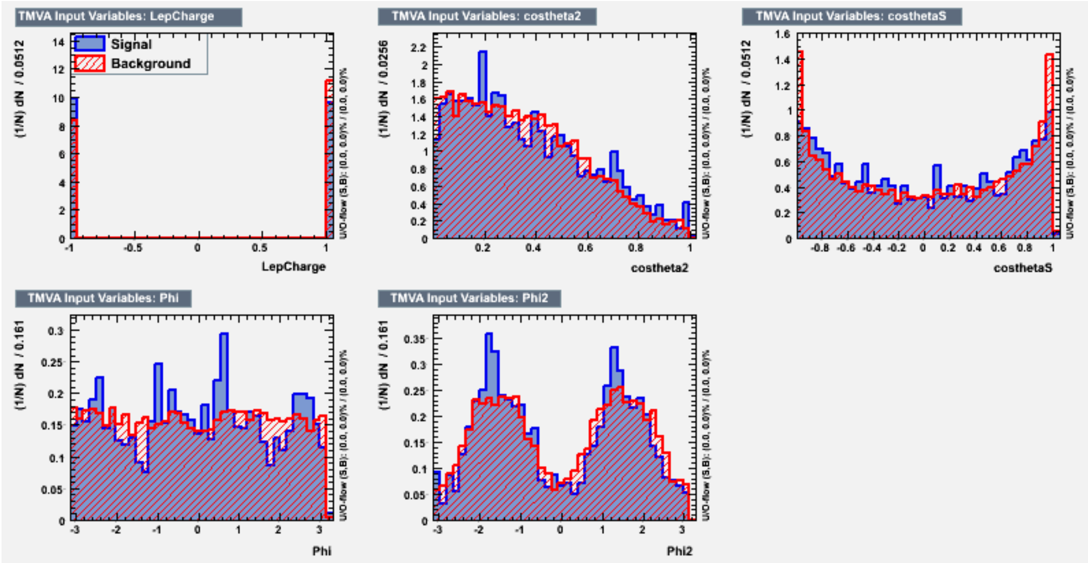


Figure 4.21: Inputs to the multivariate discriminant for M_H =170 GeV for the μ+jets category.

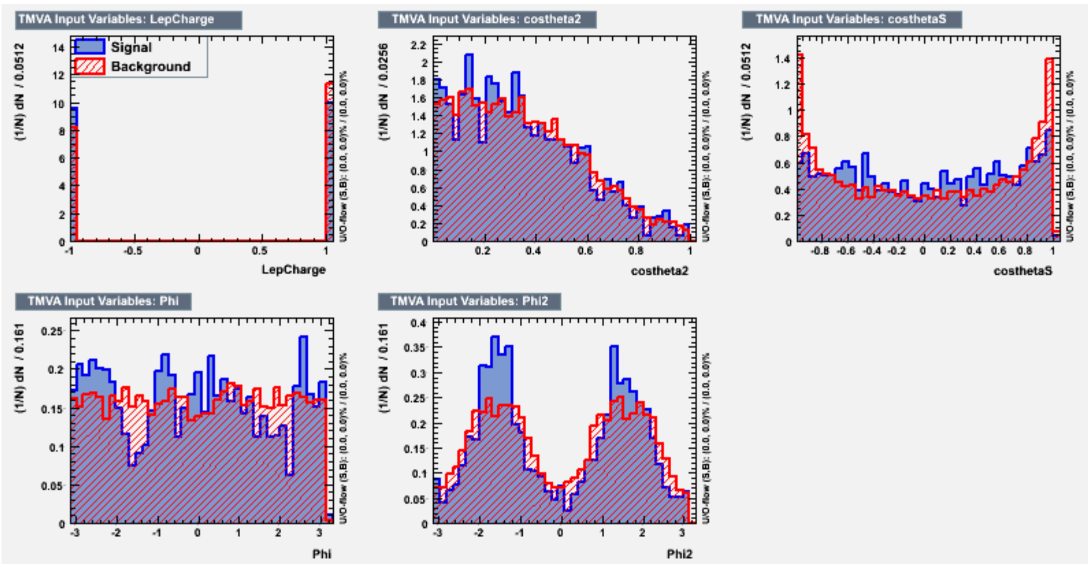


Figure 4.22: Inputs to the multivariate discriminant for M_H =180 GeV for the μ+jets category.

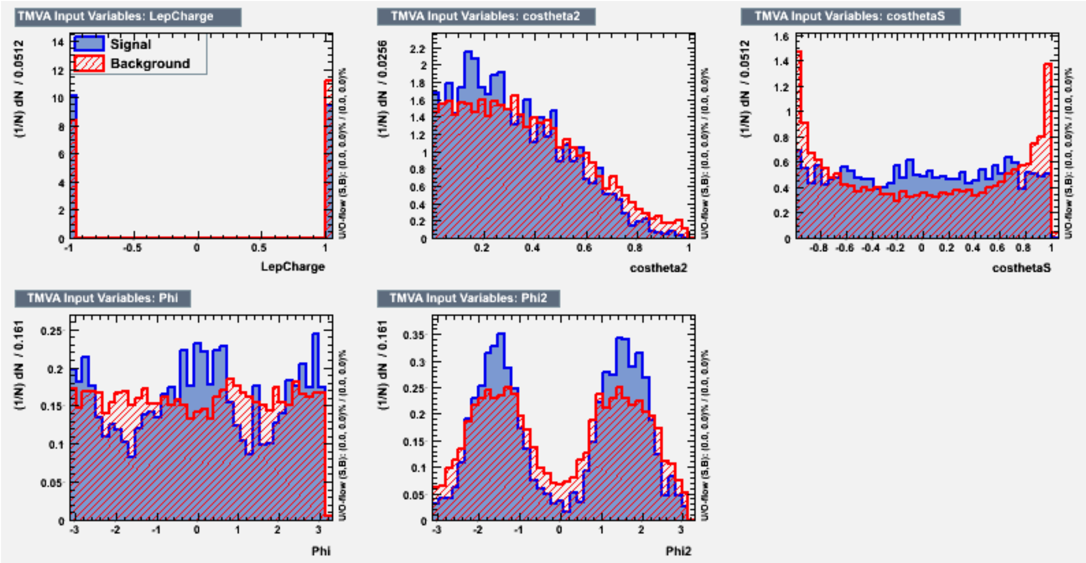


Figure 4.23: Inputs to the multivariate discriminant for M_H =190 GeV for the μ+jets category.

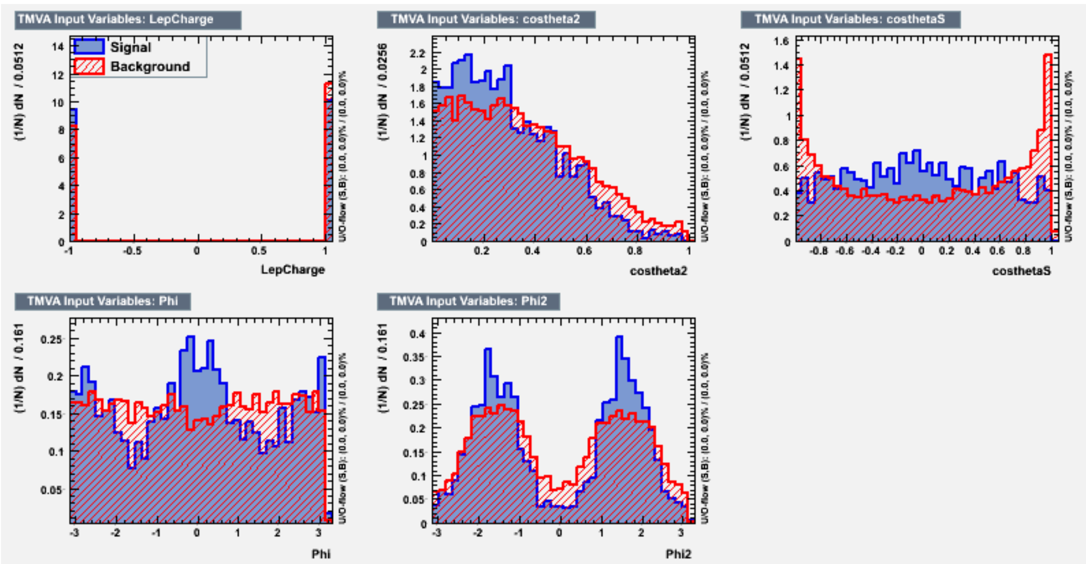


Figure 4.24: Inputs to the multivariate discriminant for M_H =200 GeV for the μ+jets category.

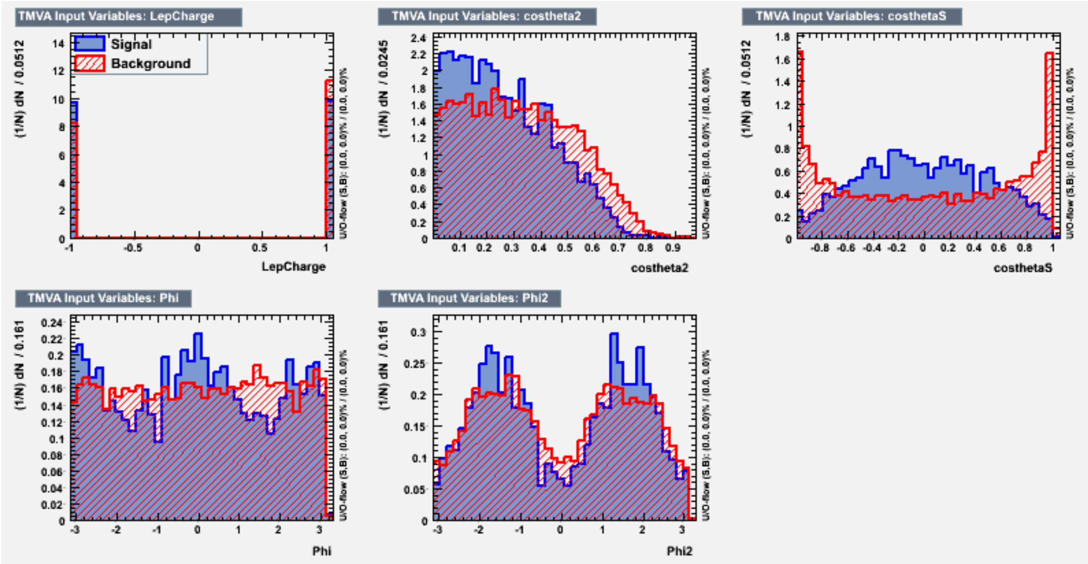


Figure 4.25: Inputs to the multivariate discriminant for M_H =250 GeV for the μ+jets category.

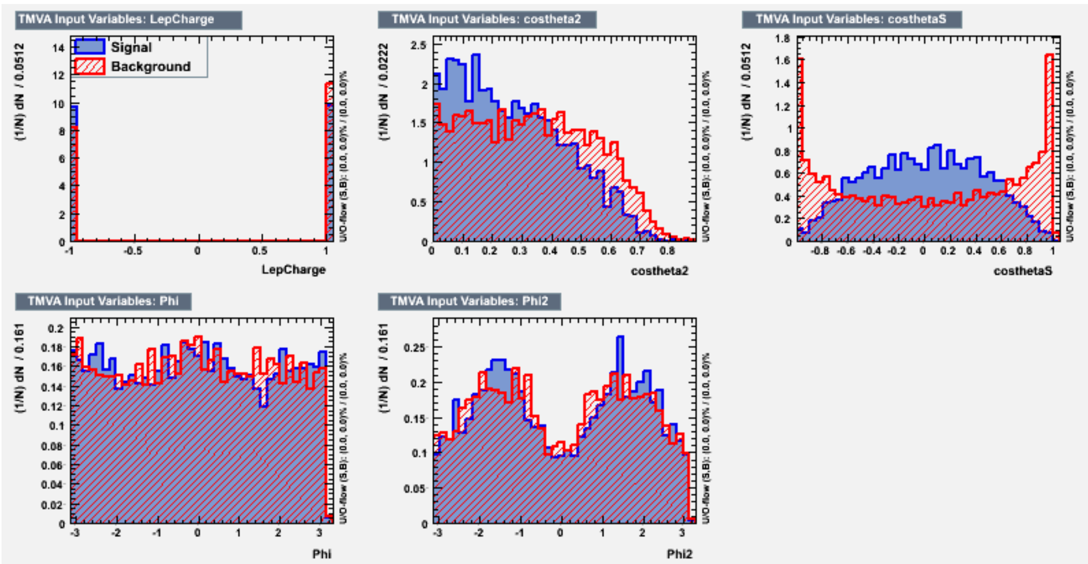


Figure 4.26: Inputs to the multivariate discriminant for M_H =300 GeV for the μ+jets category.

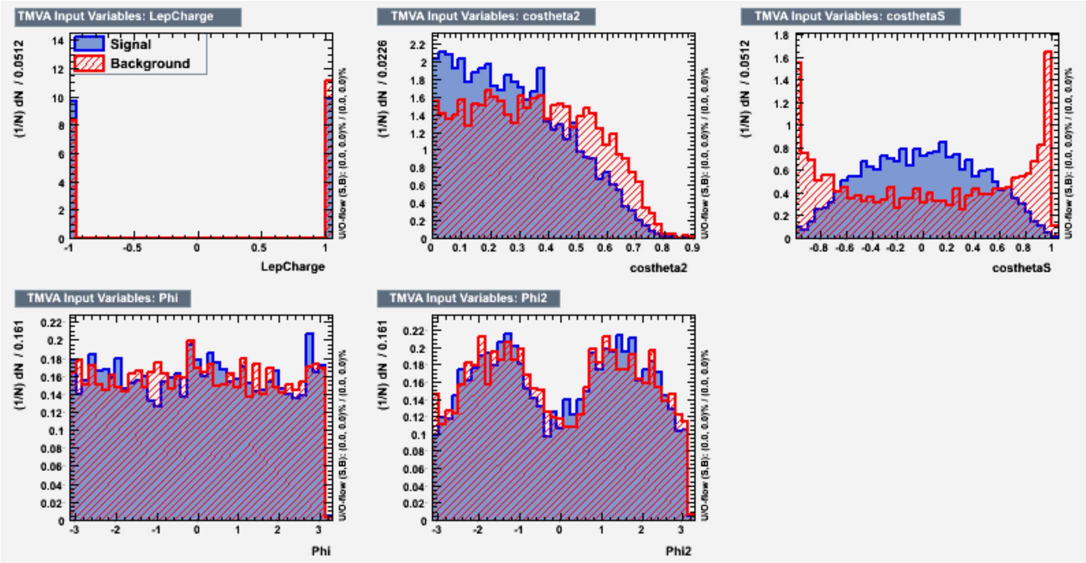


Figure 4.27: Inputs to the multivariate discriminant for M_H =350 GeV for the μ+jets category.

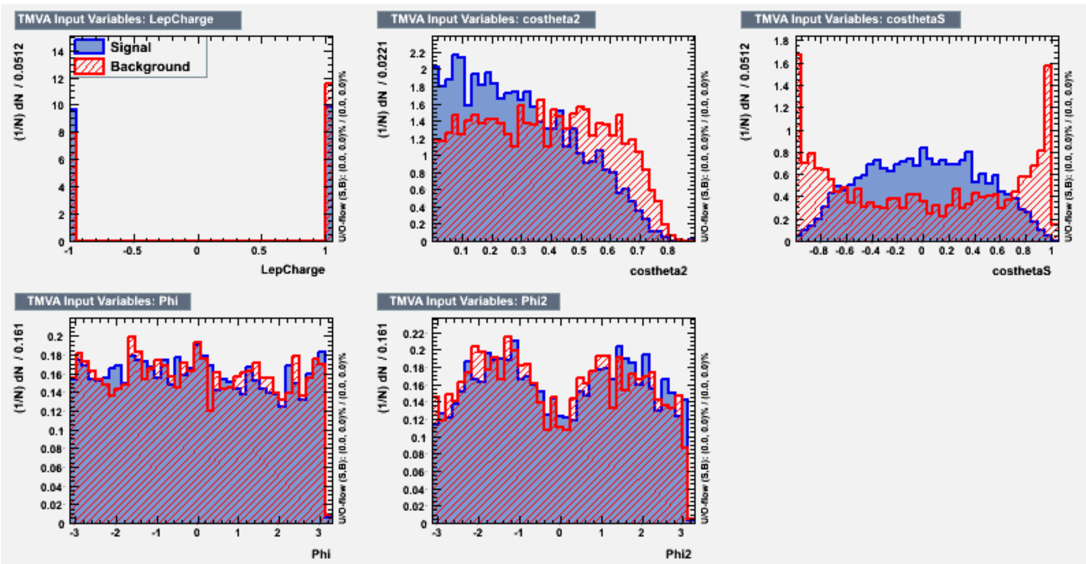


Figure 4.28: Inputs to the multivariate discriminant for M_H =400 GeV for the μ+jets category.

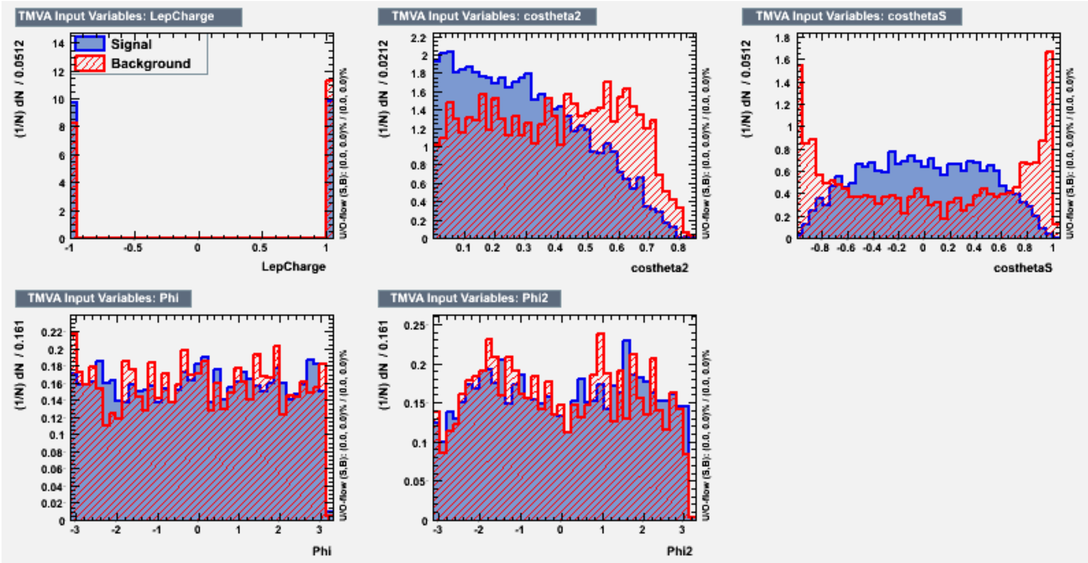


Figure 4.29: Inputs to the multivariate discriminant for M_H =450 GeV for the μ+jets category.

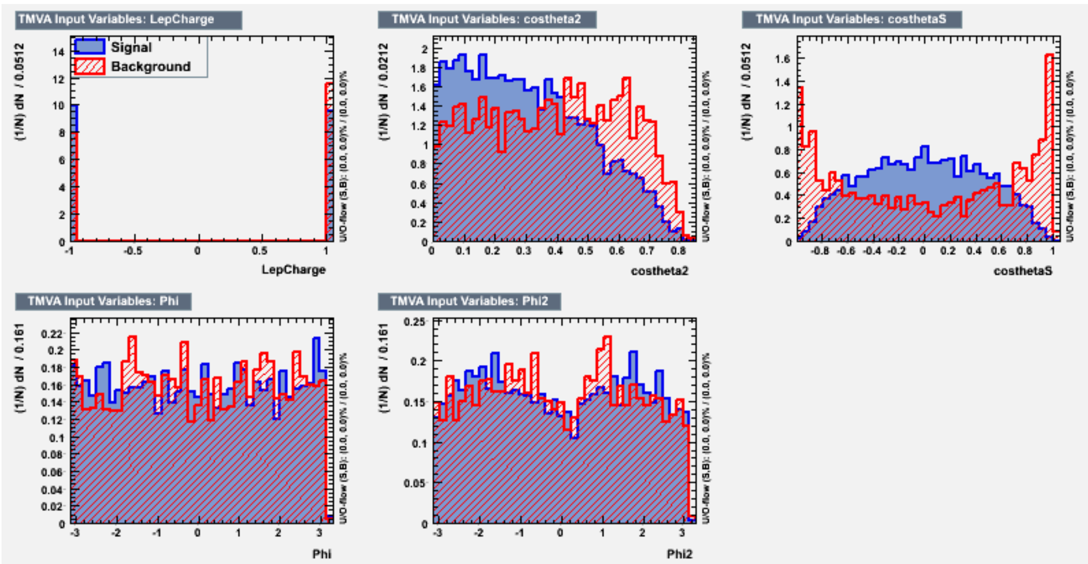


Figure 4.30: Inputs to the multivariate discriminant for M_H =500 GeV for the μ+jets category.

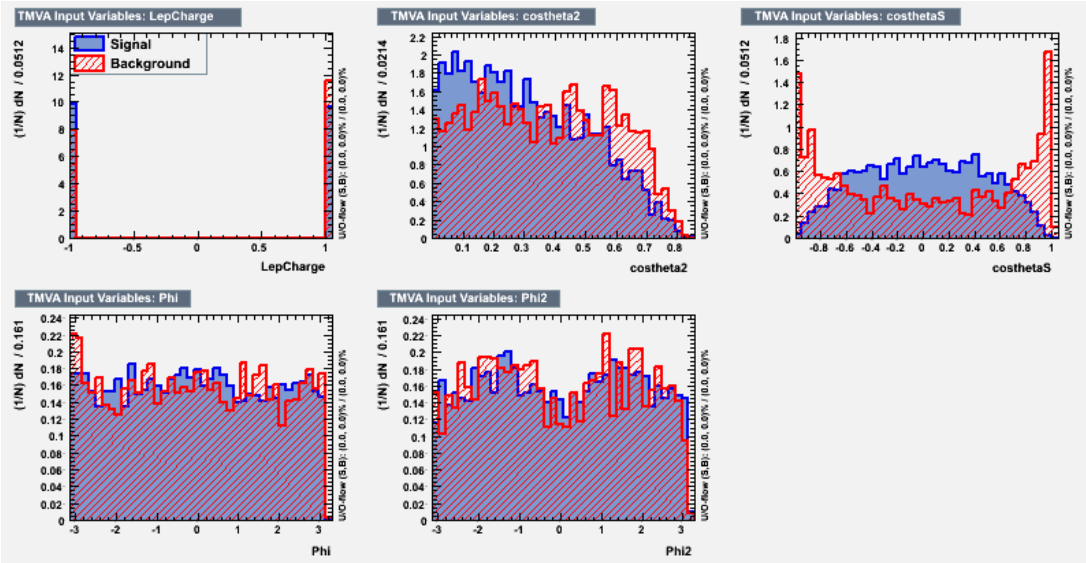


Figure 4.31: Inputs to the multivariate discriminant for M_H =550 GeV for the μ+jets category.

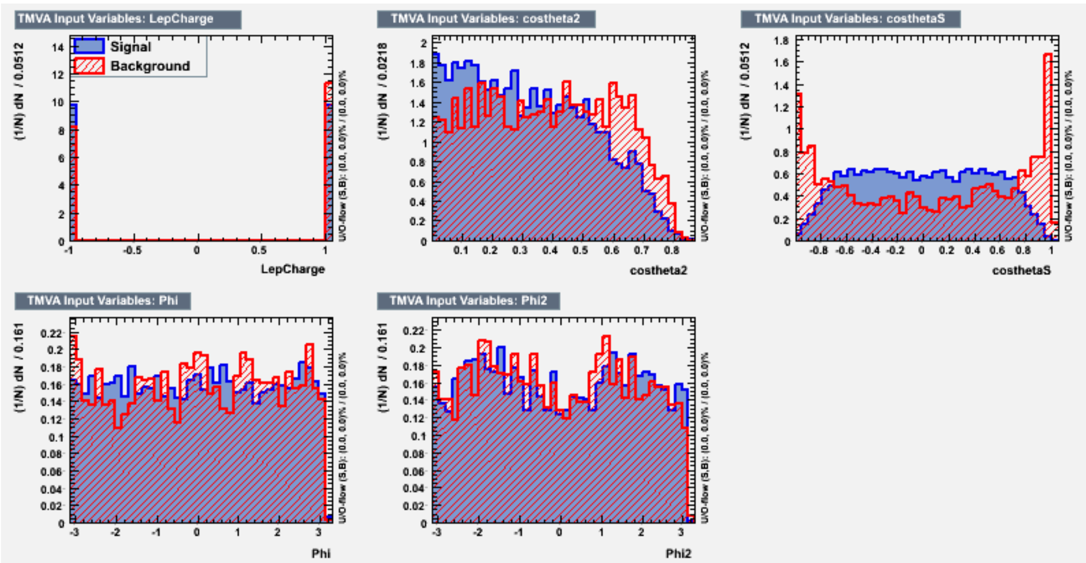


Figure 4.32: Inputs to the multivariate discriminant for M_H =600 GeV for the μ+jets category.

4.10 MVA Outputs

The classifier outputs after MVA training for the likelihood discriminant are shown in figures 4.33–4.34 for all Higgs mass points, μ+jets category. One can notice that signal and backgrounds shape get separated, signal peaks towards one while backgrounds component shape peaks towards zero. Which means by applying a requirement on MVA discriminant output above certain threshold helps us in picking more signal events and hence increasing signal over background ratio.

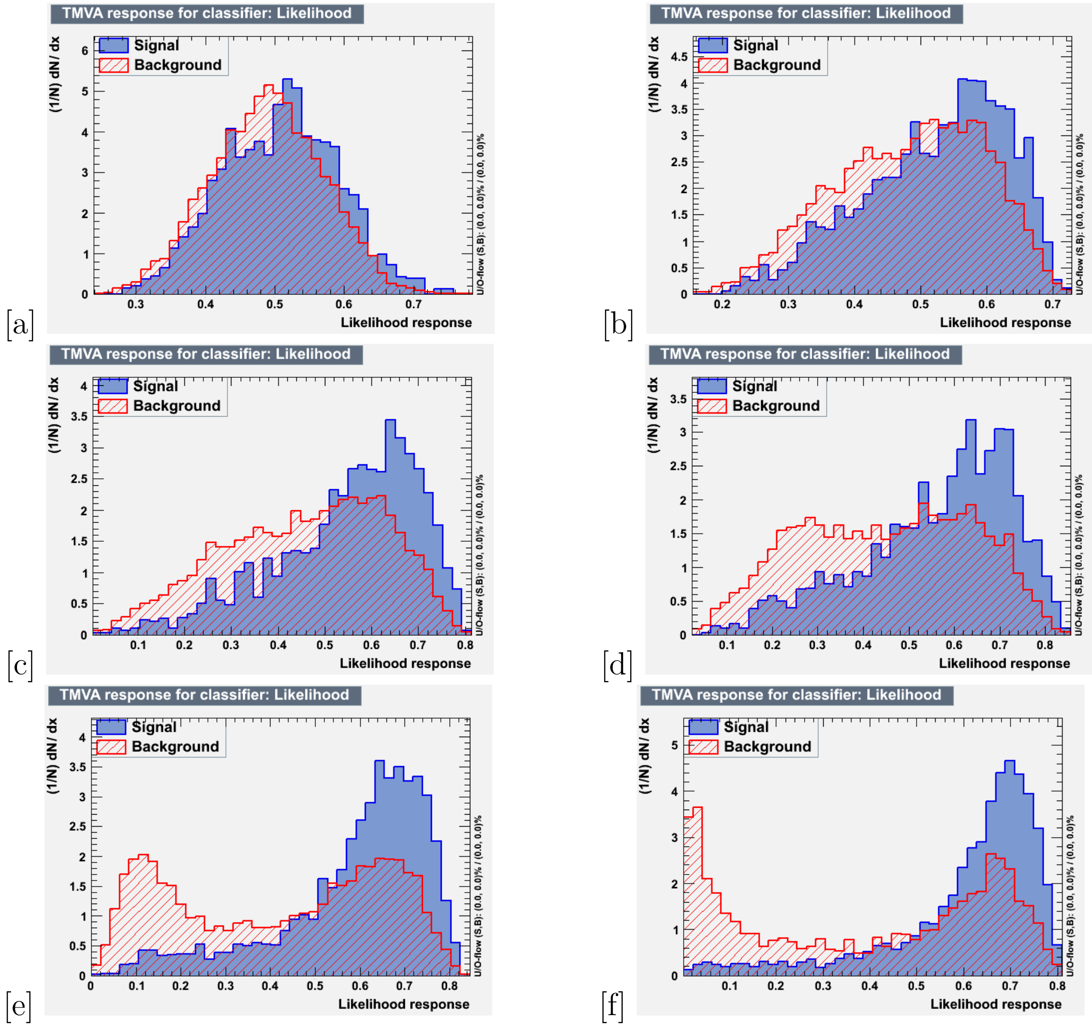


Figure 4.33: Output of the multivariate discriminant from 'a' – 'd' for M_H =170–300 GeV respectively, for μ+jets category.

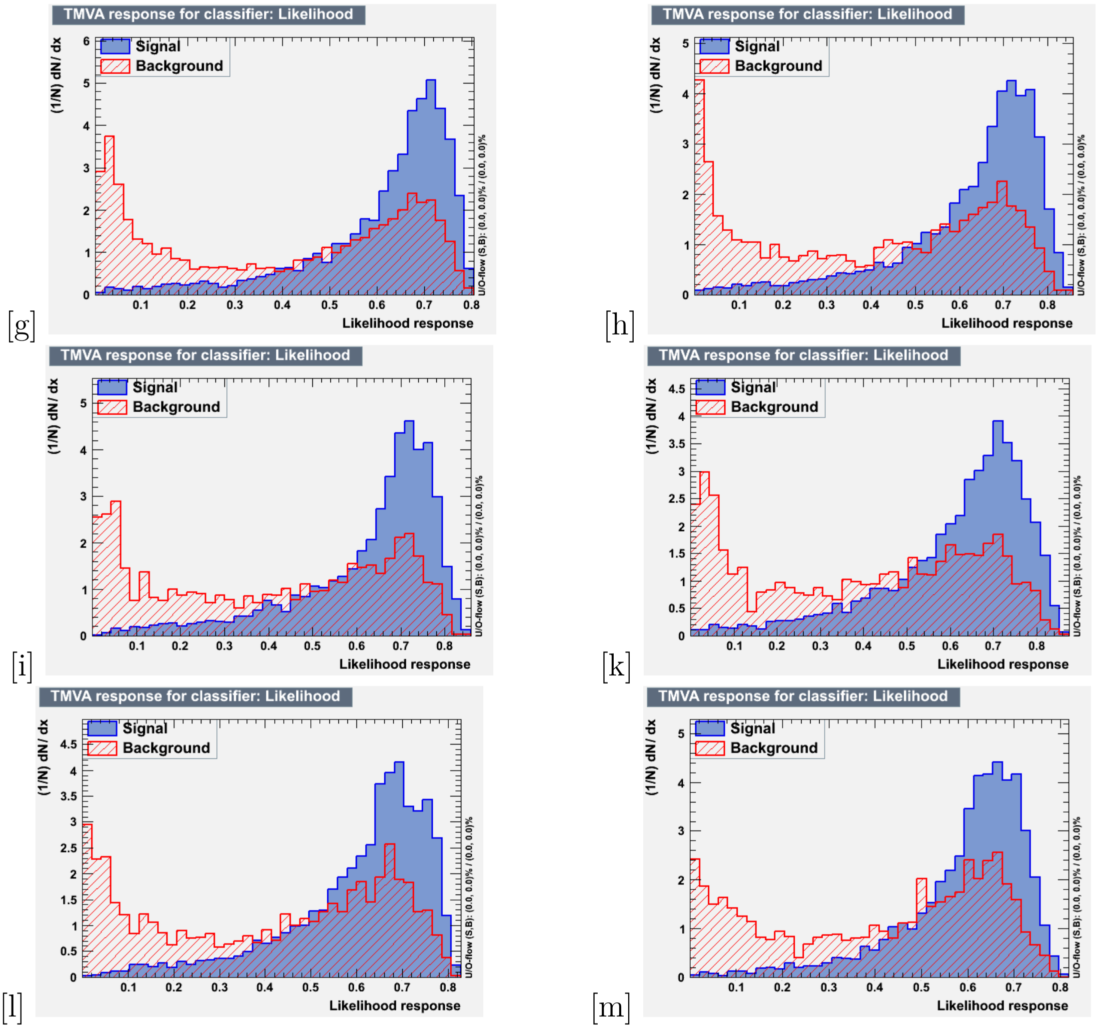


Figure 4.34: Output of the multivariate discriminant from 'g' – 'm' for M_H =350–600 GeV respetively for μ+jets category.

Also, the correlation plots for representative mass points are shown in Appendices C.1.

4.10.1 Selecting threshold values

Representative likelihood discriminants for the μ+jets category are shown in figure 4.35 after selections are optimized for Higgs mass hypotheses of 200 GeV and 500 GeV. The error bars plotted corresponds to the statistical uncertainties. The gray band in the ratio plot corresponds to a 5% uncertainty from W+jets normalization.

To select the most appropriate threshold value for the MVA discriminator, expected limits is scanned again threshold on MVA discriminant value. It should be noted that because of the nature of the MVA discriminator, the value of 0.5 is the boundary between signal-like and background-like events. It may be seem from Figure 4.36, which is a representative plot for the Higgs mass hypothesis of 250 GeV. The minimum is very broad and in almost all cases very near to 0.5. So, We adopt a threshold of 0.5 for all

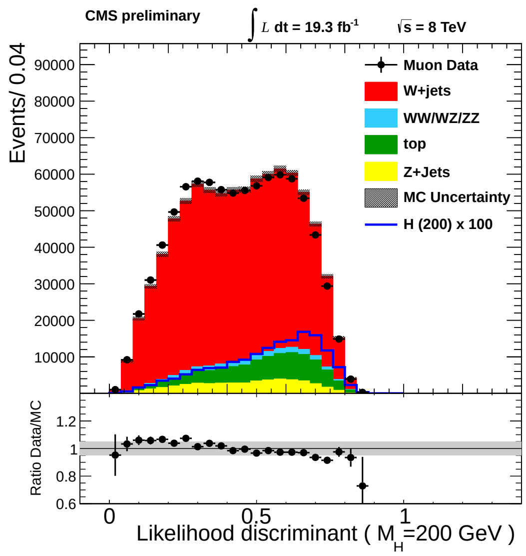


Figure 4.35: The likelihood discriminants with selections optimized for Higgs mass hypotheses of 200 GeV (left) and 500 GeV (right) for μ+jets category.

mass-hypotheses up to 550 GeV. For the masses 550 GeV and 600 GeV the threshold is slightly tighter at 0.6.

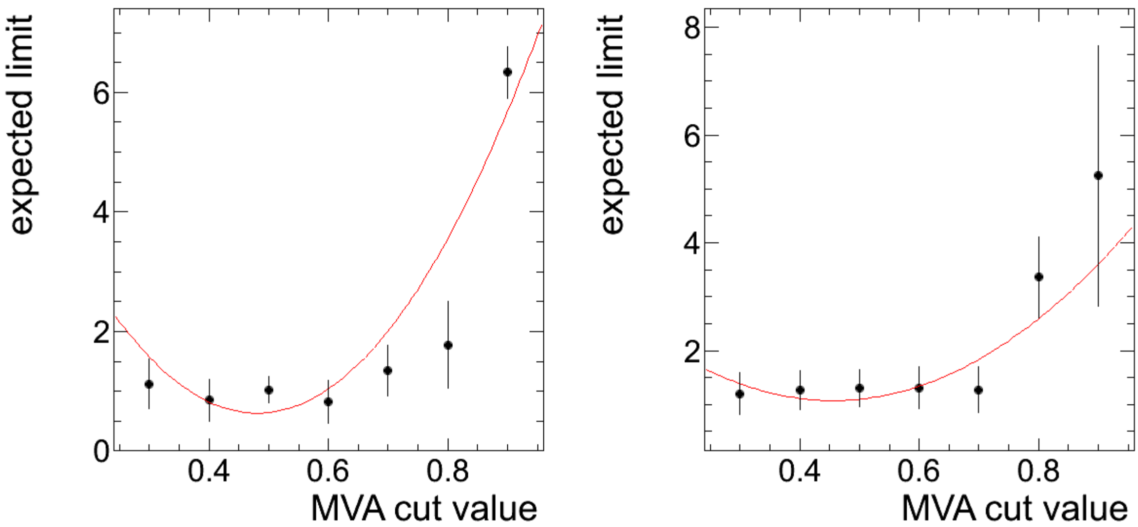


Figure 4.36: Expected limit as a function of MVA threshold value for the mass hypothesis 250 GeV. Left is μ+jets category; right is e+jets category.

4.11 Modeling of background shapes in m_{jj}

We utilize parameterizations of the shapes for the background. These are determined from fully corrected MC. There are 3 major contributions to the background, V+jets (W and Z), top ($t\bar{t}$ and single t), and diboson (WW and WZ). These are used in the fit to the

m_{jj} spectrum with the hadronic W boson signal region removed. In the next sections the shapes used for each background component, are presented in detail.

4.11.1 Diboson background shape

The diboson component is modeled as a sum of two Gaussian distributions, the means of which differ by the W & Z bosons mass difference. The widths of the Gaussian shapes are also the same fractions of the means. There is also a tail for poorly reconstructed diboson events consisting of an exponential decay times an error function. The values of the parameters are determined from MC and fixed in the fit. So, the full parametrized function can be noted as $\mathcal{F}_{\text{diboson}}$, such that:

$$\mathcal{F}_{\text{diboson}} = f_W \mathcal{G}_W + f_Z \mathcal{G}_Z + (1 - f_W - f_Z)\mathcal{F}_{\text{comb}} \tag{4.5}$$

where, f_W and $\mathcal{G}_W$ are probability and Gaussian function terms for first Gaussian of the full parametrized function for diboson, which try to model W boson mass distribution, while f_Z and $\mathcal{G}_Z$ and same term for second Gaussian which try to model Z boson peak. It may be noted that boson jets resolution is not sufficient to separate hadronically decayed W and Z mass distribution. $\mathcal{F}_{\text{comb}}$ is a combination of two functions, an exponential decay and an error function.

4.11.2 Top background shape

The top quark component is modeled as a peaking component and a combinatoric component. For Higgs mass hypotheses below 450 GeV the combinatoric background is a exponential times an error function, for 450 GeV and higher the combinatoric component is just a second Gaussian. Again the values of the parameters is determined from MC and fixed in the fit; such that

$$\mathcal{F}_{\text{top}} = f_{\text{peak}} \mathcal{G} + (1 - f_{\text{peak}})\mathcal{F}_{\text{comb}} \tag{4.6}$$

where, $\mathcal{F}_{\text{top}}$ represent the full parametrized function for top background, f_{peak} is the probability term for the peaking component (in general Gaussian) and $\mathcal{G}$ being its functional form and $\mathcal{F}_{\text{comb}}$ is the combinatoric function.

4.11.3 V+jets background shape

We employ an empirical description of the V+jets shape. This description is a kinematic turn on and a power law tail:

$$\mathcal{F}_{W+\text{jets}} = \text{erf}(m_{jj}; m_0, \sigma) \times \left[(m_{jj})^{-\alpha-\beta\ln(m_{jj}/\sqrt{s})}\right], \tag{4.7}$$

where m_0 is the value of the turn on and σ is the width of this turn on. The parameters m_0, σ, α and β are determined from the fit to the data after imposing the MVA threshold. There are some constraints on the parameters from the MC, but because of the relatively low statistics of the MC samples they are not overly constraining.

4.12 The data-driven four-body mass background determination

We want to do a fit on four-body mass to determine the precise shapes of the background components. This needs to to be done in the signal region for each Higgs mass hypothesis. The signal region is defined as 66 GeV $< m_{jj}^W <$ 98 GeV. This requirement keeps ~ 80 % signal events while substantially reduces V+Jets backgrounds. Below it is explained in detail the way to obtain four body mass shape in signal region in a data-driven way.

4.12.1 Background shapes for the four-body mass spectrum

The functional form of the background is determined from MC and validated using the data from the m_{jj} side-band. The side-band is defined as the region on m_{jj} distribution excluding signal region. There are no peaking backgrounds in the $m_{\ell\nu jj}$ spectrum which is dominated by W+jets production.

To validate the $m_{\ell\nu jj}$ spectrum arising from W+jets shape we take the following steps:

1. Fit the MC spectrum in the m_{jj} signal region with the selected functional form. If the $P(\chi^2) < 0.001$ then the functional form is rejected in favor of another functional form with more degrees of freedom.

2. Fit the MC spectrum in the m_{jj} side-band with the selected functional form. Same criteria as step 1.

3. Fit the data spectrum in the m_{jj} side-band. We use the same criteria as in step 1.

4. Fit the MC spectrum with a polynomial function with sufficient degrees of freedom to saturate the χ^2 probability.

5. Compare the background estimate from the nominal functional form with that of the polynomial. If the variation is not well within the statistical error from the nominal parameters, reject the function in favor of another functional form with more degrees of freedom.

Table 4.4: Results of the selection procedure to select a V+jets shape to use as the model for the $m_{\ell\nu jj}$ spectrum. The Bkg. rel. diff. is the relative difference in background normalization in the area of the Higgs peak from the nominal model and the polynomial model. The Bkg. diff. signif. is the significance of this difference with respect to the statistical variations of the nominal model.

Higgs mass	Nominal $m_{\ell\nu jj}$ shape	poly. order	Bkg. rel. diff.	Bkg diff. signif.
170	erf*power	6	-0.0056	0.194
180	erf*exp	6	0.0011	0.289
190	erf*exp	6	0.0001	0.033
200	erf*power	6	-0.0004	0.118
250	erf*power	6	-0.0001	0.040
300	erf*power	6	-0.0013	0.448
350	exp(quadratic)	6	0.0006	0.240
400	exp(quadratic)	5	0.0035	0.473
450	power	6	-0.0016	0.302
500	power	6	0.0011	0.193
550	power	6	0.0026	0.464
600	power	6	0.0020	0.377

This procedure is repeated for each mass point under consideration. The functional form varies by mass point. The results of this test are summarized in Table 4.4. In the fit to the data, the initial parameter values for the chosen functional form are those determined from the MC. These are nuisance parameters to the limit extraction and are subjected to loose constraints based on the MC shapes, but the ultimate values of the background shape nuisance parameters are determined during the limit extraction minimization.

4.12.2 Shape cross-checks

Before one proceeds to signal extraction, or wants to interpret any result, it is necessary to make sure that fit extraction procedure is unbiased. Since in this analysis W+Jets shape is obtained by data-driven method, any potential bias from choice background shape is assessed by generating Toy data samples set using 'nominal shape' (the parametric function which chosen initially), then these toy data samples are fitted with nominal shape or polynomial shapes with more degrees of freedom.

The results of this toy MC study is shown in figure. 4.37, which shows the mean of the pull distributions and their errors for the various mass points. From this it is seen that the polynomials actually provide a less biased background shape. Therefore polynomials are chosen as the default background parametric function. For a few of the masses, another alternative pdf were evaluated, where essentially the power law function has been replaced with an exponential and exponential function with a power law. Then again Toy data samples were generated with this alternate function and fitted with the polynomial to verify

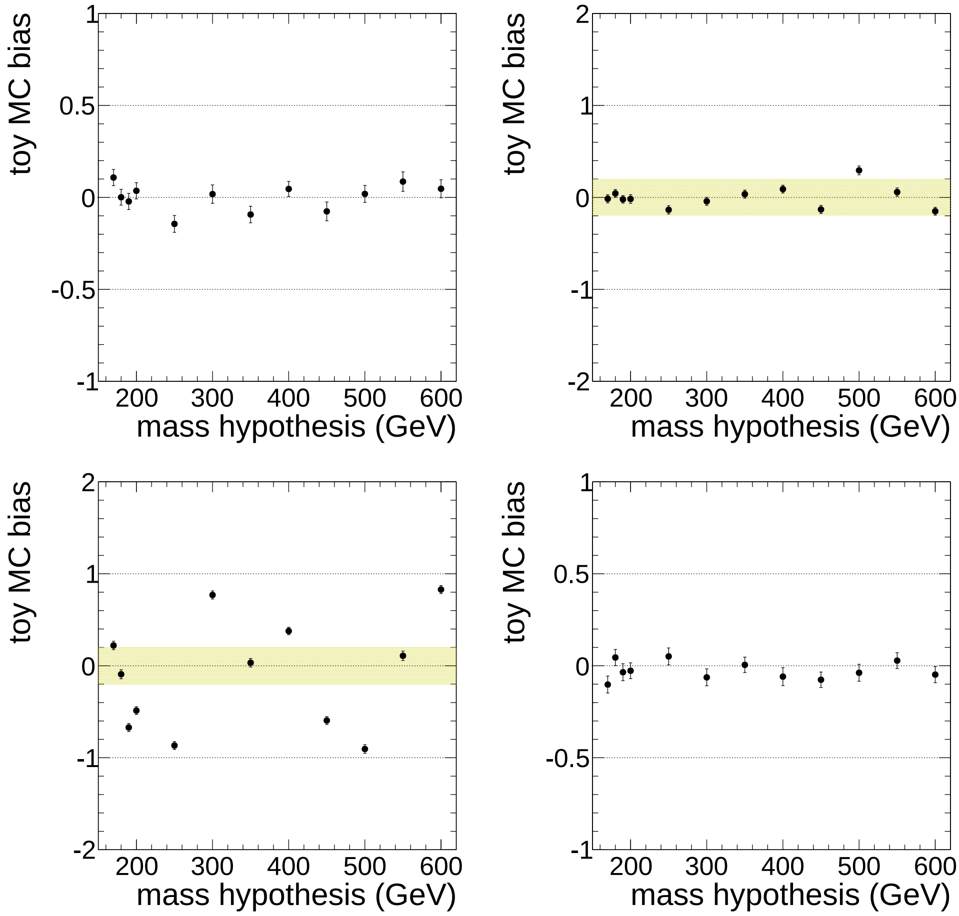


Figure 4.37: Results of the shape cross checks. The first row of plots generated using the nominal background shape, the second row with the polynomial background shape. The first column of plots is fitted with the nominal shape, where as the second column is fitted with the polynomial shape.

that the background function has enough flexibility to adapt to the different functional form. The results are shown in figure. 4.38.

4.12.3 Fit validation

Further, the fitting procedure were validated using an ensemble of pseudo-data experiments thrown from the signal and polynomial background models. The cases of 0×, 0.5× and 1× of the SM for the amount of Higgs signal included in each pseudo-data experiment were considered. The means of the Pull distributions for the signal strength parameter r are shown in figures. 4.39-4.41. In all the cases the bias is negligible, within the statistical errors returned by the Maximum likelihood (ML) fit. It may be noted that the Pull distribution is expected to look like Gaussian distribution with mean and standard deviation of zero and one respectively in an unbiased case.

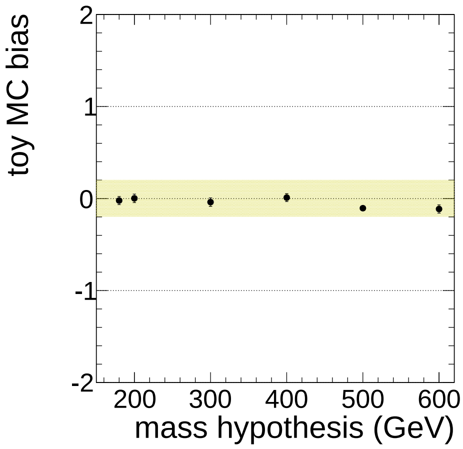


Figure 4.38: Results of a shape cross check using alternate shape.

4.12.4 Determination of the normalization from fits to dijet mass

The background yields are extracted from an unbinned maximum likelihood fit to the dijet invariant mass distribution m_{jj}, after the selection on the MVA discriminant, excluding the signal region (66 GeV $< m_{jj} <$ 98 GeV). In this fit, all the backgrounds are considered. The W+jets shape is as described in section 5.8.1, while the other background shapes come from theory expectation (i.e. from MC). The dijet mass spectrum fit provides inputs for the yields of the physics processes. Table 4.5 shows a schematic view how the shape of each component is determined, and what constraints are applied to fit for the normalization.

Table 4.5: Determination of the m_{jj} and $m_{\ell\nu jj}$ shape and normalization.

Process	Shape .	Normalization	Norm. syst.
V+jets	data/sim.	Unconstrained (m_{jj})	Unconstrained (m_{jj})
		from m_{jj} fit ($m_{\ell\nu jj}$)	m_{jj} fit uncertainty ($m_{\ell\nu jj}$)
Diboson	simulation	89.43 [95]	Lognormal: 3.4% (NLO)
top	simulation	338.74 [96]	Lognormal: 7% (NLO)

The resulting fits are shown in figures. 4.42–4.53 with the event yields of the different physics processes.

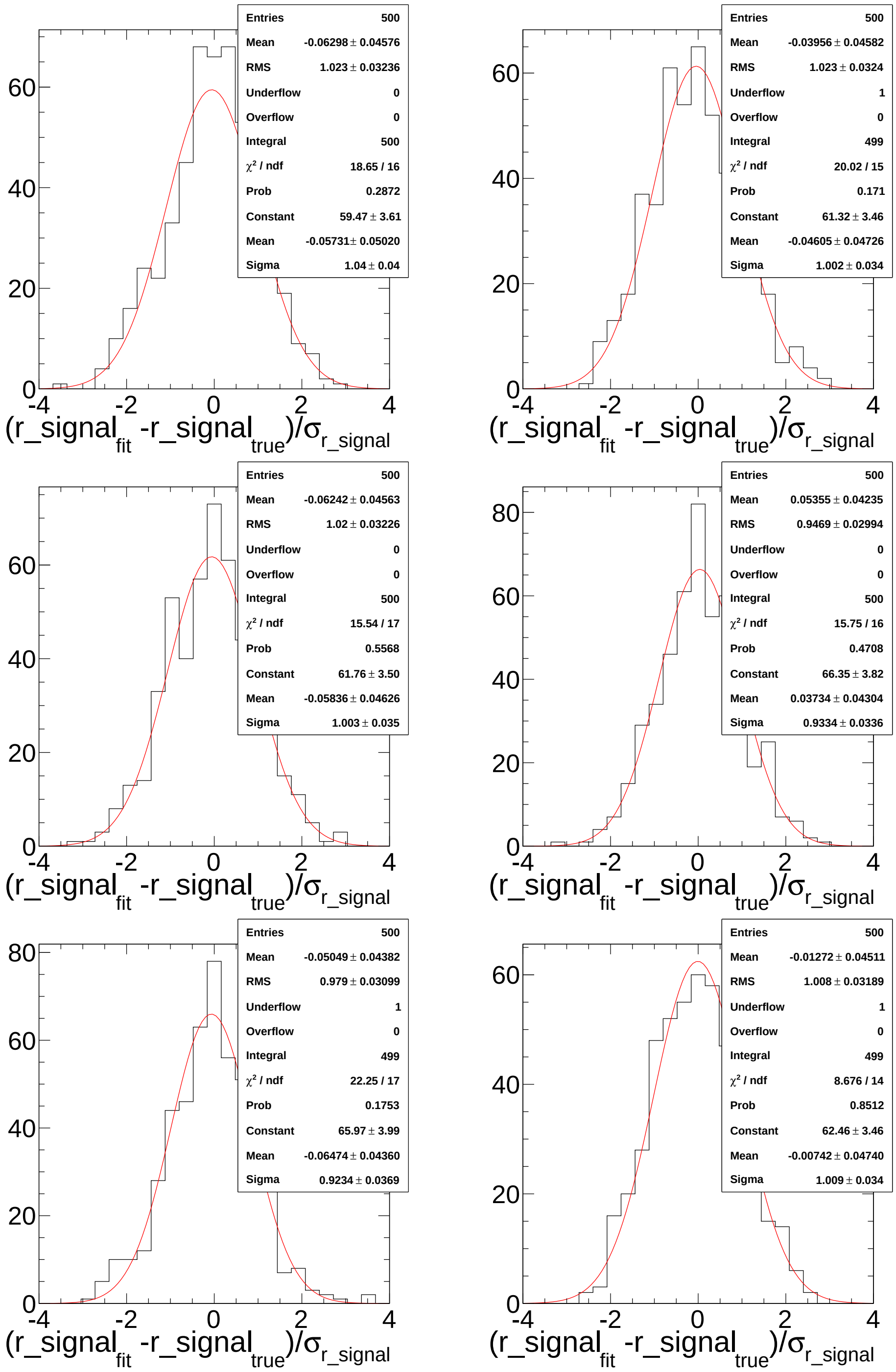


Figure 4.39: Pull distribution $(r_{\text{fit}} - r_{\text{gen}})/\sigma_r$ for the mass points beginning at the upper left 180, 200, 300, 400, 500, 600 GeV under the zero signal generation.

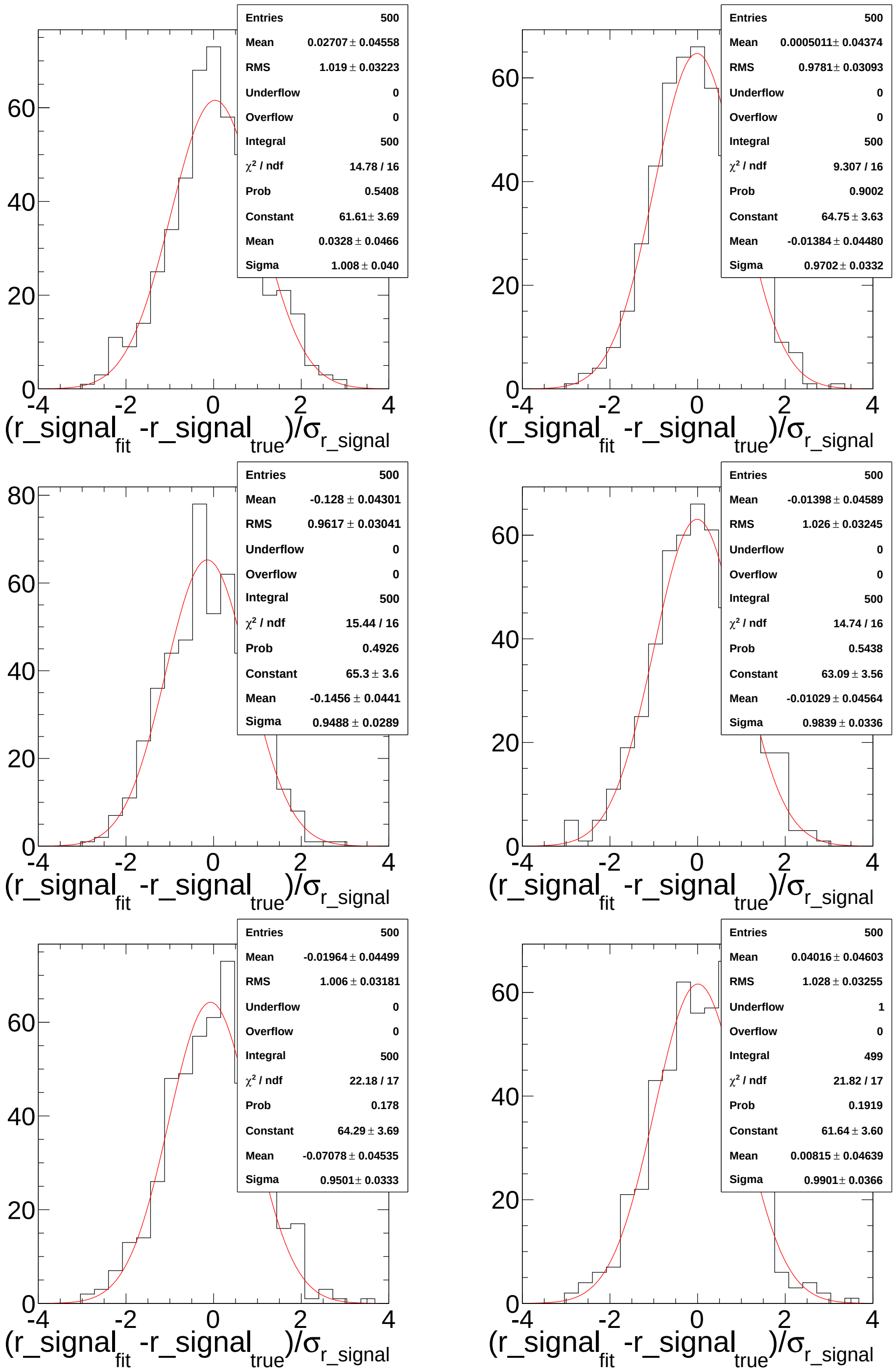


Figure 4.40: Pull distribution $(r_{\text{fit}} - r_{\text{gen}})/\sigma_r)$ for the mass points beginning at the upper left 180, 200, 300, 400, 600 GeV under the 0.5× signal generation.

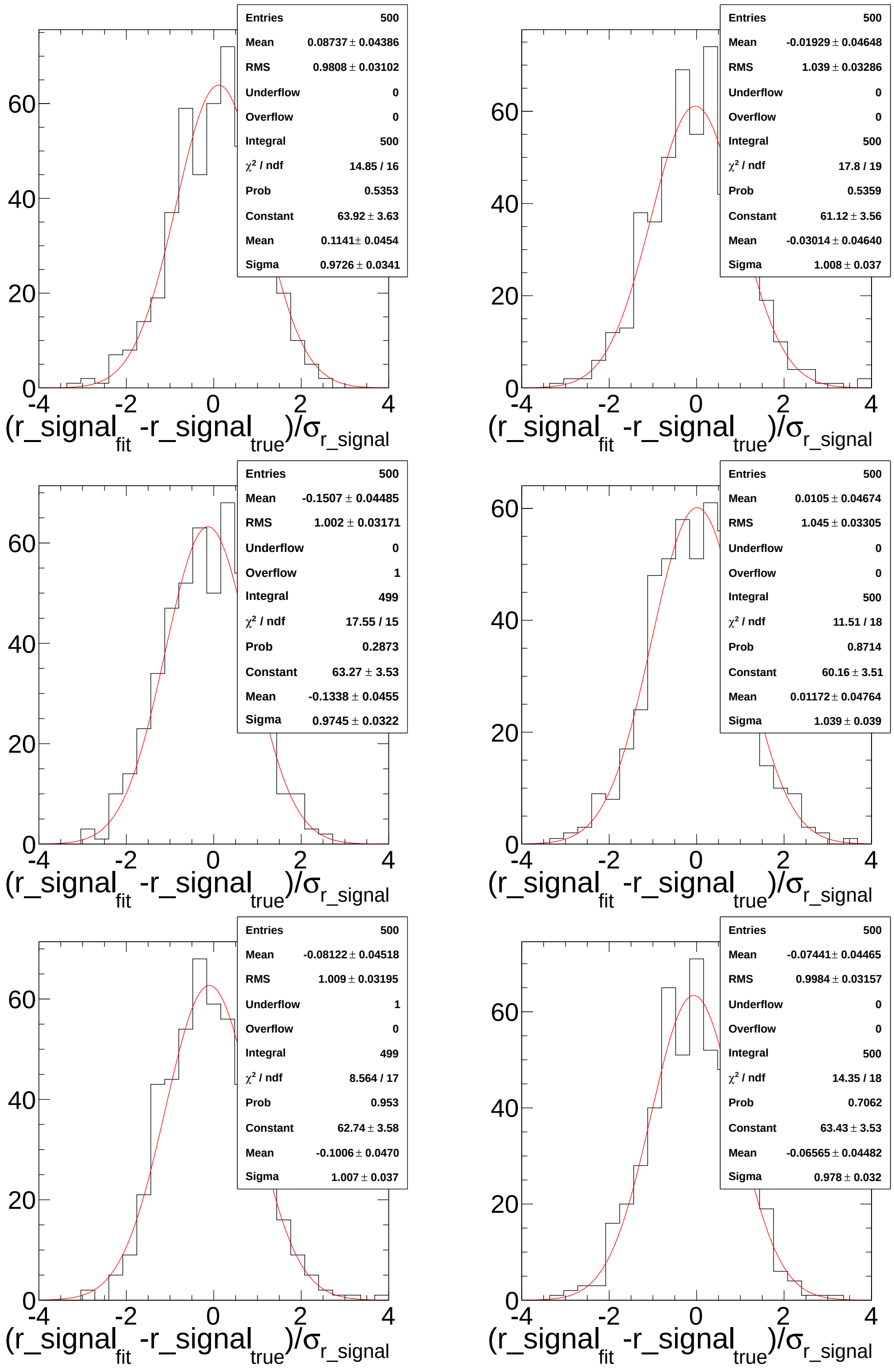


Figure 4.41: Pull distribution $(r_{\mathrm{fit}} - r_{\mathrm{gen}})/\sigma_r$) for the mass points beginning at the upper left 180, 200, 300, 400, 500, 600, under the 1× signal generation.

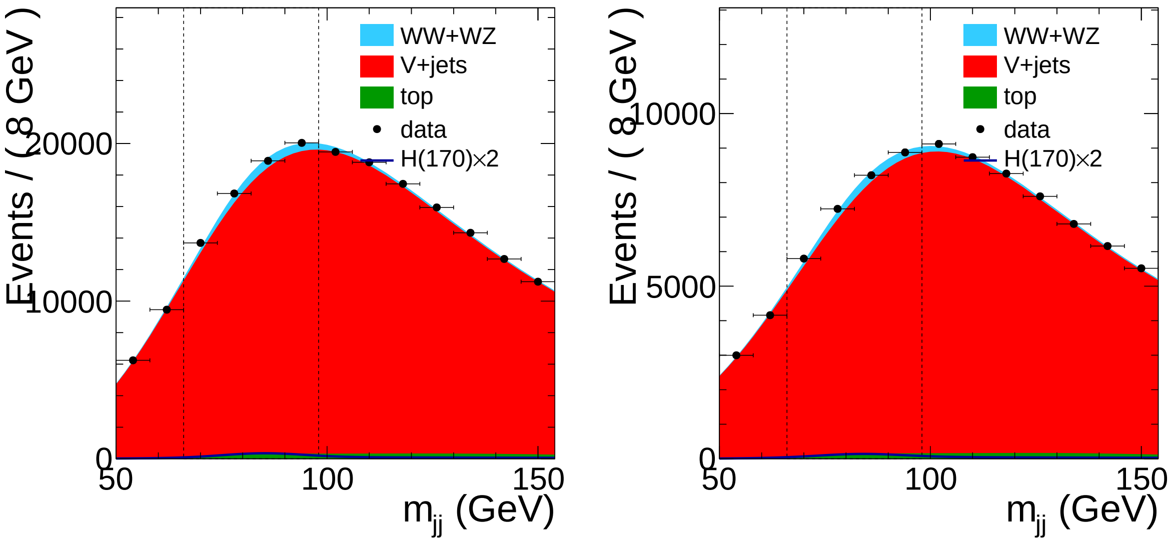


Figure 4.42: For the SM Higgs boson mass of 170 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

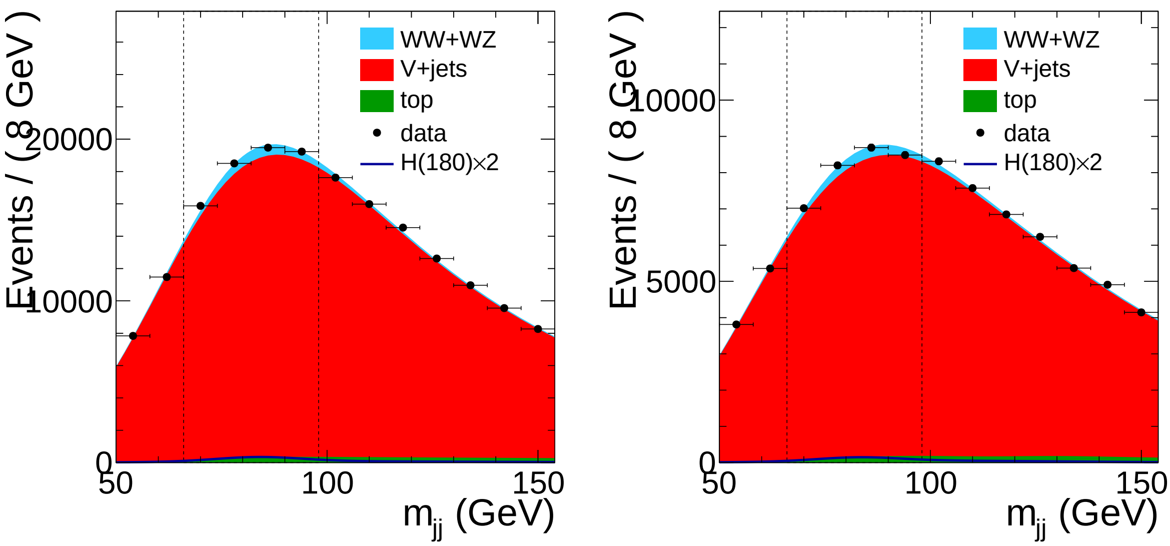


Figure 4.43: For the SM Higgs boson mass of 180 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

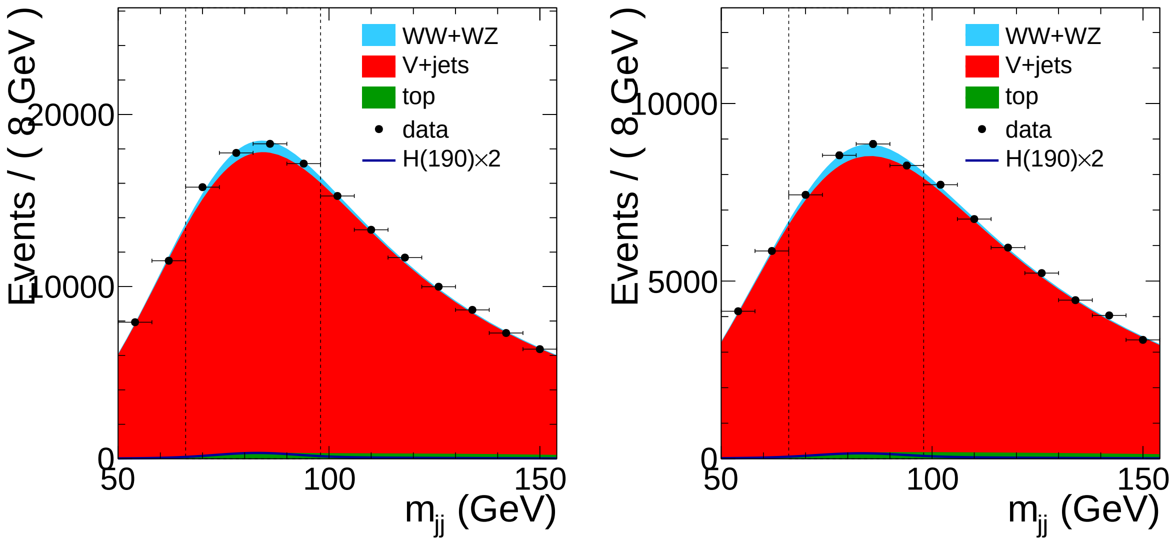


Figure 4.44: For the SM Higgs mass of 190 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

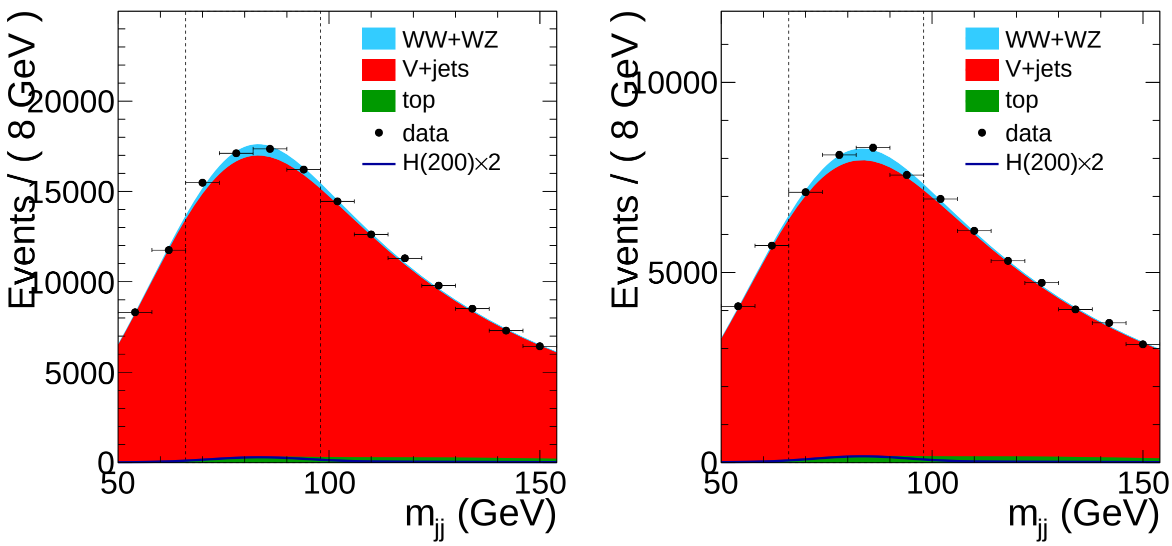


Figure 4.45: For the SM Higgs mass of 200 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

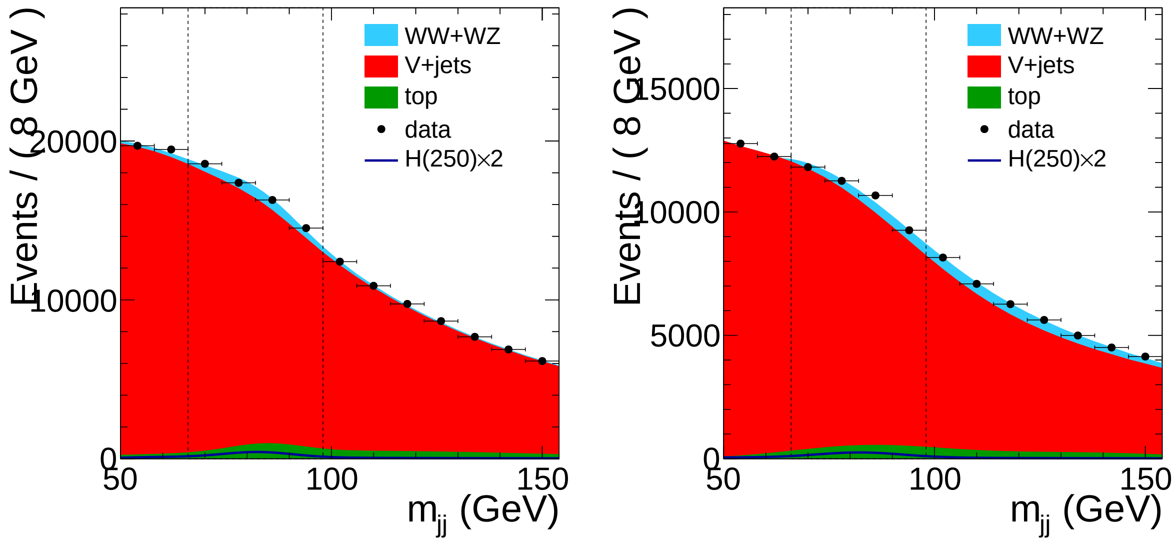


Figure 4.46: For the SM Higgs mass of 250 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

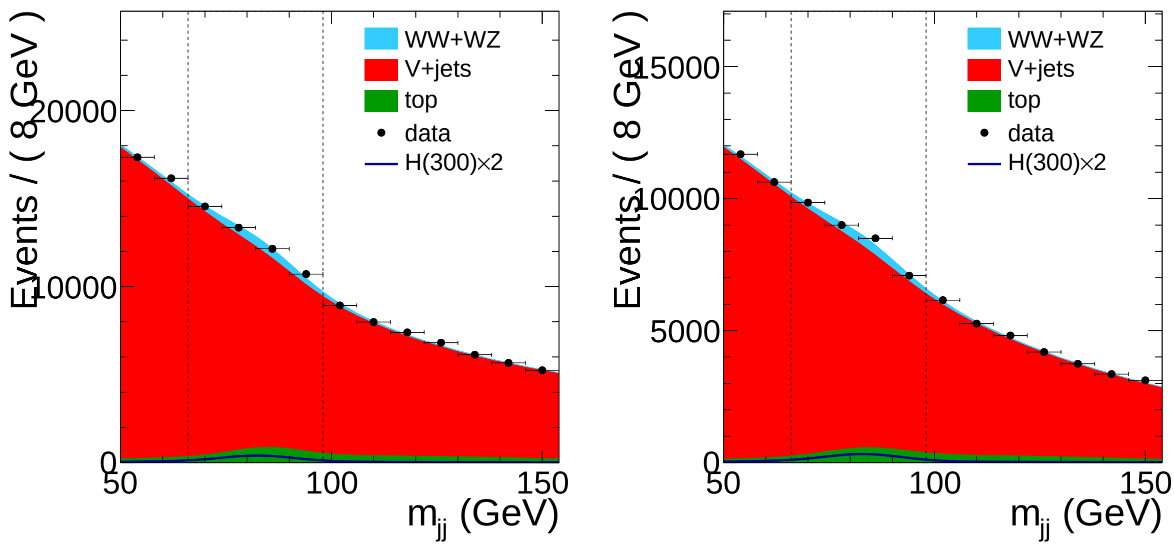


Figure 4.47: For the SM Higgs mass of 300 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

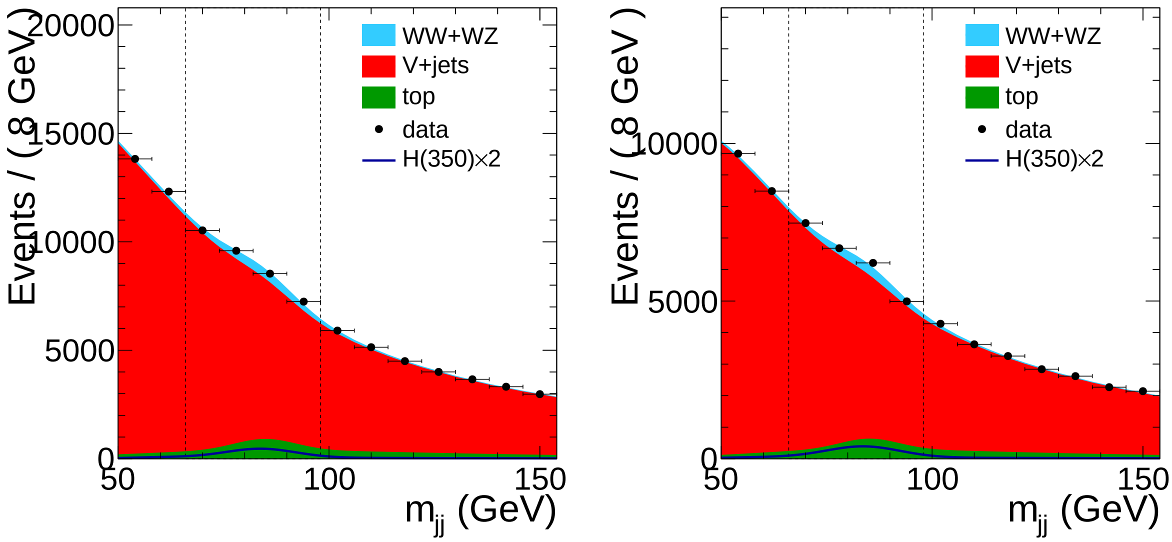


Figure 4.48: For the SM Higgs mass of 350 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

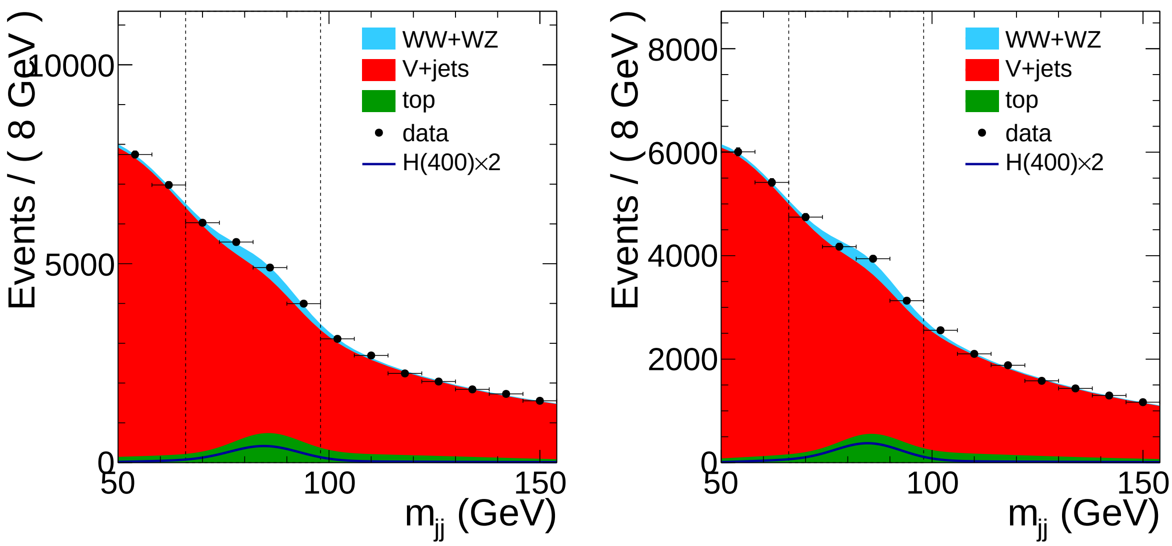


Figure 4.49: For the SM Higgs mass of 400 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

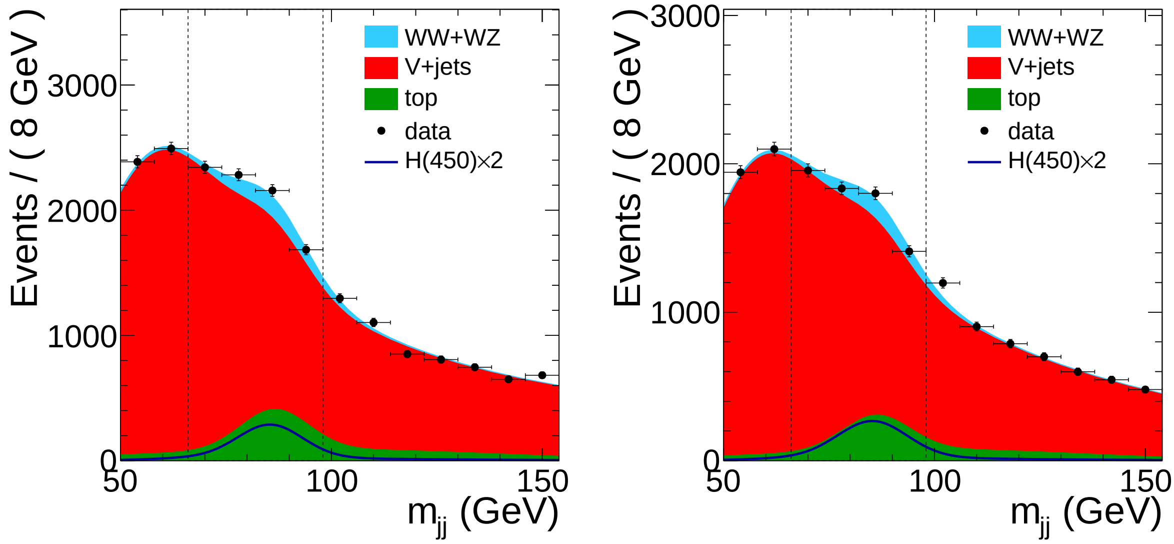


Figure 4.50: For the SM Higgs mass of 450 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

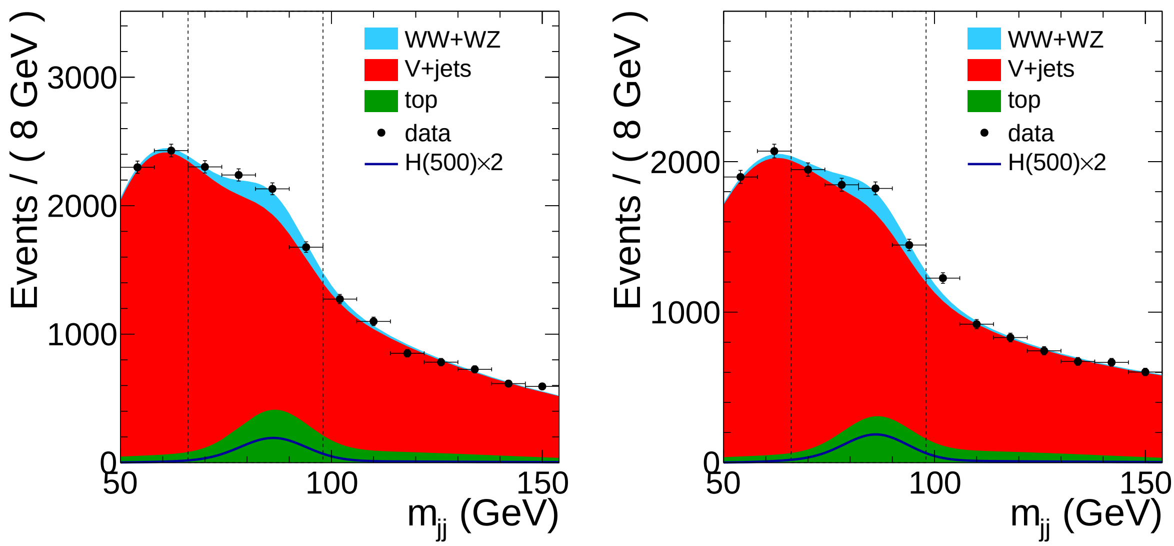


Figure 4.51: For the SM Higgs mass of 500 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

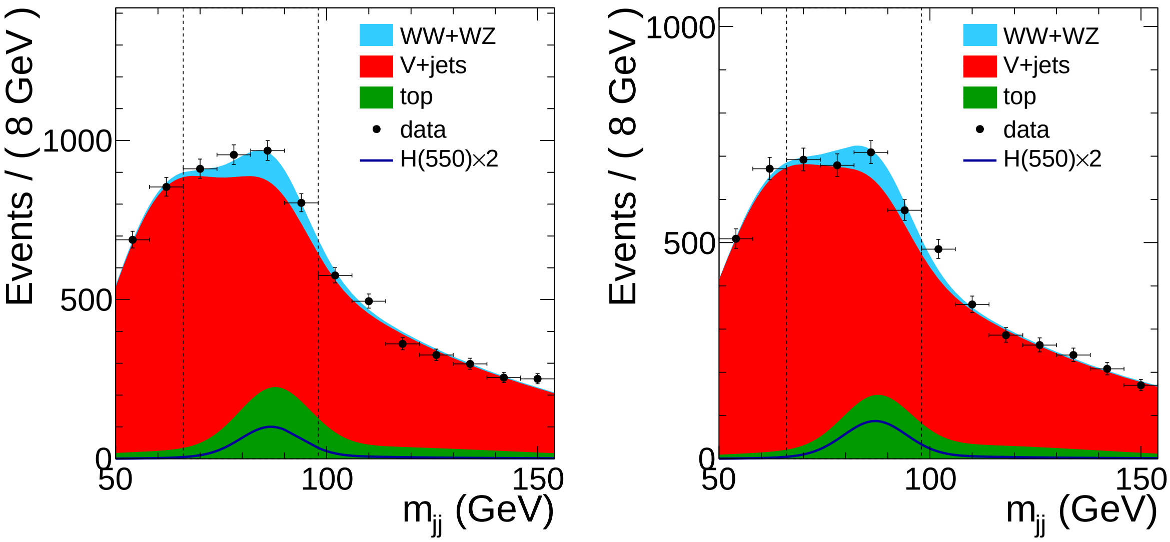


Figure 4.52: For the SM Higgs mass of 550 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

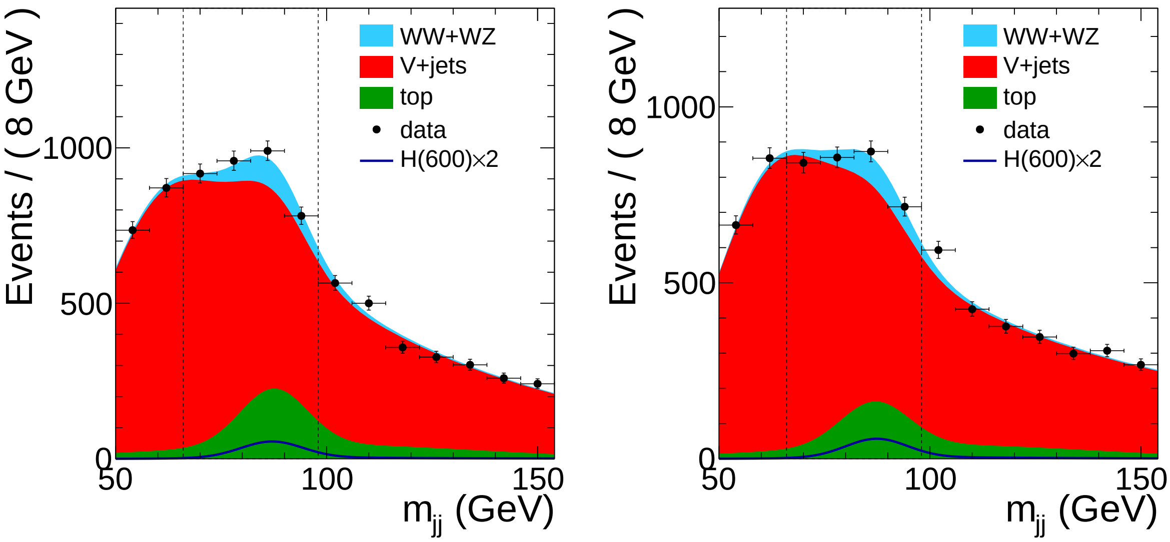


Figure 4.53: For the SM Higgs mass of 600 GeV, the distribution of the dijet invariant mass m_{jj} is shown with μ+jets category (left) and e+jets category(right).

4.12.5 Use of four-body mass to extract Higgs limits

The four-body mass background distributions in the signal regions after the maximum likelihood (ML) fit to the data under the background only hypothesis are shown in figures 4.54-4.65.

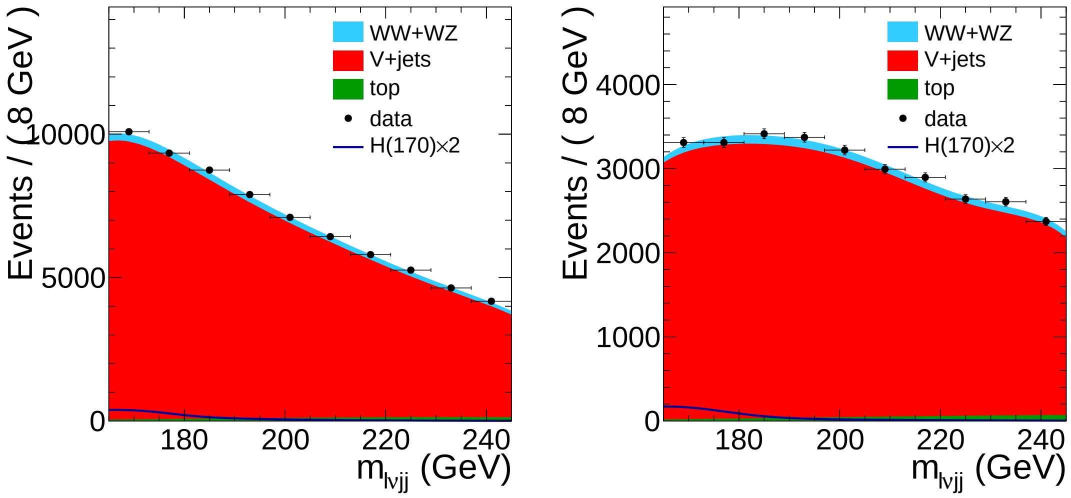


Figure 4.54: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=170 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

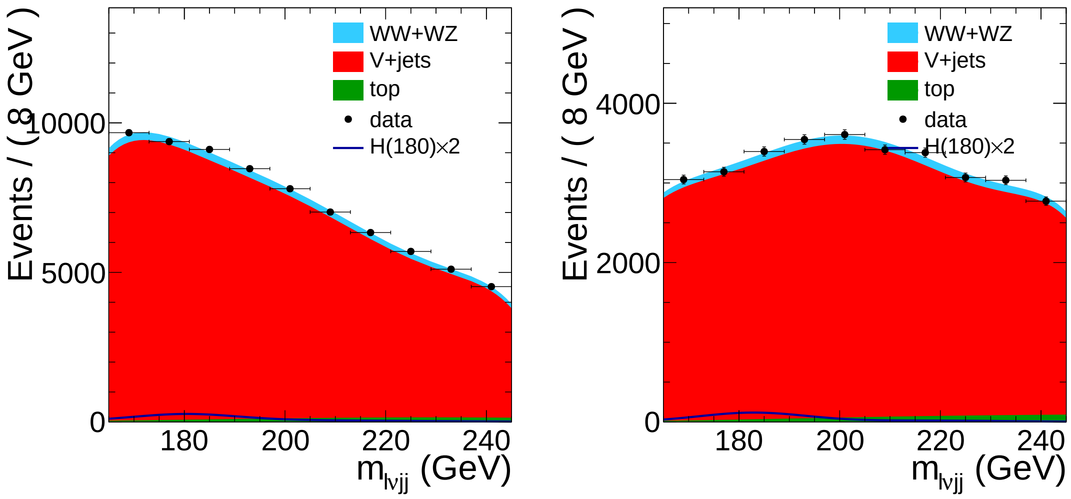


Figure 4.55: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=180 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

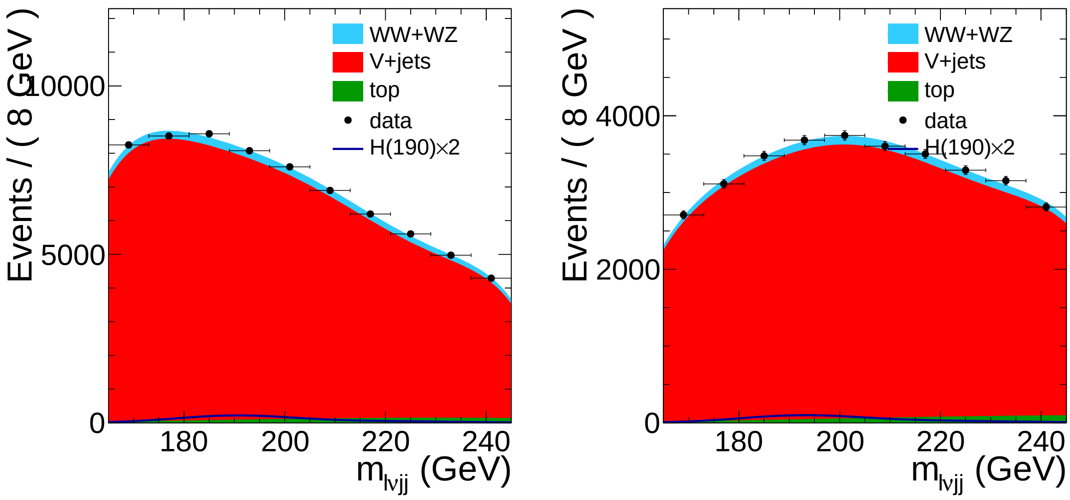


Figure 4.56: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=190 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

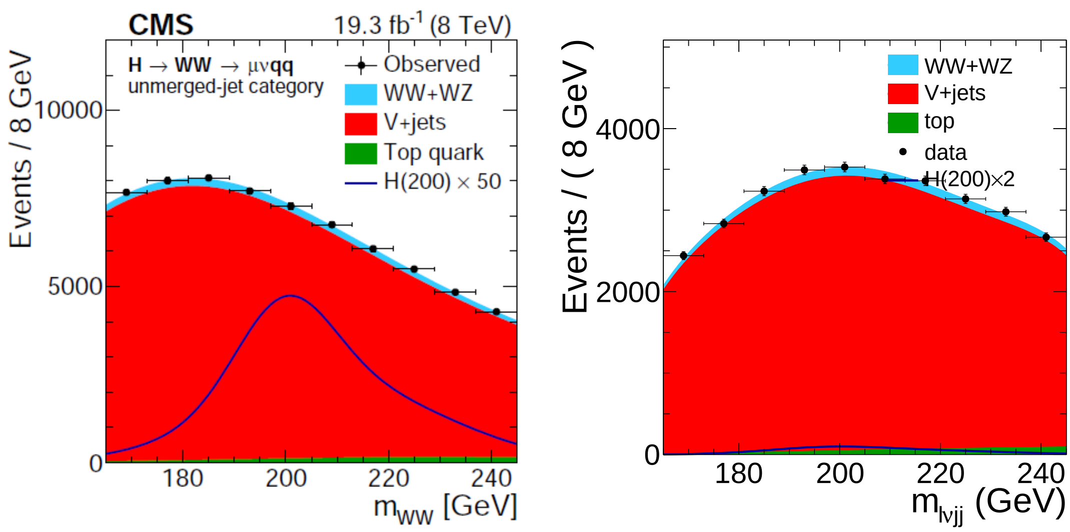


Figure 4.57: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=200 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

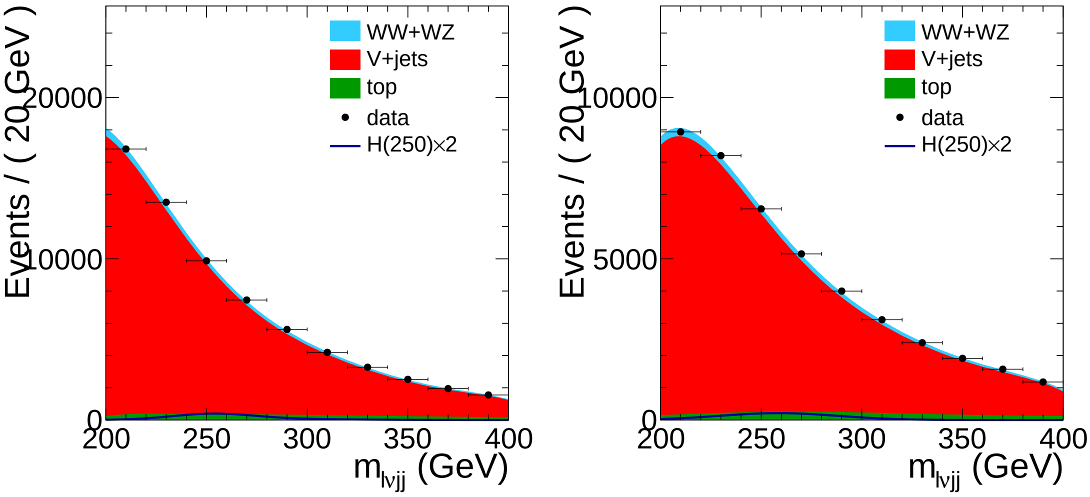


Figure 4.58: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=250 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

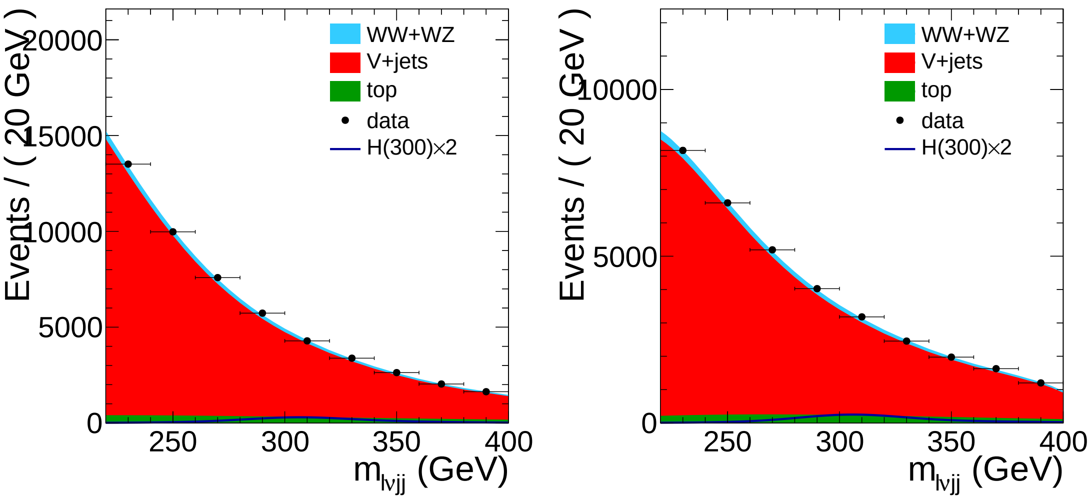


Figure 4.59: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=300 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

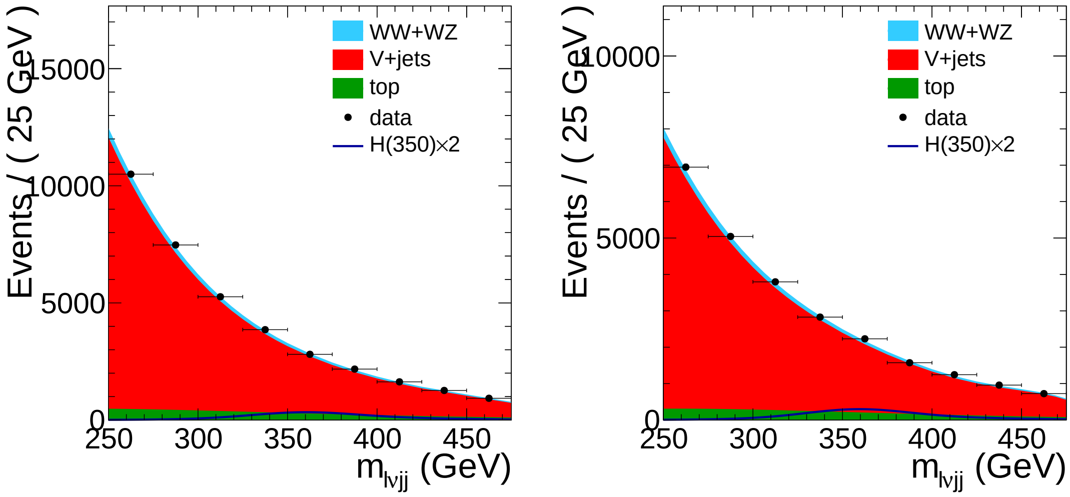


Figure 4.60: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=350 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

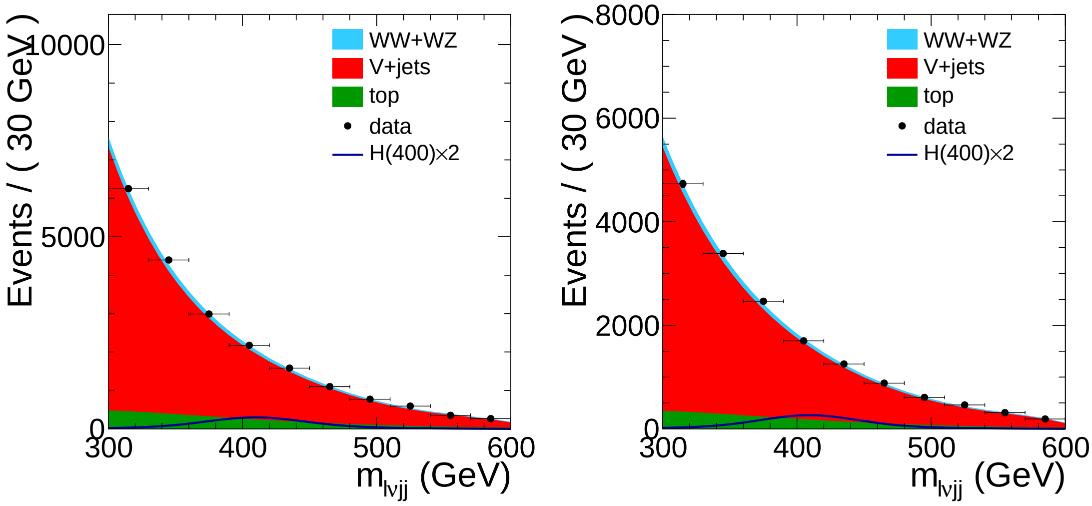


Figure 4.61: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=400 GeV. The left plot is for the μ+jets category; the right is for e+jets category.

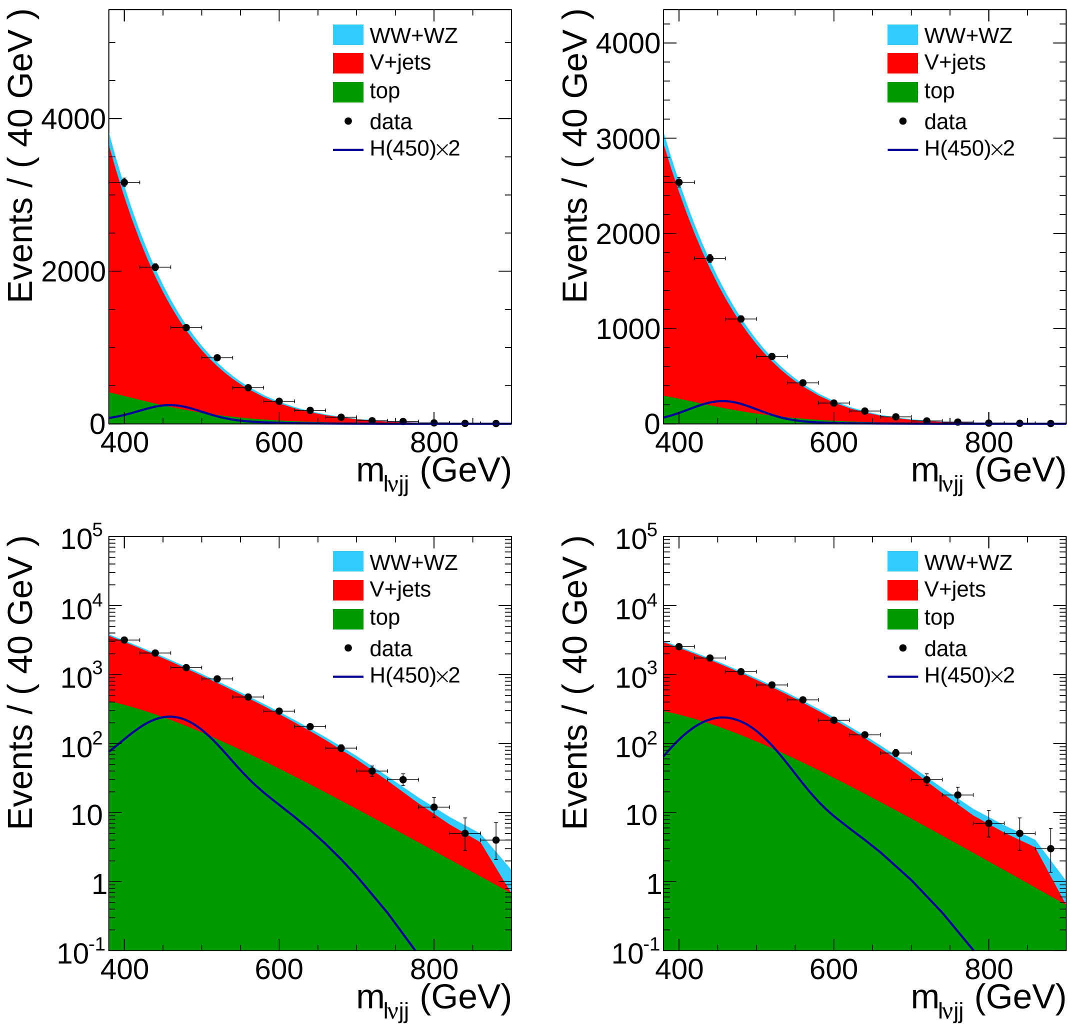


Figure 4.62: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=450 GeV. The left plot is for the μ+jets category; the right is for e+jets category. The bottom set of plots is the same plot with a log scale rather than linear.

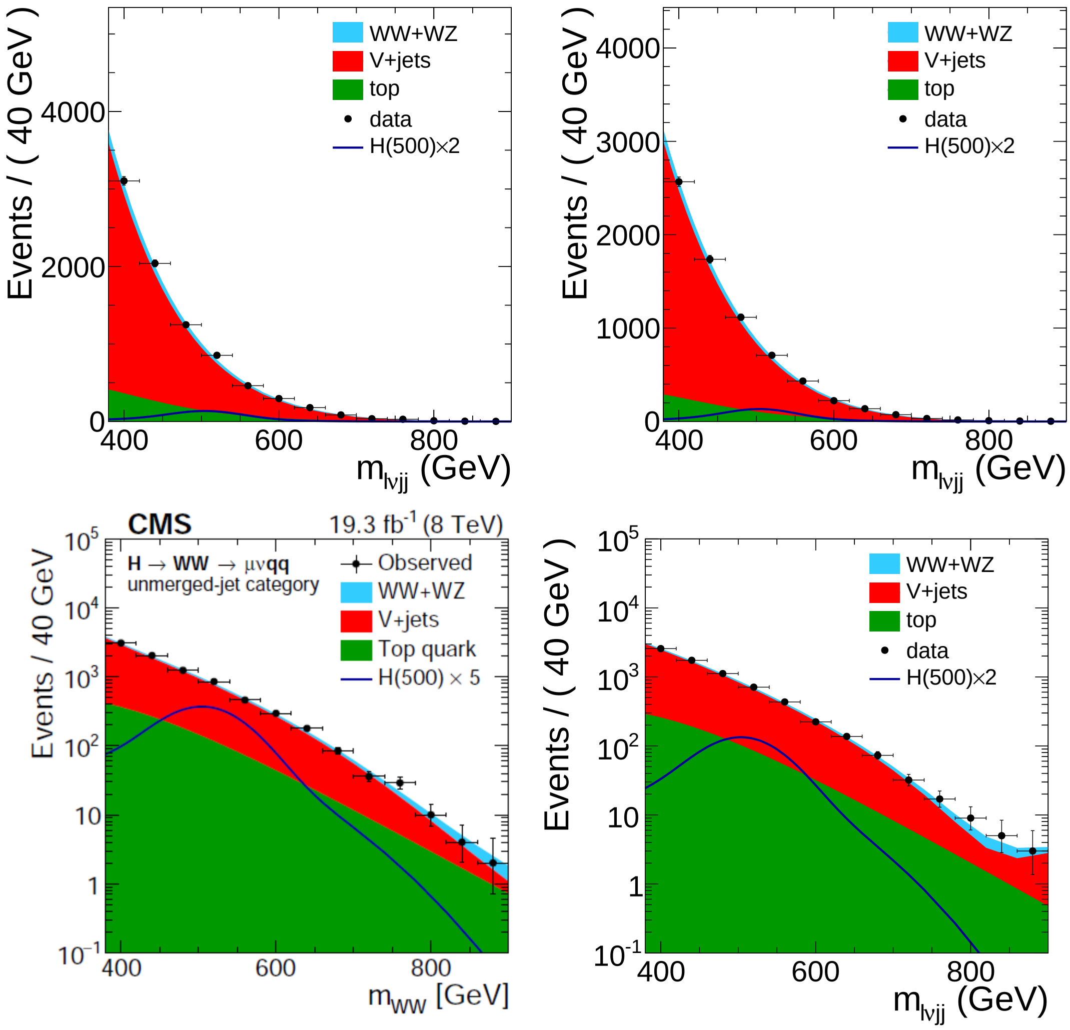


Figure 4.63: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=500 GeV. TThe left plot is for the μ+jets category; the right is for e+jets category. The bottom set of plots is the same plot with a log scale.

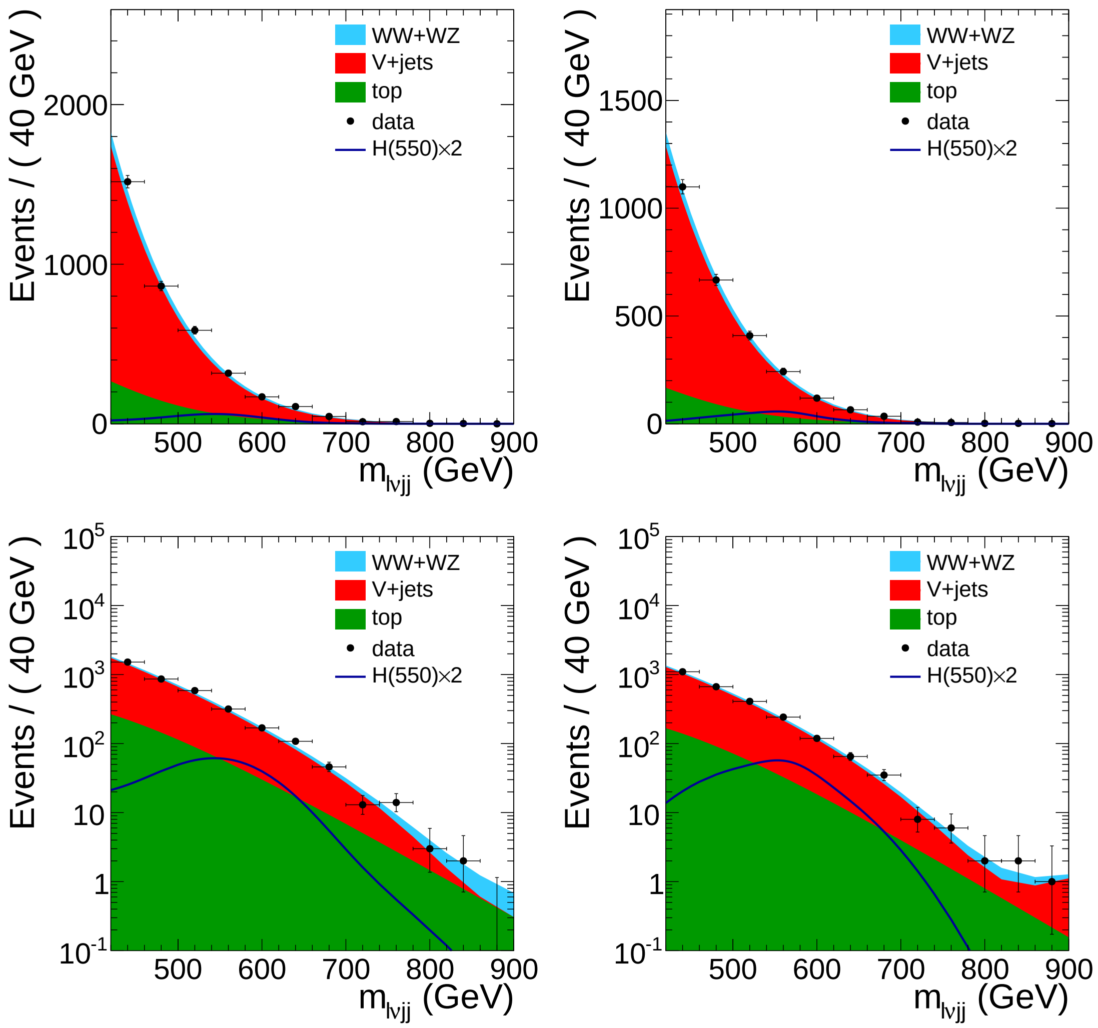


Figure 4.64: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=550 GeV. TThe left plot is for the μ+jets category; the right is for e+jets category. The bottom set of plots is the same plot with a log scale.

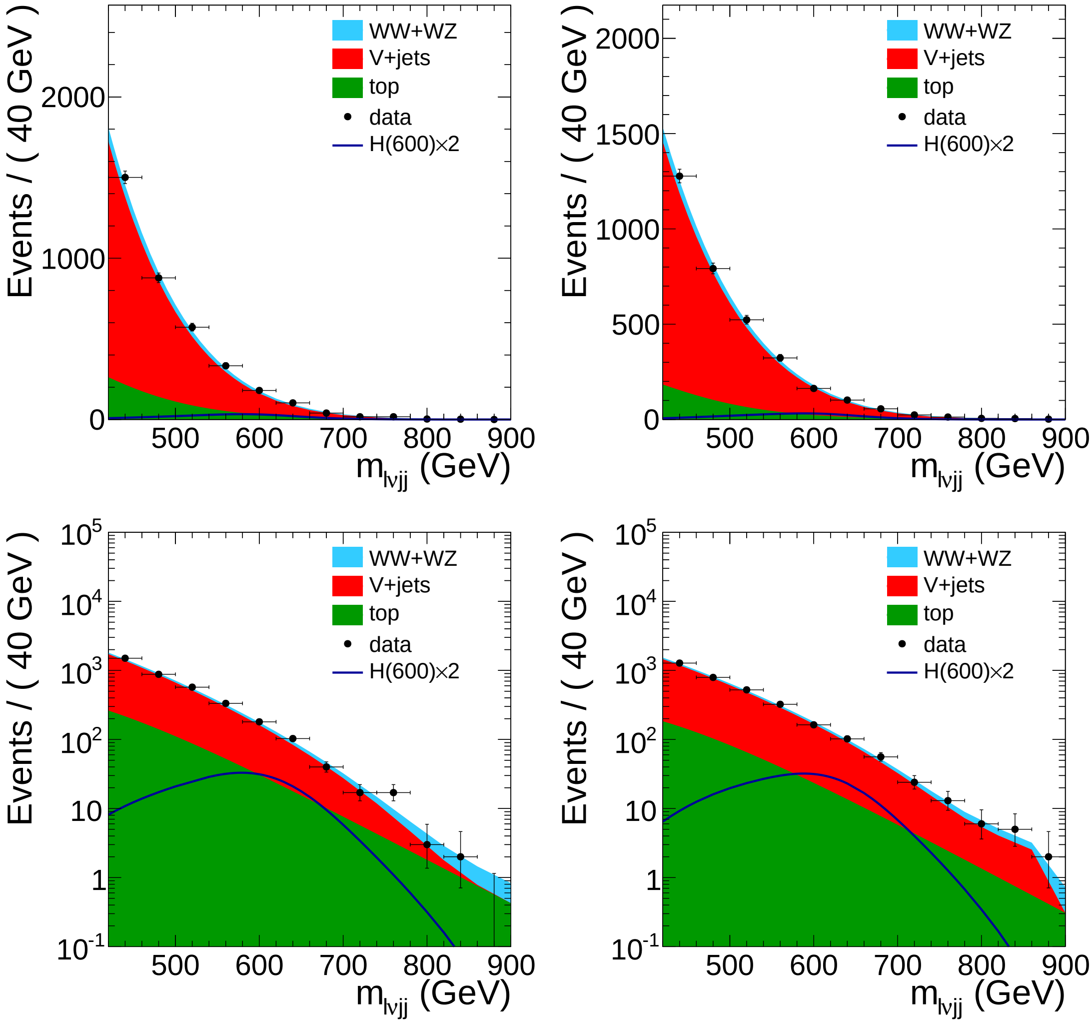


Figure 4.65: The four-body mass distribution of data-driven background events in the signal regions for Higgs mass hypothesis point M=600 GeV. The left plot is for the μ+jets category; the right is for e+jets category. The bottom set of plots is the same plot with a log scale.

4.13 Systematics uncertainties

4.13.1 V+Jets shape and normalization

The largest source of systematic uncertainty is the shape uncertainty of the V+jets in the $m_{\ell\nu jj}$ distribution. This is obtained as error on fit of four body mass described in section 4.12, while the V+jets background normalization uncertainty comes from the m_{jj} fit described in Section 4.12.4.

4.13.2 Jet energy scale

To evaluate the uncertainty due to jet energy scale in events with the topology of this analysis, we reconstruct the hadronic W boson candidate from an almost pure top quark data control sample. A semileptonic top sample is a good proxy for signal for the purposes of this study, since the top quark pairs are produced by gluon fusion and decay to two W bosons. These semileptonic top events are selected by requiring exactly four jets in the event, out of which two are b-tagged and the other two are anti-b-tagged. The hadronic W boson candidates are formed from two anti-b-tagged jets. The invariant mass of the hadronic W boson candidates in the combined channels is shown in Figure 4.66, for the data and MC, together with a gaussian fit on the peak of each distribution. The relative difference between the Gaussian means in data and MC is used as jet energy scale uncertainty, and propagated through the template fits in the backgrounds, as well as to the signal shapes in the limit setting.

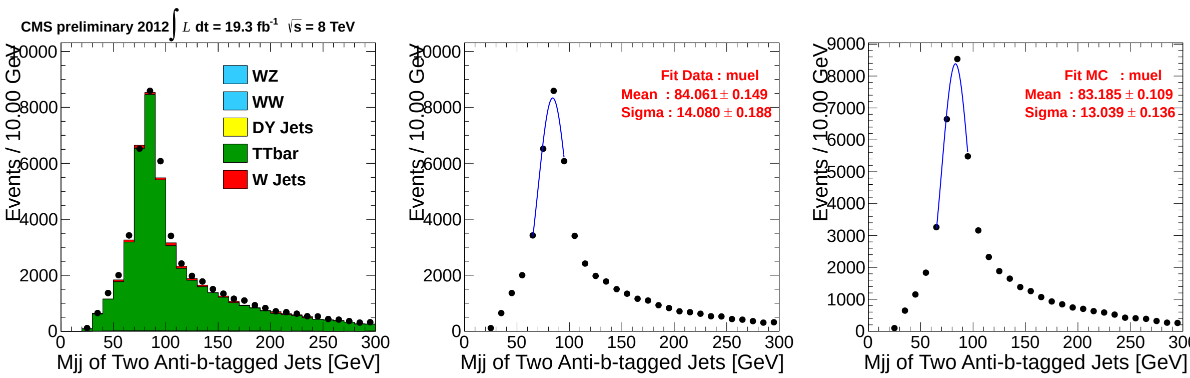


Figure 4.66: The invariant mass distribution of the hadronic W boson candidates in the semileptonic top sample (e+jets and μ+jets category combined). The left plot shows good agreement between the data and MC. The distribution is fitted with a Gaussian and peak location is extracted the for data (middle) and MC (right).

4.13.3 Final selection efficiency on the signal

The systematic associated with the efficiency on the final selection of the MVA discriminator is studied by using the same top pair events as described above. There is in reasonable agreement between the MC and the data for the top sample. The differences in selection efficiency are used to measure the potential error in the signal efficiency for each mass point / channel combination. The uncertainty is then taken as

$$100\% \times (1 - \frac{\epsilon_{data}}{\epsilon_{MC}}).$$

The distribution of measured uncertainties per mass point/channel combination is shown in Fig. 4.67. The measured efficiency uncertainties varied from less than 1% to 10%. We therefore conservatively take 10% as the signal selection efficiency uncertainty for all channels and mass points. It was verified that this conservative selection had no significant impact on the final expected limit.

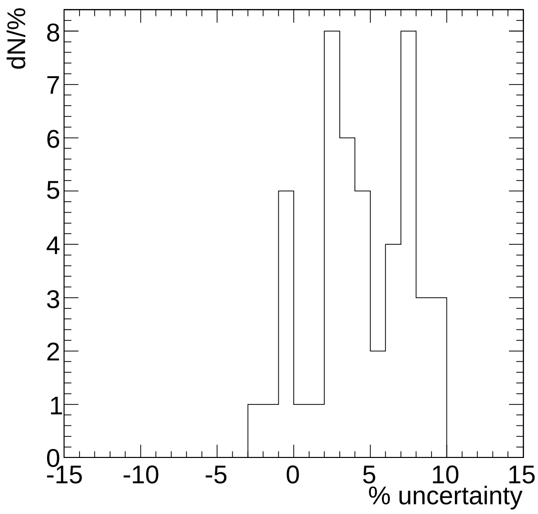


Figure 4.67: The distribution of measured uncertainties on signal selection efficiency, one entry per channel/mass point combination.

4.13.4 Lepton selection and trigger efficiency

Systematic uncertainties in the trigger efficiencies are of the order of 1%. Systematic uncertainties in the lepton reconstruction and identification efficiency scale factors are of the order of 2%. These uncertainties are accounted for in the final systematics that are input to the limit setter.

4.13.5 MET uncertainty

E_T^{miss} directly affects our signal acceptance. The uncertainty prescription is discussed in Ref. [97]. In addition, the E_T^{miss} distribution in the data is $\simeq$3% wider than the MC, and placing a hard $E_T^{miss} > 30.0$ threshold creates an uncertainty. So, it was estimated by smearing the E_T^{miss} for each event by a Gaussian with a $\sigma = 0.03{*}\text{MET}$ and observing how many events pass the cut. Specifically, (Events Passing After Smearing)/(Events Passing Before Smearing) =0.998 for both mu+jets and e+jets categories.

4.13.6 Pile-up model

The average number of pile-up interaction in a given bunch crossing BX_i is given by the following formula:

$$N_i = \frac{\mathcal{L} \cdot \sigma_{\text{min. bias}}}{\nu_{\text{orbit}}}, \tag{4.8}$$

where $\mathcal{L}$ is the instantaneous luminosity, $\sigma_{\text{min. bias}}$ is the cross-section of minimum bias interactions and ν_{orbit} is the LHC orbit frequency (11246 Hz). Source of uncertainties in the estimation of the number of pile-up interactions in the data then comes from the uncertainty on the luminosity, currently $\text{syst}_{\text{lumi}} = 2.6\%$, and the uncertainty on the minimum-bias cross-section. We have adopted $\sigma_{\text{min. bias}} = 69.3$ mb.

A total variation of 5% in the number of interactions was propagated to the re-weighting procedure for signal samples, and the obtained variation in the signal yield is used as systematics on the signal. The typical effect is less than a percent and therefore neglected.

4.13.7 Cross-section prediction

The inclusive cross-sections used for the Higgs signal at 8 TeV center-of-mass energy have been calculated by the Higgs Cross Section Working Group [77] for the gluon-gluon fusion process, and have been used in the limits extraction, together with their uncertainties, which are of the order of 15-20%.

In addition, the acceptance effect due to the PDF choice has been studied by following the PDF4LHC recipe, that considers as uncertainty the envelope of the error calculated for three sets of PDFs [98], namely CT10, NNPDF and MSTW. Table 4.6 shows the values obtained. For the purposes of the limits calculation, these systematics are added in quadrature to the inclusive cross-section uncertainties.

4.13.8 LHC luminosity

A luminosity uncertainty of 2.6% is applied to all simulation samples (signal and the minor backgrounds) [99].

Table 4.6: Acceptance uncertainty related to the PDFs variation, for the signal rate, as a function of the mass hypothesis.

ggF		VBF	
m_H	unc.	m_H	unc.
170	2.0%	170	2.0%
180	2.0%	180	2.0%
190	2.0%	190	2.0%
200	2.0%	200	2.0%
250	1.5%	250	1.1%
300	2.0%	300	0.9%
350	2.2%	350	0.8%
400	2.4%	400	0.6%
450	2.7%	450	0.7%
500	2.9%	500	0.9%
550	3.2%	550	0.9%
600	3.6%	600	0.7%

4.13.9 Uncertainties related to interference

Section B.1 describes the procedure for reweighting signal samples to correct for the effect of interference between the signal and the SM $gg \to WW$ production process. It also describes the recipe for estimating the uncertainty band associated with the weight factors.

Since the effect of interference enhances the cross-section of the signal, this uncertainty band propagates through the analysis in two ways: it changes the efficiency of the analysis selections on signal events, and it alters the shape of the signal invariant mass distribution that is used to set the upper limit. These are treated simultaneously by applying the up- and down-fluctuated weights to the signal distribution, in the same way as the nominal weights are applied. These alternative shapes are also propagated to the limit setter. Figure 4.68 shows an example of this variation for a Higgs mass of 600 GeV, for which, of all the mass hypotheses considered, the effect is largest.

4.14 Extraction of the limit on the cross-section

One strength of this analysis, with the data-driven estimation of the W+jets background, is that we can treat the full mass range, 170-600 GeV, by the same methods. We use the “Higgs Combination” package [100] for setting exclusion limits. This package is a RooStats [101]-based statistical analysis toolset recommended by the CMS Higgs PAG. Inputs to the limit setter are RooStats objects representing probability density functions for the top, diboson, and V+jets backgrounds and signal MC (gluon-gluon fusion and

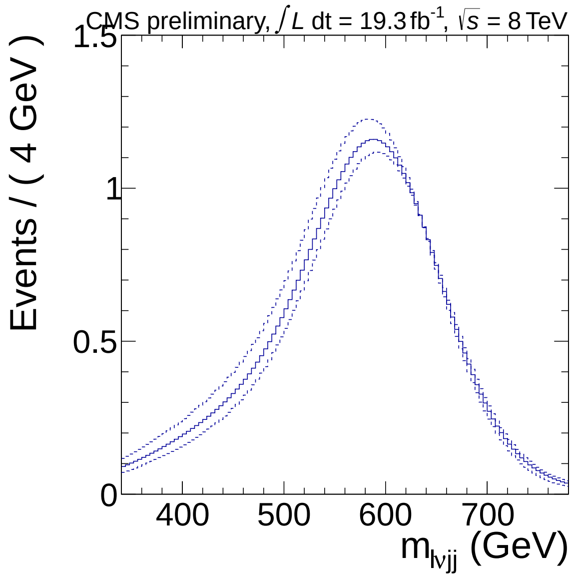


Figure 4.68: The shape variation for the signal $m_{\ell\nu jj}$ distribution, for a Higgs mass of 600 GeV, mu+jets category.

vector boson fusion), parameterized as described in Sections 4.11 and 4.12, and a dataset for data.

The Higgs bosn mass points are, nominally: 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 550, and 600 GeV. We scale the signal samples using the cross-section values taken from Ref. [78]. The SM Higgs boson cross sections are described further in Sec. 4.1.

To set limits we use the full shape information of the $m_{\ell\nu jj}$ distribution. The four-body mass window is set roughly by the position and width of the signal distributions for the different Higgs bosn mass points while trying to maintian adequate side-band for background determination. The exact values can be seen in Table 4.7. All of the distributions are segregated by lepton flavor, which represents independent channel inputs to the limit setter.

The systematics described in Section 4.13 are treated as follows when being input to the limit setter:

- The main background systematic uncertainties are total background shape uncertainty and total background normalization uncertainty, which are determined by the limit setter when performing the maximum likelihood fit. Both shape and normalization uncertainties are treated as uncorrelated across all channels, since they are derived from fits performed on independent sample sets.

- JES, E_T^{miss} uncertainty, and pileup are considered negligible for signal and are otherwise subsumed in the normalization/shape uncertainties for background, so that they are omitted from the limit setter inputs.

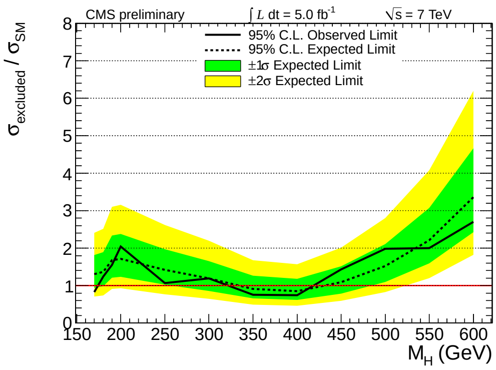
Table 4.7: Fit ranges and binning for the limit extraction. The binning was chosen because the combine tool chooses to do binned generation and fits of the "Asimov" dataset and we needed to minimize the discrepancy between the binned and unbinned fit results. In all cases the binning is about 10x finer than detector resolution. Because we have a large amount of background, we have relatively large statistics in mostly all the bins.

mass	lower limit (GeV)	upper limit (GeV)	N bins
170	165	245	100
180	165	245	100
190	165	245	100
200	165	245	100
250	200	400	100
300	220	400	90
350	250	475	90
400	300	600	100
450	380	900	130
500	420	900	120
550	420	900	120
600	420	900	120

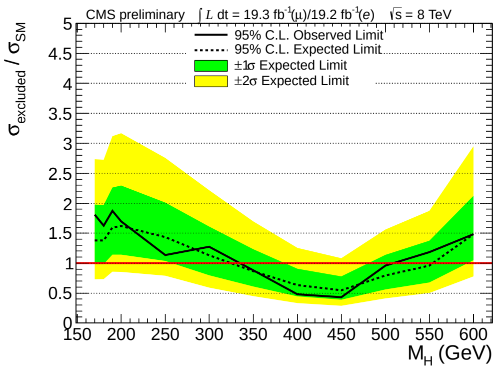
- Uncertainties on the signal deriving from lepton reconstruction and selection as well as trigger efficiency are treated as uncorrelated between μ+jets and e+jets categories.
- Uncertainties on the signal deriving from parton distribution functions (Table 4.6), luminosity, interference, and theoretical cross-section uncertainty are treated as 100% correlated across all categories.
- Uncertainties on the signal deriving from the final selection efficiency are treated as uncorrelated across different categories, but as correlated between quark-quark and gluon-gluon signal processes for the same channel.

The limit setter is then set to utilize the "asymptotic CL$_s$" [102, 103] method. The resulting median expected limit with 1- and 2-σ error bands are plotted. The limit plots for all categories combined (μ+jets and e+jets) are shown in Fig. 4.69. The background uncertainty is the limiting factor in this analysis.

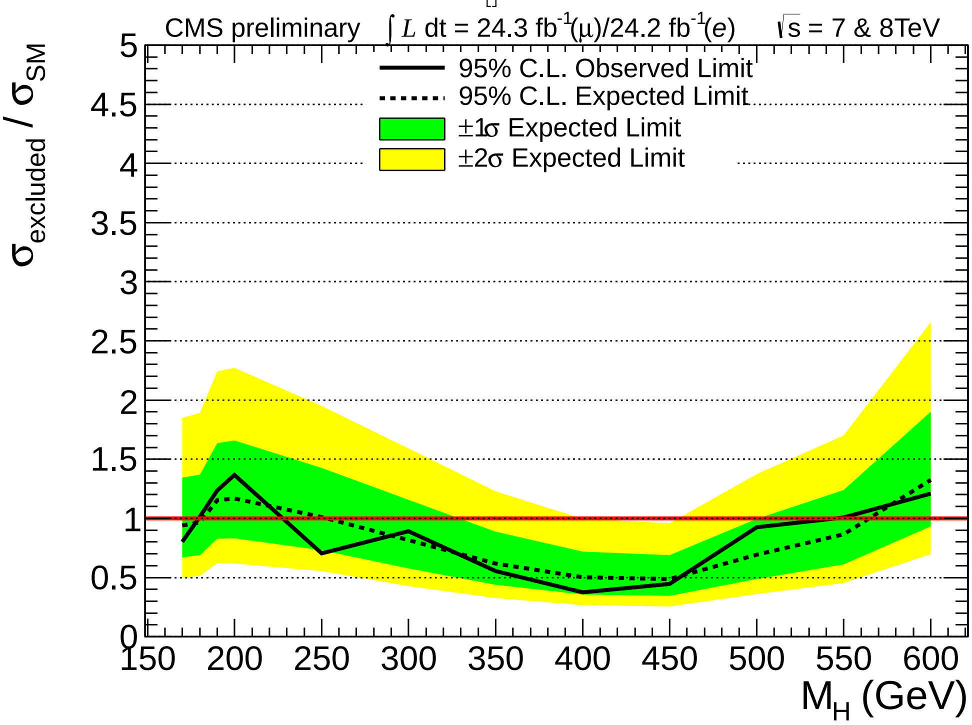
Using all categories of 8 TeV data combined, we exclude the Standard Model Higgs boson in the mass range 335–500 GeV to a confidence level of 95%. When combined with 7 TeV data [104], the exclusion range expands to 170–180 GeV and 230–545 GeV.



[]



[]



[]

Figure 4.69: The combined Higgs exclusion limit from all channels after including all systematic uncertainties (a) from 7 TeV data [104] (b) from 8 TeV data, and (c) for 7 and 8 TeV data combined.

4.14.1 Beyond the SM interpretations for Heavy Higgs

For the BSM electroweak singlet heavy Higgs hypothesis, a scan of grid points was performed in the C' vs. $\mathcal{BR}_{\mathrm{new}}$ for different value of M_H. The signal model was modified as described in Sec. 4.3.4 and then the same steps described above were performed for each grid point, using only the 8 TeV dataset. The results matched our expectation, that this analysis is most sensitive to the scenario with a lone heavy SM Higgs boson (particularly in the middle of our mass search window), which corresponds to C'=1.0 and $\mathcal{BR}_{\mathrm{new}}$=0.0. Although an electroweak singlet heavy Higgs has narrower width, the concomitant reduction in cross section and constructive interference only serves to reduce our sensitivity. Figure 4.70 shows contours in the C' vs. $\mathcal{BR}_{\mathrm{new}}$ plane for those mass hypotheses for which an exclusion contour can be drawn. These are the 95% C.L. exclusion contours for each labeled mass point; the space contained to the right and below each contour is excluded to 95% confidence.

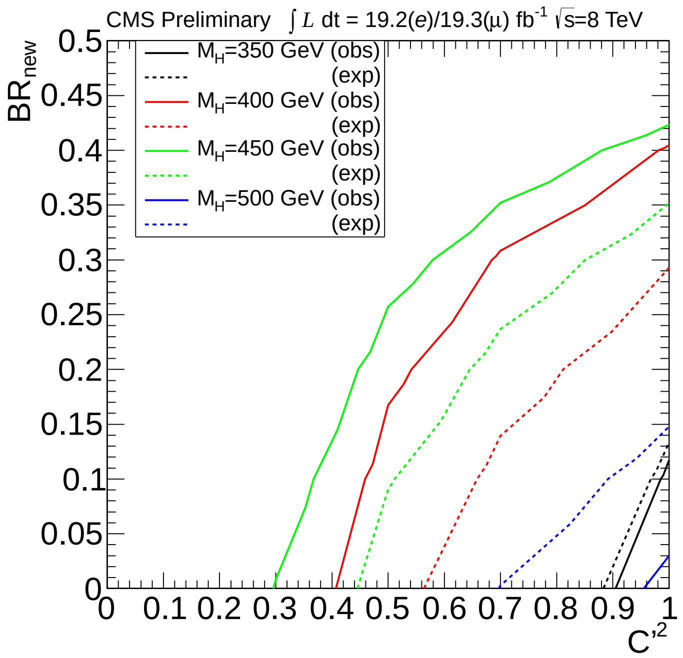


Figure 4.70: 95% confidence exclusion curves taken from a grid scan in the C' vs.$\mathcal{BR}_{\mathrm{new}}$ space. Observed curves are drawn with solid lines and expected curves are drawn with dashed lines. In both cases, the excluded region lies to the right and below the curves.

4.15 Conclusions

This study presents the search for a SM Higgs boson, in its decay into two W bosons, out of which one decays hadronically and the other one decays leptonically. The major background that contaminates the signal region is due to W+jets, and it is measured in a data-driven way from side-bands in the invariant mass of the hadronically decaying W boson. This allows us to keep the relative systematic uncertainties, which are dominated by the background shape particularly at low mass, to within a few percent for each Higgs mass point. Using 8 TeV data, SM Higgs boson in the mass range 335–500 GeV is excluded to a 95% confidence level. When this is combined with 7 TeV data, the exclusion range expands to 170–180 GeV and 230–545 GeV. The results were also interpreted within the context of a electroweak singlet heavy Higgs mixing with a light Higgs boson. Exclusion limits were set in the parameter space defining the coupling and branching ratios of the heavy Higgs.

Chapter 5

Vector Boson Scattering

The SM of particle physics has been extremely successful in predicting a vast variety of phenomena – so successful, that it is quite easy to forget the fact that some of its predictions still wide open and it needs attention. One such overwhelmingly important, intimately correlated to EW symmetry breaking, is that the gauge bosons (γ, W and Z) can interact with each other through quartic interactions. There are four such interactions allowed in the SM namely $WWWW$, $WWZZ$, $WWZ\gamma$ and $WW\gamma\gamma$. The other boson combinatoric couplings are forbidden on the ground of symmetry requirement. The first two of these interactions are of great theoretical interest as it provides a handle to explore unitarity principle non-violation in WW scattering via low mass Higgs boson unitarization. As mentioned in Chapter 1(Eq. 1.32), in a Higgs-less theory, new physics should appear ~ 1 TeV, but now since SM like Higgs boson has been discovered so unitarization of the amplitude of the longitudinally polarized component of WW may be explained. But, that needs to be verified experimentally. At the LHC, the $WWWW$ interactions can be studied through the radiation of W bosons from quarks (see figure 5.1). This interaction is

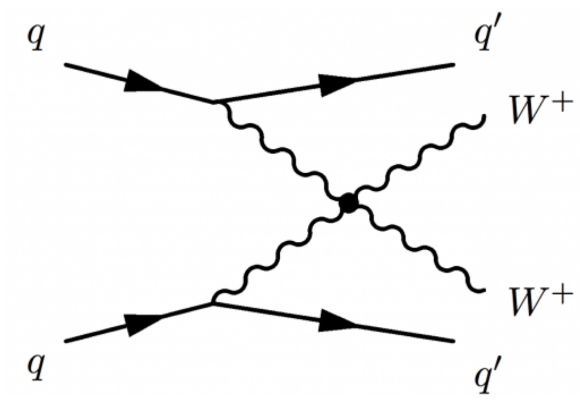


Figure 5.1: WW pair production in vector boson fusion topology, also an example of quartic gauge coupling.

a very rare process, which makes its detection extremely difficult. In fact there is already a tantalizing result, i.e. an evidence for $WWWW$ interaction in fully leptonic channel at LHC from both of the main experiments, the CMS and the ATLAS. The CMS and the ATLAS have performed analysis on selected events with two same-charge W bosons (reconstructed through their leptonic decays to electron or muon and their respective (anti)neutrinos) and two jets. The background from other SM processes are reduced due to the fact that these processes rarely produce two leptons with the same electric charge, together with the knowledge that the quarks that recoil off the radiated W bosons produce jets that are widely separated and have a particularly large invariant mass. This is the characteristic signature of these processes. It is interesting to pursue the same analysis in the semileptonic decay mode of one of the W boson. The final state is characterized by the presence of two highly energetic forward jets along with two central jets coming from decay of W or Z boson and a lepton and neutrino from other W boson. This analysis is pursued in the present Chapter using an integrated luminosity of 19.3 fb^{-1} recorded at $\sqrt{s} = 8\ TeV$ in 2012. Physics objects are selected with the standard techniques, described in Chapters 3 and 4. Almost all pre-selection thresholds are same as used in previous chapter. It may be noted that to reach the point of making a conclusive arguments in this decay mode requires large dataset which would arrive in Run2 of the LHC, but, this analysis will provide the tools and techniques and hence may be used as a baseline for any future work in this final state.

The main background components in the signal region are W+4 jets and $t\bar{t}$; its contribution is determined in a data-driven way. Minor backgrounds include, single top quark processes, diboson, and Z+Jets which are determined using Monte Carlos.

5.1 Significance of vector boson production in association with two jets

LHC allows us to probe vector boson scattering, together with triple and quartic gauge couplings 5.1. In addition vector boson pairs are also background to various searches for beyond the SM physics. Their production can be classified into two groups, namely the electroweak (EW) induced channels of order O(α^6), as shown in figure 5.2(a) and the QCD-induced processes of order O($\alpha_s^2\alpha^4$), as shown in figure 5.2(b).

Moreover, the EW contributions are divided in t- and u-channels named “vector boson fusion” (VBF) mechanisms and the s-channel corresponding to the production of three EW gauge bosons, with one off-shell gauge boson decaying into quark-antiquark pair. The advantage of reconstructing WW+WZ in $\ell\nu jj$ decay mode along with two forward jets over the purely leptonic channels is the larger branching fraction of vector boson decay to jets, however, at the expense of larger backgrounds, mainly from W+jets. Compared to the pure WW leptonic diboson decay, the semi-leptonic process has the ability to provide a direct handle on the boson’s transverse momentum (p_T). The sensitivity to very

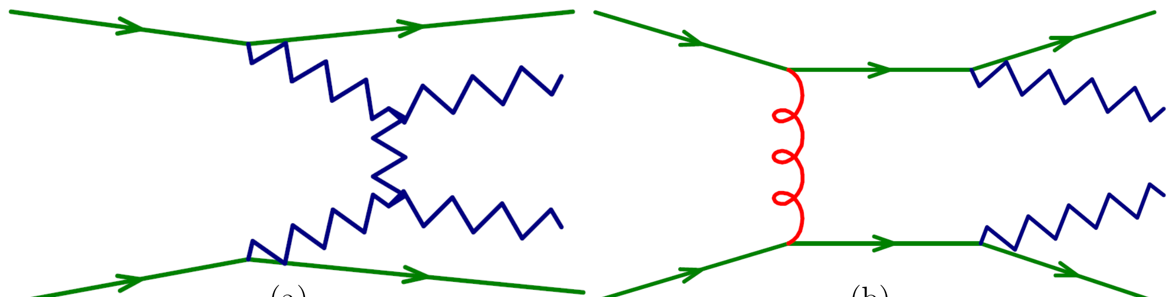


Figure 5.2: Sample diagrams for (a)electroweak and (b)QCD production mechanisms of WW+dijet production.

high boson p_T makes this process particularly useful as a probe of the gauge structure (symmetries) at high energies where we anticipate new physics. Also, it is possible to reconstruct the 4-body invariant mass ($m_{\ell\nu jj}$) of the semi-leptonic decay particles if one solves the quadratic equation to extract p_z of the neutrino. Electron or muon from the leptonic decay of the W boson is used to trigger the event (along with requirement on E_T^{miss}). The shape of dijet mass spectrum is used to discriminate the hadronic $W/Z \rightarrow jj$ decay from the W+jets, top (anti)top quark pairs, single top quark, and QCD multijet processes. On the flip side, this channel is beset by a large background from single W boson produced in association with quarks and/or gluons. However, presence of two forward jets provides an additional handle, which can be used as a “tag" by requiring large dijet invariant mass and large rapidity gap among them. The Higgs mechanism is conceived to give mass to the W and Z bosons and break the electroweak gauge symmetry. The mechanism predicts a Higgs boson particle that interacts with other particles with a strength proportional to their mass. Since the W and Z bosons are very massive, the Higgs will predominantly decay to these particles if its mass is at least twice that of the W or Z boson. The Fenyman diagram for Higgs boson production through vector boson fusion which contributes to WV (V being W/Z) scattering is shown in figure 5.3.

The W and Z bosons masses differ by about 10 GeV. However, the dijet mass resolution of the CMS detector near 80–100 GeV is about 12%, and hence it is insufficient to distinguish between W$\rightarrow jj$ and Z$\rightarrow jj$ decay processes. Therefore, our signal contains a mixture of WW and WZ events along with two forward jets.

5.2 Analysis Strategy

This section briefly provides the steps used to perform analysis in the above final states. Basically, the whole procedure can be summarized in three steps:

- a set of kinematical thresholds are defined to isolate a sample of scattering events candidate.

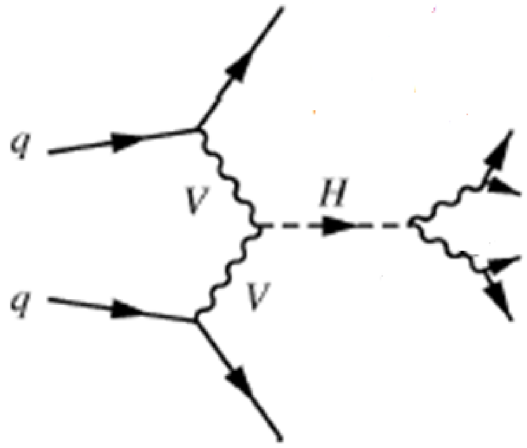


Figure 5.3: The fenyman diagram for higgs production through vector boson fusion which contribute to WV scattering.

- Vector Boson Scattering (VBS) signal is defined on this enriched sample.
- a statistical treatment for signal is developed to estimate the probability of observing an excess in terms of confidence levels.

The first step is to deal with the identification of a suitable kinematical footprint which can allows to capture the essence of WV scattering. The selection of events widely separated in pseudorapidity is a widely known technique for enhancing the scattering contributions at LHC. Looking at the topology of the diagrams included in the gauge boson scattering as a subprocess, one summaries that it is appropriate to associate the two most forward/backward jets to the tag quarks which radiate the bosons and which, in turn, initiate WV scattering; and also to relate the two most central jets to the hadronic decay of a W or a Z boson in the final state. This helps to isolate a sample of genuine WV+2 jets events while suppressing the contribution of irreducible backgrounds such as three boson production or top quark production, either from single top quark or top-(anti)top quarks pair. Having obtained a sample of scattering events, one needs to make a choice of an observable which should be sensitive to the details of the process of interest. An apparent choice could be the invariant mass of the boson pair, but this has huge QCD background, mainly coming from W+4jets. Thus, a possible way out is to look instead at the dijet invariant mass of the central jets in vector boson topology, i.e. with with large invariant mass threshold on forward jets. Once suitable set of kinematical thresholds are decided, signal over background ratio is further improved by restricting W/Z boson invariant mass within a window of 43 GeV to 147 GeV, since the EW process cross section is dominated by the W and Z peaks, while W+4jets contribution is non-resonant in this respect. This makes W+jets distribution uniform and signal can be easily recognized as peaking component. Further, once the non-resonant background has been understood, one is left to deal with resonant component which peaks like signal. Dealing with such background could be extremely challenging. However in this case the background which peaks like signal is $t\bar{t}$ and single top, which, though, have similar topology to signal but they come up with an additional handle. Since b-quark from the top quark decays can be identified, hence events containing b-quarks can be removed from the signal region. Thus, having settled with all the selection thresholds and choice of observable, an unbinned maximum likelihood fitting procedure is performed to extract signal.

5.3 Signal and backgrounds

5.3.1 The signal

Parton level study

The phase space, which this analysis deals with, is completely new and has not been explored earlier. For carrying out analysis in such a phase space, understanding of both signal and backgrounds of interest is essential. Besides, new event generators are coming up to better model these new evolving physics processes. Thus, understanding of various event generators is also mandatory. Of course, there has to be consistency in distributions irrespective of generator used and in case it's not there, then that difference has to be known, understood and quantified. This study is done precisely to achieve that goal using three different event generators. The plot in Figure 5.4 shows the di-jet invariant mass distribution from "tag jets". The event selection are chosen as close as possible to even pre-selections. It is observed that there is good agreement (shape-wise) from all three generators while, event-wise, there are more events from PHANTOM generator, which is expected because it has more Feynman diagrams contributing in it. It is inferred that signal generated by any of the three generators is satisfactory, however, since PHANTOM deals explicitly with interference effect, hence it is an ideal choice to study VBS.

Event Generators details

- PHANTOM: The PHANTOM is a tree level MC event generator for six parton final states at $p-p$, $p-\bar{p}$ and electron-positron collider at O(α_{EM}) and O (α_{EM}, α_S) including possible interferences between the two sets of diagrams. This comprehends all purely EW contributions and contributions with one virtual or two external gluons. It can generate unweighted events for any set of processes and it is interfaced to parton shower and hadronization packages through the latest Les Houches Accord protocol. This generator is specially designed to study physics like boson boson scattering, the SM Higgs boson production in boson through vector boson fusion etc. The cross-section obtained from phantom is 77.6 fb.

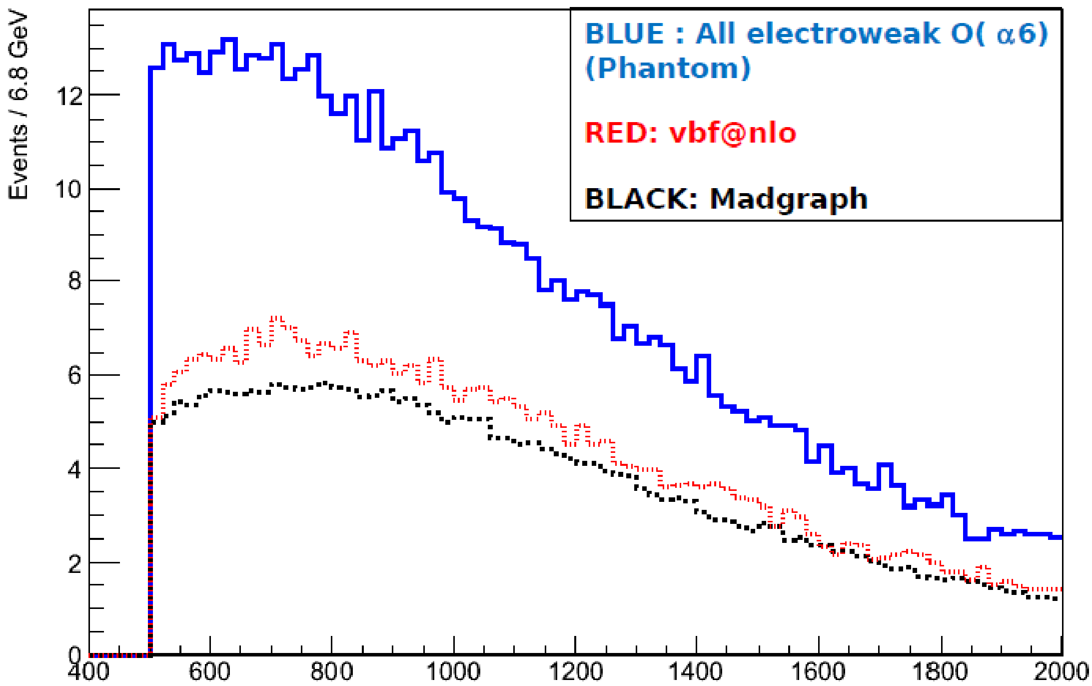


Figure 5.4: Comparison of tag jet invariant mass at generator level from different event generators. The distribution is normalized to cross-section and scaled to luminosity $19.3fb^{-1}$.

- Madgraph: MadGraph/MadEvent (MG/ME) is a software that allows to generate amplitudes and events for any process (with up to 9 external particles) in any model. MG/ME is a Matrix Element Generator, i.e. it generates the code for generating a certain process. It aims at providing all the elements necessary for SM and BSM phenomenology, such as the computations of cross sections, the generation of hard events and their matching with event generators, and the use of a variety of tools relevant to event manipulation and analysis. Processes can be simulated to LO accuracy for any user-defined Lagrangian, and the NLO accuracy in the case of QCD corrections to SM processes. Matrix elements at the tree- and one-loop-level can also be obtained. The cross-section obtained is 110.05 fb.

- VBF@NLO: VBF@NLO is a flexible parton level MC program for the simulation of VBF, QCD induced single and double vector boson production plus two jets, and double and triple vector boson production (plus jet) in hadronic collisions at NLO in the strong coupling constant, as well as Higgs boson plus two jet production via gluon fusion at the one-loop level. For the latest version, i.e. version 2.7.0, several major enhancements have been included in it. Several new production processes have been added. The implementation of anomalous gauge boson couplings has been extended to all triboson and VBF $VVjj$ processes, with an enlarged set of operators yielding anomalous couplings. Finally, semileptonic decay modes of the vector bosons are now available for many processes, including $VVjj$ in VBF, VVV

and VV production. The cross-section obtained is 39.394fb.

It may be noted that the difference in cross-section is due to the fact that generator level thresholds are different in each case, although attempts were made to make the thresholds as similar as possible. However, there are limitations from generators related with gauge invariance requirement within their framework. Tag jet invariant mass distribution is shown in figure 5.4. The Higgs contribution from 125 GeV is included in all three of them.

5.3.2 The backgrounds

All the processes with the presence of a lepton, two jets and missing energy along with two energetic forward jets have to be considered as possible sources of background. The most relevant ones are:

- W+jets: this is the production of single W vector bosons in association with quarks or gluons that mimic the final state signature. Because of its cross-section, this is by far the most important background to the analysis. The Feynman diagram at LO is shown in figure 5.5.

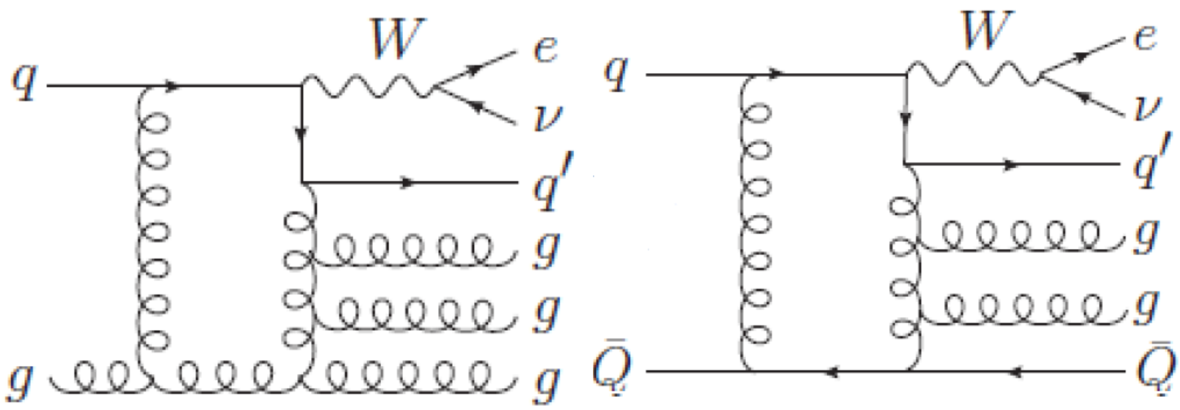


Figure 5.5: The Feynman diagram for W+4jets at leading order.

- $t\bar{t}$: Top quark pairs are produced at LHC via the gluon fusion process $gg \to t\bar{t}$ or via QCD quark annihilation $qq \to t\bar{t}$. The semi-leptonic sample, where one W decays hadronically, can be reduced by requests on the number of jets, while the fully leptonic one is reduced by requiring only one good lepton in the event. In any case, because of acceptance and inefficiencies, this background still contaminates the signal phase space. The Feynman diagram for t$\bar{t}$ production at leading order is shown in Figure 5.6.

- Single top quark production: it proceeds through three separate channels:

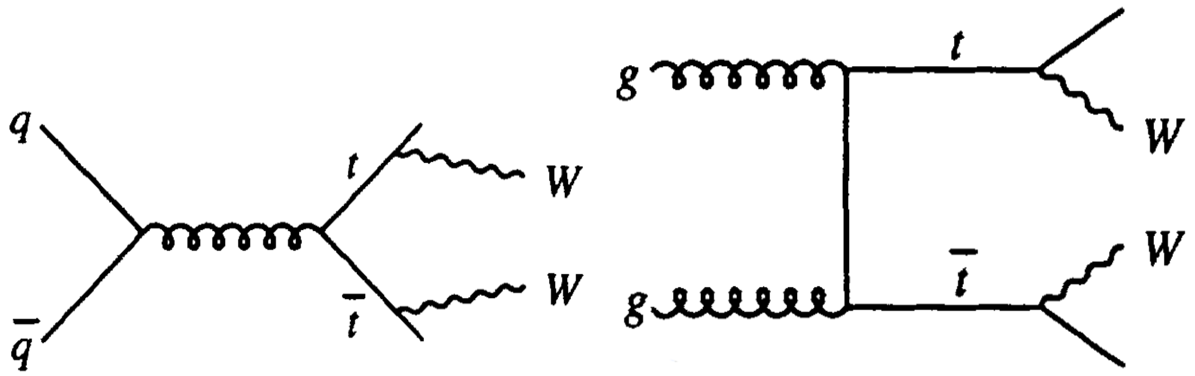


Figure 5.6: The Feynman diagram for $t\bar{t}$ production at LO.

1. t-channel: top quark is produced after a quark-gluon interaction with the exchange of a virtual W boson.
2. s-channel: top quark is produced in association with an anti-bottom quark, after the annihilation of a pair of quarks in a weak vertex.
3. tW-channel: top quark is produced in association with a charged vector boson in a weak process, from a gluon-bottom quark pair in the initial state.

This is second most significant background as its topology mimic the signal topology except that most of the jets are central.

- Drell-Yan (Z/γ^*+jets): this is the production of single Z/γ^* bosons in association with quarks or gluons, where one lepton gets undetected because of acceptance or inefficiency effects, and the hadronic activity mimics the final state signature of the signal.
- QCD multi-jet events generate a background because of the non-negligible probability of jets to be reconstructed as leptons. However, their contribution are insignificant and hence are neglected in the analysis.

The cross section for the backgrounds, multiplied by the branching ratio are reported in Table 5.1. Each different sample notation includes all the decays taken into account in the effective cross section calculation.

5.4 Datasets

5.4.1 Collision Data

The data sample used in this analysis are same as in the first analysis mentioned in chapter 4. The total statistics analyzed correspond to an integrated luminosity of $19.3 fb^{-1}$. Data

Channel	Cross-section (pb)
W+4 jets	214.0
W+3 jets	519.0
W+2 jets	1750.0
W+1 jets	5400.0
Z+ jets	3503
$t\bar{t}$+jets	225.2
t/$\bar{t}$+jets (t-channel)	85.5
t/$\bar{t}$+jets (s-channel)	5.65
t/$\bar{t}$+jets (tW-channel)	22.4

Table 5.1: The cross section for the backgrounds, multiplied by the branching.Each different sample notation includes all the decays taken into account in the calculation.

was collected using high-level single lepton triggers: a single muon trigger with a p_T threshold of 24 GeV and a single electron trigger with a p_T threshold of 27 GeV.

5.4.2 MC Simulated Data

The SM WW+2jets signal sample, as well as samples for a large variety of EW and QCD-induced background sources, have been generated and showered using different MC generators. To better reproduce the actual data-taking conditions, where there is a significant probability that more than two protons interact in the same bunch crossing, pile-up (PU) events are added on top of the hard scattering. Particle interactions with the detector were reproduced through a detailed description of CMS. Madgraph5, VBF@NLO and PHANTOM generators been used to produce signal events, and the showering has been performed with PYTHIA6 [83]. These generators are explained in section 5.3.1.

5.5 Event Selection

The final state of the WW+2jets is characterized by two forward jets a charged lepton, large missing energy and two hadronic jets that form a W boson. The analysis relies on the standard reconstruction algorithms as mentioned in chapter 3. Pre-selection is similar to Higgs analysis which is discussed in detail in section 4.4. However there is some changes in jet selection technique, this is due to the fact that in this analysis an additional pair of jets are selected which are "tag jets".

5.5.1 Special Jet Selection criterion

The most obvious choice of jets could have been choosing four leading jets however a different approach has been used for this analysis. Since this analysis is statistically limited so we make all possible combinatoric from all possible jets to have highest signal matching efficiency to generator level jets. It turns out that if we start with more than four jets then matching efficiency increases by 10%.

Tag jets threshold

Tag jets invariant mass is required to be $>$ 500 GeV, η difference between tag jets are required to be greater than 3.5 and these jets are further required to be coming from opposite hemispheres.

We explored a whole lot of methods, which were motivated by physics goals. Finally the method by which we found most effective is explained below:

Final jets selection

Firstly, We make all possible combinatoric of pairs from all jets present within $|\eta|< 4.7$, in the event. From these pair we select the one which satisfy "tag jet threshold" and if in the event there are more than one such pair then we select the pair of jets as our final tag jets which have highest invariant mass in the event.

Once we have selected the two tag jets in the events. From the remaining jets in the event we choose jets with highest and next highest p_T as the jets coming from W boson decay.

5.5.2 Top quark control region selection

The idea is to choose top quark control region which exclude signal region but still close to the signal phase space. To choose this region we start by selecting two non-b-tagged jets and two b-tagged jets from central η region.

For b-tagged jets pair:

- it is required that their invariant mass is greater than 300 GeV.
- They are required from opposite hemispheres (same selection as in signal).
- η separation between the two jets is required to be less than 3.5 (opposite to signal selection).
- $\mu(e)$: $\eta < |2.1|(|2.4|)$ (excluding transition region for electron), $p_T > 25(30)$ GeV, E_T^{miss} >25(30) GeV (same as in signal region).
- dijet invariant mass: 43 GeV $< M_{jj} <$ 127 GeV (same as in signal region).

5.6 Data - MC comparisons

To assess the quality of the modeling provided by the MC simulation we make comparisons between the MC shape, normalized to the number of collected events in data, for the background compared to the overall data yield after applying the event selection criteria. We show the distributions of various kinematic variables after applying all the selection thresholds in figures 5.7-5.14 for the μ+jets category sample. For electrons it looks similar and hence are not included. The color code used are same for all the distributions, and as it was mentioned W+jets (red in color) and top quark (sum of single and pair produced top quarks) are dominant backgrounds. The black dotted line and solid blue line drawn on top of stack is signal generated by PHANTOM and Madgraph5 generators respectively, both are scaled by 100 when being drawn. As visible all distribution looks satisfactory within statistical error (as shown in black hatched box below) . The MC has been been corrected for lepton reconstruction efficiency and the trigger efficiency and also reweighed for pile up correction. Seeing reasonable agreement gives us confidence in the qualitative aspects of the MC modeling.

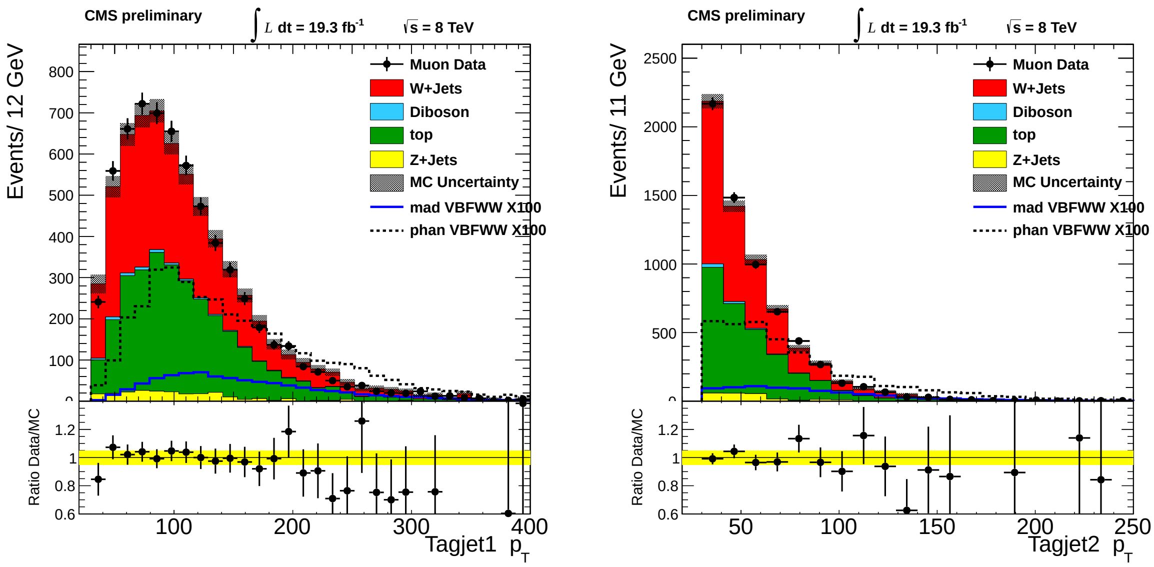


Figure 5.7: Comparison of the leading tag jet (left) and the second tag jet (right) p_T distributions from data and MC for the μ+jets category.

The data MC comparison for the various angular variables are also shown in figures 5.15-5.18 for μ+jets category sample.

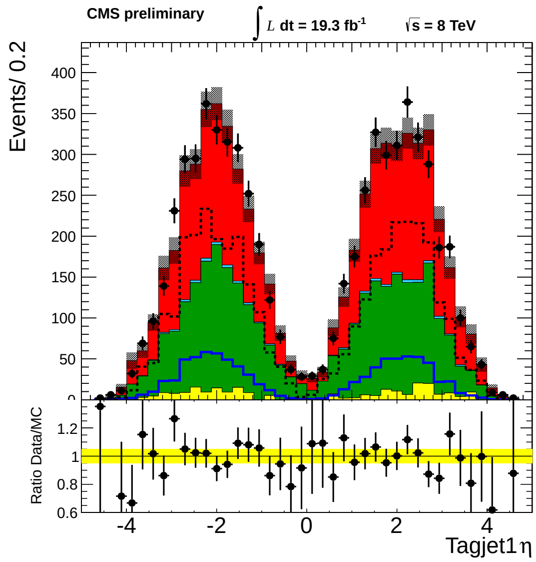


Figure 5.8: Comparison of the leading tag jet η (left) and the second leading tag jet η (right) distributions from data and MC for the μ+jets category.

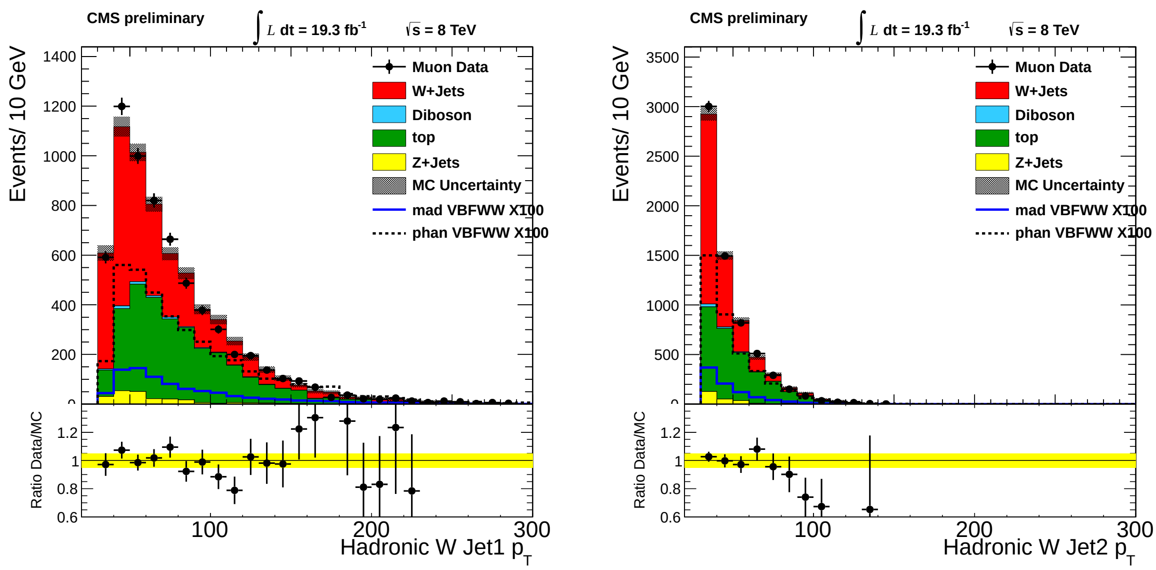


Figure 5.9: Comparison of the leading W jet (left) and the second W jet (right) p_T distributions from data and MC for the μ+jets category.

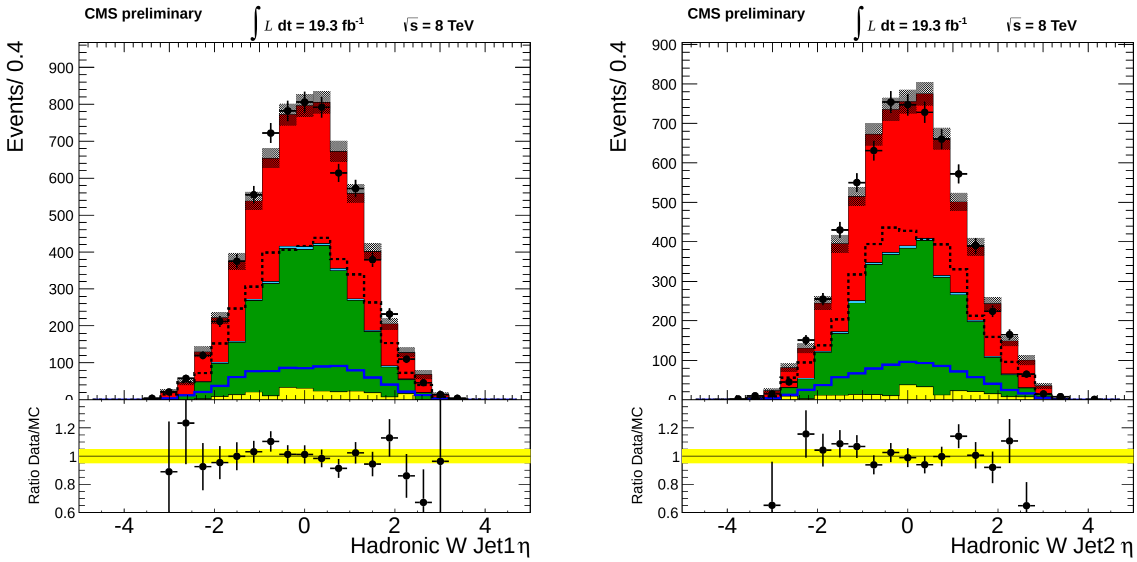


Figure 5.10: Comparison of the leading W jet η (left) and the second leading W jet η (right) distributions from data and MC for the μ+jets category.

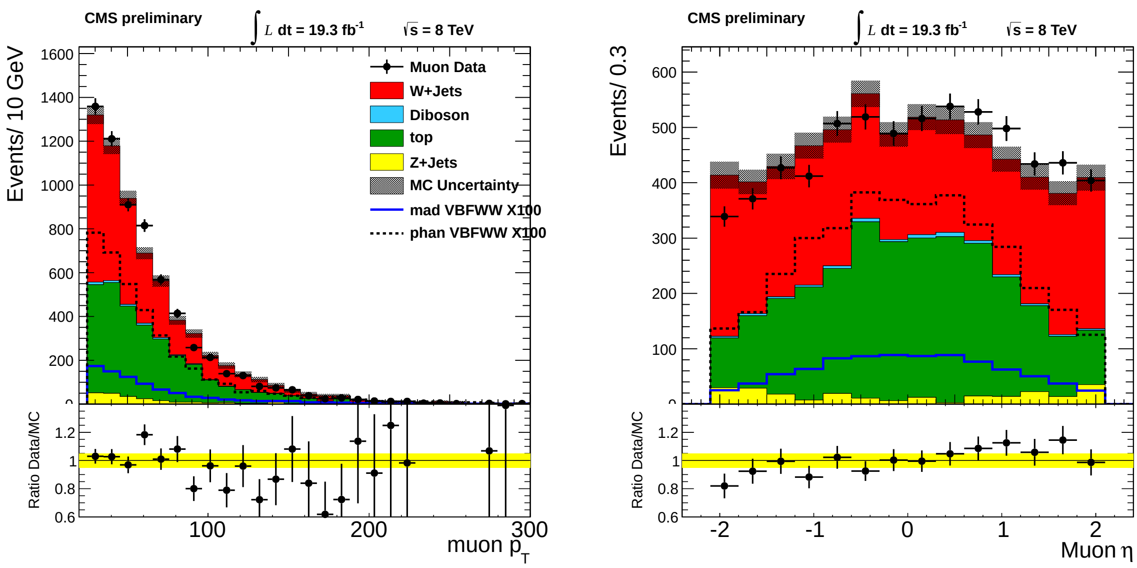


Figure 5.11: Comparison of the muon p_T (left) and the muon η (right) distributions from data and MC for the μ+jets category.

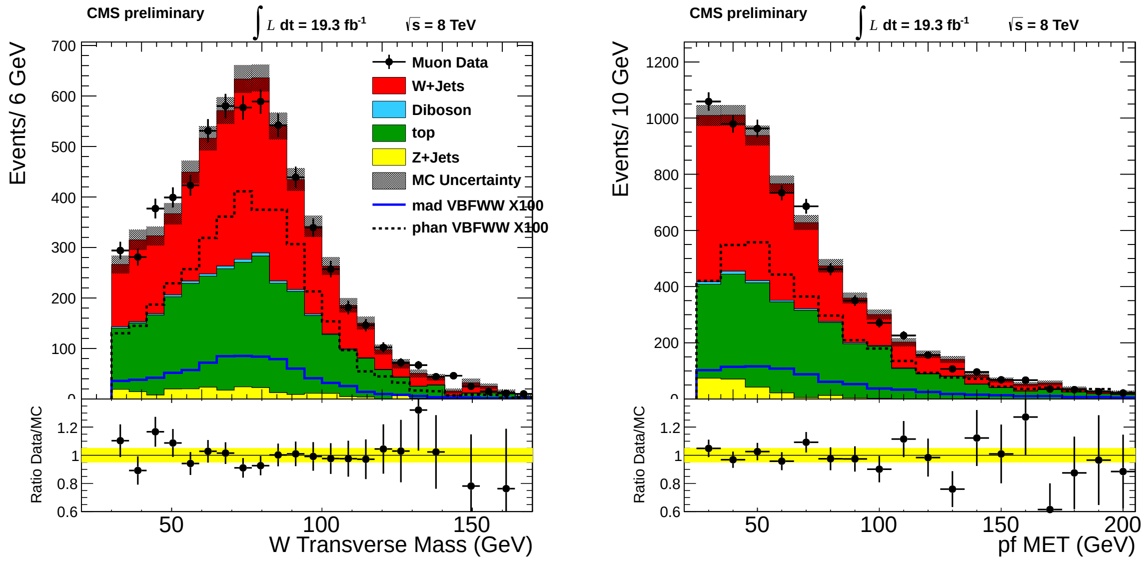


Figure 5.12: Comparison of the distributions from data and MC of the transverse mass of the muon / E_T^{miss} system (left) and the E_T^{miss} (right) for the μ+jets category.

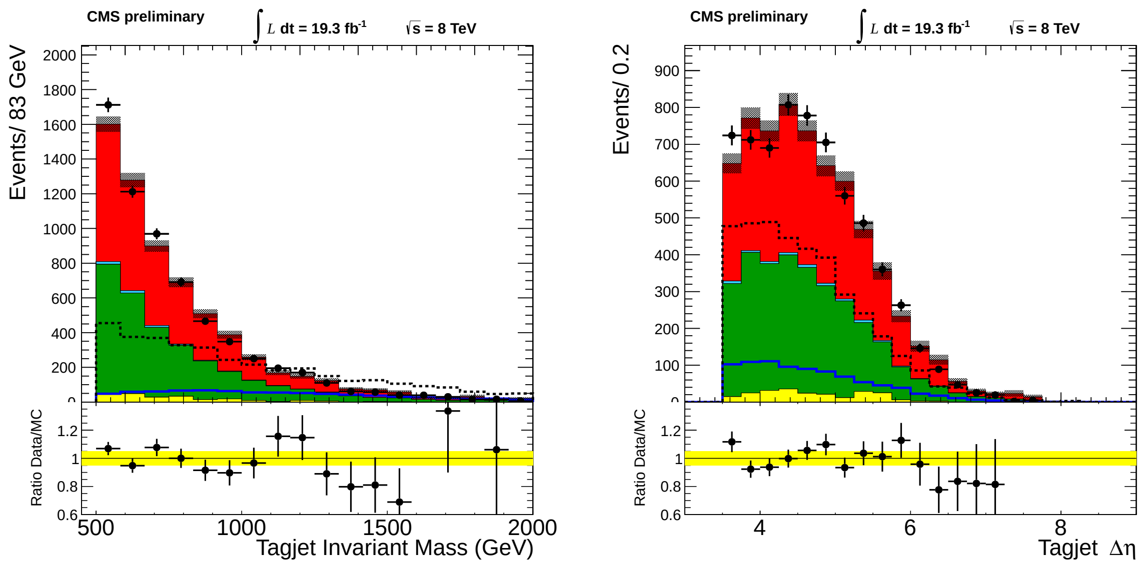


Figure 5.13: Comparison of the distributions from data and MC tag dijet system mass (left) and η diffrence (right) for the μ+jets category.

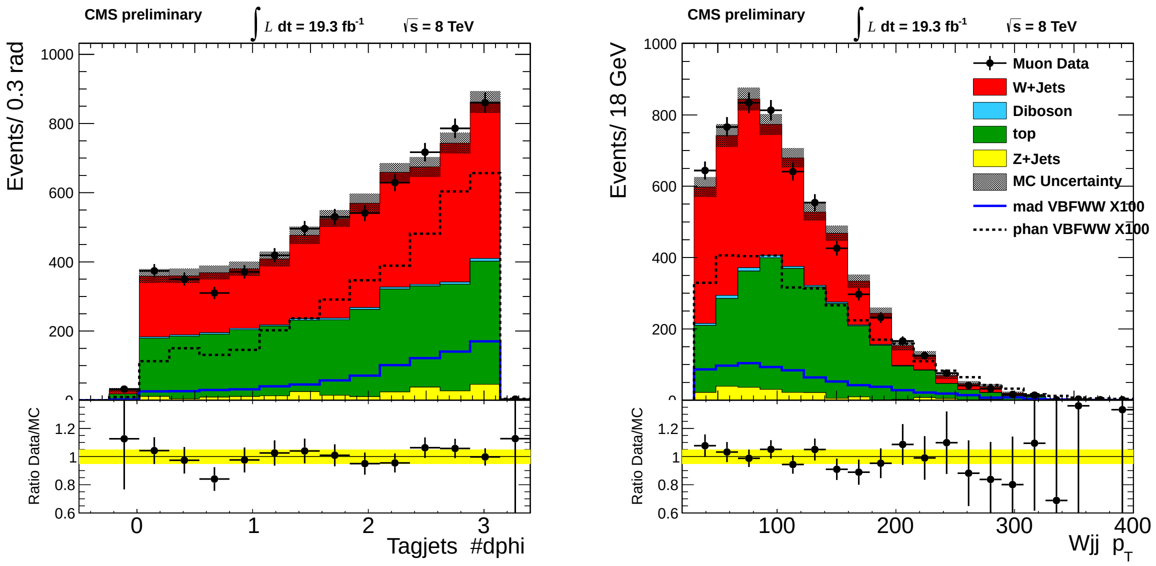


Figure 5.14: Comparison of the distributions from data and MC tag dijet system ϕ difference (left) and dijet p_T for hadronically decaying W boson on the right for the μ+jets category.

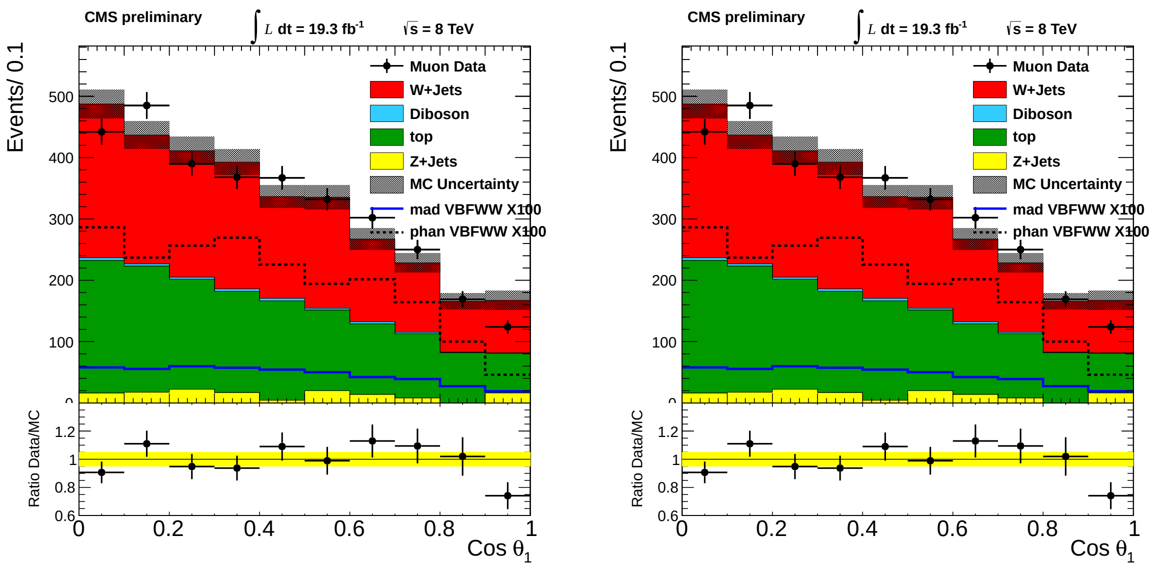


Figure 5.15: Comparison of the angular distributions for $\cos\theta_1$ (left) and $\cos\theta_2$ (right) from data and MC for the μ+jets category.

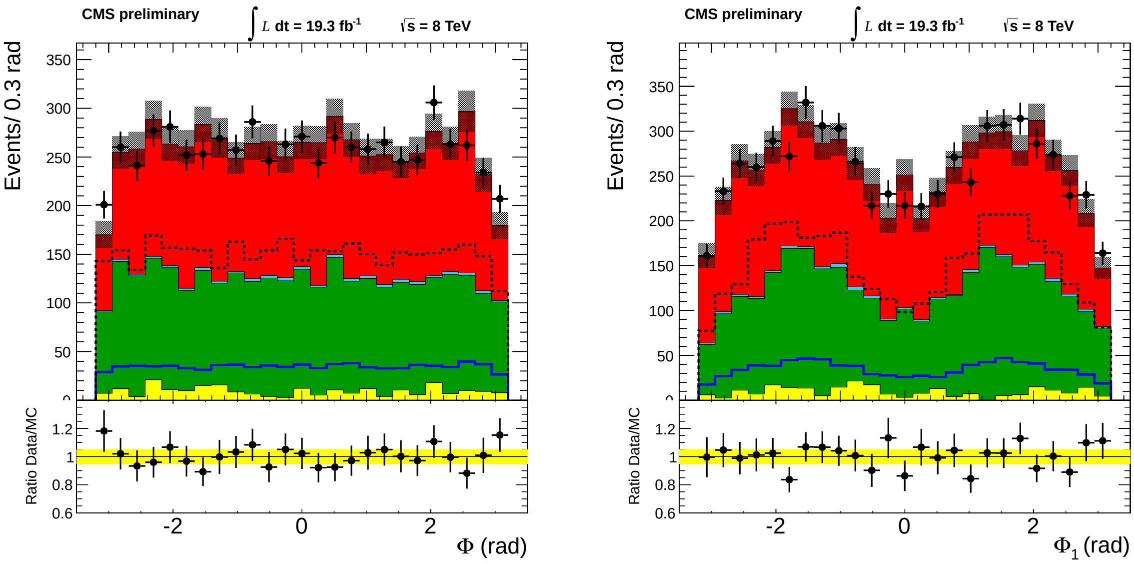


Figure 5.16: Comparison of the angular distributions for Φ (left) and Φ_1 (right) from data and MC for the μ+jets category.

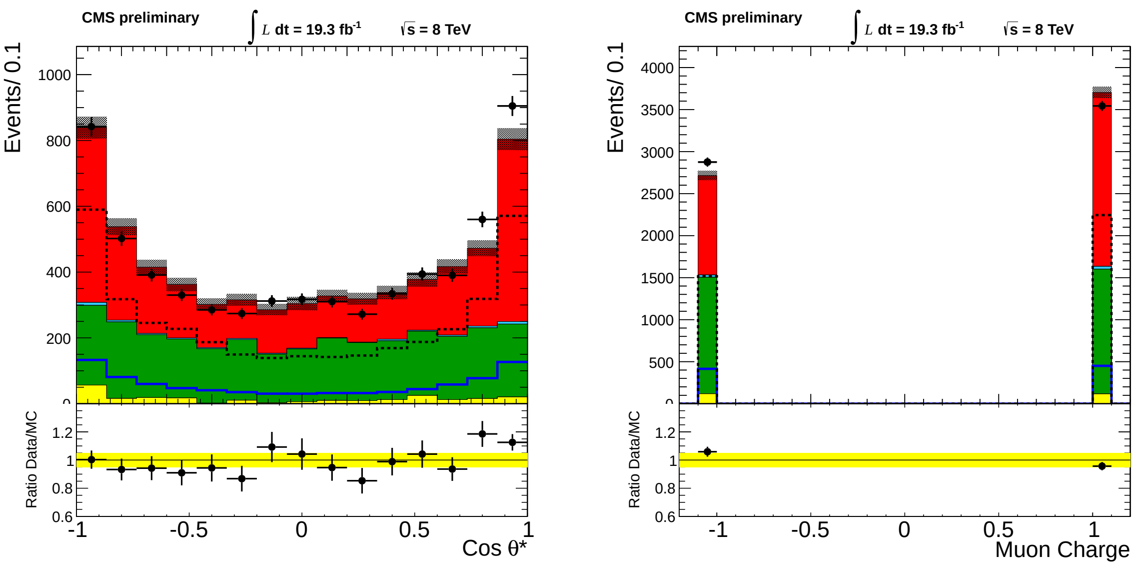


Figure 5.17: Comparison of the angular distributions for $\cos\theta^*$ (on left) and charge of the muon from data and MC for the μ+jets category.

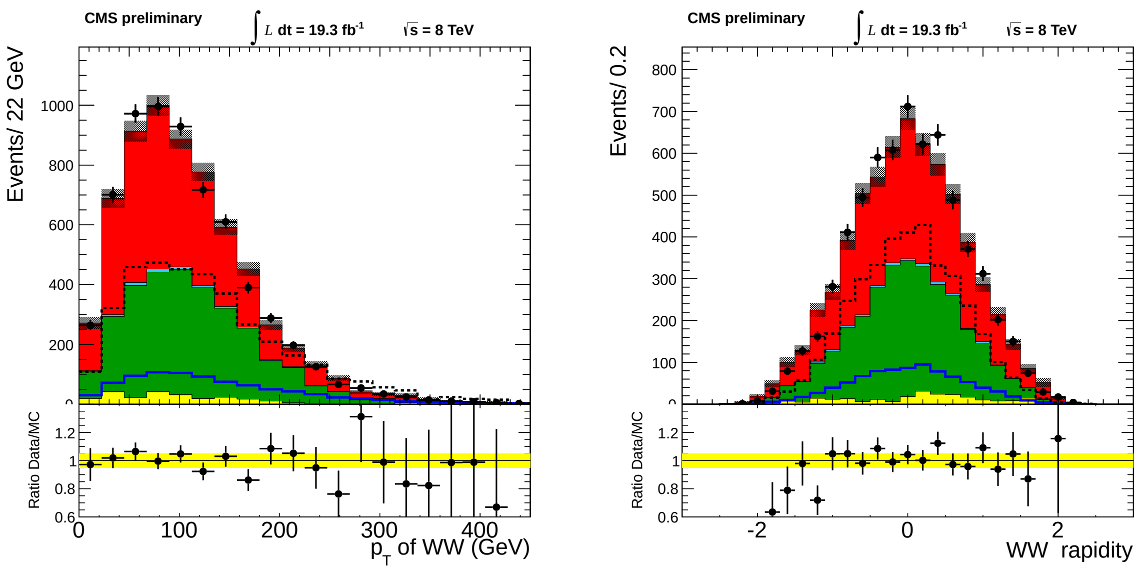


Figure 5.18: Comparison of the p_T (left) and rapidity (right) of the WW system from data and MC for the μ+jets category.

5.7 Top quark background normalization by Data-Driven Method

At the LHC the top quark pair production rate is fairly large, almost 1000 times larger than the signal production rate. According to the SM, the top quark decays into W boson and b quark with branching fraction of about 100%. The selection to define top control region is discussed in section 5.5.2. When we select events in which one W boson decays leptonically ($W \to e\nu, \mu\nu$) and the other W boson decays into quark pairs thus leading to a semileptonic final state, then the signal purity is high.

As, the the final selection for this control region consists of a high energy lepton, large missing energy E_T, and four jets of which two are b-tagged. The hadronic W candidates are formed from two anti-b-tagged jets. In the top quark control region, its normalization is estimated by maximum likelihood fit to the data in the hadronic W invariant mass distribution.

The steps involved in this fit are described below:

- Template fit to data: All background components are added forming a single probability distribution function and then it is fitted to data by keeping top quark normalization a free parameter. The other MC contribution which is W+jets is small in this region is fixed to MC expectation in the fit.
- Convoluted fit to data: Taking parameters from template fit as input, and further using gaussian function W+jets and top quark probability distribution templates are convoluted. This is to include resolution effect.
- Find top quark normalization: From the fit in which top quark background normalization was floating, we get its value at the end of the fit. This number is then propagated to next step, which is fit to the data in the signal region to extract signal fraction.

The fit result, the invariant mass of the hadronic W candidates in the μ+jets category is shown in figure 5.19 and for e+jets category it is shown in figure 5.21 for the template fit. When MC templates are convolved with a Gaussian resolution function and fit is repeated, the result is shown in figure 5.20 for μ+jets category, while for e+jets category it is shown in figure 5.22. The W mass resolution is dominated by the resolution of the jet energy measurement, with jet energy scale (JES) systematics needs to be accounted for. From the comparison between data and MC, a normalization correction factor for $t\bar{t}$ and single top quark is evaluated is tabulated in Table 5.2.

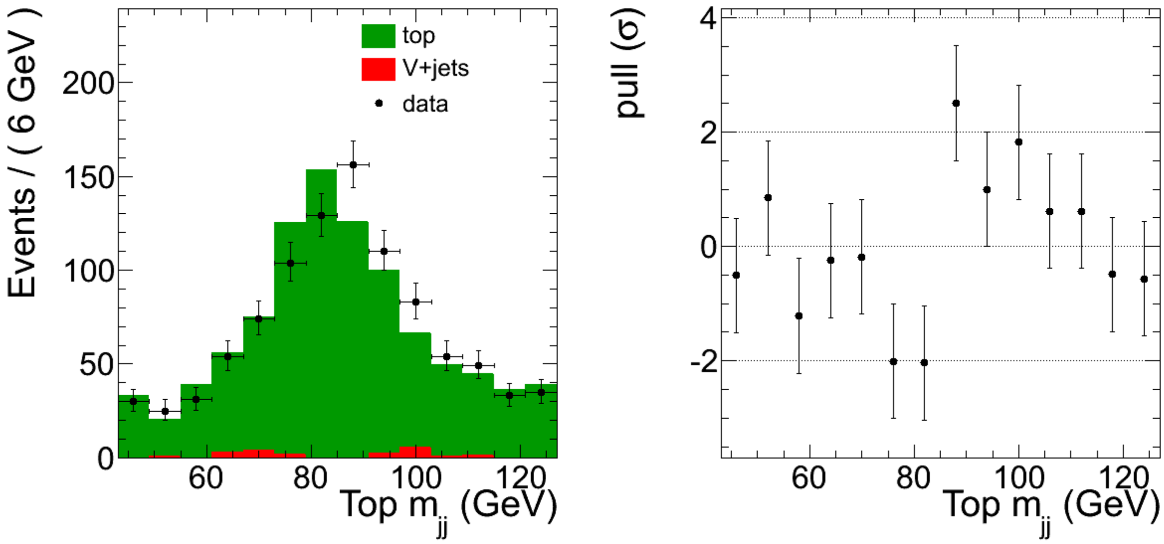


Figure 5.19: Template fit to the invariant mass distribution of the hadronic W candidates in the muon semileptonic top sample: stacked backgrounds (left) and normalized residual between data and MC (right).

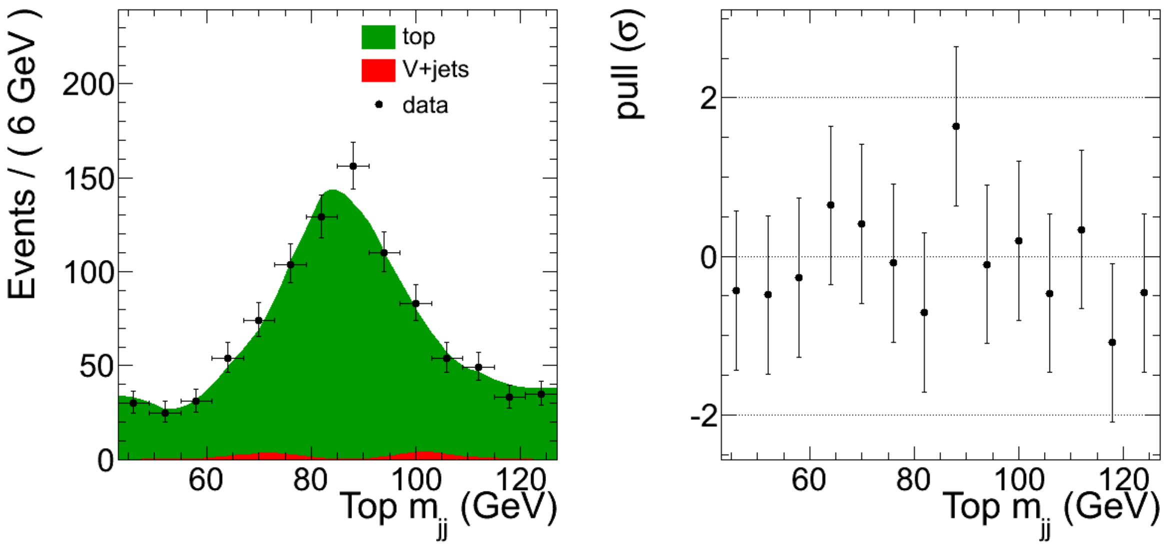


Figure 5.20: Convolved template fit to the invariant mass distribution of the hadronic W candidates in the muon semileptonic top sample: stacked backgrounds (left) and normalized residual between data and MC (right).

Channel	top normalization
μ+jets	1.12 ±0.37
e+jets	1.08 ±0.04

Table 5.2: top quark normalization obtained from top quark control region which gets applied in signal region.

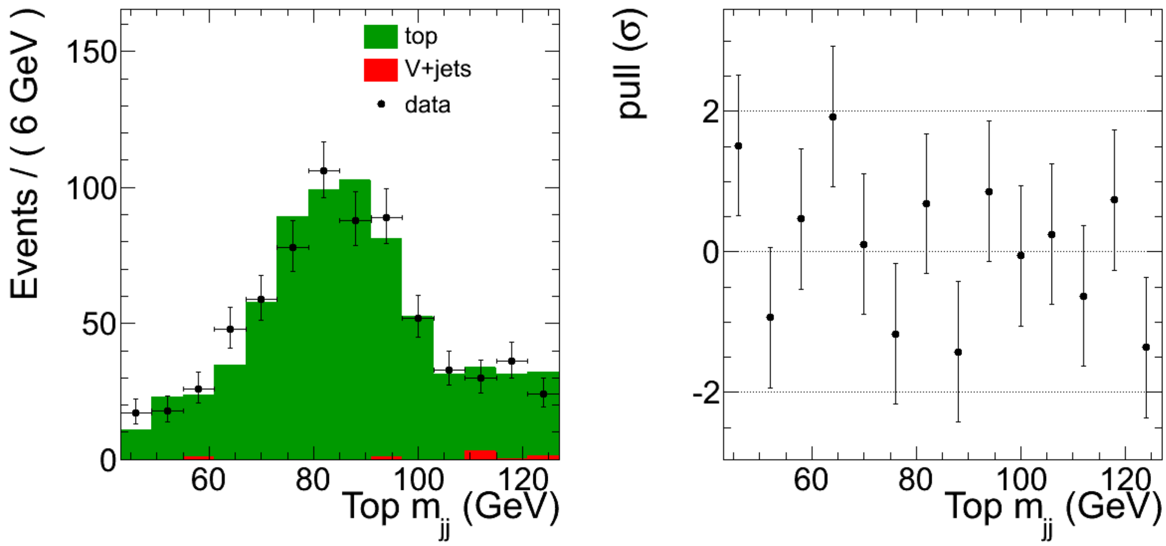


Figure 5.21: Template fit to the invariant mass distribution of the hadronic W candidates in the electron semileptonic top quark sample: stacked backgrounds (left) and normalized residual between data and MC (right).

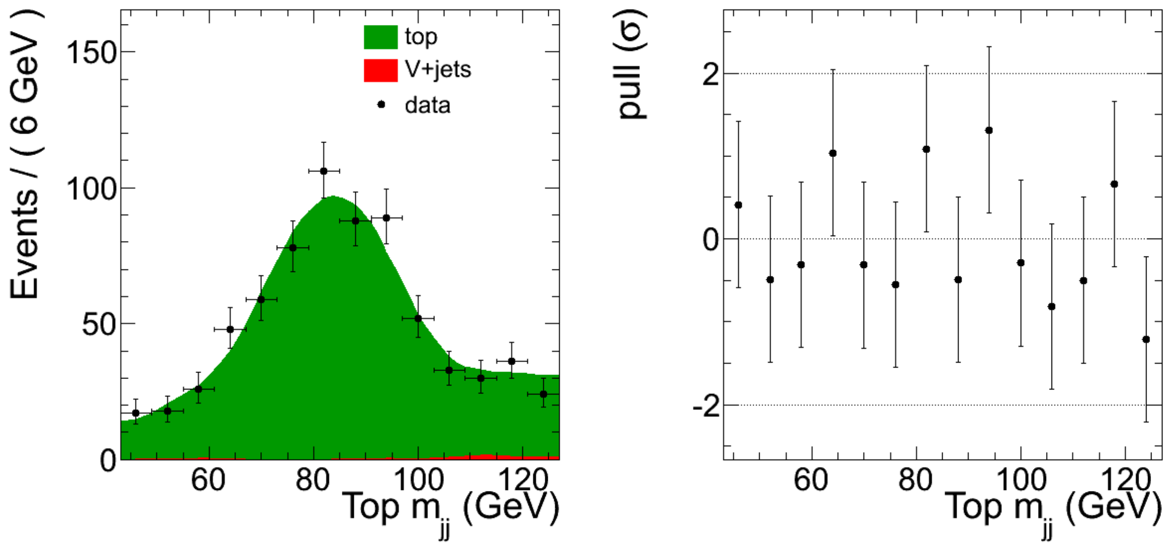


Figure 5.22: Convolved template fit to the invariant mass distribution of the hadronic W candidates in the electron semileptonic top quark sample: stacked backgrounds (left) and normalized residual between data and MC (right).

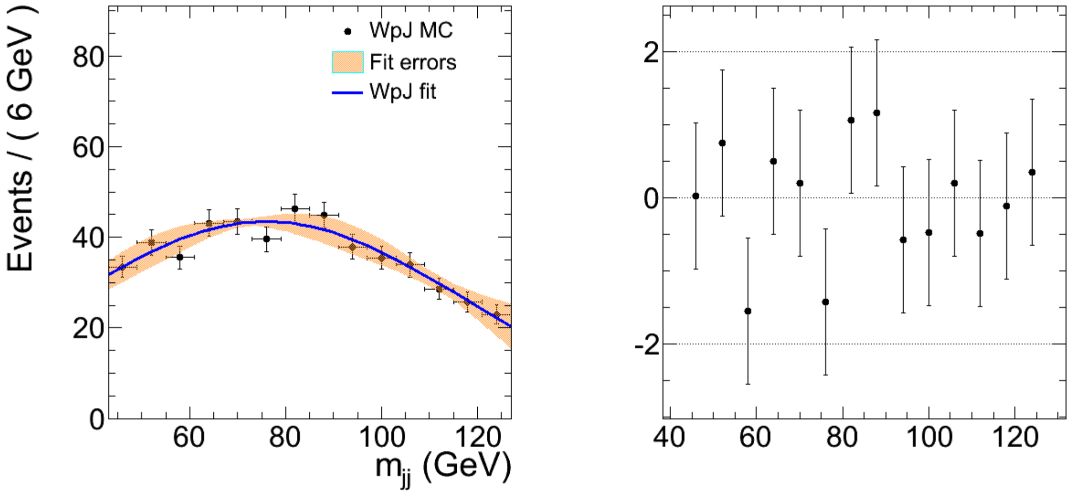


Figure 5.23: V+jets m_{jj} shape. Projection of a fit to the V+jets MC (left) and normalized residue (right) distribution for μ+jets category.

5.8 Modeling Individual MCs

5.8.1 V+jets Background Shape

Because of inadequate statistics in the V+jets (V is W/Z) MC and the overall poor agreement between V+jets MC and many data distibutions, we employ an empirical description of the V+jets shape, which consists of W+jets and Drell Yan events combined based on their predicted cross-sections.

The shape is modeled by the error function times an exponential falloff for both μ+jets and e+jets categories:

$$\mathcal{F}_{V+\text{Jet}} = \text{erf}(m_J; m_0, \sigma) \times e^{-\alpha m_J}, \tag{5.1}$$

All the parameters are fixed to MC. In figure 5.23 and figure 5.24 are shown the result of an unbinned maximum likelihood fit to the MC distributions for μ+jets and e+jets category. It may be seen, there is a good agreement between projection and the simulation.

5.8.2 VBF WW Signal Shape

Likewise, we use a parametric shape to model the signal (VBF WW) MC. The sample would contain a combination of dijet resonances at the W and Z boson masses. It is parametrized by a sum of two gaussians to catch up the W and Z mass resonance separated by a fixed offset (10.8 GeV), for both μ+jets and e+jets categories:

$$\mathcal{F}_{VBFWW} = f_W \times \text{Gaus}(m_{jj}; m_W, \sigma) + f_Z \times \text{Gaus}(m_{jj}; m_W + 10.8, \sigma), \tag{5.2}$$

where f_W, f_Z, m_W and σ are fixed based on the MC. The results of the parametric shape fit to the simulation are displayed in figure 5.25 for μ+jets category and figure 5.26 for

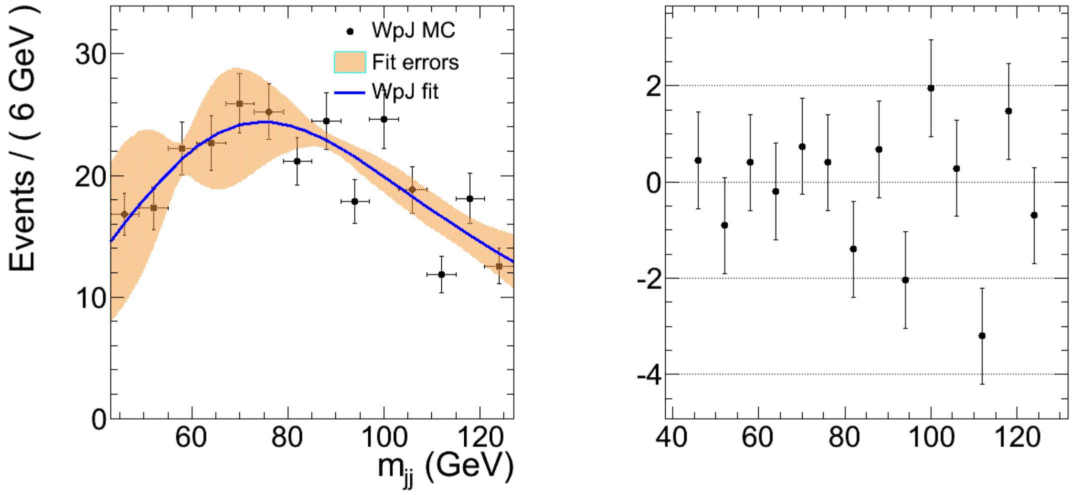


Figure 5.24: V+jets m_{jj} shape. Projection of a fit to the V+jets MC (left) and normalized residue (right) distribution for e+jets category.

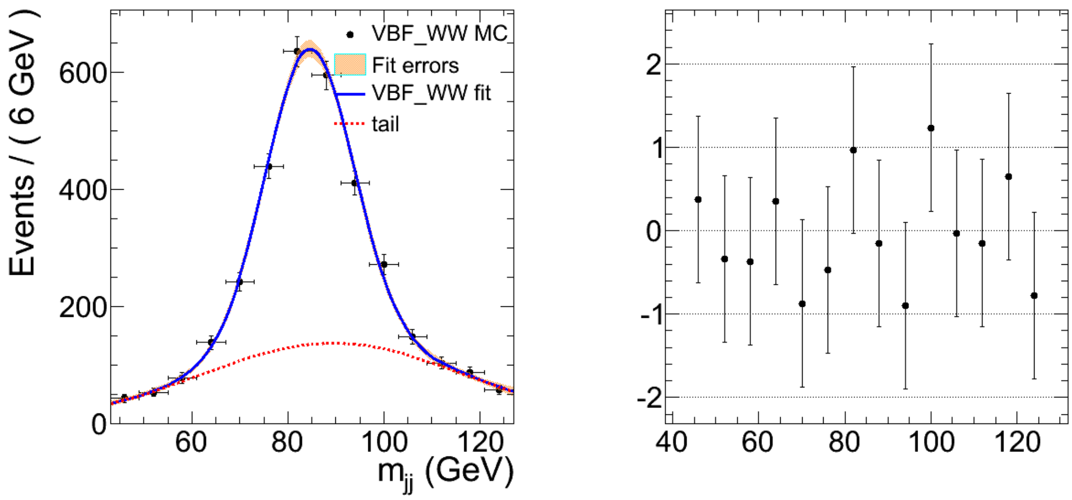


Figure 5.25: VBF WW m_{jj} shape, projection of a fit to the VBF WW MC for (on left) and normalized residue distribution (on the right) μ+jets category.

e+jets category. It may be seen that between parametrized shape and MC there is good level of agreement in both categories.

5.8.3 Top Background Shape

The top quark background is a combination of $t\bar{t}$ and single top quark processes with the MC samples combined based on the expected cross sections.The shape is modeled by a combination of a core gaussian plus a 3-order polynomial.

$$\mathcal{F}_{t\bar{t}} = f_W \text{Gaus}(m_{jj}; m_W, \sigma_W) + a_p \text{Poly}(b, c, d) \tag{5.3}$$

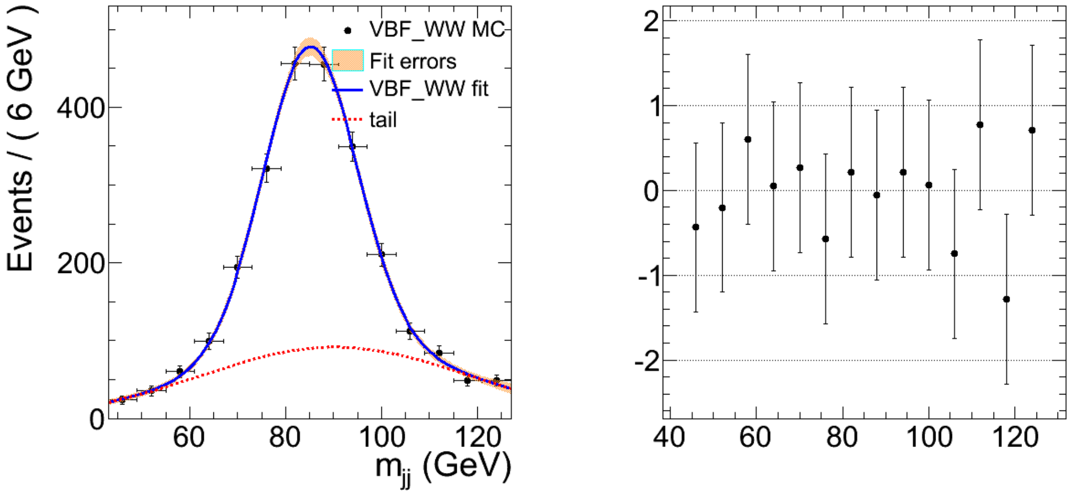


Figure 5.26: VBF WW m_{jj} shape, Projection of a fit to the VBF WW MC for (left) and normalized residue distribution (right) e+jets category.

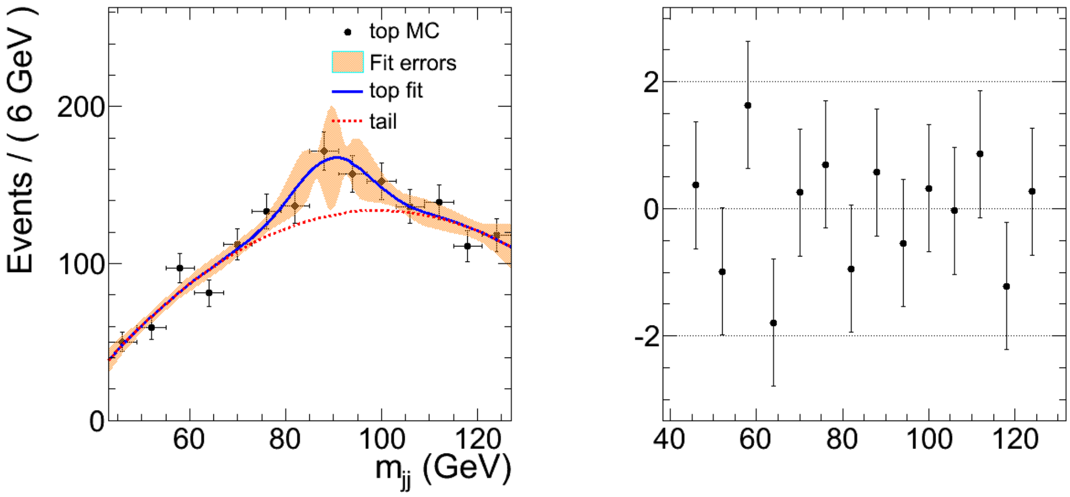


Figure 5.27: Top m_{jj} shape. Projection of a fit to the top quark MC (left) and normalized residue distribution (right) for μ+jets category.

All the parameters for gaussian and polynomial functions are fixed from MC. The results of the parametric shape fit to the simulation are displayed in figure 5.27 for mu+jets category and figure 5.28 for e+jets category. It can be seen that there is good agreement between parametric shape and MC in both categories.

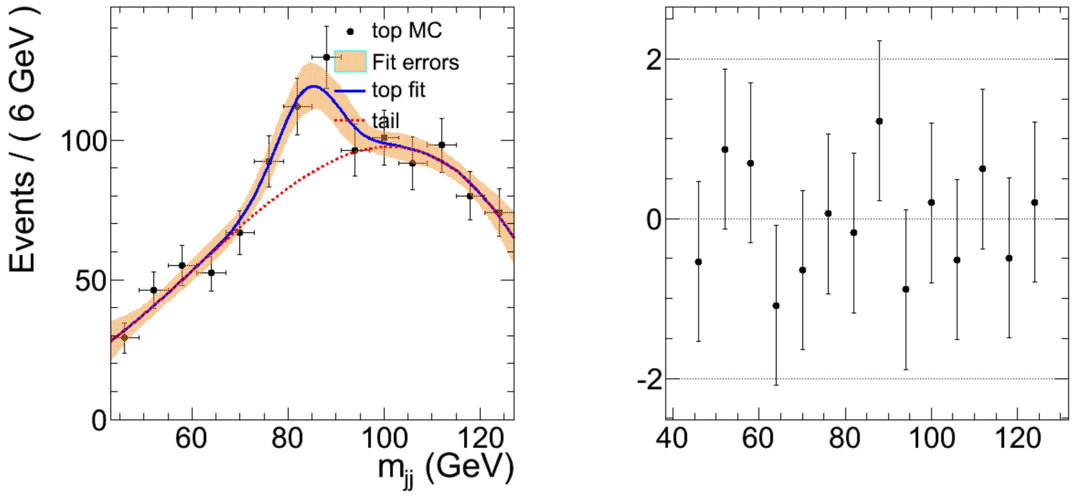


Figure 5.28: Top m_{jj} shape. Projection of a fit to the top quark MC (left) and normalized residue distribution (right) e+jets category.

5.9 Determination of signal yield from a fit to the m_{jj} distribution

The vector boson fusion signal yield is obtained from an unbinned maximum likelihood fit to the dijet invariant mass distribution in the data. The fit output is summarized in Table 5.3. The main sources of systematics error are the uncertainties in the modeling/determining parameters of individual probability distribution function as well as the jet energy scale (JES) uncertainty.

5.9.1 Fit results

The results of the fit are shown in figures 5.29, 5.30 for μ+jets category and e+jets category respectively. In this fit, top quark background normalization is obtained from top quark control region discussed in section 5.7. V+jets normalization is loosely constrained while VBF WW (signal) normalization is a free parameter. A peak from the Standard Model VBF WW production can be seen although it's very small. From the fit result we extract normalization on signal. The error on the normalization of VBF WW essentially gives us an idea about the sensitivity of this analysis. As one can see that in case of μ+jets category the error on extracted fraction is about 4.14 while extracted fraction is 2.35 and in case of μ+jets category it is 3.20 and 0.71 respectively. This shows that we are limited by statistics in this analysis. Since the statistical error dominates the result, it is understood that to extracted signal yield from this final state we need more statistics.

Consistency checks has been carried out to validate the fit procedure and take systematic errors appropriately in account. Which, give us confidence in the analysis procedures. The fit results are tabulated below.

Bin	muons		electrons	
	Predicted	Extracted Fraction	Predicted	Extracted Fraction
VBF WW	22	2.35±4.14	17	0.71±3.20
Top	1164	1.00±0.00	765	1.00±0.0
V+Jets	2556	0.99±0.04	3187	1.00±0.003
Total Yields	3773	3748	3965	3968
Data	3767	—	3966	—

Table 5.3: Fractions of the expected yields (or the total yield) determined from a likelihood fit to the data, the expected event counts and the VBF WW fraction.

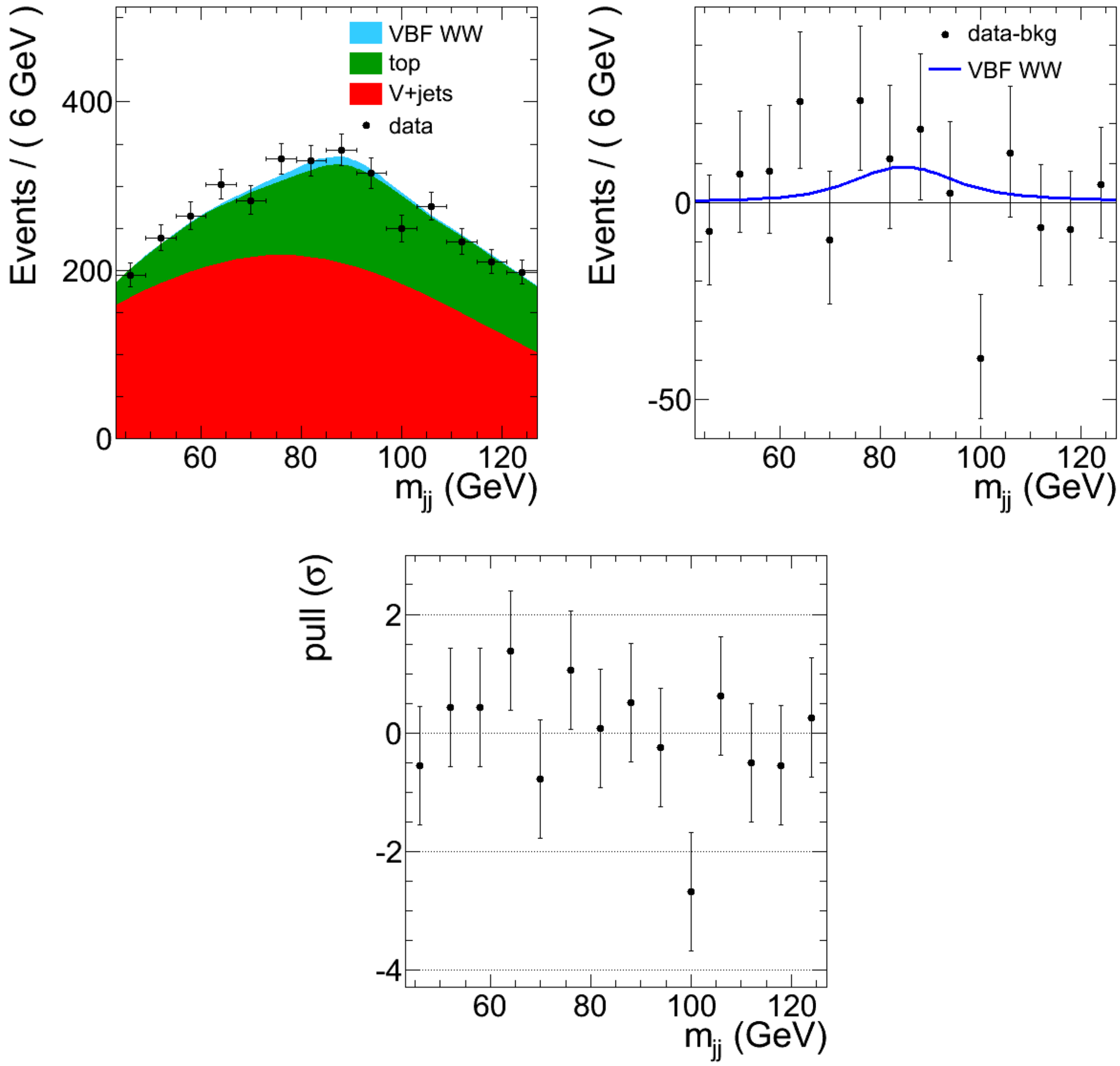


Figure 5.29: Distribution of the dijet invariant mass in muon data and corresponding fit shapes: (upper left) All background components stacked together, (upper right) Data minus all backgrounds except VBF diboson, lower row normalized residual between data and MC.

5.9.2 Fit Validation

The first step is to determine if there is any inherent bias in the fit. As such we use the yield and shape parameter values returned by the fit to MC for each processes to generate toy datasets and refit them. The principle which is employed is same as the Higgs analysis discussed in chapter 4. However we reasonably trust our MC shape which is why they are fixed in the fit to the data. A representative fit validation result is shown in figure 5.31. The normalized distribution on the right looks like Gaussian with mean zero and $\sigma = 1$. The other plot on the right is bias distribution which is fluctuating about zero as it should be.

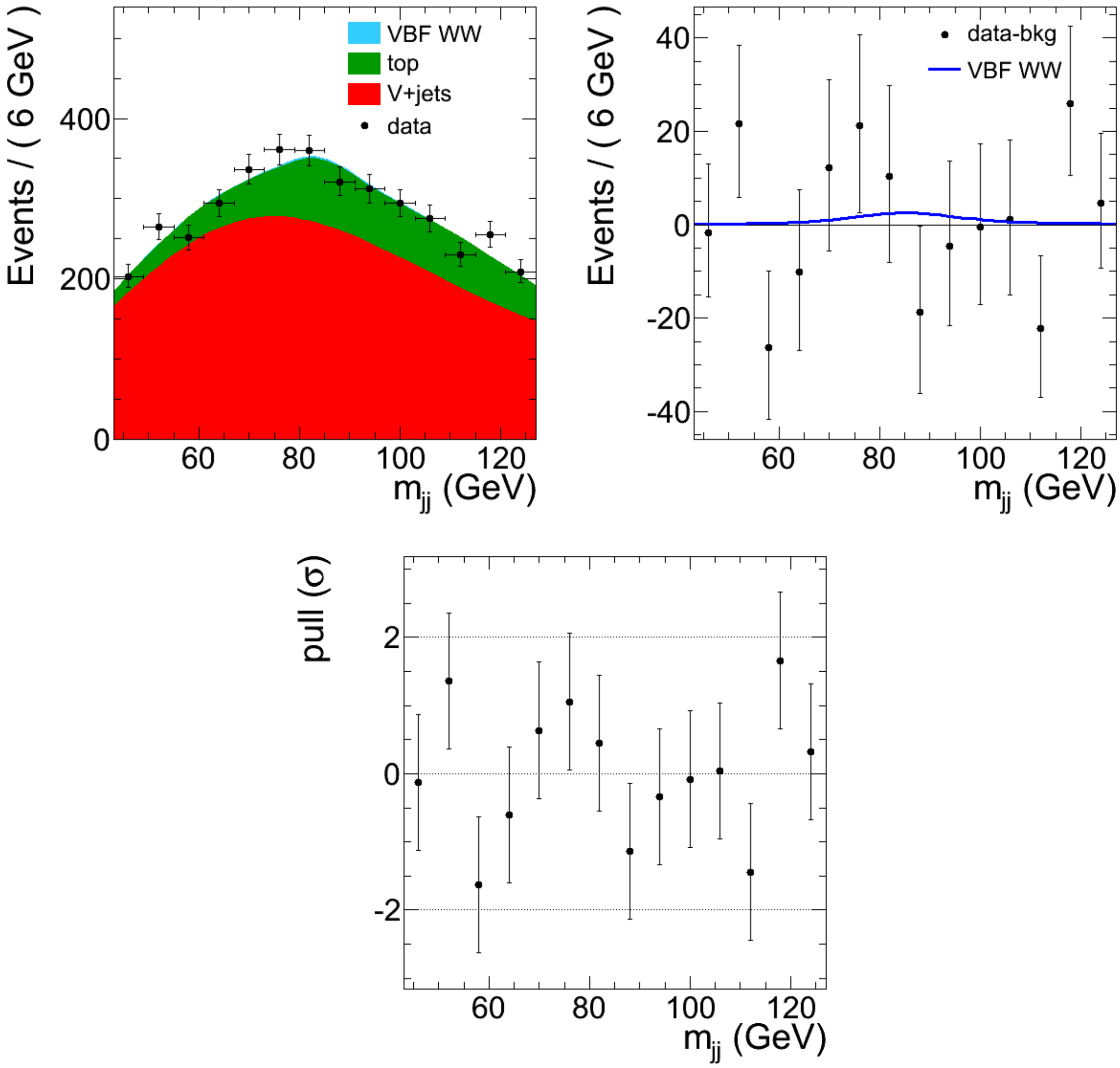


Figure 5.30: Distribution of the dijet invariant mass in electron data and corresponding fit shapes: (upper left) All background components stacked together, (upper right) Data minus all backgrounds except VBF diboson, lower row normalized residual between data and MC.

5.10 Systematic uncertainties

We consider several sources of systematic uncertainty, taking into account their effect on both the signal acceptance and on the template shapes for the signal extraction fit. The uncertainty on the normalization of the backgrounds is taken as part of the statistical uncertainty. Other sources of systematic uncertainties considered include jet energy scale (JES) as well as trigger and lepton identification efficiencies. The systematic uncertainties on signal are summarized in Table 5.4.

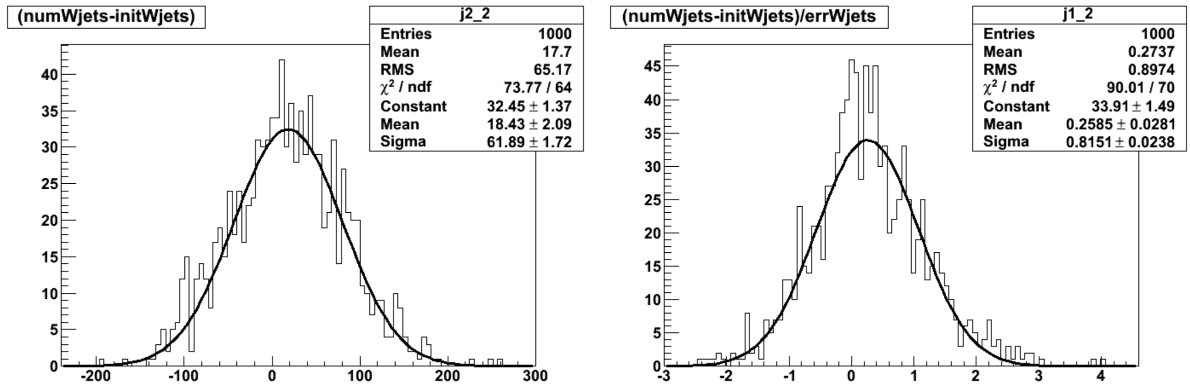


Figure 5.31: Distribution of the bias on the left and normalized residues on the right for W+jets background is shown.

Source of uncertainty	Magnitude
Luminosity	2.6%
Jet energy scale, resolution, and	<1%
Lepton trigger eff.	1%
Lepton selection eff.	2%
Pile-up	<1%

Table 5.4: Sources of signal systematics considered so far in the analysis, with the corresponding magnitude.

5.11 Conclusions

A study on measurement of WV (V being W/Z) bosons in vector boson fusion topology was presented in events with leptonically decaying W boson and a hadronically decaying W/Z bosons with events containing a well isolated lepton with high threshold on p_T, large missing traverse energy (E_T^{miss}) and four jets with high p_T, two of them required to be forward jets with large rapidity gap and large invariant mass and coming from opposite hemisphere. With the kinematic thresholds imposted in the analysis we conclude from the result that we are still about 5 times away from the SM, and we need more data to further increase the sensitivity of the analysis. This analysis can be a benchmark for future WW scattering analysis.

Chapter 6

Summary and Outlook

This thesis deals with two currently relevant topics of the SM, both of which are equally important for the SM physics. The first and the main topic is the search for the SM Higgs boson decaying to WW semileptonically ($H \rightarrow WW \rightarrow \ell\nu jj$), where one W boson decays leptonically and hence allows for triggering of events, while other W boson decays hadronically making this final state to possess second highest production cross-section times branching ratio over whole high mass range of the Higgs boson. The final state consists of a well-isolated electron or muon, missing transverse energy (E_T^{miss}), and two well-separated jets. The presence of two jets also brings in large background contributions from W+jets process and thus makes this analysis challenging at the LHC. The analyzed dataset is collected with a suite of single lepton triggers using transverse momentum (p_T) thresholds of 24 GeV for muons and 27 GeV for electrons, and is recorded by the CMS detector with an integrated luminosity of 19.3 fb^{-1} and 19.2 fb^{-1} for muon and electron channels respectively. Basic kinematic thresholds are applied to the final state objects to reduce the background contribution. The $\mu(e)$ is required to be well-isolated and have a p_T threshold greater than 25(35) GeV. Additional muons and electrons are rejected in the event to reduce backgrounds from Drell-Yan events. There is a threshold on the transverse mass of leptonic W boson, and E_T^{miss}, both greater than 30 GeV, which increases the probability to select real W bosons in the event. Further, to enhance signal over background ratio, events are chosen in the mass window 66-98 GeV, around the mass of the W boson. This preserves about 80% of signal events while significantly reducing the backgrounds. Further, to exploit the differences in kinematics between signal and background, a likelihood discriminant is constructed that consolidates even least possible discrimination in the Higgs signal from the W+jets background from a set of best modeled variables. This approach improves the expected sensitivity to Higgs mass hypothesis in this decay mode across the entire mass range.

The background normalization in the signal region is extracted independently for each Higgs mass hypothesis with an unbinned maximum likelihood fit to the dijet invariant mass distribution m_{jj} of the two leading jets in the side band of the dijet system.

The SM Higgs boson is excluded in the mass range 335–500 GeV at a confidence level of 95%. When this is combined with 7 TeV data, the exclusion range expands to 170–180 GeV and 230–545 GeV. The results were also interpreted within the context of a electroweak singlet heavy Higgs mixing with a light Higgs boson. Exclusion limits were set in the parameter space defining the coupling and branching ratios of the heavy Higgs. Further this result is combined with $H \to WW \to 2\ell2\nu$, $H \to ZZ \to 2\ell2\nu$, $H \to ZZ \to 4\ell$, $H \to ZZ \to 2\ell2\tau$ and $H \to ZZ \to 2\ell2q$ for Higgs boson mass hypotheses in the range 145 $< M_H <$ 1000 GeV. The data are interpreted both in the context of a SM-like heavy Higgs boson, as well as a search for a heavy, narrow resonance as Electroweak singlet partner of the SM Higgs at 126 GeV. No significant access over the expected SM background has been observed and exclusion limits have been set. The result is shown in figure 6.1.

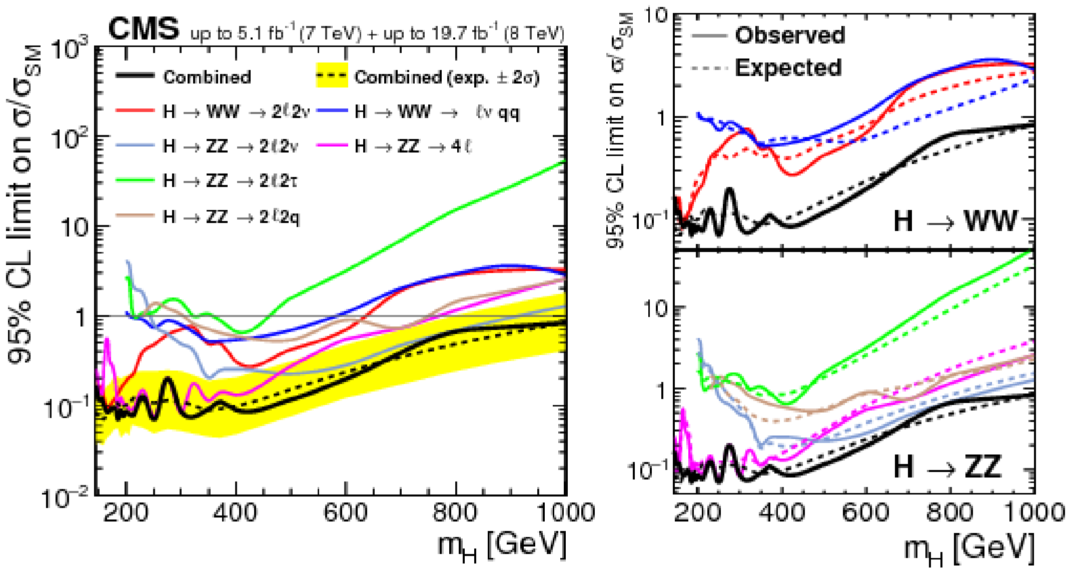


Figure 6.1: (a). On left plot upper limits at 95% C.L. limit is set for each of the contributing final states and their combination. Upper limits are displayed as a function of the SM Higgs boson mass. (b)On right upper limits at 95% C.L. are set on the electroweak singlet extension of the SM. Upper limits are displayed as a function of the heavy-Higgs-boson mass and the model parameters C'^2 and B_{new}. The theoretical cross section,σ_{BSM}, is the modified heavy-Higgs-boson cross section in the electroweak singlet model interpretation.

The results from low mass SM Higgs boson search indicates that the newly discovered boson is more and more consistent with the SM Higgs boson. In figure 6.2 a measurement of the Higgs boson mass is presented based on the combined data samples of the ATLAS and CMS experiments at the CERN LHC in the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels. The results are obtained from a simultaneous fit to the reconstructed invariant mass peaks in the two decay modes and for the two experiments. The measured masses from the individual channels and the two experiments are found to be consistent among themselves. The combined measured mass of the Higgs boson is $m_H = 125.09 \pm 0.21$ (stat.) $\pm$ 0.11 (syst.) GeV.

Further, the study of the spin-parity and tensor structure [105] of the interactions of the recently discovered Higgs boson is performed under various models and all are found to be consistent with the SM Higgs boson.

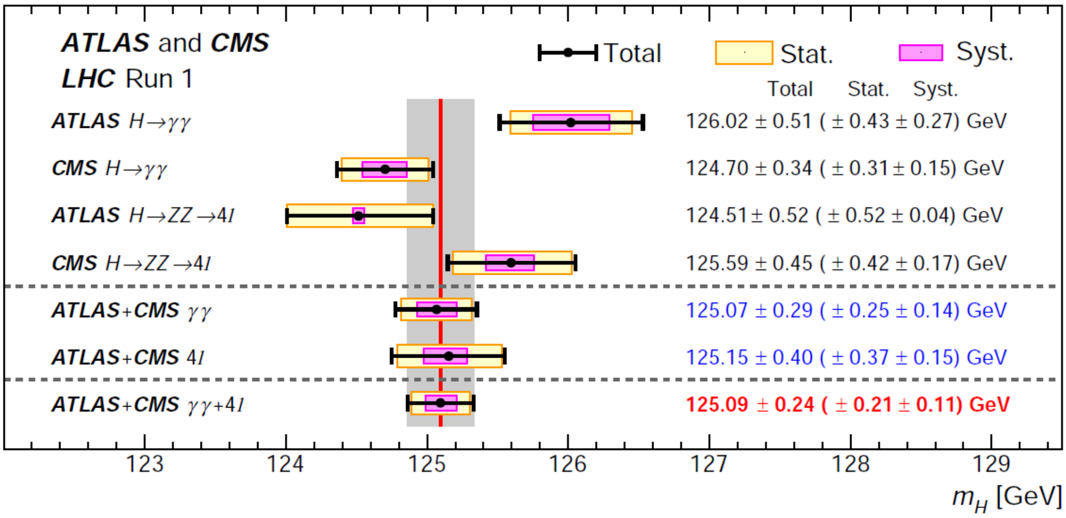


Figure 6.2: Summary of Higgs boson mass measurements from the individual analyses of ATLAS and CMS and from the combined analysis presented here. The systematic (narrower, magenta-shaded bands), statistical (wider, yellow-shaded bands), and total (black error bars) uncertainties are indicated. The (red) vertical line and corresponding (gray) shaded column indicate the central value and the total uncertainty of the combined measurement, respectively.

The second analysis covered in this thesis is the measurement of WW or WZ production with semileptonic final states in association two tag jets. Tag jets are chosen on the basis of its distinctive signature as they are produced with large rapidity gap and huge invariant mass. Since these jets arise directly from initial state quarks, hence they are very energetic. The main experimental challenge for this analysis is to control W+jets and $t\bar{t}$ backgrounds.

The data analysed and pre-selection are same as in first analysis. The additional and peculiar selection is on the jets. Jets are selected from complete detector reach (which is $|\eta| < 4.7$). Events are further selected from a loose mass window selection on dijet invariant mass (43 GeV $< m_{jj} <$ 127 GeV) to obtain enriched signal region. Top quark background normalization is estimated from top quark control region which is chosen partially exclusive to signal region but care is still taken so as to keep the kinematics as close as possible to the signal region. Signal is extracted from unbinned maximum likelihood fit to dijet invariant mass distribution. The combined result from electron and muon final states predict its sensitivity to be about 4 times away from the SM. This means that more data is needed to explore and understand this process. There has been tools and techniques such as quark-gluon jets tagging and coverage of complete η range etc.

which are in development stage and which have potential to increase the sensitivity of this analysis.

This final state provides a huge potential at LHC RUN 2. This analysis is sensitive to quartic gauge coupling, Higgs boson production through vector boson fusion etc. There is already an evidence for EW signal seen both from ATLAS and CMS [106, 107] experiment in fully leptonic final state where signal is less dominated by background to this analysis.

Appendix A

Signal Efficiency

A.1 Signal Efficiencies

Table A.1 lists signal efficiencies as a function of mass point and channel, both before and after the MVA output selection is applied.

Table A.1: Acceptance uncertainty related to the PDFs variation, for the signal rate, as a function of the mass hypothesis.

Mass (GeV)	Channel	Efficiency before MVA cut (%)	MVA cut value	Efficiency after MVA cut (%)
170	el+2jet	1.169 ± 0.032	0.300	1.143 ± 0.032
170	el+3jet	0.164 ± 0.012	0.300	0.157 ± 0.012
170	mu+2jet	3.113 ± 0.052	0.300	3.037 ± 0.052
170	mu+3jet	0.498 ± 0.021	0.300	0.490 ± 0.021
180	el+2jet	1.356 ± 0.035	0.550	0.815 ± 0.027
180	el+3jet	0.220 ± 0.014	0.300	0.214 ± 0.014
180	mu+2jet	3.029 ± 0.052	0.600	1.396 ± 0.035
180	mu+3jet	0.479 ± 0.021	0.300	0.475 ± 0.021
190	el+2jet	1.417 ± 0.036	0.550	0.916 ± 0.029
190	el+3jet	0.251 ± 0.015	0.300	0.232 ± 0.015
190	mu+2jet	3.055 ± 0.052	0.600	1.571 ± 0.037
190	mu+3jet	0.525 ± 0.022	0.300	0.464 ± 0.020
200	el+2jet	1.655 ± 0.039	0.600	1.008 ± 0.030
200	el+3jet	0.288 ± 0.016	0.400	0.219 ± 0.014
200	mu+2jet	3.145 ± 0.053	0.600	1.777 ± 0.040
200	mu+3jet	0.539 ± 0.022	0.300	0.483 ± 0.021
250	el+2jet	2.441 ± 0.047	0.650	1.533 ± 0.037
250	el+3jet	0.587 ± 0.023	0.400	0.477 ± 0.021
250	mu+2jet	4.166 ± 0.060	0.650	2.542 ± 0.047
250	mu+3jet	0.915 ± 0.029	0.400	0.720 ± 0.025
300	el+2jet	3.328 ± 0.054	0.600	2.591 ± 0.048
300	el+3jet	0.916 ± 0.029	0.500	0.759 ± 0.026
300	mu+2jet	4.730 ± 0.064	0.600	3.496 ± 0.055
300	mu+3jet	1.382 ± 0.035	0.600	0.970 ± 0.030
350	el+2jet	3.890 ± 0.058	0.600	2.978 ± 0.051
350	el+3jet	1.287 ± 0.034	0.650	0.840 ± 0.028
350	mu+2jet	5.451 ± 0.069	0.600	4.093 ± 0.060
350	mu+3jet	1.664 ± 0.039	0.650	1.037 ± 0.031
400	el+2jet	4.340 ± 0.062	0.550	3.582 ± 0.057
400	el+3jet	1.556 ± 0.038	0.600	1.144 ± 0.032
400	mu+2jet	5.678 ± 0.070	0.550	4.565 ± 0.064
400	mu+3jet	2.138 ± 0.044	0.500	1.735 ± 0.040
450	el+2jet	4.465 ± 0.063	0.500	3.948 ± 0.059
450	el+3jet	1.778 ± 0.040	0.550	1.271 ± 0.034
450	mu+2jet	5.908 ± 0.072	0.600	4.738 ± 0.065
450	mu+3jet	2.370 ± 0.046	0.550	1.662 ± 0.039
500	el+2jet	4.377 ± 0.062	0.500	3.764 ± 0.058
500	el+3jet	1.871 ± 0.041	0.600	1.049 ± 0.031
500	mu+2jet	5.838 ± 0.072	0.500	4.945 ± 0.066
500	mu+3jet	2.527 ± 0.048	0.650	1.188 ± 0.033
550	el+2jet	4.314 ± 0.062	0.550	3.531 ± 0.056
550	el+3jet	1.915 ± 0.042	0.500	1.363 ± 0.035
550	mu+2jet	5.597 ± 0.070	0.550	4.545 ± 0.063
550	mu+3jet	2.382 ± 0.046	0.500	1.677 ± 0.039
600	el+2jet	4.046 ± 0.060	0.600	2.916 ± 0.051
600	el+3jet	1.837 ± 0.041	0.600	1.058 ± 0.031
600	mu+2jet	5.260 ± 0.068	0.650	3.508 ± 0.056
600	mu+3jet	2.294 ± 0.045	0.600	1.302 ± 0.034

Appendix B

Interference

B.1 Heavy Higgs Interference Effects

With increasing Higgs mass, the interference effect becomes more and more crucial between the signal $gg \to H \to WW$ and the continuum background $gg \to WW$. As shown in Ref. [87], the effect on LO total cross sections is already over 30% for $M_H 600\,\mathrm{GeV}$. Moreover, it changes the M_{WW} spectrum substantially, with constructive behavior for $M_{WW} M_H$ and destructive for $M_{WW} M_H$. One usually can include the interference effects by reweighting the signal's M_{WW} distributions with $(1 + R2)$, where the ratio R_2 is defined as

$$R_2 \equiv \frac{Intf}{S}, \tag{B.1}$$

However, one needs then to validate the method by studying other distributions than M_{WW}. In addition, one should include conservative yet appropriate estimations on uncertainty of R_2.

In our study, we evaluate R_2 with LO $Intf$ and S obtained from the MCFM v6.3 (without any cuts applied). We then exploit the method proposed in Ref. [85] to estimate the reweighting uncertainty with 3 kinds of R_2 inspired by higher order QCD corrections:

$$R_2 = 1,\ \sqrt{K_{gg}}/K_{NNLO},\ \text{and } 1/K_{NNLO}, \tag{B.2}$$

where the K factors K_{gg} and K_{NNLO} are read from the $H \to ZZ$ numbers in Ref. [85], as the ones for $H \to WW$ are not ready yet, and the K factors should not depend on final states between $gg \to H \to WW$ and ZZ.

In Fig. B.1, we show our reweighting results for $gg \to H \to WW \to l\nu jj$ at the generator level, for $M_H = 700\,\mathrm{GeV}$ at the 8TeV LHC, with $P_{T,l} > 30\,\mathrm{GeV}$, $P_{T,j} > 30\,\mathrm{GeV}$, MET$> 30\,\mathrm{GeV}$, and $|\eta_{l,j}| < 2.4$. The reweighting based on M_{WW} spectrum turns out to decribe well the other distributions. The uncertainty band on interference effect reweighting, as mentioned above, is also provided for various distributions.

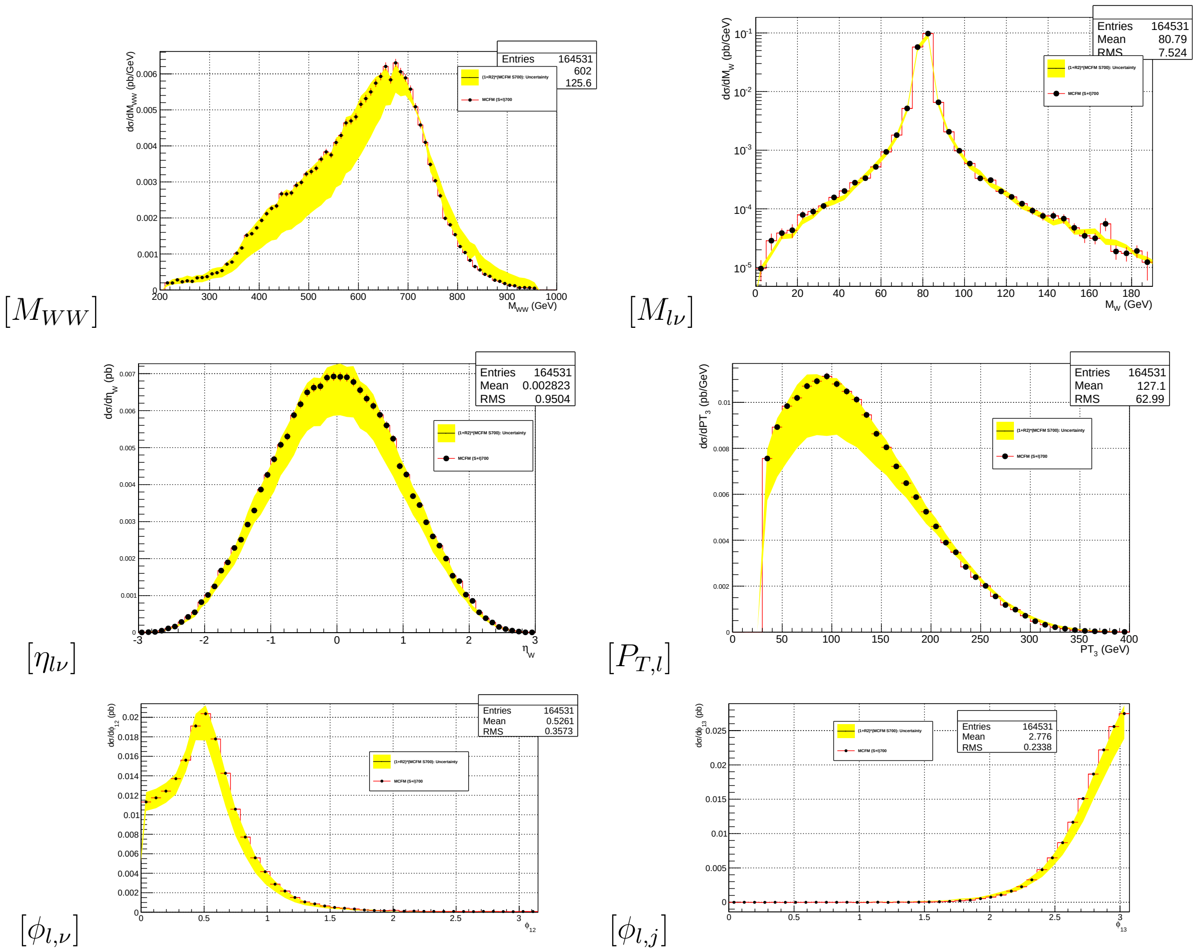


Figure B.1: Reweighting MCFM $gg \to H \to WW \to l\nu jj$ with interference effects at the generator level, for $M_H = 700\,\text{GeV}$ at the 8TeV LHC, with $P_{T,l} > 30\,\text{GeV}$, $P_{T,j} > 30\,\text{GeV}$, MET$> 30\,\text{GeV}$, and $|\eta_{l,j}| < 2.4$.

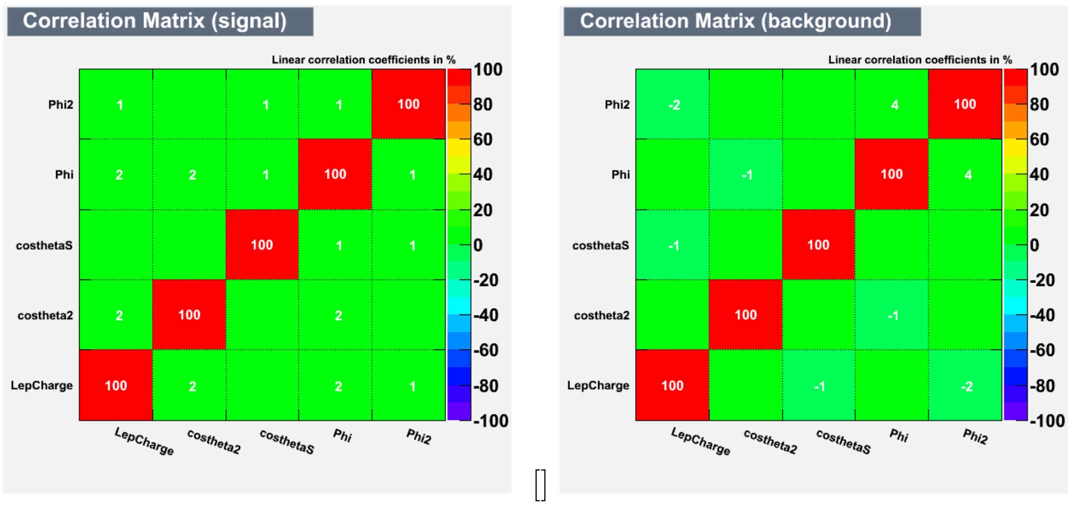
Appendix C

MVA Optimization

C.1 Appendix: MVA Correlations

we use a set of minimum number of input variables which are mostly uncorrelated. The correlation matrix for signal and background samples are shown in Figs C.1–C.12 for all Higgs mass points.

C.1.1 Correlation matrix: M_H = 170 GeV



[] []

Figure C.1: Input correlation matrix for M_H =170 GeV: (a) signal, (b) W+jets background

C.1.2 Correlation matrix: M_H = 180 GeV

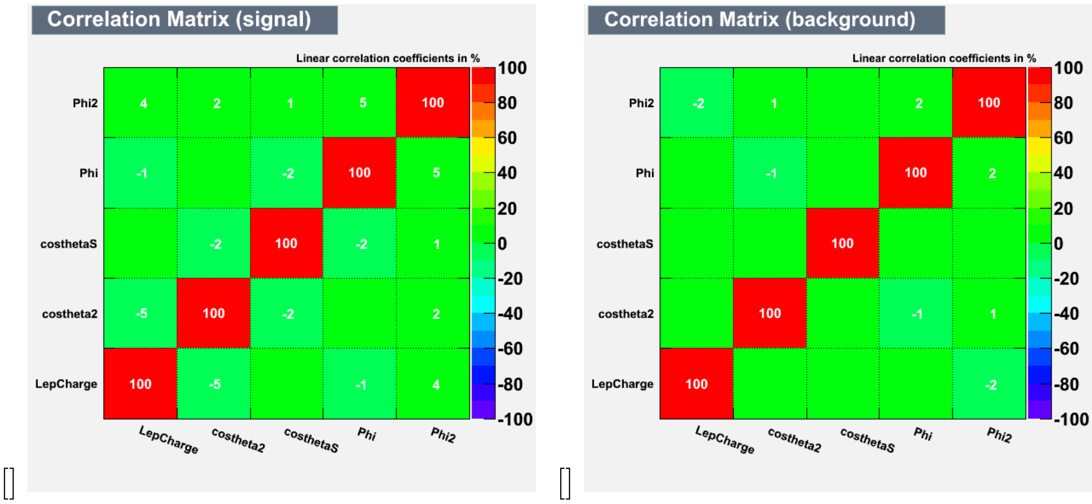


Figure C.2: Input correlation matrix for M_H =180 GeV: (a) signal, (b) W+jets background

C.1.3 Correlation matrix: M_H = 190 GeV

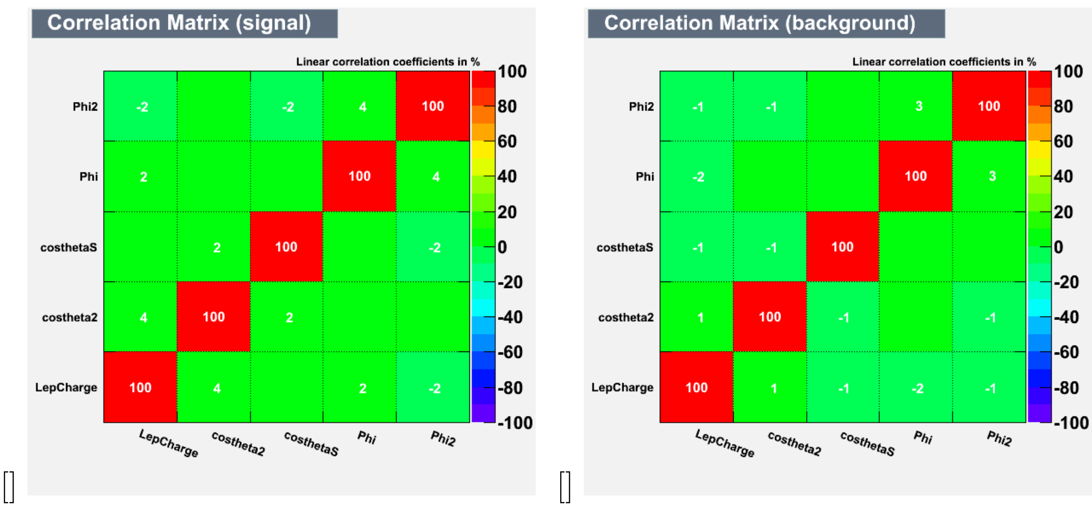


Figure C.3: Input correlation matrix for M_H =190 GeV: (a) signal, (b) W+jets background

C.1.4 Correlation matrix: M_H = 200 GeV

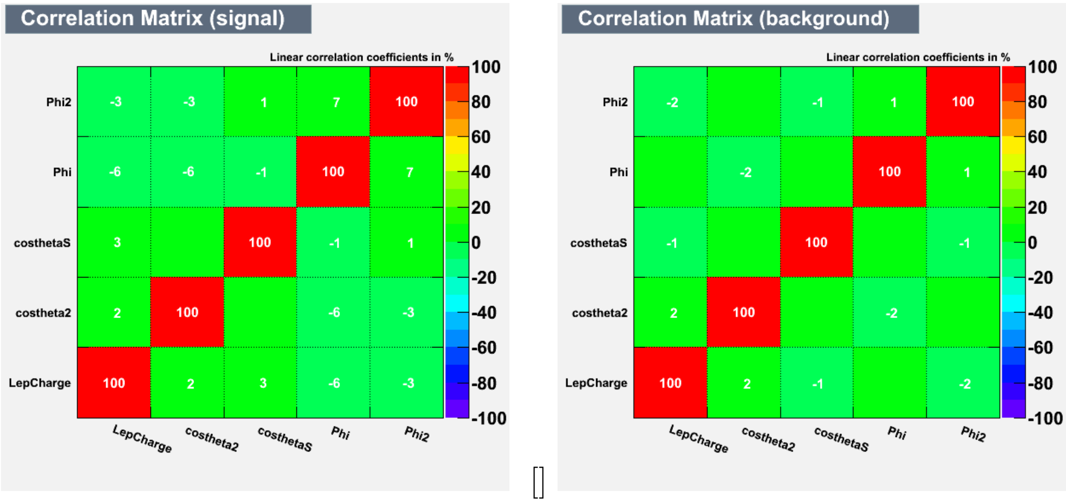


Figure C.4: Input correlation matrix for M_H =200 GeV: (a) signal, (b) W+jets background

C.1.5 Correlation matrix: M_H = 250 GeV

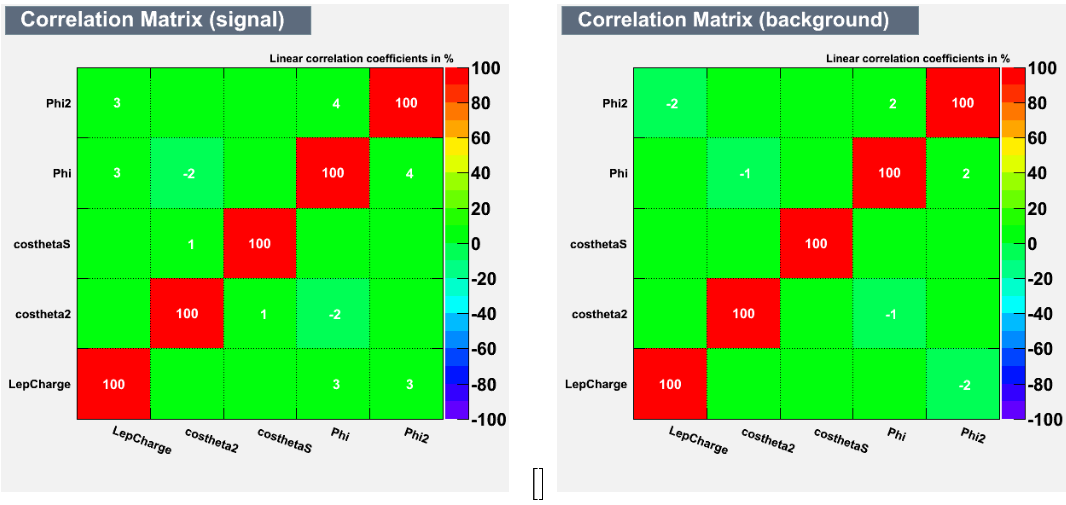


Figure C.5: Input correlation matrix for M_H =250 GeV: (a) signal, (b) W+jets background

C.1.6 Correlation matrix: M_H = 300 GeV

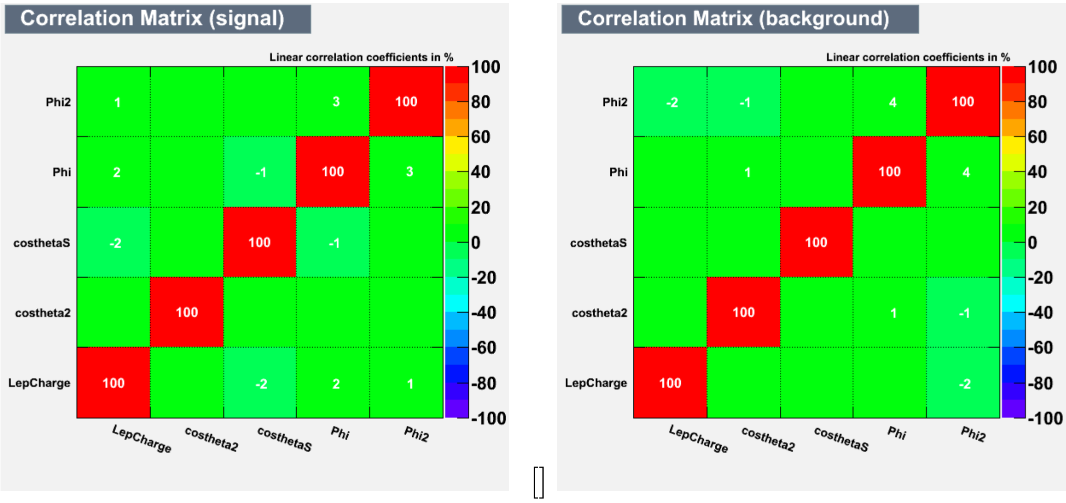


Figure C.6: Input correlation matrix for M_H =300 GeV: (a) signal, (b) W+jets background

C.1.7 Correlation matrix: M_H = 350 GeV

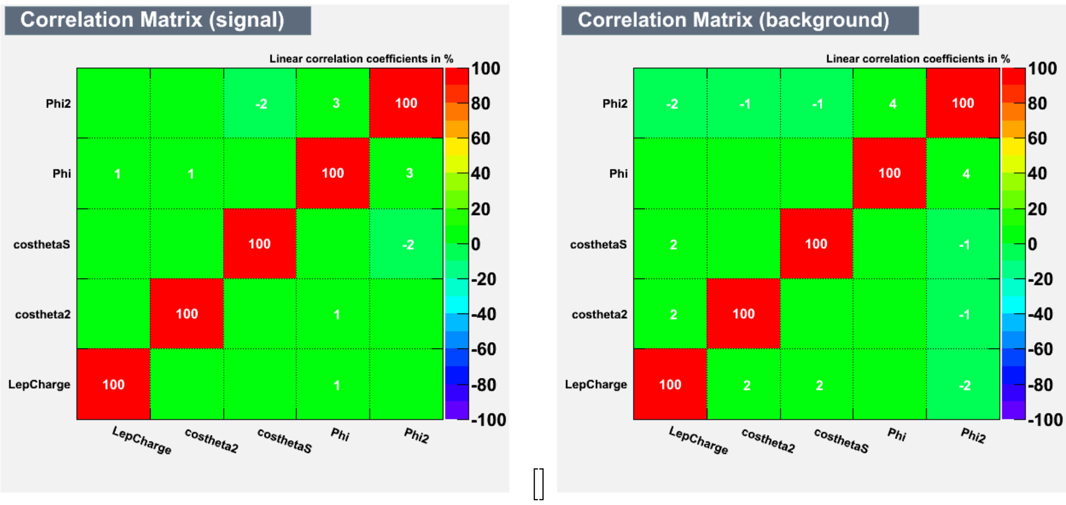
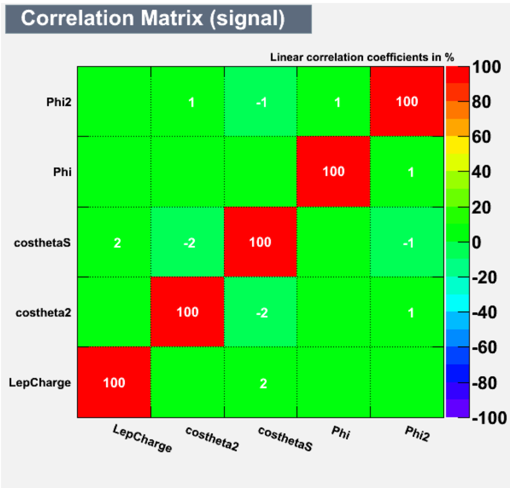


Figure C.7: Input correlation matrix for M_H =350 GeV: (a) signal, (b) W+jets background

C.1.8 Correlation matrix: M_H = 400 GeV

[] 

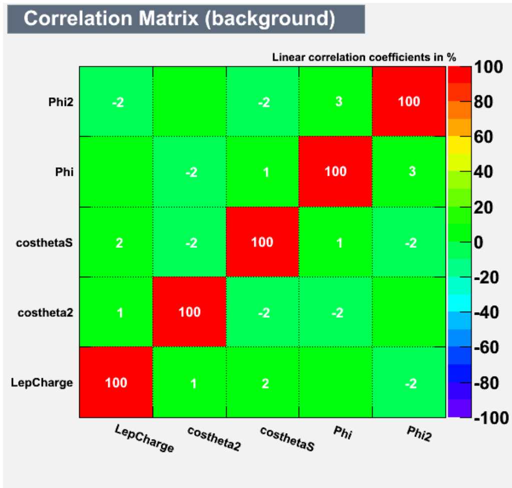
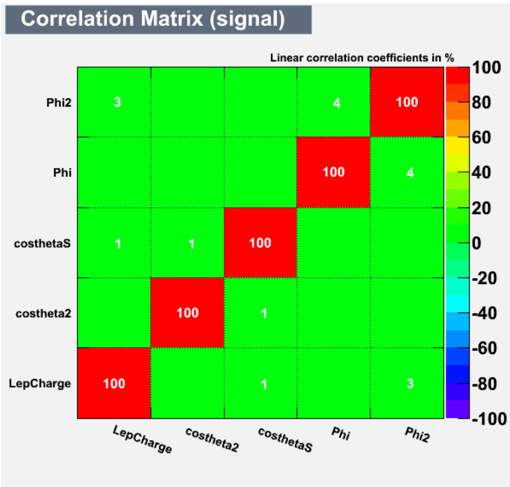
[] 

Figure C.8: Input correlation matrix for M_H =400 GeV: (a) signal, (b) W+jets background

C.1.9 Correlation matrix: M_H = 450 GeV

[] 

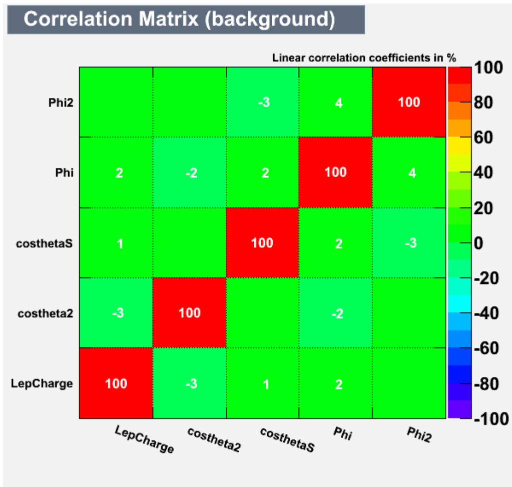
[] 

Figure C.9: Input correlation matrix for M_H =450 GeV: (a) signal, (b) W+jets background

C.1.10 Correlation matrix: M_H = 500 GeV

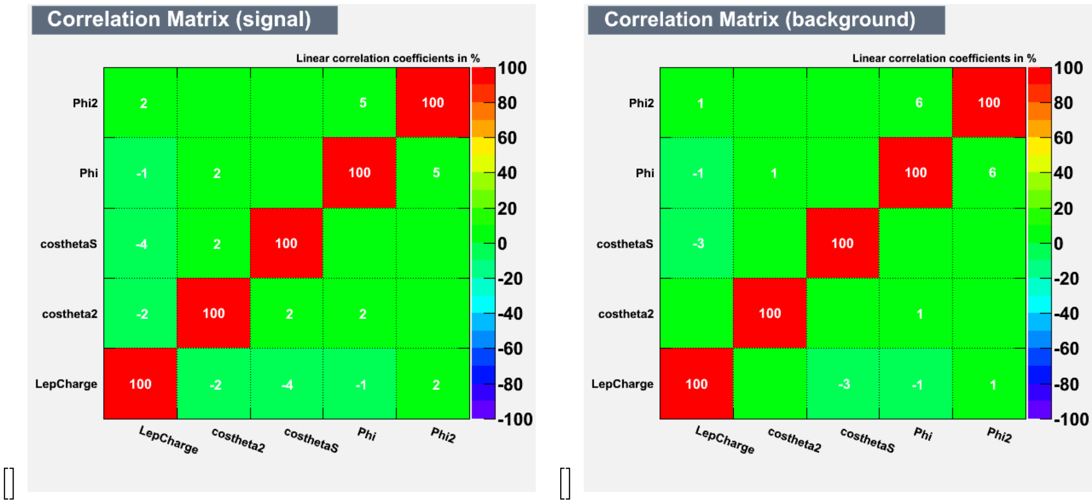


Figure C.10: Input correlation matrix for M_H =500 GeV: (a) signal, (b) W+jets background

C.1.11 Correlation matrix: M_H = 550 GeV

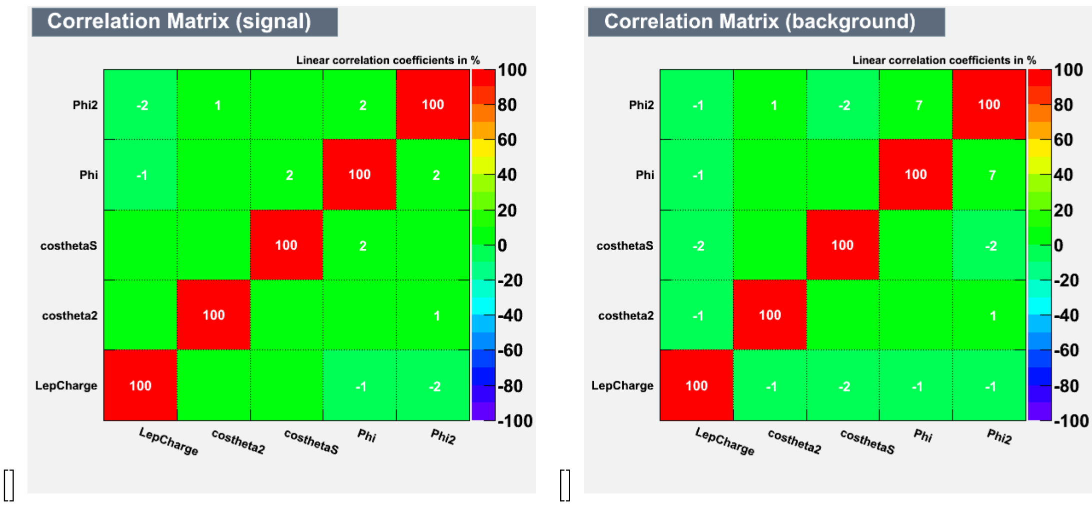


Figure C.11: Input correlation matrix for M_H =550 GeV: (a) signal, (b) W+jets background

C.1.12 Correlation matrix: M_H = 600 GeV

C.2 Data - Monte Carlo comparisons in Top Control region

Having seen disageement in data-MC comparitions in the phase space of our interest, we tried to understand it by going in top control region. The top control region is defined as events which have two btagged and two anti-btagged jets. We find that the quality of

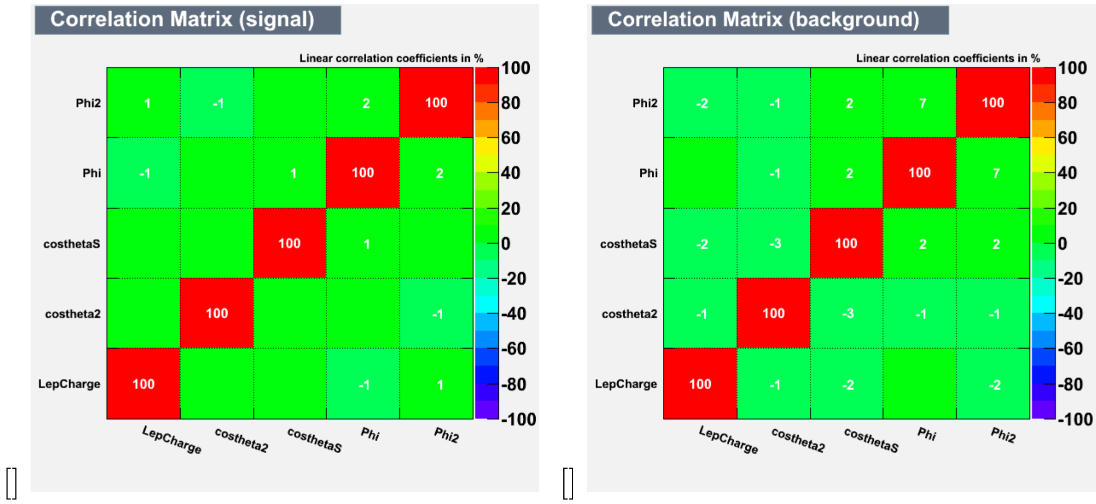


Figure C.12: Input correlation matrix for M_H =600 GeV: (a) signal, (b) W+jets background

the modeling provided by top enriched sample is very good. So, We conclude that the disagreement seen in there is because of mis-modelling of W+Jets simulated background. The MC is normalized to the number of collected events in data after applying top event selection criteria. We show the distributions of various kinematic variables in Figures C.13-C.17 for the muon+jets sample and in Figures C.18-C.23 for the electron+jets sample. Eventually, Figure C.22 shows the four-body invariant mass, for electrons (on the left) and muons (on the right). The MC has been been corrected for lepton reconstruction efficiency and the trigger efficiency. Seeing reasonable agreement gives us confidence in the qualitative aspects of the MC modeling.

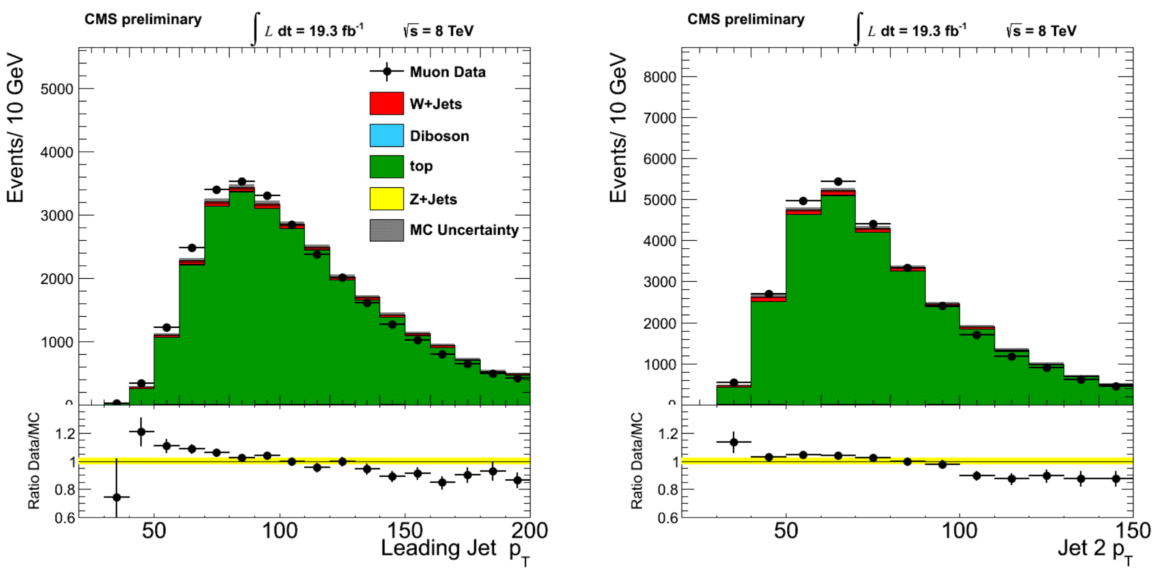


Figure C.13: Comparison of the leading jet (left) and the second jet (right) p_T distributions from data and MC for the muon+jets selection.

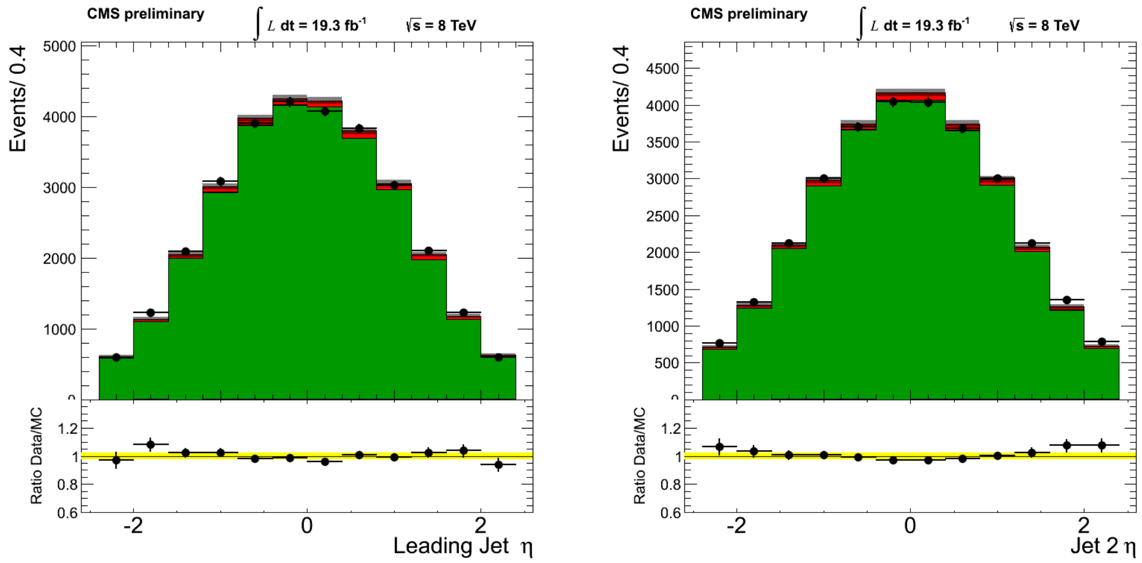


Figure C.14: Comparison of the leading jet η (left) and the second leading jet η (right) distributions from data and MC for the muon+jets selection.

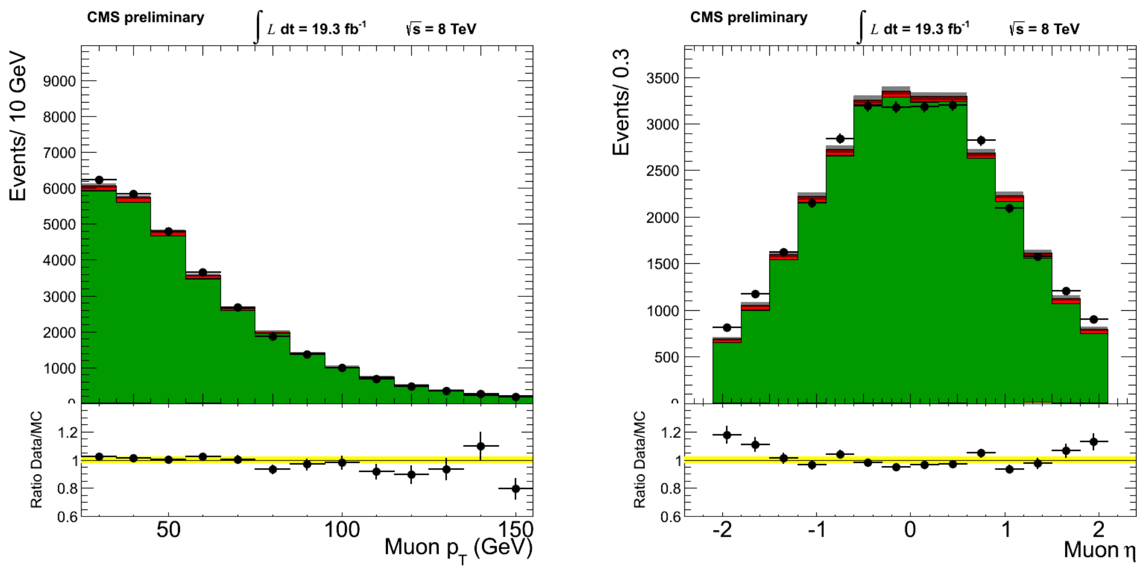


Figure C.15: Comparison of the muon p_T (left) and the muon η (right) distributions from data and MC for the muon+jets selection.

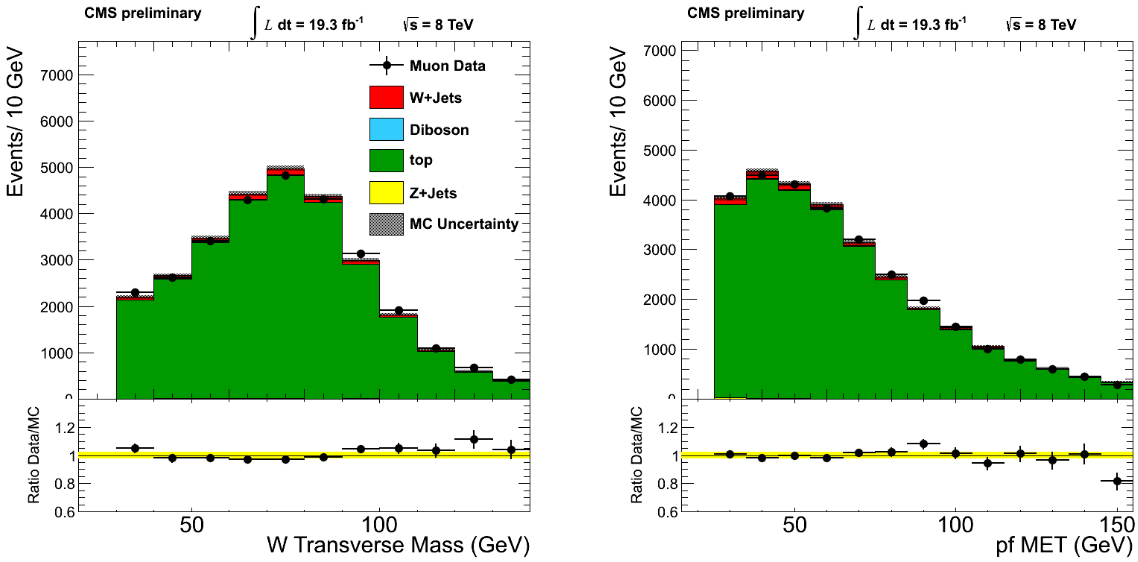


Figure C.16: Comparison of the distributions from data and MC of the transverse mass of the muon / MET system (left) and the MET (right) for the muon+jets selection.

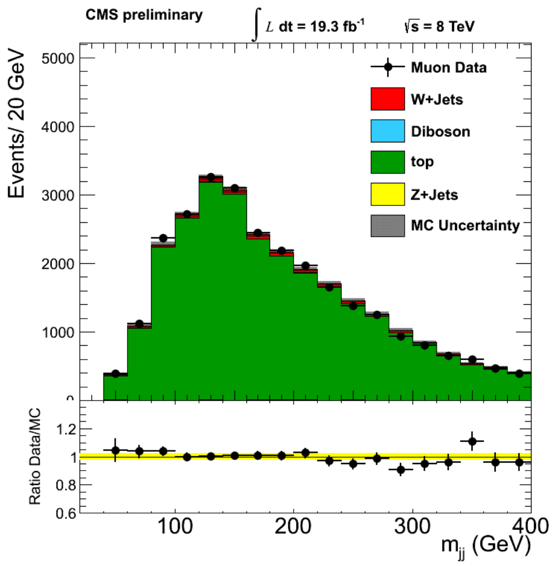


Figure C.17: Comparison of the dijet mass (m_{JJ}) distributions from data and MC for the muon+jets selection.

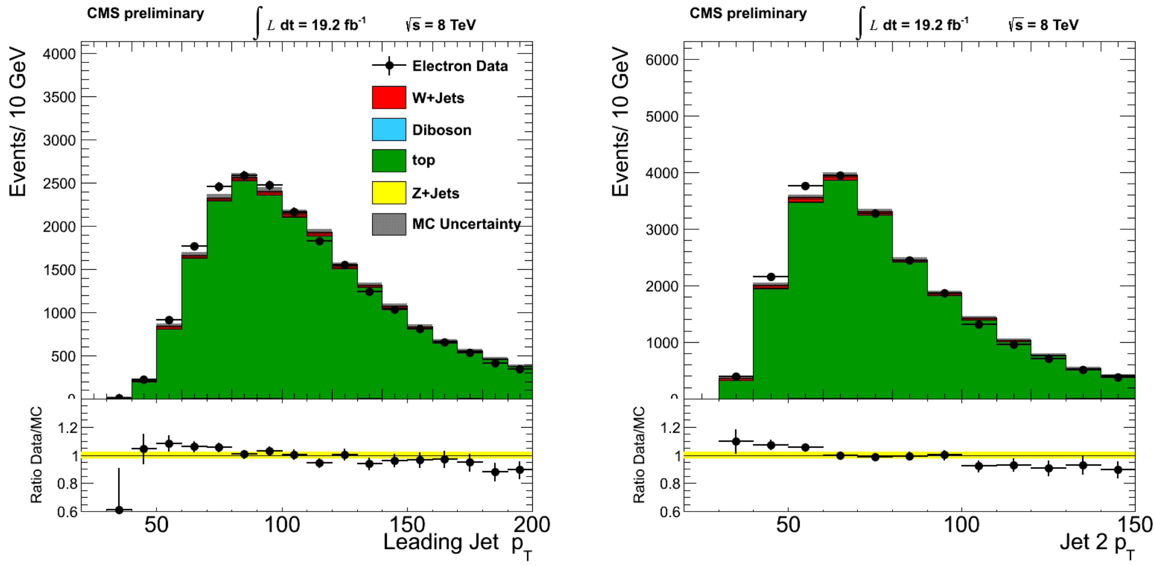


Figure C.18: Comparison of the leading jet p_T (left) and the second leading jet p_T (right) distributions from data and MC for the electron+jets selection.

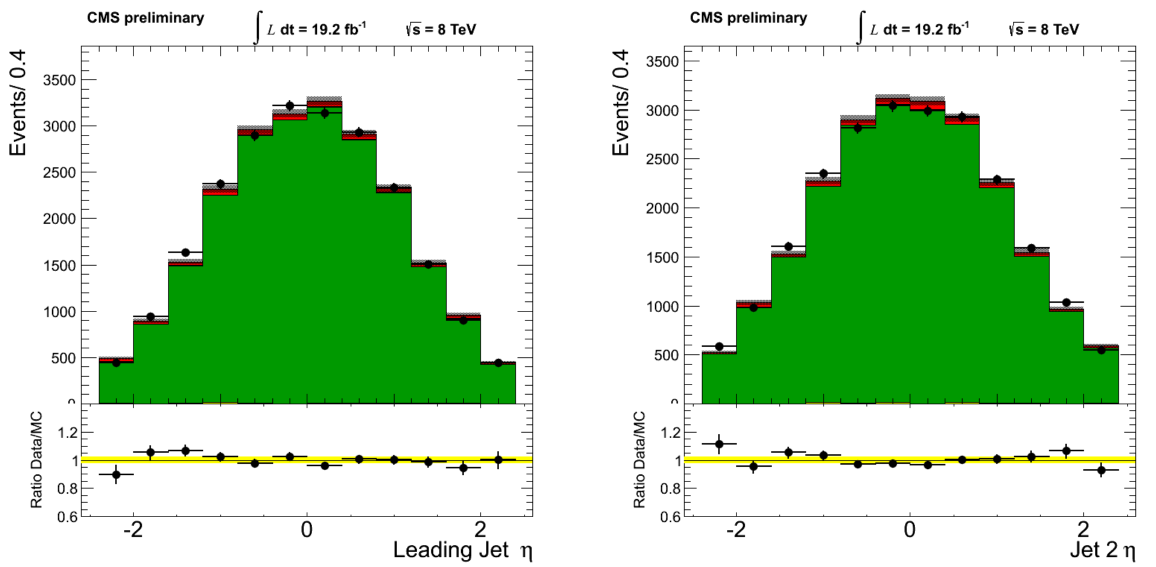


Figure C.19: Comparison of the leading jet η (left) and the second leading jet η (right) distributions from data and MC for the electron+jets selection.

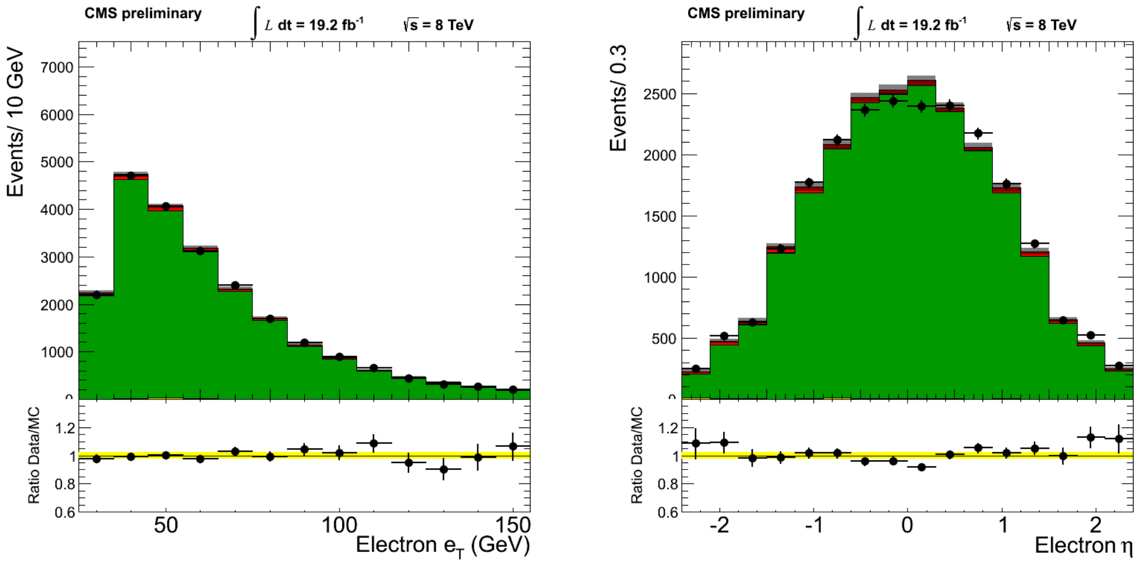


Figure C.20: Comparison of the electron E_T (left) and the electron η (right) distributions from data and MC for the electron+jets selection.

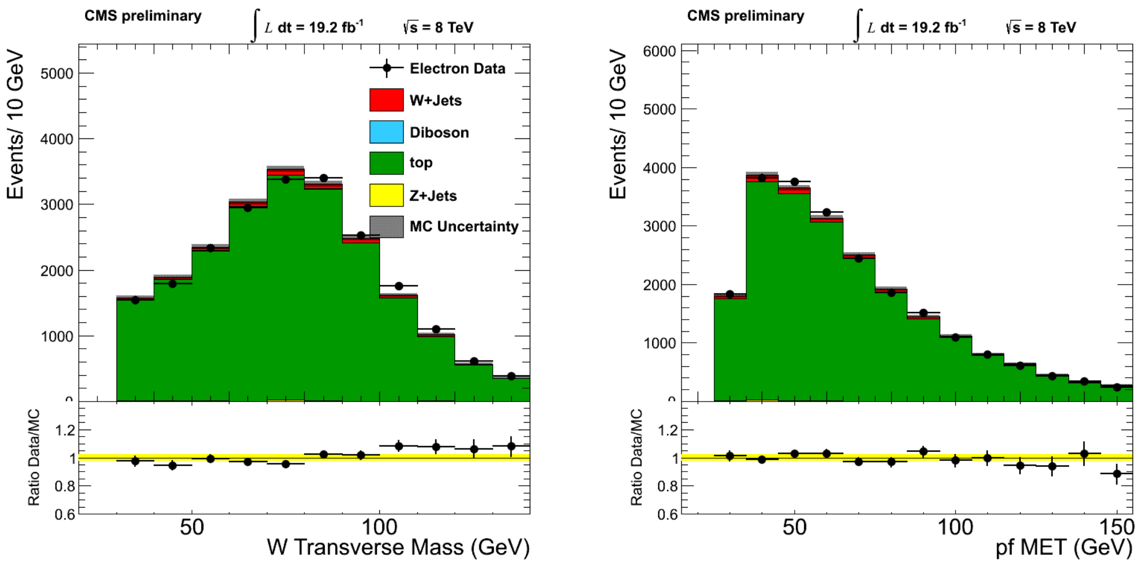


Figure C.21: Comparison of the distributions from data and MC of the transverse mass of electron / MET system (left) and the MET (right) for the electron+jets selection.

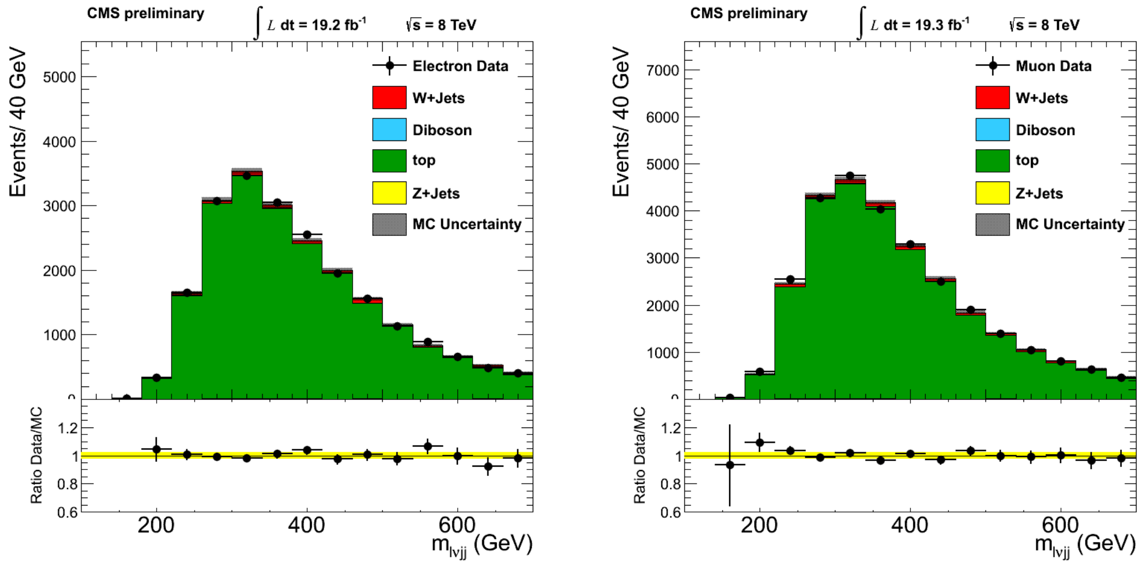


Figure C.22: Comparison of the four-body invariant mass from data and MC for the electron+jets selection (left) and muon+jets selection (right).

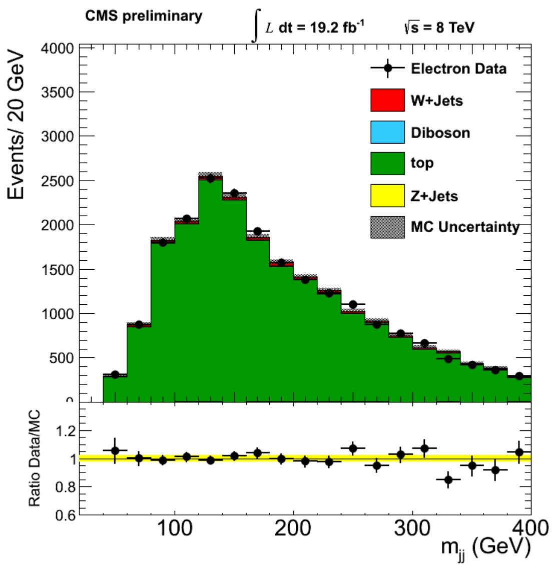


Figure C.23: Comparison of the dijet mass (m_{JJ}) distributions from data and MC for the electron+jets selection.

Bibliography

[1] S. L. Glashow. Partial-Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579, 1961.

[2] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264, 1967.

[3] Abdus Salam. Weak and electromagnetic interactions. In Nils Svartholm, editor, *Elementary particle physics: relativistic groups and analyticity*, page 367. Almqvist & Wiksell, 1968. Proceedings of the eighth Nobel symposium.

[4] S. Dittmaier and M. Schumacher. The Higgs Boson in the Standard Model - From LEP to LHC: Expectations, Searches, and Discovery of a Candidate. *Prog.Part.Nucl.Phys.*, 70:1–54, 2013.

[5] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321, 1964.

[6] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132, 1964.

[7] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585, 1964.

[9] Peter W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.*, 145:1156, 1966.

[10] T. W. B. Kibble. Symmetry breaking in non-Abelian gauge theories. *Phys. Rev.*, 155:1554, 1967.

[11] Laura Reina. TASI 2004 lecture notes on Higgs boson physics. pages 261–356, 2005.

[12] D. A. Dicus and V. S. Mathur. Upper bounds on the values of masses in unified gauge theories. *Phys. Rev. D*, 7:3111, 1973.

[13] M. J. G. Veltman. Second Threshold in Weak Interactions. *Acta Phys. Polon. B*, 8:475, 1977.

[14] Benjamin W. Lee, C. Quigg, and H B. Thacker. Weak interactions at very high-energies: The role of the higgs boson mass. *Phys. Rev. D*, 16:1519, 1977.

[15] B W. Lee, C. Quigg, and H B. Thacker. The Strength of Weak Interactions at Very High-Energies and the Higgs Boson Mass. *Phys. Rev. Lett.*, 38:883, 1977.

[16] G. Passarino. Ww scattering and perturbative unitarity. *Nucl. Phys. B*, 343:31, 1990.

[17] Michael S. Chanowitz and Mary K. Gaillard. The tev physics of strongly interacting w's and z's. *Nucl. Phys. B*, 261:379, 1985.

[18] Malcolm J. Duncan, Gordon L. Kane, and W. W. Repko. Ww physics at future colliders. *Nucl. Phys. B*, 272:517, 1986.

[19] D. A. Dicus and R. Vega. Ww production from pp collisions. *Phys. Rev. Lett.*, 57:1110, 1986.

[20] J. Bagger, V. Barger, K. Cheung, J. Gunion, T. Han, G. A. Ladinsky, and C.-P. Yuan. Cern lhc analysis of the strongly interacting ww system: Gold-plated modes. *Phys. Rev. D*, 52:3878, 1995.

[21] Alessandro Ballestrero, Giuseppe Bevilacqua, Diogo Buarque Franzosi, and Ezio Maina. How well can the lhc distinguish between the sm light higgs scenario, a composite higgs and the higgsless case using vv scattering channels? *JHEP*, 11:126, 2009.

[22] Thomas Hambye and Kurt Riesselmann. Matching conditions and Higgs mass upper bounds revisited. *Phys.Rev.*, D55:7255–7262, 1997.

[23] ALEPH, DELPHI, L3, OPAL Collaborations, and the LEP Working Group for Higgs boson searches. Search for the standard model Higgs boson at LEP. *Phys. Lett. B*, 565:61, 2003.

[24] Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data. 2012.

[25] The lep electrweak working group. *CMS twiki*, LEP EWWG, 2015.

[26] M. Ciccolini, Denner A., and Dittmaier S. Electroweak and qcd corrections to higgs production via vector-boson fusion at the cern lhc. *Phys. Rev. D*, 77:013002, Jan 2008.

[27] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN-2013-004*, CERN, Geneva, 2013.

[28] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, CERN, Geneva, 2012.

[29] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011.

[30] Brian Patt and Frank Wilczek. Higgs-field portal into hidden sectors. 2006.

[31] Vernon Barger, Paul Langacker, Mathew McCaskey, Michael J. Ramsey-Musolf, and Gabe Shaughnessy. Cern lhc phenomenology of an extended standard model with a real scalar singlet. *Phys. Rev. D*, 77:035005, Feb 2008.

[32] Gautam Bhattacharyya, Gustavo C. Branco, and S. Nandi. Universal doublet-singlet higgs couplings and phenomenology at the cern large hadron collider. *Phys. Rev. D*, 77:117701, Jun 2008.

[33] Sally Dawson and Wenbin Yan. Hiding the higgs boson with multiple scalars. *Phys. Rev. D*, 79:095002, May 2009.

[34] Sebastian Bock, Remi Lafaye, Tilman Plehn, Michael Rauch, Dirk Zerwas, and Peter M. Zerwas. Measuring hidden higgs and strongly-interacting higgs scenarios. *Physics Letters B*, 694(1):44 – 53, 2010.

[35] Patrick J. Fox, David Tucker-Smith, and Neal Weiner. Higgs friends and counterfeits at hadron colliders. *JHEP*, 1106:127, 2011.

[36] Christoph Englert, Tilman Plehn, Dirk Zerwas, and Peter M. Zerwas. Exploring the Higgs portal. *Phys.Lett.*, B703:298–305, 2011.

[37] Christoph Englert, Joerg Jaeckel, Emanuele Re, and Michael Spannowsky. Evasive higgs boson maneuvers at the lhc. *Phys. Rev. D*, 85:035008, Feb 2012.

[38] Brian Batell, Stefania Gori, and Lian-Tao Wang. Exploring the Higgs Portal with 10/fb at the LHC. *JHEP*, 1206:172, 2012.

[39] Christoph Englert, Tilman Plehn, Michael Rauch, Dirk Zerwas, and Peter M. Zerwas. LHC: Standard Higgs and Hidden Higgs. *Phys.Lett.*, B707:512–516, 2012.

[40] Rick S. Gupta and James D. Wells. Higgs boson search significance deformations due to mixed-in scalars. *Phys.Lett.*, B710:154–158, 2012.

[41] Matthew J. Dolan, Christoph Englert, and Michael Spannowsky. New Physics in LHC Higgs boson pair production. *Phys.Rev.*, D87:055002, 2013.

[42] Daniele Bertolini and Matthew McCullough. The Social Higgs. *JHEP*, 1212:118, 2012.

[43] Brian Batell, David McKeen, and Maxim Pospelov. Singlet Neighbors of the Higgs Boson. *JHEP*, 1210:104, 2012.

[44] Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV. 2013.

[45] S Heinemeyer, C Mariotti, G Passarino, R Tanaka, J R Andersen, P Artoisenet, E A Bagnaschi, A Banfi, T Becher, F U Bernlochner, S Bolognesi, P Bolzoni, R Boughezal, D Buarque, J Campbell, F Caola, M Carena, F Cascioli, N Chanon, T Cheng, S Y Choi, A David, P de Aquino, G Degrassi, D Del Re, A Denner, H van Deurzen, S Diglio, B Di Micco, R Di Nardo, S Dittmaier, M Dührssen, R K Ellis, G Ferrera, N Fidanza, M Flechl, D de Florian, S Forte, R Frederix, S Frixione, S Gangal, Y Gao, M V Garzelli, D Gillberg, P Govoni, M Grazzini, N Greiner, J Griffiths, A V Gritsan, C Grojean, D C Hall, C Hays, R Harlander, R Hernandez-Pinto, S Höche, J Huston, T Jubb, M Kadastik, S Kallweit, A Kardos, L Kashif, N Kauer, H Kim, R Klees, M Krämer, F Krauss, A Laureys, S Laurila, S Lehti, Q Li, S Liebler, X Liu, E Logan, G Luisoni, M Malberti, F Maltoni, K Mawatari, F Maierhoefer, H Mantler, S Martin, P Mastrolia, O Mattelaer, J Mazzitelli, B Mellado, K Melnikov, P Meridiani, D J Miller, E Mirabella, S O Moch, P Monni, N Moretti, A Mück, M Mühlleitner, P Musella, P Nason, C Neu, M Neubert, C Oleari, J Olsen, G Ossola, T Peraro, K Peters, F Petriello, G Piacquadio, C T Potter, S Pozzorini, K Prokofiev, I Puljak, M Rauch, D Rebuzzi, L Reina, R Rietkerk, A Rizzi, Y Rotstein-Habarnau, G P Salam, G Sborlini, F Schissler, M Schönherr, M Schulze, M Schumacher, F Siegert, P Slavich, J M Smillie, O Stål, J F von Soden-Fraunhofen, M Spira, I W Stewart, F J Tackmann, P T E Taylor, D Tommasini, J Thompson, R S Thorne, P Torrielli, F Tramontano, N V Tran, Z Trócsányi, M Ubiali, M Vazquez Acosta, T Vickey, A Vicini, W J Waalewijn, D Wackeroth, C Wagner, J R Walsh, J Wang, G Weiglein, A Whitbeck, C Williams, J Yu, G Zanderighi, M Zanetti, M Zaro, P M Zerwas, C Zhang, T J E Zirke, and S Zuberi. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group. Technical Report arXiv:1307.1347. CERN-2013-004, Geneva, 2013. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[46] BSM LHC Higgs Cross Section Working Group and Heavy Higgs Subgroup. BSM Higgs benchmarks in light of the discovery of a 126 GeV boson. *To appear*, 2012.

[47] Mike Lamont. Status of the lhc. *Journal of Physics: Conference Series*, 455(1):012001, 2013.

[48] *The CMS magnet project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[49] P. Adzic, R. Alemany-Fernandez, C.B. Almeida, N.M. Almeida, G. Anagnostou, et al. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007.

[50] S Chatrchyan et al. Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data. *JINST*, 5:T03012, 2010.

[51] D. Spiga, S. Lacaprara, W. Bacchi, M. Cinquilli, G. Codispoti, M. Corvo, A. Dorigo, A. Fanfani, F. Fanzago, F. Farina, O. Gutsche, C. Kavka, M. Merlo, and L. Servoli. Crab: the {CMS} distributed analysis tool development and design. *Nuclear Physics B - Proceedings Supplements*, 177–178(0):267 – 268, 2008. Proceedings of the Hadron Collider Physics Symposium 2007 Proceedings of the Hadron Collider Physics Symposium 2007.

[52] T. Miao, H. Wenzel, F. Yumiceva, and N. Leioatts. Beam position determination using tracks. 2007.

[53] W. Adam, B. Mangano, T. Speer, and T. Todorov. Track reconstruction in the CMS tracker. 2005.

[54] Application of vertex and mass constraints in track-based alignment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 2013.

[55] Susanna Cucciarelli, Marcin Konecki, Danek Kotlinski, and Teddy Todorov. Track reconstruction, primary vertex finding and seed generation with the Pixel Detector. Technical Report CMS-NOTE-2006-026, CERN, Geneva, Jan 2006.

[56] A. Ostaptchouk, S. Hameed-Khan, Stefan Schael, Rolf Siedling, and Bruno Wittmer. The Alignment System of the CMS Tracker. 2001.

[57] Thomas Speer, Kirill Prokofiev, R Frühwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Vertex Fitting in the CMS Tracker. Technical Report CMS-NOTE-2006-032, CERN, Geneva, Feb 2006.

[58] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the cms tracker. *Journal of Instrumentation*, 9(10):P10009, 2014.

[59] Stephanie Baffioni, Claude Charlot, Federico Ferri, David Futyan, Paolo Meridiani, Ivica Puljak, Chiara Rovelli, Roberto Salerno, and Yves Sirois. Electron reconstruction in CMS. Technical Report CMS-NOTE-2006-040, CERN, Geneva, Feb 2006.

[60] W. Adam et al. Performance studies of the CMS Strip Tracker before installation. *JINST*, 4:P06009, 2009.

[61] Wolfgang Adam, R Frühwirth, Are Strandlie, and T Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. Technical Report CMS-NOTE-2005-001, CERN, Geneva, Jan 2005.

[62] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. 2009.

[63] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett.*, B659:119–126, 2008.

[64] CMS Collaboration. Electron reconstruction and selection studies with first cms 7 tev data. *CMS Physics Analysis Summary*, CMS-PAS-EGM-10-004, 2010.

[65] Multivariate electron identification. *CMS twiki*, Electron MVA twiki, 2011.

[66] M. Cacciari, G.P. Salam and G. Soyez. The anti-kt jet clustering algorithm. *JHEP*, 04:063, 2008.

[67] Particle-flow event reconstruction in cms and performance for jets, taus, and met. *CMS PAS*, PFT-2009-001, Apr 2009.

[68] CMS Collaboration. Jet performance in pp collisions at $\sqrt{s}$=7 TeV. *CMS Physics Analysis Summary*, CMS-PAS-JME-10-003, 2010.

[69] Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 tev. *CMS PAS*, PFT-2010-002, 2010.

[70] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012.

[71] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett. B*, 716:30, 2012.

[72] Serguei Chatrchyan et al. Observation of a new boson with mass near 125 gev in pp collisions at $\sqrt{s} = 7$ and 8 tev. Submitted to JHEP, 2013.

[73] Ian Low, Joseph Lykken, and Gabe Shaughnessy. Singlet scalars as higgs imposters at the large hadron collider. *Phys. Rev. D*, 84:035027, 2011.

[74] Ian Low, Joseph Lykken, and Gabe Shaughnessy. Have we observed the higgs (imposter)? *Phys. Rev. D*, 86:093012, 2012.

[75] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-higgs-doublet models. *Phys. Rept.*, 516:1, 2012.

[76] B. Dobresceu and J. Lykken. 2009.

[77] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011.

[78] Lhc higgs cross sections twiki. *CMS twiki*, LHC Higgs Cross Sections twiki, 2012.

[79] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[80] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[81] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[82] Paolo Nason and Carlo Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *JHEP*, 02:037, 2010.

[83] T. Sjøstrand, S. Mrenna, and P.Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05 422, 2006.

[84] Nikolas Kauer and Giampiero Passarino. Inadequacy of zero-width approximation for a light Higgs boson signal. *JHEP*, 1208:116, 2012.

[85] Giampiero Passarino. Higgs Interference Effects in $gg \rightarrow ZZ$. *JHEP*, 1208:146, 2012.

[86] Stefano Goria, Giampiero Passarino, and Dario Rosco. The Higgs Boson Lineshape. *Nucl.Phys.*, B864:530–579, 2012.

[87] John M. Campbell, R.Keith Ellis, and Ciaran Williams. Gluon-Gluon Contributions to W+ W- Production and Higgs Interference Effects. *JHEP*, 1110:005, 2011.

[88] Search for a Standard Model-like Higgs boson decaying into WW to l nu qqbar in pp collisions at sqrt s = 8 TeV. Technical Report CMS-PAS-HIG-13-008, CERN, Geneva, 2013.

[89] Performance of cms muon identification in pp collisions at $\sqrt{s} = 7$ tev. *CMS PAS*, MUO-2010-002, 2010.

[90] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl.Instrum.Meth.*, A389:81–86, 1997.

[91] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, and Helge Voss. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[92] Yanyan Gao, Andrei V. Gritsan, Zijin Guo, Kirill Melnikov, Markus Schulze, et al. Spin determination of single-produced resonances at hadron colliders. *Phys.Rev.*, D81:075022, 2010.

[93] Bogdan A. Dobrescu and Joseph D. Lykken. Semileptonic decays of the standard Higgs boson. *JHEP*, 1004:083, 2010.

[94] Search for the standard model higgs boson in the decay channel h to zz to llqq at cms. *CMS PAS*, HIG-11-017, 2011.

[95] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 1107:018, 2011.

[96] Nikolaos Kidonakis. Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution. *Phys. Rev.*, D82:114030, 2010.

[97] Missing et uncertainty prescription. *CMS twiki*, MET twiki, 2011.

[98] M. R. Whalley, D. Bourilkov, and R. C. Group. The Les Houches Accord PDFs (LHAPDF) and Lhaglue. 2005.

[99] CMS Collaboration. Cms luminosity based on pixel cluster counting - summer 2013 update. CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013.

[100] ATLAS and CMS Collaborations. Procedure for the lhc higgs boson search combination in summer 2011. *ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005*, 2011.

[101] Roostats wiki. *Twiki*, RooStats Wiki, 2011.

[102] Cls - a frequentist interpretation. *CERN Indico*, ATLAS Statistics Forum, 2010.

[103] W. Perkins, M. Tygert, and R. Ward. Computing the confidence levels for a root-mean-square test of goodness-of-fit, II. *ArXiv e-prints*, September 2010.

[104] CMS Collaboration. Search for the standard model higgs boson in the h to ww to lvjj decay channel. *CMS PAS*, HIG-12-003, 2012.

[105] Vardan Khachatryan et al. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. 2014.

[106] Vardan Khachatryan et al. Study of vector boson scattering and search for new physics in events with two same-sign leptons and two jets. *Phys.Rev.Lett.*, 114(5):051801, 2015.

[107] J Fischer. Measurement of the electroweak production cross section of dijets associated with vector bosons. Technical Report ATL-PHYS-PROC-2014-075, CERN, Geneva, Jul 2014.