THESIS ENTITLED

Search for the Standard Model Higgs Boson in the H→ZZ→ ee(mu mu) nu nu channel in the CMS Experiment at LHC

SUBMITTED TO
DEPARTMENT OF PHYSICS AND ASTROPHYSICS
UNIVERSITY OF DELHI, DELHI-110 007
INDIA

FOR THE DEGREE OF PHILOSOPHY

BY

ARUN KUMAR

UNDER THE SUPERVISION OF

Dr. Kirti Ranjan

April 2014

Declaration

This thesis presents a Search for the SM Higgs Boson in the H→ZZ→ ee(mu mu) nu nu channel in the CMS Experiment at 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.

(ARUN KUMAR)

Candidate

(Dr. KIRTI RANJAN)

Supervisor

Prof. AMITABHA MUKHERJEE

Head

Department of Physics & Astrophysics

University of Delhi

Delhi 110 007

Dedicated to

My Loving Grand Parents and Parents

Acknowledgments

As I stand on the threshold of achieving an academic milestone, I would like to sincerely recall the contribution made by a number of people in building the foundation of my academic career and then helping me in developing the same till this stage.

I would like to start with the person who made the biggest difference in my life, my supervisor, Dr. Kirti Ranjan for his supervision, steady guidance and invaluable inspiration from the beginning of my Ph.D. He introduced me to the research of particle physics and helped me in growing research aptitude over the years. I am grateful to him for providing me the opportunity to work in such a stimulating environment such as the CMS collaboration and CERN. Many visits to CERN over the years helped me in participating in the cutting edge research in particle physics and also in growing my personality and confidence while collaborating internationally. His enthusiasm, curiosity and suggestions in high energy research is truly inspirational. I acknowledge him for his generosity and approachability in every step of my research. He has been a living role model to me, taking up new challenges every day, tackling them with all his grit and determination and always thriving to come out victorious.

I would like to thank Prof. R. K. Shivpuri for always being so kind, helpful and motivating. His constant guidance, cooperation and support has always kept me going ahead. I owe a lot of gratitude to him and I feel privileged to be associated with a person like him during my life.

My sincere thanks to Dr. Ashok Kumar for being there through out. His scientific inputs, constant motivation, personal helps and friendly nature has always made me feel at ease with him and I could always look back on him for any support during my course of Ph.D. I am also deeply thankful to Dr. Ashutosh Bhardwaj for useful suggestions throughout my Ph.D.

Many thanks to my Ph.D. advisory committee, Prof. D. S. Kulshrestha and Dr. Md. Naimuddin for giving me valuable suggestions. I am grateful to Prof. Amitabha Mukherjee, Head, Physics Department for providing me adequate facilities to work in the department.

I owe sincere and earnest thankfulness to Pedro Silva and Loic Quertenmont, without whose advice and suggestions this work would not have been possible. I am extremely grateful for the help, both fundamental and technical, that I received from them. I am grateful to Nicola de Filippis who helped me a lot, about software and analysis aspects, during my visit to Bari, Italy in the very beginning of my research work.

I would like to acknowledge all the teachers I learnt from since my childhood, I would not have been here without their guidance, blessing and support. My special thanks to my school teacher, Mr. Dinesh Bhardwaj, for his guidance, support and for encouraging me to continue in science after my schooling.

My acknowledgement will never be complete without the special mention of my seniors and fellow colleagues at our research lab (CDRST Lab) at University of Delhi consisting of Ruchika N. Sharma, Arun Saini, Sudha Ahuja, Shilpi Jain, Sandhya Jain, Pooja Saxena, Shivali Malhotra, Ajay Kumar, Varun Sharma, Daljeet Kaur, Rocky Bala Garg, Raman Khurana, Bhawna Gomber, Kavita Lalvani, Sumit Kesri, Geetika Jain, Priyanka Phogat, Preeti, Ankit Gaur, Aashiq Hussain, Purnendu Kumar, Mohd. Hasibuddin and Swati Mishra for helping me out with their suggestions and creating a friendly working environment in the lab. I would like to specially thank Ranjeet Dalal, Vijayender Narwal and Ramkrishna Sharma for being friends before colleagues. I could not have asked for more than what I got from them. I would like to thank Anil Singh for helping me in understanding the CMS software. I also thank my colleagues and friends from Panjab University for providing the useful discussions and for making my stay enjoyable at CERN, Geneva. My thanks to the staff of CDRST lab, Sanjay Singh, Aamir and Puran Singh for their innumerable help. A special word of gratefulness to Mohd. Yunus who has helped me with several official and personal matters.

I acknowledge the Council of Scientific and Industrial Research (CSIR), Government of India for providing me with the necessary funding and fellowship to pursue research. I am thankful to the administrators of Compact Muon Solenoid experiment (CMS) at CERN, Geneva for inviting me and providing the excellent atmosphere for the research work. I would also like to thank INDIA-CMS collaboration for providing a platform to discuss our work by arranging regular collaboration meetings. Also, my sincere thanks to Department of Science and Technology, Govt. of India, for providing the financial support for my visits to CERN.

I would like to acknowledge my friends for their moral support and motivation, which drives me to give my best. My special gratitude to Ravi, Santhust, Tarun, Himanshu and Vibhore for their wonderful company and the shared laughters. I met some wonderful people in my Masters degree, Piyush, Vikram Ravi, Amit Jawla, Alokesh and Samoday who went on to become my close friends. I thank them for their encouragement and for all those unforgettable moments shared together during those years.

I am very grateful to Mr. Hariom Jamdagni for inspiring me through out my academic life and for motivating me to do higher studies. I never forget his kind and encouraging words for me whenever I feel down.

I would like to specially mention my best friend and now my fiancee, Nishu Naib. She has been the pillar of support for me, emotionally and otherwise. She is the axis of my life. She taught me to remain calm and have patience. I thank her from depth of my heart for bearing with me in the crucial times when I may be frustrating and annoying.

Apart from the physics discussions we had, we enjoyed a lot during our CERN visits and being with you is the best feeling ever. I am deeply grateful for your invaluable help in writing this thesis from the very beginning till the end. I consider myself very fortunate to have a life partner like you.

Finally, I would like to acknowledge the people who mean the world to me, my father Mr. Surender Kumar and my mother Mrs. Sushila. I extend my respect to my parents, my paternal and maternal grand parents and all elders to me in the family. I don't imagine a life without their love and blessings. Thank you maa and papa, for showing faith in me and giving me liberty to choose what I desired. You sacrificed a lot of comfort for studies of your children. No one can believe that I have reached at this level of study by still staying in a village. This is possible only because of your ever increasing care, love and belief. I have always been a hero in your eyes. I like to mention my younger brother, Akhil and my younger sister, Akanksha with whom I spent my childhood and shared magical moments in all these years. I can always feel their love, care and respect for me which encourages me to work hard and do my best. I consider myself the luckiest in the world to have such a supportive family, standing behind me with their love and support.

I thank the Almighty for giving me the strength and patience to work through all these years. I could not have done anything without His blessings.

List of Publications

Publication in Peer Reviewed Journals

Publications as Primary Author

(i) "Search for the standard-model-like Higgs boson with a mass in the range 145 to 1000 GeV at the LHC", **Arun Kumar**, K. Ranjan et al, **European. Physics Journal C, 73 (2013).2469**, CMS-HIG-12-034, CERN-PH-EP-2013-050. Published 14^{th}June 2013: **European. Physics Journal C, 73 (2013).2469**, CMS-HIG-12-034, CERN-PH-EP-2013-050.
Corresponding to CMS AN-2012/138
Primary Authors: **Arun Kumar**, Kirti Ranjan et al.

(ii) "Search for the standard model Higgs boson in the $H \to ZZ \to 2\ell 2\nu$ channel"
Published 13^{th}March 2012: **Journal of High Energy Physics 03 (2012).040**, CMS-HIG-11-026, CERN-PH-EP-2012-033.
Corresponding to CMS AN-2011/407
Primary Authors: **Arun Kumar**, Kirti Ranjan et al.

Total Number of Publications as of $30^{th} April$ 2014 **:** 267

International conferences

Oral Presentations

- 46th Recontres de Moriond, La Thuile, Italy, March 2014, **Young Scientist Talk** om 'Search for a heavy Higgs boson in $H \to ZZ \to 2\ell 2\nu$ channel in pp collisions with CMS detector at the LHC'.
- XXXII Physics In Collision (PIC), Strbske Pleso, Slovakia, September 2012, 'Search for the Standard Model Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ decay channel in the CMS Experiment'.

Poster Presentations

- Hadron Collider Physics (HCP) Symposium, Paris, France, November 2011, Poster Presentation on the 'Search for Standard Model Higgs boson using the $H \to ZZ \to \ell^+\ell^-\nu\nu$ channel in pp collisions at CMS'.

- Lepton Photon 2011, Mumbai, India, August 2011, Poster Presentation on the 'Search for Higgs boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ channel at CMS'.

Proceedings

- 'Search for a heavy Higgs boson in $H \to ZZ \to 2\ell 2\nu$ channel in pp collisions with CMS detector at the LHC', Arun Kumar, Recontres de Moriond 2014, (to be published).

- 'Search for the Standard Model Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ decay channel in the CMS Experiment', Arun Kumar, K. Ranjan, published in the proceedings of the Physics In Collision 2012 Conference.
  CMS CR-2012/314, arXiv:1212.0288.

- 'Search for Standard Model Higgs boson using the $H \to ZZ \to \ell^+\ell^-\nu\nu$ channel in pp collisions at CMS', Arun Kumar, K. Ranjan, published in the proceedings of the Hadron Collider Physics 2011 Conference.
  CMS CR-2012/019, EPJ Web of Conferences Volume 28, 12065(2012).

- 'Search for Higgs boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ channel at CMS', Arun Kumar, K. Ranjan, published in tthe proceedings of the Lepton Photon 2011 Conference.
  CMS CR-2011/322, Pramana Journal of Physics Vol. 79, (2012) pg. 821-824.

National conferences

- Oral Presentation in XX DAE-BRNS High Energy Physics Symposium, Visva-Bharati, Santiniketan, West Bangal (India), on the 'Search for the Standard Model Higgs boson in the $H \to ZZ \to 2\ell 2\nu$ decay mode in the CMS Experiment'.

CMS Public Documents

(i) "Constraints on the Higgs boson width from off-shell production and decay to $ZZ \to \ell\ell\ell'\ell'$ and $\ell\ell\nu\nu$", **Arun Kumar**, Kirti Ranjan et al., **CMS PAS HIG-14-002 (2014)**.

(ii) "Search for a heavy Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ channel in pp collisions at $\sqrt{s} = 7$ TeV and 8TeV", **Arun Kumar**, Kirti Ranjan et al., **CMS PAS HIG-13-014 (2013)**.

(iii) "Search for the Standard Model Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ channel in pp collisions at $\sqrt{s} = 7$ TeV and 8TeV", **Arun Kumar**, Kirti Ranjan et al., **CMS PAS HIG-12-023 (2012)**.

(iv) "Search for the Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ channel in pp collisions at $\sqrt{s} = 7$ TeV", **Arun Kumar**, Kirti Ranjan et al., **CMS PAS HIG-11-026 (2011)**.

(v) "Search for the Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ channel in pp collisions at $\sqrt{s} = 7$ TeV", **Arun Kumar**, Kirti Ranjan et al., **CMS PAS HIG -11-016 (2011)**.

(vi) "Search for the Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ channel in pp collisions at $\sqrt{s} = 7$ TeV", **Arun Kumar**, Kirti Ranjan et al., **CMS PAS HIG -11-005 (2011)**.

CMS Internal Documents

(i) " Bounding the Higgs width using $ZZ \to \ell\ell$ events with high missing transverse energy", **Arun Kumar**, Kirti Ranjan et al., **CMS-AN-2013/411**.

(ii) "Search for a high-mass Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ decay channel in pp collisions with in CMS detector", **Arun Kumar**, Kirti Ranjan et al., **CMS-AN-2012/371**.

(iii) "Search for the Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ decay channel in pp collisions with the CMS detector", **Arun Kumar**, Kirti Ranjan et al., **CMS-AN-2012/138**.

(iv) "Search for the Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ decay channel in pp collisions at $\sqrt{s} = 7$ TeV at CMS", **Arun Kumar**, Kirti Ranjan et al., **CMS-AN-2011/407**.

(v) "Search for the Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu\nu$ decay channel in pp collisions at $\sqrt{s} = 7$ TeV at CMS", **Arun Kumar**, Kirti Ranjan et al., **CMS-AN-2011/119**.

(vi) "Higgs Trigger Strategy Note", **Arun Kumar** et al., **CMS-AN-2011/065**.

ABSTRACT

The Standard Model (SM) is a framework which describes the fundamental particles and their interactions at the most fundamental level. The Large Hadron Collider (LHC) is a superconducting proton-proton collider designed to test the SM at high energies, to search for the SM Higgs boson (H) and also to search for any possible physics beyond the SM.
In this Thesis, a search has been presented for the SM Higgs boson in $H \rightarrow ZZ \rightarrow \ell^{+}\ell^{-}\nu_{\ell}\bar{\nu}_{\ell}$ (ℓ = electron or muon) decay mode using the data collected by the Compact Muon Solenoid experiment from proton-proton collisions in the LHC at European Center for Nuclear Research (CERN), Geneva, Switzerland. The analysed data is collected over two years, 2011 and 2012, which correspond to an integrated luminosity of 5.0 fb^{-1}at center of mass energy of 7 TeV and 19.7 fb^{-1}at center of mass energy of 8 TeV, i.e. the full data recorded by the CMS experiment till the LHC stopped for its first long shutdown in 2013. This decay mode is mostly sensitive in high Higgs mass (M_H) range from 200 GeV to 1000 GeV. Two most dominant production modes of the Higgs boson, gluon-gluon fusion and vector boson fusion (VBF), are considered. Main backgrounds to this analysis come from $Z + jets$, $t\bar{t} + jets$, WW, WZ and ZZ. Selection criteria is based on the properties of the signal final state, which is high transverse missing energy (E_T^{miss}) arising from two undetected neutrinos and two same flavor leptons (electrons or muons) consistent with Z boson decay. The selection procedure is divided into preselection, event categorization and final selection. Preselection includes the kinematical selection of leptons and topological selection based on minimum azimuthal separation of jets and E_T^{miss}. To reduce top quark related backgrounds, events having jets arising from b-quarks are rejected. To enhance the sensitivity of the search, the events are categorized on the basis of jet multiplicity. A separate selection is applied to identify VBF type events and then rest of the events are divided into exclusive 0-jet and inclusive 1-jet ($\geq$ 1-jet) categories. Finally, Higgs mass dependent selection is applied based on E_T^{miss} and transverse mass of the Higgs (M_{T}). All the major backgrounds like $Z + jets$, $t\bar{t} + jets$ and WW are estimated using data-driven techniques. All sources of the systematic uncertainties are also considered.
No significant excess is observed over the SM background prediction over the entire Higgs mass range. A statistical analysis is carried out to set upper limits on the Higgs boson production cross-section. The baseline approach is the shape analysis based on final discriminant (M_{T} for non-VBF category and E_T^{miss} for VBF category) in which the full distribution of the final discriminant is used to impose limits. A cross-check has also been made by imposing a selection requirement on final discriminant (instead of using the full distribution), commonly referred to as the cut-and-count analysis.
In order to explore the possible improvements in the analysis sensitivity and to cross-check the baseline analysis using an independent technique, a multivariate analysis is introduced

based on the Boosted Decision Tree (BDT) technique. Training of BDT is performed for each Higgs mass hypothesis and is done separately for $e^+e^-\nu_e\bar{\nu_e}$ and $\mu^+\mu^-\nu_\mu\bar{\nu_\mu}$ decay modes. The optimization of BDT parameters is done keeping the best expected limit on ratio of measured Higgs boson production cross-section to the SM cross-section. Background estimation techniques are same as used in baseline analysis and all systematic sources are also considered. A similar statistical analysis is carried out and exclusion limits are imposed on M_H.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The notion that a fundamental simplicity lies beneath the observed diversity of the universe has carried physics far. Mankind has pursued this quest from time immemorial that what are the ultimate constituents of matter and what holds them together. Elementary particle physics tries to answer these questions at the very fundamental level. The era of particle physics started with the discovery of electron by J.J. Thomson in 1897. Thomson showed that electron was about 2,000 times lighter than hydrogen ion, which was the lightest thing around. In 1911, the amazing results of the existence of atomic nucleus came up from the α-particle scattering experiment by Ernest Rutherford. In 1913, Niels Bohr came up with atomic model, where electrons move about the nucleus in circular orbits, separated from each other like rungs in a ladder. Historically, the list of particles and forces considered to be elementary has changed continually as closer scrutiny of matter and its interactions revealed microcosms within microcosms: atoms within molecules, nuclei and electrons within atoms, and successively deeper levels of structure within the nucleus. With the rise of high energy experiments, modern science has helped us to understand several phenomena but has also posed various questions. What is the origin of mass? Why are some particles light but some massive? What happened at the time of the Big Bang, and where did all the anti-matter go?, to name a few. To answer these basic questions, by the 1970s, a model was emerged capturing all the theoretical and experimental information available in particle physics, known as the *'Standard Model'* (SM) [1–3]. The SM is a theoretical framework which describes the fundamental particles and the interactions that act between them. It has withstood close to four decades of experimental tests and has provided a very good description of all currently observed phenomena. Predictions associated with the SM have been confirmed by many experiments to a spectacular degree of precision [4]. It is used as a basis for building more exotic models which tries to find solutions to other problems like dark matter particles, hierarchy and unification of strong and electroweak forces etc. The SM is essentially a gauge theory and the notion of symmetry is central to gauge theories.

The development of particle physics is closely associated with the development in technology for accelerating particles to higher and higher energies and detecting them with more and more sophisticated detectors. To probe deeper into matter one needs high energy probes according to the de Broglie's wave-particle duality relation $\lambda = h/p$, where, h = planck's constant, p is momentum of particle and λ is the wavelength of the associated wave with particle. Thus, if p is higher then λ is smaller and since resolution is directly proportional to wavelength so one can probe more with shorter wavelength. Another motivation to go for higher energy comes from Einstein's famous $E = mc^2$ equation. This has been used as a basis for creation of new particles, i.e. if particles are accelerated to very high energies, making the head on collisions with each other, the energy of their motion can transform and result into production of new particles which do not exist in nature in ordinary conditions. The modern accelerators called '*Synchrotron*' takes particle bunches at a time, and accelerates these bunches. The first major synchrotron was built at Brookhaven in the USA and CERN in Geneva, with energies up to 30 GeV by 1960. With the idea of quarks came in 1960s, more effort was put to reach energies up to 100 GeV. Improvements in technology led to more powerful magnets, and by placing them in an almost circular ring with a diameter of over a kilometre, by the middle of the 1970s, Fermilab (near Chicago in USA) and CERN (near Swiss-French border in Geneva, Switzerland) had achieved proton energies of some 500 GeV. By 1982, Fermilab had achieved 1,000 GeV, or '1 TeV', and became known as the '$Tevatron$'. Today superconducting magnets enable even more powerful magnetic fields to be achieved. The latest of them is Large Hadron Collider (LHC), at CERN, Geneva.

With the increase in energy and rate of particles there was a need to develop the detection system to detect the scattered particles. There has been a significant technological advancement in the particle detectors, starting from cloud chambers to bubble chambers to the usage of silicon detectors. Modern day detectors collect various individual pieces of detectors and combine them in cylindrical detectors that surround the interaction point at a collider. The collision happens on the central axis of the detector. As many particles streams out, it encounters a series of different pieces of detector, each with its own speciality in recognizing particles.

The Large Hadron Collider (LHC) has been designed and built with the purpose of exploring energies which have never been probed before this in a laboratory. It aims to search for the Higgs boson and for new physics beyond the SM, and also to provide precision measurements of processes predicted by the SM. The LHC is a superconducting proton-proton collider, designed to produce collisions at a centre of mass energy up to 14 TeV. The LHC recorded the first proton-proton collisions at a center of mass energy of 7 TeV on March 30th, 2010 and it continued to deliver data at 7 TeV in 2011, which correspond to integrated luminosity of 6.13 fb^{-1}. In 2012, the center of mass energy was increased to 8 TeV and the total delivered luminosity by the LHC was 23.3 fb^{-1}. Along with the proton-proton collisions, the LHC also collides lead (Pb) ions for searches related to quark-gluon plasma. The "Higgs hunting" has been one of the main objectives of the physics program at the LHC. The collisions produced are recorded by four major detectors, A Toroidal LHC ApparatuS (ATLAS), Compact Muon Solenoid (CMS), A Large Ion Collider Experiment (ALICE) and Large Hadron Collider beauty (LHCb). The Higgs boson search is the central part of the physics programs of the two general purpose experiments, CMS and ATLAS. On 4^{th} July 2012, the first observation of a Higgs boson like particle has been announced by the two experiments CMS and ATLAS, with a mass of about 125 GeV [5, 6]. After the discovery of this boson, the Nobel Prize in Physics 2013 was awarded jointly to Francois Englert and Peter Higgs for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles.

In this Thesis, a search for the SM Higgs boson is described in $H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ decay channel. This decay channel[1] is sensitive to higher Higgs boson masses (≥ 200 GeV).

In this chapter, a short overview of the SM (Section 1.1) and its representation as gauge theory (Section 1.2) is given with focus on the ElectroWeak Symmetry Breaking (EWSB), the Higgs mechanism and the Higgs boson properties (Section 1.3). In Sections 1.4 and 1.5, theoretical and experimental constraints on the Higgs boson mass are presented. In Sections 1.6-1.8, a brief description is given for Higgs phenomenology at the LHC. Section 1.9 presents the organization of this Thesis.

1.1 The Standard Model

The fundamental constituents of matter and their interactions are accurately described by the Standard Model (SM) of Particle Physics. The SM passes all the precision tests conducted so far by many high energy physics experiments through out the years. The SM includes fermionic fundamental particles called '*matter particles*' and bosonic particles called '*vector bosons*' which are the carriers of fundamental interactions. Figure 1.1 lists all these particles of the SM along with their properties like charge, mass and spin. Certain patterns are observed among the known elementary particles that allow us to group them into three families or generations; each generation is made up of a quark doublet, a charged lepton and its associated neutrino. The only difference between three generations is that of "mass", i.e. as we go to higher generation the masses of particles increase with respect to their counterparts in lower generation.

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

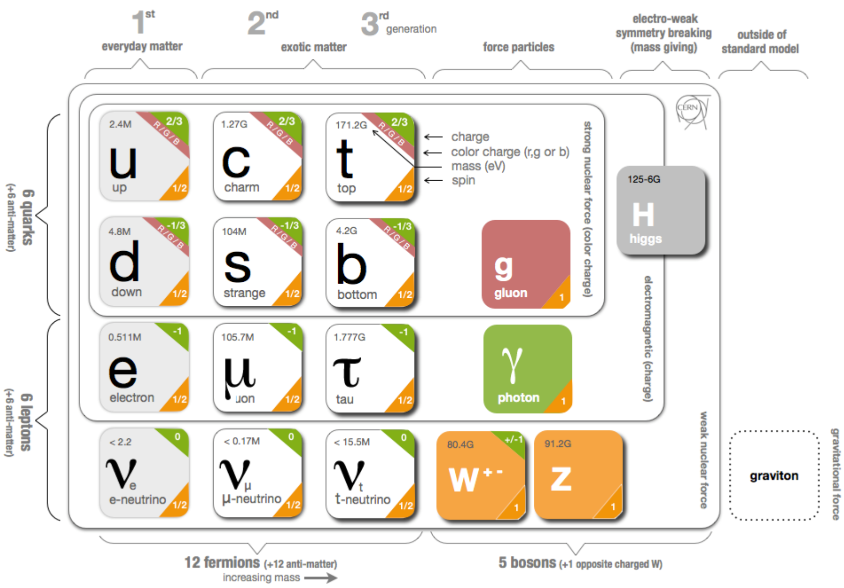


Figure 1.1: Schematic diagram of the particles in the SM and their associated symbols and groupings.

1.1.1 Matter Particles

All the matter particles are divided into *quarks* and *leptons*. Quarks carry fractional electric charge of +2/3 times charge of electron or -1/3 times charge of electron. Leptons carry either integral electric charge in units of electron charge or they do not carry any electric charge. Each quark and lepton has an associated anti-quark and anti-lepton with same mass, same spin but opposite electric charge (and other quantum numbers).

Quarks: As mentioned earlier, three generations of matter particles exist. Quarks are arranged in three doublets: *up* and *down*, (u, d), stable constituents of the atomic nuclei, with masses of a few MeV; *charm* and *strange*, (c, s), with masses of about 1.3 and 0.1 GeV; *top* and *bottom* (t, b), with masses of about 172 and 4.2 GeV [7]. Quarks carry a new kind of charge called 'color charge', on which the strong force acts. Each quark can carry one of the three kinds of color charge: red, green and blue. The color charge of quarks is completely unrelated to visual perception of color. Hadrons which are combinations of quarks or anti-quarks must be color neutral. Therefore, the only allowable combinations are those of quark and its antiquark, giving rise to mesons, and of a red, a green and a blue quark or anti-quarks with anti-colors yielding baryons or anti-baryons. Colored states are never seen in isolation.

Leptons: The three generations of leptons are the electron e, the muon μ and the tau τ along with their associated neutrinos ν_e, ν_μ and ν_τ. Masses of e, μ, and τ are about 0.5 MeV, 0.1 GeV and 1.8 GeV respectively. Mass limit on three neutrinos ν_e, ν_μ, ν_τ are; $<$ 2 eV, $<$ 0.19 MeV and $<$ 18.2 MeV [7] respectively. While only the electron is stable the muon decays to electron and neutrinos with a long lifetime about 2.2 μs. Because of the shorter tau lepton lifetime, about 2.9×10^{-13}s, τs are observed only through their decay products.

1.1.2 Fundamental Interactions

The description of the structure of matter can not be complete without considering the interactions (forces) that "join" the particles and more generally that regulate the interactions between them. Four fundamental interactions along with their mediators 'vector bosons' are discussed below:

Strong interactions are responsible for binding the quarks in the hadrons like neutron and proton. The inter-quark force is mediated by *gluon*. In total, there are 8 gluons in the SM. *Electromagnetic interactions* are responsible for almost all the extra-nuclear phenomena. These interactions bind the electrons with the nuclei to form atoms and molecules. These interactions are mediated by exchange of *photon*. They are long range forces. *Weak interactions* are typified by nuclear β-decay. These interactions are mediated by massive $W^{\pm}$ and Z^0 bosons. Since, the force carriers are massive so these are short range forces (fm) are active only within nuclear dimensions. *Gravitational interactions* act between all kind of particles. But since, in particle physics one deals with particles with very low masses (compared to macroscopic objects), gravitational interactions are too small and are not included in the SM. Leptons interact via electromagnetic and weak interactions, whereas quarks, along with these two interactions, also interact via strong interactions. Table 1.1

Force	**Mediating boson**	**Electric Charge**	**Spin**	**Mass (GeV)**	**Range**	**Relative Strength**
Strong	8 gluons(g)	0	1	0	∞	1
EM	Photon(γ)	0	1	0	∞	$\frac{1}{137}$
Weak	$W^{\pm}$	±1	1	80.403$\pm$0.029	10^{-18}m	10^{-5}
	Z^0	0	1	91.187$\pm$0.002		

Table 1.1: Fundamental interactions included in the SM and their properties.

shows a list of forces, various properties of the mediating bosons, range and their relative strength. Besides the ultimate fermionic constituents and vector bosons that mediate the fundamental interactions, the SM, via Higgs mechanism [8–10] predicts the existence of the *scalar Higgs boson*, with spin zero, which is necessary for the spontaneous symmetry breaking mechanism through which fermions and $W^{\pm}$ and Z bosons acquire mass. The SM Higgs boson also interacts with a fermion-antifermion pair, giving them masses. In total, there are 61 fundamental particles: 24 fermions + 24 anti-fermions + 12 vector bosons + 1 Higgs boson.

1.2 Standard Model as gauge theory

The term *gauge* refers to redundant degrees of freedom that exist in the Lagrangian. The transformations between possible gauges, called gauge transformations, form a group known as the *symmetry group* or *the gauge group* of the theory. These transformations can be *global* or *local* depending on the choice of *gauge*. Global transformations are continuous transformations and according to Noether's theorem there exists a conserved quantity for every continuous global symmetry.

A gauge theory is a type of field theory in which the Lagrangian is invariant under a continuous group of local transformations. In this process of local gauge transformations, a new field is introduced (the gauge field) which has its own dynamics and couples to the particles/fields which have the symmetry. Those particles are then said to be charged under the gauge field. On the same steps, the SM of particle physics is a gauge theory based on unified framework to describe electroweak (EW) interactions between quarks and leptons together with strong interactions between quarks. It is a Yang-Mills[2] theory based on the EW symmetry group $SU(2)_L \times U(1)_Y$[3] proposed by Glashow, Weinberg and Salam and $SU(3)_C$ group of QCD [11]. The Lagrangian $\mathcal{L}_{SM}$ of the SM can be divided into following different parts [12]:

$$\mathcal{L}_{SM} = \mathcal{L}_{YM} + \mathcal{L}_{ferm} + \mathcal{L}_H + \mathcal{L}_{Yuk} \tag{1.1}$$

The Yang-Mills part $\mathcal{L}_{YM}$ describes the dynamics of the gauge fields, i.e. their free propagation as well as 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

$$\begin{aligned}
W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g_2\epsilon^{ijk}W^j_\mu W^k_\nu, \qquad i,j,k = 1,2,3 \\
B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \\
G^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc}G^b_\mu G^c_\nu, \qquad a,b,c = 1,..,8
\end{aligned} \tag{1.3}$$

are the field strength tensors of the gauge fields W^i_μ for the $SU(2)_L$ group of weak isospin, B_μ for $U(1)_Y$ of weak hypercharge and G^a_μ for the $SU(3)_C$ of color. The respective gauge couplings are denoted as g_2, g_1 (described later) and g_s and the structure constants of the groups $SU(2)$ and $SU(3)$ are ϵ^{abc} and f^{abc} respectively.

The interaction of the gauge fields with the fermions is encoded in the term $\mathcal{L}_{ferm}$[4],

$$\mathcal{L}_{ferm} = \bar{\psi}_L iD_\mu\gamma^\mu\psi_L + \bar{\psi}_{\ell_R} iD_\mu\gamma^\mu\psi_{\ell_R} + \bar{\psi}_Q iD_\mu\gamma^\mu\psi_Q + \bar{\psi}_{u_R} iD_\mu\gamma^\mu\psi_{u_R} + \bar{\psi}_{d_R} iD_\mu\gamma^\mu\psi_{d_R} \tag{1.4}$$

where L = $(\nu_{\ell_L}, \ell_L)^T$ are left-handed $SU(2)$ doublets of charged leptons $\ell = e$, μ, τ and neutrinos $\nu_\ell = \nu_e$, ν_μ, ν_τ; Q = $(u_L, d_L)^T$ are the left-handed $SU(2)$ doublets of up-type

[2]YangMills theory is a gauge theory based on the $SU(N)$ group.
[3]where, L represents left-handedness, Y is weak hypercharge and C is color charge
[4]ψ and $\bar{\psi}$ are Dirac-fields (matter fields) and γ^μ are Dirac matrices for relativistic fermions.

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.

This part of the Lagrangian $\mathcal{L}_{YM} + \mathcal{L}_{ferm}$ is invariant under local $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge transformations for fermion and gauge fields. The matter fields ψ couple to the gauge fields through the covariant derivative

$$D_\mu \psi = (\partial_\mu - ig_s T_a G^a_\mu - ig_2 T_i W^i_\mu - ig_1 \frac{Y}{2} B_\mu)\psi \tag{1.5}$$

which leads to unique couplings between the fermion and gauge fields $-g_i \bar{\psi} V_\mu \gamma^\mu \psi$. Here, T_i, Y and T_a are the generators of the respective gauge groups in the representation of the fermions they act on, 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. The weak hyper charge Y is related to electric charge Q by Gell-Mann-Nishijima relation:

$$Q = T_3 + \frac{Y}{2} \tag{1.6}$$

The Lagrangian described so far does not include any mass term and hence the gauge and fermion fields have been kept massless. However, if one adds standard mass term for the gauge bosons, $M_W^2 W_\mu W^\mu$ to the Lagrangian, one immediately violates the local $SU(2) \times U(1)$ gauge invariance. Also, for fermions if one includes explicitly the mass term $-m_f \bar{\psi}_f \psi_f$ for the SM fermion f in the Lagrangian then, for instance, one would have for the electron

$$-m_e \bar{e} e = -m_e (\bar{e_R} e_L + \bar{e_L} e_R) \tag{1.7}$$

which is non-invariant under the weak isospin symmetry transformations, since e_L is a member of SU(2) doublet, while e_R is the SU(2) singlet and therefore they change under transformations in a different manner.

Therefore, the mass terms for gauge bosons and fermions induced "by-hand" lead to an obvious breakdown of the local $SU(2) \times U(1)$ gauge invariance. The unbroken symmetry means that all fundamental particles have to be massless. But we know experimentally that the weak gauge bosons are massive and also fermions have masses. So, the mass generation mechanism for an elementary particle in the SM is strongly connected with symmetry violation. The introduction of particle masses, while still maintaining the gauge invariance of the dynamics of the model, requires extensions of the theory.

1.3 Electroweak symmetry breaking and the Higgs mechanism

In the $SU(2) \times U(1)$ case of the SM one needs to generate masses for the three gauge bosons $W^{\pm}$ and Z but the photon should remain massless. Therefore, one needs at least three degrees of freedom for the scalar fields. So, the simplest choice is to use an iso-vector state with exactly three scalar fields, but in this case one lacks for massless field and it is impossible to generate all above mentioned masses in the SM. Hence, the Higgs mechanism is implemented in the SM by introducing a complex scalar field Φ, doublet of SU(2) with hypercharge $Y = 1/2$,

$$\Phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.8}$$

where, ϕ_i are 4 real scalar fields (4 degrees of freedom). The relevant scalar Lagrangian has the form

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

The self-interaction of the field Φ is given by potential V(Φ)

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

Where, μ is a parameter related to mass and λ is the self-coupling constant.

Both the product $\Phi^\dagger \Phi$ and the potential $V(\Phi)$ are invariant under the local gauge

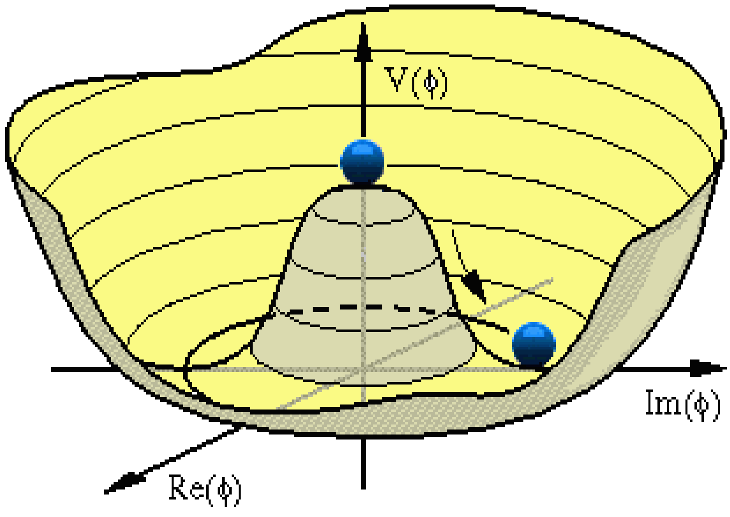


Figure 1.2: The Higgs potential $V(\Phi)$ for $\lambda > 0$ and $\mu^2 < 0$ condition.

transformations

$$\Phi(x) \to \Phi^{'}(x) = e^{i\alpha_i(x)\sigma_i/2}\Phi(x), \tag{1.11}$$

where σ_i are Pauli matrices and $\alpha_i(x)$ are transformation parameters.

When $\lambda > 0$ and $\mu^2 > 0$, one gets a unique minima. When $\lambda > 0$ and $\mu^2 < 0$ one obtains a continuous set of minima for the potential. The fields satisfy the condition:

$$\Phi^\dagger\Phi = \frac{1}{2}\left(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\right) = -\frac{\mu^2}{2}\lambda = v^2 \tag{1.12}$$

The shape of the potential for this case looks like a Mexican hat in the space described by various components of Φ shown in Figure 1.2. Choosing the ground state of the theory to be a particular Φ among the many satisfying the equation of the minima, and expanding the potential in the vicinity of the chosen minimum, transforms the Lagrangian in such a way that the original gauge symmetry is now hidden or spontaneously broken. We choose a minimum like: $\phi_1 = 0$, $\phi_2 = 0$, $\phi_4 = 0$, and ϕ_3^2 is given by, $\phi_3^2 = -\frac{\mu^2}{\lambda} = v^2$. We now expand $\Phi(x)$ about this particular vacuum:

$$\Phi_0 \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.13}$$

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

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

Moving to the unitary gauge by means of a proper gauge transformation of the field, one "gauges away" three θ_a fields, chooses only one direction, violates three global symmetries of the Lagrangian and leaves only one invariant.

$$\frac{1}{\sqrt{2}} e^{i\theta_a(x)\tau^a(x)/v} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.15}$$

1.3.1 Masses of the Gauge Bosons

In the Lagrangian the mass terms are given by representations like $M^2\Phi^\dagger\Phi$. By inserting the vacuum expectation value into the Lagrangian of Equation (1.9) one is able to study its structure. The covariant derivative for the EW part is defined as:

$$D_\mu \Phi = \left(\partial_\mu - ig_2 W^i_\mu \frac{T^i}{2} - ig_1 \frac{Y}{2} B_\mu \right) \Phi \tag{1.16}$$

Using Equation (1.15) and Equation (1.16) in Equation (1.9), the kinetic term $(D_\mu\Phi)^\dagger(D^\mu\Phi) \equiv |D_\mu\Phi|^2$ is given as:

$$\begin{aligned} |D_\mu\Phi|^2 &= \left| \left(\partial_\mu - ig_2 W^a_\mu \frac{\tau^a}{2} - ig_1 \frac{Y_H}{2} B_\mu \right) \Phi \right|^2 && (1.17) \\ &= \frac{1}{2}(\partial_\mu h)^2 + \frac{g_2^2}{8}(v+h)^2 (W_1^\mu + iW_2^\mu)(W^1_\mu - iW^2_\mu) + \frac{1}{8}(v+h)^2 (g_2 W_3^\mu - g_1 Y_H B^\mu)^2 \\ &= \frac{g_2^2 v^2}{8}(W_1^\mu + iW_2^\mu)(W^1_\mu - iW^2_\mu) + \frac{v^2}{8}(g_2 W_3^\mu - g_1 Y_H B^\mu)^2 + \frac{1}{2}(\partial_\mu h)^2 + ... && (1.18) \end{aligned}$$

The first term in Equation (1.18) is proportional to the mass term $M_W^2 W_\mu^+ W^{-\mu}$ for the charged gauge boson field

$$W^\pm = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu) \quad with \quad M_W = \frac{1}{2} v g_2 \tag{1.19}$$

In particular, the last relation allows one to fix the vacuum expectation value v in terms of the W boson mass M_W and coupling g_2. Also, it can be shown in terms of the Fermi constant ($G_F = 1.667 \times 10^{-5}\text{GeV}^{-2}$) as:

$$M_W = \frac{g_2 v}{2} = \left(\frac{\sqrt{2}g_2^2}{8G_F}\right)^2 \quad and \quad v = 2\frac{M_W}{g_2} = \frac{1}{(\sqrt{2}G_F)^{1/2}} \approx 246 \text{ GeV}. \tag{1.20}$$

The second term in Equation (1.18) mixes two neutral components of the gauge fields W_3^μ and B^μ in such a way that the physical states (mass eigenstates) are:

$$Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, \qquad A_\mu = \frac{g_2 W_\mu^3 + g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}} \tag{1.21}$$

This mixing can also be represented as:

$$\begin{aligned} A_\mu &= B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \qquad && massless \rightarrow \gamma \\ Z_\mu &= -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W \qquad && massive \rightarrow Z^o \end{aligned} \tag{1.22}$$

Here θ_W is the Weinberg or weak mixing angle. Its value is found experimentally as $\sin^2\theta_W = 0.231$ [7]. So, the mass term $\frac{1}{2}M_Z^2 Z_\mu^+ Z^{-\mu}$ can be interpreted as:

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

Using above expressions one can show that the mass of W and Z bosons are connected through the following relation:

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

No mass term appears for neutral field A_μ, which is orthogonal to field Z_μ and thus the photon field (A_μ) remains massless.

1.3.2 Fermion mass generation

The arrangement of 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 fields with fermions, being only the SU(2) doublets or singlets. The most general form of these interactions is:

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

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 EWSB 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). Consider the case of electron, i.e. the first term and its complex conjugate in Equation (1.25). With the Higgs field in the unitary gauge Equation (1.15), we get

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

The first term in Equation (1.26) looks exactly as a mass term for fermions $-m_f\bar{\psi}\psi$, with the electron mass (and in complete analogy for u-quark and d-quark)

$$m_e = \frac{\lambda_e v}{\sqrt{2}}, \quad m_u = \frac{\lambda_u v}{\sqrt{2}}, \quad m_d = \frac{\lambda_d v}{\sqrt{2}} \tag{1.27}$$

It is evident from Equation (1.27) that the mass of the fermion depends on the strength of interaction with the Higgs field.

1.4 Higgs Boson mass

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 (1.18)), while the Higgs mass and Higgs self-interaction parts, come from the scalar potential $V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$ which, after EWSB, takes the form [5]

$$V(h) = \frac{\mu^2}{2}(v+h)^2 + \frac{\lambda}{4}(v+h)^4 \tag{1.28}$$

[5] as defined earlier, λ is self coupling constant taking care of the self-interaction of the Higgs field.

Finally, with the relation $\mu^2 = -v^2\lambda$ the pure SM Higgs Lagrangian is given by

$$\mathcal{L}_H = \frac{1}{2}(\partial^\mu h)^2 - \frac{2\lambda v^2}{2}h^2 - \lambda v h^3 - \frac{\lambda}{4}h^4 + \frac{\lambda v^4}{4} \tag{1.29}$$

From this equation we can see that

$$M_H^2 = 2\lambda v^2 \tag{1.30}$$

The strength of the Higgs self-interaction is proportional to the square of the Higgs mass (M_H).

$$g_{h_3} = \lambda v = \frac{M_H^2}{2v}, \quad g_{h_4} = \frac{\lambda}{4} = \frac{M_H^2}{8v^2} \tag{1.31}$$

The mass of the Higgs boson is generated by the Higgs self-interaction and is defined by the parameter λ. There is no clear idea within the SM concerning the source of λ, and its value stays, in principle, undefined (and hence also the Higgs boson mass). However, there are constraints on the Higgs mass from theory and experimental searches which are described in the next section.

1.4.1 Theoretical Constraints on Higgs Mass

The mass of the Higgs boson is not a prediction of the SM, but rather a free parameter of the theory. It depends on v and on λ. The former has already been calculated in terms of the Fermi coupling constant but the latter is characteristic of the scalar Φ-field and cannot be determined other than measuring the Higgs mass itself. However, both theoretical and experimental constraints exist, including those from direct searches at colliders, in particular at Large Electron-Positron (LEP) and Tevatron colliders. The recent results on Higgs mass which are given by the LHC, after the first observation of Higgs boson at the LHC, are given in Section 1.8. In this Section the most important theoretical constraints that are imposed on the mass of the SM Higgs boson by the consistency of the theory up to a given energy scale Λ (new physics scale) are explained. These constraints are: unitarity, triviality and vacuum stability. More details about the theoretical constraints can be found in Ref. [13].

Unitarity Constraint

The scattering amplitudes for longitudinal gauge bosons grow as the square of the Higgs boson mass ($V_L V_L \rightarrow V_L V_L$) where V is W/Z boson. This can be calculated using the EW equivalence theorem [14], 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 a partial wave decomposition, one can write the Scattering Amplitude $\mathcal{A}$ as:

$$\mathcal{A} = 16\pi \sum_{l=0}^{\infty} (2l+1) P_l(cos\theta) a_l \tag{1.32}$$

where a_l is the partial wave amplitude for orbital angular momentum l partial wave and $P_l(cos\theta)$ are the Legendre polynomials. In terms of a_l, one can write the scattering cross-section (σ) as:

$$\sigma = \frac{1}{s} Im\left[\mathcal{A}(\theta = 0)\right] = \frac{16\pi}{s} \sum_{l=0}^{\infty} (2l+1)|a_l|^2 \tag{1.33}$$

Using the optical theorem and applying the unitarity requirements, we can impose that:

$$|a_l|^2 = Re(a_l)^2 + Im(a_l)^2 = Im(a_l) \rightarrow |Re(a_l)| \leq \frac{1}{2} \tag{1.34}$$

$$M_H < 870 \ \ \text{GeV}. \tag{1.35}$$

By considering coupled channels like $W_L^+ W_L^- \rightarrow Z_L Z_L$ the bound can be lowered to

$$M_H < 710 \ \ \text{GeV}. \tag{1.36}$$

This means that for heavier Higgs boson masses, perturbation theory is not valid. In this case the Higgs boson is either not heavier or another mechanism has to regulate the $W_L^+ W_L^-$ scattering cross-section. By studying the case that $s << M_H^2$ one obtains the critical scale $\sqrt{s_c}$ at which perturbative unitarity has to be violated:

$$\sqrt{s_c} < 1.8 \text{ TeV}. \tag{1.37}$$

And using more constraining channels this bound can be reduced to

$$\sqrt{s_c} < 1.2 \text{ TeV}. \tag{1.38}$$

Here one again observes a strong indication for new physics at the TeV scale.

Triviality and Vacuum Stability

The dependence of the quartic coupling λ on the energy scale Q is regulated by the renormalization group equation:

$$\frac{d\lambda(Q)}{dQ^2} = \frac{3}{4\pi^2}\lambda^2(Q) \tag{1.39}$$

This equation states that the λ decreases for small energies and vice versa. Hence the coupling vanishes in the low energy regime and the theory becomes trivial (non-interactive). In the high energy regime, on the other hand, 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 made trivial. In the SM, the running of λ is governed by more interactions. λ increases with increasing M_H as:

$$\lambda = \frac{M_H^2}{2v^2} \tag{1.40}$$

In the case when M_H becomes large, the equation for running of $\lambda(Q)$ can be written as:

$$\frac{d\lambda}{dt} = \frac{3\lambda^2}{4\pi^2} \tag{1.41}$$

where $t = \ln(Q^2/Q_0^2)$ and Q_0 is some reference energy scale. By solving Equation(1.41) one sees that $\lambda(Q)$ 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 $\lambda(Q) > 0$. When requiring the quartic coupling to be finite at high scales of Λ:

$$\frac{1}{\lambda(\Lambda)} > 0 \tag{1.42}$$

This can be interpreted as a bound on the Higgs boson mass.

$$M_H^2 = \frac{8\pi^2 v^2}{3\log\left(\Lambda^2/v^2\right)} \tag{1.43}$$

The behaviour of M_H and Λ is shown in Figure 1.3. By requiring 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 [15]:

$$M_H < 160\ GeV \tag{1.44}$$

As the scale Λ becomes smaller the limit on the M_H becomes progressively weaker.

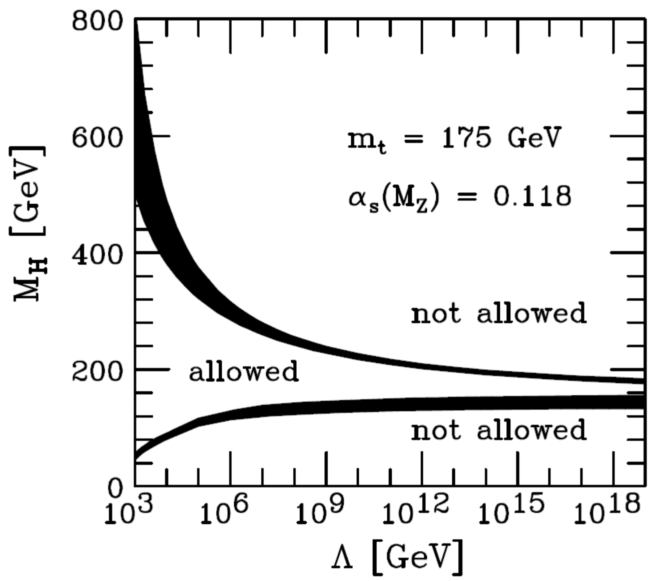


Figure 1.3: Lower and upper limits on the SM Higgs boson mass as a function of the scale of new physics, Λ, relying on the theoretical arguments. The allowed region is between the curves [15].

Λ is often interpreted as the scale of new physics because the SM cannot be extended beyond that scale. Another bound can be derived from the top quark Yukawa coupling which causes the scalar self-coupling to decrease, eventually making it negative if the Higgs boson mass is too low. Then the value of the effective potential becomes negative, dropping lower than the SM vacuum. This is equivalent to the statement 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 any more. It can be shown that if the SM is valid up to the scales of $\sim 10^{16}$ GeV one obtains:

$$M_H(GeV) = 130 + 2(m_t - 170) \tag{1.45}$$

where m_t is the mass of top-quark. Therefore, the mass of the Higgs boson is restricted between 126 to 160 GeV. It should be pointed out that this is exactly the mass range preferred by the EW precision measurements as discussed in section 1.5.2.

1.5 Experimental Searches for SM Higgs Boson

1.5.1 Direct Searches for Higgs Boson

LEP Searches

At e^+e^- colliders the importance of the various Higgs production mechanisms strongly depends on the center of mass energy ($\sqrt{s}$) of the collider and to some extent, also, on the Higgs boson mass. Up to LEP-II energies, $\sqrt{s} \leq 209$ GeV, ZH production via "Higgs-strahlung" with a variety of Higgs and Z boson decays was the main search channel. The actual Higgs-strahlung signal that was searched for at the LEP was mainly based on the $H \to b\bar{b}$ decay and various Z boson decays, $Z \to f\bar{f}$.

- **LEP-I era:** From autumn 1989 to summer 1995 the LEP accelerator was operated at center of mass energies in the vicinity of the mass of the Z boson, and in total 15.5 million hadronic Z boson decays ($Z \to q\bar{q}$) and 1.7 million leptonic decays ($Z \to \ell^+\ell^-$) were collected by the four LEP experiments, ALEPH, DELPHI, L3, and OPAL [16]. At the end of LEP1 data taking a peak luminosity of 2×10^{31} $cm^{-2}s^{-1}$ was reached, corresponding to approximately 1000 Z bosons being recorded per hour by each experiment. Three Higgs boson mass ranges can be roughly distinguished: (a) $0 \leq M_H \leq 2m_\mu$, (b) $2m_\mu \leq M_H \leq 3$ GeV, and (c) $M_H \geq 3$GeV, where m_μ

is the mass of the muon. These mass ranges differ in the dominant Higgs boson branching ratios and the robustness of their theoretical predictions, the kinematics and topology, the composition of the dominant backgrounds, and the search strategies applied. The searches performed by the four experiments allowed to exclude all Higgs boson mass hypotheses below 63.9 GeV at 95% CL (Confidence Level).

- **LEP-II era:** Starting in summer 1995 at 130/136 GeV, the center of mass energy of the LEP collider was continuously increased over the next years up to 209 GeV in 2000. Approximately $0.5(2.5) fb^{-1}$of integrated luminosity at center of mass energies of 206(189) GeV were collected in total by the four experiments. At LEP2 center of mass energies the dominant production process is again Higgs-strahlung. But in contrast to LEP1 the Z boson in the final state is dominantly produced on its mass shell. The hadronic decays of the Z boson were utilized in the search. Searches were performed in the "four jet channel" $ZH \to q\bar{q}b\bar{b}$ ($ZH \to q\bar{q}q\bar{q}$), "missing energy channel" $ZH \to \nu\bar{\nu}b\bar{b}$ ($ZH \to \nu\bar{\nu}q\bar{q}$), "tau channels" $ZH \to q\bar{q}\tau^+\tau^-$ and $ZH \to \tau^+\tau^-b\bar{b}$ ($ZH \to \tau^+\tau^-q\bar{q}$), and "lepton channels" $ZH \to e^+e^-b\bar{b}$ and $ZH \to \mu^+\mu^-b\bar{b}$ ($ZH \to e^+e^-q\bar{q}$ and $ZH \to \mu^+\mu^-q\bar{q}$).
  All SM Higgs boson mass hypotheses below 114.4 GeV were excluded with at least 95% CL from the combination of the four experiments. The expected combined limit on M_H was 115.3 GeV [16].

Tevatron searches

The Tevatron Run-II started in 2001 with $\sqrt{s} = 1.96$ TeV. From spring 2002 until its shutdown in autumn 2011 it provided proton-antiproton collisions corresponding to a data set of up to 10 fb^{-1} to each of the two experiments, CDF and DØ. In contrast to the searches at LEP, where the a priori signal-to-background ratio was at the level of 10^4 to 10^6 as both signal and background processes are mediated by the EW interaction, the huge total inelastic cross-section due to processes mediated by the strong force is more than ten orders of magnitude higher than the cross-section for Higgs boson production.

The Higgs boson mass range 100-200 GeV was investigated by both the experiments. For $M_H \leq 130$ GeV, the most abundant and most sensitive search channels are $H \to b\bar{b}$ produced in association with a weak gauge boson W or Z, which decay leptonically via $W \to \ell\nu$, $Z \to \ell^+\ell^-$, or $Z \to \nu^+\nu^-$. For M_H in the range 130-200 GeV the sensitivity is dominated by the search for $H \to W^+W^- \to \ell^+\nu\ell^-\bar{\nu}$ produced in gluon fusion. The mass regions M_H = 100-120 GeV and M_H = 139-184 GeV are expected to be excluded as shown in Figure 1.4 [17]. The observed excluded mass range are weaker and extend over the range M_H = 100-103 GeV and M_H = 147-180 GeV. The largest deviation from the background-only hypothesis is observed for M_H = 120 GeV with a local p value[6] of 1.5×10^{-3}, corresponding to a local significance of 3.0 σ[7].

1.5.2 Indirect searches for Higgs Boson

Once a Higgs field is introduced in the SM, its virtual excitations contribute to several physical observables, from the mass of $W^{\pm}$ boson, to various leptonic and hadronic asymmetries to many other EW observables that are usually considered in precision tests of the SM. Higgs mass can thus be extracted indirectly from the precision fits of all the measured EW observables. This is an important test that illustrates the predictability of the SM itself. Hence, precision measurements of the EW parameters can be used to obtain constraints on the Higgs boson mass. The *one-loop* effects of the top quark and Higgs boson on the W boson mass are shown in Figure 1.5.

The three main EW observables are the EW ρ - parameter, the weak mixing angle $\sin^2\theta_W$ and the W/Z boson mass defined as:

$$\rho = \frac{M_W^2}{M_Z^2}\left(1 - \sin^2\theta_W\right) = 1 + \Delta r \tag{1.46}$$

The term Δr is sensitive to the existence of heavy particles in the SM, in particular the top quark and the Higgs boson. These radiative corrections Δr can be written as:

[6] p value is the probability of the background to fluctuate at least as large as to describe the observed excess in events under background only hypothesis.

[7] σ is the measure of significance of any excess of events observed.

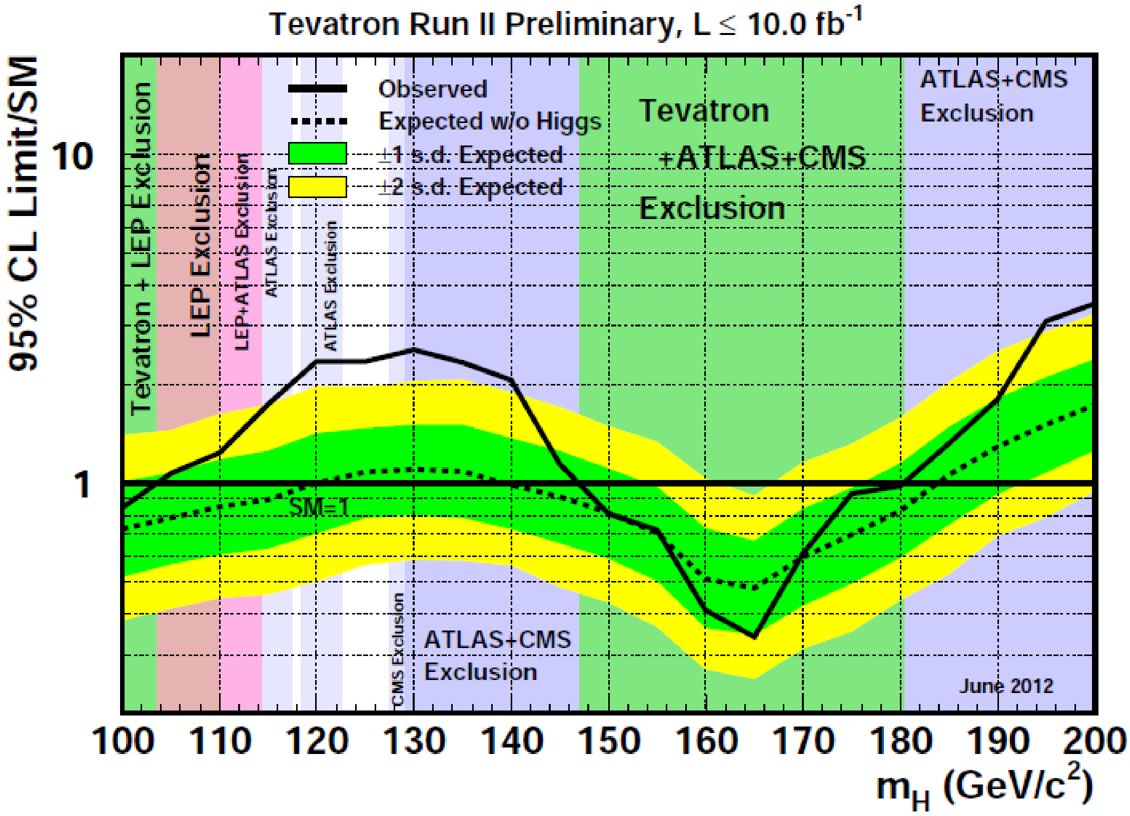


Figure 1.4: Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross-section, as function of the Higgs boson mass for the combined CDF and DØ analyses. The limits are expressed as a multiple of the SM prediction for test masses (every 5 GeV) for which both experiments have performed dedicated searches in different channels. The points are joined by straight lines for better readability. The bands indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal.

$$\Delta r = \frac{3G_F}{8\pi^2\sqrt{2}} m_t^2 + \frac{\sqrt{2}G_F}{16\pi^2} m_t^2 \left[\frac{11}{3} \ln\left(\frac{M_H^2}{M_W^2}\right) + \cdots\right] + \cdots \tag{1.47}$$

Hence one can deduce the Higgs boson mass from precision measurements of the top quark and W boson masses. These quantities have been measured in several experiments at the LEP, SLC and the Tevatron as shown in Figure 1.6.

The correlation between the Higgs boson mass, top quark and W boson masses is shown in the contours in Figure 1.7. Increased precision in the measured mass of the top quark and W boson have further reduced the uncertainty on these constraints over the last few years. Therefore, assuming no new physics beyond the SM, all available EW precision data favours a light Higgs boson. The actual value of Higgs mass emerging from

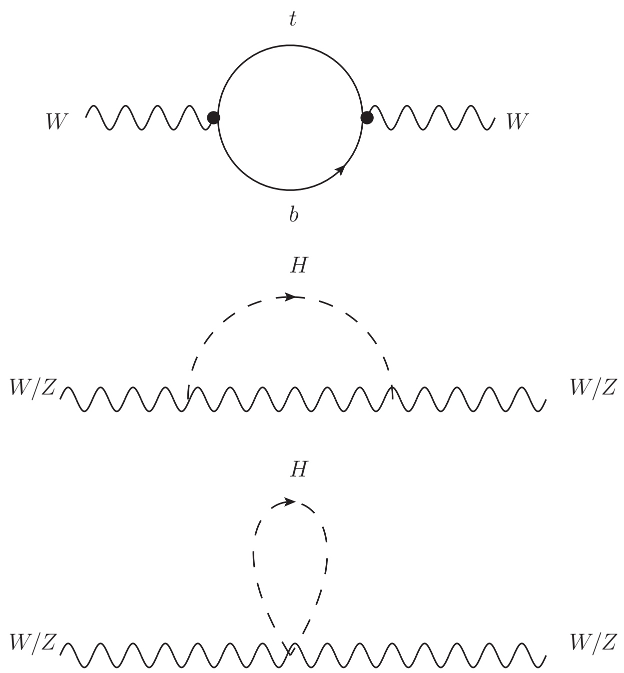


Figure 1.5: The masses of top quark, W and Higgs bosons are related via radiative loop corrections.

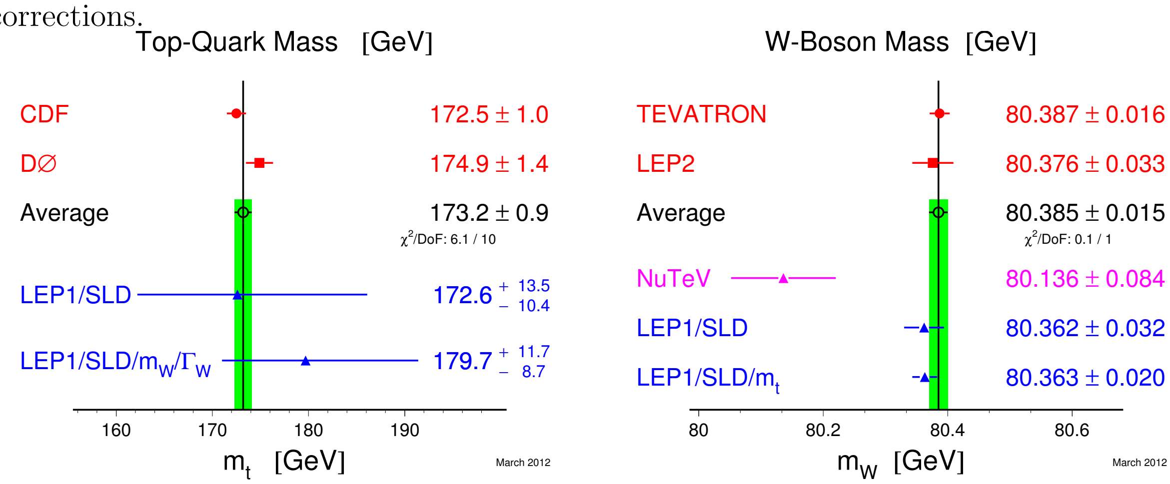


Figure 1.6: World average and individual measurements of the top quark (left) and W boson mass (right) [18] including both direct measurements of the top quark mass as well as indirect measurements via higher order corrections to the W boson mass.

the EW precision fits strongly depends on theoretical predictions of physical observables that include different orders of strong and EW corrections. As an example, in the contours of Figure 1.7 (right plot), the magenta arrow shows how the yellow band will move for one standard deviation variation in the QCD fine structure constant $\alpha(M_Z^2)$. The shaded band in the Figure 1.7 (left plot) shows the SM prediction based on the value of G_F for various values of the Higgs boson mass and fixed $\Delta\alpha_{had}(M_Z)$. Varying the hadronic

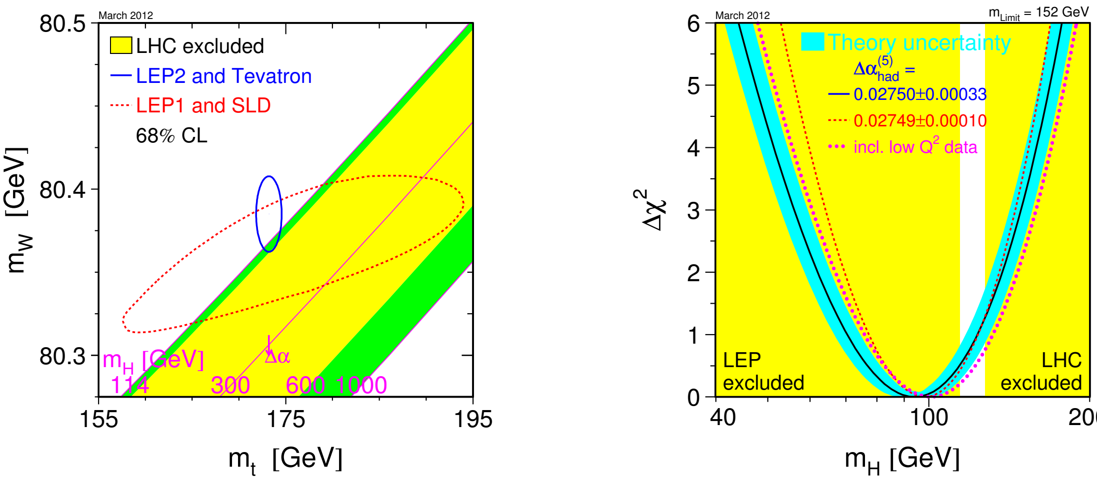


Figure 1.7: Combined results between m_t and M_W. The left plots shows the 68% C.L. for the indirect measurements of LEP1 and SLD (dashed ellipses) and direct observations from LEP2 and Tevatron (solid ellipses). The right plot shows the constraints from the EW fits on a potential SM Higgs boson mass as $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ vs. M_H curve.

vacuum polarization yields an additional uncertainty on the SM prediction indicated by the arrow labeled $\Delta\alpha$.

In the Figure 1.7 (right plot) the solid black line is the result of the fit using high-Q^2 data while the blue band represents an estimate of the theoretical error due to missing higher order corrections. The yellow shaded area shows the direct exclusion limits by LEP-II, Tevatron and LHC. The dotted line shows the corresponding fits using low-Q^2 data and the dashed line shows the contributions of the hadronic correction from light quark to $\alpha(M_Z^2)$. The SM Higgs boson mass preferred by the fit is [18]:

$$M_H = 94^{+29}_{-24} \ \text{GeV} \tag{1.48}$$

1.6 Higgs Production at the LHC

In the SM, Higgs boson production in proton-proton collisions [19–21] can happen through four main modes: gluon fusion $gg \to H$; vector boson fusion $qq \to H + 2$ jets, associated production of a Higgs boson with a W or Z boson, and associated production with a $t\bar{t}$ pair. The hierarchy of the cross-sections and their dependence on the Higgs boson mass is shown in Figure 1.8. In general all production cross-sections decrease with increasing

Higgs boson mass, except for $gg \to H$ around the $M_H \sim 2m_t$ threshold.

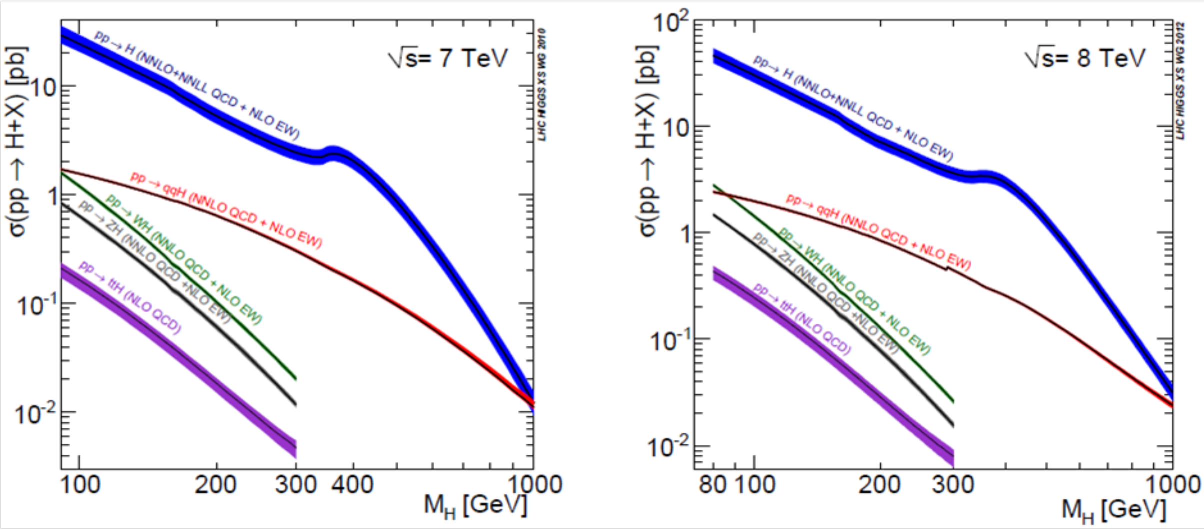


Figure 1.8: Theoretical predictions for the Higgs boson production cross-sections in proton-proton collisions at $\sqrt{s} = 7\ TeV$ (left) and $\sqrt{s} = 8\ TeV$ (right).

1.6.1 Gluon-Gluon Fusion

The main Higgs boson production mechanism at the LHC is through gluon fusion, for which the leading Feynman diagram involves a quark loop (Figure 1.9). A remarkable property of the amplitude for this process at leading order is that it is zero for massless quarks, and saturates to a constant for increasing m_q/M_H (m_q is the quark mass), so that the 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, but 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 corrections to the process are positive and 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, whose size is nonetheless small. Because of the slow convergence of the QCD radiative corrections, the theoretical

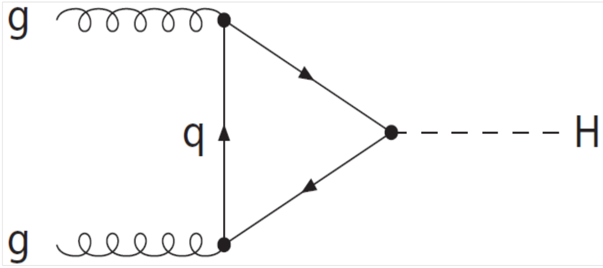


Figure 1.9: Leading diagram for $gg \rightarrow H$ production.

uncertainty on the predicted cross-section from neglected higher order terms is 8% - 12% depending on the Higgs boson mass even for the most precise NNLO + NNLL result. The theoretical uncertainties from the knowledge of the gluon parton density function are likewise sizable, about 8% up to Higgs boson masses below 250 GeV and increasing 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.

1.6.2 Vector Boson Fusion

With a production cross-section of about one tenth of the gluon fusion one, vector boson fusion (VBF) is the second largest production mode relevant at the LHC. The three leading order diagrams for the process in the t, u and *s-channels* are shown in Figure 1.10. In the first two, either or both of the incoming quarks can also be replaced by anti-quarks. From an experimental point of view, the t and *u-channels* are the most relevant ones, because the momentum transfer between the two scattering quarks tends to be of order M_V (M_V is the mass of the vector boson), small compared to $\sqrt{s}$, resulting in final states with two jets at high rapidities. As the two outgoing quarks are not connected by any quark or gluon line, QCD radiation in the region between the two jets is suppressed, another experimentally viable feature, which is useful in separating $qqH \rightarrow qqWW \rightarrow 2l2\nu + 2jets$ signal from the larger $t\bar{t} \rightarrow b\bar{b}WW \rightarrow 2l2\nu + 2jets$ background. For this process, QCD radiative corrections are significantly smaller, $O(10\%)$ at NLO [22]. As a consequence, the theoretical uncertainty from higher order terms on the most accurate prediction available,

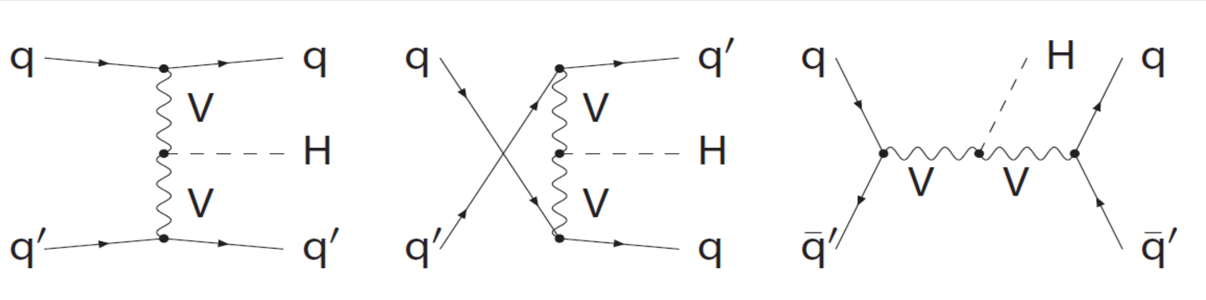


Figure 1.10: Leading order Feynman diagrams for $qq \to H + 2jets$ production through Vector Boson Fusion. *t-channel left*, *u-channel middle*, *s-channel right*.

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, ranging from about 2.5% at low M_H (200 GeV) to about 7.5% at 600 GeV.

1.6.3 Associated Production

The associated production of a Higgs boson and a W or Z boson is characterized by an even smaller production cross-section than VBF. This mode is nonetheless experimentally viable when considering Higgs boson decays to bottom quark pairs or to τ pairs, since leptons and neutrinos from the W or Z bosons decay can provide handles to select events, while the all-hadronic final states from $gg \to H \to b\bar{b}, qq \to qqH \to qqb\bar{b}$ suffer larger backgrounds from QCD multijet production. The leading order Feynman diagrams for this process is closely related to the VBF ones since they both rely on the WWH and ZZH vertices (Figure 1.11). Associated $t\bar{t}H$ (Figure 1.12) is an important channel, allowing for a direct measurement of the Yukawa coupling between top quark and Higgs boson. But the very low production cross-section and the complex final state with a large number of hadronic jets make it very challenging to reconstruct properly.

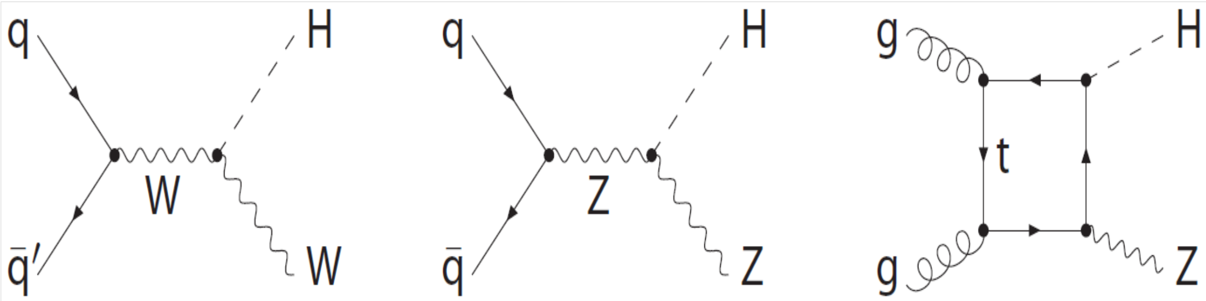


Figure 1.11: Leading order diagrams for WH and ZH associated production, and higher-order contribution to ZH with gluons in the initial state.

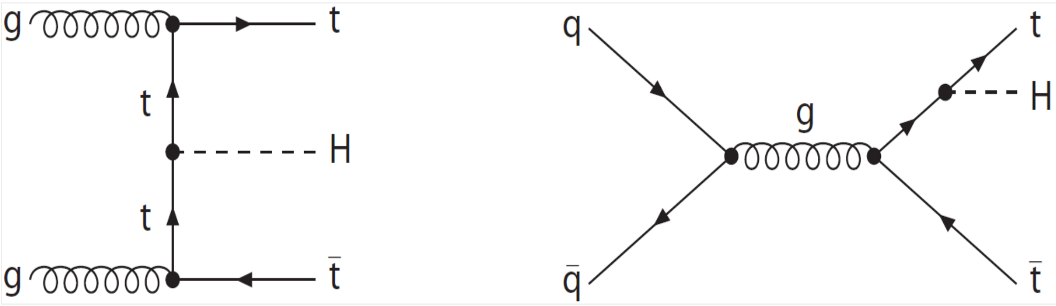


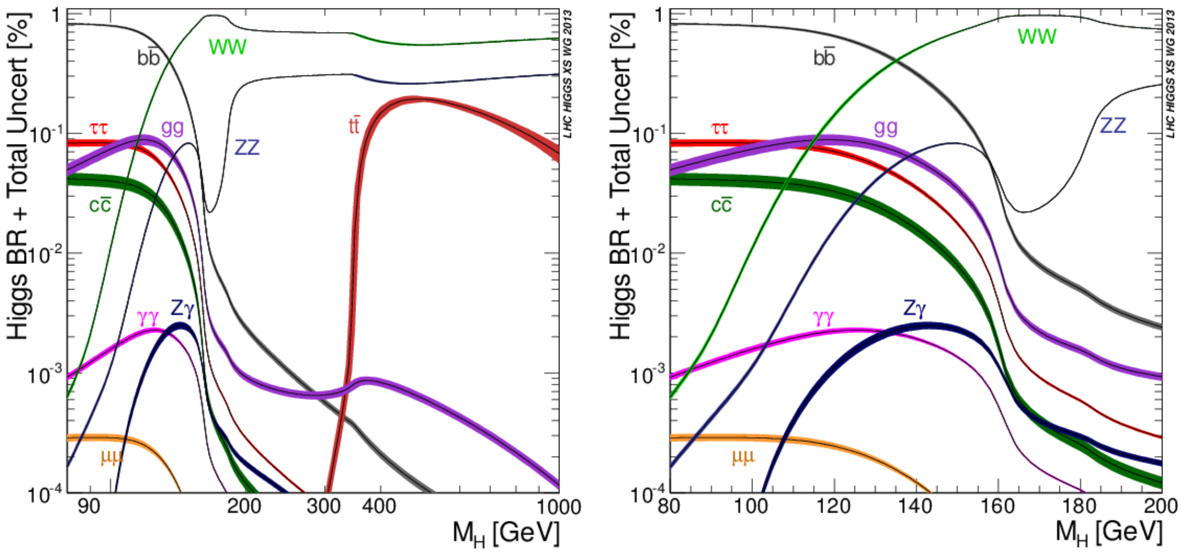
Figure 1.12: Leading order diagrams for $t\bar{t}$H associated production with gluons and quarks in the initial state.

1.7 Higgs Decays at the LHC

The Branching Ratios (BR) of the different Higgs decay channels [23–26] are shown in Figure 1.13 as a function of M_H. The Higgs decay processes are divided into the complementary low, intermediate and the high mass regions. In the low mass region (up to $\sim$ 130 GeV), the Higgs boson decaying into fermions is a dominating process, in particular, the $H \to b\bar{b}$ channel. In the intermediate region (M_H = 130-180 GeV), the Higgs boson decays into vector boson pairs dominate. A peak in the $H \to W^+W^-$ decay is visible around 160 GeV, where the production of two on mass-shell W bosons become possible and the production of a real ZZ pair is still not allowed. At high masses ($M_H >$ 180 GeV), on-shell $t\bar{t}$ pairs ($\sim$ 350 GeV) can also be produced. Indeed, the most promising decay channels for the Higgs discovery do not only depend on the corresponding branching ratios, but also on the capability of experimentally detecting the signal while rejecting the backgrounds.

1.7.1 Low Mass Region

In the low mass range, $M_H < 130$ GeV, the channel with the highest BR is $H \to b\bar{b}$. The search for a Higgs boson in this channel is quite challenging due to the overwhelming amount of QCD background. In spite of its low cross-section, the channel with a Higgs boson produced via Higgs-strahlung or in association with a $t\bar{t}$ decaying to $b\bar{b}$, can give interesting results because of its distinctive signature. The most relevant decay channel in this mass region is the decay of Higgs boson into a pair of photons, $H \to \gamma\gamma$. Despite its low cross-section times BR the two high energy photons provide a very clear signature of the event and a narrow peak is expected in the two photons invariant mass distribution. It only suffers from the $q\bar{q} \to \gamma\gamma$ and γ + jets backgrounds. $H \to ZZ^* \to 4\ell$ is also very



(a) (b)

Figure 1.13: The branching ratios for the different Higgs decay channels as a function of M_H (a) in whole mass range (b) zoomed at low mass region.

important channel because of very less background and high resolution. In this channel, four leptons are present in the final state which can be reconstructed and identified very well.

1.7.2 Intermediate Mass Region

In the mass range from 130-180 GeV, the decay of Higgs boson into a pair of W and Z bosons are the main processes. The most interesting channels in this region are: $H \to WW^* \to 2\ell 2\nu$ and $H \to ZZ^* \to 4\ell$. The first one has greater cross-section, and it is very important in the range of mass where a Higgs boson can decay into two real W bosons. Nevertheless, due to the presence of neutrinos, the reconstruction of the Higgs boson in this case is quite difficult. The signal event signature is two isolated high transverse momentum (p_{T}) leptons with a small opening angle in the transverse plane and undetected neutrinos resulting in high missing transverse energy. The decay mode $H \to ZZ^* \to 4\ell$ has a lower BR with respect to WW decay, but, as mentioned earlier, its final state is very clean and almost background free and hence it is also called "Golden Channel" for search of the Higgs boson. This decay channel allows a precise reconstruction of the Higgs boson due to the presence of isolated charged leptons with high transverse momenta.

1.7.3 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 either into two real Z bosons or two real W bosons. Since Higgs production cross-section decreases very fast at higher masses, channels with high BR becomes more important. Hence $H \to ZZ \to 2\ell 2\nu$, $H \to ZZ \to 2\ell 2q$ and $H \to WW \to \ell\nu qq$ are the dominant channels at high mass searches. Figure 1.14 shows the cross-section times BR of various high mass channels studied at the LHC. Even if the discovered new boson around mass of 125 GeV is the Higgs boson, it is still necessary to probe the high mass region. These searches at high mass can be a good benchmark for the other beyond the SM scenarios where SM Higgs-like resonances can appear.

1.7.4 Higgs Total 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 1.15. The total decay width is very narrow in

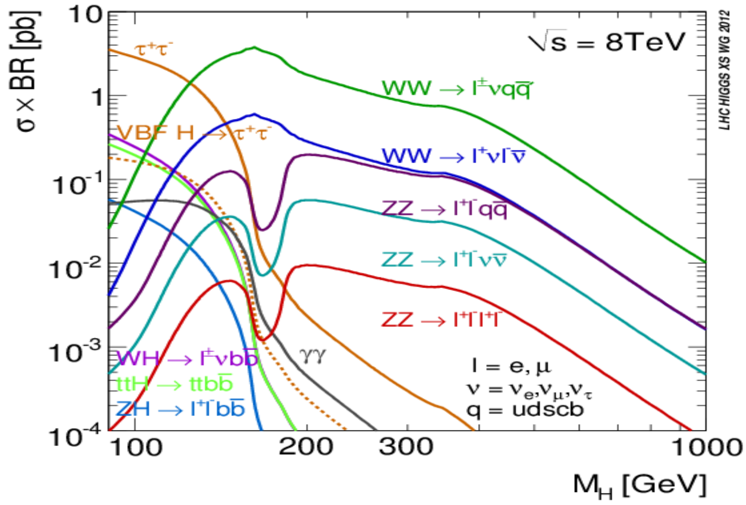


Figure 1.14: SM Higgs boson production cross-section times BR at $\sqrt{s}$= 8 TeV.

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

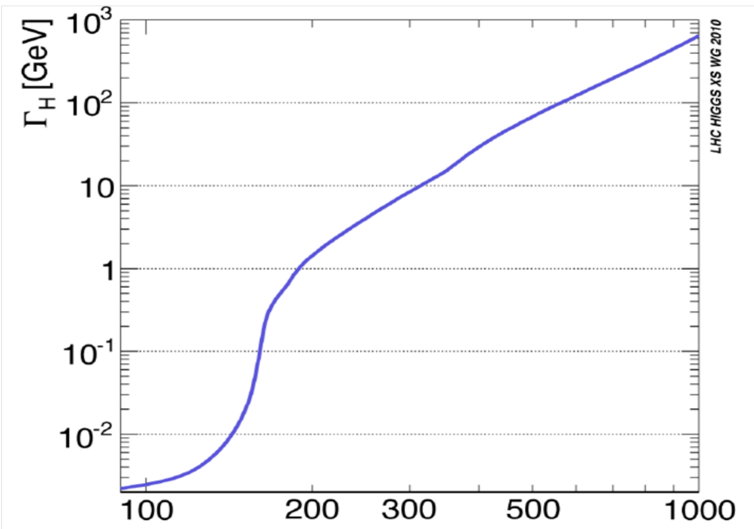


Figure 1.15: The SM Higgs boson total decay width as a function of M_H (GeV).

total decay width comprises of the ZZ and the WW partial widths. For the higher Higgs 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.

1.8 Recent Results of Higgs Searches at the LHC

On 4^{th} July 2012, the first observation of Higgs boson was announced worldwide by two experiments, CMS and ATLAS. The results were also updated with the full data from 2011 and 2012, which correspond to integrated luminosities of 5.0 fb^{-1} at $\sqrt{s} = 7$ TeV and up to 19.7 fb^{-1} at $\sqrt{s} = 8$ TeV in proton-proton collisions at the LHC [27]. Five decay modes are studied: $\gamma\gamma$, ZZ, W^+W^-, $\tau^+\tau^-$, and $b\bar{b}$ quarks. The mass of the new boson is measured to be 125.7 $\pm$ 0.3(stat.) $\pm$ 0.3(syst.) GeV. The event yields obtained by the different analyses targeting specific decay modes and production mechanisms are consistent with those expected for the SM Higgs boson. The best-fit signal strength for all channels combined, expressed in units of the SM Higgs boson cross-section (σ_{SM}), is 0.80 ± 0.14 at the measured mass. The consistency of the couplings of the observed boson with those predicted for the SM Higgs boson is tested in various ways, and no significant deviations are found. Table 1.2 summarizes the median expected and observed local significance (σ) for a SM Higgs boson mass hypothesis of 125.7 GeV from the individual decay modes. The expected significance is evaluated assuming the expected background and signal rates.

Decay Mode	**Expected σ**	**Observed σ**
ZZ	7.1	6.7
$\gamma\gamma$	3.9	3.2
WW	5.3	3.9
$b\bar{b}$	2.2	2.0
$\tau\tau$	2.6	2.8

Table 1.2: The median expected and observed significances of the excess in the combination of individual decay modes for a SM Higgs boson mass hypothesis of 125.7 GeV.

The low probability for an excess at least as large as the observed one to arise from a statistical fluctuation of the background leads to the conclusion that the observation of a new particle is confirmed with a mass near 125 GeV in the $\gamma\gamma$, $ZZ \rightarrow 4\ell$ and

$WW \to \ell\nu\ell\nu$ channels and that the excess is consistent with expectation in the $b\bar{b}$ and $\tau\tau$ channels. The $\gamma\gamma$ and $ZZ \to 4\ell$ decay modes indicate that the new particle is a boson, and the diphoton decay implies that its spin is different from unity. The mass of the observed state is measured using the $ZZ \to 4\ell$ and $\gamma\gamma$ channels that have excellent mass resolution. Figure 1.16 shows the values of ratio of upper limit on predicted cross-section for Higgs boson at 95% confidence level to the SM cross-section (σ/σ_{SM}) for the combination (solid vertical line) and for sub-combinations grouped by decay mode. The band shows the overall σ/σ_{SM} uncertainty. The horizontal bars indicate the ± 1 standard deviation uncertainties (both statistical and systematic) in the σ/σ_{SM} values for the individual modes.

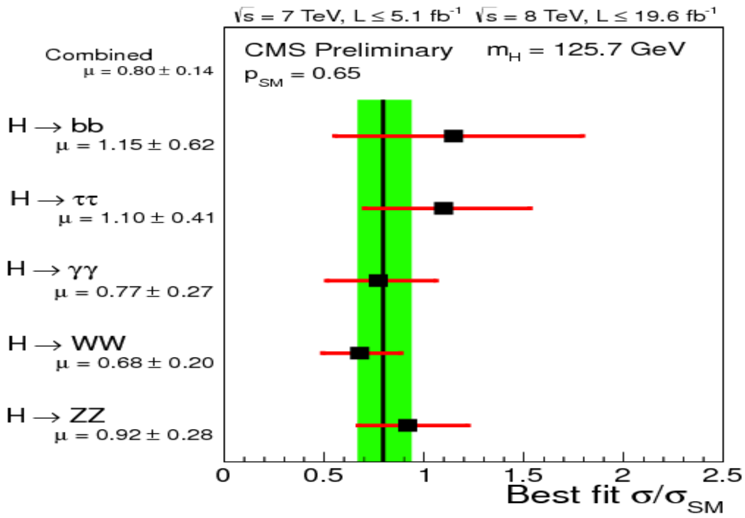


Figure 1.16: Values of σ/σ_{SM} for the sub-combinations by decay mode.

1.9 Thesis Organization

In this Thesis, a search for the SM Higgs boson is described in $H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ decay mode. This decay mode is sensitive to higher Higgs boson masses ($\geq$ 200 GeV). This is one of the leading channels in the CMS experiment for search of the SM Higgs boson for $M_H \geq$ 400 GeV. This mode has six times higher branching ratio than $H \to ZZ \to 4\ell$ channel and hence this benefits from the larger statistics. But, this comes at the cost of worse resolution, because of the presence of transverse missing energy (E_T^{miss}) arising from the undetected neutrinos. Even after the possible discovery of the SM Higgs boson at around 125 GeV, this channel is important to study because it is still not clear if the discovered particle is SM Higgs boson or possibly the lightest of several bosons predicted in some theories that go beyond the SM. It is also necessary to rule out the presence of any other SM Higgs-like resonance in high mass range. Also, SM-like Higgs search using this channel can be a benchmark for searching of Higgs-like resonances at higher energies suggested by some other models which are beyond SM [28].

The total inelastic cross-section (σ_{inel}) is about 10 orders of magnitude larger than the signal production cross-section. Even individual production cross-sections of typical background processes like $t\bar{t}$ or Z boson production are factors or even orders of magnitude larger than the signal production cross-section as shown in Figure 1.17. Hence selection requirements on the kinematic variables have to be applied to suppress the background and enrich the relative signal contribution in the data sample.

The Thesis is organized in **six** chapters. A brief description of each chapter is given below:

Chapter 1 provides an introduction of particle physics and overview of the SM of particle physics. Starting from the gauge structure of the SM, this chapter describes the EW interactions and spontaneous symmetry breaking in this sector to provide masses to vector bosons. Also, it describes the mass generation for fermions by Yukawa interactions with Higgs field. Higgs boson self interactions are also discussed which are responsible for its own mass, which is a free parameter of the theory. Various theoretical and experimental constraints on Higgs mass are described in this chapter. Experimental constraints include the searches at the LEP and the Tevatron experiments and continued with the LHC Higgs search program. Higgs phenomenology including its production and decay in various channels is discussed according to the LHC Higgs boson physics program. Latest results on the Higgs boson from CMS are discussed briefly. At last, motivation for $H \to ZZ \to 2\ell 2\nu$ analysis is described, which is discussed in detail in this Thesis.

Chapter 2 describes in detail the experimental apparatus used for the collection of data, which is studied in this Thesis. These are the LHC accelerator and the CMS detector.

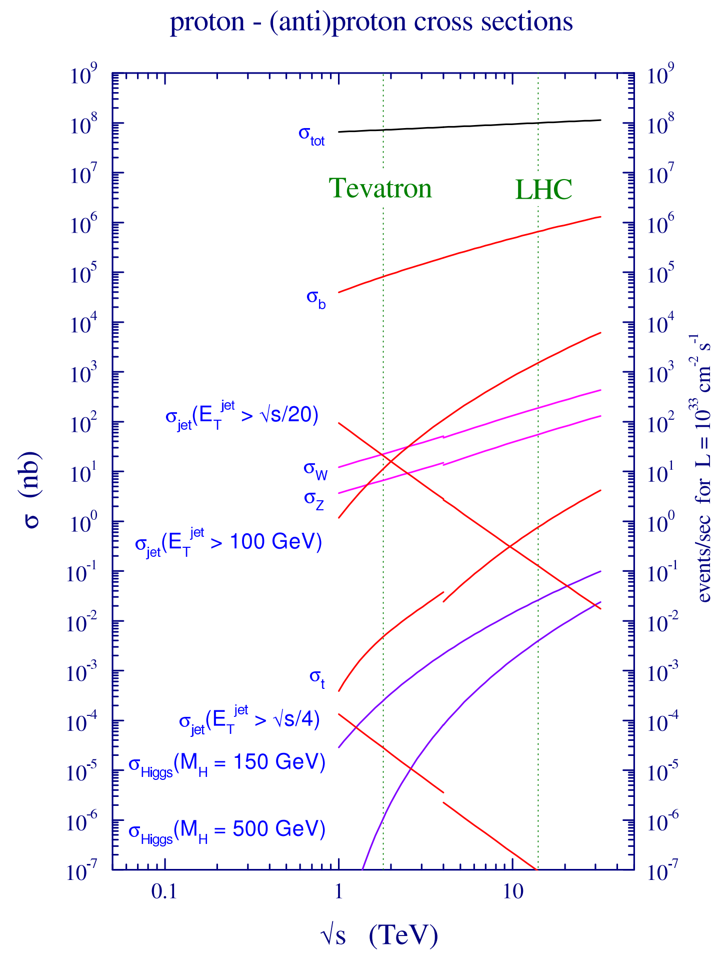


Figure 1.17: Comparison of the inclusive production cross-section for various processes depending on the center of mass energy. The dashed green lines indicate the center of mass energies at the Tevatron and at the Large Hadron Collider.

This chapter starts with an introduction of the LHC accelerator and then a brief overview is given for its designed and operating parameters. The CMS detector is then discussed starting from its components to its performance. Various sub detectors in the CMS are explained briefly. The analysis discussed in this Thesis has leptons (electrons and muons) and missing transverse energy (E_T^{miss}) in the final state and also jets from the initial state radiations (ISR). So, every sub detector of CMS plays a critical role. Hence, the performance for each sub detector is also discussed. The CMS trigger system and data acquisition system are briefly discussed along with the physics requirements of the trigger system. In the end, some 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 outlines the reconstruction of various physics objects like primary vertex, electrons, muons, jets, E_T^{miss} etc. Along with that the identification criterion are explained for each one of them. The performance of these objects is also discussed. Various algorithms are given for b-quark tagging used in the CMS as well.

Chapter 4 describes a search for the SM Higgs boson in the $H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ decay channel. The analysis uses the data collected by the CMS in years 2011 and 2012, which correspond to full integrated luminosity of 5.0 $fb^{-1}$and 19.7 fb^{-1}respectively. The analysis starts with an introduction of signal and various background processes. The Higgs mass range considered in this analysis is 200-1000 GeV. The whole selection procedure is then described starting with the selection of physics objects and then the kinematical selection of Z boson candidates. Since one of the Z decays to neutrinos, so high E_T^{miss} is expected in signal events. Hence, a high E_T^{miss} threshold between 80 GeV to 110 GeV is applied, depending on the Higgs mass hypothesis. Events are divided into exclusive categories based on jet-multiplicity. A separate selection criteria is set up for identifying events produced by vector boson fusion (VBF) production mode. After selecting these events, remaining events are categorized into exclusive 0-jet and inclusive 1-jet. Final discrimination is done by using transverse mass of Higgs (M_{T}) variable in non-VBF category and E_T^{miss} for VBF category. All the major backgrounds like $Z + Jets$ and non-resonant backgrounds like $t\bar{t} + jets$, WW etc are estimated from data using data-driven methods. A statistical analysis is then carried out based on the final selection variable. Shape analysis (baseline analysis) is carried out in which no selection criteria is applied on final discriminant and full shape is used to set up upper limits on the production cross-section for Higgs boson at various masses. A cross-check cut and count analysis is also done where all selection criteria are applied and then the event count is used to set up upper limits on the Higgs boson production cross-section. No excess was observed over the SM background prediction and the exclusion limits were set on the mass of Higgs boson.

Chapter 5 describes the search in same decay channel i.e. $H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$, using Multivariate technique, Boosted Decision Tree (BDT). This method was introduced to explore further possible improvements in the baseline analysis and to cross-check the existing method with a different approach. This analysis also uses full data collected by the CMS detector in 2011 and 2012. The basic event selection like physics objects selection and selection of Z candidates is same as in baseline analysis. E_T^{miss} cut is relaxed in order to provide more phase space for BDT and also to avoid statistical fluctuations in the results. The optimization of BDT parameters is done keeping in mind the best expected limit on Higgs production cross-section. BDT is trained and tested for each Higgs mass hypothesis and for $e^+e^-\nu_e\bar{\nu_e}$ and $\mu^+\mu^-\nu_\mu\bar{\nu_\mu}$ channels. Events are divided according to jet-multiplicity only for training and then later on combined before doing statistical analysis. By dividing these events, we get a gain of around 15-20% depending on Higgs mass. The background estimation techniques are same as used in baseline analysis. A Shape analysis has been carried out on final BDT output to set up upper limits on Higgs production cross-section.

Chapter 6 summarizes the results presented in this Thesis.

Chapter 2

Large Hadron Collider and the CMS Experiment

The Large Hadron Collider (LHC) [29] is an accelerator and collider, of 27 km circumference, proposed and realized by the European Organization for Nuclear Research (CERN). This is located in the already existing LEP tunnel on the Swiss-French border, at about 100 m depth underground. It has been designed to collide protons, as well as lead ions, at an unprecedented energy and rate, in order to answer some of the most fundamental questions of physics. The choice to have a proton-proton collider, rather than electron-positron one, is mainly driven by the need to reach very high energies without significant loss due to synchronization radiation, which is proportional to the fourth inverse power of mass of accelerated particles. LHC has been designed to produce proton-proton collisions at a center of mass energy ($\sqrt{s}$) of 14 TeV, which is about seven times higher than the highest energy reached at the Tevatron, and with instantaneous luminosity of 10^{34} $cm^{-2}s^{-1}$, which is about two orders of magnitude more than the luminosity of any previous hadron collider machines. This designed center of mass energy will not be reached until after a series of hardware updates which are currently underway during the first shutdown of machine. The LHC is expected to be started again in 2015.

On 23 November 2009, LHC produced the first proton-proton collisions. After few pilot

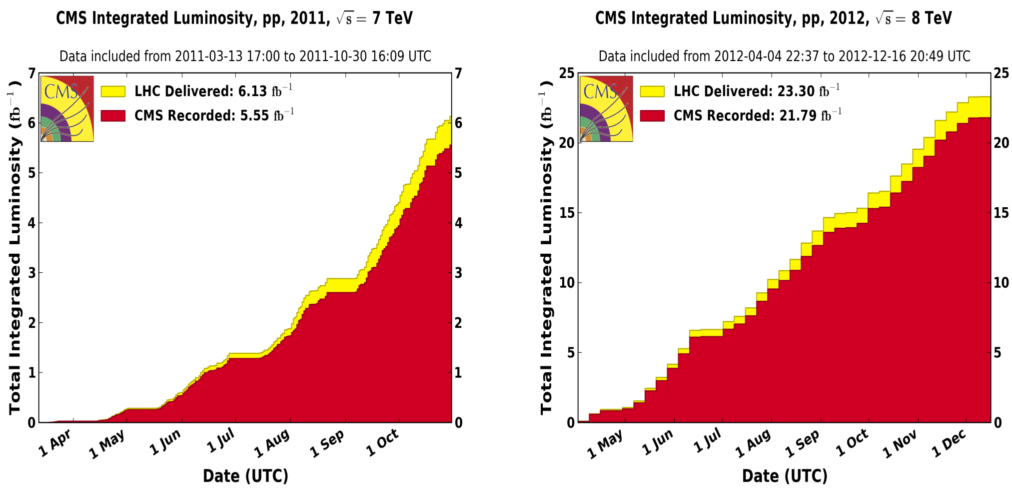


Figure 2.1: Integrated luminosity delivered by LHC and recorded by CMS in proton-proton collisions in 2011 (left) and in 2012 (right).

proton runs at beam energies of 450 GeV and 1.18 TeV, the energy was ramped up to 3.5 TeV and, on 30 March 2010, the first collisions occurred at $\sqrt{s} = 7$ TeV. This was the highest ever $\sqrt{s}$ reached at a particle collider, and the debris of the collisions are recorded by the four experiments: ALICE, ATLAS, CMS and LHCb. During these years the LHC has worked beyond expectations and the total integrated luminosity delivered in 2011 is 6.13 fb^{-1}, as shown in Figure 2.1. In 2012 the center of mass energy has been incremented to 8 TeV, with an integrated luminosity of 23.3 fb^{-1}.

2.1 LHC Design

The proton beams which feed the collisions in the LHC start in a modest hydrogen tank at the beginning of the accelerator. The hydrogen atoms are stripped of their electrons, becoming protons that are accelerated along the linear accelerator (Linac 2) up to an energy of 50 MeV. The protons are then injected into the Proton Synchrotron Booster (PSB), where they are further accelerated to 1.4 GeV. The next steps are the Proton Synchrotron (PS) and Super Proton Synchrotron (SPS), which accelerate protons to 25 GeV and 450 GeV respectively.

After the SPS, the proton beams are injected into the LHC ring, where they circulate in opposite directions through the 27 km circumference ring. Here, they are accelerated to their final energy, stored and steered toward collision. The accelerator chain is shown in Figure 2.2. The acceleration of the protons at each stage is done using RF (radio frequency) cavities. Inside of these cavities, the time oscillations of electromagnetic fields 'push' the proton bunches to accelerate them. The timing of these oscillations also help keep the bunches together, as protons lagging slightly behind the majority receive a larger relative boost than protons in the bulk. There are sixteen such cavities in the LHC ring, with eight being used for each of the beams. The operating frequency (40 MHz) of LHC is driven by the nominal LHC bunch spacing, allowing the bunch spacing of 25 ns.

The proton beams are steered using a series of 1232 superconducting dipole magnets, with magnetic fields up to 8.3 Tesla. They are maintained at 1.9 K of temperature, making LHC the largest cryogenic system in the world. The cooling of the magnets is achieved by superfluid helium. The cooling process happens in a sequential way, i.e. first helium is cooled down to 4.5 K, then it is injected into the magnet cold masses, and, once the magnets are filled, the refrigeration units bring the temperature to 1.9 K. Because there are two proton beams within the LHC ring circulating in opposite directions, the dipoles have a twin-bore design. In this setup, there are two beam pipes, with each pipe having its own set of coils to produce steering field. In addition to the dipoles, there are about 850 quadrupole magnets and about 6000 correcting magnets which focus the proton bunches.

Table 2.1 shows the designed parameters of the LHC, which are required to reach a peak luminosity of $10^{34}\ cm^{-2}s^{-1}$ for proton-proton collisions at $\sqrt{s} = 14$ TeV. The number of events per second (N_{event}) generated in the LHC collision is given by:

$$N_{event} = \mathcal{L} \times \sigma_{event} \tag{2.1}$$

where σ_{event} is the cross-section for the event type under study and $\mathcal{L}$ is the instantaneous luminosity which depends upon the beam parameters and can be written for a Gaussian beam distribution as:

$$\mathcal{L} = \frac{N^2 k_b f \gamma}{4\pi\epsilon_n\beta^*} F \tag{2.2}$$

where

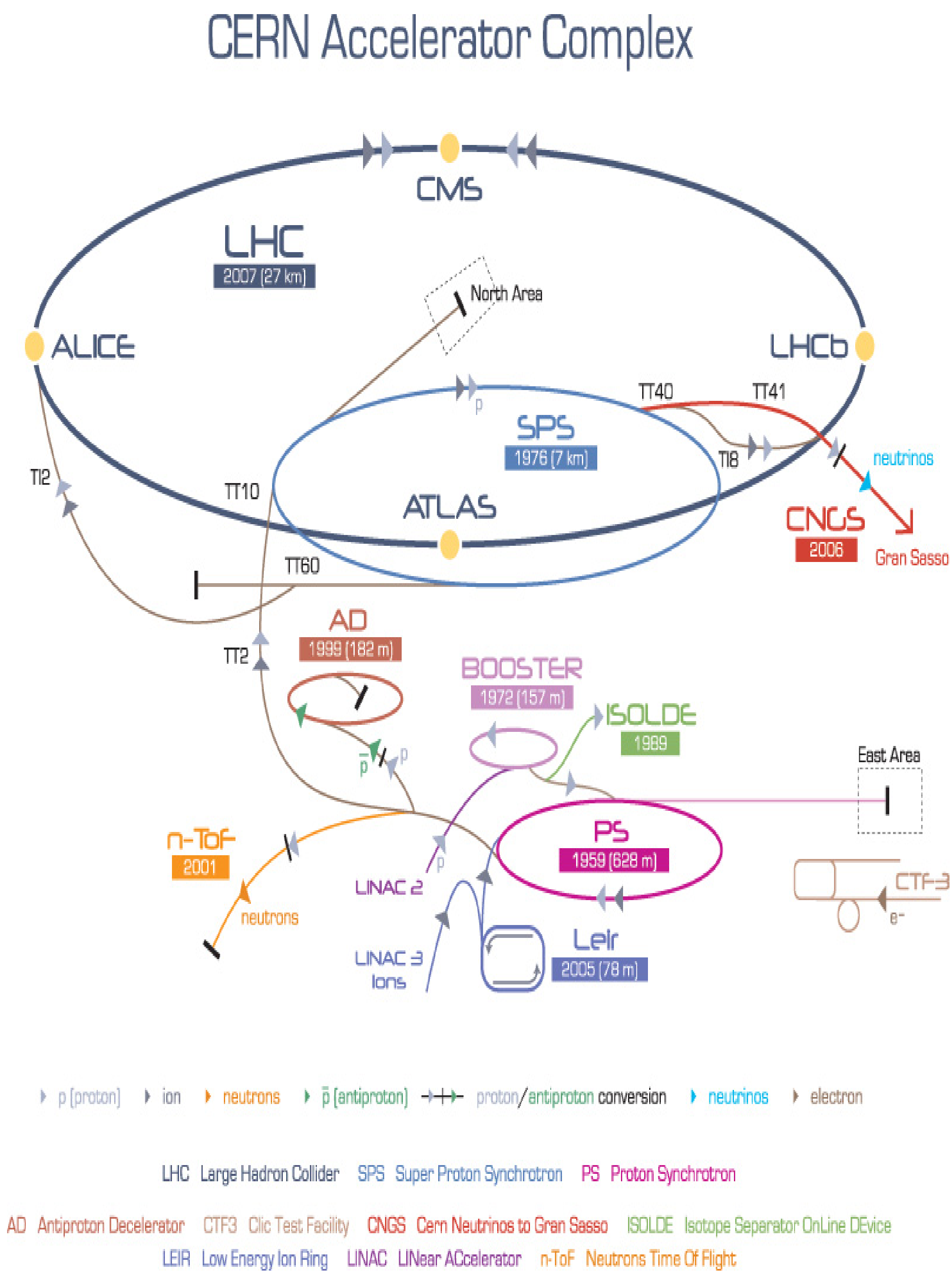


Figure 2.2: Schematic layout of the LHC accelerator.

Parameter	**Value in 2012**	**Design Value**
Beam Energy (TeV)	4	7
β^* in IP (m)	0.6	0.55
Bunch spacing (ns)	50	25
Number of bunches	1374	2808
Average bunch intensity (protons per bunch)	1.6-1.7$\times10^{11}$	1.15$\times10^{11}$
Normalized emittance at start of fill (mm.mrad)	2.5	3.75
Peak Luminosity ($cm^{-2}s^{-1}$)	7.7$\times10^{33}$	1$\times10^{34}$ $cm^{-2}s^{-1}$
Max. mean number of events per-bunch crossing	≈40	19
Stored beam energy (MJ)	≈140	362

Table 2.1: Design performance parameters for the LHC.

- N is the number of protons per bunch
- k_b is the number of bunches
- f is the revolution frequency
- γ is the relativistic factor
- ϵ_n is the normalized transverse beam emittance
- β^* is the value of beta function at the collision point
- F is the geometric reduction factor arising from the crossing angle

2.1.1 Terminology

It is useful to introduce some general definitions and variables commonly used to describe the phenomenology at hadron colliders. The reference frame used in this work is the standard CMS reference frame which is a Cartesian system centered at the interaction point and with the z-axis tangent to the beam line. The x-axis is chosen to be horizontal and pointing towards the center of ring, and the y-axis is vertical and pointing upwards as shown in Figure 2.3. The direction of the z-axis, i.e. the direction of the beam is referred to as *longitudinal*. The $x - y$ plane, orthogonal to the beam line, is called *transverse* plane. Based on these definitions, the momentum of a particle can be divided into two components: the *longitudinal momentum* p_z and the *transverse momentum* p_T, defined as

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{2.3}$$

The *rapidity* (y) of a particle of total energy E is defined as

$$y = \frac{1}{2} \times ln \frac{E + p_z}{E - p_z} \tag{2.4}$$

Rapidity has the property of being additive under Lorentz boosts along the z direction, i.e. it is simply shifted by a constant when subjected to such transformations. For high energy or massless particles, rapidity can be approximated by *pseudorapidity*(η)

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

which only depends on the polar angle θ of the particle momentum, i.e. its angle with respect to the z-axis. The angle ϕ is measured up from the x-axis in the $x - y$ plane. We have $\eta = 0$ for the particles moving perpendicular to the z-axis and $\eta = \pm\infty$ is in $\pm z$-direction. The distance between two particles is defined in the form of a Lorentz invariant ΔR variable:

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

Detectors are installed at the collision points around the LHC ring. There are six experiments installed at the LHC, which are: ALICE (A Large Ion Collider Solenoid) [30], ATLAS (A Toroidal LHC ApparatuS) [31], CMS (Compact Muon Solenoid) [32], LHCb (Large Hadron Collider beauty) [33], LHCf (Large Hadron Collider forward) [34] and TOTEM (TOTal Elastic and diffractive cross-section Measurement) [35]. ALICE, AT-

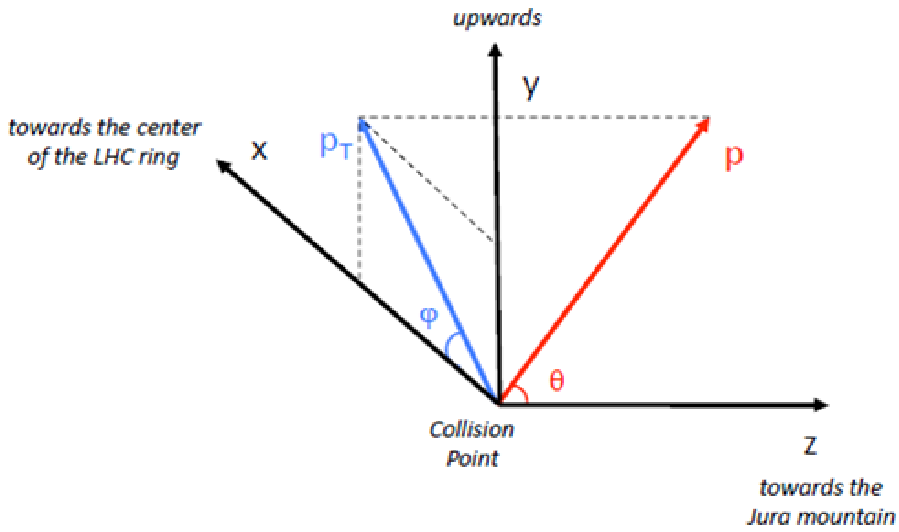


Figure 2.3: The CMS coordinate system.

LAS, CMS and LHCb are installed in four huge underground caverns built around the four collision points of the LHC, while LHCf and TOTEM are placed close to the two detectors, ATLAS and CMS respectively. ATLAS and CMS are the two general purpose experiments at the LHC, with the same physics goals but different technical solutions and design. The various detectors of LHC machine are shown in Figure 2.4. This Thesis presents the analysis of data delivered by the LHC and recorded by the CMS detector during the years 2011 and 2012.

2.2 The Compact Muon Solenoid Experiment

The CMS detector is one of the two general purpose detectors at LHC. Various components of the CMS detector were primarily built on surface before being lowered into the detector cavern, about 100 m below ground. The CMS detector sits at the LHC "Point 5", located near French village Cessy, between Lake Geneva and the Jura mountains. In the LHC ring, this is almost on the opposite side of ATLAS. The prime motivation of LHC is to explore the nature of electroweak symmetry breaking because of Higgs mechanism and search for the Higgs boson. Along with that, searches for new physics beyond SM and

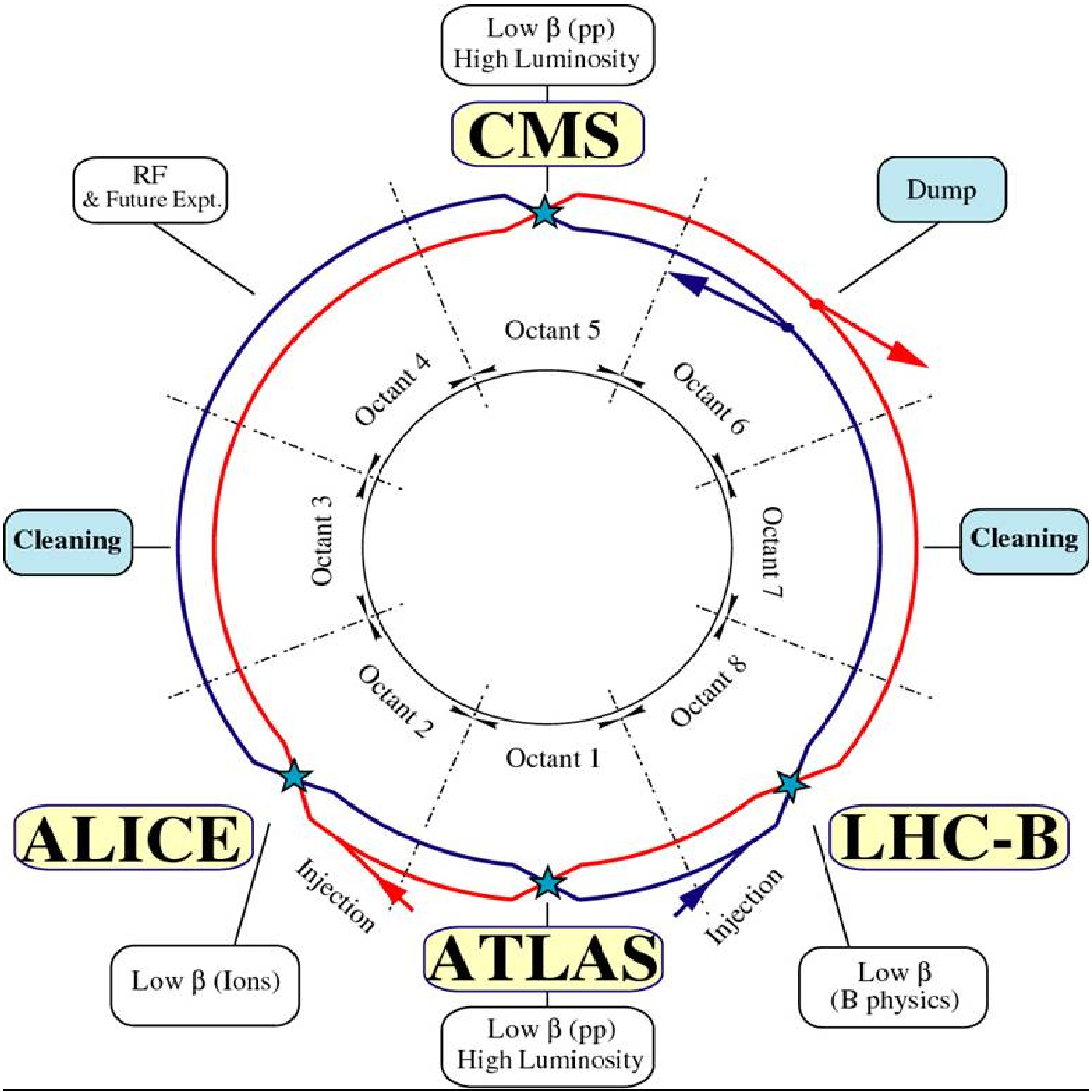


Figure 2.4: The various detector systems at the LHC.

precision measurements of already known physics at TeV scale are also the main highlights of LHC physics program. Detector requirements for the CMS experiment to achieve these goals of LHC physics program can be summarized as:

- A highly efficient and high momentum resolution inner tracker to reconstruct and measure the charged-particle momentum. Efficient triggering and tagging of τ leptons and b-quark jets, require pixel detectors close to the interaction region.
- Good electromagnetic energy resolution, good di-photon and di-electron mass resolution ($\approx$ 1% at 100 GeV), wide geometric coverage, π^0 rejection, and efficient

photon and lepton isolation at high luminosities.

- Good missing transverse energy and dijet mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

- Good muon identification and momentum resolution over a wide range of momenta and angles, good di-muon mass resolution ($\approx 1\%$ at 100 GeV), and the ability to determine the charge of muons with momentum up to 1 TeV.

The design of the CMS detector meets these requirements. The main distinguishing features of the CMS are a 4 Tesla superconducting solenoid which allows a compact design with a strong magnetic field, a full silicon-based high quality inner tracking system, a homogeneous scintillating-crystal-based electromagnetic calorimeter, a large hermetic hadron calorimeter and a robust and redundant muon system. The overall structure of the CMS detector consists of several cylindrical layers coaxial to the beam axis, closed at both ends by detector disks orthogonal to the beam direction to have full geometrical coverage. The overall length of the CMS detector is 21.6 m, the diameter is 14.6 m and the total weight is about 12500 tonnes. The overall layout of CMS is shown in Figure 2.5 and is described below.

2.2.1 The Superconducting Magnet

An important aspect driving the detector design and layout is the choice of the magnetic field configuration for the measurement of the momentum of muons and other high energy charged particles. Large bending power is needed for this purpose which forces a choice of superconducting technology for the magnets. The superconducting magnet for CMS has been designed to reach 4 Tesla field in a cylindrical region of 6 m diameter and 12.5 m length with a stored energy of 2.6 GJ at full current [36]. The magnet coil is large enough to accommodate the inner tracker and the calorimetry inside. A layout of magnetic field is shown in (Figure 2.6). The CMS magnet operates at a temperature of 4 K, ensured by a

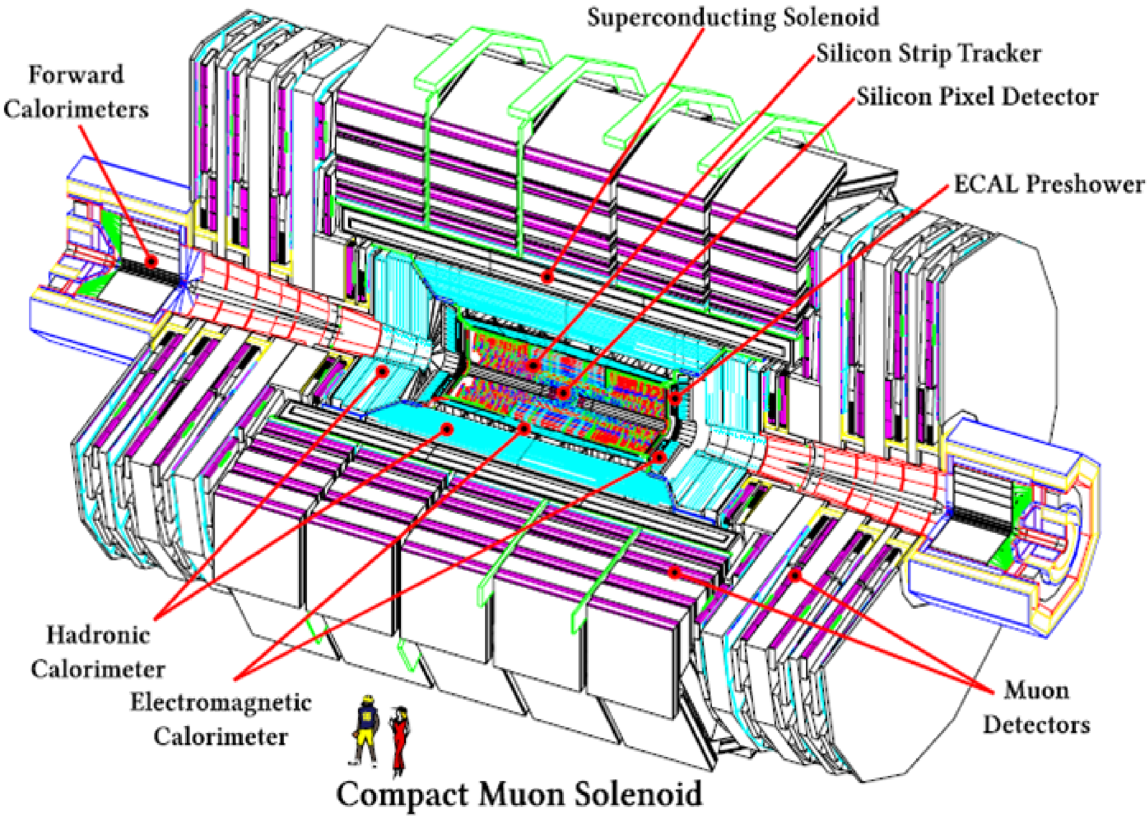


Figure 2.5: A Perspective view of the CMS detector.

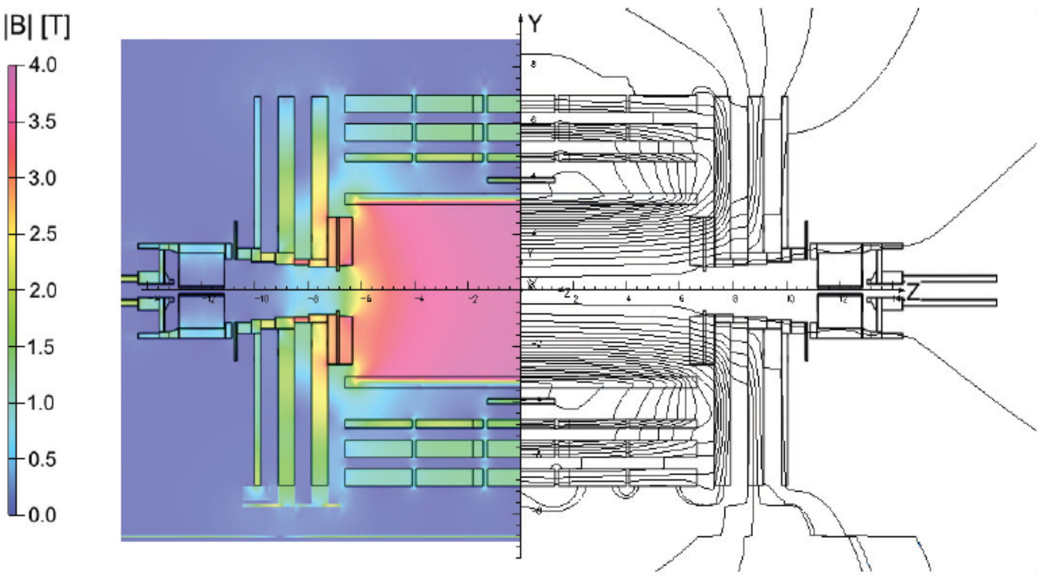


Figure 2.6: Layout of the magnetic field of CMS.

complex helium cooling system. At such temperature, the flat Niobium-Titanium (NbTi) cable becomes superconducting, allowing a 20 kA current to flow without appreciable loss. The whole magnet is contained in an enormous vacuum cylinder which isolates it from the external environment. Outside the magnet, an iron structure composed of five barrel layers and three disks for each endcap constitutes the iron yoke, which is needed to bridge the return magnetic flux. This provides a large bending power of 12 Tesla·m before the muons bending angle is measured by the muon system.

2.2.2 The Tracker

Tracker [32] is the innermost subdetector of the CMS and is closest to the interaction point. It is designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions, as well as a precise reconstruction of secondary vertices. It surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m. The tracker covers a pseudorapidity range of $|\eta| < 2.5$. The CMS solenoid provides a homogeneous magnetic field of 4 Tesla over the full volume of the tracker. The layout of the CMS tracker is shown in Figure 2.7.

The tracker is further subdivided into two parts, the pixel tracking detector and the strip tracking detector.

Pixel Detector: The pixel system is part of the tracker and is closest to the interaction region. It contributes to precise tracking points in r-ϕ and z and hence responsible for a small impact parameter resolution that is important for good secondary vertex reconstruction. It consists of three cylindrical layers pattern in the barrel region at the radial distance of 4.4 cm, 7.3 cm, 10.2 cm, with a total length of 53.3 cm, and two disks in each endcap region, at $|z| = 34.5$ cm and 46.5 cm, with radii ranging from 6 to 15 cm, as shown in Figure 2.8.

It consists of 65 million pixels, each with dimension of 100×150 μm^2 with around 47 million in barrel and 18 million in endcap region. The pixel detector delivers the spatial resolution between 10 -20 μm. As charged particles pass through the silicon pixels, they

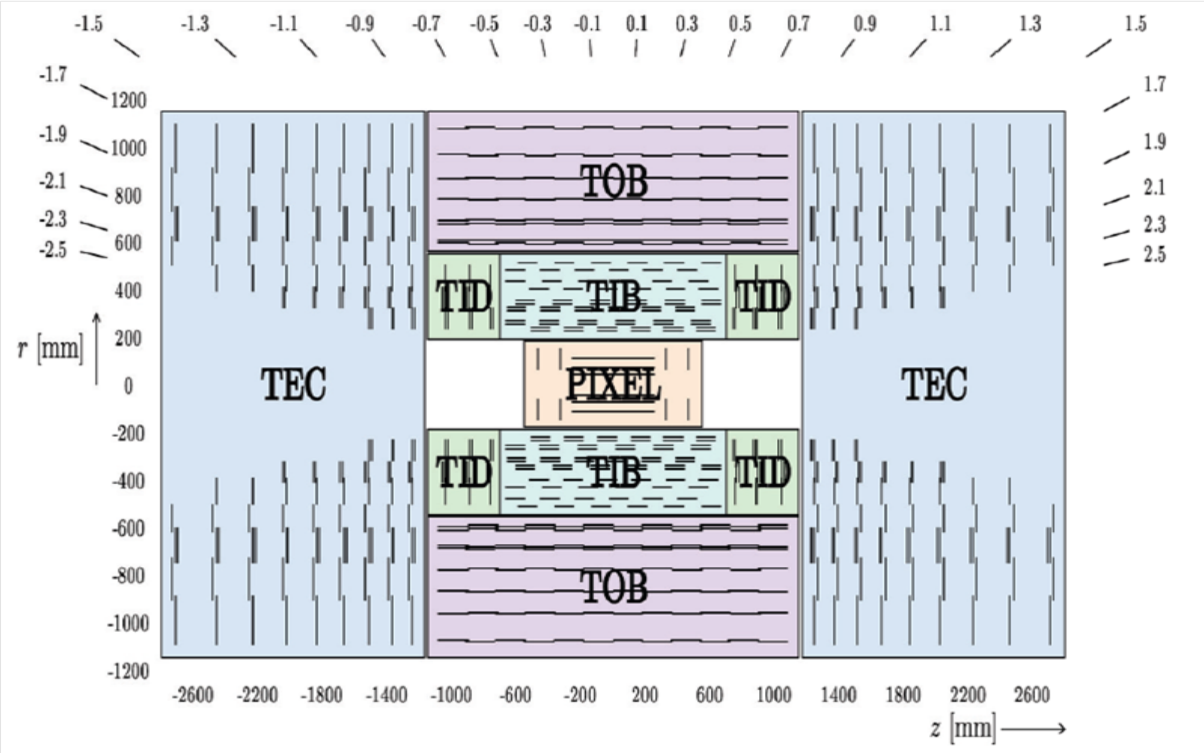


Figure 2.7: Schematic cross-section through the CMS tracker, each line represents a detector module and double lines indicate back-to-back modules which deliver stereo hits.

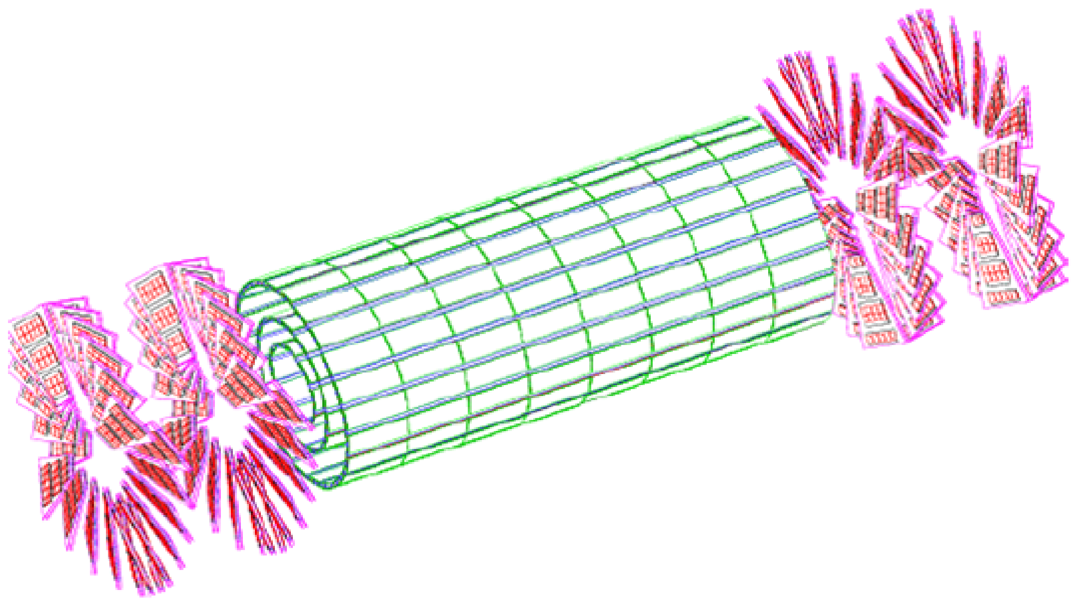


Figure 2.8: Layout of the CMS silicon pixel detector.

create small, detectable currents by the excitation of electrons to the conduction band of the silicon. The currents produced are then modified by the electronics.

Silicon Microstrip Detector: The silicon strip detector layout is shown in Figure 2.7. The inner region is made of 4 barrel layers (tracker inner barrel or TIB) and 3 disks at each side (tracker inner disks or TIDs). The outer system consists of 6 barrel layers (tracker outer barrel or TOB) and 9 disks for each endcap (tracker endcaps or TECs). It covers a radial region between 20 and 120 cm and the pseudorapidity region $|\eta| < 2.5$. All four regions (TIB, TID, TOB, TEC) are provided with both single-sided and double-sided microstrip modules. The strips are oriented along the z direction in the barrel and along the r coordinate in the endcaps. The microstrip silicon detector is designed to provide a spatial resolution of about 40 - 60 μm in the r-ϕ plane and about 500 μm along z. The occupancy is lower than 1%.

2.2.3 The ElectroMagnetic Calorimeter (ECAL)

The electromagnetic calorimeter of the CMS detector is a hermetic homogeneous calorimeter [37–39] made of 61200 lead tungstate ($PbWO_4$) crystals mounted in the central barrel part, closed by 7324 crystals in each of the two endcaps. A preshower detector is placed in front of the endcap crystals. ECAL plays an essential role in the study of the physics of electroweak symmetry breaking, particularly through the exploration of the Higgs sector. The search for the Higgs boson at the LHC strongly relies on information from the ECAL: by measuring the two-photon decay mode for $M_H \leq 150$ GeV, and by measuring the electrons and positrons from the decay of Ws and Zs originating from the $H \rightarrow ZZ^{(*)}$ and $H \rightarrow WW^{(*)}$ decay chain for 115 GeV $\leq M_H \leq$ 700 GeV.

The ECAL is also an important detector element for a large variety of the SM and other new physics processes. The reconstruction of a background-free $Z \rightarrow ee$ data sample is desirable for any new high mass object with one or more Zs in the subsequent decay chain. It is also crucial for other measurements such as supersymmetric searches, or the leptonic decay of new heavy vector bosons (W', Z') in the multi-TeV mass range. The use of high density crystals has allowed the design of a calorimeter which is fast, has fine granularity and is radiation resistant, which are all the important characteristics in the LHC environment. Following are the main features of $PbWO_4$ crystals:

- The high density (8.28 g/cm^3), short radiation length X_0 (0.89 cm) and small Moliere radius (2.2 cm) results in a fine granularity and a compact calorimeter as limited by the magnet radius;
- Scintillation decay time is almost consistent with the nominal LHC bunch crossing rate; about 80% of the light is emitted in 25 ns.
- It is sufficiently radiation hard, allowing it to sustain several years of high luminosity running with tolerable degradation of the crystals transparency, which can be corrected with a light monitoring system.

The barrel section (EB) has an inner radius of 129 cm. It is structured as 36 identical "supermodules", each covering half the barrel length and corresponding to a pseudorapidity interval of $|\eta| < 1.479$. The crystals are quasi-projective (the axes are tilted at 3^o with respect to the line from the nominal vertex position) and cover 0.0174 (i.e. 1^o) in $\Delta\phi$ and $\Delta\eta$. The crystals have a front face cross-section of about 22×22 mm^2 and a length of 230 mm, corresponding to 25.8 X_0. The endcaps (EE), at a distance of 314 cm from the vertex and covers a pseudorapidity range of $1.479 < |\eta| < 3.0$. The endcap consists of identically shaped crystals grouped in mechanical units of 5×5 crystals (supercrystals, or SCs) consisting of a carbon-fibre alveolar structure. Each endcap is divided into 2 halves, or *Dees*. Each Dee holds 3662 crystals. The endcap crystals, like the barrel crystals, are off-point from the nominal vertex position, and are arranged in an x-y grid (i.e. not an η-ϕ grid). They are all identical and have a front face cross-section of 28.6 × 28.6 mm^2 and a length of 220 mm (24.7 X_0). In addition, a preshower detector is used to provide $\pi^0 - \gamma$ separation having strong photon identification ability, and for the vertex reconstruction for photons. It is present in endcap region in front of EE. It consists of two lead radiators and of two planes of silicon strip detectors, for a total radiation length

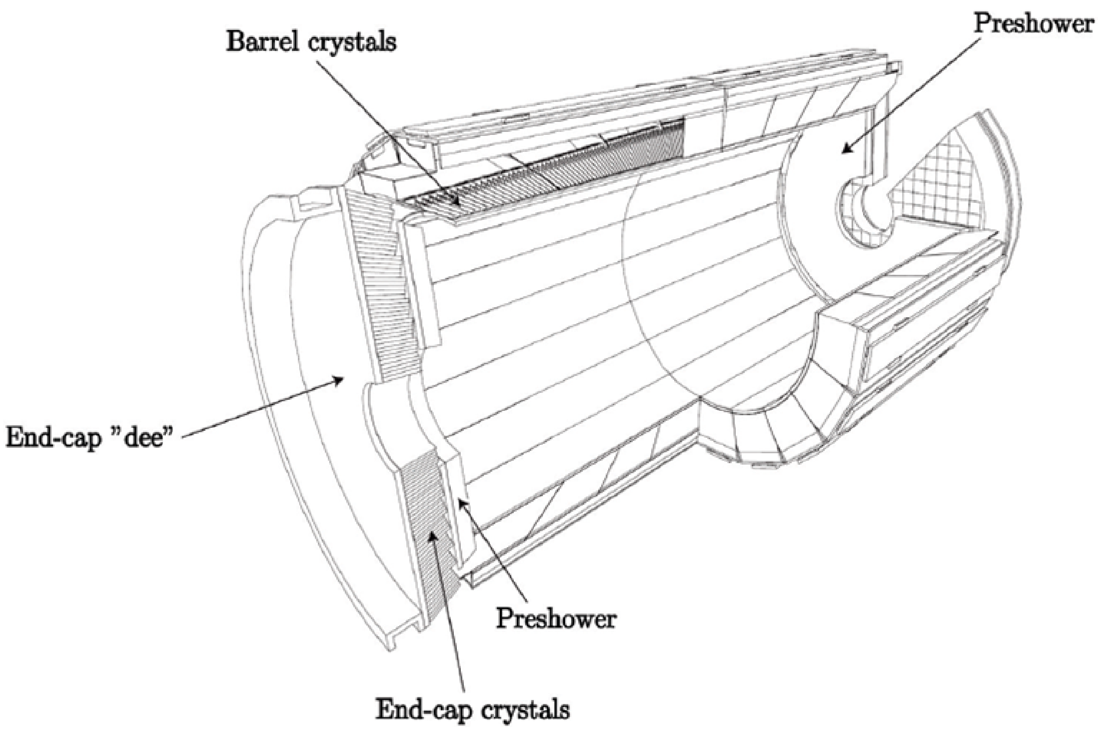


Figure 2.9: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps with the preshower detector in front of it.

of $3X_0$ and it allows the use of slightly shorter crystals of 220 mm in EE. It covers the pseudorapidity range of $1.653 < |\eta| < 2.6$. University of Delhi has played a major role in fabricating and installing this detector along with Bhabha Atomic Research Center (BARC), Mumbai. The layout of ECAL is shown in Figure 2.9. The longitudinal view of ECAL is shown in Figure 2.10. The relatively low light yield of the crystals (about 30 γ/MeV) requires photodetectors with intrinsic high gain that can operate in a magnetic field. Silicon Avalanche Photodiodes (APDs) and Vacuum Phototriodes (VPTs) are used to collect the scintillation light in the barrel and in the endcaps, respectively. The energy resolution of a calorimeter [40] can be parametrised as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \tag{2.7}$$

where

- a is called stochastic term which represents the statistical error coming from elec-

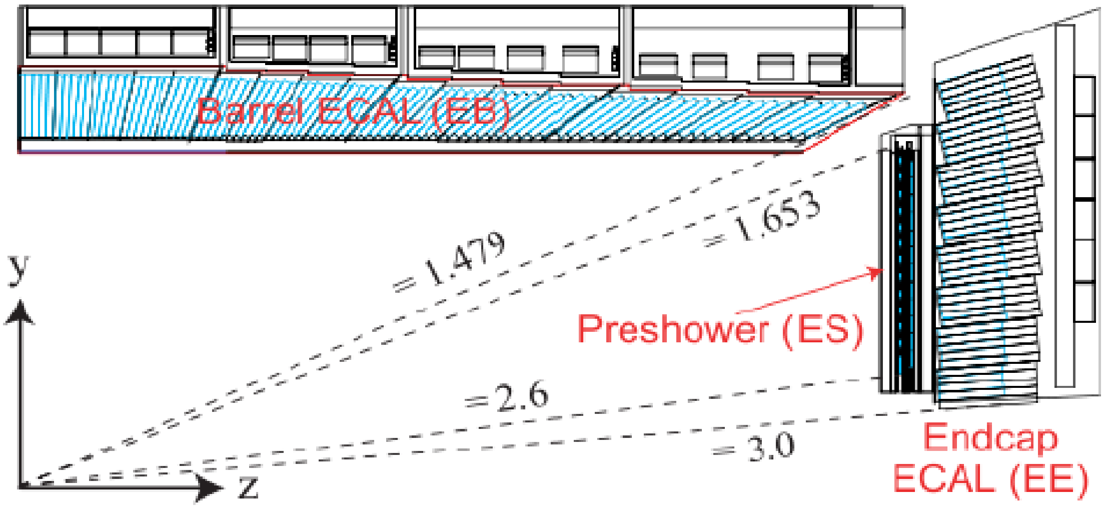


Figure 2.10: Longitudinal view of a part of CMS ECAL showing the barrel and endcap with preshower in front.

tromagnetic shower evolution.

- b is the error in energy measurement due to noise from the electronics and pile-up[1].

- c is a constant term related to the calibration of the calorimeter.

The study fitted the measured resolution and found for the ECAL the following values: a = 2.8% of $\sqrt{E}$ in units of $\sqrt{GeV}$, $b = 0.0415$ GeV and $c = 0.3\%$. The contributions change at different pseudorapidities and with the luminosity of the machine. The constant term c is the dominating term at high energies and it includes many different contributions. The most important among them are following:

- the stability of the operating conditions, like high voltage and temperature.

- transparency loss of crystals.

- presence of dead material between the crystals.

- the inter-calibration errors.

[1]at LHC, one single bunch crossing may have multiple inelastic collisions of protons, which are called pile-up events.

2.2.4 The Hadron Calorimeter

A schematic view of HCAL is shown in Figure 2.11. The hadron calorimeter (HCAL) [32] plays an important role in measurements of hadrons by measuring energy and direction of jets and missing transverse energy (E_T^{miss}). An important requirement of HCAL is to minimize the non-Gaussian tails in the energy resolution of jets and to provide good containment and hermeticity for the E_T^{miss} measurement. Hence, the HCAL design maximizes material inside the magnet coil in terms of interaction lengths. Another requirement is high hermeticity. In particular, the HCAL angular coverage must include the very forward region, since the identification of forward jets is very important for the rejection of many backgrounds. Therefore, CMS HCAL is divided into 4 regions, which provide a good segmentation, a moderate energy resolution and a full angular coverage up to $|\eta|$ of 5.

Hadron Barrel: The hadron barrel (HB) part of HCAL consists of 32 towers covering the pseudorapidity region -1.4 < η < 1.4, resulting in 2304 towers with a segmentation $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. These are the sampling calorimeters consisting of plastic scintillators as active material inserted between copper absorber plates. The absorber plates are 5 cm thick and plastic scintillator tiles are 4 mm thick. The scintillator tiles are read out individually by green wavelength shifting fibers coupled to clear plastic fibers which transport light to the transducers outside the calorimeter. Since calorimeter is in a high magnetic field, Hybrid Photo Diodes (HPD) are used for readout.

Hadron Endcap: Each hadron endcap (HE) of HCAL consists of 14 η towers with 5^o ϕ segmentation, covering the pseudorapidity region $1.3 < |\eta| < 3.0$. The granularity in the endcap is $\Delta\phi \times \Delta\eta = 0.17 \times 0.17$. The total number of HE towers is 2304. Both HB and HE sit inside the 4 Tesla magnetic field provided by the CMS solenoid magnet.

Hadron Outer: The hadronic outer (HO) is located just behind the HB and outside of the solenoid. HO uses iron as the absorber material, acts as tail catchers that effectively increases the thickness of the calorimeter in the central pseudorapidity region and ensures that no energy leaks out at the back of the HB. It is divided into 5 sections along η, called

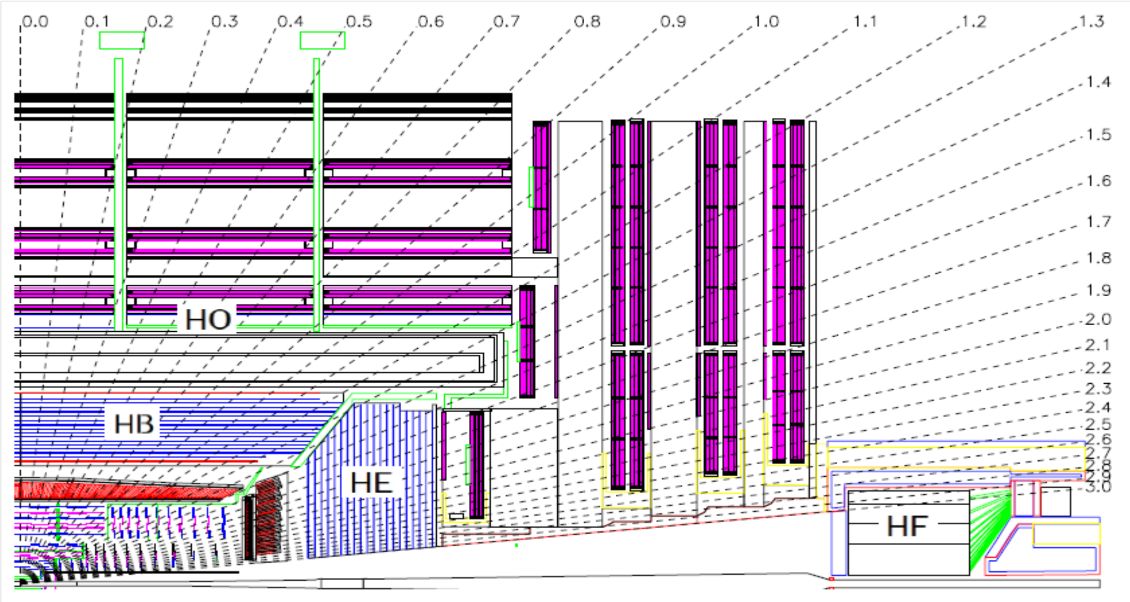


Figure 2.11: Schematic of one-quarter of the CMS HCAL system, in the r-z view, showing the locations of HB, HE, HO and HF.

"rings" -2, -1, 0, 1, and 2. One of the major contribution in fabrication of Scintillation Counters for the HO is by the Panjab University High Energy physics group along with Tata Institute of fundamental research (TIFR), Mumbai.

Hadron Forward: Coverage between pseudorapidities of 3.0 and 5.0 is provided by the steel/quartz fibre Hadron Forward (HF) [41] calorimeter. The HF calorimeter is useful to identify and reconstruct the very forward jets. It is Cerenkov light subdetector which is made up of quartz fibres embedded within a long steel absorber. The HF calorimeter is located about 11 m away from the interaction point. The HF calorimeter improves the measurement of the E_T^{miss} and enables very forward jets to be identified and reconstructed. The energy resolutions for different parts of HCAL are given in Equations 2.8 and 2.9 [42]:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{84.7\%\sqrt{GeV}}{\sqrt{E}}\right)^2 + (7.4\%)^2 \quad for\ HB,\quad HE\ \ and\ \ HO \tag{2.8}$$

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{198\%\sqrt{GeV}}{\sqrt{E}}\right)^2 + (9.0\%)^2 \quad for\ HF \tag{2.9}$$

2.2.5 Muon System

Muon detection is one of the main task of the CMS detector, as muons appear in many interesting decays of new physics or in Higgs boson decay signatures that can be produced by the LHC, like the one that is studied in this Thesis. As muons can penetrate several meters of material without interacting, they are not stopped by any of the CMS calorimeters. The muon spectrometer [43] is placed outside the solenoid magnet and is embedded in the iron return yoke, so that the 1.8 Tesla average magnetic field can be used as bending field to enhance muon p_T measurements. It covers the pseudorapidity region $|\eta| < 2.4$ and it performs three main tasks: triggering on muons, identifying muons, and assisting the tracker in measuring the momentum and charge of high p_T muons.
Three types of gaseous detectors (see Figure 2.12) are used to identify and measure muons. The choice of the detector technologies has been driven by the very large surface to be covered and by the different radiation environments. The reason for these different technologies lies in the different particle rates and occupancies (both higher in the endcaps) and in the intensity of the stray magnetic field, which is lower in the barrel. In the barrel region ($|\eta| < 1.2$), where the neutron induced background is small, the muon rate is low and the residual magnetic field in the chambers is low, Drift Tube (DT) chambers are used. In the two endcaps, where the muon rate as well as the neutron induced background rate is high, and the magnetic field is also high, Cathode Strip Chambers (CSC) are deployed and they cover the region up to $|\eta| < 2.4$. In addition to this, Resistive Plate Chambers (RPC) are used in both the barrel and the endcap regions. These RPCs are operated in avalanche mode to ensure good operation at high rates. They produce a fast response, with good time resolution but coarser position resolution than the DTs or CSCs. They also help to resolve ambiguities in attempting to make tracks from multiple hits in a chamber.

Drift Tubes Chambers: In the barrel region, the expected occupancy is low (< 10 Hz/cm^2), allowing for the use of DTs as detection element. In Figure 2.13, a section of a drift tube cell is illustrated. The DT chamber design is very redundant. Each cham-

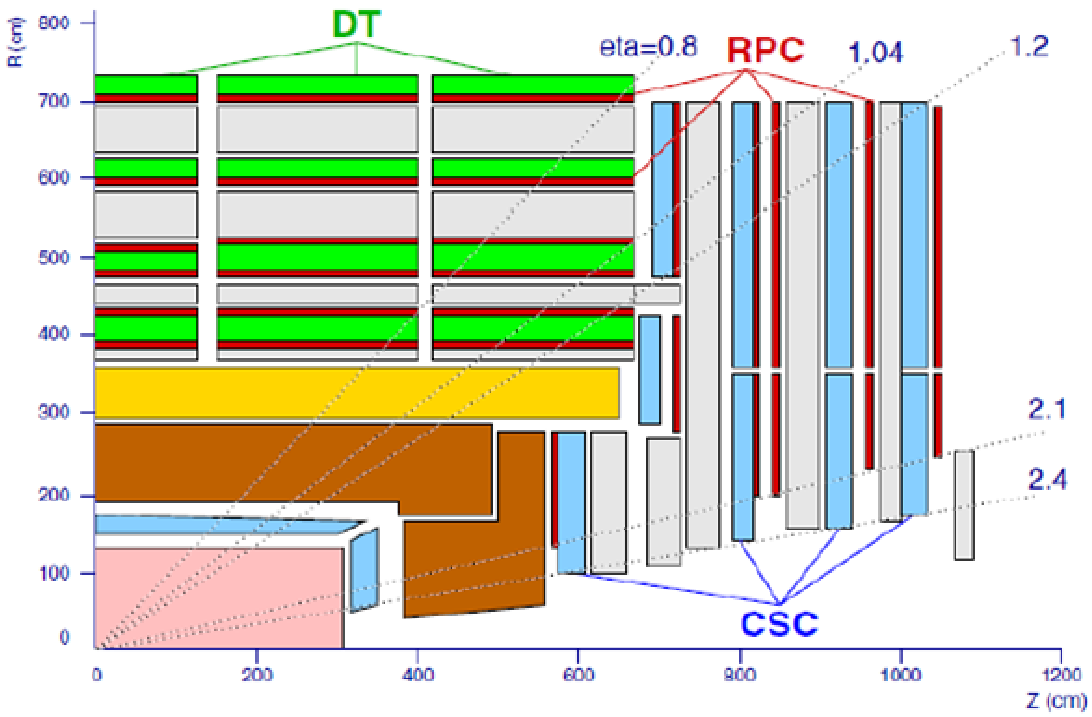


Figure 2.12: Layout of the CMS muon system.

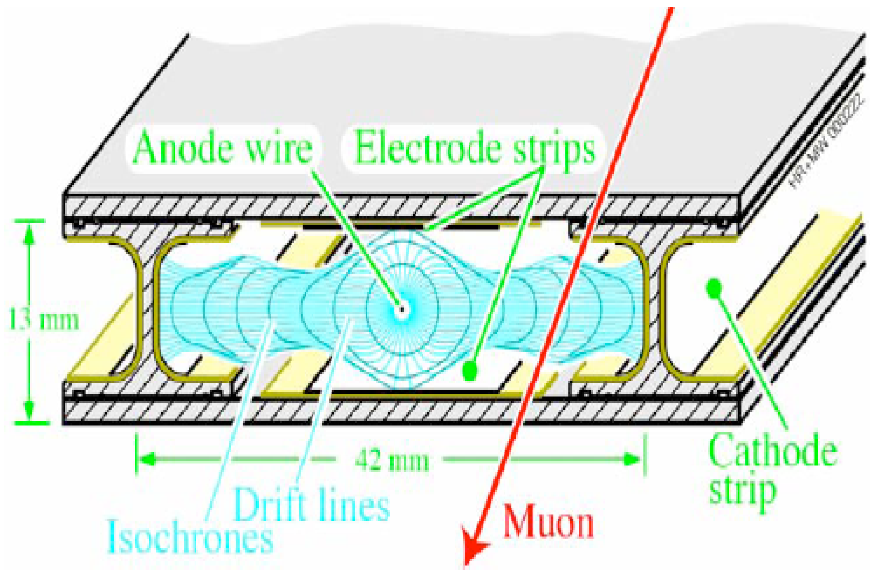


Figure 2.13: Section of a drift tube cell. The drift lines of the electrons are shown, along with isochrone lines which connects plates from which it takes the same time to drift to the anode. In the presence of a strong magnetic field these lines become distorted.

ber is made up of twelve layers of DTs, grouped in three independent subunits, called superlayers (SLs) [44]. Two of them have the anode wires parallel to the beam axis, to measure the transverse coordinate r-ϕ; the third SL is placed between the other two and is orthogonal to them, in order to determine the longitudinal coordinate z. The DTs are made of parallel aluminium plates, with cells obtained with perpendicular I-shaped aluminium cathodes. The anodes are 50 mm diameter steel wires placed between the cathodes. The internal volume is filled with a gas mixture of Ar (85%) and CO_2 (15%) at atmospheric pressure, because this gas is non-flammable and can be safely used in underground operations in large volumes, as required in the CMS.

Cathode Strip Chambers: Due to the larger occupancy of the endcap regions, from few Hz/cm^2 to more that 100 Hz/cm^2, and the intense and non-uniform magnetic field, CSCs have been chosen in this region. The CSCs have multiwire proportional chambers with one cathode plane segmented in strips running orthogonal to the wires [45]. CSC chambers are arranged in four disks (stations) placed between the iron disks of the yoke (Figure. 2.12). The innermost stations consist of three concentric rings, the first (ME1/1) being closer to the interaction point than the other two. The other stations are composed of two disks only. There are 36 chambers in each of the rings, i.e. ME±1/1, ME±1/2, ME±1/3, ME±2/2 and ME±3/2, and 18 chambers in each of the rings ME±2/1, ME±3/1 and ME±4/1 (Figure 2.14). The extra five chambers are mounted on the ME±4/2 ring. Each chamber (except for those in the ME±1/3 rings) overlaps slightly with its neighbour to provide contiguous coverage in ϕ and each successive station of chambers is positioned so that there is nearly complete coverage of muon detection on both endcaps from $0.9 < |\eta| < 2.4$. Each chamber is formed by six trapezoidal layers, with strips in the radial direction for a precise measurement of the azimuthal coordinate ϕ. Wires have a resolution of about 0.5 cm and strips have that of 5 cm. The full chamber spatial resolution varies from about 50 mm in the first CSC station to about 250 mm in the fourth. Inside each CSC, there are twelve alternating planes of metal strips that run through the length of the chamber with an array of anode wires between two cathode planes (the global

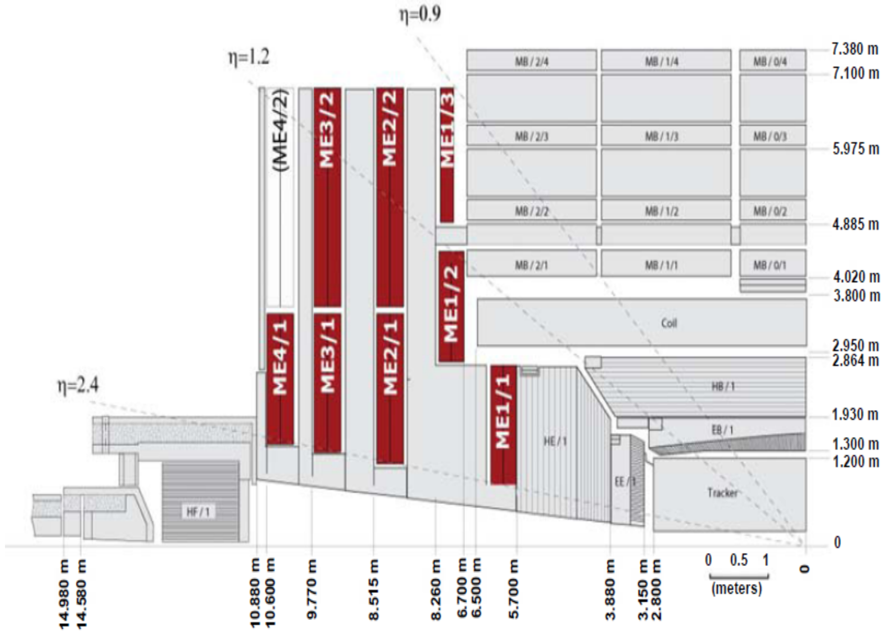


Figure 2.14: Quarter-view of the CMS detector. Cathode strip chambers of the Endcap Muon system are highlighted.

ϕ-direction). The gap is 9.5 mm thick and is filled with a mixture of 40% Ar, 50% CO_2 and 10% CF_4 gas. As a muon passes through the chamber, it ionizes the gas. The wires are kept at a potential of -3000 to +3600 V with respect to the strips, so that the freed electrons are strongly attracted to the anode wires which induces current in the cathode strips and this current is read out using on-chamber signal amplifiers and shapers. Studies of both collision and cosmic-ray data show that a pre-layer position resolution of about 50 μm in the r-ϕ direction and about 250 μm in the r direction are in good agreement with simulations. Figure 2.15 shows the internal working of the CSCs as explained above. The crucial characteristic of the DT and CSC subsystems is that they can each trigger on the p_{T} of muons with good efficiency and high background rejection, independent of the rest of the detector. The Level-1 trigger p_{T} resolution is about 15% in the barrel and 25% in the endcap.

Resistive Plate Chambers: Resistive plate chambers are installed both in the barrel and the endcap regions. They have a limited spatial resolution, but an excellent time resolution, thanks to their fast response of about 3 ns. This feature makes the RPCs

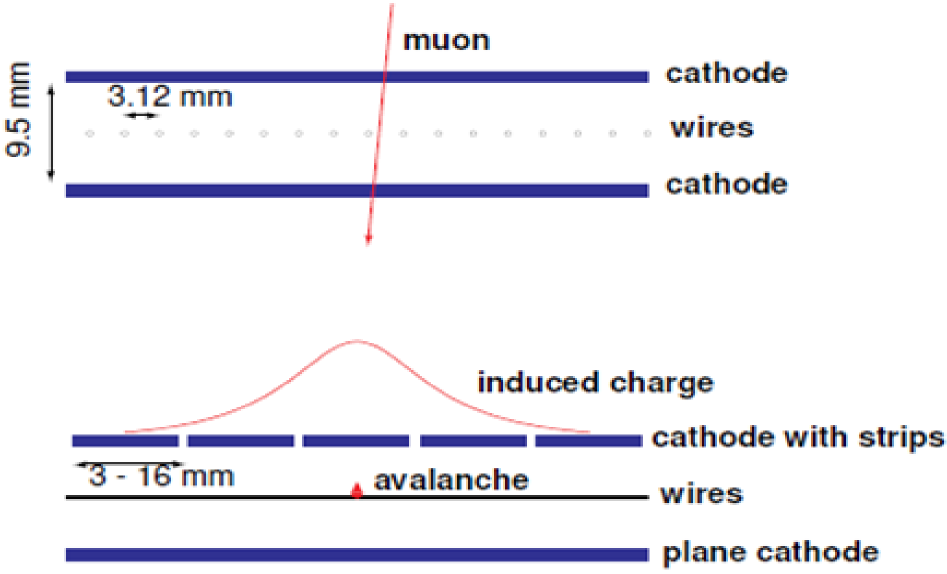


Figure 2.15: Orthogonal sections of a cathode strip chamber.

suitable for trigger purposes: they are used as a dedicated trigger subsystem, mainly for unambiguous bunch crossing identification [46]. The RPCs (Figure 2.16) used in the CMS are double-gap RPCs, made of four bakelite planes forming two coupled gaps 2 mm thick, filled with a $C_2H_2F_4$ and C_4H_{10} gas mixture. They operate in avalanche mode with common pick-up read-out strips in between. The total induced signal is the sum of the 2 single-gap signals. This allows the single-gaps to operate at lower gas gain (lower high voltage) with an effective detector efficiency higher than that for a single-gap. In order to suppress the noise rate due to discharges caused by the irregularities in the bakelite surfaces, the surfaces of bakelite are treated with linseed oil. Six layers of RPC chambers are embedded in the barrel iron yoke, 2 located in each of the first and second muon stations and 1 in each of the 2 last stations. The redundancy in the first 2 stations allows the trigger algorithm to perform the reconstruction always on the basis of 4 layers, even for low p_T particles, which may stop inside the iron yoke. In the endcap region, the baseline design foresees the instrumentation of the iron disks with 4 layers of RPCs to cover the region up to η of 2.1.

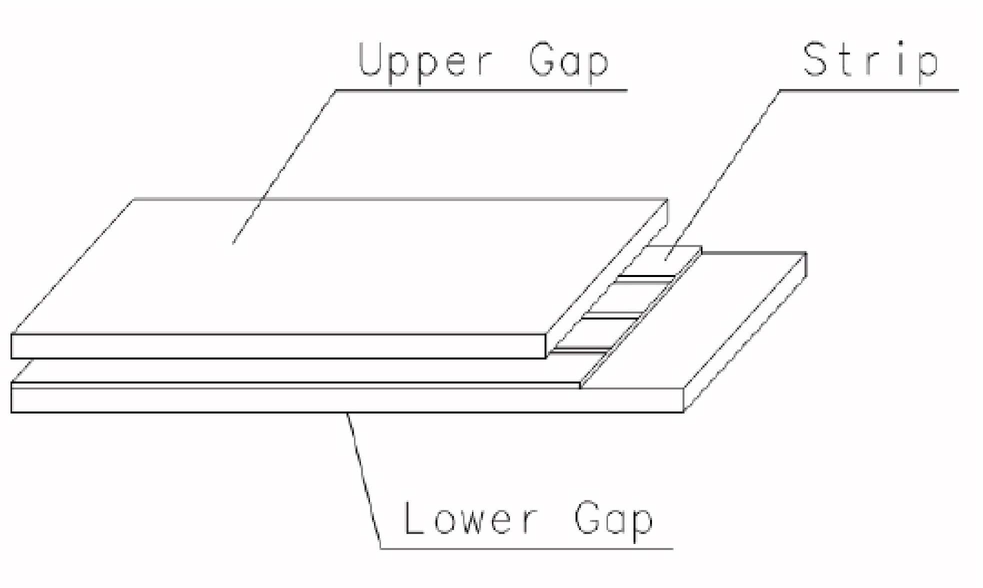


Figure 2.16: Layout of a double-gap RPC.

2.3 CMS Trigger and Data Acquisition System

The nominal rate of proton-proton collisions at the LHC is 40 MHz. Keeping in mind the typical event size for the CMS data (1 MB), a huge amount of data is delivered by the LHC. It is almost impossible to handle, store and process this much data so there is a need of reduction in this rate to make it equal to the offline storage limit, $O(100Hz)$. Hence, one of the most important and difficult aspects of the experiment is designing and implementing a Trigger system and Data Acquisition (DAQ) system to transfer this data to storage and for further processing.
The selection of events is based on their physics content, so that online algorithms must have a level of sophistication comparable to that of the offline reconstruction keeping the high efficiency for physics objects. The following section briefly describe the physics requirements of trigger system.

2.3.1 Physics Requirements

The CMS experiment is a general purpose experiment so a variety of physics searches and measurements carried out covering the low to high $p_{\rm T}$ physics. Because of that, CMS trigger system should be flexible and in a sense complete to cover the whole kinematic phase space. The CMS trigger system is based on the identification, kinematic and topological requirement on individual physics objects like muons, electrons, taus, photons, jets, and E_T^{miss} and on the combination of them. Also, since the LHC luminosity continued to increase, the evaluation of physics processes also changed with increasing luminosity. Lower p_{T} thresholds and removal or relaxation of thresholds were useful to maximize the physics output during lower luminosity. The physics requirements on the trigger are:

- In order to do searches for high p_{T} physics like Higgs Physics, SuperSymmetry (SUSY), and other exotic Beyond Standard Model (BSM) signatures, the CMS trigger system should be capable of selecting leptons and jets over the pseudorapidity range $|\eta| < 2.5$ and high p_{T} threshold like 20-40 GeV with high efficiency.

- Other than searches, there are measurements like cross-section measurements of Z boson, W boson, top-quark and other diboson processes. For that, one needs single leptons triggers with efficiencies greater than 95% in the pseudorapidity range $|\eta| < 2.5$ with a threshold of $p_T >$ 20-40 GeV.

- Searches where we have multileptons in the final state, dilepton and even trilepton triggers can be used. For these triggers the p_{T} thresholds are little lower than single lepton triggers like $p_T >$20 GeV and 15 GeV for the first and second lepton respectively.

- Single photon and diphoton triggers are required to have thresholds similar to those of the leptons.

- Single and multiple jet triggers along with E_T^{miss} triggers are required with a well defined efficiency over the entire η range of 5.0 in order to reconstruct high E_T^{miss} signals like SUSY etc.

- Triggers are also required for low p_{T} physics like b-quark physics. These triggers involve the requirements on combination of physics objects and also topological requirements like the leptons in the central region etc. In case the rates become too high then trigger prescales are applied on these triggers.

The CMS trigger has two levels: The Level-1 Trigger (L1) [47] and the High Level Trigger (HLT) [48].

2.3.2 Level 1 Triggers

The Level 1 (L1) Trigger, which is at the hardware level, reduces the rate of selected events down to 100 kHz for the high luminosity runs. A diagram of the L1 trigger system is shown in Figure 2.17. The L1 Trigger system is organized into three major subsystems: the L1 calorimeter trigger, the L1 muon trigger, and the L1 global trigger. Waiting for the trigger decision, data are stored in pipelines of processing elements. Within a latency period of 3.2 μs, if the L1 accepts the event, the data is moved to be processed by the HLT. Because of the high bunch crossing rate, the full readout of the detector is not possible. Only the calorimetric and muons information are employed. The Calorimeter Trigger identifies, from the shape of the deposited energy, the best four candidates of electrons and photons, central jets, forward jets and so on. The information about these objects is sent to the Global Calorimetric Trigger (GCT), together with the measured E_T^{miss}. The Muon triggering is performed independently by each of the muon subdetector, i.e. DTs, CSCs and RPCs. The DT and CSC triggers carry out a local muon reconstruction by comparing the slopes of track segments built in subsequent detector layers. The RPC trigger compares a given muon track with predefined hit patterns depending on the track p_{T}. The four best muon candidates are passed to the Global Muon Trigger (GMT) system which discards the low quality tracks after matching those from DTs and CSCs with those from RPCs. The L1 electronics is installed partly directly on the detectors, partly in the underground control room about 20 m far from the experimental cavern.

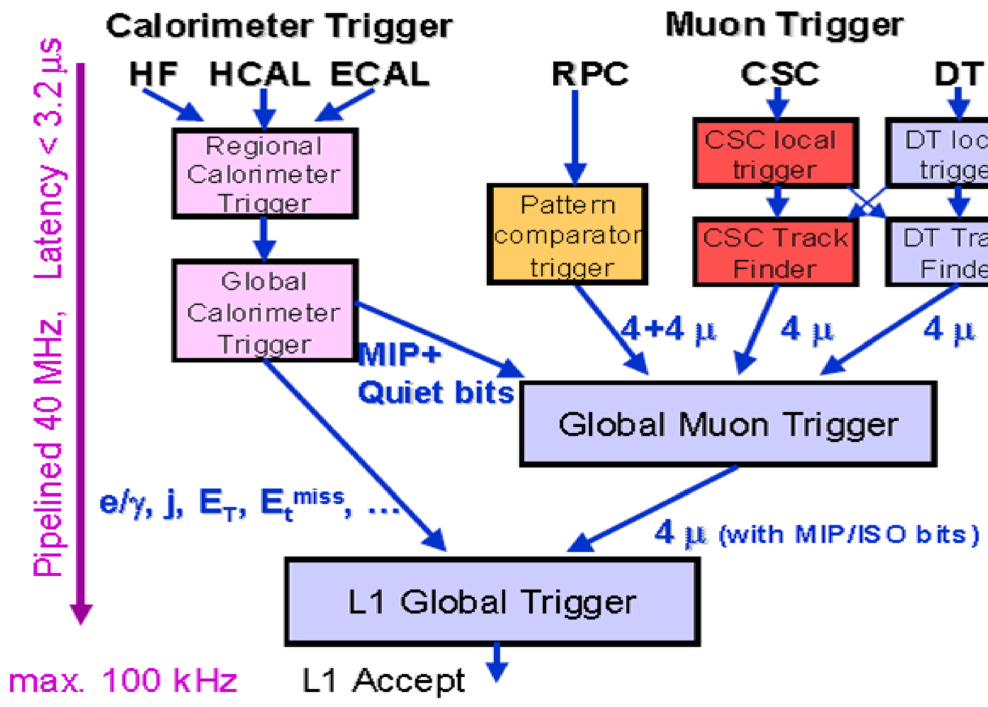


Figure 2.17: Overview of CMS L1 trigger.

2.3.3 The High Level Trigger

The subsequent reduction in the event rate for storage is obtained by software implemented HLT. The HLT reduces the rate of events from 100 kHz to $O(100Hz)$ that can be written on storage tape. The HLT processes all events accepted by the Level-1 trigger in a single processing farm containing 1000 CPUs. The reconstruction and selection in HLT takes place in steps and the selection is optimized step by step in order to reject events in each step. The basic idea is to reconstruct those part of each physics object that can be used for selection while minimizing the overall CPU usage.

The basic principles on which HLT works are: regional reconstruction and fast event rejection. Regional reconstruction tries to avoid the complete event reconstruction, which consumes lot of time, but rather focuses on the detector regions close to where L1 trigger has found interesting activity. Fast event rejection means that the uninteresting events are rejected as fast as possible; therefore emptying the buffers for next set of events. There are two virtual levels in HLT system: Level-2 trigger, which has input from Level-1 trigger system, considers information only from calorimeters and muon detectors. This is the first level of selection in HLT. Next is Level-3 where the information from tracker is included in the selection process of HLT. Because of the high number of channels and the complex pattern recognition, track reconstruction is a CPU consuming process. Hence, tracks are reconstructed only for events which pass Level-2 trigger. Final step is to read full event information from all subdetectors and then make final decision for the event. A list of reconstruction algorithms and filters for one or more physics objects that is triggered by L1 trigger is called *HLT Path*. For each luminosity scenario the set of trigger paths that, if enabled, contribute to a final OR of decisions which determines whether to reject or store an event and it is called, in the CMS jargon, *HLT Menu*. A single trigger path can require the presence of one or more physics objects of a particular type that pass specific kinematic thresholds, and it can also mix different type of physics objects (cross triggers). The rate of expected events for a particular set of trigger paths, given a luminosity scenario, is studied using simulation. With the increase of the instantaneous luminosity the requirements on the trigger paths became more and more stringent due to rate constraints, and the selection algorithms became more complex to cope with the requests of each physics analysis. Trigger paths with lower threshold than those necessary to face the event rate are kept in the HLT Menu with a prescale factor applied. This is done to have on storage the data sample required to measure and monitor the efficiency of the higher threshold trigger paths.

2.4 CMS Computing Model

Despite the excellent Trigger System, CMS still produces a huge amount of data that must be analyzed when running at peak performance. To meet this challenge, the LHC employs the Worldwide LHC Computing Grid (WLCG), a distributed computing and data storage infrastructure. It gives access to data to thousands of scientists all over the world in a highly distributed way. It starts in the Tier 0 centre at CERN where first and quick reconstruction of full data takes place for the data analysis. The CMS computing model [49] is highly distributed computing system, with a primary Tier-0 centre at CERN which is supplemented by Tier 1, Tier 2 and Tier3 computing centers at national laboratories and universities worldwide.

The Tier-0 Center (T0): is the first tier in the CMS model located at CERN. This is a common facility of data handling, data storage and grid. It performs well organized sequential processing of RAW data from the CMS detector and then divide them into primary datasets depending on the HLT information. It also transfers one copy of the RAW data to each of Tier-1 centers and one copy is kept at CERN. It also runs prompt calibration needed for quick reconstruction of data. T0 does not provide analysis resources and performs only scheduled activities.

Tier-1 Centers (T1): Once CERN has made a primary backup of the data it is then sent to large Tier-1 computer centers, located in seven countries around the world: France, Germany, Italy, Spain, Taiwan, UK and US. Tier-1 centers perform a mixture of tasks which include, organized sequential processing and more chaotic analysis activities. Since a Tier-1 center may have the only available copy of some data samples, any CMS user must be allowed to access it. However different users, or groups of users may have different priorities. These priorities are set by the CMS and by the physics analysis organization. CMS physicists perform selection, skims, reprocessing etc. on data that has been pre-located at the Tier-1 centers by CMS by using the Tier-1 center computers. Some local users may have long term local storage at the center while others would move the results to another computer center (typically a Tier-2 center). Some analyses are expected to be fully carried out at Tier-1 centers. There are likely to be groups of local users who use the Tier-1 in the usual way but who also have access to local storage and local batch and interactive facilities for their analysis activities. Most CMS users are expected to use a Tier-2 or Tier-3 center as their computing interface to CMS analysis.

Tier-2 Centers (T2): are more responsive to local priorities rather than central CMS requirements. Thus, Tier-2 centers are not generally obliged to respond to computing requests from all CMS users. Most of the Tier-2 facility is used by some groups of CMS

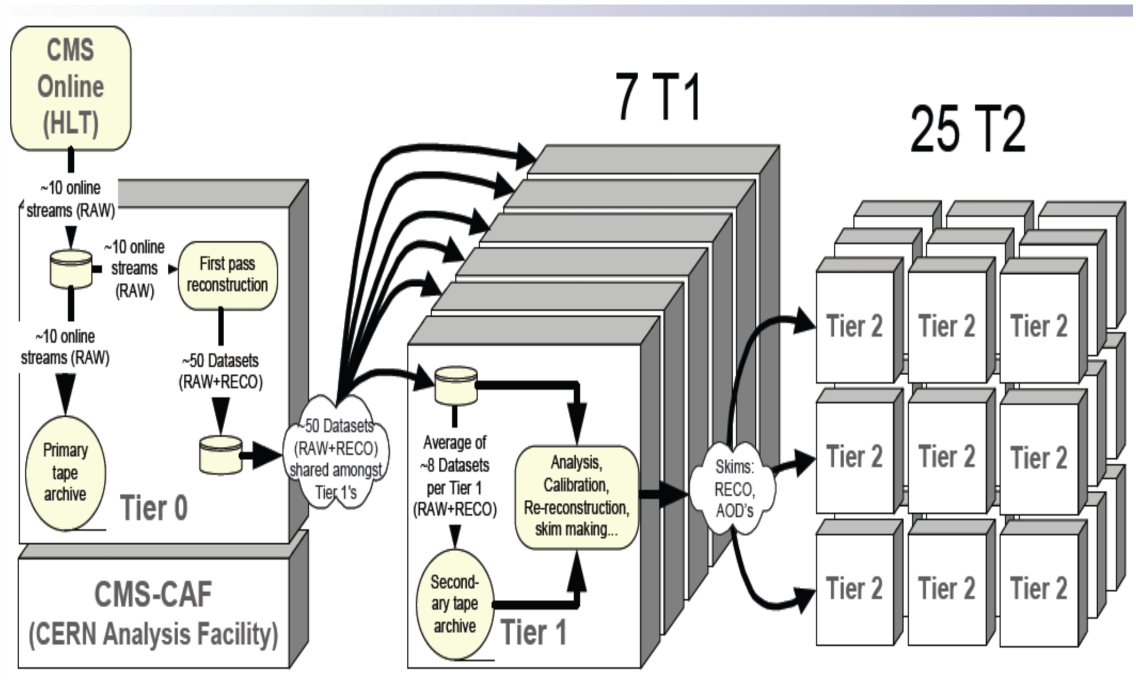


Figure 2.18: Event data flow in the CMS Computing System.

physicists to perform their iterative analyses. A signicant fraction of the resources of a Tier-2 center is, however, dedicated towards scheduled MC simulation of common CMS samples. CMS does not centrally mandate where skim data is kept. However the Tier-2 centers do have to facilitate access to their published data products so that other CMS physicists can analyse it. CMS has the responsibility to supply a solid basis to this analysis activity by assuring that consistently reconstructed datasets can be used by the collaboration.

Tier-3 Centers (T3): are modest facilities at institutes for local use. Such computing facility is not generally available for any coordinated CMS use but is valuable for local physicists. In this way information branches out from each Tier, and the analysis of the final data can be performed locally all around the world. The Event data flow in the CMS Computing System is shown in Figure 2.18.

2.5 CMS Software Components

The goals of the CMS Software (CMSSW) are to process and select events inside the HLT $farm$, to deliver the processed results to the experimentalists within the CMS collaboration and to provide tools for them to analyze the processed information and produce physics results. The overall collection of software, now referred to as CMSSW, is built around a Framework, an Event Data Model (EDM), and services needed by the simulation. The physics and utility modules are written by detector groups. The modules can be plugged into the application framework at run time, independently of the computing environment. The software is developed keeping in mind not only performance but also modularity, flexibility, maintainability, quality assurance and documentation. CMS has adopted an object-oriented development methodology, based primarily on the C++ programming language. The primary goal of the CMS Framework and EDM is to facilitate the development and deployment of reconstruction and analysis software. The EDM is centered around the Event class, which holds all data that was taken during a triggered physics event as well as all data derived from the data taking (e.g. calibration and alignment constants). The detailed CMS detector and physics simulation is currently based on the simulation toolkit GEANT 4 [50] and the CMS object-oriented framework and event model. GEANT 4 provides a rich set of physics processes describing electromagnetic and hadronic interactions in detail. It also provides tools for modeling the full CMS detector and geometry and the interfaces required for retrieving information from particle tracking through these detectors and the magnetic field. The validation of GEANT 4 in the context of CMS is described in detail in Ref. [51]. The CMS GEANT 4 based simulation program uses the standard CMS software framework and utilities, as used by the reconstruction programs. The simulation is implemented for all CMS subdetectors in both the central and forward region, including the field map of the 4 Tesla solenoid. In addition, several test-beam prototypes and layouts have been simulated. The full simulation program implements the sensitive detector behavior, track selection mechanism, hit collection and digitization (i.e. detector electronic response). The detailed simulation workflow is as follows:

- An appropriate Monte Carlo (MC) event generator (several are used) is configured to produce the data samples of interest;
- The production team/system runs the generator software to produce generator event data files;
- The generator data samples are then validated and a configuration for the GEANT based simulation of CMS is selected, with generator events as input, to produce (using the standard CMS framework) persistent hits in the detectors;
- These hits data are then validated and are used as input to the subsequent digitization step, allowing for pile-up to be included. This step converts hits into digitizations which correspond to the output of the CMS electronics.

The digitization step, following the hit creation step, constitutes the simulation of the electronic readout used to acquire data by the detector. It starts from the hit positions and simulated energy losses in the detectors and produces an output that needs to be as close as possible to real data coming from CMS. Information from the generation stage (e.g. particle type and momentum) is preserved in the digitization step. The output of this step has the same format of real collision events, and therefore can be fed to the same reconstruction software chain. Collision events are reconstructed and stored if they satisfy at least one of the High Level Trigger paths employed online. Depending on the type of HLT path which was passed, an event is stored in a given primary dataset, which therefore collect events with similar topologies. Examples of Primary Datasets are *Photon*, *DoubleMuon*, *DoubleElectron*, *SingleMuon* etc. In this Thesis, all of the mentioned primary datasets are used. To maintain the required level of data reduction while keeping the flexibility, the CMS uses several event formats with different levels of detail and precision. Other specialized event formats are used for heavy-ion data. The process of data reduction and analysis takes place in steps, typically carried out at different computer centers. The main types of data format are referred to as, RAW, RECO, and AOD. RAW events contain the full recorded information from the detector, and a record of the trigger decision and other information. The RAW data is permanently archived in safe storage, and is designed to occupy around 1.5 MB/event (2 MB/event for simulated data, due to additional MC generator level information). The Reconstructed format, RECO data, is produced by applying several levels of pattern recognition and compression algorithms to the RAW data. Reconstruction is the most CPU intensive activity in the CMS data processing chain. The resulting RECO events contain physics objects, and a full record of the reconstructed hits and clusters used to produce them. Sufficient information is retained to allow subsequent application of new calibrations or algorithms without recourse to RAW data. RECO events are foreseen to occupy around 0.5 MB/event. Finally, Analysis Object Data (AOD) is the compact analysis format which is designed to allow a wide range of physics analyses whilst occupying sufficiently small storage so that very large event samples may be held at many centers. AOD data is produced by filtering of RECO data, and requires around 100 kB/event, small enough to allow a complete copy of the experimental data in AOD format to be held at computing centers outside CERN. In this Thesis, the analysis has been performed using AOD data. The Physics Analysis Toolkit (PAT) is a high level analysis layer providing the Physics Analysis Groups (PAGs) with easy access to the algorithms developed by Physics Objects Groups (POGs) in the framework of the CMSSW offline software. It aims at fulfilling the needs of most CMS analyses, providing both ease of use for beginners and flexibility for advanced users. PAT is fully integrated into CMSSW and an integral part of any release of CMSSW. To facilitate development, CMS consolidates its code base regularly into releases. Different releases are grouped into release cycles dedicated to a specific purpose (i.e. data taking and reconstruction in LHC collision period, integration of a new ROOT version, etc.) to aggregate specific feature sets of the software stack. The rapid development cycle of CMS and the resulting high number of releases requires a thorough quality assurance process. To guarantee stable and performant releases while supporting a high turn around and diverse development, CMS implemented an advanced central release validation process.

2.5.1 CMS Remote Analysis Builder

CMS Remote Analysis Builder, known as CRAB [52], is a Python program intended to simplify the process of creation and submission of CMS analysis jobs into a computing grid environment. In order to store and manage a huge quantity of data, while assuring data access to physicists of CMS collaboration and computing power for analysis and simulation, CMS uses a distributed architecture based on grid infrastructure that guarantees resources and data availability to allowed users. The user's point of view involves analyzing the distributed data as they were in the local computing centre, but the analysis in a distributed environment is a complex computing task. Hence, the CMS collaboration has developed CRAB to simplify the work of users to create, submit and manage the analysis job in grid environments to handle data discovery, to provide resources availability, to create and submit job, to monitor status and to retrieve output. CRAB is a user friendly tool that helps users to run their analysis code on data available at remote sites, hiding grid and CMS infrastructure details.

Chapter 3

Physics Object Reconstruction and Identification at CMS

Reconstruction is the method of constructing physics quantities from the raw data collected in the experiment [53]. The reconstruction process can be divided into 3 steps: local reconstruction within an individual detector module, global reconstruction within a whole detector, and combination of these reconstructed objects to produce physics objects like electron, muon, jets etc. The reconstruction units providing local reconstruction in a detector module use raw data from the DAQ system or simulated data as input. These data in either case are called "digis". The output from the reconstruction units are called "RecHits", i.e. reconstructed hits which are typically position measurements (from times or clusters of strips or pixels) in tracking-type detectors (Muon and Tracker systems) and calorimetric clusters in Calorimeter systems. In the global reconstruction step, information from the different modules of a subdetector are combined, although information from different subdetectors is not grouped. For example, Tracker RecHits are used to produce reconstructed charged particle tracks and Muon RecHits are used to produce muon tracks. The final reconstruction step combines reconstructed objects from individual subdetectors to produce reconstructed objects suitable for physics analysis. For example, tracks in the Tracker system and tracks in the Muon system are combined to provide final muon candidates, and electron candidates from the Calorimeter system are matched to tracks in the Tracker system. In this chapter, an overview for reconstruction of physics objects which are used in the physics analyses is provided. After the reconstruction, identification of these physics objects is also explained. Identification of any physics object is a method to improve the separation of the 'real' object from 'fake' one. 'Fake' means the mis-reconstructed or mis-measured physics object, for example a jet deposits its energy both in ECAL and HCAL and hence it can be reconstructed as electron. Every physics object has a set of identification characteristics or variables which can be used to identify it. The selection on these variables is decided based on the signal efficiency and background (fake) rejection. In the case of searches for physics signals having very low cross-sections, one cares more about the efficiency for signal than the background rejection and hence the selection criteria is less stringent and it is generally called 'loose' criteria. Another case is precision measurements where more interest is to reject the background and hence a more stringent identification criteria is applied and it is generally called 'tight' criteria. In between of these two criteria, there exists a 'medium' criterion as well. In this chapter, Section 3.1 and 3.2 gives the description about the interaction vertex and beamspot reconstruction. In Section 3.3, a discussion is given for track reconstruction. Section 3.4 and 3.5 give the details about the reconstruction, identification and isolation for electrons and muons respectively. Section 3.6 describes briefly the photon reconstruction and identification. In the Section 3.7 various jet reconstruction algorithms are discussed along with their identification criteria. Section 3.8 describes the various missing transverse energy definitions. Finally, Section 3.9 and 3.10 give the overview of τ reconstruction and various b-tagging algorithms respectively.

3.1 Vertex Reconstruction

The identification of vertices plays an important role in event reconstruction. Precise coordinates of the primary event vertices are needed to assign tracks to collisions and to determine the event kinematics. Secondary vertices are tools for identifying long-lived particles like heavy flavor hadrons and τ leptons. Multiple overlapping events with high track density and particle interactions in the tracking volume are the main challenges for vertex reconstruction.

Vertex reconstruction usually involves two steps, vertex finding and vertex fitting. Vertex finding involves grouping tracks into vertex candidates, such as the tracks of a jet in case of quark flavor (b-quark or light quarks) tagging or the full event in case of primary vertex finding. Vertex finding algorithms can be very different for the different physics cases (primary or secondary vertex finding, reconstruction of exclusive decays, etc.). This is done via deterministic annealing [54] algorithm. Vertex fitting involves determining the best estimate of the vertex parameters for a given set of tracks. The goodness of fit may be used to accept or discard a vertex hypothesis. In addition the vertex fit is often used to improve the measurement of track parameters at the vertex. The CMS reconstruction uses two different vertex fitting methods. The first one is the Kalman Filter [55], which uses a global least-square minimization. It works best for Gaussian distributed measurements and may include an update of the involved track parameters. The Gaussian Sum filter (GSF) [56] treats tails by modeling the non-Gaussian resolution functions more precisely. The second method, the Adaptive Vertex Fitter [57], is an iterative re-weighted fit, which down-weights tracks according to their significant distance from the vertex. This enlarges the robustness by decreasing the sensitivity to outlying or mis-measured tracks.

Once a vertex is found and fit, the tracks are again recalculated, using the vertex as an additional position in the track fit algorithm. In order to separate the vertices associated with pile-up events with the interaction of interest, each event is given a primary vertex, defined to be the one with the highest sum of squared track transverse momenta.

3.2 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 the following reasons:

- Important input for physics, for example quark flavor tagging of jets, lifetime, etc.
- For pattern recognition (input for 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 calculated the transverse beam position is the so called $d_0\phi$ algorithm. The $d_0\phi$ algorithm has been used successfully in previous experiments [58].

3.3 Track Reconstruction in CMS

Different algorithms are used in CMS for track reconstruction. All methods use the reconstructed positions (hits) of the passage of charged particles in the silicon detectors to determine the helix trajectories of the charged tracks and therefore measure their directions and momenta. The standard algorithm designed for the reconstruction of proton-proton collisions is the Combinatorial Track Finder (CTF). The CTF proceeds in three stages: (1) seeding, (2) finding and (3) fitting. In the seeding stage pairs of hits, that are compatible with the interaction region above a lower p_{T} threshold, are considered as possible candidates of charged tracks. Pixel hits provide the best track seeding, given their three dimensional position information and lower occupancy. The seeding efficiency with pixel hits drops in the $2 < |\eta| < 2.5$ forward region where a mixed seeding of hits from pixels and inner strips is needed to achieve a fully efficient track finding in the whole tracker acceptance. The track finding stage is based on a Kalman Filter pattern recognition approach [55]. Starting with the seeded parameters, the track trajectory is extrapolated to the neighboring tracker layers and compatible hits are assigned to the track. Final Track fit module estimates the parameters of trajectories with ultimate precision by means of a Kalman filter and smoother. During the iterative reconstruction process, a lot of ambiguities are created in the case of multiple tracks per seed or multiple seeds per track. Trajectories are cleaned to reduce those ambiguities by examining the number of shared hits between track pairs. In the case of equal number of hits, the track with the highest pearson chi-square (χ^2) is discarded and if the number of shared hits is larger than 50% of the hits of the track with the lower numbers of hits, the track with the lower number of hits is discarded. Finally, track selection is performed using several tracking quality criteria, for example the number of hits, normalized χ^2 and vertex compatibility. The fake tracks, especially in the case of high multiplicity in jets are rejected using these quality criteria. In the central region $|\eta| < 1$, the p_{T} resolution is better than 1% for tracks with $p_{\mathrm{T}} \leq 10$ GeV [59]. At higher p_{T} the momentum resolution worsens approximately as $\Delta(1/p_{\mathrm{T}}) \sim 0.2\ TeV^{-1}$. The resolution on the transverse (Δd_0) and longitudinal (ΔZ_0) impact parameters in the central region is $\Delta d_0 \sim 10 \mu m$ and $\Delta z_0 \sim 20 \mu m$, for single muons of very high $p_{\mathrm{T}} > 100$ GeV. At lower p_{T} the track impact parameter resolution worsens approximately with an additional $\Delta d_0 \sim \Delta z_0 \sim 100\ \mu m/p_{\mathrm{T}}$ (GeV).

3.4 Electron Reconstruction and Selection

An electron in the CMS is defined as clusters of ECAL energy deposits matched to a track from the silicon tracker. Electron reconstruction in CMS [60] is complicated due to the large amount of material (Figure 3.1) that it must traverse before it reaches electromagnetic calorimeter resulting in high rates of bremsstrahlung as shown in Figure 3.2.

3.4.1 ECAL Clustering and Superclustering

Electron and photon showers deposit their energy in several crystals in the ECAL. Approximately 94% (97%) of the incident energy of a single electron or photon is contained in a 3×3 (5×5) crystal array (Figure 3.2). Summing the energy measured in such fixed arrays provides the best performance for photons or electrons in a field-free environment that excludes material in front of the crystals. However, in the realistic detector envi-

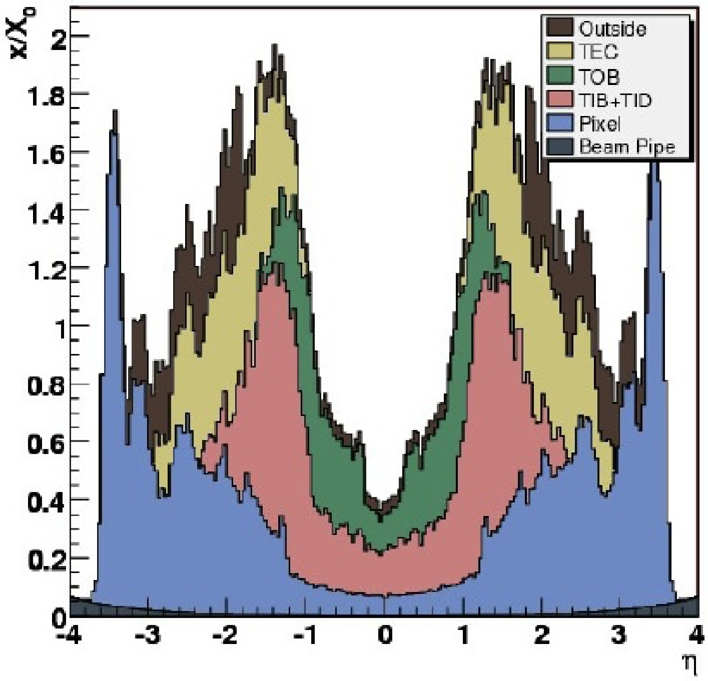


Figure 3.1: The distribution of tracker material in front of ECAL. η is shown in x-axis and y-axis shows the distance in terms of Radiation length (X_0).

ronment material in front of the calorimeter leads electrons to undergo bremsstrahlung and photons to convert, and the energy reaching the calorimeter is spread in ϕ due to the magnetic field. This spread in energy is recovered for electrons and photons by building a cluster of clusters of such energy depositions that is extended in ϕ, referred to as a "supercluster".

Superclustering Algorithms

Details of the superclustering algorithms used have been given elsewhere [61]. Here, only a brief idea description is given. The reconstruction of electrons starts with building clusters seeded by highest energy deposited crystals in the ECAL. These clusters are then used to form superclusters that further collect the energy radiated by bremsstrahlung in the tracker volume.

- **Hybrid supercluster algorithm:** This algorithm uses the $\eta - \phi$ geometry of the barrel crystals to exploit the knowledge of the lateral shower shape in the η direction (taking a fixed bar of 3 or 5 crystals in η), while searching dynamically for separated energy in the ϕ direction. The algorithm makes superclusters which are then decomposed into sub-clusters ("basic clusters").

- **Island basic cluster and supercluster algorithm:** The Island algorithm starts with a seed crystal having 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 (i.e. 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.

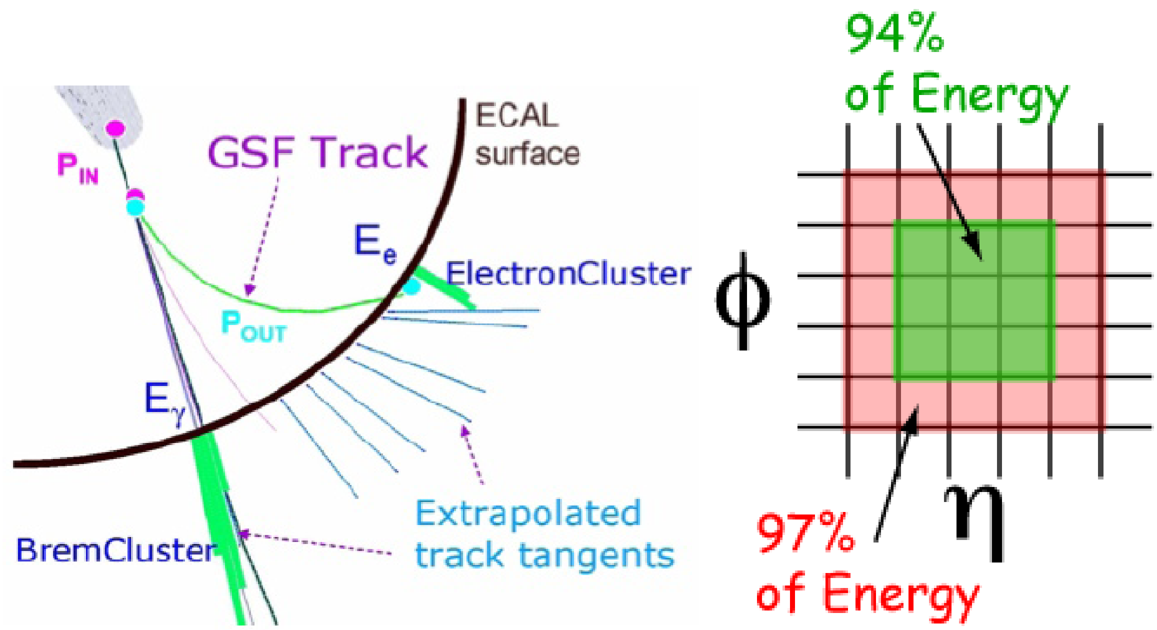
  For a crystal to be added to a cluster the following must be true:

  1. The crystal contains a RecHit with positive energy.
  2. The crystal has not been assigned to another cluster already.
  3. 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.

3.4.2 Electron Tracking and Charge determination

After the clustering, the supercluster is propagated back towards the pixel detector, accounting for the influence of the magnetic field, and matched with compatible track segments reconstructed in the pixel detector. Next, the electron track is built by proceeding through each layer of the silicon tracker starting from the pixel track segment, and adding measured hits in subsequent layer that are consistent with the propagated trajectory. The propagation and track fitting algorithm used is based on the Kalman filter [55]. To account for the high rate of bremsstrahlung and to maximize electron track finding efficiency,



(a) (b)

Figure 3.2: Reconstruction of electron candidates at CMS experiment.

looser track hit position compatibility requirements are imposed in the track building process.

The Kalman filter technique of track fitting is known to be optimal only if all energy loss distributions are Gaussian. The best reference model for electron energy loss is the Bethe-Heitler model [62] and the energy loss distribution described by this model is not particularly well approximated by a Gaussian distribution. We generalize the Kalman filter technique by representing all track trajectory states by a sum of Gaussians rather than a single Gaussian. This track fitting technique is called the Gaussian Sum Filter (GSF) technique [63]. This method results in an electron track trajectory that is composed of multiple weighted trajectories. The trajectory of the electron is taken as the trajectory with the largest weight. The electron momentum is measured as a weighted mean of the supercluster energy measurement and the track momentum measurement, whose weights are based on the uncertainties of each measurement. Three different charge measurements are used together to determine the charge of the electron: the charge of the GSF track, the charge of a track reconstructed from the algorithm used for generic minimum ionizing particles matched to the electron candidate, and the charge measured from the relative position of the innermost track hit and the supercluster. The charge of the electron candidate is taken as the majority decision of these three methods. This choice reduces the rate of charge mis-identification.

3.4.3 Electron energy scale corrections

The electron energy measurement provided by the superclustering algorithm is obtained by simply adding the contributions from the energy deposits in each crystal. The ϕ spread of the energy deposition in the ECAL, and hence the energy containment of the supercluster, will in general depend on the pattern of bremsstrahlung emission and on p_{T}. The variations of the supercluster energy containment associated to bremsstrahlung can be correlated with the volume of the shower integration, as measured approximately by the number of contributing crystals. In the case of showering electrons for which the bremsstrahlung recovery is not fully efficient and which are affected by a significant fraction of energy lost in the tracker material, the correction based on number of crystals must be complemented by a residual parametrized η-dependent correction.

3.4.4 Electron Identification

In physics analysis with multi-electron final states high identification efficiency is needed to enhance signal selection, in particular at low transverse energy (E_T) where the background increases and the fake rate is much higher. The CMS detector has features that greatly impact on electron identification like the high magnetic field, a thick tracker and the lower ECAL response to pions with respect to electrons. Electrons, for example from W boson and Z boson decay, can be distinguished from other particles due to their unique characteristics that are primarily measured in the ECAL and tracker. Ideally, the electron track would match well with the cluster of energy found in the ECAL, both in position and momentum. The track would also emanate directly from the event vertex and be isolated from other tracks and calorimeter energy deposits. Electron identification makes use of a complete set of variables to distinguish between real electrons and background electrons. A brief description of the variables is given below:

- H/E: Hadronic to electromagnetic energy ratio.

- E/p_{in}, E_{seed}/p_{in} and E_{seed}/p_{out}: Energy-momentum matching variables between the energy of the supercluster or of the supercluster seed and the electron track momentum measured at the vertex or at the calorimeter. Here, p_{in} is the electron track momentum at the closest position from the nominal vertex. p_{out} is the electron track momentum at the farthest position from nominal vertex and closest to the ECAL supercluster. E_{seed} is the highest energy crystal and E is the total energy contained in supercluster.

- $\Delta\phi_{in}$, $\Delta\eta_{in}$: Geometrical matching between the electron track parameters at the vertex extrapolated to the super cluster and the measured super cluster position.

- Calorimeter shower shape variables: $\sigma_{i\eta i\eta}$ - The width of the ECAL cluster along the η direction computed for all the crystals in the 5×5 block of crystals centered on the highest energy crystal of the seed cluster.

- d_0: Transverse impact parameter of the electron track computed with respect to the reconstructed primary vertex.

- d_z: Longitudinal impact parameter of the electron track computed with respect to the reconstructed primary vertex.

Considering the high photon conversion rates due to the material budget in front of ECAL, a selection criterion is designed to reject electrons from conversion:

- Missing Hits: Number of crossed layers without compatible hits in the back-propagation of the track to the beam-line.

3.4.5 Electron Isolation

Electrons are required to be geometrically isolated from other charged particles and energy deposits in order to reduce the contribution from fakes arising from QCD. Each track is matched to calorimeter deposits with very fine granularity. This allows for individual tracks to be identified as electrons, muons, or charged hadrons, and for the unmatched calorimeter deposits to be identified as photons or neutral hadrons. This procedure results in a global description of the event with all particles reconstructed, and is referred to as Particle Flow (PF) [64]. Such reconstructed "PF candidates" are used as the basis for defining isolation observables for an electron. A "cone" is defined in the η and ϕ coordinate space around the direction of the electron candidate by the requirement that $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < R_0$, for some particular value of R_0. R_0 is typically chosen to be a number between 0.3 and 0.5. In this Thesis, ΔR is chosen to be 0.4. Next, the p_{T} of all reconstructed particles within this cone is summed to obtain the isolation observable. Three different isolation observables based on the type of the reconstructed particles have been defined: charged particles, photons, and neutral hadrons. The PF charged particle isolation is obtained by summing the momenta of all particles inside the cone that are charged, imposing the 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 ϕ coordinate defined by $|\Delta\eta| < 0.025$. In the PF electron reconstruction procedure, an attempt is made to collect all PF photon candidates that are emitted from the electron undergoing bremsstrahlung and to combine their momenta with the momentum of the electron. Due to some inefficiency of this bremsstrahlung procedure, some fraction of such PF photons do not get collected. The veto of a narrow strip in the ϕ coordinate attempts to make the isolation definition more robust against this particular situation. Finally the PF neutral hadron isolation is computed by summing the momenta of all neutral hadron PF candidates inside the cone, but vetoing those which fall within an inner cone defined by $\Delta R < 0.07$. This veto is intended to address the situation where the electromagnetic shower induced by the electron is not fully contained within the ECAL and spills into the HCAL. In the case that the particle flow reconstruction fails to identify such an object as an electron, the energy will be assigned to a neutral hadron, and included in the isolation sum. To ensure robustness against such failures, this inner cone veto is imposed. For the PF photon isolation and the PF neutral hadron isolation, a minimum p_{T} threshold is imposed on the particle flow candidates that are considered in order to increase robustness against the influence of particles from pile-up events. The minimum p_{T} threshold is typically chosen to be 1.0 GeV. Once the three isolation variables are defined, they are summed to form a combined isolation quantity and this is required to be less than some fraction of the electron p_{T}. Typical maximum values for this fraction are between 10% and 15%. A correction is applied to the neutral flux by using the average energy density due to pile-up and underlying event in the central region of the detector (ρ) and an effective area (A_{eff}) correction which normalizes this estimator in such a way that the isolation is independent of the number of pile-up interactions. The electron isolation (I_{rel}) is therefore defined as:

$$I_{\mathrm{rel}} = \frac{1}{p_{\mathrm{T}}} \left[I_{\mathrm{ch}} + \max(I_{\mathrm{nh}} + I_{\mathrm{g}} - A_{\mathrm{eff}} \cdot \rho, 0) \right] \tag{3.1}$$

where

- I_{ch} is the sum of p_{T} from charged hadrons within the isolation cone defined.
- I_{nh} is the sum of p_{T} from neutral hadrons within the isolation cone defined.
- I_{g} is the contribution from photons.

3.4.6 Electron Performance in CMS

The latest results from CMS experiment for electron performance are documented in [65]. Figure 3.3 shows expected resolution for prompt and isolated electrons in the ECAL barrel as a function of the initial electron energy from the ECAL, the tracker and the combined estimates. The resolution is evaluated as half the minimal width that contains 68.3% of the reconstructed energy or momentum distribution (effective resolution) and using a Gaussian fit in the core of the momentum distribution. Figure 3.4 shows relative difference between the di-electron mass scales in data and simulation extracted from J/ψ, Υ and Z decays, as function of the average electron p_{T} and for different pseudorapidity regions.

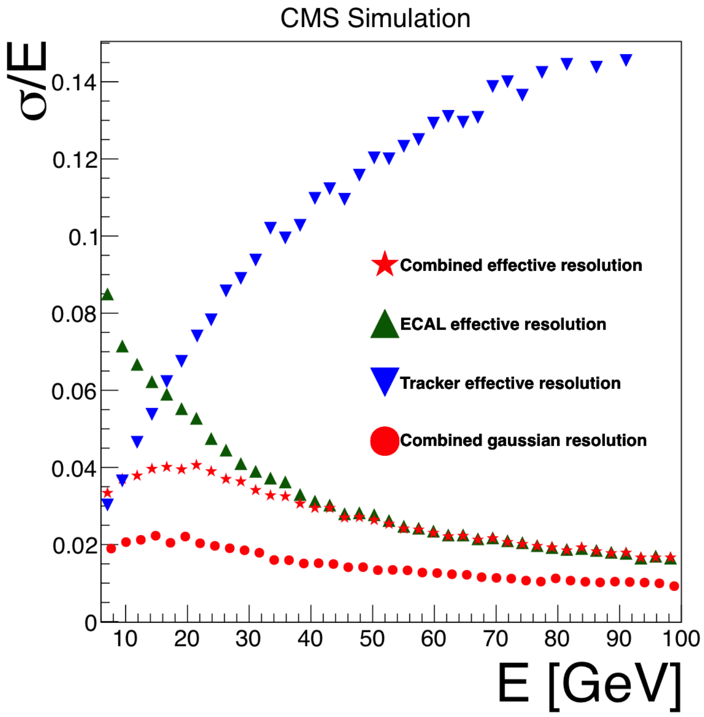


Figure 3.3: Electron resolution as a function of electron energy.

3.5 Muon Reconstruction and Selection

3.5.1 Muon Reconstruction

Muon reconstruction [66] begins with track segment reconstruction using hits in the DT and CSC muon detector systems. First, track segments are reconstructed within each muon chamber, which are loosely matched and combined to form seeds for muon reconstruction. Each seed track is extended using an iterative tracking procedure, where the candidate trajectory is propagated from one muon detector layer to the next, account-

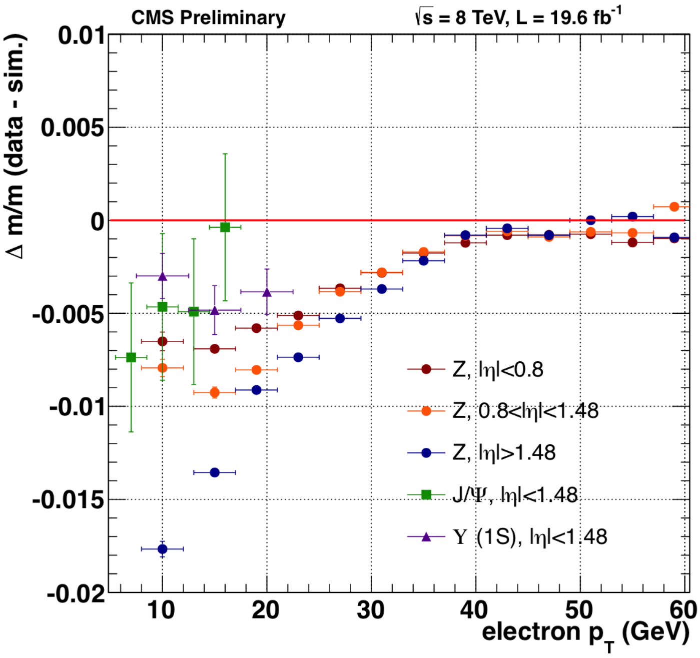


Figure 3.4: Di-electron mass resolution comparison between data and simulation at $\sqrt{s} = 8\text{TeV}$.

ing for energy loss and multiple scattering in material. This procedure results in a set of muon candidates, referred as "standalone" muons, because only information from the muon detectors are used for their reconstruction.

Next, one attempts to match standalone muons with tracks reconstructed in the silicon tracker to form muon candidates, referred as "global" muons, because they use the full information from the muon system and the tracking system. A loose rectangular region defined by the measured uncertainties in the direction of the standalone muon is used to find geometrically matching tracks. The trajectory defined by these tracks and the trajectory defined by the standalone muons are propagated to a common surface, typically chosen to be the tracker outer boundary or the muon system inner boundary, where various matching criteria, like number of matched muon stations are applied. A number of observables including spatial parameters, momentum parameters, and combined variables are used to match the standalone muons with the standard reconstructed tracks. Finally, for each pair of standalone muon matched to a silicon track, a global fit is performed

using the individual hits that comprise the standalone muon and the silicon track. This procedure results in a global muon candidate. For each standalone muon, a unique global muon is chosen with the best global fit pearson chi-square (χ^2) result.
Muons with badly measured segments in the muon chambers may fail the standalone reconstruction sequence entirely. In order to recover such muons an alternate muon reconstruction sequence is applied which begins by considering all well reconstructed silicon tracks. These tracks are propagated through the calorimeter and into the muon stations, accounting for possible energy loss and multiple scattering. Track segments reconstructed in each muon station are uniquely associated to candidate tracks by finding the nearest geometric match. Muons reconstructed by this alternative algorithm are referred as "tracker muons". In order to reduce backgrounds, requirements are placed on the compatibility of the measured calorimeter energy deposition with a muon hypothesis as well as the quality of the geometric matching of the associated muon segments. Finally, to suppress background muons to sufficiently low levels we enforce that the muon penetrate through the last muon station. This is done by requiring that there is a reconstructed track segment in the outermost muon station. Furthermore, the silicon track trajectory is required to match the segment in the outermost muon station geometrically to within three standard deviations in both transverse local coordinates. The tracker muon reconstruction is particularly useful for recovering muons with low transverse momentum, which fail the global muon reconstruction sequence more often because they are unable to penetrate fully through the material to the last muon stations. All the three types of muons and their reconstruction are shown in Figure 3.5.

3.5.2 Muon Identification

The reconstructed muons contain a significant amount of (undecayed) charged hadrons, such as kaons and pions, which are mis-identified like muons. In order to have a relatively pure sample of muons, the identification requirements must be applied. Muon identification algorithm starts with all reconstructed tracks from the silicon inner track-

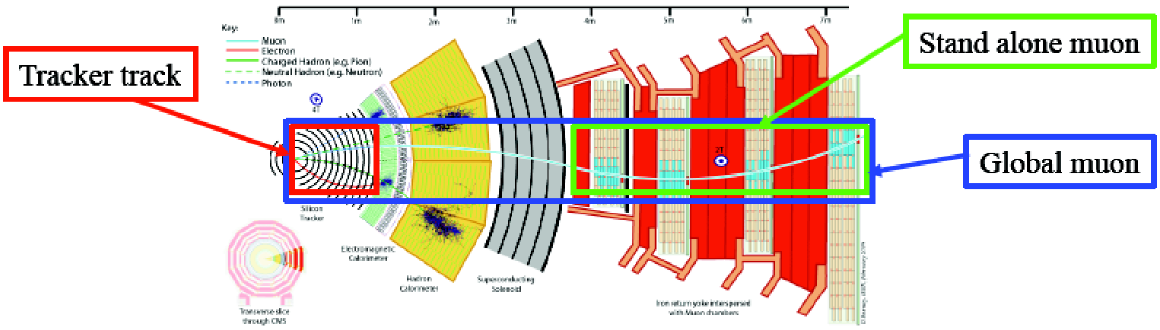


Figure 3.5: Different types of Muon reconstruction.

ing detector and attempts to quantify a muon compatibility for each track, including those tracks with no matching standalone tracks in the muon detectors. Muon identification also takes advantage of information from other detectors not used in reconstruction, such as associated energy deposits in the calorimeters for determining the compatibility of a given track with the muon hypothesis. In addition, the algorithm is able to make use of hit information from individual layers of the muon detectors, even in cases where those hits are not associated with a reconstructed, standalone muon track. Extrapolation is done for each reconstructed silicon track outward to its most probable location within each detector of interest (ECAL, HCAL, muon system). The track extrapolation takes into account the magnetic field and energy loss via the Geane software package [67]. The algorithm then searches in a cone around that most probable location for any signals that might be associated with the track of interest. There are many identification variables based on information from various subdetectors [68].

- χ^2/number of degrees of freedom (n_{dof}) of the global muon track fit : This variable is used to suppress hadronic punch-through and muons from decays in flight.

- Number of valid muon hits: These are the number of hits included in the global muon track fit. This also helps in reducing the muons which do not come from the primary vertex.

- Number of matched stations: This variables ensures that the muon has passed through muon segments in muon stations. This implies that the muon is also an arbitrated[1] tracker muon. This is used to suppress punch-through and accidental track-to-segment matches.

- d_0: Transverse impact parameter with respect to the primary vertex. This variable is required to reject cosmic muons.

- d_z: Longitudinal distance of muon track in inner tracker from primary vertex. This is used for the selection of prompt muons.

- Number of pixel hits: This counts the number of hits in pixel layers. This variable along with d_0 and d_z ensures the promptness of muons.

- Number of tracker layers: Minimum threshold on this variable is to ensure a good p_{T} measurement, for which some minimal number of measurement points in the tracker is needed.

3.5.3 Muon isolation

Muons are required to be geometrically isolated from other charged particles and energy deposits. The procedure used is analogous to the case for electrons, as discussed in Section 3.4.5. The PF candidate-based isolation variable is used to reduce the contamination from the non-isolated muons originating from jets. The muon isolation is defined as the scalar sum of the p_{T} of the PF candidates within $\Delta R < 0.3$. A correction is applied to correct for pile-up contamination. This correction is achieved by subtracting half the sum of the p_{T} of the charged particles in the cone of interest but with particles not originating from the primary vertex (I_{chPU}). The muon isolation is therefore defined as:

$$I_{\mathrm{rel}} = \frac{1}{p_{\mathrm{T}}}\left[I_{\mathrm{ch}} + \max(I_{\mathrm{nh}} + I_{\mathrm{g}} - 0.5 \cdot I_{\mathrm{chPU}}, 0)\right] \tag{3.2}$$

All the symbols have the meanings as described in Section 3.4.5. The factor 0.5 corresponds to an average of neutral to charged particles and has been measured in jets in

[1]Arbitration is the pattern recognition process which assigns each muon segment uniquely to a single Tracker Muon.

Ref. [69].

3.5.4 Muon momentum scale and resolution

The measurement of the muon transverse momentum is highly sensitive to the alignment of the tracker and of the muon chambers, to the composition of material and its distribution inside the tracking volume, and to the knowledge of the magnetic field inside and outside the solenoid volume. The momentum scale and resolution of muons are studied using different approaches in different p_{T} ranges. At low and intermediate p_{T} (≤ 100 GeV), the mass constraint of di-muon decays from the J/ψ and Z resonances is used to calibrate the momentum scale and measure the momentum resolution. In the high p_{T} range ($\geq$100 GeV), the muon momentum scale and resolution can be measured using cosmic-ray muons (with the exception of the high $|\eta|$ region). In the intermediate p_{T} range, two approaches have been developed to study the muon p_{T} measurement. The first, referred to as MuScleFit (Muon momentum Scale calibration Fit), produces an absolute measurement of momentum scale and resolution by using a reference model of the generated Z lineshape convoluted with a Gaussian function. The second, called SIDRA (SImulation DRiven Analysis), compares the data with the full simulation of the Z decay to two muons in the CMS detector and provides a way to directly modify the simulation to better match the data. Averaged over the whole acceptance, the relative bias in the muon momentum scale is measured with a precision of better than 0.2% and is found to be consistent with zero up to p_{T} values of 100 GeV [70].

The relative p_{T} resolution in the intermediate p_{T} range obtained using MuScleFit is found to be in the range from 1.3% to 2.0% for muons in the barrel and up to 6% for muons in the endcaps, which are in good agreement with the results obtained from simulation. The $\sigma(p_T)/p_{\mathrm{T}}$ averaged over ϕ and η varies in p_{T} from (1.8)% at $p_{\mathrm{T}} = 30$ GeV to (2.3)% at $p_{\mathrm{T}} = 50$ GeV, which are again in good agreement with the expectations from simulation. These results are shown in Figure 3.6.

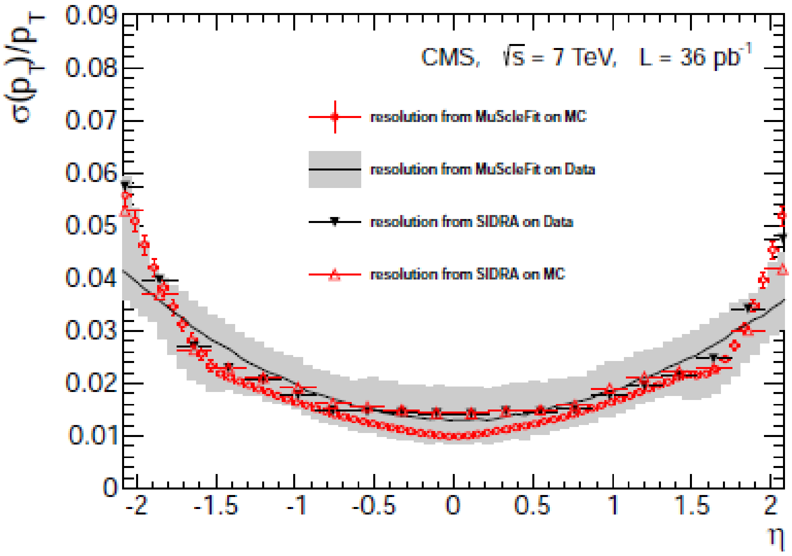


Figure 3.6: Relative transverse momentum resolution $\sigma(p_T)/p_{\mathrm{T}}$ in data and simulation measured by applying the MuScleFit and SIDRA methods to muons produced in the decays of Z bosons.

3.6 Photon Reconstruction and Identification

Photon reconstruction [71] is very similar to electron reconstruction as it also deposits its energy in ECAL. The photon energy is reconstructed by summing the energy deposited in a cluster of crystals. The clusters are 5×5 crystal arrays, or the superclusters found by the hybrid algorithm, in barrel, or the island algorithm, in endcaps. The clustering algorithms have been discussed in Section 3.4.1. The choice of clustering algorithm depends on the spread of the deposited energy. For compact energy deposits, originating mainly from the single showers of unconverted photons, the best energy measurement is achieved using a 5×5 crystal array, while superclustering algorithms provide better measurement for multiple showers originating from conversions where the bending of the electron and positron tracks, and their radiation in tracker material, spreads the energy over a larger area of the calorimeter. But they do not have tracks in tracker so the identification mainly depends on the shower shapes and lateral spread of showers. Variables which are used to identify photons are:

- R_9: The ratio of energy contained in a 3×3 array of crystals centred on the crystal with the highest deposited energy to the total supercluster 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 5×5 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.

A very convenient measure of the lateral spread of energy deposition is provided by the R_9 variable. Values approaching unity are obtained for unconverted photons, or for photons that have converted very close to the ECAL. Smaller values are obtained for increasing distances of the conversion vertex from the ECAL.

3.7 Jet Reconstruction

Jets are the main ingredient to reconstruct the hadronic content in an event. Quarks and gluons cannot be observed directly because of quark confinement, although they can be detected through the tracking and calorimeter systems as they fragment into stable hadrons. They appear in the detector as jets of particles [72]. The jet reconstruction and identification in the CMS is a complicated and sensitive task that involves the calorimeters and the tracking systems.

3.7.1 Reconstructed jet types at CMS

Three different types of jets are used in the CMS, characterized by the way the inputs from subdetectors are used during the jet finding procedure: calorimetric jets (CaloJets), jet-plus-tracks jets (JPT Jets) and particle flow jets (PFJets).

Calorimeter Jets - CaloJets: In the case of calorimeter jets [73], the input objects for the jet clustering are formed from the measurements of the energy depositions in the electromagnetic (ECAL) and the hadronic (HCAL) calorimeters. The ECAL crystals and HCAL cells are combined into calorimeter towers. A calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL crystals. The energy of these towers is calculated by summing up the energy of all contributing readout cells. As calorimeter jets are reconstructed from only one type of detectors, they are the simplest reconstructed jet type. They have been used at previous experiments and have been used extensively for physics studies at the CMS [74]. However, their performance suffers from the relatively poor energy resolution of the very compact hadronic calorimeter of CMS.

Jet-Plus-Tracks (JPT): The JPT algorithm [75] corrects the energy and the direction of a calorimeter jet. It exploits the excellent performance of the CMS tracking detectors [76] to improve the p_{T} response and resolution of calorimeter jets. Charged particle tracks are associated with each calorimeter jet based on the spatial separation in $\eta - \phi$ between the jet axis and the track momentum measured at the interaction vertex. The associated tracks are classified as in-cone tracks if their projection onto the surface of the calorimeter (ECAL) falls within the jet cone. Conversely, if they are bent outside the cone by the magnetic field, then they are called out-of-cone tracks. The expected calorimeter measurements of the in-cone tracks are subtracted from the calorimeter measurement and replaced by the precise momentum measurement of the tracker. The measurement of out-of-cone tracks is also taken into account so that it is possible to improve the jet energy resolution significantly.

Particle Flow Jets (PF): PF algorithm aims to reconstruct, identify and calibrate each individual particle in the event by combining the information from all the CMS subdetector systems. PF particles are reconstructed as a combination of charged particle tracks and clusters in the electromagnetic and hadronic calorimeters, as well as signals in either of the two CMS preshower detectors and the muon system. Depending on which of the detector systems contribute to a single particle, it is defined as either an electron (track+ECAL), muon (track+ECAL+HCAL+Muon System), photon(ECAL), charged hadron (track+ECAL+HCAL), or neutral hadron (HCAL). The algorithm employs strategies to handle ambiguities stemming from the overlapping detector signals to avoid double counting. Based on the particle type, the energy of each particle is calibrated. Charged hadrons are treated under the assumption that they are pions. The reconstruction of jets from these reconstructed particles leads to an improved performance with respect to the jet energy resolution. In particle flow approach, charged hadrons are identified to carry, on the average, 65% of the jet energy in the tracker covered region, photon 25% and neutral hadrons 10% of the jet energy. These fractions ensure that 90% of the jet energy can be reconstructed with good precision by the particle flow algorithm, while only 10% of the energy is affected by the poor hadron calorimeter resolution.

3.7.2 Jet Algorithms

Depending on the reconstruction approach, jet clustering algorithms can be classified into two major groups. In the first group, cone-based algorithms take particles or calorimeter towers as seeds in an iterative search for stable cones in an event. In the second group, a distance is defined between pairs of particles and successive re-combinations of the pair of closest particles are performed, which is stopped when all resulting objects are too far apart. The jet algorithm used in CMS is the 'anti-k_t' with $\Delta R = 0.5$. In the anti-kt algorithm [77], one introduces distances d_{ij} between entities i and j and d_{iB} between object i and the beam (B). For clustering, smallest of the distances is calculated; if it is a d_{ij} , the objects i and j are recombined by adding their four-momenta, while if it is d_{iB} it is labeled as a jet and is removed from the list of objects. Distance measures are defined as:

$$d_{ij} = min(k_{t,i}^{2p}, k_{t,j}^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{3.3}$$

$$d_{iB} = k_{t,i}^{2p} \tag{3.4}$$

where, $\Delta_{i,j}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and $k_{t,i}$, y_i and ϕ_i are the transverse momentum, rapidity and azimuth of particle i respectively. In addition to the usual radius parameter R, a parameter p has been added to govern the relative power of the energy versus geometrical (Δ_{ij}) scales. The anti-k_t algorithm is a sub case of the above set of equations where p = -1, while for p= 0 and p = 1 we get the so called k_t [78, 79] and Cambridge/Aachen [80] algorithms respectively.

3.7.3 Jet-energy corrections (JEC)

Since jets are clusters of PF candidates, where certain thresholds are applied, the jets underestimate the total contained energy. Also, the reconstructed jet has the energy usually different than that of the corresponding particle because of the non-uniform and non-linear response of the calorimeters. Moreover, electronic noise and multiple proton-proton interactions in the same bunch crossing (pile-up events) can cause this extra amount of energy. The goal of the jet energy calibration is to find the relation between the energy measured in the detector jet and that of the corresponding particle jet. Thus it is required to implement the additional corrections to the jet energies as explained here.
CMS has adopted a factorized solution to the problem of jet energy corrections [81], where each level of correction takes care of a different effect. The set of corrections used are as follows:

- **L1 Offset Correction**: The goal of the L1 correction is to remove the energy from pile-up events, electronic noise and jet energy lost by thresholds. In principle this will remove any dataset dependence on luminosity so that the following corrections are applied upon a luminosity independent sample.

- **L2 Relative Correction**: The goal of the L2 Relative correction is to make the jet response uniform with η. Essentially, the uniformity in pseudorapidity is achieved by correcting a jet in arbitrary η relative to a jet in the central region ($|\eta| < 1.3$).

- **L3 Absolute Correction**: The goal of the L3 Absolute correction is to make the jet response uniform with p_{T}. Once a jet has been corrected for η dependence, it is corrected back to particle level, i.e. the corrected jet p_{T} is equal, on average, to the generator-level jet p_{T}.

- **L2 L3 Residual corrections**: 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 depending on η. Therefore, the L2 L3 Residual corrections are provided as described below:

  – first, the L2 and L3 JEC calibrations are 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.

3.7.4 Jet identification

A set of quality cuts is applied to every reconstructed jet in order to reduce fake jets from calorimeter noise [82]. Various identifications used are:

- Number of particles in jet (n)

- Fraction of energy deposited in HCAL by neutral hadron constituents (nhf)

- Fraction of energy deposited in ECAL by neutral constituents (nef)

- Fraction of energy deposited in ECAL by charged constituents (cef)

- Fraction of energy deposited in HCAL by charged hadron constituents (chf)

- Number of charged particles (nch)

Apart from these variables, a separate identification criteria is applied to suppress the pile-up contamination in jets. Following variables are used:

- Fraction of tracks associated to other vertices than primary vertex (β^*)
- Jet contains many particles in it. So, root mean square (RMS) of the constituents p_{T} is also used.

3.8 Missing Transverse Energy (E_T^{miss})

The lack of electromagnetic and strong interaction of neutrinos with the materials of the detectors renders them invisible to the detector. They do not leave any physical signal and their presence in a collision has to be inferred indirectly by studying energy and momentum conservation. Due to the hermeticity of the CMS detector, almost any particle in an ample angle coverage can be detected. The event should therefore be balanced energetically. An imbalance in the momentum in the plane transverse to the beam direction signifies the presence of undetected particles (which can be neutrinos, but also objects escaping the fiducial volume of the detector). Reconstruction of $\sum E_T$ (scalar sum of all the energy in the event in the transverse plane) and $\overrightarrow{E_T}^{miss}$ (vectorial imbalance in the transverse plane) is challenging. It is very sensitive to detector miscalibration and noise. In addition to this it can be mismeasured if otherwise identifiable particles pass through poorly instrumented parts of the detector. The magnitude $E_T^{miss} = |\overrightarrow{E_T}^{miss}|$ is one of the most important variables for many processes. In particular, it is a clear discriminator for any physical process involving the production of neutrinos, or in any supersymmetric search involving weakly interacting particles. In this Thesis, E_T^{miss} reconstruction plays a key role, as a large E_T^{miss} is involved because of Z decay into neutrinos in the signal. CMS has designed three main algorithms to reconstruct E_T^{miss}, involving all of the subdetectors to ensure that the required performance is obtained. They are briefly described in this section.

3.8.1 E_T^{miss} Algorithms

The basis for a global energy reconstruction in a detector like CMS is the sum of all the energy in the event. This energy can be merely measured in the calorimeters, or incorporate corrections using other subdetectors. In particular, CMS uses the tracker, given its excellent momentum determination.

Calorimetric $\overrightarrow{E_T}^{miss}$ (caloMet): It uses exclusively the energy deposited in the ECAL and HCAL towers. The $\overrightarrow{E_T}^{miss}$ variable is computed as the negative vector sum of these calorimeter depositions. The energy collected in the calorimeters does not take into account muon energy, since muons are minimum ionising particles that traverse calorimeters almost unaffected (on average, they deposit only about 2 GeV). Therefore, the contribution of momenta of all the muons in the event has to be subtracted vectorially for the final calculation.

$$\overrightarrow{E_T}^{miss} = -\sum_{CaloTowers} \overrightarrow{E_T} - \sum_{muons} \overrightarrow{p_T}^{\mu} + \sum_{muons} \overrightarrow{E_T}^{\mu}$$

Further corrections can be applied to this caloMet. Final states with jets require that the E_T^{miss} is corrected by the Jet Energy Scale. These corrections take the measured raw energy values and adjust them for the difference between corrected and raw energy scales. Uninstrumented regions of the detector, and effects of underlying event, are also considered.

Track-Corrected $\overrightarrow{E_T}^{miss}$ (tcMet): The performance of the CMS tracker is superior to that of the calorimeters for the measurement of the momentum of charged particles. The overall resolution and tails of the $\overrightarrow{E_T}^{miss}$ reconstruction is greatly improved by combining the information provided by the CaloTowers as a basis with momentum measurement in the tracker, in the cases where this is possible. '$tcMet$' is calculated replacing for all well reconstructed tracks, the expected energy deposition in the calorimeters with the corresponding measured track momentum. This correction is applied to all tracks not identified as muons or electrons. Charged particles are assumed to be pions, and their expected energy deposit is determined from a response function derived from a single pion MC sample, mapped in intervals of η and p_{T}. Tracks with $p_T < 2$ GeV are fully compensated for assuming no response from the calorimeters ($\overrightarrow{E_T}^{miss}= 0$). Tracks with $p_T > 100$ GeV receive no correction since at that energy the measurement of calorimetric energy for charged hadrons is already good enough not to gain anything from the tracker resolution. Identified muons are corrected in a similar way as in '$caloMet'$. No correction is applied for identified electrons as their energy is already correctly measured in the electromagnetic calorimeter.

$$\overrightarrow{E_T}^{miss} = -\sum_{CaloTowers}\overrightarrow{E_T} - \sum_{muons}\overrightarrow{p_T}^{\mu} + \sum_{muons}\overrightarrow{E_T}^{\mu} - \sum_{goodtracks}\overrightarrow{p_T}^{track} + \sum_{goodtracks}\overrightarrow{E_T}^{track}$$

The aim behind the design of this algorithm is the control of the tails of the $\overrightarrow{E_T}^{miss}$ distributions (high E_T^{miss} events) in events with no real E_T^{miss}.

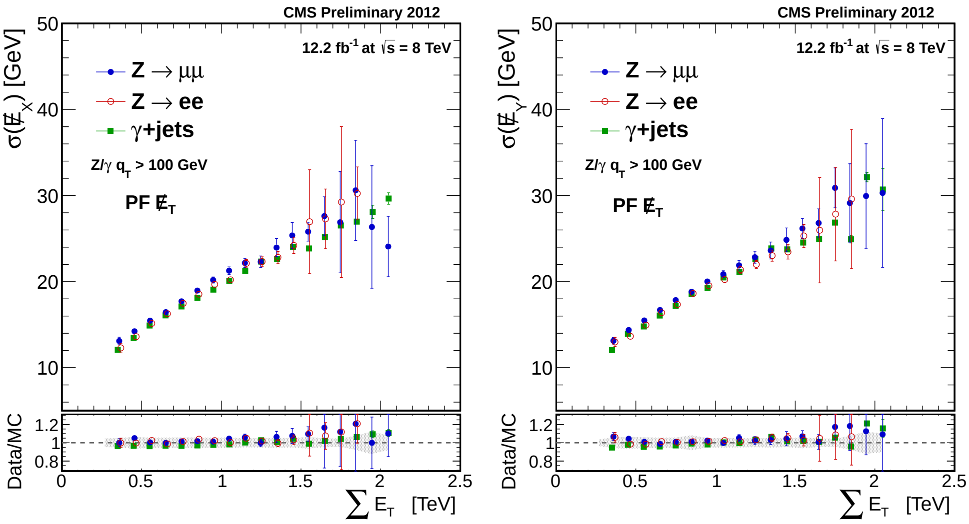
Particle-Flow $\overrightarrow{E_T}^{miss}$ (pfMet): A final step comes from the full reconstruction and interpretation of the event through PF techniques [83], which aim to identify all stable particles in the event (electrons, muons, photons, neutral hadrons and charged hadrons) through the combination of all CMS subdetectors. Jets and τs are built from these identified objects. Once all particles in the event are identified by the algorithm, $\overrightarrow{E_T}^{miss}$ is determined by building the vectorial sum of transverse momenta over all reconstructed particles in the event and then taking the opposite of this vector.

$$\overrightarrow{E_T}^{miss} = -\sum_{AllPF-Particles}\overrightarrow{p_T}$$

There is no need to apply corrections for muons or tracks since they are directly taken into account in this inclusive approach. In this Thesis, particle-flow reconstructed E_T^{miss} is used.

3.8.2 Performance of E_T^{miss}

The performance of E_T^{miss} reconstruction algorithms have been studied using proton-proton collision data collected with the CMS detector. Events with anomalous missing transverse momentum are studied to find the performance of algorithms used to identify and remove those events. The scale and resolution for E_T^{miss}, including the effects of multiple proton-proton interactions in the same, earlier, and later bunch crossings (pile-up interactions), are measured using events with an identified Z boson or isolated photon. Figure 3.7 [84] shows the resolution of the PF E_T^{miss} projections along the x- and y-axes as a function of $\sum E_T$. The $\sum E_T$ is the scalar sum of E_T of all the particles reconstructed by the particle-flow reconstruction except for dileptons from Z's or photons. The results here are shown for $\sqrt{s} = 8$ TeV data corresponding to an integrated luminosity up to 12.2 $\pm$ 0.5 fb^{-1}.



(a) (b)

Figure 3.7: Resolution of the PF E_T^{miss} projection along the x-axis (a) and the y-axis (b) as a function of $\sum E_T$ for events with Z and γ. Results are shown for $Z \to \mu^+\mu^-$ events (full blue circles), $Z \to e^+e^-$ events (open red circles), and photon events (full green squares). The upper frame of each figure shows the response in data; the lower frame shows the ratio of data to simulation.

3.9 Tau Reconstruction

The tau lepton is the heaviest among the three families of leptons. Therefore, τ leptons produced in collider experiments typically decay in flight before interacting with any of the detectors. The purely leptonic decays account for about 35% of the total decay width; the rest is absorbed by semi-leptonic decays into one or three charged pions (prongs) and additional neutral pions. In about 53% of the cases, tau leptons decay to final state with more than one hadron. An additional challenge in reconstructing tau leptons is posed by the presence of neutrinos in the final state, which escape undetected. The p_{T} spectrum of the visible particles from the tau decay is therefore softer than that of original τ lepton. CMS has adopted a PF approach towards τ reconstruction [85]. First, all PF particles are clustered into jets using the anti-k_t algorithm with radius R = 0.5. For each jet, the particle content is scanned: π^0 candidates are built from the photons, if any, and finally assembled with the charged hadrons to form a hypothesized decay mode, from which the hadronically decaying tau (τ_h) four-momentum and the compatibility with a τ_h hypothesis can be measured. Both cut-based Hadronic Plus Strips (HPS) and a multivariate Tau Neutral Classifier (TaNC) algorithm are currently used for this purpose [86].

3.10 *b*-tagging

b-tagging [53] is one of the most important tool to identify the top-quark related events as top quark decays almost exclusively into W boson and *b*-quark. Top-related processes like $t\bar{t}$ and single-top production are major backgrounds in many searches including Higgs boson searches. *b*-tagging gives a way to suppress these events. Inclusive tagging of b jets, as opposed to lighter flavour jets, mainly relies upon relatively distinct properties of *b*-hadrons such as large proper lifetime ($\tau = 1.5$ ps, c$\tau = 450$ μm), large mass, decay to final states with high charged track multiplicities (on average 5 charged tracks), relatively large semileptonic branching ratios (in about 20% of the cases, *b*-hadrons decay into muons or electrons), and a hard fragmentation function. At the CMS, *b*-tagging is typically applied to jets and hence accurate jet reconstruction algorithms are necessary. A brief introduction of various methods of b-tagging is as follows:

3.10.1 Track impact parameter based tags

The track impact parameters of charged particles as measured in the tracking detector provide a powerful means to determine whether a track originates from the primary interaction vertex or from the decay of a particle that has travelled a significant distance from the primary interaction vertex. Since *b*-quark has higher lifetime than other charged particles so impact parameter can be a good tool to identify it. The impact parameter can be computed in the transverse plane or in three dimensions. Three dimensional impact parameters yield better overall *b*-tagging performance.

Track-counting b-tagging

In this method, for each selected track in a jet, the impact parameter significance is computed. The jet is tagged as b-jet if the number of tracks with an impact parameter significance exceeding a given threshold is greater than a value which can be optimized for various purposes. Tracks are ordered in decreasing impact parameter significance. The significance of the track impact parameter of the N^{th} track thus serves as the discriminator for this algorithm. In the CMS, there are b-tagging algorithms available based on this method like Track-Counting High Efficiency (TCHE) tagger [53].

Probability b-tagging

The probability method is also based on track impact parameters: it entails computing the compatibility of a set of tracks which come from the primary vertex [87]. The jet probability estimation is performed through several steps. For each track, the probability of coming from the primary vertex is computed and these probabilities are combined together to provide the jet probability. In the CMS, jet probability taggers (JPT) are used with different working points. In this Thesis, jet probability tagger is used to tag b-jets for reducing top-quark related backgrounds.

3.10.2 Combined secondary vertex tag (CSV)

In this technique [88] different topological and kinematic vertex variables are combined together with track impact parameter significances into a discriminating variable to distinguish b-quark jets from non-b jets.

3.10.3 Soft lepton tags

The soft lepton b-tagging algorithm is based on the relatively high inclusive b-quark decay branching ratio to electrons and muons which is about 19% for each lepton family if both direct and cascade decays are taken into account. The key element that is required in order to take advantage of this situation is the identification of leptons among the tracks associated to each jet. To increase the purity of the selection, additional requirements are applied to parameters associated with these leptons. A detailed description can be found in [89].

Chapter 4

$H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ Analysis

In order to achieve high sensitivity over the entire search range of Higgs masses (M_H), different analyses, investigating all possible decay channels of the Higgs boson and corresponding final states, need to be combined. This effort is also very crucial in the mass region beyond $M_H > 2M_Z$, which is not explored at other colliders such as the LEP and the Tevatron. $H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ is very sensitive for $M_H = 200$ GeV to $M_H = 1000$ GeV. Higgs boson production cross-section decreases very rapidly with increase in its mass, so to compensate that deficit, there is a need to explore decay channels with higher branching fractions. The branching ratio of this channel is six times larger than full leptonic $H \to ZZ \to 4\ell$ channel. This way one gets more statistics; however, at the cost of more background and worse resolution than fully leptonic channel.

After the observation of new boson with mass close to 125 GeV [5, 6] by the ATLAS and CMS experiments at the LHC, many questions seek experimental answers namely, if the new boson is the long-sought Higgs boson as favoured by the EW fits [18], if it is the only of its kind and if the scattering of vector bosons, in particular longitudinally-polarized vector bosons, is affected by this newly observed state. While these and other related questions will naturally find answers from the direct study of the boson properties, it is compulsory to complete the search for a SM-like high mass Higgs boson and to pave the road for the study of EW vector boson scattering in the high $\sqrt{s}$ regime. With an increased integrated luminosity the previous searches carried in the bosonic channels [90] can be extended to higher masses. Ultimately, by including a dedicated study of the vector boson fusion (VBF) production in high mass regime search we are indirectly testing one of the outstanding issues of the SM: the unitarity of the cross-section for the scattering of longitudinally-polarized vector bosons. The search can be extended, as well, to Beyond the Standard Model (BSM) scenario. Two examples are a fermiophobic-like Higgs or a second heavy EW singlet state [25, 28]. Also, with this analysis one can constrain the SM Higgs boson decay width using off-shell production at higher masses [91].

This chapter describes the search for the SM Higgs boson production at the CMS experiment in the decay channel $H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ (ℓ = e,μ). The Higgs boson mass range considered is 200 GeV $\leq M_H \leq$1000 GeV. Both gluon-gluon fusion and VBF mode of Higgs boson production are included as signal component. The search is optimised separately for both production mechanisms to extract the maximum sensitivity. The results shown are outcome of the analysis performed using a total integrated luminosity of 24.7 fb^{-1} collected at $\sqrt{s}$ = 7 TeV and $\sqrt{s}$ = 8 TeV. This is the full data recorded by the CMS experiment till the LHC stopped for its first long shutdown in 2013. The main background processes for this final state are Drell-Yan production $Z + Jets$ and the subsequent decay of Z boson in leptonic final states, $t\bar{t}$, diboson production (ZZ, WZ, WW), single top-quark production, and $W + Jets$ production. Two approaches are considered. Baseline or Shape analysis is the one where full distribution or shape of the final discriminating variable is used (instead of imposing any threshold). A cross-check is made where a selection criterion is applied on the final discriminant, referred as cut-and-count analysis. These are described in this Chapter. An independent cross-check is made using a multivariate analysis, which can utilize correlations among discriminating variables in a better way, which is described in Chapter 5.

A description of the dataset together with the signal and the background samples is given in the Section 4.1. Various reweightings used in the analysis are explained in Section 4.2. The physics object selection used in the analysis is discussed in Section 4.3 and the full

event selection is described in Section 4.4. All the dominant backgrounds are estimated using data-driven techniques and they are described in Section 4.5. Overview of selection efficiencies is given in Section 4.6. Systematic uncertainties are explained in Section 4.7 and the results are described in Section 4.8.
All the distributions shown in this chapter are for $\sqrt{s} = 8$ TeV and all corresponding plots for $\sqrt{s} = 7$ TeV are given in Appendix B. However, results for $\sqrt{s} = 7$ TeV analysis are shown in Section 4.8.

4.1 Data and Monte Carlo Samples

In order to search for a rare signal like SM Higgs boson in the billions of proton-proton collisions taking place at the LHC, understanding the signal and background processes becomes very important. The analysis strategy, including the event selection, is decided by looking at the distributions of these processes using Monte Carlo (MC) samples. There are various MC event generators, which, after passing through the detector simulation, take into account various effects like interaction of particles with the detector, resolution effects, data reconstruction etc. In this analysis most of the backgrounds are taken from data-driven techniques, but still to be confident about the method of estimation of background, validation of these tests have been done using MC samples.

4.1.1 Collision Data

The data sample studied in the analysis is collected from the proton-proton collisions at $\sqrt{s} = 7$ TeV between April 2011 and October 2011 and at $\sqrt{s}$= 8 TeV between April 2012 and December 2012, using the CMS detector. The data in 2011 correspond to an integrated luminosity of 5.0 $fb^{-1}$and in 2012 it corresponds to about 19.7 fb^{-1}. The data recorded is ordered in units called *event, luminosity sections* and *runs.* Runs or individual luminosity sections affected by hardware failures or noise for any of the subdetectors are marked as bad and the events belonging to these bad runs or luminosity sections are removed from the analysis. Since, there are two leptons in the final state so the datasets triggered by dilepton triggers have been considered for this analysis. However, there are other datasets for lepton efficiency measurements and background estimations. The details of these datasets are given in Table A.1 in Appendix A.

4.1.2 Signal Processes

For this search of the SM Higgs boson, both the dominant production modes at the LHC, gluon-gluon fusion and VBF, are considered. Thirteen Higgs boson mass points are considered from M_H= 200 GeV - M_H= 1000 GeV. The signature of the signal consists of two isolated, well reconstructed and identified leptons consistent with Z boson mass with high E_T^{miss} coming from another Z boson decaying to neutrinos. The cross-sections of these processes at the LHC are shown in Figure 1.8 and their tree level Feynman diagrams are shown in Figures 1.9 and 1.10. It can be seen that the VBF production contributes about 10% to the total cross-section in almost a full mass range, but tends to become comparable with gluon-gluon fusion when M_H approaches TeV scale. Higgs production through the gluon-gluon fusion and the VBF processes is simulated using the next to leading order (NLO) POWHEG [92] generator. The signal samples used in this analysis are listed in Table A.2 in Appendix A along with the cross-section values assumed for each mass point of the Higgs boson. The signal cross-section values are taken directly from the recommendations of the LHC Higgs Cross Section Working Group (LHC Higgs XS WG) [23, 24].

4.1.3 Background Processes

There are the SM processes that have same topology as signal and hence they can mimic the signal signature and thus pose as a background to the expected signal. A precise understanding of their contribution is crucial for an effective background rejection. The MC datasets for various background processes relevant to this analysis are listed in Table A.3 in Appendix A along with the respective cross-sections [93]. The diboson cross-sections are based on MCFM (Monte Carlo for FeMtobarn processes at Hadron Colliders) computations from Ref. [94]. All samples are generated using the MADGRAPH [95], PYTHIA [96] or POWHEG [92] generators and pass through full simulation of the CMS detector based on Geant4 [50].

Z+Jets Production

The main background contribution at the preselection stage is due to the Drell-Yan (DY) processes, i.e. $q\bar{q} \rightarrow Z/\gamma^* \rightarrow \ell\ell$. The cross-section of this process is six orders of magnitude higher than the cross-section for the signal. These final states contain two well identified leptons with ideally no E_T^{miss}. Any E_T^{miss} that may appear in the final state is instrumental and hence arises from some source of mis-measurement. Since there is a large E_T^{miss} present in the signal so a minimum threshold on E_T^{miss} is very important criterion against this background. There are other topological distributions which differ significantly from the signal and hence this background process can be separated from the signal. The leading order Feynman diagrams of this process is displayed in Figure 4.1.

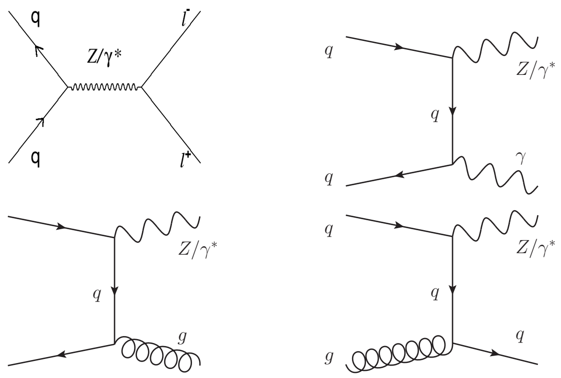
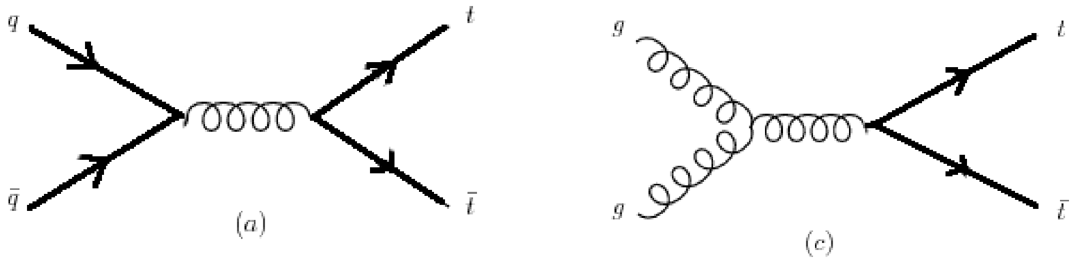


Figure 4.1: Leading order Feynman diagrams for Z+jets production.

Top-AntiTop quark pair production

This is the second most dominating background. Figure 4.2 shows the production of a top-antitop quark pair ($t\bar{t}$). At the LHC, the dominant top quark production channel is pair-production of the $t\bar{t}$ quarks, and the top quark always decays into a W boson in association with a bottom (b) quark. Hence, the final state consists of E_T^{miss} and two well identified leptons from the decay of the W bosons and jets. The properties of b-quark (long life time) along with the identification of soft (low-p_T) non-isolated muon coming from b-quark decay help in distinguishing this background from the signal events.



(a) (b)

Figure 4.2: Leading order Feynman diagrams for $t\bar{t}$ production.

DiBoson Production

There are few diboson processes which are background to this analysis. These include the irreducible ZZ, WZ and WW production. However, their cross-sections are also small so they are not large background at initial stage of analysis. But they dominate the background contribution at the level of final selection. The production diagram from $q\bar{q}$ is shown in Figure 4.3. A reweighting is applied for including the contribution from gluon-gluon fusion production of ZZ. This is discussed in detail in Section 4.2.

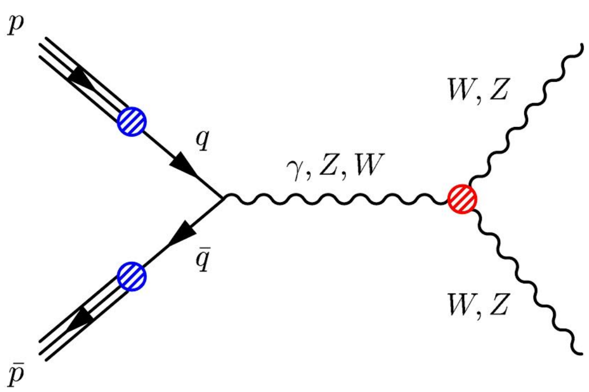
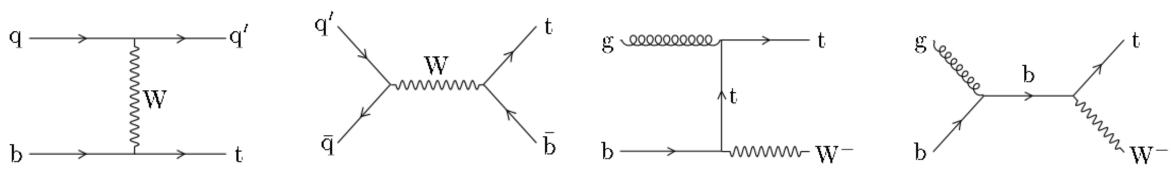


Figure 4.3: Leading order Feynman diagrams for diboson production.

Single-top quark production

These are minimal backgrounds which include the s-channel, t-channel and tW channel production of single-top quark. The production diagrams of these backgrounds are shown in Figure 4.4. tW process contributes the most because top quark can decay to b-quark and a W boson and then leptonic decays of two W bosons can lead to two leptons and E_T^{miss}.

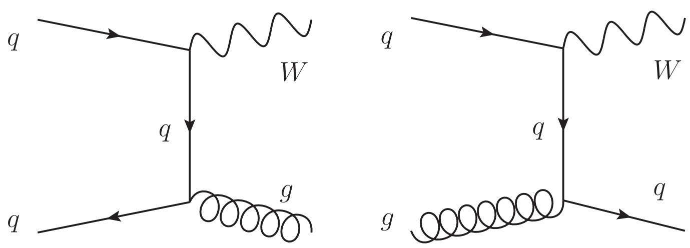


(a) (b) (c) (d)

Figure 4.4: Leading order Feynman diagrams of the single top processes. From left to right: t-channel, s-channel and two diagrams for tW-channel.

W+jets production

This is also a sub-dominant background which is reduced by lepton selection and Z boson mass requirement. The main feynman diagrams for this production are shown in Figure 4.5. In this background, W boson can decay leptonically to give a lepton and E_T^{miss} and one jet can fake as lepton giving rise to two leptons and E_T^{miss} which mimic the signal process. Other MC and data samples used for background estimation are also



(a) (b)

Figure 4.5: Leading order Feynman diagrams for W+jets production.

listed in Table A.4 in Appendix A.

4.2 Reweightings

The various reweightings which are applied to signal and background samples are discussed in this Section.

4.2.1 Pile-up reweighting

Multiple proton-proton collisions in a single bunch crossing is called pile-up. Simulated MC samples are generated with distributions for the number of pile-up interactions which is meant to roughly cover, though not exactly match, the conditions expected for each data taking period. As the luminosity of the LHC changed frequently, the simulation samples must be reweighted to mimic the LHC conditions. The number of pile-up events simulated in the MC samples (in the same bunch crossing, 'in time', or in the previous or following one, 'out of time' pile-up) have been generated using poisson sampling from a distribution similar to what is expected from data. For a given range of analyzed runs, the mean number of pile-up interactions per bunch crossing is estimated for given luminosity section using the instantaneous luminosity provided by the LHC, integrated over the entire run range. This distribution is then used to reweight the simulated pile-up distribution. Figure 4.6 shows the estimate of the "observed" number of pile-up interactions in data assuming a Minimum Bias (MB) cross-section of 70.3 mb at $\sqrt{s} = 8$ TeV. The "observed" pile-up corresponds to observed number of pile-up interactions, i.e. as drawn from a poisson distribution where the expected mean number of interactions is the "true" pile-up. The MB cross-section values were chosen in order to have the best possible agreement with the simulation with respect to the distribution of the reconstructed vertices. The level of agreement has been assessed using the value of the pearson chi-square (χ^2) between data and simulation. The vertex multiplicity distributions before and after pile-up reweighting the dilepton sample (i.e. $ee/\mu\mu$ before any Z boson invariant mass requirement) are shown in Figure 4.7.

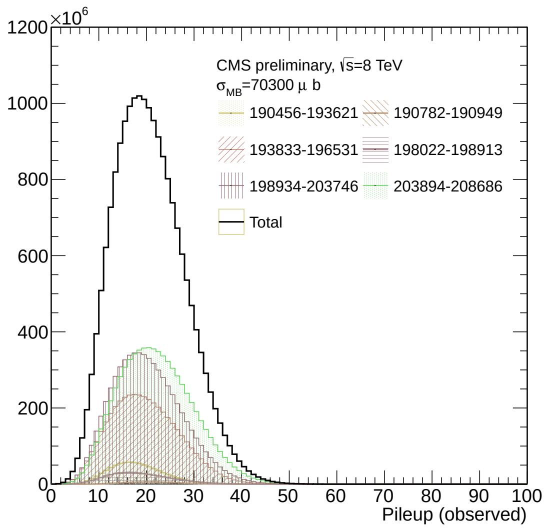


Figure 4.6: Distribution of the pile-up interactions estimated in data at $\sqrt{s} = 8$ TeV for the different run-range (given by numbers) and also for the full run-range.

4.2.2 Higgs Lineshape Reweighting

The lineshape of the generated mass distribution for the double vector boson (V) system, $M(VV)$, has been corrected to match the results presented in Ref. [97–99] where the Complex-Mass Scheme (CPS) for the Higgs propagator has been used. Moreover, in the gluon-gluon fusion production channel the effects on the lineshape due to the interference between Higgs boson signal and the $gg \to VV$ background have been included, as suggested in Ref. [100]. A reweighting is applied to signal events to get the proper Higgs mass lineshape for heavy Higgs ($M_H \geq$ 400 GeV). The first step is to reweight POWHEG events for the CPS related effect. This is performed for both gluon-gluon and VBF samples. The second step is only applied to the gluon-gluon sample and it embeds the interference with $gg \to ZZ$. The effect of such reweighting is shown in Figure 4.8.

For ZZ production the contribution from the gluon-gluon box diagram [101] is included in the NLO calculation used to scale this process. It contributes about 14% of the total cross-section. It was also investigated if a reweighting is needed to take into account the difference in the p_{T} of the dilepton system predicted at leading order (LO) and at NLO. But it was observed that the variations induced are minor and are covered by the statistical uncertainty on p_{T} distribution shape from dilepton system. Thus this reweighting is not

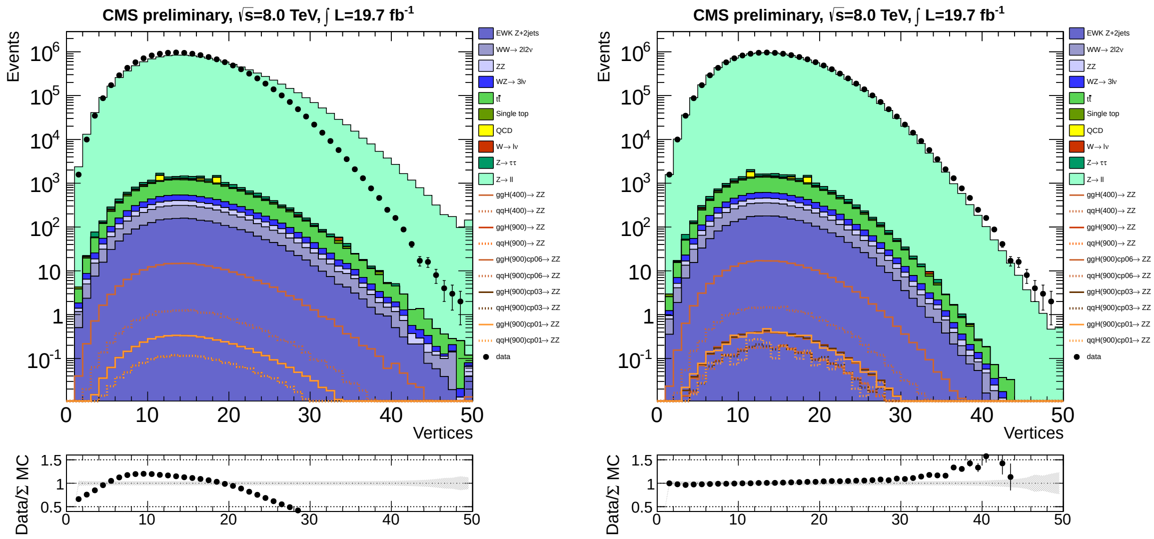


Figure 4.7: Vertex multiplicity distribution in data and in simulation before (*left*) and after (*right*) reweighting according to the estimated distribution of the number of interactions in each event.

applied.

4.3 Selection of Physics Objects

This section explains the selection of various physics objects including triggers, electrons, muons, jet, E_T^{miss} etc.

4.3.1 Trigger

The list of triggers used to search for the signal are listed in Table A.5 in Appendix A. Dilepton triggers are used in the analysis because of two leptons in the final state. As the instantaneous luminosity goes higher, the single lepton trigger's p_{T} thresholds become very high to reduce the data taking rates. However, the p_{T} thresholds of dilepton triggers is lower, so they are used in this analysis. Along with that, a complementary set of triggers is used to select events for efficiency measurements or to analyze background dominated control regions for data-driven background estimation. These monitoring triggers are also listed in the Table A.5. As mentioned in Section 2.3 of Chapter 2, the CMS uses two level trigger system, L1 and HLT. In the HLT, module instances providing the event selection are arranged in "trigger paths" or 'HLT Path'. HLT path consists of

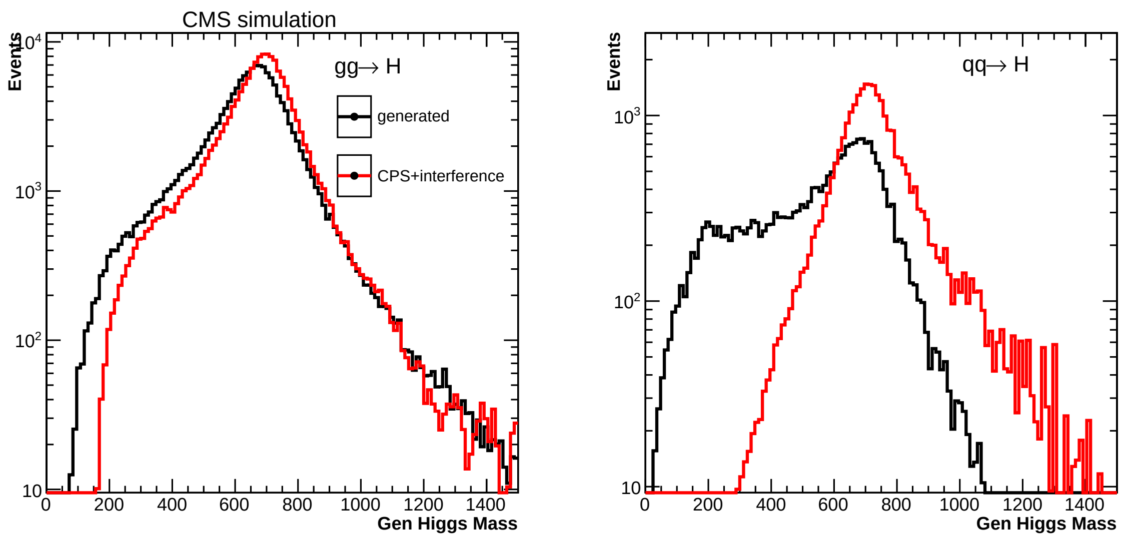


Figure 4.8: Mass line-shapes obtained after reweighting M_H = 700 GeV (referred to as 'Gen Higgs Mass' in the x-axis) Higgs produced via gluon-gluon fusion (*left*) and VBF (*right*). The distributions are weighted according to the cross-section and integrated luminosity.

a sequence of reconstruction and filter modules of increasing complexity and sophistication, designed to identify a certain category of events. This HLT path with the name *Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL* uses L1 trigger of *L1_DoubleEG_13_7*, i.e. at L1 this trigger requires two ECAL superclusters with energy $\geq$ 13 GeV and 7 GeV. At the HLT level, since now electron track is also reconstructed so the p_{T} requirement of 17 GeV and 8 GeV are imposed, along with the identification and isolation requirements in both calorimeter and tracker.

4.3.2 Electron Identification and Isolation

Electrons must have a p_{T} >20 GeV and are reconstructed in the tracker acceptance, i.e. $|\eta| < 2.5$. Electron identification requirements have been applied based on distribution of electromagnetic shower in the ECAL, a matching trajectory of an electron track with the clusters in the ECAL and consistency with origin of track from the primary vertex to reduce the misidentification rate. Both ECAL and tracker driven electrons, which are not reconstructed in the barrel to endcap transition region (i.e. $1.442 < |\eta| < 1.556$), are considered in this analysis.

Requirement	Barrel		Endcap	
	Loose	Medium	Loose	Medium
$\Delta\eta_{in}$	0.007	0.004	0.01	0.007
$\Delta\phi_{in}$	0.8	0.06	0.7	0.03
$\sigma_{i\eta i\eta}$	0.01	0.01	0.03	0.03
H/E	0.15	0.12	-	0.1
d_0 (cm)	0.04	0.02	0.04	0.02
d_z (cm)	0.2	0.1	0.2	0.1
$\|1/E - 1/p\|$	-	0.05	-	0.05
Missing hits	-	1	-	1

Table 4.1: Electron identification requirements. Symbols used are explained in Section 3.4.4 in Chapter 3.

Table 4.1 summarizes the identification criteria used for electrons. In this analysis medium working point is used for electron identification. Also, electrons are required to pass particle flow isolation as described in Section 3.4.5 of Chapter 3. The isolation requirement is $I_{rel} < 0.15$.

4.3.3 Muon Identification and Isolation

Muons are required to have p_{T} >20 GeV and $|\eta| < 2.4$. Muon candidates can be reconstructed as either tracker or global muons. Requirements are furthermore imposed on the quality of the track fit, number of hits in the pixel, tracker and number of matched muon segments for the muons to be considered for the dilepton candidate. The requirements are summarized in Table 4.2 and muons with tight identification criteria are used in the analysis.

Requirement	Loose	Tight
Global		true
PF	true	true
χ^2/dof (global)		<10
Valid pixel hits		>0
Pixel layers with measurement	>1	
Tracker layers with measurement	>5	>5
Valid muon hits		>0
Matched muon stations		>1
d_0 (cm)	3	0.2
d_z (cm)	30	0.5

Table 4.2: Muon identification requirements. Symbols used are explained in Section 3.5.2 in Chapter 3.

In this analysis, particle flow isolation is used for the muons as described in Section 3.5.3 of Chapter 3. The isolation requirement is $I_{rel} < 0.20$.

4.3.4 Jet Selection

Jets are reconstructed from all the particle flow candidates using the AK5 algorithm, described in Section 3.7. In this analysis, all jets with p_{T} >30 GeV and $|\eta| < 5$ passing the identification requirements summarized in Table 4.3 are considered. In addition to the above, additional identification criteria is applied to remove pile-up contamination in jets.

Variable	Threshold
n	>1
nhf	<0.99
nef	<0.99
cef	<0.99 if $\|\eta\| < 2.4$
chf	>0 if $\|\eta\| < 2.4$
nch	>0 if $\|\eta\| < 2.4$

Table 4.3: Jet identification requirements. Symbols used are explained in Section 3.7.4 in Chapter 3.

4.3.5 b-Tagging

The selected jets are expected to be mostly coming from gluon or light-quarks due to ISR or from the hard process in the VBF case. We reject events in which at least one jet is b-tagged with the jet probability algorithm (JP), explained in Section 3.10. The maximum value of the b-quark discriminator used is 0.275, i.e. events for which the discriminator has value ≥ 0.275 are rejected.

4.3.6 Missing Transverse Energy E_T^{miss}

E_T^{miss} is estimated from the imbalance of the transverse momentum of all the reconstructed particle flow candidates as described in Section 3.8. Particle-Flow E_T^{miss} is used in this analysis. The details of E_T^{miss} selection will be mentioned in Section 4.4.3.

4.3.7 Photon Selection

Photons are not directly used in the analysis but as a control sample for a data-driven estimate of the $Z + jets$ background. Photons are required to have a $p_T > 55$ GeV and $|\eta| <2.5$. Likewise for electrons, photons reconstructed in the barrel to endcap transition region are not considered. Photon is identified if it fulfills a set of basic requirements mentioned in Table 4.4.

Requirement	Barrel	Endcap
$\sigma_{i\phi,i\phi}$	> 0.01	> 0.01
$\sigma_{i\eta,i\eta}$	> 0.001 and < 0.011	< 0.03
R_9	> 0.9	> 0.9
H/E	< 0.05	< 0.05
Pixel Veto	True	True

Table 4.4: Photon identification requirements. Variables are defined in Section 3.6 in Chapter 3.

Photon events are required to have a single photon selected in the barrel region and the jets in photon events are required to be reconstructed with at least $\Delta R >0.4$ from the selected photon.

4.3.8 Primary Vertex

Vertices are required to be reconstructed inside a cylindrical region of length 24 cm and radius 2 cm surrounding the nominal interaction point. They are, in addition, required to have a minimum of four degrees of freedom related with the vertex fit. When multiple vertices are present in an event, the vertex with the highest $\sum_{\text{tracks}} p_T^2$ is chosen as Primary Vertex.

4.4 Event Selection

The selection has been divided into mainly three parts: preselection, event categorization and final selection. Preselection includes the lepton selection and some kinematic requirements related to Z boson. After the preselection, events are categorized based on jet multiplicity and then final selection is based on two observables: E_T^{miss} and transverse mass of Higgs (M_{T}).

4.4.1 Preselection

The preselection of the events used in this analysis consists of the following requirements:

1. Sample cleanup filters : They are applied to minimize the contamination from beam-gas interactions, beam-halo and calorimeter noise. The usage of these filters is advisable to minimize misreconstruction effects and the detailed description can be found elsewhere [102].

2. One of the signal triggers which are summarized in Table A.5;

3. Two same flavor leptons (ee or $\mu\mu$) with selection requirements described in Section 4.3.

4. Invariant mass ($M_{\ell\ell}$) of the dilepton system should be consistent with Z boson mass, $|M_{\ell\ell} - 91| <$15 GeV.

5. p_{T} of the dilepton system $p_{\mathrm{T}}^{\ell\ell} >$55 GeV.

6. No third lepton should be found in the event. The requirements used to identify a third lepton candidate are: Loose-id electrons with $I_{rel} <$0.15 and $p_{\mathrm{T}} >$10 GeV or loose muon with $I_{rel} <$0.2.

7. b-quark jet veto, i.e. no b-quark tagged jet identified by JP algorithm should be present in an event, i.e. JP discriminator is less than 0.275.

8. Soft muon veto is also applied to suppress top-quark related background. Top-quark decays mainly into b-quark and W boson and b-quark can, in-turn, decay to muon with low p_{T}. Thus, events with such muons are rejected. These muons have low p_{T} and do not pass isolation criteria. The p_{T} requirement for such muons is > 3 GeV and no isolation requirement is applied.

9. The minimum azimuthal difference between jets and E_T^{miss}, $\min\Delta\phi(\text{jet},E_T^{miss})$ should be greater than 0.5. This ensures that the E_T^{miss} should be well separated from jets so that there are less chances of jet mismeasurement contributing to E_T^{miss}.

Figure 4.9 shows basic kinematic distributions after preselection (Step 3) for 19.7 fb^{-1} of integrated luminosity at $\sqrt{s}$ = 8 TeV. A fair agreement between data and direct MC prediction is found for both channels. The residual discrepancy in the shape and normalization of the dilepton mass peak is due to the fact that no extra corrections based on momentum scale have been applied on electrons and muons. These are not expected to have significant impact in the final result given that it is not possible to fully reconstruct the mass of the ZZ pair in this analysis.

Figures 4.10 depicts dilepton η and E_T^{miss} distribution for muon and electron channels. The level of agreement between data and simulation in E_T^{miss} distribution is poor because MC is unable to model the instrumental E_T^{miss} arising from the mismeasurement of jets in data. This requires us to build a control region to model the Z+jets from data and will be discussed in detail in Section 4.5.

The number of events surviving each preselection criterion is shown in Figure 4.11. An overall good agreement is found for both $e^+e^-\nu_e\bar{\nu}_e$ and $\mu^+\mu^-\nu_\mu\bar{\nu}_\mu$ channels.

4.4.2 Event categorization

After the preselection, events are categorized based on the number of selected jets with p_{T} > 30 GeV (see Table 4.3 for the corresponding requirements). Events which have at least two jets are further examined and if two jets are found to constitute a so-called VBF-like pair (VBF-tagging), the event is counted in a separate category - the VBF category. If the event fails the 'VBF tag', it is counted in the inclusive jet multiplicity category.

VBF-tagging:

The selection of the VBF category is done after ranking the jets by decreasing p_{T} values. The two leading jets are required to have $p_{\mathrm{T}} > 30$ GeV; pseudorapidity separation $|\Delta\eta| > 4$ and invariant mass of these two jets (M_{jj}) > 500 GeV. The decay products of

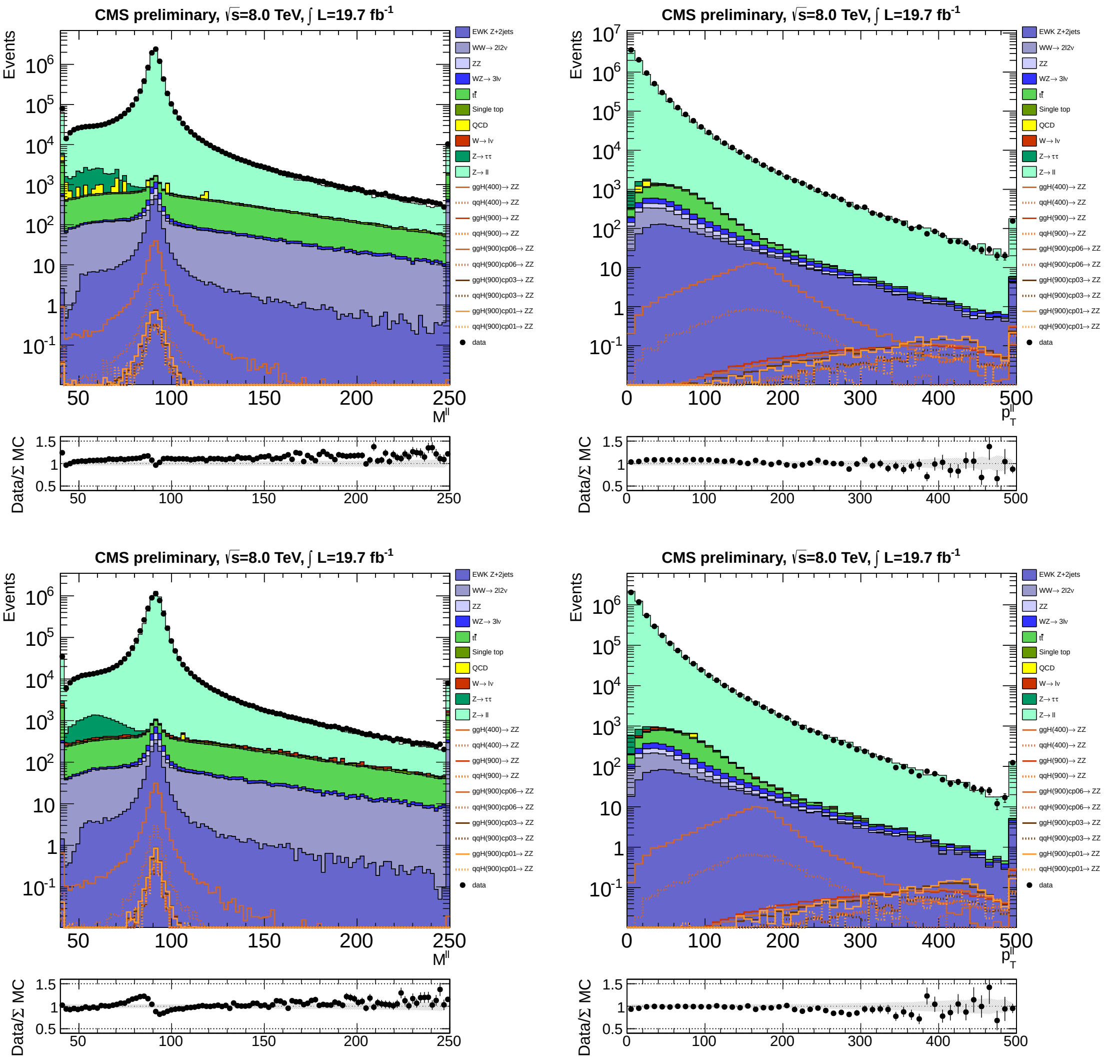


Figure 4.9: From *left* to *right*: dilepton invariant mass distribution and p_{T} distribution after preselection (Step 3). The *top* (*bottom*) distributions correspond to the di-muon (di-electron) channel. Reference signal distributions for M_H =300, 600 and 1000 GeV are superimposed to all distributions.

the Higgs boson are expected to be found in the $\Delta\eta$ region defined by two leading jets. So, no other jet with $p_{\mathrm{T}} > 30$ GeV should be present in this $\Delta\eta$ region.

After the VBF tagging, all the remaining events are divided into exclusive 0-jet and inclusive 1-jet categories ($\geq$ 1-jet). For the VBF selection, control distributions are shown in Figure 4.12 after selecting two jets with $p_{\mathrm{T}} > 30$ GeV but which have not yet passed the VBF tag. Figure 4.12 shows the η distribution of the selected jets and $\Delta\eta$ of two selected jets. There is some disagreement in the dijet invariant mass M_{jj} distribution, however, the statistics in VBF category is small after imposing $|\Delta\eta| > 4$. Figures 4.13

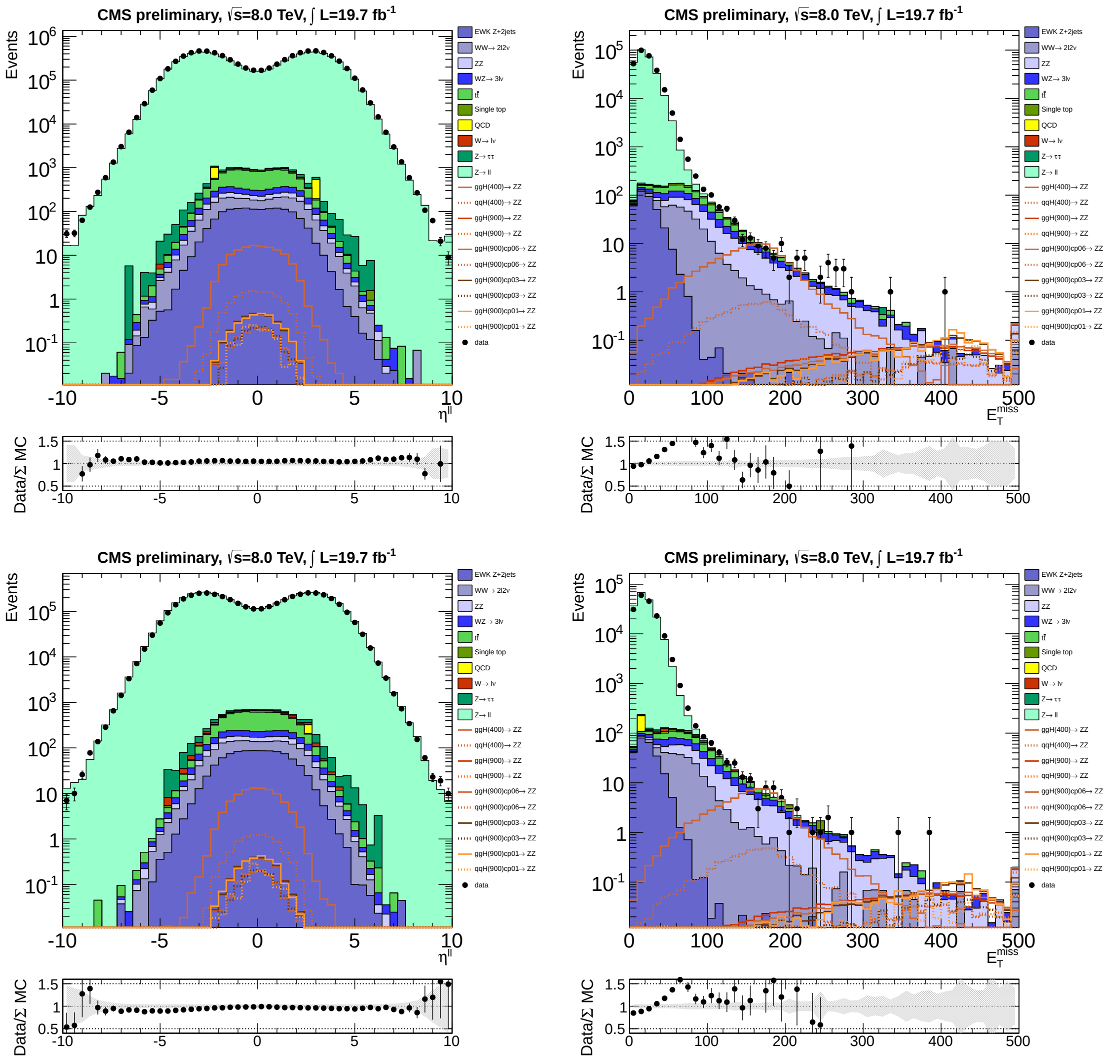


Figure 4.10: From *left* to *right*: dilepton η and E_T^{miss} distribution after preselection (step 3). The *top* (*bottom*) distributions correspond to the di-muon (di-electron) channel. Reference signal distributions for M_H =300, 600 and 1000 GeV are superimposed to all distributions.

show the distribution of the M_{jj} system selected with $|\Delta\eta| > 4$.

4.4.3 Final Selection

As mentioned before, after the preselection, events are categorized based on jet multiplicity. After the selection of VBF-type events, rest of the events are divided into 0-jet and $\geq$ 1-jet category. Final selection consists of applying requirements on the two most significant variables, the E_T^{miss} and the reconstructed transverse mass of the dilepton and E_T^{miss} system ($M_{\rm T}$). The $M_{\rm T}$ variable is defined as follows:

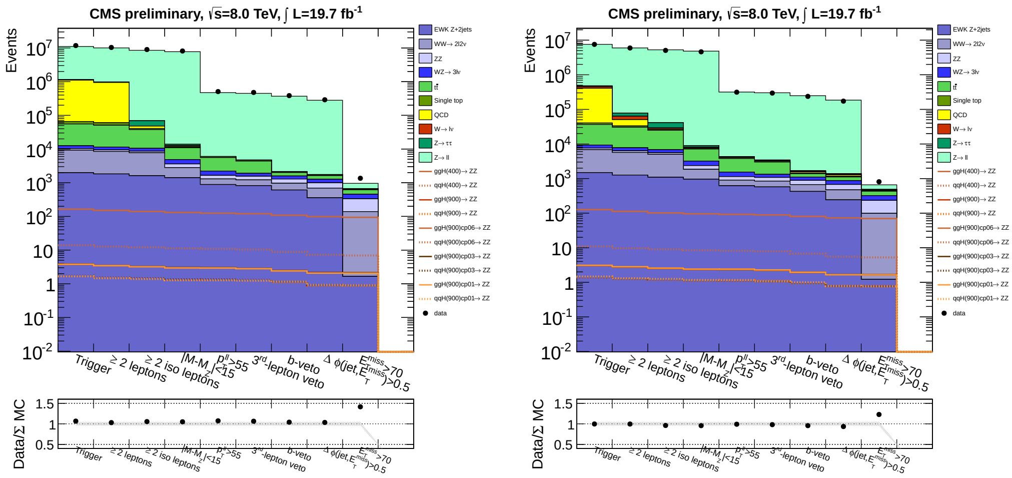


Figure 4.11: Number of events passing each preselection cuts in the di-muon (*left*) or di-electron (*right*) sample. Reference signal distributions for M_H =300, 600 and 1000 GeV are superimposed to all distributions.

$$M_T^2 = \left[\sqrt{p_{T,\ell\ell}^2 + M_{\ell\ell}^2} + \sqrt{E_T^{miss,2} + M(\ell\ell)^2}\right]^2 - \left[\vec{p}_{T,\ell\ell} + \vec{E}_T^{\rm miss}\right]^2 \tag{4.1}$$

where $p_{T,\ell\ell}$ and $M_{\ell\ell}$ are the transverse momentum and invariant mass of the dilepton system, respectively. Final discriminant was chosen to be $M_{\rm T}$ in non-VBF category and E_T^{miss} in the VBF category. Two different approaches are then used. In the baseline (or shape) method, the full distribution of the final discriminating is used without imposing any selection requirement. However, the results are cross-checked with a simpler method where, a selection requirement is imposed on the final discriminant. Requirement on $M_{\rm T}$ in non-VBF category is chosen to be Higgs-mass dependent and is given, together with E_T^{miss}, in Table 4.5. Uniform minimum threshold on E_T^{miss} of 90 GeV is imposed in VBF category without any $M_{\rm T}$ requirement. This method is referred to as the cut-and-count method. The optimisation of these requirements on E_T^{miss} and $M_{\rm T}$ for non-VBF category has been performed using thirteen simulated M_H points. The final selection criteria obtained after optimisation are listed in Table 4.5.

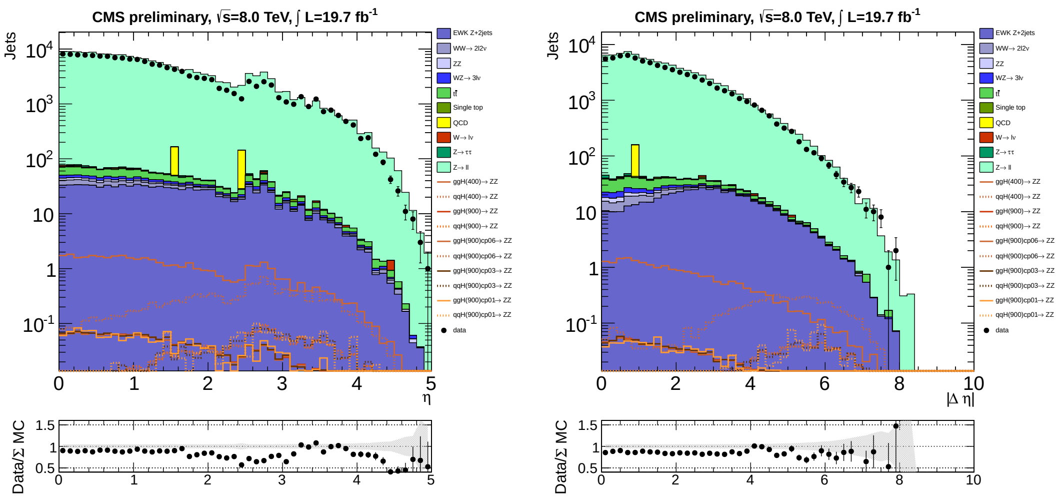


Figure 4.12: *Left* : pseudo-rapidity distribution of the two leading jets. *Right*: rapidity gap of the two leading jets for combination of di-electron and di-muon channels.

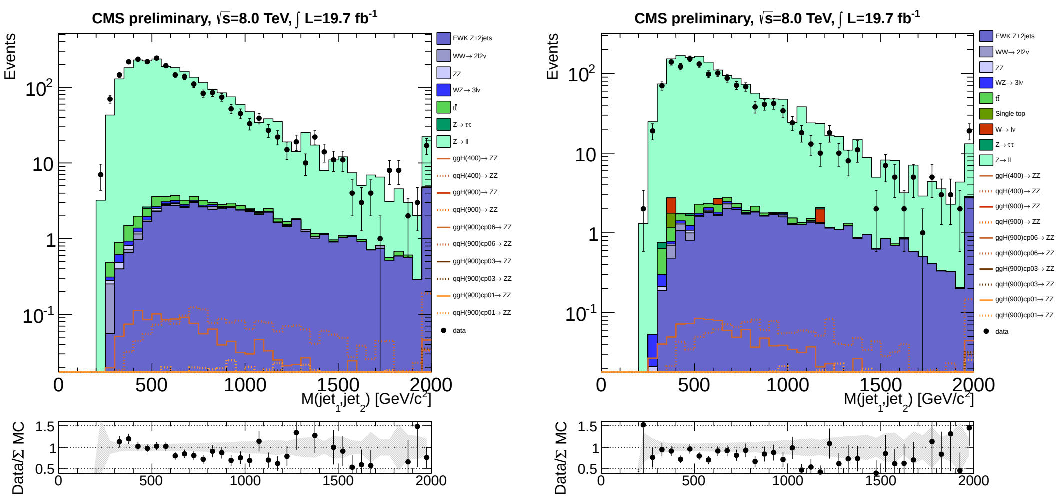


Figure 4.13: Invariant mass of the dijet systems selected with a rapidity gap of $|\Delta\eta| > 4$ in di-muon (*left*) and di-electron (*right*) events.

M_H(GeV)	200	300	400	500	$\geq$600
E_T^{miss} (GeV)	> 80	> 100	> 110	> 110	> 110
M_{T} (GeV)	$180 - 220$	$180 - 270$	$350 - 450$	$400 - 600$	$500 - \infty$

Table 4.5: Higgs boson mass-dependent selection for E_T^{miss} and M_{T} variables for the non-VBF category.

4.5 Background estimation

As described in Section 4.4, there are discrepancies in data-MC distributions particularly in E_T^{miss}. It is very hard to simulate instrumental effects like fake-MET in MC and hence most dominant backgrounds like Z+jets and non-resonant backgrounds like $t\bar{t}$, single-top quark and WW are estimated using data-driven techniques. In this section, these techniques are described.

4.5.1 Non-resonant background

This type of background includes the ones which do not have a real Z boson. These include top quark production (single or in pairs), WW and W+jets. Opposite flavor dilepton events (i.e. $e\mu$ pairs) that pass the signal event selection are used to estimate this background. For W+jets, the second lepton is usually a fake electron which will be present in the $e\mu$ sample as well. Additional contributions from WZ, and part of Z and ZZ are also there involving leptonically decaying taus. We rescale the number of $e\mu$ events passing preselection cuts (except the same flavour requirement on dileptons) with a factor α_ℓ, which is calculated from the side-bands. The procedure consists of two steps.
Step 1: α_ℓ is the correction factor which takes into account the difference in the lepton selection between the same flavor channel - $\ell\ell$ (ee or $\mu\mu$) and the opposite flavor channel - $e\mu$. This factor is calculated from the sidebands of $M_{\ell\ell}$ distribution, defined as: 55< $M_{\ell\ell}$ <70 GeV and 110< $M_{\ell\ell}$ < 200 GeV. α_ℓ is given as:

$$\alpha_\ell = N_{\ell\ell}^{\text{out}}/N_{e\mu}^{\text{out}} \tag{4.2}$$

where

- $N_{\ell\ell}^{\text{out}}$ = Number of ee or $\mu\mu$ events in $M_{\ell\ell}$ side-band regions passing the preselection, except $M_{\ell\ell}$ requirement and also reversing the b-quark jet veto (i.e. we select events with at least one b-tagged jet).

- $N_{e\mu}^{\text{out}}$ = Number of $e\mu$ events in $M_{\ell\ell}$ side-band regions passing the preselection, except $M_{\ell\ell}$ requirement and reversing the b-quark jet veto.

E_T^{miss} >80 GeV is also applied to reduce Z+jets contamination. Table 4.6 lists the values of α_ℓ used in the analysis, which are computed both from data and MC. Values from data are used to estimate this background.

α_ℓ	α_e	α_μ
Data	0.429 ± 0.015	0.632 ± 0.019
MC	0.425 ± 0.004	0.616 ± 0.006

Table 4.6: α_ℓ computation.

Step 2: In the second step, the value of α_ℓ is used to estimate the non-resonant background $N_{\ell\ell}^{\text{NR}}$ as given by:

$$N_{\ell\ell}^{\text{NR}} = \alpha_\ell \cdot N_{e\mu}^{\text{in}} \tag{4.3}$$

where, $N_{e\mu}^{\text{in}}$ = Number of $e\mu$ events after full preselection, except the same flavor requirement. Using the α_ℓ values mentioned in Table 4.6, a validation test also referred as 'closure test' is performed for this method. Events are counted directly from the background MC samples after the preselection, except that the b-quark jet requirement is reversed to enrich the sample with top-quark events. These are called 'Expected' number of events. Then events are predicted from same MC samples using the above defined method, which gives 'Predicted' number of events. Table 4.7 shows the expected and predicted number of events for non-resonant backgrounds. The expected and predicted numbers agree well within the statistical errors, which gives us confidence in the method.

Channel	$N_{\ell\ell}^{\text{NR}}$(Expected)	$N_{\ell\ell}^{\text{NR}}$(Predicted)
ee	541 ± 6	554 ± 10
$\mu\mu$	790 ± 8	811 ± 10

Table 4.7: Closure test for expected and predicted non-resonant background.

Further cross-checks are performed for the validation of this method by using both the b-vetoed and the b-tagged sample as well as the inclusive and right-side sidebands. The test is performed for each channel (ee and $\mu\mu$) inclusively or for the different categories (depending on the jet multiplicity). For each E_T^{miss} requirement, we compare the prediction from this method to the expectation and evaluate the '*bias*' as:

$$bias = 100 \cdot \left(\frac{N_{\text{predicted}}}{N_{\text{expected}}} - 1 \right) \tag{4.4}$$

Figure 4.14 shows the result of the closure test. Although the results for different E_T^{miss} requirements are correlated, given that they partially use the same subset of events, we don't observe any particular *bias* within 5-10%. We assign therefore a 10% uncertainty on the prediction of the method. Possible biases due to differences in the trigger paths are expected to be included in this uncertainty given that trigger was required in the simulation to perform these closure tests. It must also be noticed from Table A.5 that the legs of the $e\mu$ triggers are similar to the ones used to select the $ee/\mu\mu$ samples. The only exception is the single muon trigger. For the VBF category the statistics is not sufficient to perform this validation. In this channel we expect, however, that non-resonant background contribution is negligible.

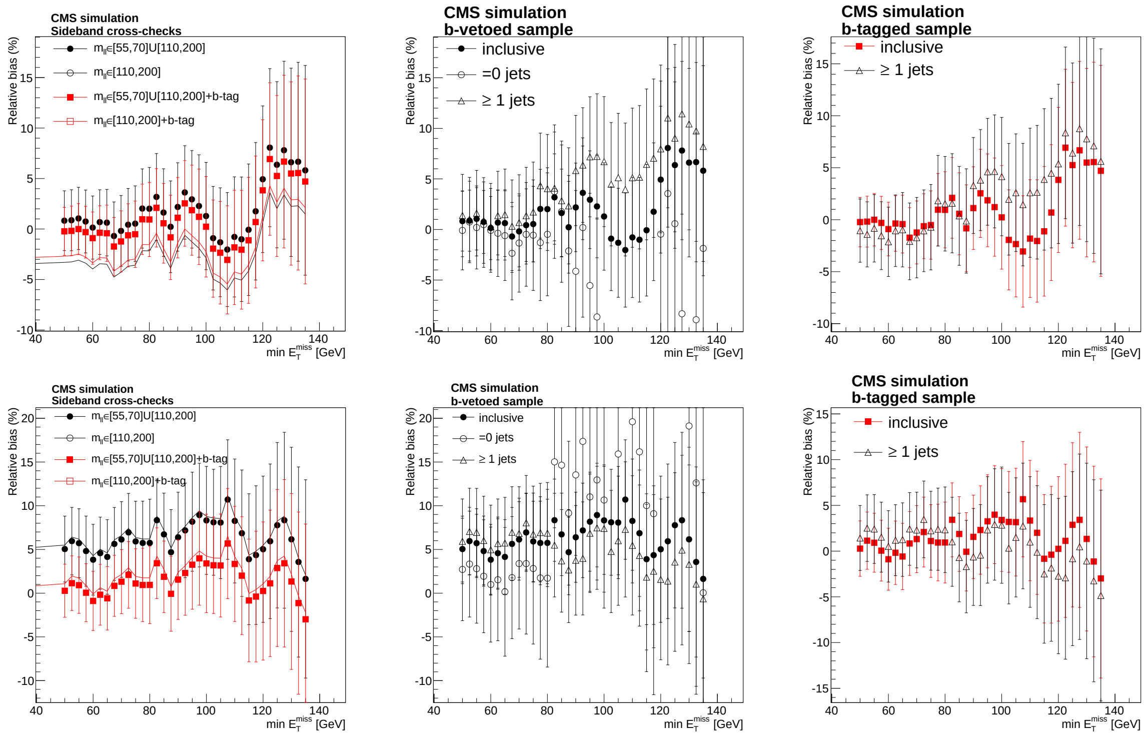


Figure 4.14: Validation of the non-resonant background estimation method for the di-electron(di-muon) channel is shown on the top(bottom) three plots as a function of E_T^{miss}. *Left* : inclusive sample using different side-band definitions. *Center* : exclusive samples with b-jet veto. *Right* : exclusive samples requiring a b-tagged jet.

4.5.2 $Z + jets$

In this background process, fake E_T^{miss} arises because of the instrumental effects, and is very hard to model in MC. The discrepancy is visible in Figure 4.10. So, this background is estimated from data. To estimate this background a control sample was chosen from single isolated photon sample from data. Single isolated photons are generated through the same processes as Z bosons and are expected to preserve the similar kinematics. Since the cross-section of γ+jets process is an order of magnitude higher than Z+jets so, even being acquired through pre-scaled triggers, the statistics of the γ+jets sample is significantly higher than that of the dilepton sample. The procedure adopted is the following:

1. Select events triggered by a single photon trigger and categorize the photon event according to the number of jets.

2. Reweight the p_{T} spectrum of the photon sample to match the p_{T} spectrum of the dilepton sample in each jet multiplicity bin.

3. The vertex distributions is also reweighted in order to minimize bias from trigger pre-scaling or due to the efficiency of the photon selection.

4. After reweighting the distributions of interest, i.e. E_T^{miss} and M_{T} are computed from the photon sample and used as models of the Z+jets.

5. The final normalization for the shapes is obtained for each jet multiplicity bin from the region with E_T^{miss} <50 GeV.

In order to minimize the threshold effects and statistical fluctuations we begin by building a weighted photon p_{T} spectrum by using the trigger prescales as event weights. Then both, the photon p_{T} and the dilepton spectra are fitted with the following function:

$$f(p_T) = \sum_{i=1}^{N} \mathcal{S}_i(p_T) \cdot \left(\frac{p_T}{a_i}\right)^{b_i} \cdot \frac{e^{-p_T/a_i}}{1 + e^{\frac{c_i - p_T}{d_i}}} \tag{4.5}$$

where the sum is taken over the different trigger p_{T} thresholds used (22,36,50,75,90) GeV; $\mathcal{S}_i$ is a step function which takes the value of 1 if the p_{T} is higher than the i^{th} trigger threshold and lower than the i^{th}+1 and 0 otherwise; and a_i, b_i, c_i and d_i are free parameters to be determined from the fit. The motivation for Equation (4.5) is as follows: a power law describes the p_{T}-falling boson spectrum and the Fermi functions describe the activation of a new trigger threshold. Given that the usage of the step function generates discontinuities in the vicinity of each threshold an ad-hoc regularization procedure was applied by interpolating a second order polynomial to the p_{T} spectrum using the two intervals preceding the interval corresponding to each threshold. In order to evaluate the validity of the method a validation test using simulated photon and DY MC samples is performed by applying the steps described above. The resulting distributions for E_T^{miss} are shown in Figure 4.15 for the different categories considered in the analysis and for the inclusive samples. Each figure shows below the prediction bias which is the relative difference between the number of events predicted by the reweighted γ+jets MC sample and the one expected in the DY MC sample. In the high E_T^{miss} tails we run out of statistics but one can conclude overall that the prediction is well inside a 50% uncertainty band.

Further details of the validation are given in Appendix B. Figures 4.16, 4.17, 4.18 show that the most important control distributions are reproduced where the main background (DY) is now modeled using the predictions from the γ+jets data sample as templates. The E_T^{miss} distributions after applying the above procedure are shown in Figure 4.16. A good agreement is found between data and prediction.

Figure 4.17 shows the distribution of the $\min\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{miss})$ variable for each jet multiplicity interval after applying $E_T^{miss} > 50$ GeV. A slight deficit (excess) is found in the ee ($\mu\mu$) channel with at least one jet at high values of $\min\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{miss})$. These discrepancies are covered by the final uncertainty assigned to the method.

An overall good agreement is found for the M_{T} distributions in all the different categories as shown in Figure 4.18.

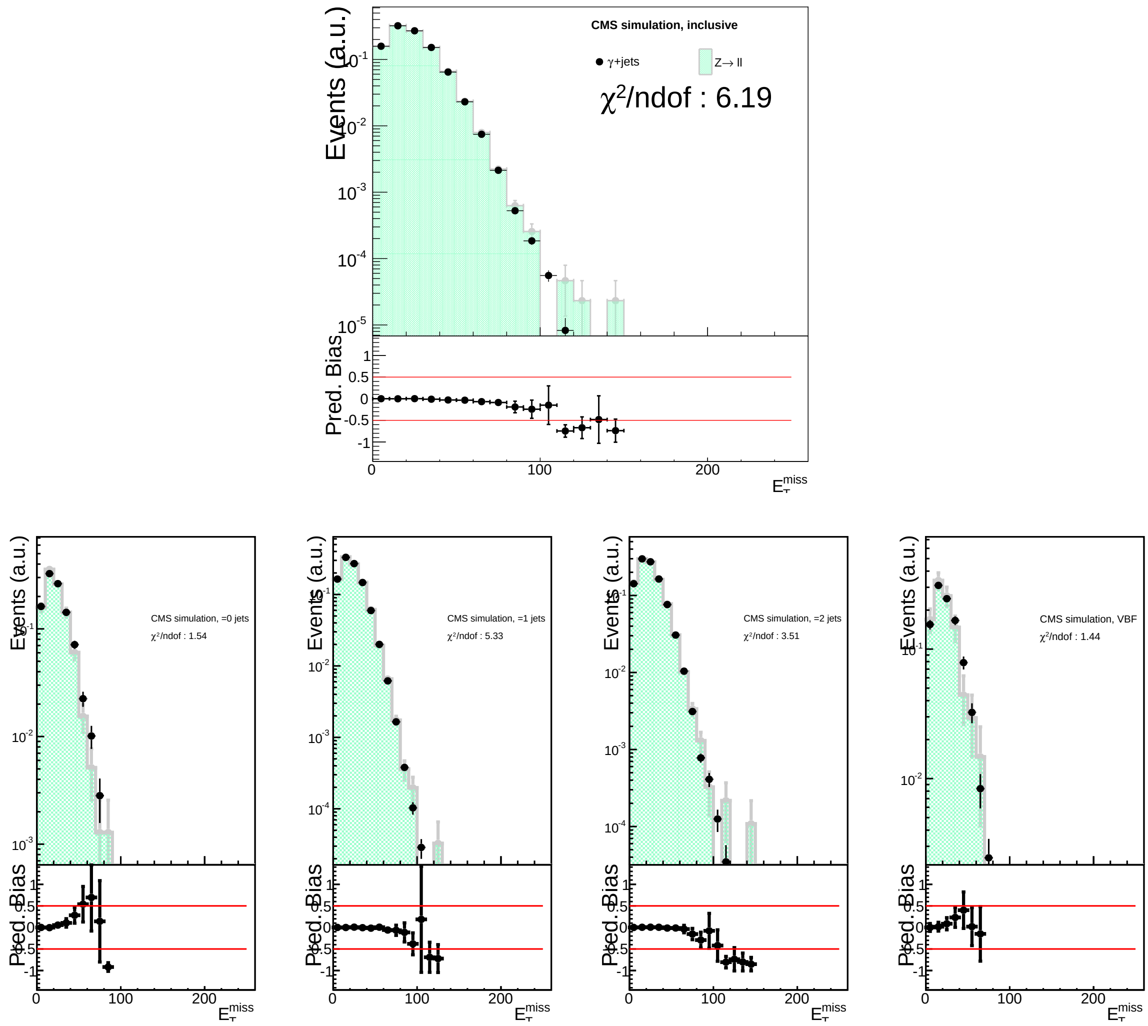


Figure 4.15: Validation test of the Z+jets background estimation method for the E_T^{miss} variable using MC samples. The grey distributions correspond to the DY MC sample and the points to the $\gamma + jets$ MC sample after applying the reweighting procedure. In bottom plots from *left* to *right* the exclusive jet categories correspond to events with 0, 1, $\geq$ 2 and VBF-like.

Figure 4.19 depicts that although MC prediction might not be accurate in particular in phase space corners of the selection such as the VBF tag, still a fair, modeling is obtained from the $\gamma + jets$ sample. It may be noted that jets with low η tend to be overestimated by this method. This is related to the fact that the selection phase space of the $\gamma + jets$ sample is not fully matched to the Z phase space due to the barrel-only requirements. Given that VBF jets tend to have higher η (and $\Delta\eta$), this small discrepancy is not significant for the final result.

Other backgrounds like ZZ and WZ are taken directly from the MC prediction.

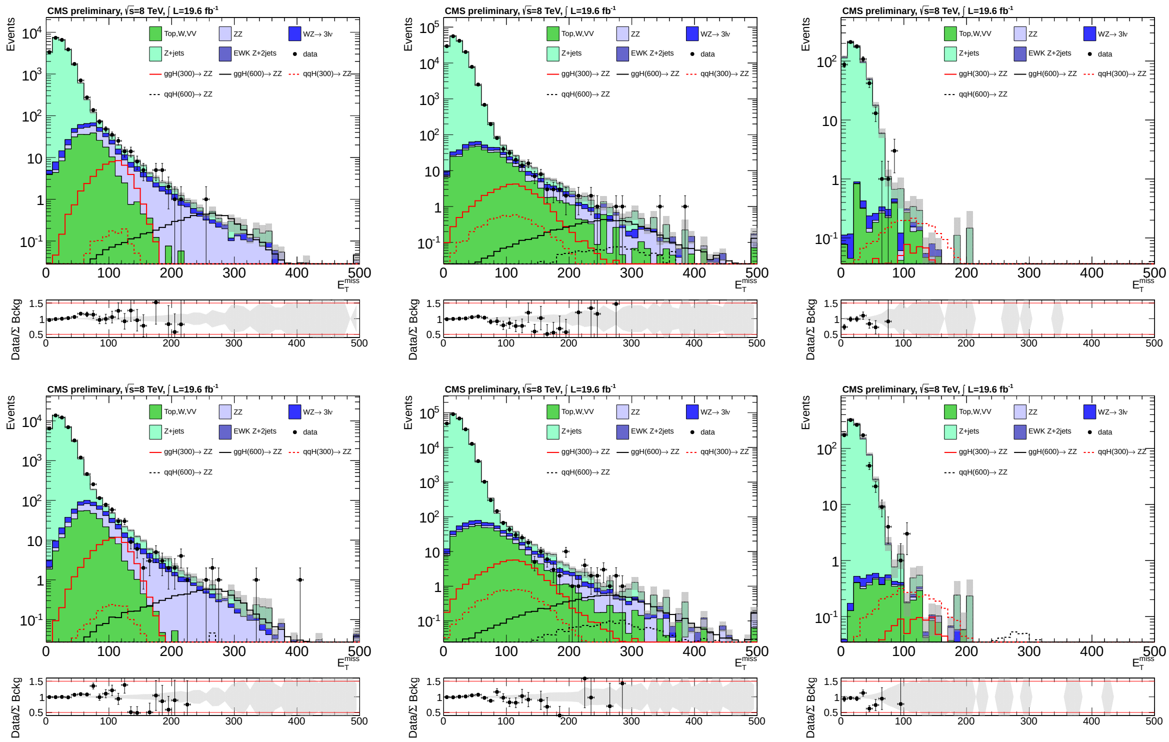


Figure 4.16: From *left* to *right*: E_T^{miss} spectrum in di-electron (*top*) and di-muon (*bottom*) events with 0, at least 1 or VBF tagged jets.

4.6 Selection efficiencies

This section describes the various efficiencies related to the leptons and also to other selection requirements which are used in the analysis. For every step of the lepton selection, i.e. trigger, lepton identification and isolation, there is a probability to mis-identify a lepton which has been poorly reconstructed. There is also the probability that CMS full simulation does not take fully into account how pile-up or material effects distort the lepton identification. In order to correct for these effects, the efficiencies of our selections have been computed from data using the tag and probe [103] technique. This method consists in selecting events where a very good electron/muon candidate (called the tag) and another loosely identified electron/muon candidate (called the probe) can be found. The invariant mass of the reconstructed pair must be compatible with a well measured SM particle and for the region of kinematical interest used in this analysis, we choose the dilepton to be compatible with a Z boson decay. This condition enforces the probability that the probe is also a true electron/muon. By studying the probe selection after addi-

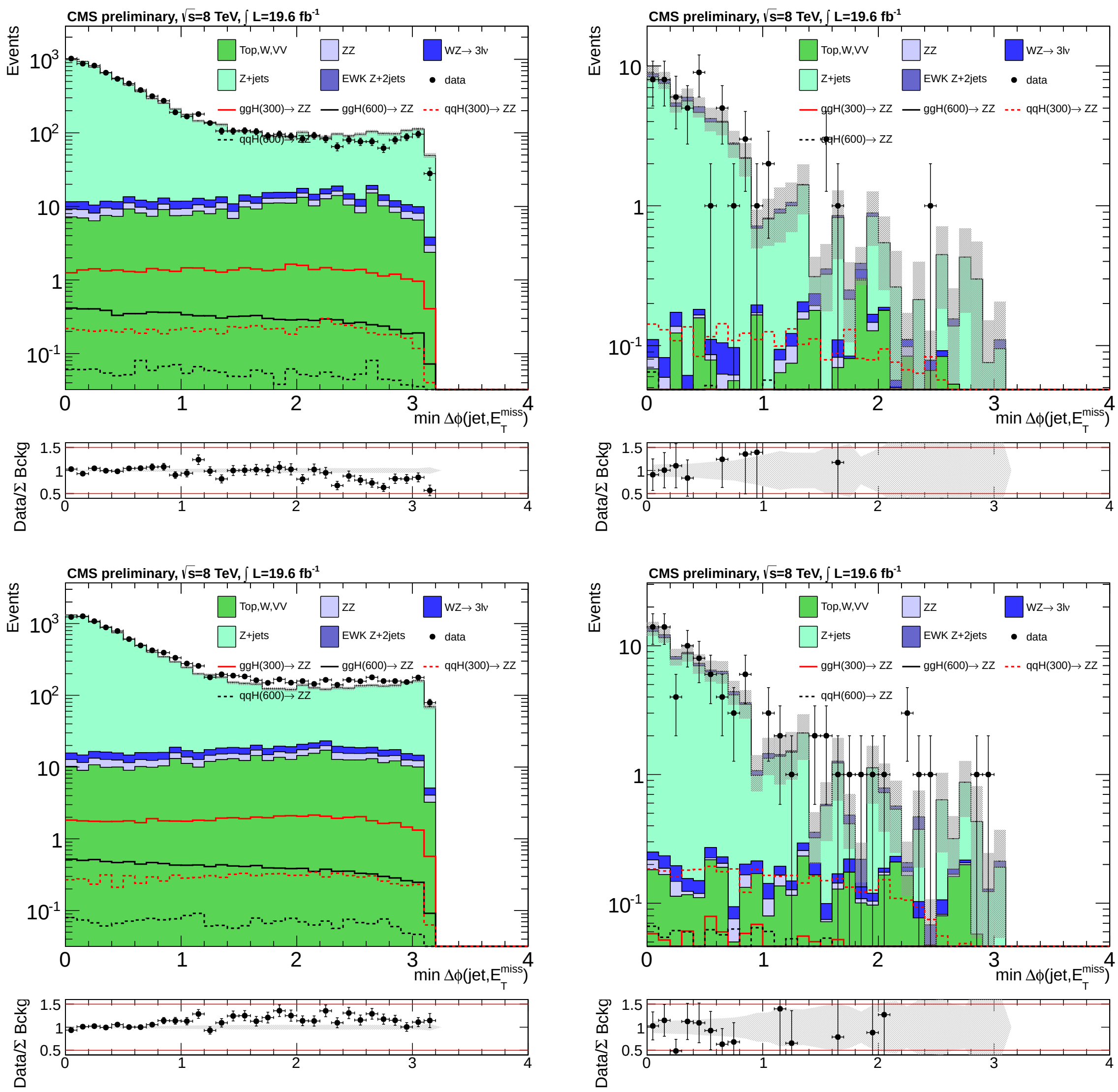


Figure 4.17: From *left* to *right*: min $\Delta\phi(\text{jet}, E_{\mathrm{T}}^{miss})$ variable in di-electron (*top*) and di-muon (*bottom*) events with E_T^{miss} >50 GeV and with at least 1 or VBF-tagged jets.

tional criteria are imposed, it is possible to deduce the efficiencies of the selection and the necessary corrections to be applied to our simulation of the signal (and the backgrounds assumed from MC).

4.6.1 Trigger efficiencies

Dilepton triggers used in this analysis are mentioned in Table A.5. The efficiency of the triggers used in this analysis is evaluated with a tag and probe method, and reweighting is applied to the MC in the limit computations. The results are shown in Figure 4.20 for

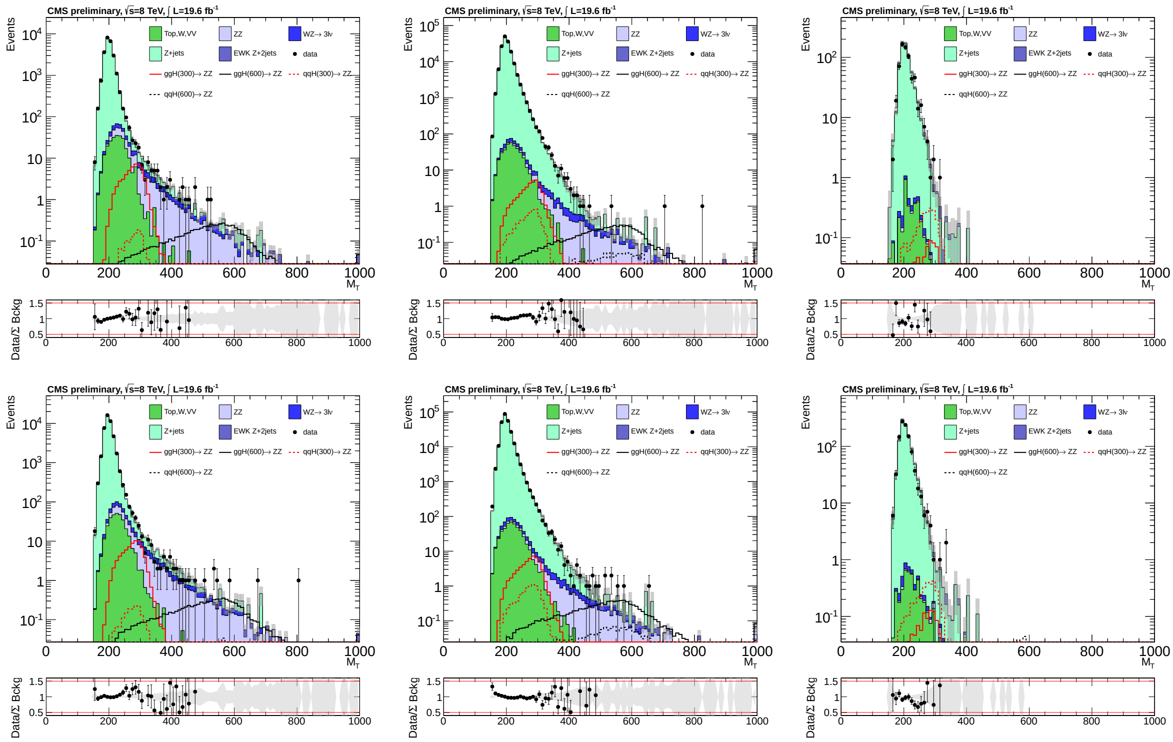


Figure 4.18: From *left* to *right*: M_{T} spectrum in di-electron (*top*) and di-muon (*bottom*) events with exactly 0, at least 1 or VBF tagged jets.

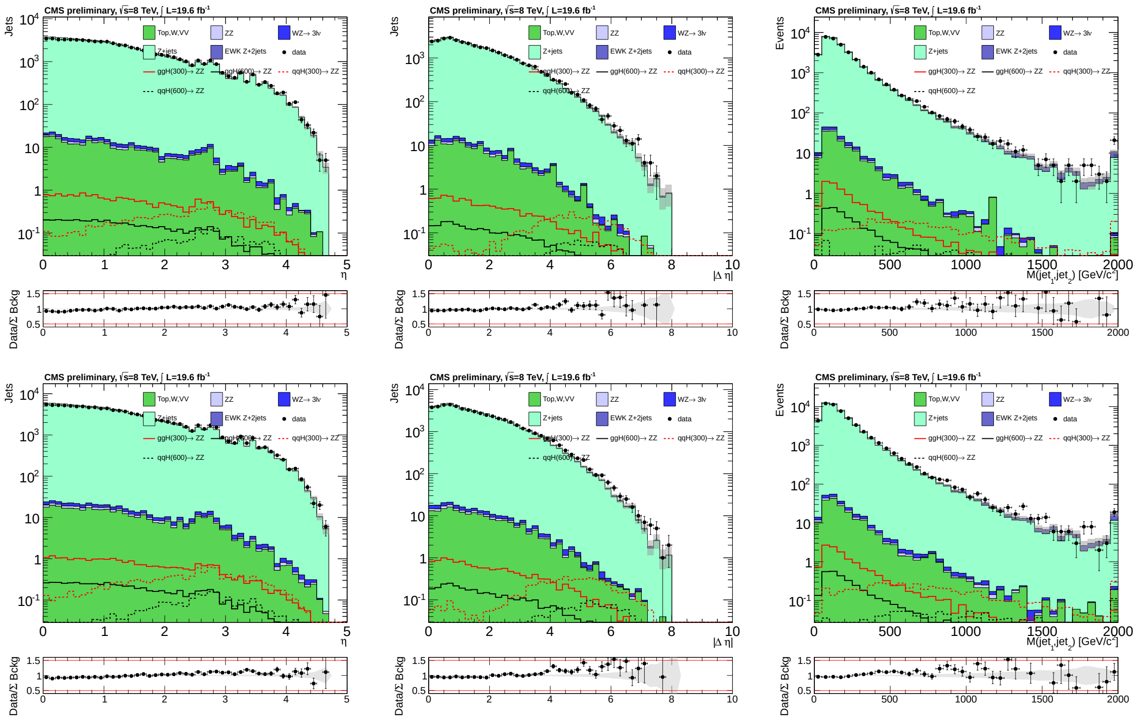


Figure 4.19: VBF selection distributions including the data-driven prediction for the Z+jets contribution in the $\sqrt{s}$ = 8 TeV dataset. The di-electron (di-muon) channel is shown on the *top* (*bottom*) plots. From *left* to *right*: rapidity distribution of the two leading jets, rapidity gap of the two leading jets and invariant mass of the dijet system.

electron triggers, and in Figure 4.21 for muon triggers.

Electrons: For the di-electron triggers two electrons are selected whose invariant mass is found to be compatible with the Z boson mass, i.e. $|M(\ell\ell) - M_Z| <$15 GeV. For both electrons, identification and isolation cuts were applied as specified in Section 4.3. If the pair survives the selection criteria then we choose randomly one to be the tag electron and the other to be the probe electron. To increase the purity of sample the tag electron is required to pass another trigger. The geometrical matching to the trigger object was done with the electron by using a cone of ΔR =0.2. For the probe electron separate checks were done for both legs of the di-electron trigger listed in Table A.5. The trigger efficiency measured with respect to the offline selection is therefore defined as function of the p_{T} and η of the reconstructed probe as follows:

$$\varepsilon_{trig}(p_{\mathrm{T}}, \eta) = \frac{N^{pass}_{probes}}{N_{probes}} \tag{4.6}$$

where N_{probes} is the total number of probe electrons and N^{pass}_{probes} is the number of probes passing the trigger requirement. Figure 4.20 shows the results obtained for the trigger efficiency measured in data. The overall efficiency is above 96% for both legs.

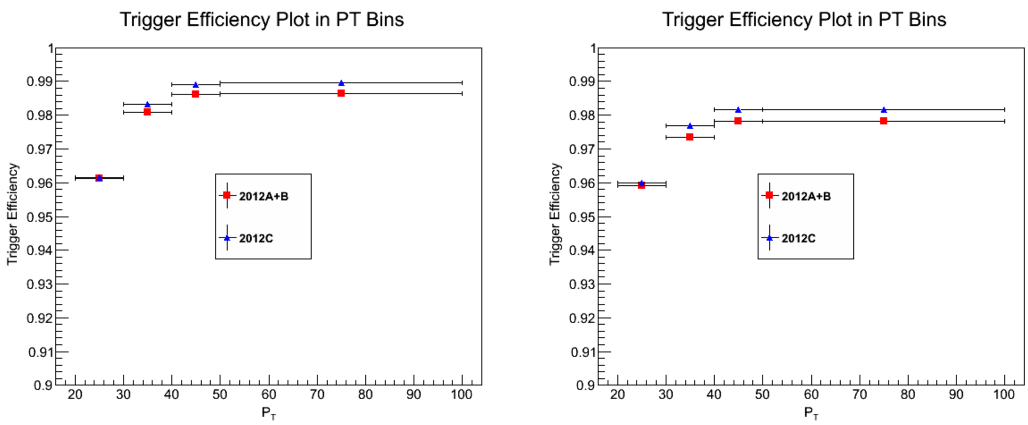


Figure 4.20: Di-electron trigger efficiency measured in data as function of the p_{T} of each trigger candidate. The *left* (*right*) plot correspond to the first (second) leg of the trigger. The different points correspond to the different data periods, 2012A data period corresponds to an integrated luminosity of around $1fb^{-1}$, 2012B corresponds to around 4.4 $fb^{-1}$and 2012C corresponds to around 6.9 fb^{-1}.

Muons: The double muon trigger efficiency is also calculated per leg. Single muon trigger is used to gain in trigger efficiency. The tag muon is required to pass the tight muon identification and the isolation criteria as described in Section 4.3 and is required to pass another trigger with $p_{\mathrm{T}} > 24$ GeV requirement. The matching between the reconstructed muon and the trigger leg is done using a cone of $\Delta R = 0.2$. The probe is found after requiring the standard muon selection chosen for analysis and imposing the Z mass constraint, i.e. $|M(\ell\ell) - M_Z| < 15$ GeV. Figure 4.21 shows the results obtained for the trigger efficiency measured in data. The overall efficiency is above 91%.

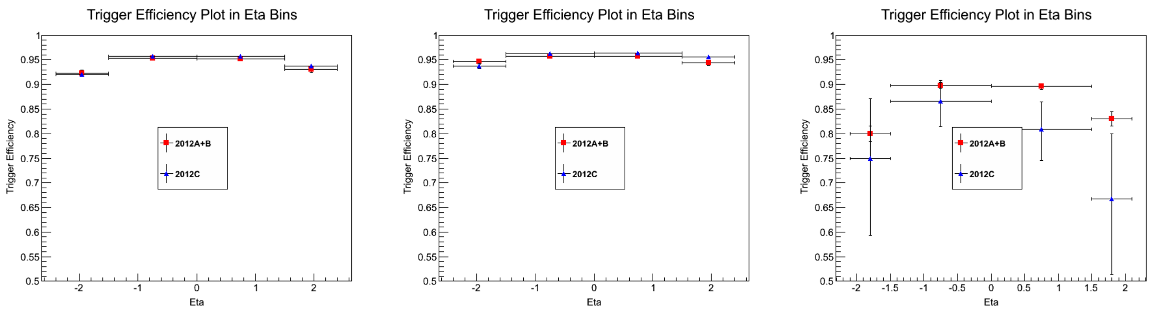


Figure 4.21: Trigger efficiency for the first (*left*) and second (*center*) leg of the di-muon trigger and for the single muon trigger (*right*) used in the analysis as function of the pseudo-rapidity of the muon. The different points correspond to the different data periods.

The efficiencies measured above are applied to analysis using a combination of each leg. Each simulated event is reweighted using the following formula:

$$\varepsilon_{trig}(p_{\mathrm{T}}^1, \eta^1, p_{\mathrm{T}}^2, \eta^2) = \varepsilon_d(p_{\mathrm{T}}^1, \eta^1)\varepsilon_d(p_{\mathrm{T}}^2, \eta^2) + \varepsilon_s(p_{\mathrm{T}}^1, \eta^1)[1 - \varepsilon_d(p_{\mathrm{T}}^2, \eta^2)] + [1 - \varepsilon_d(p_{\mathrm{T}}^1, \eta^1)]\varepsilon_s(p_{\mathrm{T}}^2, \eta^2) \tag{4.7}$$

where ε_d (ε_s) denotes the efficiency measured for a (the) leg of a double (single) trigger and $({p_{\mathrm{T}}}^i, \eta^i)$ denotes the transverse momentum and pseudo-rapidity of the ith lepton.

4.6.2 Lepton identification efficiencies

Leptons identification and isolation efficiencies are also calculated using tag and probe method. These efficiencies are measured from data and also from MC and then a ratio of data over MC is calculated, which is called 'scale factor'. Later on, MC are reweighted by these scale factors. These are calculated in various η and p_{T} ranges of leptons. The

Scale Factors for $0. < \|\eta\| \leq 1.556$			
Range	$0. < \|\eta\| \leq 0.8$	$0.8 < \|\eta\| \leq 1.442$	$1.442 < \|\eta\| \leq 1.556$
$20 < p_T \leq 30$	1.010 ± 0.003	0.981 ± 0.006	1.046 ± 0.003
$30 < p_T \leq 40$	1.006 ± 0.000	0.987 ± 0.000	1.011 ± 0.001
$40 < p_T \leq 50$	1.009 ± 0.001	0.993 ± 0.001	0.994 ± 0.002
$50 < p_T \leq 200$	1.008 ± 0.002	0.995 ± 0.001	0.997 ± 0.002

Table 4.8: Scale factors for the electron identification+isolation efficiency vs. p_{T} for barrel region of the ECAL. The statistical uncertainties are shown.

Scale Factors for $1.556 < \|\eta\| \leq 2.5$		
Range	$1.566 < \|\eta\| \leq 2.0$	$2.0 < \|\eta\| \leq 2.5$
$20 < p_T \leq 30$	0.992 ± 0.009	1.045 ± 0.005
$30 < p_T \leq 40$	0.993 ± 0.000	1.031 ± 0.000
$40 < p_T \leq 50$	1.008 ± 0.000	1.019 ± 0.000
$50 < p_T \leq 200$	1.009 ± 0.000	1.014 ± 0.001

Table 4.9: Scale factors for the electron identification+isolation efficiency vs. p_{T} for endcap region of the ECAL. The statistical uncertainties are shown.

scale factors for electrons identification and isolation that are applied in analysis are summarized in Tables 4.8 and 4.9. The scale factors for muons are shown in Table 4.10. All the scale factors are consistent with one, within associated uncertainties.

4.6.3 VBF efficiencies

In VBF type of events, we have two jets in the final state, which are affected by pile-up. In order to ensure that the VBF selection is independent of pile-up we need to check if the VBF efficiencies are robust against pile-up. So, VBF tagging efficiency is studied as a function of pile-up for standard PF jets as described in Section 4.3.4, before and after requiring a cut and count pile-up jet identification as described in [104]. Although we use jets with at least 30 GeV, however, after applying a requirement of dilepton $p_{\mathrm{T}} > 55$ GeV the jet p_{T} distribution is biased to larger p_{T} values. This makes the VBF selection in this analysis somehow robust to pile-up given that higher the jet p_{T} the less probable it is to be a jet due to the overlap of pile-up energy deposits in the calorimeters. This effect is shown in Figure 4.22 where one can see that both the VBF tagging efficiency in VBF Higgs signal at M_H= 200 GeV and in Z+jets (mistagging efficiency) are constant

Scale Factors for Muon identification				
Range	$0 < \|\eta\| < 0.9$	$0.9 < \|\eta\| \leq 1.2$	$1.2 < \|\eta\| \leq 2.1$	$2.1 < \|\eta\| \leq 2.4$
$20 < p_T \leq 25$	0.989 ± 0.002	0.987 ± 0.003	1.002 ± 0.001	0.994 ± 0.001
$25 < p_T \leq 30$	0.994 ± 0.001	0.994 ± 0.001	0.999 ± 0.001	
$30 < p_T \leq 35$	0.994 ± 0.001	0.991 ± 0.001	0.998 ± 0.001	
$35 < p_T \leq 40$	0.994 ± 0.000	0.990 ± 0.001	0.997 ± 0.001	
$40 < p_T \leq 50$	0.992 ± 0.000	0.990 ± 0.000	0.997 ± 0.000	
$50 < p_T \leq 60$	0.991 ± 0.001	0.991 ± 0.001	0.998 ± 0.001	
$60 < p_T \leq 90$	0.990 ± 0.001	0.986 ± 0.002	0.994 ± 0.002	
$90 < p_T \leq 140$	1.004 ± 0.003	1.012 ± 0.007	1.019 ± 0.006	
$140 < p_T \leq 300$	1.028 ± 0.019	0.956 ± 0.034	1.016 ± 0.034	
$300 < p_T \leq 500$	1.000 ± 0.162	1.000 ± 0.255	0.609 ± 0.415	
Scale Factors for Muon isolation				
Range	$0 < \|\eta\| < 0.9$	$0.9 < \|\eta\| \leq 1.2$	$1.2 < \|\eta\| \leq 2.1$	$2.1 < \|\eta\| \leq 2.4$
$20 < p_T \leq 25$	0.987 ± 0.001	0.992 ± 0.002	1.000 ± 0.001	1.028 ± 0.000
$25 < p_T \leq 30$	1.001 ± 0.001	1.004 ± 0.001	1.006 ± 0.001	
$30 < p_T \leq 35$	0.999 ± 0.000	1.002 ± 0.001	1.005 ± 0.000	
$35 < p_T \leq 40$	0.999 ± 0.000	1.001 ± 0.000	1.002 ± 0.000	
$40 < p_T \leq 50$	0.998 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	
$50 < p_T \leq 60$	0.999 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	
$60 < p_T \leq 90$	1.001 ± 0.000	1.000 ± 0.000	1.001 ± 0.000	
$90 < p_T \leq 140$	1.001 ± 0.001	0.999 ± 0.001	1.000 ± 0.001	
$140 < p_T \leq 300$	1.001 ± 0.001	1.000 ± 0.002	1.002 ± 0.002	
$300 < p_T \leq 500$	1.011 ± 0.000	1.000 ± 0.016	1.000 ± 0.012	

Table 4.10: Scale factors for Muon identification and isolation efficiencies vs. η for muon transverse momenta between 20 and 500 GeV. Uncertainties are of statistical nature.

as function of the number of vertices independently of requiring or not the pile-up jet identification.

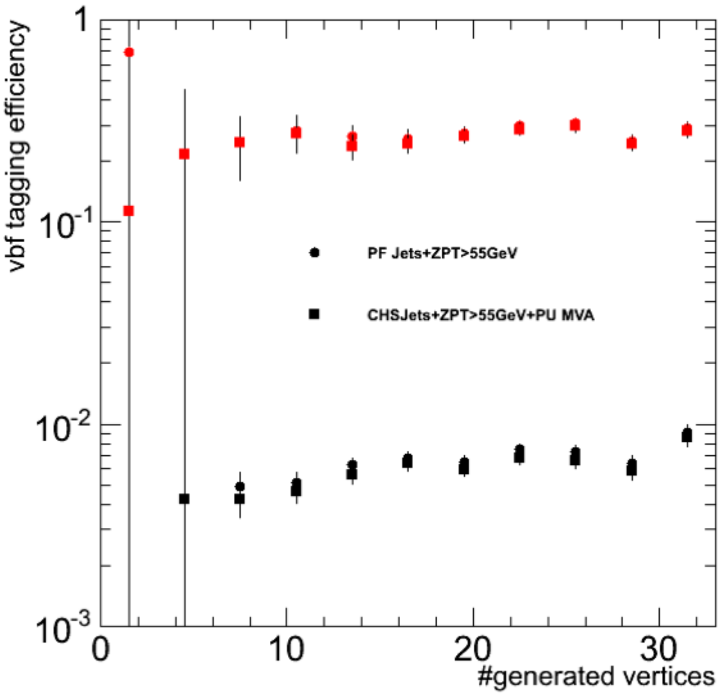


Figure 4.22: VBF efficiency versus number of vertices for Z+Jets (black) and VBF signal at M_H= 200 (red). Circular dots are without pile-up jet identification and square dots are with pile-up jet identification.

4.7 Systematic uncertainties

The sources of systematic uncertainties in estimating signal and background yields are discussed in this section. The sources are categorized into two main categories: instrumental (comprising energy scales, efficiencies, etc.) and theoretical (comprising signal cross-section and shape among others). The description of each one is given below.

4.7.1 Instrumental Uncertainties

1. **Luminosity**

   The uncertainty on the luminosity measurement is estimated to be 4.4% in 2012 data and 2.2% in 2011 data [105].

2. **Trigger**

The uncertainty on the trigger efficiency is estimated to be 1% for the di-electron sample. The di-electron trigger efficiency is slightly below 100% and the systematic uncertainty is estimated at 2%. Following studies of the di-muon trigger efficiency as a function of the p_{T} of the Z boson as well as the opening angle between the two muons, we assign a 2% systematic uncertainty for the muon trigger efficiency.

3. **Lepton identification and isolation**

Identification and isolation efficiencies for leptons are calculated by tag and probe method in two ways. One is counting method in which number of total probe electrons and those which pass the identification and isolation criteria are counted after applying Z boson selection and then calculate the efficiency. In second method, instead of applying selection criteria, dilepton invariant mass distribution is fitted both for signal having reconstructed and identified leptons consistent with Z boson decay and background having fake leptons. The difference in efficiencies from these two methods is quoted as systematic uncertainty. We have found that this uncertainty does not exceed 1% for each muon and 1% for each electron. Therefore, we assume 2% uncertainty for dilepton events.

4. **Lepton momentum scale**

The uncertainty on the scale of the muon momentum measurement is assumed to be 1%. In the electron case we apply an uncertainty of 2% (4.5%) to the electron p_{T} when reconstructed in ECAL barrel (endcap) region. The uncertainty due to the lepton p_{T} is estimated by shifting the nominal p_{T} by one standard deviation and by propagating this change to the measured E_T^{miss}. The resulting variation in yields is found to be very small in both cases (≈1-2%) but it is found to have a non-negligible effect on the shape of the distributions as shown in Figure 4.23. Table 4.11 shows the uncertainty (in %) obtained per event category. Very high uncertainty for VBF type of events is mainly because of low statistics.

5. **Jet energy scale, resolution and E_T^{miss}**

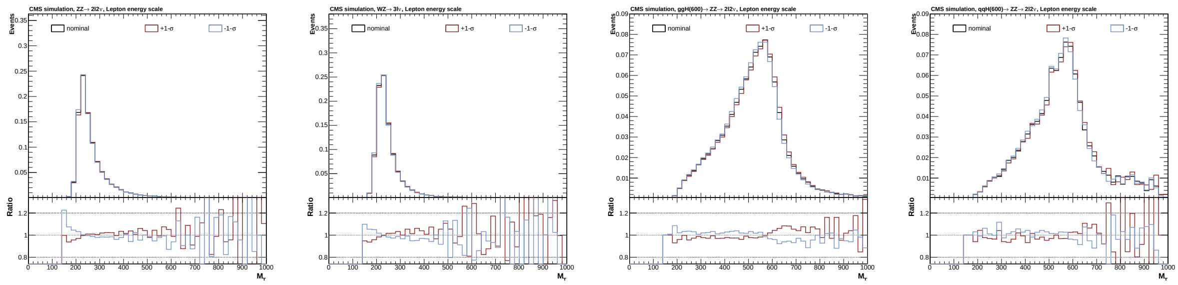


Figure 4.23: From *left* to *right*: systematic variations of the M_{T} distribution for ZZ, WZ, gluon-gluon fusion signal at M_H= 600 and VBF signal at M_H= 600 due to the variation of the lepton momentum scale. The bottom plots show the relative variation with respect to the nominal shape.

Process	Inclusive	Category		
		=0-jet	≥1-jets	VBF
ZZ	1.0	1.4	2.5	29.2
WZ	1.4	2.6	3.1	6.2
gg→ H	0.3	0.2	0.5	0.8
qq→ qqH	0.5	0.7	0.4	0.1

Table 4.11: Systematic uncertainty (in %) due to the uncertainty on the lepton momentum scale.

The uncertainty on the jet energy scale and resolution as well as of the unclustered E_T^{miss} scale affects this analysis in several distinct ways:

- the categorization of the events in the 0, ≥1 and VBF categories, including jet bin-to-bin migration of the events;
- the efficiency of the b-jet veto which is applied for jets reconstructed within the tracker acceptance region with p_{T} >30 GeV;
- the requirement on $\min \Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{miss})$ which is applied considering only the jets with p_{T} >30 GeV;
- the requirement on values of E_T^{miss} and M_{T}

To estimate the effects described above, we vary the nominal jet energy scale (or resolution) by one standard deviation [106, 107]. The effect due to the uncertainty on the energy scale of the unclustered component of the E_T^{miss} is estimated by subtracting the leptons and jets (with p_{T} >10 GeV) from the E_T^{miss} estimation and by varying the residual recoil by ±10%. The clustered component is added back in order to reassess the E_T^{miss} estimation.

Each one of the variations described above is used to recompute the E_T^{miss} and M_{T} estimation for each event and the jets are also checked for b-tagged if the new corrected p_{T} is greater than 30 GeV having $|\eta|$ <2.5. The category in which the event falls is also reassessed taking into account the recomputed p_{T}. We observe that the effect on events with real E_T^{miss} is small for each one of these instrumental uncertainties. Figure 4.24 represents the expected variations and these effects are further quantified in Table 4.12. The high uncertainty numbers for VBF category of events is because of low statistics.

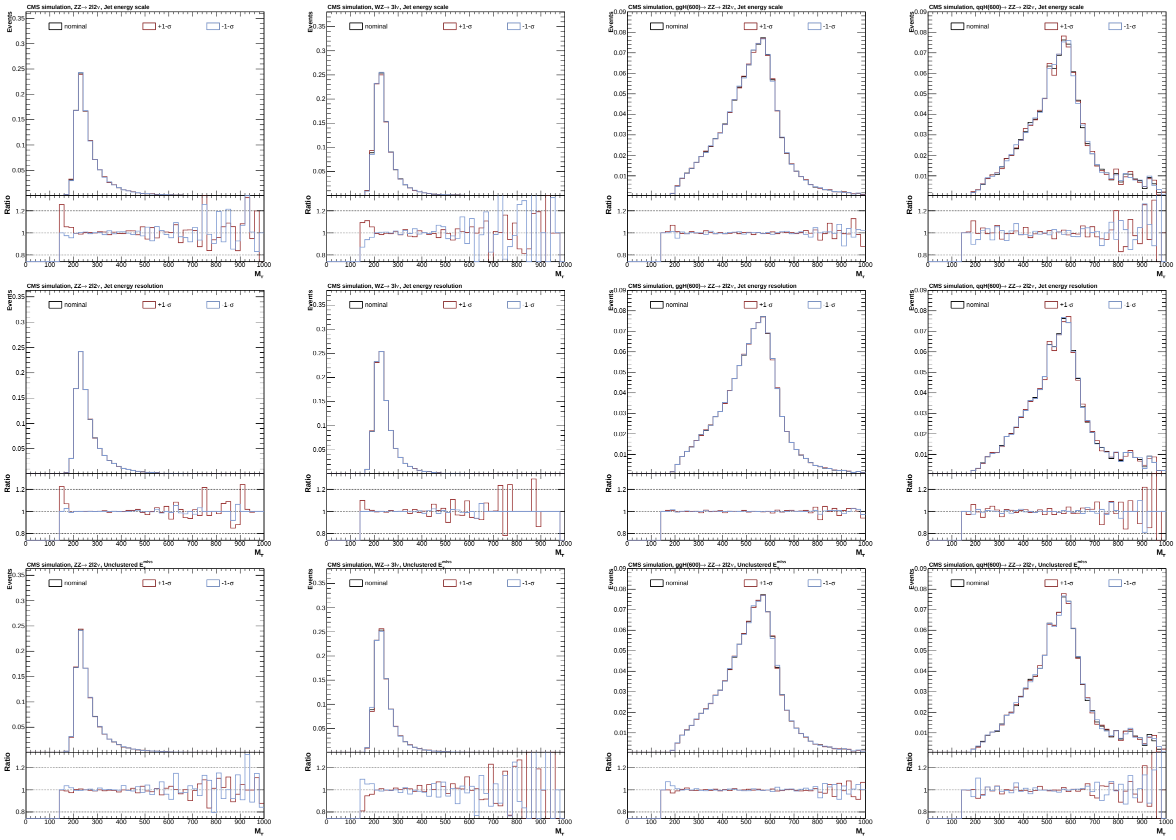


Figure 4.24: From *left* to *right*: systematic variations of the M_{T} distribution for ZZ, WZ, gluon-gluon fusion signal at M_H= 600 and VBF signal at M_H= 600 due to different systematic variations. *Top, center* and *bottom* rows shows the variations for jet energy scale, jet energy resolution and unclustered E_T^{miss}. For each distribution the bottom panels show the relative variation with respect to the nominal shape.

6. **b-quark jet veto**

Process	Inclusive			Category =0-jet			≥1 jets			VBF		
	JES	JER	uE_T^{miss}	JES	JER	uE_T^{miss}	JES	JER	uE_T^{miss}	JES	JER	uE_T^{miss}
ZZ	0.7	0.6	0.5	2.9	0.7	0.7	8.0	2.5	1.4	29.2	0.7	0.4
WZ	1.1	0.9	1.1	3.8	1.2	1.1	2.7	0.9	1.4	21.7	3.3	1.2
gg→ H	0.2	0	0.1	3.4	0.7	0.1	6.2	2.3	0.1	19.3	4.3	0.3
qq→ H	0.1	0.1	0.1	14.5	2.8	0.1	5.6	1.0	0.2	6.2	0.5	0.3

Table 4.12: Systematic uncertainty (in %) due to the uncertainty on the jet energy scale (JES), jet energy resolution (JER) and unclustured E_T^{miss} (uE_T^{miss}).

In order to propagate the uncertainty on the event yields from the variation of the b-tagging efficiency uncertainty, the following approach is taken using simulated events:

- the original quark from which the jet is generated is accessed for each event,
- we check if the jet is tagged or not according to the nominal Jet Probability discriminator requirement discussed in Section 4.3.5,
- if the jet is matched to a b-quark or a c-quark jet the probability of being tagged is downgraded by 2%,
- likewise, if the jet is not matched to a b-quark or c-quark the probability of being tagged is upgraded by 11%

The values are envelopes for the uncertainty on the b-tag and mistag rate[1] scale factors according to the values derived in Ref. [108]. We find the uncertainty on b-jet veto efficiency on MC driven processes - namely signal, WZ, ZZ to be <1% following this procedure. Figure 4.25 summarizes the result of the variation of the b-tag/mistag rate on the M_T shape. As expected, almost negligible effect is observed in the shape of the M_T.

7. **Soft muon veto**

The systematic uncertainty related to the soft muon veto is relevant for the MC based estimates such as ZZ, WZ and Higgs signal. For these processes we find

[1]mistag is the probability of tagging a non b-quark jet as b-quark jet

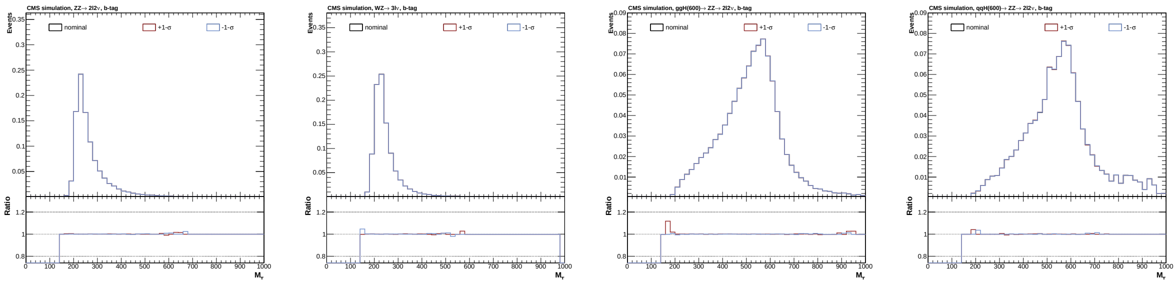


Figure 4.25: From *left* to *right*: systematic variations of the M_{T} distribution for ZZ, WZ, gluon-gluon fusion signal at M_H= 600 and VBF signal at M_H= 600 due to the variation of the b-tag uncertainty.

Process	Inclusive	Category		
		=0-jet	≥2-jets	VBF
ZZ	0.1	-	0.3	0.2
WZ	0.2	-	0.4	0
gg→ H	0.2	-	0.4	0.3
qq→ qqH	0.1	-	0.1	0.2

Table 4.13: Systematic uncertainty (in %) due to the uncertainty on the b-tag and mistag rates.

that the efficiency of the soft muon veto is high (between 96-99%). Given such high efficiency the systematic uncertainty is expected to be small and it is ignored.

8. **Top quark and WW estimate**

For the Top/WW backgrounds we assign a systematic uncertainty related to the data-driven method described in Section 4.5.1. The systematic uncertainty on the background estimate stems from the statistical uncertainty on the estimation of the α_ℓ parameters, which, in turn, is driven by the statistical uncertainty on the observed input events and the systematic uncertainty obtained from MC validation tests. A total uncertainty of 25% is assigned to α_ℓ.

9. $Z + jets$ **estimate**

To estimate the systematic uncertainty on the $Z + jets$ background estimate (see Section 4.5.2) we have studied two different approaches. We have subtracted the contribution of EW processes like $W + \gamma$, $Z + \gamma$, from the data driven estimate of $\gamma + jets$. The systematic uncertainty on this background was taken as a 100%

uncertainty on this subtracted amount due to limited MC statistics. The result was compared with the second method where we let the event yields to float in the computation of the limits between 0 and the direct prediction from the γ+jets method. The results are observed to be very similar in practice from these two methods. We assign 100% uncertainty on half of the γ+jets prediction.

4.7.2 Theoretical uncertainties

1. **MC statistics:**

Although not being strictly a "theoretical" uncertainty it reflects an intrinsic uncertainty in the modelling of processes taken from simulation. In the signal, WZ and ZZ the limited statistics of the MC is taken into account as a systematic uncertainty. Figure 4.26 illustrates the effect, which the statistical uncertainty is expected to have on the prediction of the M_{T} shape for the signal and the background.

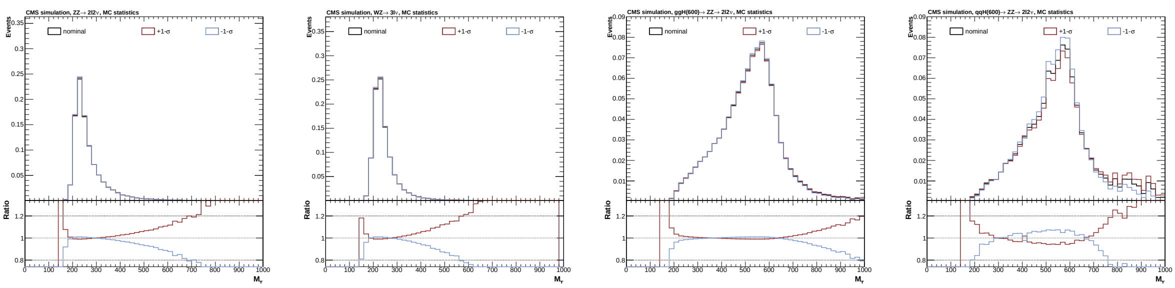


Figure 4.26: From *left* to *right*: systematic variations of the M_{T} distribution for ZZ, WZ, gluon-gluon fusion signal at M_H= 600 and VBF signal at M_H= 600 due to the limited statistics in simulation.

Table 4.14 summarizes the size of this uncertainty for each event category. Again, because of the lack of statistics in VBF category, the uncertainty is very high.

Process	Inclusive	Category =0-jet	=1-jet	VBF
gg→ H	2.9	3.7	5.7	37.7
qq→ qqH	5.2	13	8.3	9.7
ZZ	5.2	5.6	13.8	93
WZ	5.3	6.7	9.9	63

Table 4.14: Systematic uncertainty (in %) due to the limited size of the simulated samples.

2. **Higgs cross-section** We use the QCD scale and PDF uncertainties recommended by the LHC Higgs Cross Section Working Group [23] for the signal cross-sections.

3. **Higgs line shape** The Higgs line shape, which is re-weighted to match the CPS prediction and to take into account the interference with the ZZ and WW background, has a theoretical uncertainty which ranges typically between 5-15% as depicted in Figure 4.27. This uncertainty is taken into account depending on the M_T selection applied.

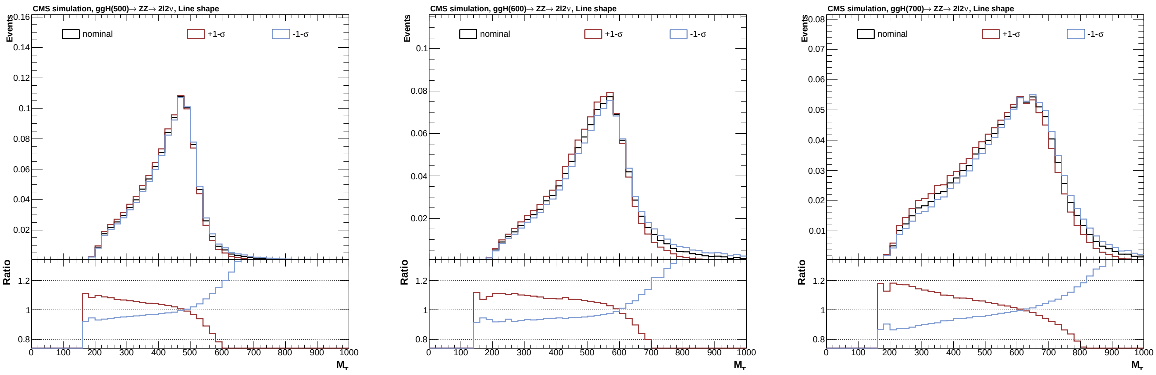


Figure 4.27: Systematic variations of the M_T distribution for gluon-gluon fusion signal at different generated masses taking into account the one standard deviation of the Higgs line shape reweighting. From *left* to *right* M_H= 500,600,700 GeV.

4. ***ZZ* and *WZ* cross-section** We use the following procedure to estimate the systematic uncertainty on the cross-section predictions for the ZZ and WZ processes. The cross-sections for $ZZ \to 2\ell 2\nu$ and $W^\pm Z \to 3\ell\nu$ are calculated using the program MCFM, version 6.2 [94]. MCFM runs using parton density functions (PDF) from the Les Houches accord PDF (LHAPDF) program. The following requirements are applied to mimic those used in the full analysis:

   - 76 GeV $< M_{\ell\ell} <$ 106 GeV,
   - p_T of the leptons $>$ 20 GeV,
   - $E_T^{miss} >$ 70 GeV,
   - $|\eta|$ of the leptons $<$ 2.5.

Calculations for each process was performed using three different PDFs: CT10, MSTW2008 and NNPDF. Additionally, the QCD renormalization and factorization scales ($\mu_R = \mu_F$) were both varied and chosen to be μ_R = 45 GeV, 90 GeV, 180 GeV for ZZ and μ_R = 45 GeV, 85 GeV, 170 GeV for WZ. For each scale, central value of cross-section and PDF uncertainty was obtained from an "envelope" of results when using different PDFs. This calculation was done using the PDF4LHC recommendations [109–113]. The renormalization and factorization scale uncertainties are estimated using the maximal relative difference between the central cross section value for different values of μ_R. The total uncertainty on the cross-section is given by $\Delta PDF + \Delta\alpha_S$, where, ΔPDF is the uncertainty from PDF and $\Delta\alpha_S$ is the uncertainty from changing the μ_R or μ_F. The results for $ZZ \to 2\ell 2\nu$ and $W^{\pm}Z \to 3\ell\nu$ at $\sqrt{s} = 8$ TeV are shown in Table 4.15.

Process	ΔPDF	$\Delta\alpha_S$	Systematic Uncertainty
$ZZ \to e^+e^- + 3(\nu\bar{\nu})$	3.12%	6.69%	9.8%
$W^+Z \to (\mu^+ + \nu) + (e^- + e^+)$	2.68%	7.67%	10.3%
$W^-Z \to (\mu^- + \nu) + (e^- + e^+)$	4.55%	6.90%	11.5%

Table 4.15: Theoretical uncertainties for ZZ and WZ production cross-sections at $\sqrt{s} =$ 8 TeV.

4.7.3 Summary of the systematic uncertainties

Table 4.16 summarizes the individual systematic uncertainties on the final event yields which are assigned per source.

4.8 Results

In this Section, results are discussed with the final control plots and background estimations. These results are then further used to perform a statistical analysis in order to search/exclude SM Higgs-like boson.

Source		Uncertainty [%]
Experimental	b-veto	1-3
	Lepton ID+Isolation	2
	Lepton momentum scale	1-2
	Jet energy scale/resolution	1
	Pile-up effects, E_T^{miss}	1-3
	Trigger	2
	Non-resonant background estimation from data	25
	Z+jets estimation from data	100
	Simulation statistics	1-2
Luminosity		2.2 (7 TeV), 4.4 (8 TeV)
Theory	PDF, gluon-gluon initial state	6-11
	PDF, quark-quark initial state	3.3-7.6
	QCD scale, gluon-gluon initial state (ggH)	7.6-11
	QCD scale, quark-quark initial state (VBF)	0.2-2
	QCD scale, gluon-gluon initial state (ggZZ)	20
	QCD scale, quark-quark initial state (qqVV)	5.8-8.5
	Line shape uncertainty	2-8
	Contribution from WW interference	2–13 (100% in BSM)

Table 4.16: Summary of systematic uncertainties on event yields of signal and background processes. Ranges refer to uncertainties that depend on the M_H hypothesis.

4.8.1 Final event yields

Figures 4.28 show the final inclusive distributions for min $\Delta\phi(\text{jet}, E_{\text{T}}^{miss})$ (after preselection and requirement on $E_T^{miss} > 70$ GeV) and $M(jj)$ (after preselection and a requirement on $|\Delta\eta_{\text{jj}}| > 4.0$). Figure 4.29 shows the inclusive E_T^{miss} and M_{T} distributions after preselection. An overall good agreement is observed between observed data and predicted SM background.

The expected number of signal and background events obtained after applying the full selection is summarized in Tables 4.17 and 4.18.

4.8.2 Statistical Analysis

Two analyses techniques are used to get the upper limits on the SM Higgs production cross-section. First is the baseline analysis or shape analysis in which full distribution for final discriminant is used without applying any selection on it. This is cross-checked with second method in which selection is applied on final discriminant (cut-and-count method).

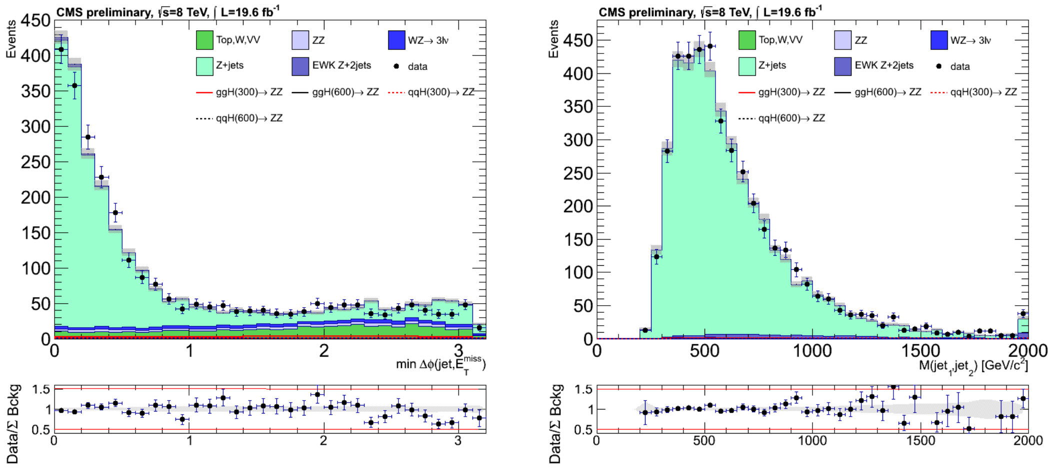


Figure 4.28: Inclusive distribution for the $\min \Delta\phi(\text{jet}, E_{\text{T}}^{miss})$ variable (*left*) and dijet invariant mass (*right*) for combined *ee* and $\mu\mu$ channels.

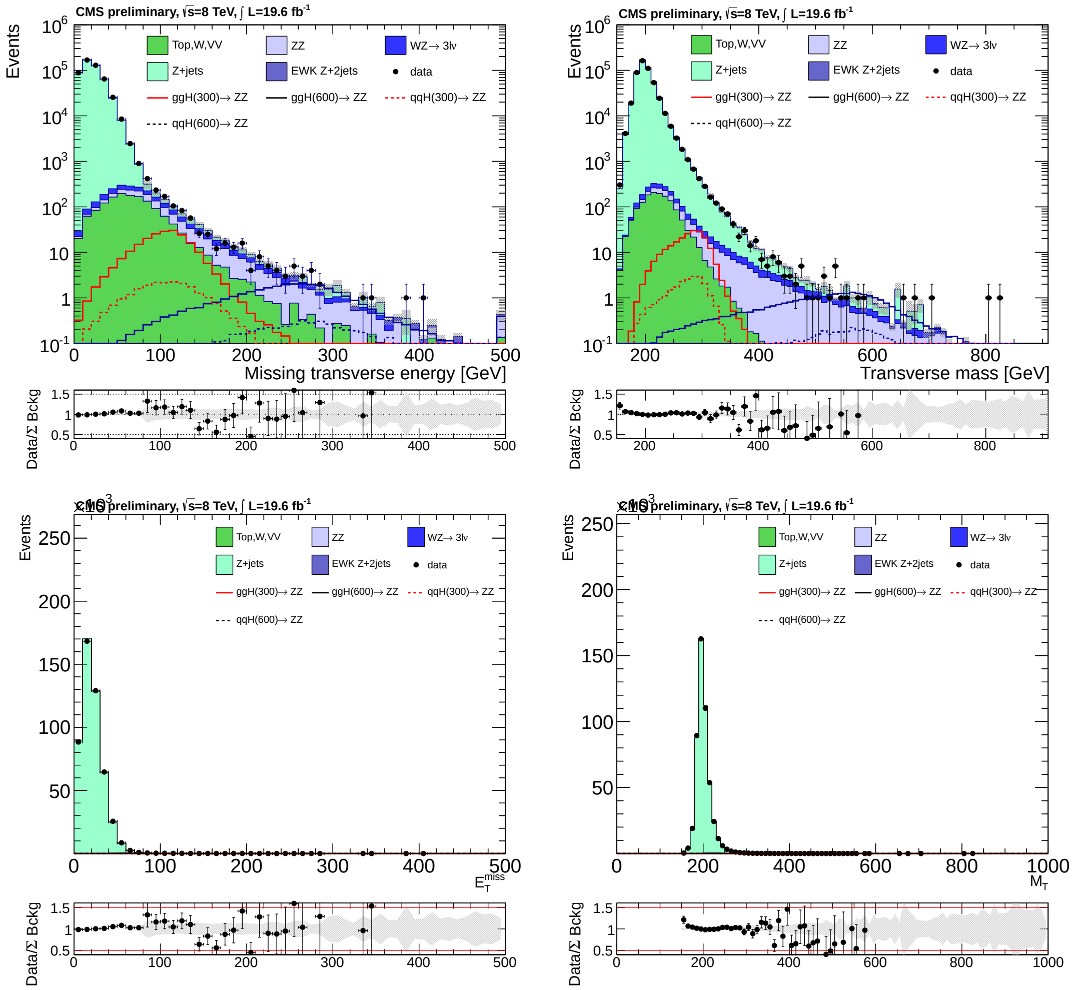


Figure 4.29: *Left:* E_T^{miss} variable. *Right:*M_{T} variable. The distributions are shown in logarithmic scale (*top*) and linear scale (*bottom*). These distributions are for combined *ee* and $\mu\mu$ channels.

Channel	ZZ	$WZ \to 3\ell\nu$	$Z \to \ell\ell$ (data)	$Top/WW/W$ (data)	$Total$	ggH	qqH	$Data$
M_H =200 GeV								
$ee,\ =0jets$	0.05 ± 0.02	0.05 ± 0.02	$0.01 \pm 0.01 \pm 0.01$	$0.4 \pm 0.4 \pm 0.1$	$0.5 \pm 0.4 \pm 0.1$	0.006 ± 0.004	$(0.004 \pm 0.004) \times 10^{-1}$	1
$\mu\mu,\ =0jets$	0.04 ± 0.02	0.17 ± 0.04	$0.02 \pm 0.02 \pm 0.02$	$0.7 \pm 0.7 \pm 0.2$	$0.9 \pm 0.7 \pm 0.2$	0.009 ± 0.005	0.004 ± 0.002	3
$ee,\ \geq 1jets$	0.34 ± 0.05	1.4 ± 0.1	$1.3 \pm 0.7 \pm 1.3$	$6.6 \pm 1.7 \pm 1.6$	$9.6 \pm 1.9 \pm 2.1$	0.59 ± 0.03	0.24 ± 0.01	28
$\mu\mu,\ \geq 1jets$	0.55 ± 0.06	2.2 ± 0.1	$2.2 \pm 1.2 \pm 2.2$	$10.6 \pm 2.8 \pm 2.7$	$15.6 \pm 3.0 \pm 3.5$	1.02 ± 0.04	0.35 ± 0.02	25
$ee,\ vbf$	0.007 ± 0.007	0.04 ± 0.02	$0.02 \pm 0.02 \pm 0.02$	$0.4 \pm 0.3 \pm 0.1$	$0.5 \pm 0.3 \pm 0.1$	0.014 ± 0.005	0.100 ± 0.009	1
$\mu\mu,\ vbf$	0.03 ± 0.01	0.09 ± 0.03	$0.03 \pm 0.03 \pm 0.03$	$0.7 \pm 0.5 \pm 0.2$	$0.9 \pm 0.5 \pm 0.2$	0.014 ± 0.005	0.14 ± 0.01	0
M_H =400 GeV								
$ee,\ =0jets$	1.7 ± 0.1	0.91 ± 0.09	$1.0 \pm 0.4 \pm 1.0$		$3.6 \pm 0.4 \pm 1.0$	4.00 ± 0.08	0.068 ± 0.004	5
$\mu\mu,\ =0jets$	2.4 ± 0.1	0.84 ± 0.08	$0.03 \pm 0.02 \pm 0.03$		$3.29 \pm 0.15 \pm 0.03$	5.35 ± 0.09	0.105 ± 0.005	3
$ee,\ \geq 1jets$	0.59 ± 0.06	0.84 ± 0.08	$0.3 \pm 0.1 \pm 0.3$		$1.8 \pm 0.2 \pm 0.3$	2.95 ± 0.06	0.36 ± 0.01	1
$\mu\mu,\ \geq 1jets$	0.72 ± 0.07	1.3 ± 0.1	$0.2 \pm 0.1 \pm 0.2$		$2.2 \pm 0.2 \pm 0.2$	4.34 ± 0.08	0.46 ± 0.01	4
$ee,\ vbf$	0.007 ± 0.007	0.04 ± 0.02	$0.02 \pm 0.02 \pm 0.02$	$0.4 \pm 0.3 \pm 0.1$	$0.5 \pm 0.3 \pm 0.1$	0.08 ± 0.01	0.240 ± 0.009	1
$\mu\mu,\ vbf$	0.03 ± 0.01	0.09 ± 0.03	$0.03 \pm 0.03 \pm 0.03$	$0.7 \pm 0.5 \pm 0.2$	$0.9 \pm 0.5 \pm 0.2$	0.12 ± 0.01	0.35 ± 0.01	0
M_H =600 GeV								
$ee,\ =0jets$	0.48 ± 0.05	0.16 ± 0.04	$0.1 \pm 0.1 \pm 0.1$		$0.8 \pm 0.1 \pm 0.1$	0.73 ± 0.02	0.034 ± 0.004	0
$\mu\mu,\ =0jets$	0.75 ± 0.07	0.17 ± 0.04	$0.0011 \pm 0.0008 \pm 0.0011$		$0.922 \pm 0.075 \pm 0.001$	0.93 ± 0.02	0.042 ± 0.003	1
$ee,\ \geq 1jets$	0.23 ± 0.04	0.17 ± 0.04	$0.19 \pm 0.07 \pm 0.19$		$0.60 \pm 0.09 \pm 0.19$	0.84 ± 0.02	0.152 ± 0.007	0
$\mu\mu,\ \geq 1jets$	0.35 ± 0.05	0.26 ± 0.05	$0.006 \pm 0.003 \pm 0.006$		$0.619 \pm 0.065 \pm 0.006$	1.08 ± 0.02	0.196 ± 0.008	2
$ee,\ vbf$	0.007 ± 0.007	0.04 ± 0.02	$0.02 \pm 0.02 \pm 0.02$	$0.4 \pm 0.3 \pm 0.1$	$0.5 \pm 0.3 \pm 0.1$	0.030 ± 0.003	0.099 ± 0.005	1
$\mu\mu,\ vbf$	0.03 ± 0.01	0.09 ± 0.03	$0.03 \pm 0.03 \pm 0.03$	$0.7 \pm 0.5 \pm 0.2$	$0.9 \pm 0.5 \pm 0.2$	0.034 ± 0.003	0.142 ± 0.007	0
M_H =800 GeV								
$ee,\ =0jets$	0.48 ± 0.05	0.16 ± 0.04	$0.1 \pm 0.1 \pm 0.1$		$0.8 \pm 0.1 \pm 0.1$	0.175 ± 0.003	0.017 ± 0.002	0
$\mu\mu,\ =0jets$	0.75 ± 0.07	0.17 ± 0.04	$0.0011 \pm 0.0008 \pm 0.0011$		$0.922 \pm 0.075 \pm 0.001$	0.241 ± 0.004	0.022 ± 0.002	1
$ee,\ \geq 1jets$	0.23 ± 0.04	0.17 ± 0.04	$0.19 \pm 0.07 \pm 0.19$		$0.60 \pm 0.09 \pm 0.19$	0.238 ± 0.004	0.064 ± 0.003	0
$\mu\mu,\ \geq 1jets$	0.35 ± 0.05	0.26 ± 0.05	$0.006 \pm 0.003 \pm 0.006$		$0.619 \pm 0.065 \pm 0.006$	0.340 ± 0.004	0.097 ± 0.004	2
$ee,\ vbf$	0.007 ± 0.007	0.04 ± 0.02	$0.02 \pm 0.02 \pm 0.02$	$0.4 \pm 0.3 \pm 0.1$	$0.5 \pm 0.3 \pm 0.1$	0.0072 ± 0.0006	0.039 ± 0.002	1
$\mu\mu,\ vbf$	0.03 ± 0.01	0.09 ± 0.03	$0.03 \pm 0.03 \pm 0.03$	$0.7 \pm 0.5 \pm 0.2$	$0.9 \pm 0.5 \pm 0.2$	0.0095 ± 0.0007	0.055 ± 0.003	0

Table 4.17: Expected number of signal and background events for an integrated luminosity of 5.0 fb^{-1} at $\sqrt{s} = 7$ TeV after applying the full Higgs selection compared to the number of observed events in data. Uncertainties for $ZZ/WZ/Z + jets$ and non-resonant backgrounds as well as for signal include the statistical and systematic components.

Channel	ZZ	$WZ \to 3\ell\nu$	$Z \to \ell\ell$ (*data*)	$Top/WW/W$ (*data*)	$Total$	ggH	qqH	$Data$
				M_H **=200 GeV**				
$ee,\ =0jets$	0.19 ± 0.04	0.37 ± 0.06	$1.0 \pm 0.7 \pm 1.0$	$4.8 \pm 1.5 \pm 1.2$	$6.4 \pm 1.6 \pm 1.6$	0.07 ± 0.02	0.012 ± 0.007	5
$\mu\mu,\ =0jets$	0.35 ± 0.05	0.65 ± 0.08	$1.5 \pm 0.9 \pm 1.5$	$7.8 \pm 2.4 \pm 2.0$	$10.3 \pm 2.6 \pm 2.4$	0.09 ± 0.02		14
$ee,\ \geq 1jets$	2.2 ± 0.1	6.6 ± 0.3	$16.3 \pm 8.4 \pm 16.3$	$33.0 \pm 4.1 \pm 8.2$	$58.2 \pm 9.3 \pm 18.3$	3.7 ± 0.1	1.17 ± 0.07	50
$\mu\mu,\ \geq 1jets$	3.6 ± 0.2	9.4 ± 0.3	$23.4 \pm 12.1 \pm 23.4$	$53.2 \pm 6.4 \pm 13.3$	$89.9 \pm 13.7 \pm 26.9$	4.9 ± 0.2	1.47 ± 0.07	96
$ee,\ vbf$	0.14 ± 0.03	0.29 ± 0.06	$0.6 \pm 0.3 \pm 0.6$	$1.8 \pm 0.4 \pm 0.4$	$2.7 \pm 0.5 \pm 0.7$	0.12 ± 0.02	0.52 ± 0.04	0
$\mu\mu,\ vbf$	0.17 ± 0.03	0.27 ± 0.05	$0.9 \pm 0.4 \pm 0.9$	$2.8 \pm 0.5 \pm 0.7$	$4.2 \pm 0.6 \pm 1.1$	0.12 ± 0.02	0.64 ± 0.05	5
				M_H **=400 GeV**				
$ee,\ =0jets$	10.1 ± 0.3	4.8 ± 0.2	$2.8 \pm 1.0 \pm 2.8$		$17.7 \pm 1.1 \pm 2.8$	21.2 ± 0.2	0.42 ± 0.03	16
$\mu\mu,\ =0jets$	13.4 ± 0.3	6.7 ± 0.3	$5.0 \pm 1.9 \pm 5.0$		$25.1 \pm 2.0 \pm 5.0$	27.5 ± 0.3	0.51 ± 0.03	21
$ee,\ \geq 1jets$	5.0 ± 0.2	4.1 ± 0.2	$2.9 \pm 1.2 \pm 2.9$	$1.3 \pm 0.8 \pm 0.3$	$13.3 \pm 1.4 \pm 2.9$	17.8 ± 0.2	1.89 ± 0.05	9
$\mu\mu,\ \geq 1jets$	6.8 ± 0.2	5.1 ± 0.2	$4.2 \pm 1.7 \pm 4.2$	$2.1 \pm 1.2 \pm 0.5$	$18.2 \pm 2.1 \pm 4.2$	22.7 ± 0.2	2.45 ± 0.06	14
$ee,\ vbf$	0.14 ± 0.03	0.29 ± 0.06	$0.6 \pm 0.3 \pm 0.6$	$1.8 \pm 0.4 \pm 0.4$	$2.7 \pm 0.5 \pm 0.7$	0.69 ± 0.04	1.45 ± 0.05	0
$\mu\mu,\ vbf$	0.17 ± 0.03	0.27 ± 0.05	$0.9 \pm 0.4 \pm 0.9$	$2.8 \pm 0.5 \pm 0.7$	$4.2 \pm 0.6 \pm 1.1$	0.84 ± 0.04	1.92 ± 0.05	5
				M_H **=600 GeV**				
$ee,\ =0jets$	2.6 ± 0.1	1.2 ± 0.1	$0.5 \pm 0.2 \pm 0.5$		$4.2 \pm 0.2 \pm 0.5$	4.29 ± 0.05	0.20 ± 0.02	2
$\mu\mu,\ =0jets$	3.3 ± 0.2	1.4 ± 0.1	$0.8 \pm 0.3 \pm 0.8$		$5.4 \pm 0.4 \pm 0.8$	5.43 ± 0.05	0.26 ± 0.02	7
$ee,\ \geq 1jets$	2.0 ± 0.1	1.1 ± 0.1	$0.8 \pm 0.4 \pm 0.8$		$3.9 \pm 0.4 \pm 0.8$	5.00 ± 0.05	0.95 ± 0.05	3
$\mu\mu,\ \geq 1jets$	2.4 ± 0.1	1.1 ± 0.1	$1.2 \pm 0.5 \pm 1.2$		$4.7 \pm 0.5 \pm 1.2$	6.37 ± 0.06	1.20 ± 0.05	6
$ee,\ vbf$	0.14 ± 0.03	0.29 ± 0.06	$0.6 \pm 0.3 \pm 0.6$	$1.8 \pm 0.4 \pm 0.4$	$2.7 \pm 0.5 \pm 0.7$	0.21 ± 0.01	0.67 ± 0.03	0
$\mu\mu,\ vbf$	0.17 ± 0.03	0.27 ± 0.05	$0.9 \pm 0.4 \pm 0.9$	$2.8 \pm 0.5 \pm 0.7$	$4.2 \pm 0.6 \pm 1.1$	0.27 ± 0.01	0.89 ± 0.03	5
				M_H **=800 GeV**				
$ee,\ =0jets$	2.6 ± 0.1	1.2 ± 0.1	$0.5 \pm 0.2 \pm 0.5$		$4.2 \pm 0.2 \pm 0.5$	1.29 ± 0.01	0.13 ± 0.01	2
$\mu\mu,\ =0jets$	3.3 ± 0.2	1.4 ± 0.1	$0.8 \pm 0.3 \pm 0.8$		$5.4 \pm 0.4 \pm 0.8$	1.63 ± 0.02	0.16 ± 0.01	7
$ee,\ \geq 1jets$	2.0 ± 0.1	1.1 ± 0.1	$0.8 \pm 0.4 \pm 0.8$		$3.9 \pm 0.4 \pm 0.8$	1.78 ± 0.02	0.51 ± 0.02	3
$\mu\mu,\ \geq 1jets$	2.4 ± 0.1	1.1 ± 0.1	$1.2 \pm 0.5 \pm 1.2$		$4.7 \pm 0.5 \pm 1.2$	2.24 ± 0.02	0.67 ± 0.02	6
$ee,\ vbf$	0.14 ± 0.03	0.29 ± 0.06	$0.6 \pm 0.3 \pm 0.6$	$1.8 \pm 0.4 \pm 0.4$	$2.7 \pm 0.5 \pm 0.7$	0.064 ± 0.003	0.34 ± 0.02	0
$\mu\mu,\ vbf$	0.17 ± 0.03	0.27 ± 0.05	$0.9 \pm 0.4 \pm 0.9$	$2.8 \pm 0.5 \pm 0.7$	$4.2 \pm 0.6 \pm 1.1$	0.077 ± 0.003	0.42 ± 0.02	5

Table 4.18: Expected number of signal and background events for an integrated luminosity of 19.7 fb^{-1} at $\sqrt{s} = 8$ TeV after applying the full Higgs selection compared to the number of observed events in data. Uncertainties for $ZZ/WZ/Z + jets$ and non-resonant backgrounds as well as for signal include the statistical and systematic components.

Figures 4.30 and 4.31 show the M_{T} and E_T^{miss} distributions used for the statistical analysis in each of the two categories. Using the $\mathrm{CL_s}$ [114–117] method, limits were set on the Higgs boson cross-section production. To fully define the method, one needs to make a choice of the test statistic, which is used to determine how signal or background-like the experimental observations are. It is based on the profile likelihood ratio. The likelihood $\mathcal{L}(data|\mu,\theta)$ used in constructing the test statistic is defined as:

$$\mathcal{L}(data|\mu,\theta) = Poisson(obs|\mu \cdot s(\theta) + b(\theta)) \cdot p(\bar{\theta}|\theta) \tag{4.8}$$

where s stands for the expected signal under the SM Higgs hypothesis, μ is a signal strength modifier and it scales the signal SM cross-section of all production channels by the same factor, i.e. $\sigma = \mu \cdot \sigma_{SM}$, b stands for the backgrounds, and θ are the nuisance parameters which describe the uncertainties. The probability density function $p(\bar{\theta}|\theta)$ is the probability of measuring a set of nuisance parameter $\bar{\theta}$, given its true value θ. And *obs* can be the actual *observation* or pseudo-data from MC simulated experiments. The explicit form of the poissonian term in Equation (4.8) for the product of probabilities of observing n_i events in the i^{th} interval is given by:

$$Poisson(obs|\mu \cdot s(\theta) + b(\theta)) = \prod_i \frac{(\mu \cdot s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} \cdot e^{\mu \cdot s_i(\theta) + b_i(\theta)} \tag{4.9}$$

Case of Discovery

The test statistic q_0 used to quantify the statistical significance of an excess of events over the background-only expectation is defined as:

$$q_0 = -2ln \frac{\mathcal{L}(obs|b(\hat{\theta_\mu}))}{\mathcal{L}(obs|\hat{\mu} \cdot s(\hat{\theta}) + b(\hat{\theta}))} \qquad (\hat{\mu} \geq 0) \tag{4.10}$$

Both the denominator and numerator are maximized. In the numerator, μ remains fixed and only the nuisance parameters θ are allowed to float. Their values at which $\mathcal{L}$ reaches the maximum are denoted as $\hat{\theta_\mu}$. In the denominator, both μ and θ are allowed to float in the fit, and $\hat{\mu}$ and $\hat{\theta}$ are parameters at which $\mathcal{L}$ reaches its maximum. The value of the test statistic for the actual observation is denoted as q_0^{obs}. The probability of getting a q_0 value as large as q_0^{obs} or larger in the background-only hypothesis is called p-value. The significance Z is obtained from a p-value via the Gaussian one-sided tail integral:

$$p_0 = P(q_0 > q_0^{obs}|b) = \int_Z^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \tag{4.11}$$

This local p-value is used to search for a signal excess in the background-only hypothesis across the whole explored M_H range. **Case of Exclusion**

As the $\mu = 0$ subscript indicates, q_0 is used to look for a signal under the hypothesis of pure background. On the contrary, if one wants to exclude the presence of a given signal one can assume that the signal strength is μ and define q_μ as:

$$q_\mu = -2ln \frac{\mathcal{L}(obs|\hat{\mu} \cdot s(\hat{\theta_\mu}) + b(\hat{\theta_\mu}))}{\mathcal{L}(obs|\hat{\mu} \cdot s(\hat{\theta}) + b(\hat{\theta}))} \qquad (\mu \geq \hat{\mu} \geq 0) \tag{4.12}$$

Similar to Equation (4.11), two p-values can be defined under the hypothesis of signal plus background $(s + b)$ and of background only (b):

$$CL_{s+b} = P(q_\mu \geq q_\mu^{obs}|\mu \cdot s + b) \tag{4.13}$$

$$CL_b = P(q_\mu \geq q_\mu^{obs}|b) \tag{4.14}$$

The ratio of these p-values is a quantity called CL_s. The following criterion is used for excluding the presence of a signal at the 1-α confidence level (CL).

$$CL_s(\mu) = \frac{CL_{s+b}}{CL_b} \leq \alpha \tag{4.15}$$

A signal with a cross-section $\sigma = \mu \cdot \sigma_{SM}$ is defined to be excluded at 95% CL if $CL_s(\mu) \leq 0.05$. Here, σ_{SM} stands for the SM Higgs boson cross-section.

Figure 4.32 shows the expected signal strength-μ as a function of the Higgs boson mass M_H in the range 200 – 1000 GeV, obtained using the shape analysis and cut-and-count analysis. Limits on signal strength μ are shown separately for data at $\sqrt{s}$ = 7 TeV, at $\sqrt{s}$ = 8 TeV and for the combination of data. Using the shape based analysis, we expect to exclude the SM Higgs boson in the M_H range [295,581] GeV at $\sqrt{s}$ = 7 TeV, [265,870] GeV at $\sqrt{s}$ = 8 TeV and [254,898] GeV for the combined dataset at 95% CL. And the observed exclusion of the SM Higgs boson at 95% CL is in the M_H range [265,458] GeV, [280, 886] GeV and [248, 930] GeV for the $\sqrt{s}$ = 7 TeV, $\sqrt{s}$ = 8 TeV and combined dataset respectively. We notice a good agreement between observation and expectation. Similarly the exclusion limits on M_H are calculated from cut-and-count analysis. In this analysis, we expect to exclude the SM Higgs boson in the M_H range [290,533] GeV at $\sqrt{s}$ = 7 TeV, [275,745] GeV at $\sqrt{s}$ = 8 TeV and [265,761] GeV for the combined dataset at 95% CL. The observed exclusion of the SM Higgs boson at 95% CL from cut-and-count analysis is in the M_H range [290,454] GeV and [543,554] GeV for $\sqrt{s}$ = 7 TeV, [276, 754] GeV for $\sqrt{s}$ = 8 TeV and [268, 756] GeV for combined dataset. Observed and expected limits are in agreement with each other.

All these exclusion limits are listed in the Tables 4.19 and 4.20.

Analysis	Median Expected Exclusion limit on M_H in GeV at 95% CL		
	$\sqrt{s}$ =7 TeV	$\sqrt{s}$ =8 TeV	Combined
Shape based	[295,581]	[265,870]	[254,898]
Cut-and-count	[290,533]	[275,745]	[265,761]

Table 4.19: Median Expected Exclusion limits on M_H for different data and both analyses at 95% CL.

Analysis	Median Observed Exclusion limit on M_H in GeV at 95% CL		
	$\sqrt{s}$ =7 TeV	$\sqrt{s}$ =8 TeV	Combined
Shape based	[265,458]	[280,886]	[248,930]
Cut-and-count	[290,454] & [543,554]	[276,754]	[268,756]

Table 4.20: Median Observed Exclusion limits on M_H for different data and both analyses at 95% CL.

It is interesting to note that the performance of the shape-based and of the cut-and-count analysis is not always the same in relative terms. The shape-based analysis improves the overall sensitivity in the following regions: the low mass region (M_H < 300 GeV), due to the VBF category where the background is significantly reduced and where the selection is less stringent for the signal; the high mass region ($M_H > 500$ GeV) due to the fast-falling M_{T} distribution for the remaining backgrounds and due to the broader signal shape which has a large width. In the intermediate region, both cut-and-count and shape-based analyses perform similarly given that the M_{T} regions used in the cut-and-count analysis capture efficiently the region where the signal is expected to peak.

It is also instructive to compare, with respect to the observed data, how much the different categories considered in this analysis are expected to contribute to the final limit. Figure 4.33 shows the expected and observed limits on the signal strength per event category. Events with zero jets are expected to drive the final result in the 300-400 GeV region where the limits of analysis are more stringent. Due to the tight selection used at low mass ($M_H < 300$ GeV) the limits are driven by the VBF and $\geq$1 jets-categories where either the background is lower or the Higgs is expected to be more boosted. In the high mass region ($M_H > 400$ GeV) the 0 jets- and $\geq$1 jets-categories contribute most significantly to the analysis given the higher cross section for gluon-gluon fusion and the fast falling background spectra in the M_{T} variable. There is an overall fair agreement between the expected and the observed limits in each category.

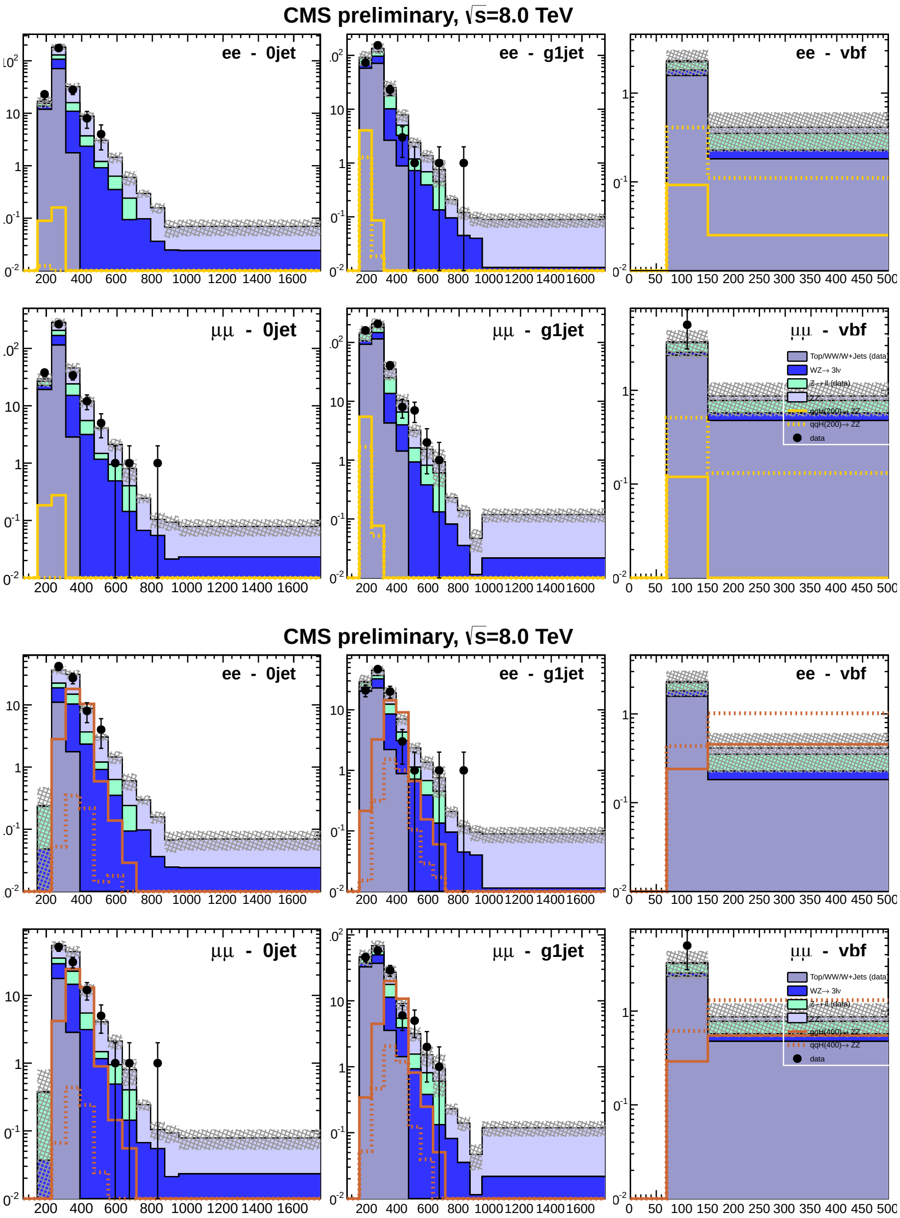


Figure 4.30: *Left* and *center*: M_{T} distribution in the 0 and at least 1-jet categories correspondingly. *Right*: E_T^{miss} distribution in the VBF category. The distributions correspond to 200 GeV (*top*) and 400 GeV (*bottom*) scenarios.

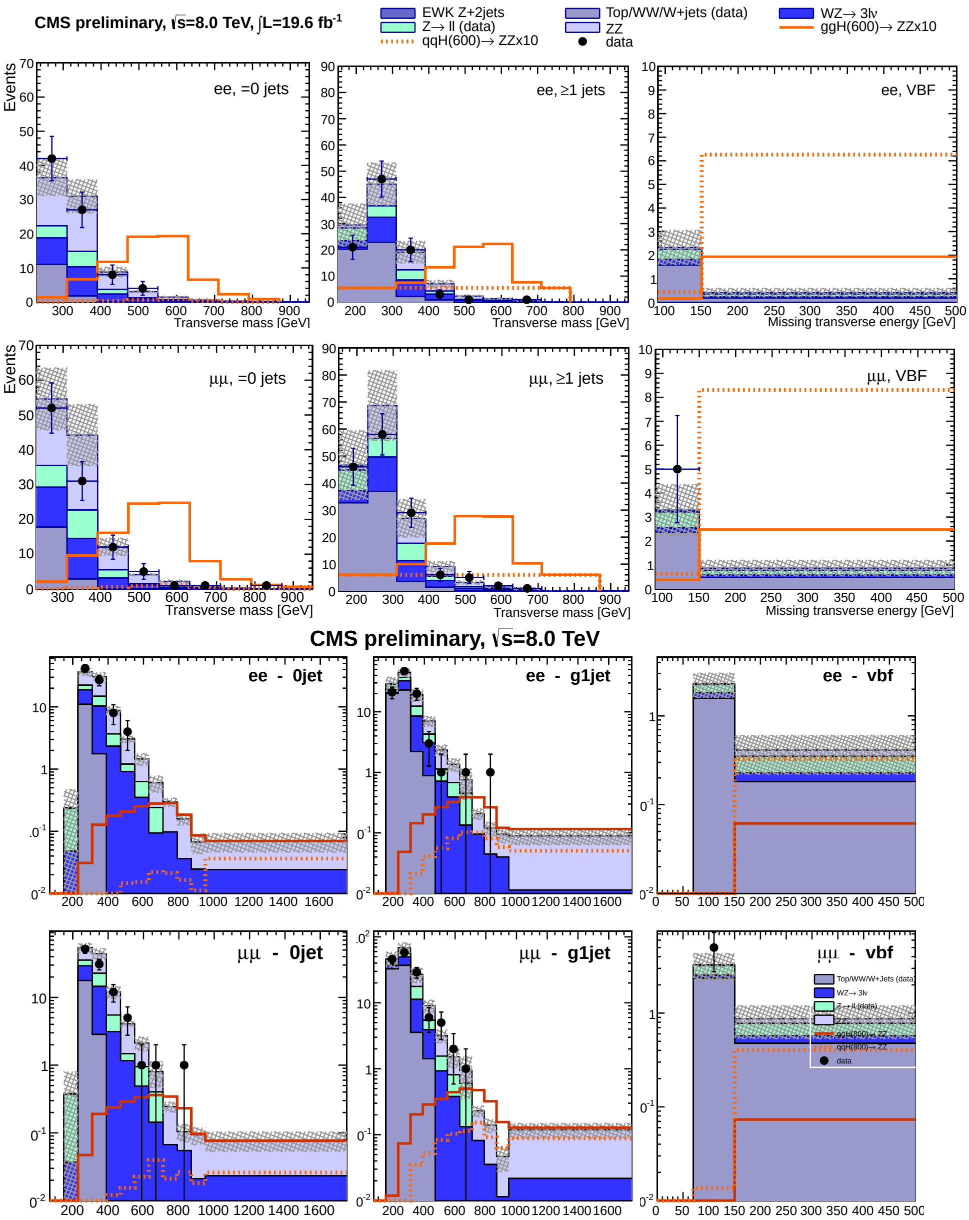


Figure 4.31: *Left* and *center*: M_{T} distribution in the 0 and at least 1-jet categories correspondingly. *Right*: E_T^{miss} distribution in the VBF category. The distributions correspond to 600 GeV (*top*) and 800 GeV (*bottom*) scenarios.

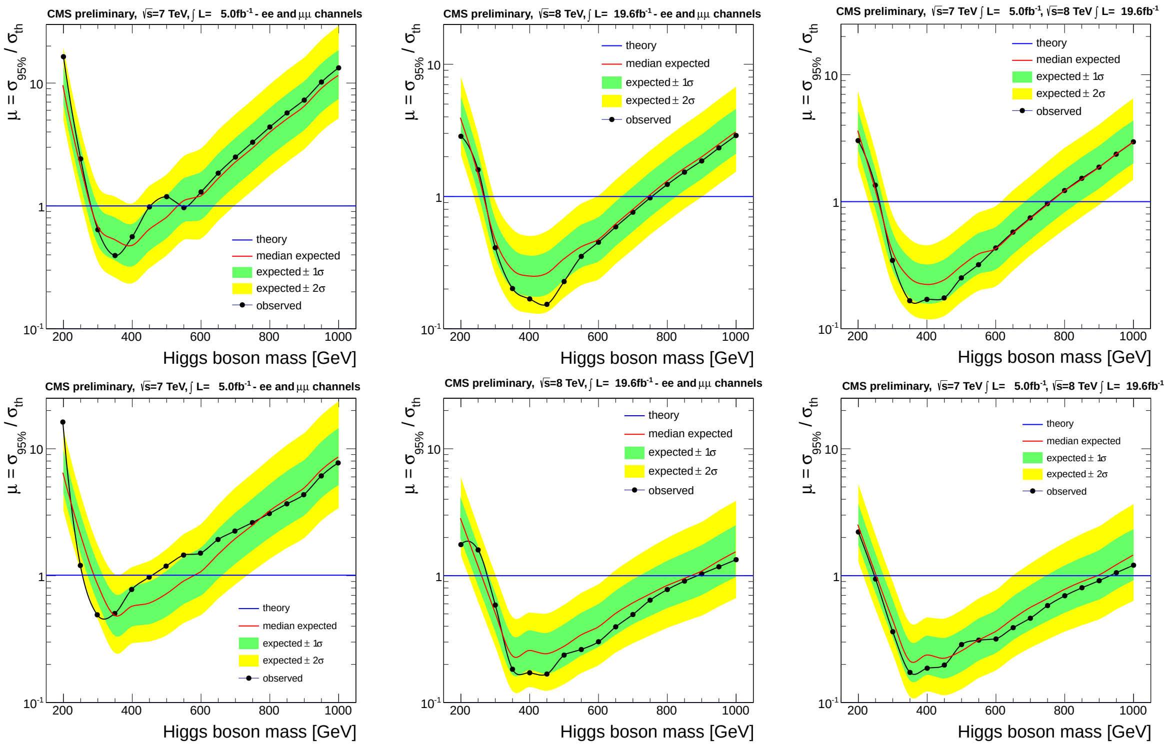


Figure 4.32: The expected 95% C.L. upper limit on $\mu = \sigma/\sigma_{SM}$ is shown as a function of the Higgs mass (M_H) for 5.0 fb^{-1} of integrated luminosity at $\sqrt{s} = 7$ TeV (*left*), for 19.7 fb^{-1} of integrated luminosity at $\sqrt{s} = 8$ TeV (*center*) and for 7+8 TeV datasets combined (*right*). Each plot combines the limits from the different categories considered in the analysis. Cut-and-Count analysis (*top*) and shape-based (*bottom*) limits are shown.

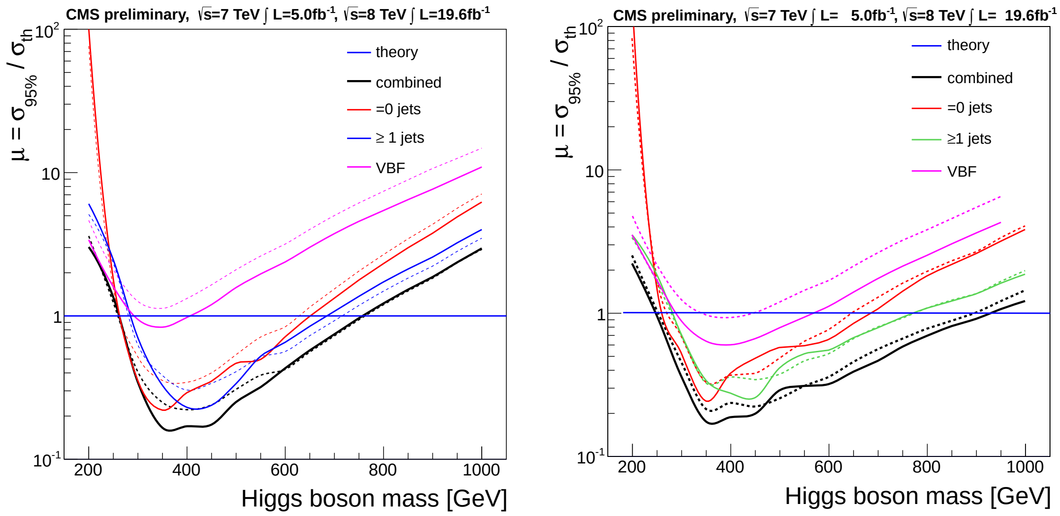
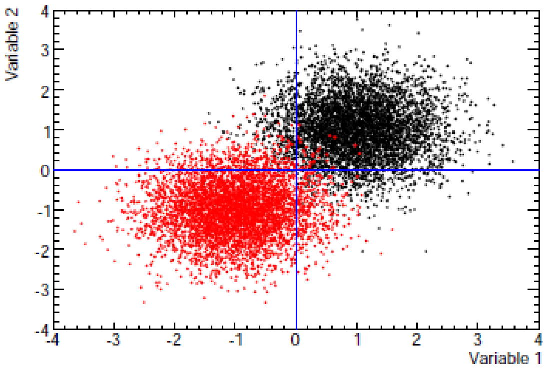


Figure 4.33: Evolution of the upper limit on the signal strength, μ, as a function of the Higgs boson mass for the different categories. The combination of both data periods is shown using the cut-and-count analysis (*left*) and shape-based analysis (*right*).

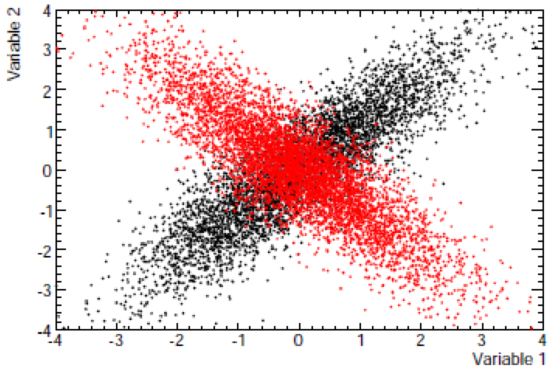
Chapter 5

$H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ Analysis using Multivariate Technique

The baseline $H \to ZZ \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ analysis was discussed in Chapter 4. However, to explore further improvements in sensitivity of this decay channel and to cross-check the analysis using an independent method, a Multivariate Analysis (MVA) was introduced and full analysis was carried out. Analyses based on sequential selection requirement on discriminating variables, like the one presented in Chapter 4, are straightforward and robust. However, they have their limitations owing to the possible correlations among discriminating variables. We consider here an artificial example to illustrate this. In Figure 5.1(a) it may be seen that the optimal rectangular cuts are quite easy to determine. However, in Figure 5.1(b) the variables are correlated differently for signal and background. Projecting both background and signal on the variable axes hides these correlations and hence they can not be exploited using the chosen variables with rectangular cuts. Although, in this artificial example it is rather simple to define two new variables accounting for the correlations, which can be achieved effectively rotating Figure 5.1(b) by 45°. Confronted with more realistic problems, this quickly becomes excessively complicated, as there is usually a rather high number of possible variables, resulting in many combinations to be tested for correlations. The situation is even worse if there are cor-



(a)



(b)

Figure 5.1: An example of a distribution of a signal (red) and background (black) with (*left*) and without (*right*) correlations. Trivial cuts are shown for the left plot.

relations between three or more variables. In addition all correlations would have to be re-investigated for different model parameters. For example, in data analyses like the one described in this Thesis, various discriminating variables are used to disentangle signal from background processes. And hence it is rather difficult to handle the correlations between these variables by applying rectangular cuts.

In this analysis we have used a multivariate method called Boosted Decision Tree (BDT). In Section 5.1 the concept of BDT along with various technical parameters is discussed. Section 5.2 describes in detail various steps of BDT analysis, including the selection BDT input trees, choice of BDT input variables and explanation of BDT performance distributions. Section 5.3 gives a brief overview of various systematics uncertainties. Results are discussed in Section 5.4.

5.1 Boosted Decision Tree

A decision tree (DT) uses the multiple rectangular cut technique in a sophisticated way in order to salvage the signal events which would otherwise be lost and remove background events that would normally pass. It is designed to maximize signal and background separation by recycling events that both pass and fail various requirements for further analysis. DT is a machine-learning technique which combines several weak classifiers [1]

[1]classifier is the variable which can discriminate between signal and background

to create a more powerful multivariate discriminant. In general, a DT is a structure of requirements (or cuts) organized into nodes. A node is a decision point in the tree in which a variable and cut value are provided and the event is determined to either pass or fail it. Figure 5.2 shows an exemplary cut based analysis as a DT. One can see an event being classified as signal, if and only if it passes all selection requirements. Here, some signal events are bound to fail the series of cuts and some background events will pass. The reason for that is by applying rectangular cuts, exactly one region is selected, whose borders are perpendicular to the variable axes. Ideally this region is located and sized such that the separation of signal and background is maximized. However, depending on the physics involved, the best region may be irregularly shaped or even separated in numerous unconnected regions. Thus, this results both in a loss in signal events and in an impure final state. To avoid that, the simple DT of Figure 5.2 can be extended to a more complex one like shown in Figure 5.3.

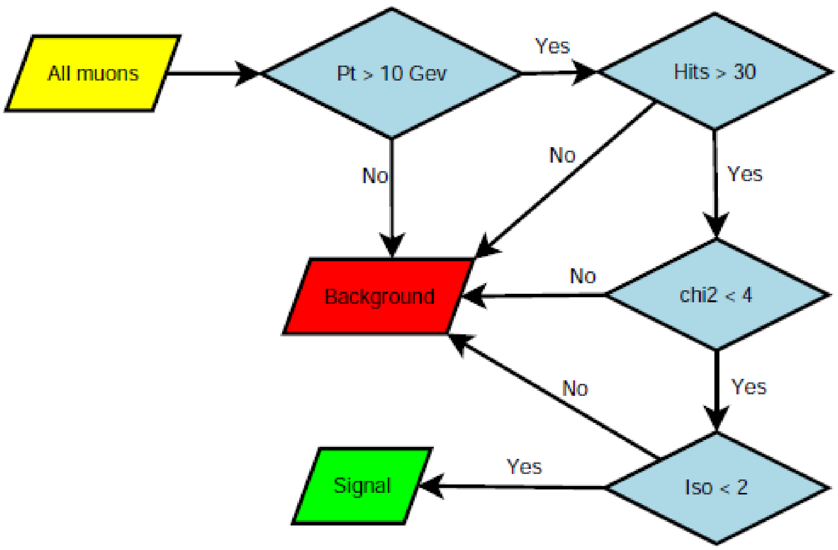


Figure 5.2: Conventional analysis based on one dimensional selection is shown as a simple decision tree: Events failing one cut are considered background.

In the tree shown in Figure 5.3, a series of cuts are applied starting from the parent (root) node. Then the root node branches off to two daughter nodes corresponding to the pass or fail of the root node cut. This division of events into daughter nodes goes on till it reaches at terminal node called 'leaf'.

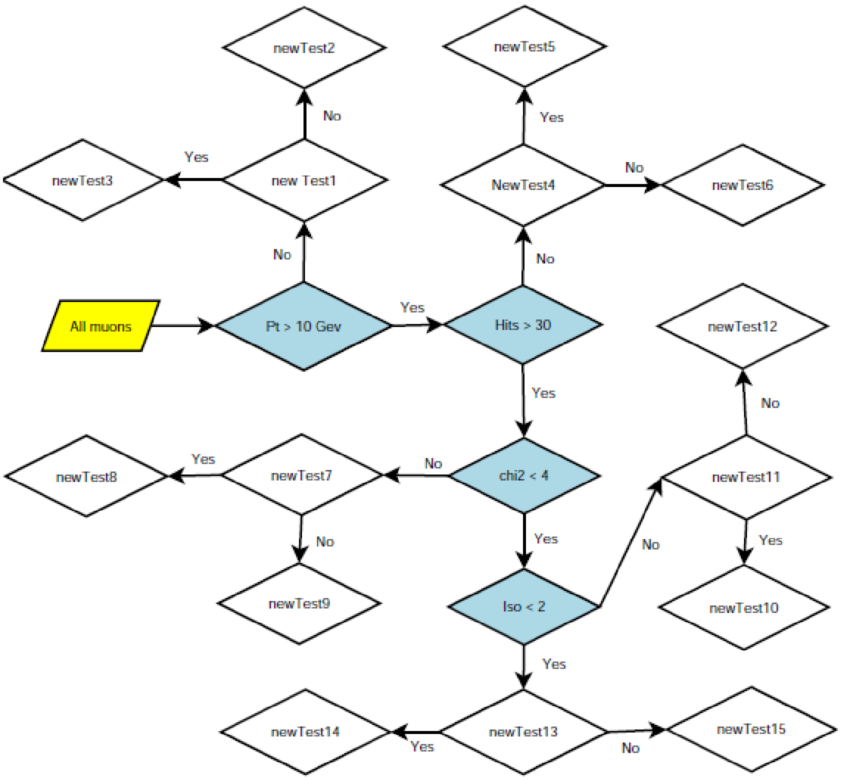


Figure 5.3: Extended decision tree: Leaves may be considered as signal or background.

The stopping criterion can be defined in advance, for example the number of events at the node should be above a threshold value. Each leaf would ideally contain only signal or only background, but realistically, we associate with each leaf a "purity", defined as the fraction of events in the leaf that are signal. The final discriminant depends on the weighted sum of fraction of these signal events at each leaf.

In this method, whole set of events are divided into two sets. One set is the "training set", which comprises the input variables and the desired output for each event and is used to optimize the parameters of the DT and the other set is the "testing set" which is used to test the performance of the DT. These two set of events must be exclusive to avoid any optimization bias. In this analysis, the events which are used in testing are also used for evaluation of limits. During 'training' of a DT, whole set of training set events are present at root node and then they are divided into daughter nodes. The function used in this division to quantify the separation between signal and background at any given node is called "Gini index", defined as:

$$G = p(1 - p) \tag{5.1}$$

where the purity 'p' is given as:

$$p = \frac{S}{S + B} \tag{5.2}$$

and $S(B)$ is the weighted total number of signal(background) events on the node. Thus, the Gini index for a node is 0(minimal) when the purity is either 1 or 0 (pure signal or pure background) and maximized when the purity is 0.5 (maximally mixed sample). Starting with the root node containing all the events, the variable and corresponding cut which maximizes separation is calculated and executed. The maximum separation is defined as the maximum change in the Gini index between the parent node and the two daughter nodes. That is, for each potential cut, the weighted change in Gini index

$$\Delta Gini = G_{parent} - f_1 \cdot G_1 - f_2 \cdot G_2 \tag{5.3}$$

is calculated, where 1(2) represents the first(second) daughter node and $f_{1(2)}$ is the weighted fraction of events in the daughter node. The cut corresponding to the highest $\Delta Gini$ is chosen. Following this initial branching, the best cuts according to the Gini separation are calculated and executed for both the first and second daughter nodes. The tree then grows as the cut selection process continues at each node. A node is no longer split when a split would result in a less than a minimum number of events (200, in this case) landing in the node. Such an unsplit node is called "leaf" as described earlier. Each leaf node is assigned a purity value p (as defined above). A test event works through the cut conditions beginning with the root node and follows the path along the tree according to pass or fail until it lands on a terminal node. The result of decision tree (or classifier value) $D(i)$ for an event tested on a single tree is equal to the purity of the leaf on which testing terminates. To gain information about the signal-likeness of an event and reduce the vulnerability to statistical fluctuations, not only one but several decision trees have to be used, whose outputs are weighted according to their overall performance. In this process of classifying events into branches and leaves containing only signal or background, some events may be misclassified. Thus, a technique called 'boosting' is introduced in which the misclassified events get a 'boost' or higher weight so that in the next decision tree they can take priority. Because of this technique the name of this method becomes "Boosted Decision Trees (BDT)". For this analysis, we have used a boosting technique called *AdaBoost* (Adaptive Boost). In this technique, the misclassified events are multiplied by a common boost weight α. The boost weight is derived from the misclassification rate, err, of the previous tree,

$$\alpha = \frac{1 - err}{err} \tag{5.4}$$

The weights of the entire event sample are then renormalized such that the sum of weights remains constant. This factor α is calculated for each tree and the result of final decision tree assigned to an event i is:

$$D(i) = \sum_{m=1}^{M_{tree}} \alpha_m D_m(i) \tag{5.5}$$

Where, m denotes the tree number. In the analysis, we train 1000 decision trees using *AdaBoost* to select the training sample for each tree. The output of the BDT is just the average output from all of the trees. We also realized that it was optimal to further randomize the trees by considering only a random subset of variables while maximizing $\Delta Gini$ at each node in each tree.

With any type of multivariate classifier, one must be cautious to minimize sensitivity to statistical fluctuations in the training sample, because the signal and background samples used for training are discrete and finite. The phenomenon in which a classifier becomes trained to respond to statistical fluctuations is known as "overtraining", and it generally leads to sub-optimal classifiers. The lower limit on the number of events in a node helps to avoid overtraining. To further reduce overtraining we also limit the depth of each tree to 4 nested decisions. To ensure the validity of results, it is of utmost importance that the BDT output should be well modeled. This, in turn, requires that every variable input to the classifier is well-modeled. The training of BDT is done separately for different jet multiplicity events, i.e. 0 jet, 1 jet and ≥ 2 jets, which results in an increase in sensitivity of around 20 %. BDTs are trained for each of thirteen Higgs boson mass point considered 200 - 1000 GeV in 50 GeV increments) and for each jet multiplicity. The SM background processes trained against are WW, WZ, ZZ, $t\bar{t}$, single top, and $Z + jets$. Both the signal and the backgrounds input to training are normalized to the SM expectations. In this analysis the ROOT [118] based multivariate analysis framework TMVA [119] is used. Similar to previous Chapter, all the distributions in this Chapter are for 19.7 fb^{-1}at $\sqrt{s}$ = 8 TeV. All corresponding plots for $\sqrt{s}$ = 7 TeV are in Appendix C. Results for $\sqrt{s}$ = 7 TeV analysis are shown in Section 5.4.

5.2 Analysis Workflow

In this Section, analysis strategy is described in various steps. The first step is the selection for BDT input trees and selection of various physics objects. After that, a description is given for input variables for BDT training and then various standard output distributions for BDT are shown.

5.2.1 Selection used for BDT input trees

All the physics object selection including the kinematic and identification etc. are same for leptons, jets and E_T^{miss} as used in the baseline analysis in Chapter 4. A brief summary of BDT input tree selection is given here:

1. one of the triggers summarized in Table A.5;
2. two leptons (medium-id electrons or tight-id muons) with p_{T} >20 GeV, isolated and with an invariant mass $|M_{\ell\ell} - 91|$ <15 GeV. The identification and isolation criteria of the leptons can be found in Section 4.3;
3. $p_{\mathrm{T}}^{\ell\ell}$ >55 GeV for the Z candidate;

4. no 3^{rd} lepton must be found in the event. The requirements used to identify a 3^{rd} lepton candidate are: loose-id electrons with I_{rel} <0.15 and p_{T} >10 GeV, or loose muon with I_{rel} <0.2, or soft muon with p_{T} >3 GeV.

5. no b-tagged jet must be found in the event;

6. $E_T^{miss} > 80$ GeV.

7. The nominal p_{T} threshold for the jets is required to be >30 GeV.

Instead of applying Higgs-mass dependent cut on E_T^{miss}, a flat cut of 80 GeV is applied. No selection criterion on M_{T} and $\min\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{miss})$ is applied. All the reweightings like pile-up and Higgs lineshape are same as described in Section 4.2.

5.2.2 BDT Input Variables

We mainly select the variables which were already used in the analysis Section 4.4.1. Events are categorized based on jet multiplicity while training. Thus, for events having atleast 2 jets some specific di-jet variables are used which were used as VBF tagging variables in baseline analysis Section 4.4.1. Here is the list of input variables used in the BDT:

1. $M_{\ell\ell}$: dilepton invariant mass;

2. p_{T} $(\ell\ell)$: p_{T} of the dilepton system;

3. E_T^{miss}: Transverse Missing Energy;

4. M_T : Transverse mass of Higgs;

5. $\min\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})$;

6. $\Delta\eta(\mathrm{jj})$: Pseudorapidity separation of two leading jets (used for atleast 2-jet category only);

7. $\Delta\phi(\mathrm{jj})$: ϕ separation of two leading jets (used for atleast 2-jet category only);

8. $M(jj)$: Invariant mass of two leading jets (used for atleast 2-jet category only).

The validation plots for collision data and the background prediction for all the input variables are shown in Appendix C in Section C.1 for $\sqrt{s} = 8$ TeV.

5.2.3 Background Estimation

In this analysis, all the background estimation techniques are same as used in baseline analysis described in Chapter 4. $Z+jets$ background is estimated using $\gamma+jets$ events by reweighting $\gamma+jets$ events as described in Section 4.5.2. Non-resonant background is also estimated by using $e-\mu$ events in data which passed full selection cuts. Control region is built using at least one b-tagged jet and then a factor α_ℓ is calculated as described in Section 4.5.1. Figure 5.4 show the E_T^{miss} distributions after data-driven estimation of $Z+jets$ and non-resonant background.

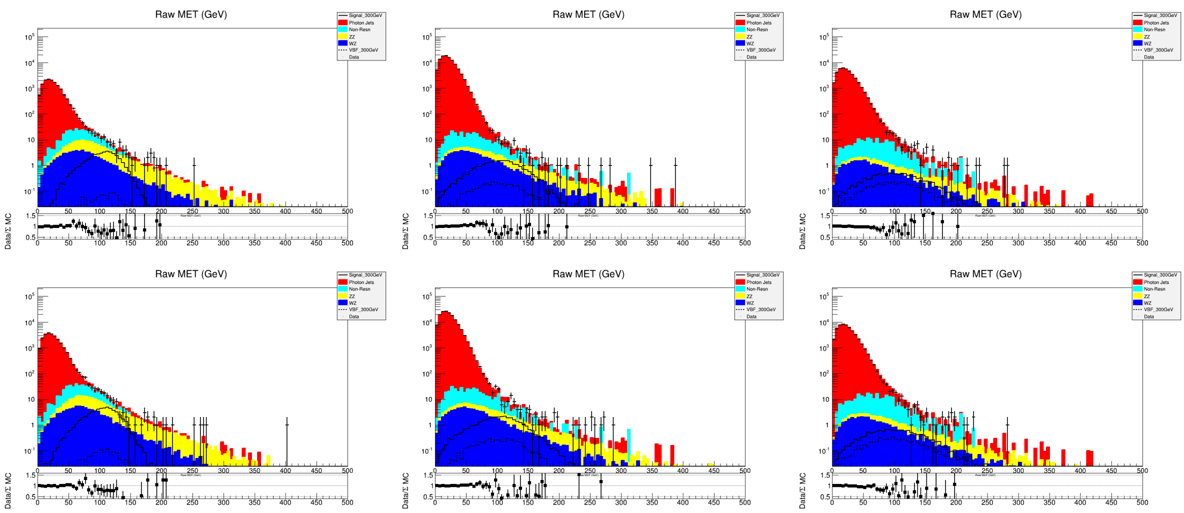


Figure 5.4: From *left* to *right*: E_T^{miss} (referred as RawMET in the plots) spectrum in di-electron (*top*) and di-muon (*bottom*) events with 0, 1 and ≥2 jets.

Figure 5.5 shows the BDT output plots for e-μ events in data and in MC. The agreement in these distributions shows the consistency of BDT output shapes in data and MC and thus, gives us confidence in estimating the non-resonant background.

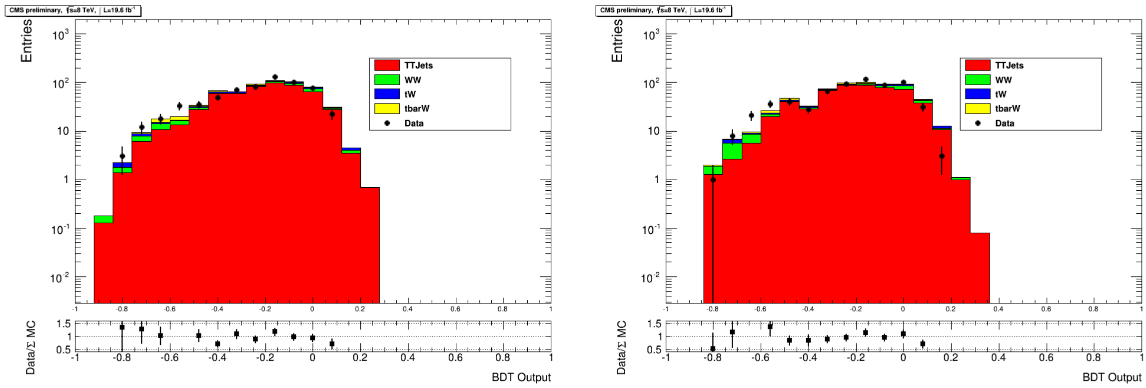


Figure 5.5: BDT Output distributions for di-electron channel (*left*) and di-muon channel (right) for $M_H = 200$ GeV.

5.2.4 BDT distributions

In this subsection various distributions are shown which are related to performance of the BDT.

Input Variables After Training

Figures 5.6 to 5.9 describe the distributions of all BDT input variables after training for two representative Higgs mass points, 200 GeV and 600 GeV, for ee and $\mu\mu$ channels. For all other mass points the plots are given is Appendix C in Section C.2. The event weights are renormalized such that, the sum of all weighted signal training events equals the sum of all weights of the background training events. It can be seen from Figure 5.6 and Figure 5.7 that VBF related variables like $\Delta\eta(\mathrm{jj})$, $\Delta\phi(\mathrm{jj})$ and $M(jj)$ along with $\min\Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})$ are provide the maximum discrimination at low mass (i.e. 200 GeV) whereas, dilepton kinematic variables and E_T^{miss} with M_T provide the discrimination in full mass range but most dominantly at higher mass. The trend is same for both ee and $\mu\mu$ channels.

Overtraining check distributions

Since we have divided the whole event set into two equal parts for training and testing of the BDT so, due to lack of statistics, the training of BDT may be prone to statistical fluctuations. This can be checked by checking the BDT output shapes for training and

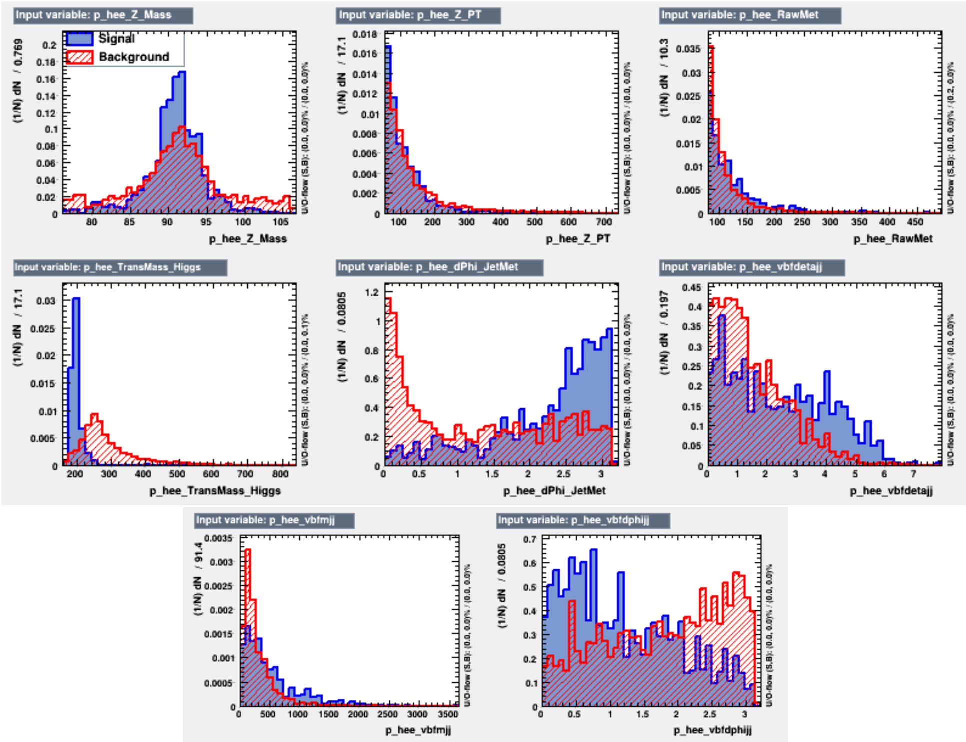


Figure 5.6: BDT Input variables distribution for di-electron channel for M_H= 200 GeV after training. Blue is the signal and Red is the combined background.

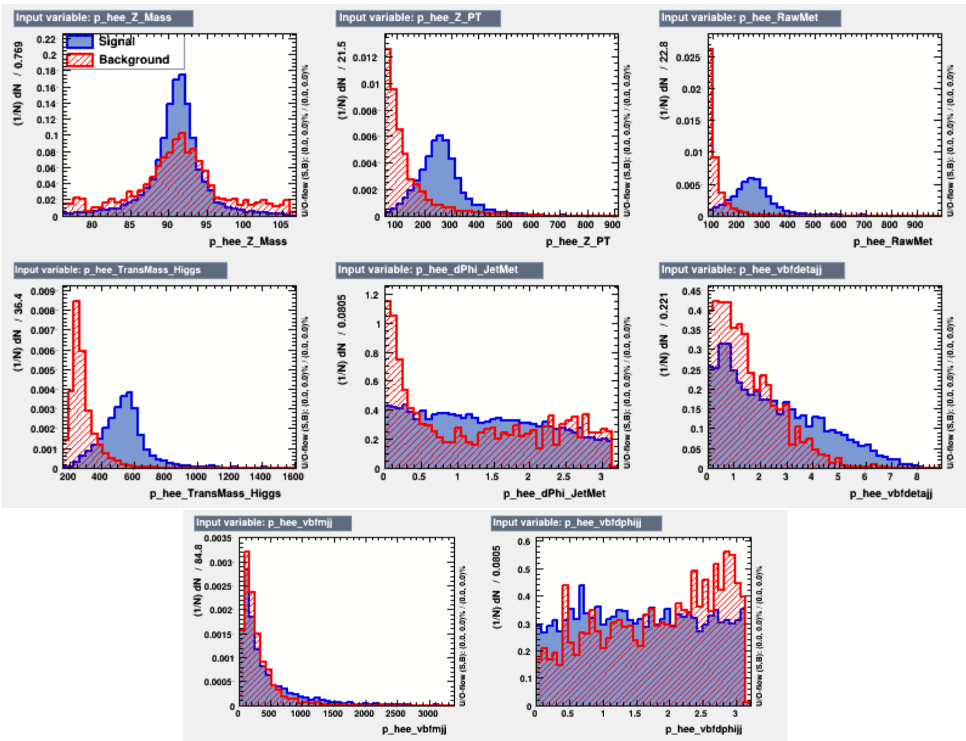


Figure 5.7: BDT Input variables distribution for di-electron channel for M_H= 600 GeV after training. Blue is the signal and Red is the combined background.

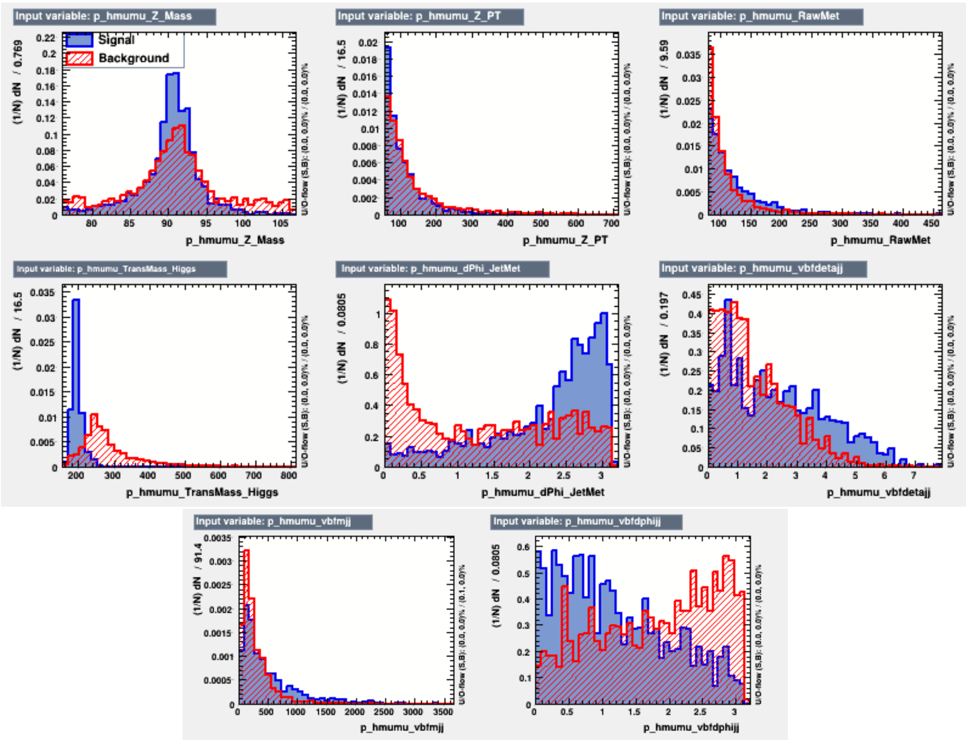


Figure 5.8: BDT Input variables distribution for di-muon channel for M_H= 200 GeV after training. Blue is the signal and Red is the combined background.

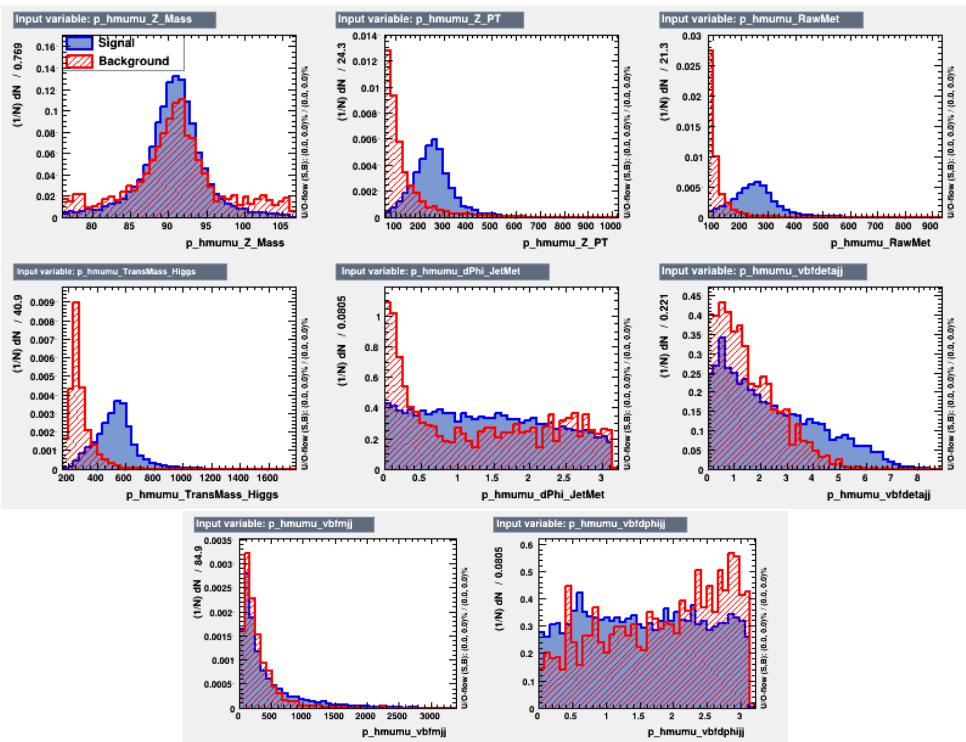


Figure 5.9: BDT Input variables distribution for di-muon channel for M_H= 600 GeV after training. Blue is the signal and Red is the combined background.

testing samples. This is what is shown in Figure 5.10 and Figure 5.11. These figures show the overtraining check distributions for different mass points for ee and $\mu\mu$ channels respectively. It can be seen from the distributions that training (dots) and testing (shaded portion) of BDT is matching within errors. Hence, the possibility of overtraining is taken care of in this analysis.

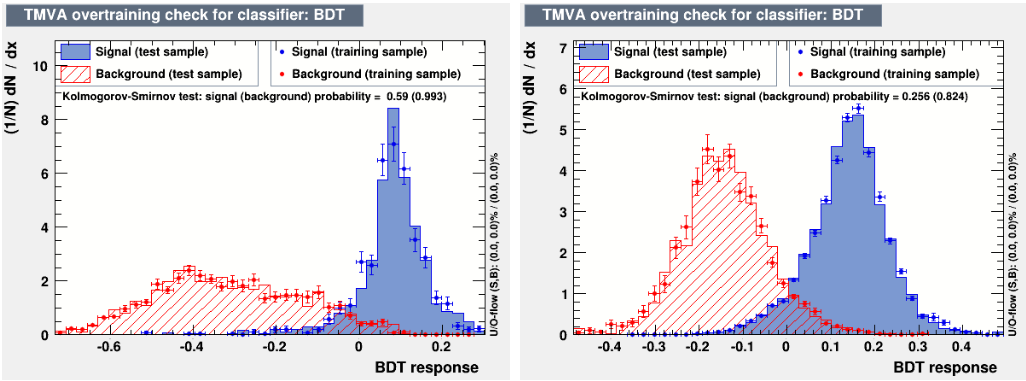


Figure 5.10: Overtraining Check plots for di-electron channel for M_H= 200 GeV ($left$) and M_H= 600 GeV ($right$). The shaded portion in the distributions show the testing sample and dots shows the training sample.

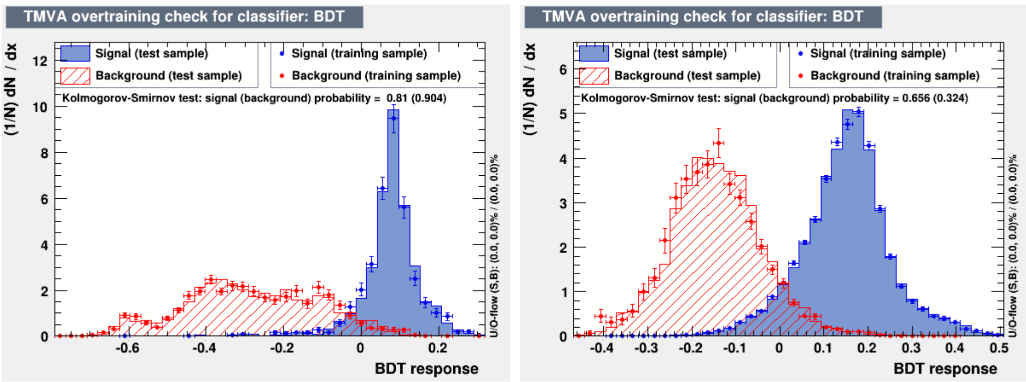


Figure 5.11: Overtraining Check plots for di-muon channel for M_H= 200 GeV ($left$) and M_H= 600 GeV ($right$). The shaded portion in the distributions show the testing sample and dots shows the training sample.

Correlation Plots

Figure 5.12 and Figure 5.13 show the correlation matrices for all the BDT input variables. The distributions are both for signal and backgrounds for two representative Higgs mass points, i.e. M_H = 200 GeV and M_H = 600 GeV.

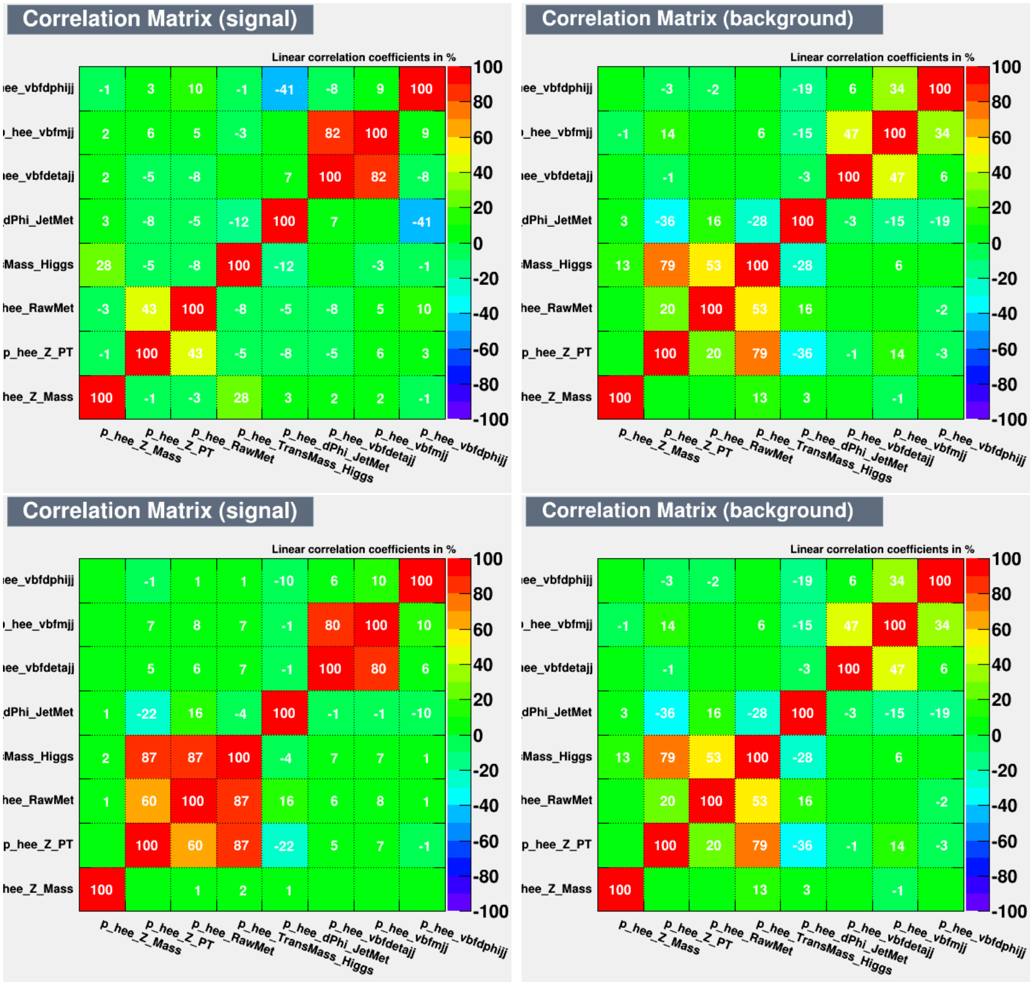


Figure 5.12: Plots showing the correlation matrices for di-electron channel for signal (*left*) and background (*right*) for M_H= 200 GeV (*top*) and M_H= 600 GeV(*bottom*).

Figure 5.12 shows that for M_H = 200 GeV, $M_{\rm T}$ (referred as $p\_hee\_Transmass\_Higgs$) and E_T^{miss} (referred as p_hee_RawMet) are correlated and also some jet related variables like $\Delta\eta_{jj}$ (referred as $p\_hee\_vbfdetajj$) and $\Delta\phi_{jj}$ (referred as $p\_hee\_vbfdphijj$) are correlated. And the trend is same in other distributions as well. BDT deals with these correlations and gives an optimal performance.

Receiver Operating Characteristic (ROC) Curves

ROC distributions in their standard form describe the true positive rate versus the false positive rate for the different possible cut points of a hypothesis test. Here, they describe the signal efficiency vs. background rejection for each BDT output value. Figure 5.14 and Figure 5.15 show the ROC curves for different Higgs mass points in ee and $\mu\mu$ channels

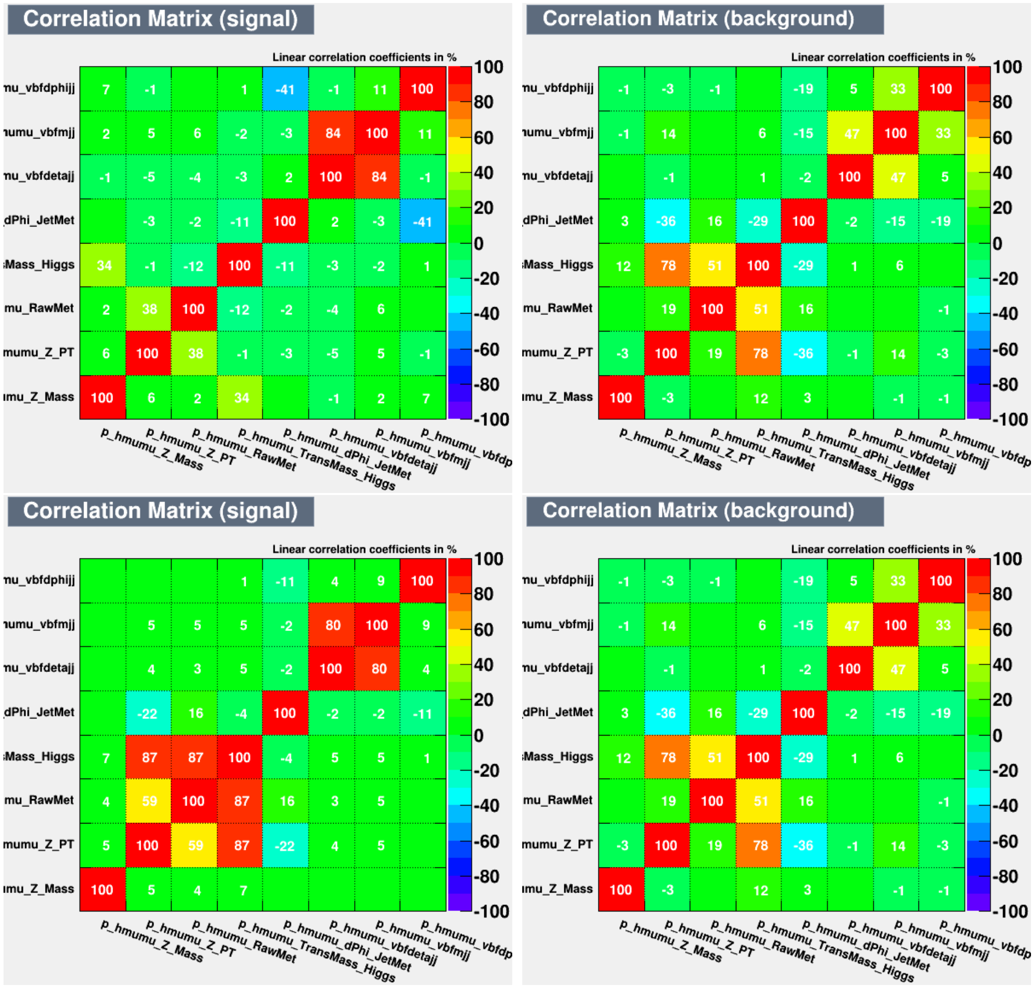


Figure 5.13: Plots showing the correlation matrices for di-muon channel for signal (*left*) and background (*right*) for M_H= 200 GeV (*top*) and M_H= 600 GeV(*bottom*).

respectively. At M_H= 200 GeV, we observe a very high background rejection even for larger signal efficiency but for M_H= 400 GeV, there is a decrease in performance. This is mainly because VBF related variables are no longer effective at 400 GeV (see Figure C.6 and Figure C.9 in Appendix C) and hence almost full discrimination is done by dilepton, E_T^{miss} and M_{T} variables. Thus, there is a kind of balancing going on at these masses resulting in relatively poor BDT performance. Whereas, when we go to higher masses like M_H= 600 GeV and M_H= 800 GeV then again we get high performance because here these variables are really distinctive as can be seen in Figure 5.7 for *ee* and in Figure 5.9 for $\mu\mu$.

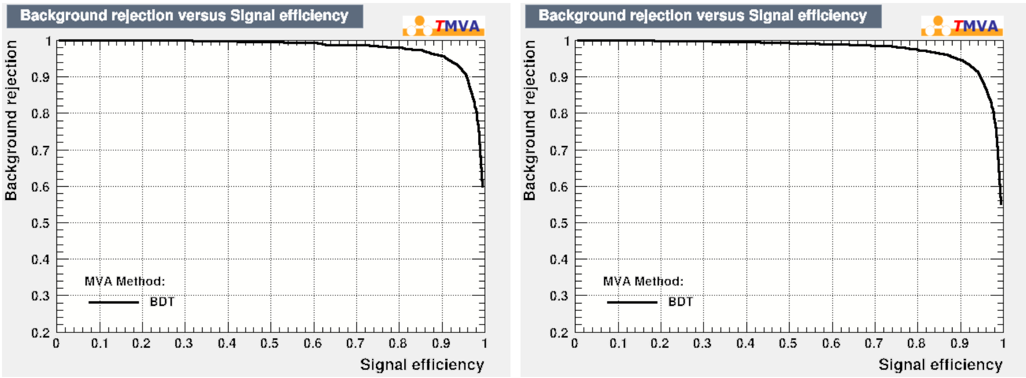


Figure 5.14: ROC curves for di-electron channel for M_H= 200 GeV (*left*) and M_H= 600 GeV (*right*).

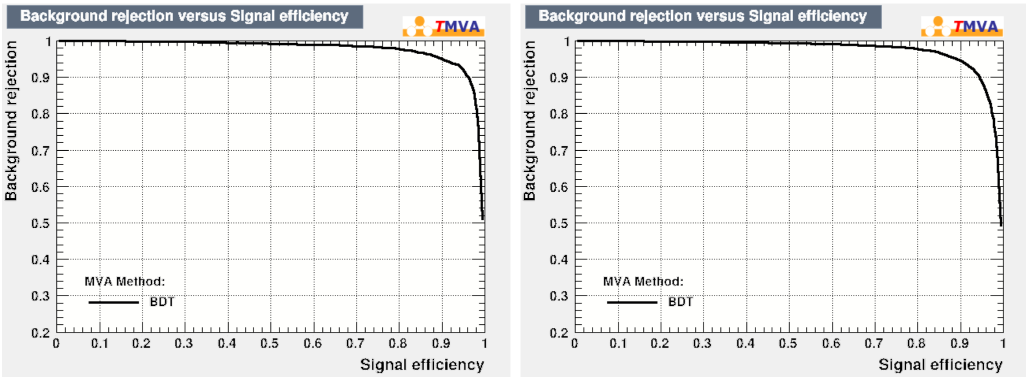


Figure 5.15: ROC curves for di-muon channel for M_H= 200 GeV (*left*) and M_H= 600 GeV (*right*).

5.3 Systematic Uncertainties

All sources and methods used for estimating systematic uncertainties are similar to that in Section 4.7.

- **Lepton momentum scale:** The uncertainty due to the lepton p_{T} is estimated by shifting the nominal energy by one standard deviation and by propagating this change to the measured E_T^{miss}. Since this change in lepton p_{T} is propagated to input variables for the BDT and hence BDT output shows variations. Figure 5.16 shows the variation in BDT shape by changing the lepton momentum scale. The change in the BDT shape is quite minimal.

- **Jet energy scale, resolution and E_T^{miss}:** Variation in jet energy scale, resolution and E_T^{miss} changes many input variables of BDT. And hence this change is visible in

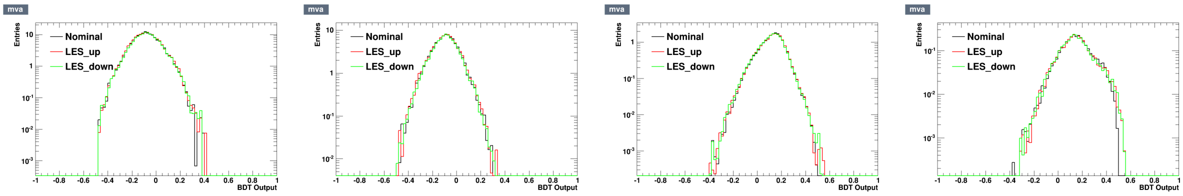


Figure 5.16: From *left* to *right*: systematic variations of the BDT Output for ZZ, WZ, gluon-gluon fusion signal at M_H= 600 GeV and VBF signal at M_H= 600 GeV due to the variation of the lepton energy scale.

BDT output plots. Figure 5.17 represents the expected variations in BDT output shape by change in jet energy scale, jet energy resolution and unclustered E_T^{miss} scale.

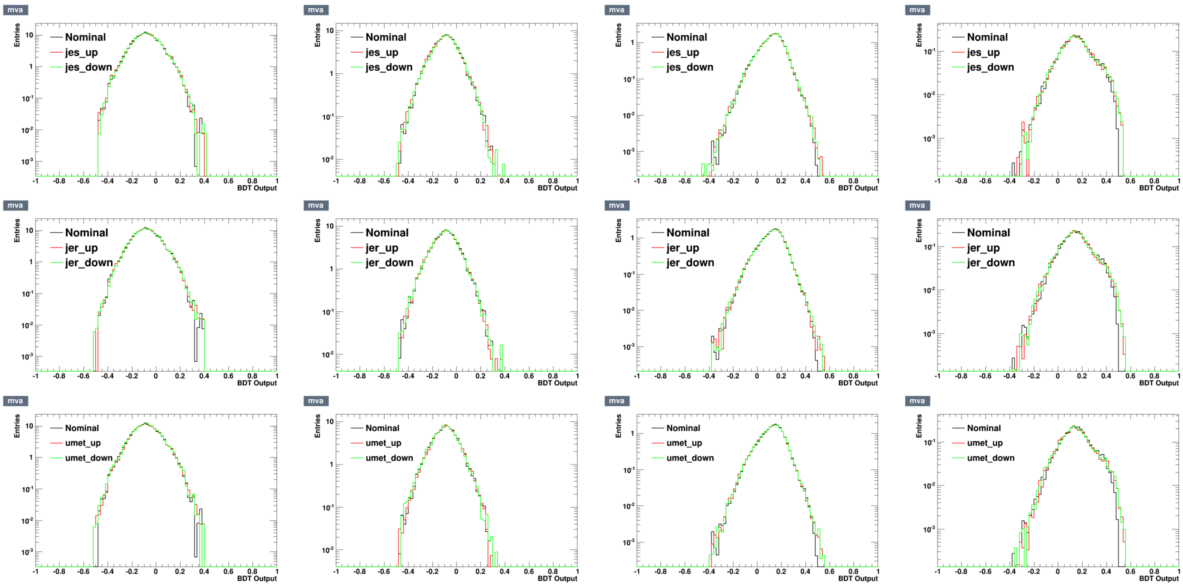


Figure 5.17: From *left* to *right*: systematic variations of the BDT Output distribution for ZZ, WZ, gluon-gluon fusion signal at M_H= 600 GeV and VBF signal at M_H= 600 GeV due to different systematic variations. *Top, center* and *bottom* rows shows the variations for jet energy scale, jet energy resolution and unclustered E_T^{miss}.

- ***b*-quark jet veto:** The effect of changing the threshold of b-tag discriminator by one standard deviation is shown on the BDT output shape in Figure 5.18. The effect is very small.

- **Higgs line shape:** Changing the Higgs line shape from the nominal by one standard deviation changes the BDT output shape as shown in Figure 5.19.

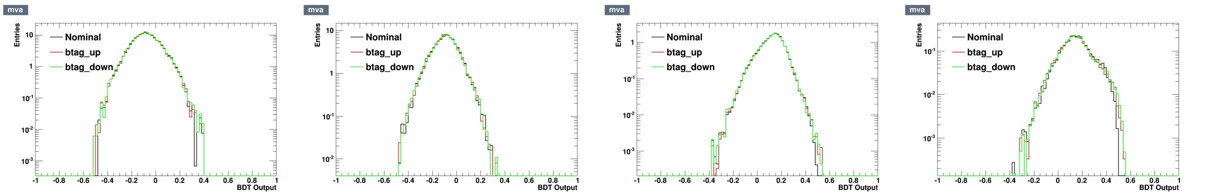


Figure 5.18: From *left* to *right*: systematic variations of the BDT Output distribution for ZZ, WZ, gluon-gluon fusion signal at M_H= 600 GeV and VBF signal at M_H= 600 GeV due to the variation of the b-tag uncertainty.

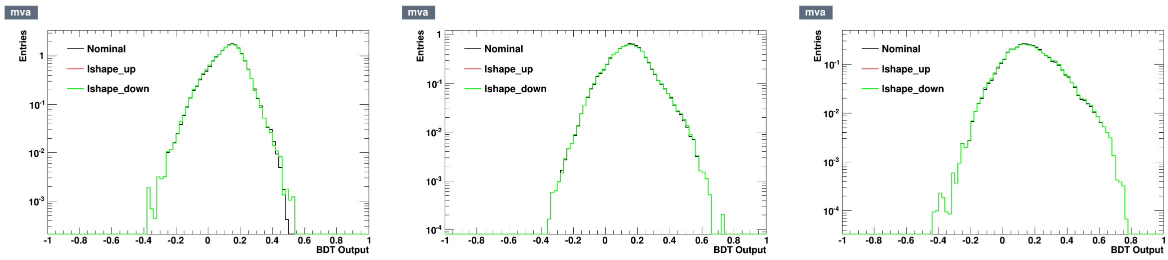


Figure 5.19: Systematic variations of the BDT Output distribution for gluon-gluon fusion signal at different generated masses taking into account one standard deviation of the Higgs line shape reweighting. From (*left* to *right*) M_H = 600, 700, 800 GeV.

5.3.1 Summary of the Systematic Uncertainties

Table 5.1 summarizes the individual systematic uncertainties on the final event yields which are assigned per source.

5.4 Results

Final BDT output plots are shown in Figure 5.20 and Figure 5.21 for two representative Higgs mass points M_H = 200 GeV and 600 GeV for an integrated luminosity of 19.7 fb^{-1}at $\sqrt{s} = 8$ TeV. Corresponding plots for $\sqrt{s} = 7$ TeV data are shown in Appendix C, in Section C.3. These distributions are further used to perform a statistical analysis for setting upper limits on the SM Higgs boson production cross-section. From these distributions it can be observed that for M_H= 200 GeV, $Z + jets$ background is still present in high BDT output value region (where we expect mostly signal). Whereas, for M_H = 600 GeV, di-boson backgrounds ZZ and WZ start appearing in signal region and hence become the dominant backgrounds.

Source		Uncertainty [%]
Experimental	b-veto	1-4
	Lepton ID+Isolation	2
	Lepton momentum scale	1-3
	Jet energy scale	1-4
	PU effects, E_T^{miss}	4-8
	Trigger	2
	Non-resonant background estimation from data	25
	Z+jets estimation from data	100
Luminosity		4.4
Theory	PDF, gluon-gluon initial state	6-11
	PDF, quark-quark initial state	3.3-7.6
	QCD scale, gluon-gluon initial state (ggH)	7.6-11
	QCD scale, quark-quark initial state (VBF)	0.2-2
	QCD scale, gluon-gluon initial state (ggZZ)	20
	QCD scale, quark-quark initial state (qqVV)	5.8-8.5
	Contribution from WW interference	2-13
	Line shape uncertainty	2-8

Table 5.1: Summary of all systematic uncertainties on final event yields.

5.4.1 Statistical Analysis for the SM Higgs boson

No excess is observed over the SM background prediction over entire Higgs mass range scanned. A statistical method, same as described in Section 5.4.1, is used to set upper limits on the Higgs boson production cross-section. A shape analysis is done on final BDT output shape, i.e. full BDT distribution has been used to impose the limits. Figure 5.22 shows the expected and observed signal strength-μ at 95% CL as a function of M_H in the range 200 – 1000 GeV, obtained using the shape-based BDT analysis for $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV data. Using the $\sqrt{s} = 8$ TeV data and BDT shape-based analysis, the SM Higgs boson is expected to be excluded at 95% CL in the mass range [245,781] GeV and [837,910] GeV; and [274,507] GeV for $\sqrt{s} = 7$ TeV data. The observed limit of exclusion is [260,954] GeV at $\sqrt{s} = 8$ TeV and [290,498] GeV for $\sqrt{s} = 7$ TeV.

Figure 5.23 shows the combined limit from $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV data. The SM Higgs boson is expected to be excluded in mass range [244,913] GeV and observed exclusion is in the mass range [248,963] GeV at 95% CL. The exclusion limit obtained by BDT shape analysis is very much compatible with the baseline analysis as described in

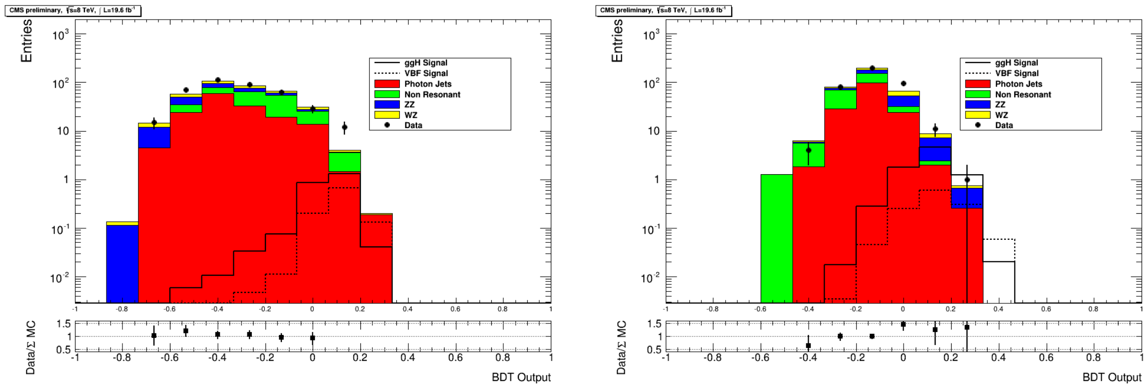


Figure 5.20: Final BDT Output distributions for di-electron channel for $M_H = 200$ GeV (*left*) and $M_H = 600$ GeV (*right*) Higgs mass points. Black dots represent data points and solid black line shows the gluon-gluon fusion signal and dotted black line shows VBF signal.

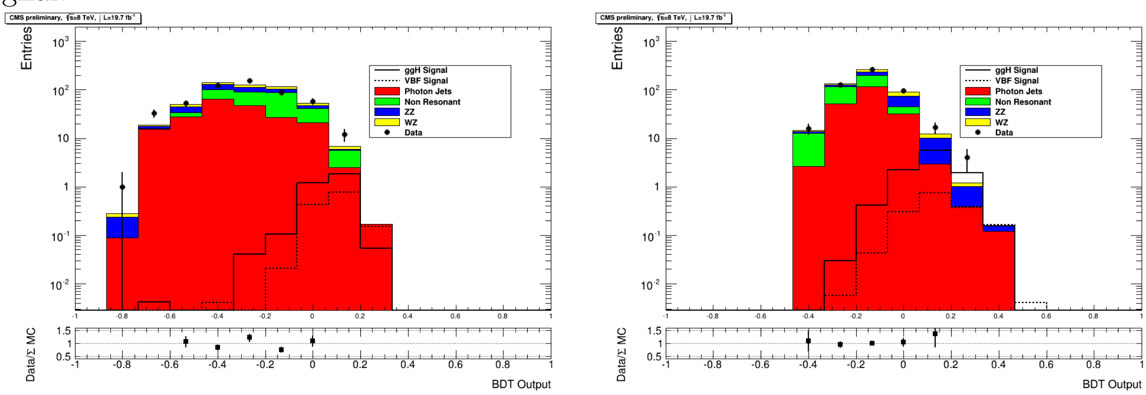


Figure 5.21: Final BDT Output distributions for di-muon channel for $M_H = 200$ GeV (*left*) and $M_H = 600$ GeV (*right*) Higgs mass points. Black dots represent data points and solid black line shows the gluon-gluon fusion signal and dotted black line shows VBF signal.

Table 4.19 and 4.20. However, there is slight improvement in the exclusion limit towards high M_H.

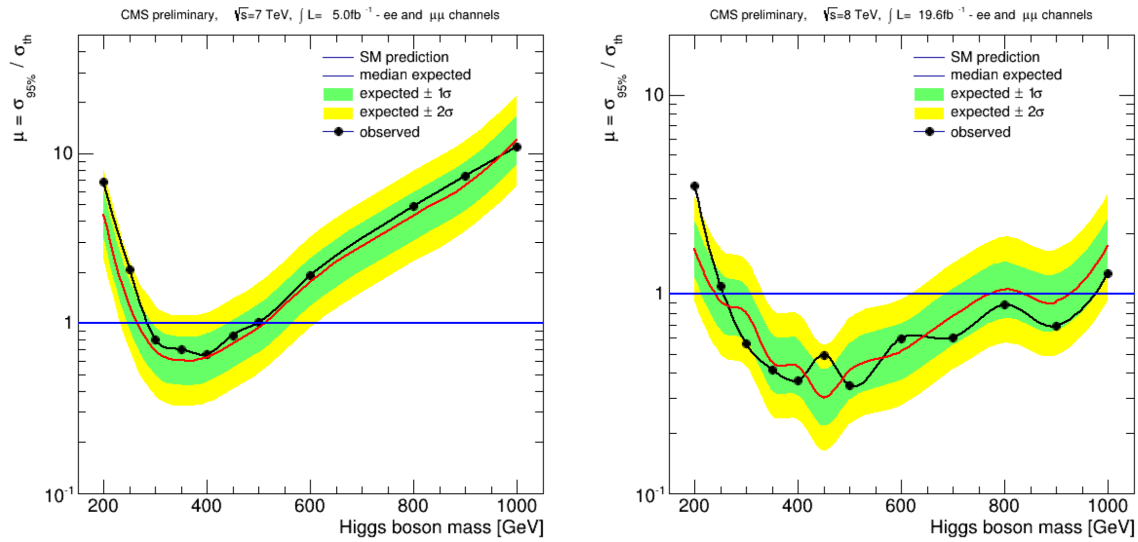


Figure 5.22: The expected and observed 95% C.L. upper limit on $\mu = \sigma/\sigma_{SM}$ is shown as a function of the M_H for 5.0 fb^{-1}at $\sqrt{s} = 7$ TeV (*left*) and 19.7 fb^{-1} integrated luminosity at $\sqrt{s} = 8$ TeV (*right*) for BDT shape based Analysis. This plot combines the limits from ee and $\mu\mu$ channels.

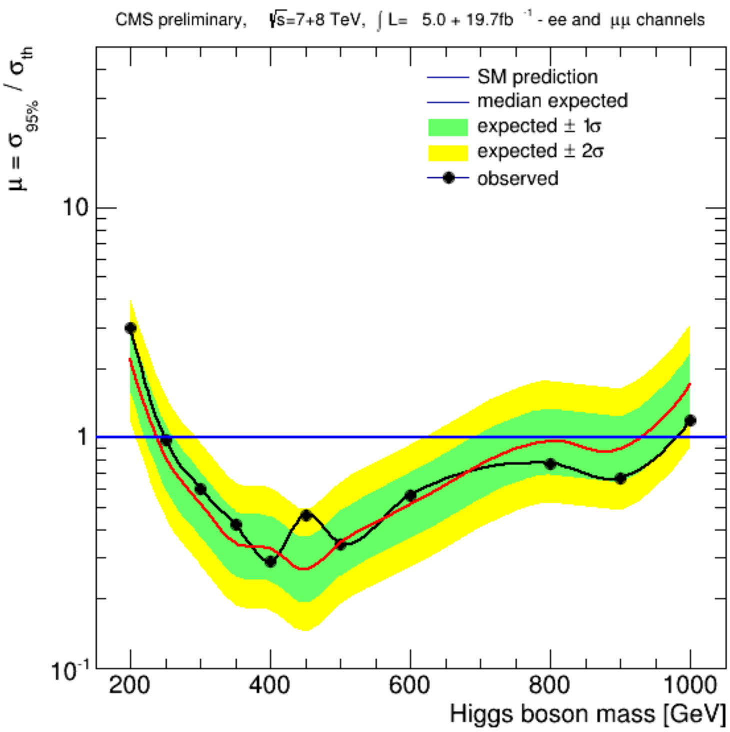


Figure 5.23: The combined expected and observed 95% C.L. upper limit on $\mu = \sigma/\sigma_{SM}$ is shown as a function of M_H for 5.0 fb^{-1}at $\sqrt{s} = 7$ TeV and 19.7 fb^{-1} integrated luminosity at $\sqrt{s} = 8$ TeV for BDT shape based Analysis.

Chapter 6

Summary and Outlook

In this Thesis, the results are presented for the search of the SM Higgs boson in $H \rightarrow ZZ \rightarrow \ell^+\ell^-\nu_\ell\bar{\nu}_\ell$ decay channel, where $\ell = e$ or μ. The search is performed with data collected at the CMS experiment in the years 2011 and 2012 which correspond to an integrated luminosity of 5.0 fb^{-1}at $\sqrt{s} = 7$ TeV and 19.7 fb^{-1}at $\sqrt{s} = 8$ TeV. This is the full data recorded by the CMS experiment till the LHC stopped for its first long shutdown in 2013. This decay mode is particularly important for high masses (M_H) of the Higgs boson. Its branching ratio is about six times larger than $H \rightarrow ZZ \rightarrow 4\ell$ channel, though its resolution is worse than $H \rightarrow ZZ \rightarrow 4\ell$ decay mode because of the presence of E_T^{miss}. However, this channel contributes significantly in a region which has never been probed in any high energy physics experiment before the LHC, i.e. from M_H= 200 to 1000 GeV. Signal events are characterized by the presence of two same flavor leptons consistent with the Z boson decay and high E_T^{miss} coming from two undetected neutrinos in the final state. With this topology of final state, there are many backgrounds to this search which can mimic the signal. These backgrounds mainly include $Z + Jets$, $t\bar{t} + jets$, WW, WZ and ZZ. Other than that, there are some sub-dominant backgrounds arising from single top quark and $W + jets$. An analysis strategy is devised in order to reduce these backgrounds. The whole selection procedure is divided into three parts: *preselection*, *event categorization* and *final selection*.

Preselection includes the lepton selection by applying kinematical selection criteria based on p_{T} of leptons together with p_{T} and invariant mass of the dilepton system. Along with that, some specific selection is applied for reducing particular backgrounds. Events having b-quark jets are rejected (using b-tagging and soft-muon veto) for reducing top-quark related backgrounds. Minimum azimuthal difference between jets and E_T^{miss}, $\Delta\phi(jet,E_T^{miss})$, distribution is used to reduce $Z+jets$ background. After the preselection all the events are divided into categories based on jet multiplicity to increase the search sensitivity. A separate selection criteria based on topology of VBF type events is chosen to identify these events. After selecting VBF type events, rest of them are divided into exclusive 0-jet and inclusive 1-jet ($\geq$ 1-jet) categories. Final selection is then applied based on two variables, E_T^{miss} and M_{T}. This selection is Higgs mass dependent for non-VBF category events and optimized for best expected limit on ratio of measured Higgs boson production cross-section to the SM cross-section. All the major backgrounds in this analysis like $Z+jets$, $t\bar{t}+jets$ and WW are estimated using data-driven techniques. No excess over the SM background prediction is observed. A statistical analysis is carried out using CL_s technique to set up upper limits on the Higgs boson production cross-section at various Higgs mass points and then to set exclusion limit on the Higgs boson mass range. All sources of systematic uncertainties are considered. The baseline analysis is based on M_{T} distribution in non-VBF category and E_T^{miss} distribution in VBF category, i.e. the upper limits are calculated using full distribution without imposing any selection requirement on the final discriminant. This analysis is cross-checked with cut-and-count analysis where whole selection is applied including minimum threshold on M_{T} (E_T^{miss}) for non-VBF(VBF) events. Using baseline analysis, the SM Higgs boson is excluded at 95% CL in Higgs mass range between 248 and 930 GeV. From cut-and-count analysis, the exclusion range is between 268 to 756 GeV.

A multivariate analysis based on BDT statistical technique is introduced to explore further possibility of improving the sensitivity of this decay channel and to cross-check the baseline analysis using an independent technique. BDT deals with the correlations of various input variables and get the best discrimination out of them. This analysis also considered the full data collected by the CMS experiment. All the analysis selection procedures are same till preselection except that $\Delta\phi(jet, E_T^{miss})$ threshold criterion is not applied and in fact it is used as one of the input variables. Also, in the final selection, only $E_T^{miss} > 80$ GeV is applied and selection criterion on M_T is relaxed. The choice of input variables is same as that in the baseline analysis to have a direct comparison of BDT with baseline analysis. Various parameters related with BDT are optimized based on best expected limit on ratio of measured Higgs boson production cross-section to the SM cross-section, while avoiding the over-training. Training of BDT is done for each Higgs mass hypothesis and separately for $e^+e^-\nu_e\bar{\nu_e}$ and $\mu^+\mu^-\nu_\mu\bar{\nu_\mu}$ decay modes. A boosting technique is used to take care of mis-classified events while training BDT. All the backgrounds are used in the training and events are classified based on jet multiplicity in 0-jet, 1-jet and >1-jet categories. The data-driven background estimation techniques are same as that used in baseline analysis. No excess over the SM background is observed and a statistical analysis is carried out to impose upper limits on the Higgs boson production cross-section. All sources of the systematic uncertainties are considered and a BDT shape analysis is carried out on final discriminant. The SM Higgs boson is excluded in the mass range between 248 to 963 GeV.

In the future, this analysis can be a good benchmark for studies of many other physics problems beyond the SM. One of them is to explore the hidden sector SM singlet fields which appear in many theories such as string inspired particle physics models containing many more gauge groups and scalar sectors than that present in the SM [28]. There can be mixing of Higgs bosons from these fields with the discovered boson at around 125 GeV mass, and a new SM-like Higgs boson appears with very high mass and narrower width. Since this decay channel is highly sensitive in high mass range so it will be an ideal scenario to search for such signatures at higher LHC energy and higher luminosity. Along with that, with this channel, one can study anomalous triple gauge coupling in ZZ final state and search for signatures of new physics. This analysis is also very useful in constraining the SM Higgs decay width using its off-shell production at higher masses [91]. Hence, keeping in mind the higher center of mass energy at the LHC after the technical shutdown in 2015, a variety of physics signals can be probed using this decay channel and one can explore the unexplored regions of particle physics.

Bibliography

[1] S. Weinberg, "A model of leptons", Phys. Rev. Lett., vol. 19, pp. 12641266, 1967.

[2] A. Salam, "Weak and electromagnetic interactions", in Elementary particle physics: relativistic groups and analyticity (N. Svartholm, ed.), p. 367, Almquvist & Wiskell, 1968. Proceedings of the eighth Nobel symposium.

[3] S. L. Glashow,"Partial Symmetries of Weak Interactions", Nucl. Phys., vol. 22, pp. 579588, 1961.

[4] M. K. Gaillard, P.D. Grannis, and F. J. Sciulli, "The Standard Model of Particle Physics", Rev. Mod. Phys. 71, S96 (1999).

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

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

[7] K. Nakamura et al., "Particle Data Group", J. Phys. G 37, 075021, 2010.

[8] P.W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett., vol. 13, p. 508, 1964.

[9] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", Phys. Rev. Lett., vol. 13, pp. 321323, 1964.

[10] G.S. Guralnik, C.R. Hagen, and T.W.B. Kibble, "Global conservation laws and massless particles", Phys. Rev. Lett., vol. 13, pp. 585587, 1964.

[11] V.A. Bednyakov et al., "On Higgs mass generation mechanism in the Standard Model", arxiv:hep-ph/0703280v1.

[12] S. Dittmaier and M.Schumacher, "The Higgs Boson in the Standard Model - From LEP to LHC: Expectations, Searches, and Discovery of a Candidate", arxiv:hep-ph/1211.4828v3.

[13] "TASI 2004 Lecture Notes: Higgs Boson Physics", arXiv:hep-ph/0512377v1.

[14] M. E. Peskin and D. V. Schroeder,"An Introduction to quantum field theory", (June 1, 1995).

[15] T. Hambye and K. Riesselmann, "Matching conditions and Higgs mass upper bounds revisited", Phys. Rev. D 55 (1997) 7255.

[16] 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, vol. 565, p. 61, 2003.

[17] The TEVNPH Working Group, "Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data", arXiv:1207.0449 [hep-ex].

[18] The LEP Electroweak Working Group. *http://lepewwg.web.cern.ch/LEPEWWG/.*

[19] M. Spira et al., "Higgs boson production at the LHC", Nucl. Phys. B, vol. 453, pp. 1782, 1995.

[20] S. Actis et al., "NLO electroweak corrections to Higgs boson production at hadron colliders", Phys. Lett. B, vol. 670, pp. 1217, 2008.

[21] C. Anastasiou and K. Melnikov, "Higgs boson production at hadron colliders in NNLO QCD", Nucl. Phys. B, vol. 646, pp. 220256, 2002.

[22] M. Ciccolini, A. Denner, and S. Dittmaier, "Electroweak and QCD corrections to Higgs production via vector-boson fusion at the LHC", Phys. Rev. D, vol. 77, p. 013002, 2008.

[23] LHC Higgs Cross Section Working Group and Dittmaier, S. and Mariotti, C. et al, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", CERN-2011-002, arxiv:1101.0593.

[24] LHC Higgs Cross Section Working Group and Dittmaier, S. and Mariotti, C. et al, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", CERN-2011-002, arxiv:1201.3084.

[25] LHC Higgs Cross Section Working Group and Dittmaier, S. and Mariotti, C. et al, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", CERN-2011-002, arxiv:1307.1347.

[26] LHC Higgs Cross Section Working Group *https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.*

[27] The CMS Collaboration,"Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV", CMS-PAS-HIG-13-005.

[28] M. Bowen, Y. Cui, and J. D. Wells, "Narrow trans-TeV Higgs bosons and $H \to hh$ decays Two LHC search paths for a hidden sector Higgs boson", JHEP 0703 (2007) 036, arXiv:hep-ph/0701035.

[29] Lyndon Evans and Bryant (editors), "LHC Machine", JINST 3 S08001 (2008).

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

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

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

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

[34] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", JINST 3 S08006 (2008).

[35] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", JINST 3 S08007 (2008).

[36] CMS Collaboration, "CMS, the magnet project: Technical design report", CERN/LHCC 97-10, CMS TDR 1 (1997).

[37] CMS Collaboration, "A lead/scintillator electromagnetic calorimeter for CMS", (1993) 34.

[38] CMS Collaboration, "The CMS electromagnetic calorimeter project : Technical Design Report", (1997) 34.

[39] CMS Collaboration, "Changes to CMS ECAL electronics : addendum to the Technical Design Report", (2002) 34.

[40] 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.

[41] CMS-HCAL Collaboration, S.Abdullin et al., "Design performance and calibration of CMS forward calorimeter wedges", Eur. Phys. J. C53 (2008) no. 139-166.

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

[43] CMS Collaboration, G,Bayatian et al., “The CMS muon project: Technical Design Report”, CERN-LHCC-97-032; CMS-TDR-003, 1997.

[44] CMS Collaboration, “Performance of the CMS Drift Tube Chambers with Cosmic Rays”, JINST 5 (2010) -03015, 0911.4855.

[45] CMS Collaboration, “Performance of the CMS cathode strip chambers with Cosmic Rays”, JINST 5 (2010) no. 03, T03018, arxiv:0911.4992v2 [physics.ins-det]. http://cdsweb.cern.ch/record/1224837. CMS-CFT-09-014.

[46] M.Abbrecia et al., “Resistive plate chambers performances at cosmic rays fluxes”, Nucl. Instrum. Methods A 359 (1995) no. 3, 603.

[47] CMS Collaboration, “CMS TriDAS Project, Technical Design Report, Volume 1 The Trigger Systems”, CERN/LHCC 2000-38, CMS TDR 6.1 (2000).

[48] W. Adam et al., “The CMS high level trigger”, Eur. Phys. J., vol. C46, pp. 605667, 2006.

[49] CMS Collaboration, “CMS Computing Project, Technical Design Report”, CERN/LHCC 2005-023, CMS TDR 7, 2005.

[50] Agostinelli, S. et al., “Geant 4: A Simulation Toolkit”, Nucl. Inst. Meth., A506 (2003), 250.

[51] CMS Collaboration, “CMS physics Technical Design Report, Volume II: Physics Performance”, oai:cds.cern.ch:942733. J. Phys. G, 34 (2006), 995. Revised version submitted on 2006-09-22 17:44:47.

[52] D. Spiga et al., “CRAB: the CMS distributed analysis tool development and design”, Nuclear Physics B Proceedings Supplements, Volume 177 (2008), p. 267-268.

[53] CMS Collaboration,“CMS Physics Technical Design Report Volume I: Detector Performance and Software”, CERN-LHCC-2006-001, *http://cmsdoc.cern.ch/cms/cpt/tdr/ptdr1 final colour.pdf, 2006.*

[54] Kenneth Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, Proceedings of the IEEE, 86(11):2210-2239, 1998.

[55] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 262 (1987), no. 2-3, 444 450. doi:10.1016/0168-9002(87)90887-4.

[56] T. Speer, R. Fruhwirt, “A Gaussian-sum filter for vertex reconstruction”, 2006 Comp. Phys. Comm. 174 935.

[57] R. Fruhwirth, W.Waltenberger and P.Vanlaer, “Adaptive Vertex Fitting”, Technical Report CMS-NOTE-2007-008, CERN, Geneva, March, 2007.

[58] CDF Collaboration, F. Abe et al., “Evidence for top quark production in $p - \bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV”, Phys.Rev. D50(1994) 2966. doi:10.1103/PhysRevD.50.2966.

[59] Paolo Azzurri, “Track Reconstruction performance in CMS”, arXiv:0812.5036v1.

[60] CMS Collaboration, “Electron reconstruction and identification at $\sqrt{s} = 7$ TeV”, CMS-PAS-EGM-10-004, 2010.

[61] CMS Collaboration, “The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger”, CERN/LHCC 2002-26 (2002). CMS TDR 6.2.

[62] H. Bethe and W. Heitler, “On the Stopping of fast particles and on the creation of positive electrons”, Proc.Roy.Soc.Lond., vol. A146, pp. 83112, 1934.

[63] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC”, eConf, vol. C0303241, p. TULT009, 2003.

[64] CMS Collaboration, “Particleflow event reconstruction in CMS and performance for jets, taus, and E_T^{miss}”, CMS-PAS-PFT-09-001, 2009.

[65] CMS Collaboration, “Electron performance with 19.6 fb^{-1} of data collected at $\sqrt{s} = 8$ TeV with the CMS detector”, CMS-DP-2013/003.

[66] G. Abbiendi et al., “Muon Reconstruction in the CMS Detector”, CMS AN-2008/097, July 2009.

[67] V. Innocente, M. Maire, and E. Nagy, “GEANE: Average Tracking and Error Propagation Package”, CERN Program Library, IT-ASD W5013-E (1991).

[68] M. Mulders, et al., “Muon Identification in CMS”, CMS Note 2008/098 (2008).

[69] CMS Collaboration, “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-PFT-10-002 (2010).

[70] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, CMS-PAS-MUO-10-004, arXiv:1206.4071v2, CERN-PH-EP/2012-173.

[71] CMS Collaboration, “Photon reconstruction and identification at $\sqrt{s} = 7$ TeV”, CMS-PAS-EGM-10-005.

[72] G. P. Salam, “Towards Jetography”, hep-ph/0906.1833v2.

[73] CMS Collaboration, “Jet Performance in pp Collisions at $\sqrt{s} = 7$ TeV”, CMS PAS JME-10-03(2010).

[74] CMS Collaboration, “CMS Technical design report, volume II:Physics performance”, J.Phys. G34(2007)995.

[75] CMS Collaboration, “Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS”, CMS PAS JME-09-02(2009).

[76] CMS Collaboration, “Tracking and Vertexing Results from First Collisions”, CMS PAS TRK-10-001(2010).

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

[78] S. Catani et. al., "Longitudinally-invariant k-clustering algorithms for hadron-hadron collisions", Nucl. Phys. B 406 (1993) 187-224.

[79] Stephen D. Ellis, Davision E. Soper, "Successive Combination Jet Algorithm For Hadron Collisions", Phys.Rev. D 48 (1993) 3160-3166, [hep-ph/9305266].

[80] Yu.L. Dokshitzer, G.D. Leder, S. Moretti, B.R. Webber, "Better Jet Clustering Algorithms", JHEP 9708:001 (1997) [hep-ph/9707323].

[81] CMS Collaboration, "Plans for Jet Energy Corrections at CMS", CMS-PAS-JME-07-002, (2007).

[82] Niki Saoulidou, "Particle Flow Jet Identification Criteria", CMS AN-2010/003, (2010).

[83] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/ψ and W events at 7 TeV", CMS PAS PFT-2010-003 (2010).

[84] CMS Collaboration, "Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV with the CMS Detector", CMS-PAS-JME-12-002.

[85] CMS Collaboration, "τ Reconstruction using the Particle Flow Technique", CMS PAS PFT-08-01, 2008.

[86] CMS Collaboration, "Performances of τ-lepton reconstruction and identification in CMS", JINST 7, P01001, 2012.

[87] A. Rizzi, F. Palla, and G. Segneri, "Track impact parameter based b-tagging with CMS", CMS Note 2006/019 (2006).

[88] C.Weiser, “A Combined Secondary Vertex Based B-Tagging Algorithm in CMS”, CMS Note 2006/014 (2006).

[89] A. Bocci, P. Demin, et al., “Tagging b jets with electrons and muons at CMS”, CMS Note 2006/043 (2006).

[90] CMS Collaboration, “Search for a standard-model-like Higgs boson with a mass of up to 1 TeV at the LHC”, arXiv:1304.0213.

[91] CMS Collaboration, “Constraints on the Higgs boson width from off-shell production and decay to $ZZ \to \ell\ell\ell'\ell'$ ”, CMS-PAS-HIG-14-002.

[92] S. Alioli, P. Nason, C. Oleari et al., “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, JHEP 1006 (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[93] Lenzi, P. and Padhi, S. and Ceballos, G. andWuerthwein, F., “Standard Model Cross Sections for CMS at 8 TeV”, twiki, (2012).

[94] J. M. Campbell, R. K. Ellis, and C.Williams, “Vector boson pair production at the LHC”, JHEP 1107 (2011) 018, arXiv:1105.0020.

[95] J. Alwall, M. Herquet, F. Maltoni et al., “MadGraph 5 : Going Beyond”, JHEP 1106 (2011) 1211 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

[96] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, JHEP 05 (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[97] G. Passarino, C. Sturm, and S. Uccirati, “Higgs pseudo-observables, second Riemann sheet and all that”, Nucl. Phys. B834 (2010) 77115, doi:10.1016/j.nuclphysb.2010.03.013, arXiv:1001.3360.

[98] S. Goria, G. Passarino, and D. Rosco, “The Higgs Boson Lineshape”, Nucl.Phys. B864(2012) 530579, doi:10.1016/j.nuclphysb.2012.07.006, arXiv:1112.5517.

[99] N. Kauer and G. Passarino, "Inadequacy of zero-width approximation for a light Higgs boson signal", JHEP 1208 (2012) 116, doi:10.1007/JHEP08(2012)116, 1234 arXiv:1206.4803.

[100] G. Passarino, "Higgs Interference Effects in $gg \to ZZ$ and their Uncertainty", JHEP 1208 (2012) 146, doi:10.1007/JHEP08(2012)146, arXiv:1206.3824.

[101] T. Binoth, N. Kauer, and P. Mertsch, "Gluon-induced QCD corrections to $pp \to ZZ \to l\bar{l}l'\bar{l}'$", arXiv:0807.0024.

[102] Schoefbeck, R. and others, "MET Optional Filters", (2011) *https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFilters.*

[103] G. Petrucciani et al. *https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideTagProbeFitTreeProducer.*

[104] P. Musella, "Jet identification in high pile-up environment", *https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsJME13005*, (2012).

[105] CMS Collaboration, "Updated results on the new boson discovered in the search for the Standard Model Higgs boson in the $H \to ZZ \to 4\ell$ channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV", HIG-12-041, 2012.

[106] CMS Collaboration, "Jet Energy Resolution in CMS at $\sqrt{s}$ =7 TeV", CMS-PAS-JME-10-014.

[107] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", JINST 6 (2011) 11002, arXiv:1107.4277v1.

[108] Bloch, D. and others, "Combination of b-tagging efficiency measurements in 2012 data at 8 TeV pp collision", CMS Analysis Note CMS AN 2012/470 (2012).

[109] M. Botje, J. Butterworth, A. Cooper-Sarkar et al., "The PDF4LHCWorking Group Interim Recommendations", arXiv:1101.0538.

[110] S. Alekhin, S. Alioli, R. D. Ball et al., "The PDF4LHCWorking Group Interim Report", arXiv:1101.0536.

[111] H.-L. Lai, M. Guzzi, J. Huston et al., "New parton distributions for collider physics", Phys.Rev. D82 (2010) 074024, doi:10.1103/PhysRevD.82.074024, arXiv:1007.2241.

[112] A. Martin, W. Stirling, R. Thorne et al., "Parton distributions for the LHC", Eur.Phys.J. C63 (2009) 189285, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

[113] NNPDF Collaboration, "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology", arXiv:1101.1300.

[114] A. L. Read, "Presentation of search results:the CL_s technique", J.Phys.G:Nucl.Part.Phys. 28(2002)2693.

[115] A. L. Read, "Modified frequentist analysis of search results (the CLs method)", CERN Report CERN-OPEN-2000-005, (2000).

[116] T. Junk, "Confidence level computation for combining searches with small statistics", Nucl. Instrum. Meth. A 434 (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.

[117] ATLAS and CMS Collaborations, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", ATL-PHYS-PUB/CMS NOTE 1271 2011-11, 2011/005, (2011).

[118] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", Nucl. Instrum. Meth. A 389, 81, 1997. *http://root.cern.ch/*.

[119] A. Hocker et al., "TMVA: Toolkit for multivariate data analysis". *http://arxiv.org/abs/physics/0703039v4*. *http://tmva.sourceforge.net/*.

Appendix A

A.1 Collision Datasets

Table A.1 shows the datasets for the collision data which are used in the analysis.

Dataset	Global tag	$\int \mathcal{L}$ (pb^{-1})	Run range
/PD/Run2012A-13Jul2012-v1/AOD	FT_53_V6_AN2	890	190645-193621
/PD/Run2012A-recover-06Aug2012-v1/AOD	FT_53_V6C_AN2	82	190782-190949
/PD/Run2012B-13Jul2012-v1/AOD	FT_53_V6_AN2	4430	193834-196531
/PD/Run2012C-24Aug2012-v1/AOD	FT_53_V10_AN2	490	198049-198522
/PD/Run2012C-PromptReco-v2/AOD	FT_53_V10_AN2	6390	198941-203002
/PD/Run2012D-PromptReco-v1/AOD	GR_P_V42_AN2	7380	203894-204567
Total		**19662**	

Table A.1: Datasets used for analysis. The primary datasets (PDs) correspond to DoubleElectron, DoubleMu, SingleMu, MuEG, Photon and PhotonHad The integrated luminosity and the run-ranges are shown for each data period.

A.2 Simulation datasets – Signal

Table A.2 shows the signal samples for all Higgs mass points. These are generated using POWHEG MC generator.

A.3 Simulation datasets – Background

Table A.3 shows the various background MC samples along with their cross-section values used in the analysis. For the study of the photon sample we use the samples described in

Higgs mass	Dataset	σ (pb)	$BR(H \to ZZ)$
200	/GGtoH200_8TeV-powheg/S12_S53-v1 /VBFtoH200_8TeV-powheg/S12_S53-v1	7.127 0.8685	0.256
250	/GGtoH250_8TeV-powheg/S12_S53-v1 /VBFtoH250_8TeV-powheg/S12_S53-v1	4.802 0.6225	0.297
300	/GGtoH300_8TeV-powheg/S12_S53-v1 /VBFtoH300_8TeV-powheg/S12_S53-v1	3.606 0.4408	0.307
350	/GGtoH350_8TeV-powheg/S12_S53-v1 /VBFtoH350_8TeV-powheg/S12_S53-v1	3.406 0.3200	0.307
400	/GGtoH400_8TeV-powheg/S12_S53-v1 /VBFtoH400_8TeV-powheg/S12_S53-v1	2.924 0.2543	0.269
450	/GGtoH450_8TeV-powheg/S12_S53-v1 /VBFtoH450_8TeV-powheg/S12_S53-v1	2.003 0.2002	0.259
500	/GGtoH500_8TeV-powheg/S12_S53-v1 /VBFtoH500_8TeV-powheg/S12_S53-v1	1.283 0.1561	0.261
550	/GGtoH550_8TeV-powheg/S12_S53-v1 /VBFtoH550_8TeV-powheg/S12_S53-v1	0.8141 0.1223	0.267
600	/GGtoH600_8TeV-powheg/S12_S53-v1 /VBFtoH600_8TeV-powheg/S12_S53-v1	0.5230 0.09688	0.272
700	/GGtoH700_8TeV-powheg/S12_S53-v1 /VBFtoH700_8TeV-powheg/S12_S53-v1	0.2288 0.06330	0.285
800	/GGtoH800_8TeV-powheg/S12_S53-v1 /VBFtoH800_8TeV-powheg/S12_S53-v1	0.1095 0.04365	0.296
900	/GGtoH900_8TeV-powheg/S12_S53-v1 /VBFtoH900_8TeV-powheg/S12_S53-v1	0.0568 0.03164	0.304
1000	/GGtoH1000_8TeV-powheg/S12_S53-v1 /VBFtoH1000_8TeV-powheg/S12_S53-v1	0.03163 0.02399	0.311

Table A.2: List of signal samples used in analysis. GGtoH, VBFtoH, powheg, S12-S53 are used as an abbreviation for: GluGluToHToZZTo2L2Nu_M, VBF_ToHToZZTo2L2NU_M, powheg-pythia6, Summer12_DR53X-PU_S10_START53_V7A.

Table A.4.

A.4 Various triggers used in analysis

Table A.5 shows the various triggers used in the analysis. It includes dilepton triggers for signal selection, γ triggers for selecting $\gamma + jets$ events and $e - \mu$ triggers for selecting events for non-resonant background estimation.

Process	Dataset	$\sigma \cdot k$ (pb)
	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/S12_S53-v1 †	
	/W1JetsToLNu_TuneZ2Star_8TeV-madgraph/S12_S53-v1	
$W \to \ell\nu$	/W2JetsToLNu_TuneZ2Star_8TeV-madgraph/S12_S53-v1	36257
	/W3JetsToLNu_TuneZ2Star_8TeV-madgraph/S12_S53-v1	
	/W4JetsToLNu_TuneZ2Star_8TeV-madgraph/S12_S53-v1	
	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/S12_S53-v1 †	
	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/S12_S53-v1	
$Z \to \ell\ell$	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/S12_S53-v1	3533
	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/S12_S53-v1	
	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/S12_S53-v1	
EWK $\ell\ell$+2 jets	/DYJJ01JetsToLL_M-50_MJJ-200_TuneZ2Star_8TeV-madgraph_tauola/S12-53x-v1	0.75
	/TTJets_FullLeptMGDecays_8TeV-madgraph/S12_S53-v2	23.89
$t\bar{t}$+X	/TTWJets_8TeV-madgraph/S12_S53-v1	0.232
	/TTZJets_8TeV-madgraph_v2/S12_S53-v1	0.208
	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/S12_S53-v1 ($\bar{t}$)	11.2
	/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/S12_S53-v1 (t)	11.2
	/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/S12_S53-v1 ($\bar{t}$)	30.7
Single top	/T_t-channel_TuneZ2star_8TeV-powheg-tauola/S12_S53-v1 (t)	56.4
	/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/S12_S53-v1 ($\bar{t}$)	1.76
	/T_s-channel_TuneZ2star_8TeV-powheg-tauola/S12_S53-v1 (t)	3.89
	/WZJetsTo3LNu_TuneZ2_8TeV-madgraph-tauola/S12_S53-v1	1.057
Dibosons	/WWJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola/S12_S53-v1	5.995
	/ZZJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola/S12_S53-v3	0.320

Table A.3: List of the SM dilepton MC samples used in the comparison with data. For the different processes (signal and background) considered the expected cross sections are quoted. S12_S53 is used as an abbreviation for Summer12_DR53X-PU_S10_START53_V7A.

Process	Dataset	σ (pb)
γ+jets	G_Pt-30to50_TuneZ2star_8TeV_pythia6_Z2/S12-S53-v1	1.99×10^{4}
	G_Pt-50to80_TuneZ2star_8TeV_pythia6/S12-S53-v1	3.32×10^{3}
	G_Pt-80to120_TuneZ2star_8TeV_pythia6/S12-S53-v1	5.58×10^{2}
	G_Pt-120to170_TuneZ2star_8TeV_pythia6/S12-S53-v1	108
	G_Pt-170to300_TuneZ2star_8TeV_pythia6/S12-S53-v1	30.1
	G_Pt-300to470_TuneZ2star_8TeV_pythia6/S12-S53-v1	2.14
	G_Pt-470to800_TuneZ2star_8TeV_pythia6/S12-S53-v1	0.212
	G_Pt-800to1400_TuneZ2star_8TeV_pythia6/S12-S53-v1	7.08×10^{-3}
QCD	QCD_Pt-30to80_EMEnriched_TuneZ2star_8TeV-pythia6/S12-S53-v1	4615893
	QCD_Pt-80to170_EMEnriched_TuneZ2star_8TeV-pythia6/S12-S53-v1	183295
	QCD_Pt-170to250_EMEnriched_TuneZ2star_8TeV-pythia6/S12-S53-v1	4587
	QCD_Pt-250to350_EMEnriched_TuneZ2star_8TeV-pythia6/S12-S53-v1	557
	QCD_Pt-350_EMEnriched_TuneZ2star_8TeV-pythia6/S12-S53-v1	89

Table A.4: Data and MC samples analyzed for the estimation of the $Z + jets$ background. The expected cross section is quoted for each MC process. S12-S53 is used as an abbreviation for Summer12_DR53X-PU_S10_START53_V7A.

Primary dataset	Trigger path	Type
DoubleElectron	Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL _Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL	signal
	Ele32_CaloIdT_CaloIsoT _TrkIdT_TrkIsoT_SC17_Mass50	eff. measurement
	Photon22_R9Id90_HE10_Iso40_EBOnly	control region
	Photon36_R9Id90_HE10_Iso40_EBOnly	control region
	Photon50_R9Id90_HE10_Iso40_EBOnly	control region
	Photon75_R9Id90_HE10_Iso40_EBOnly	control region
	Photon90_R9Id90_HE10_Iso40_EBOnly	control region
DoubleMu	Mu17_Mu8	signal
	IsoMu24_eta2p1	signal and eff. measurement
MuEG	Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL	control region
	Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL	control region
PhotonHad	Photon135	control region
	Photon150	control region
	Photon160	control region
	Photon250_NoHE	control region
	Photon300_NoHE	control region

Table A.5: Triggers used for the analysis and their provenance with respect to the primary datasets. The last column shows the purpose with which the trigger is used: signal - for the search of our signal; eff. measurement - for the measurement of trigger/selection efficiencies; control region - to build specific control regions for data-driven background estimation.

Appendix B

B.1 γ fits and weights

Figure B.1 shows the regularized fits for parametrizations obtained from Equation 4.5 for p_{T} distributions for $\gamma + jets$ and dilepton samples from data. The weights for reweighting $\gamma + jets$ sample are calculated using these fits for various event categories. The distributions for these weights are shown Figure B.2 for ee and $\mu\mu$ channels.

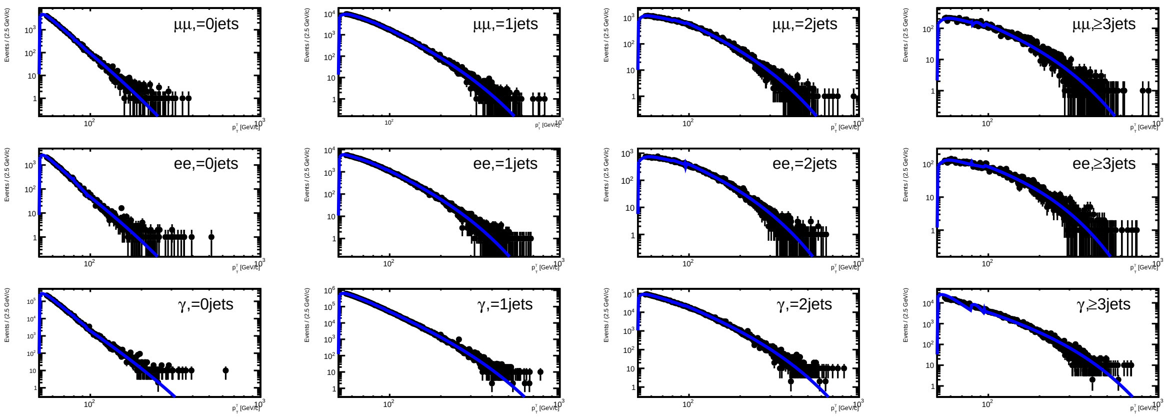


Figure B.1: Di-muon (*top*), di-electron (*center*) and photon (*bottom*) p_{T} spectra for different jet multiplicity bins. The blue lines correspond to the regularized fit of Equation. 4.5.

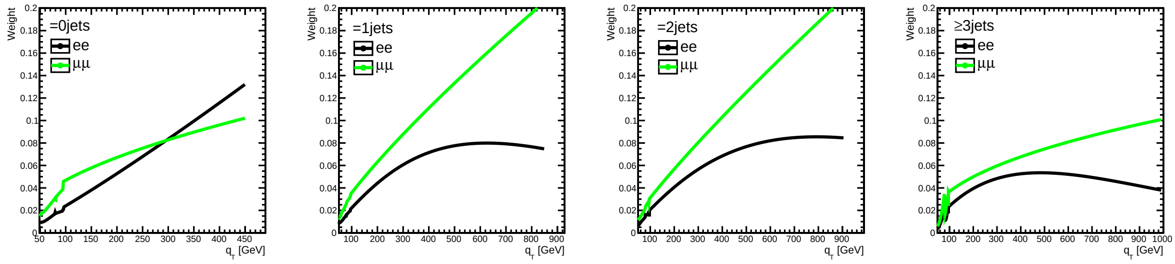


Figure B.2: Photon weights derived from the parametrization of the p_{T} spectra.

B.2 Closure Tests

Figure B.3 shows the M_{T} plots for $Z\!+\!jets$ MC and reweighted $\gamma\!+\!jets$ MC sample. Good agreement is found between two samples which gives us confidence in this background estimation method. The distributions are shown for inclusive as well as in various event categories.

Further closure test is performed using $\min\Delta\phi(\text{jet}, E_{\mathrm{T}}^{miss})$ distribution for various E_T^{miss} requirements. The distributions are in Figure B.4. The discrepancy is enhanced by larger E_T^{miss} requirements. However it is not expected to yield significant impact on the analysis given that in the tail the shape is correctly described and therefore the bias of the prediction is small ($< 5\%$).

In data the level of agreement between the γ+jets and Z+jets samples can be first of all measured by the profile of the E_T^{miss} and vertex distribution as function of the different data periods for $\sqrt{s} = 8$ TeV data. These distributions are profiled and shown as a function of Run number and of the average instantaneous luminosity. Given that these distributions are profiled, the main contribution is indeed instrumental and the result is, in practice, insensitive to the contamination of the tails by processes with real E_T^{miss} which have much lower cross sections in comparison with γ+jets and Z+jets. Figure B.5 shows the average number of vertices as function of the run number (*left*) and as function of the average instantaneous luminosity (*right*). It is concluded that all the three samples analyzed, i.e. γ, ee and $\mu\mu$, are in good agreement with each other in the run range analyzed and for the different pile-up conditions included in the dataset. Notice that for simplicity the γ curve represented in this plots corresponds to the one obtained after re-weighting to

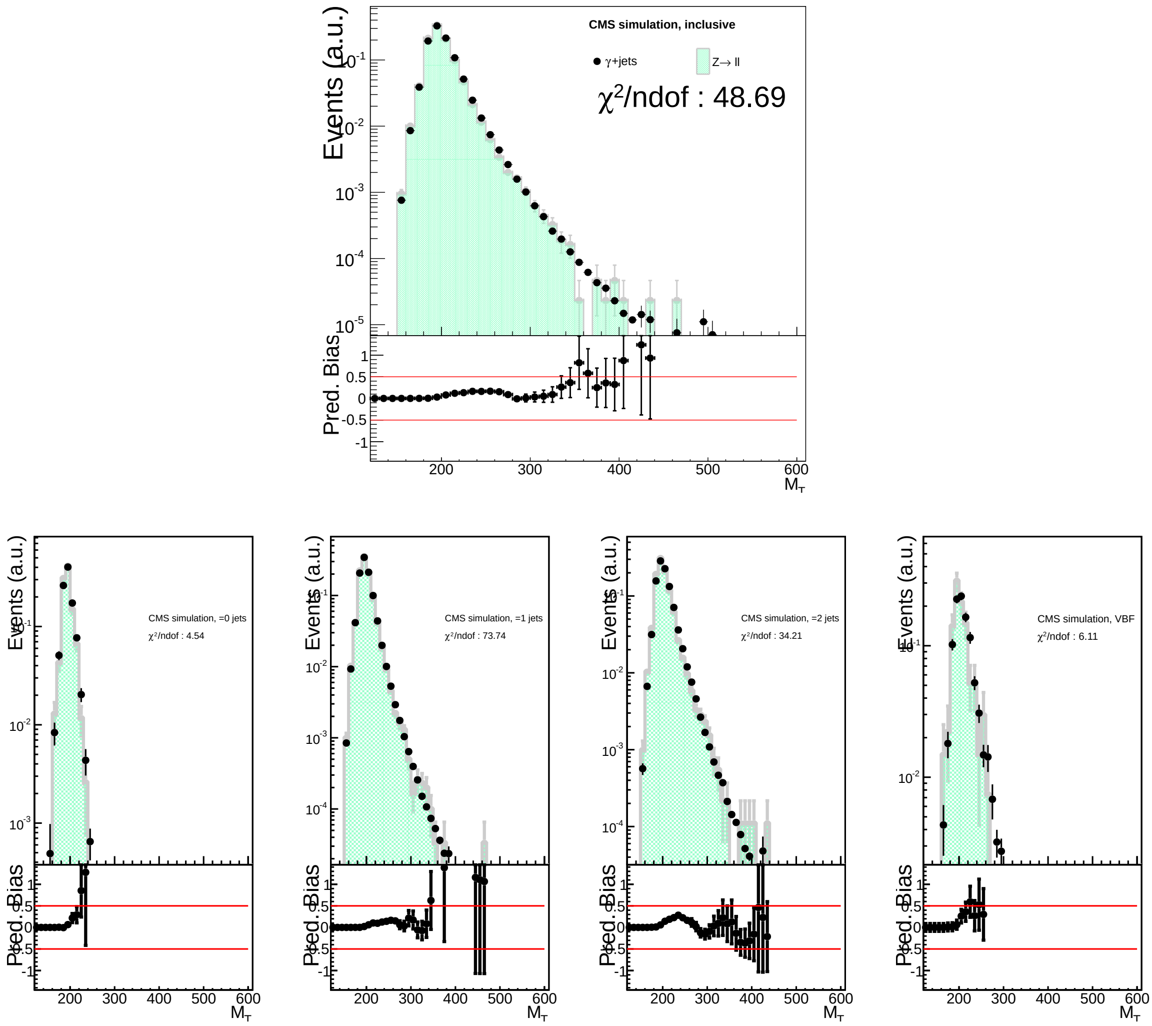


Figure B.3: Closure test of the γ+jets method for the M_{T} variable using MC samples. (*top*) distribution is for inclusive case and (*bottom) distributions show the M_{T} for different event categories. The distributions combine ee and $\mu\mu$ channels.*

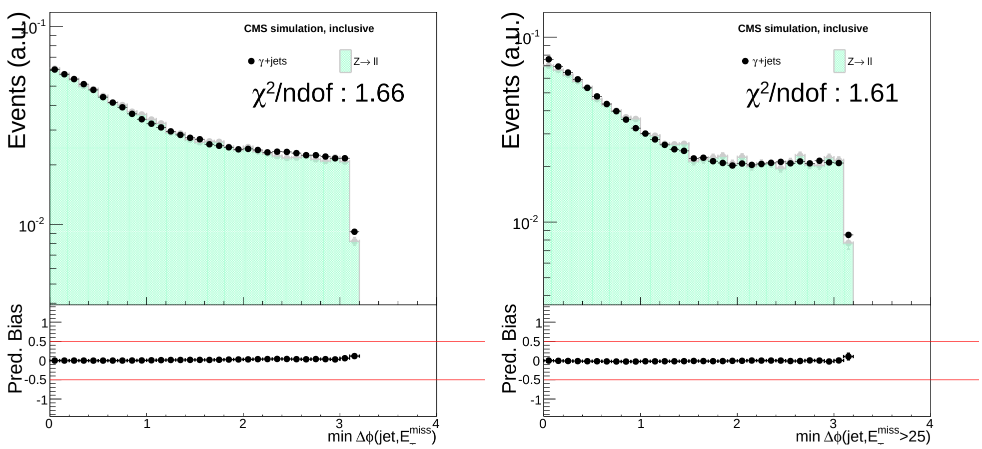


Figure B.4: Closure test of the γ+jets method for the $\min \Delta\phi(\mathrm{jet}, E_{\mathrm{T}}^{\mathrm{miss}})$ variable using MC samples considering E_T^{miss} >0 (*left*) and E_T^{miss} >25 GeV(*right*) events.

the ee spectrum. The γ+jets vertex multiplicity distribution is reweighted to match the dilepton one in each jet multiplicity category. A similar level of agreement is obtained for the average E_T^{miss} reconstructed in each sample throughout the run range analyzed and for different pile-up conditions (see Figure B.6). This observation re-inforces the confidence in the extraction of the instrumental background description from the γ+jets control sample.

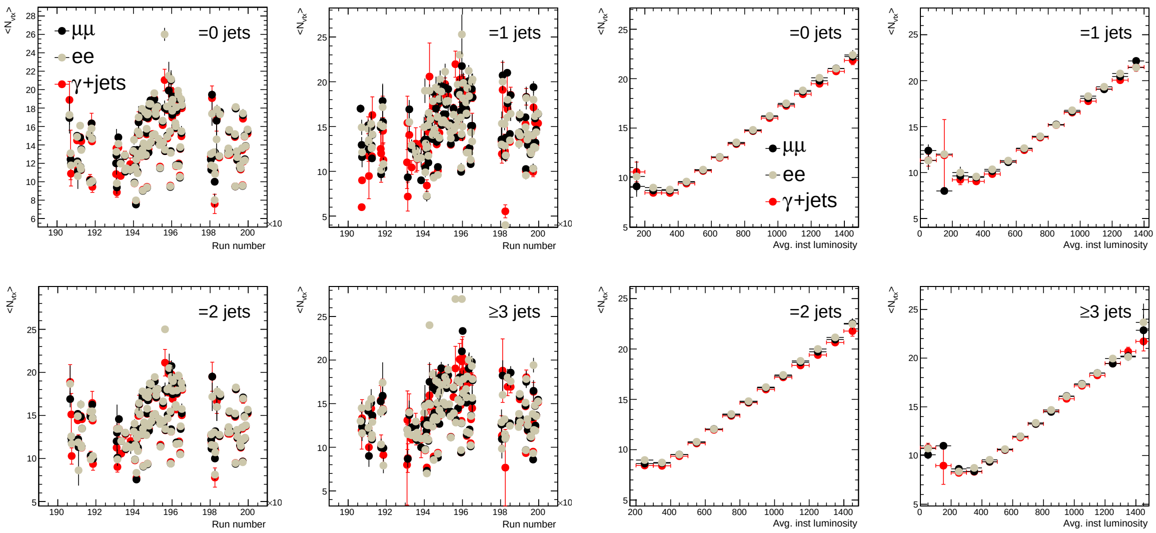


Figure B.5: Average number of vertices in photon and dilepton events as function of the run number (4 plots on the *left*) and of the average instantaneous luminosity (4 plots on the *right*). Different event categories are shown for comparison. In clockwise order starting from (*top left*): inclusive, 0 jets, 1 jets, at least 2 jets.

B.3 Data-MC validation distributions for $\sqrt{s} = 7$ TeV data

Figure B.7 show the E_T^{miss} distributions for various event categories after preselection. A good agreement is found between data and background prediction. Figure B.8 shows the distribution for the $\min\Delta\phi(\text{jet}, E_{\mathrm{T}}^{miss})$ variable for $\geq$ 1-jet and VBF event category after $E_T^{miss} > 50$ GeV. A slight deficit (excess) is found in the ee ($\mu\mu$) channel with at least one jet at high values of $\min\Delta\phi(\text{jet}, E_{\mathrm{T}}^{miss})$. These discrepancies are covered by the final uncertainty assigned to the method of $Z + jets$ background estimation.

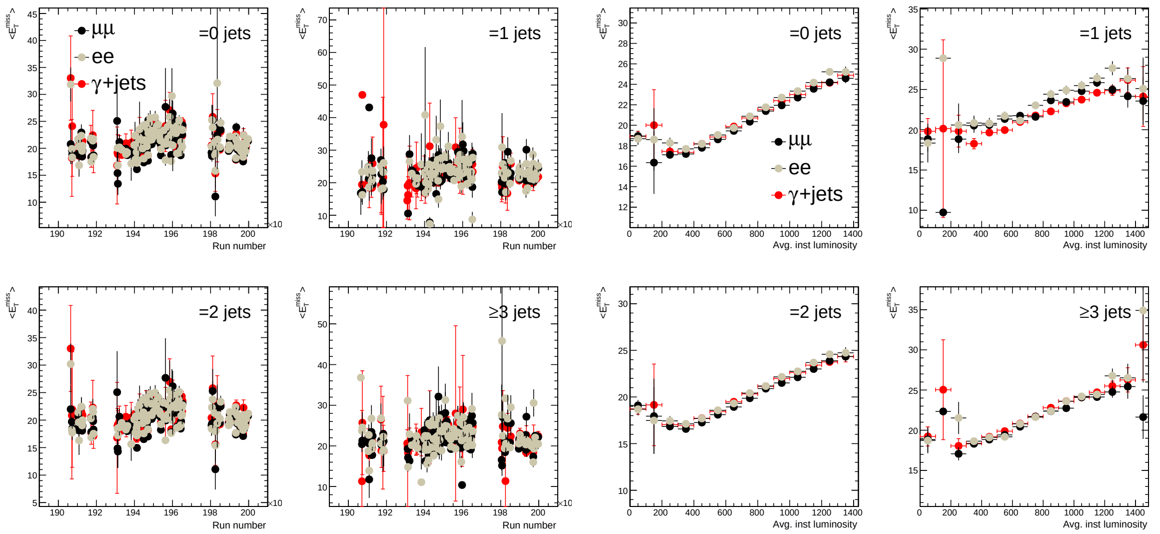


Figure B.6: Average E_T^{miss} in photon and dilepton events as function of the run number (4 plots on the *left*) and of the average instantaneous luminosity (4 plots on the *right*). Different event categories are shown for comparison. In clockwise order starting from *top left*: inclusive, 0 jets, 1 jets, at least 2 jets.

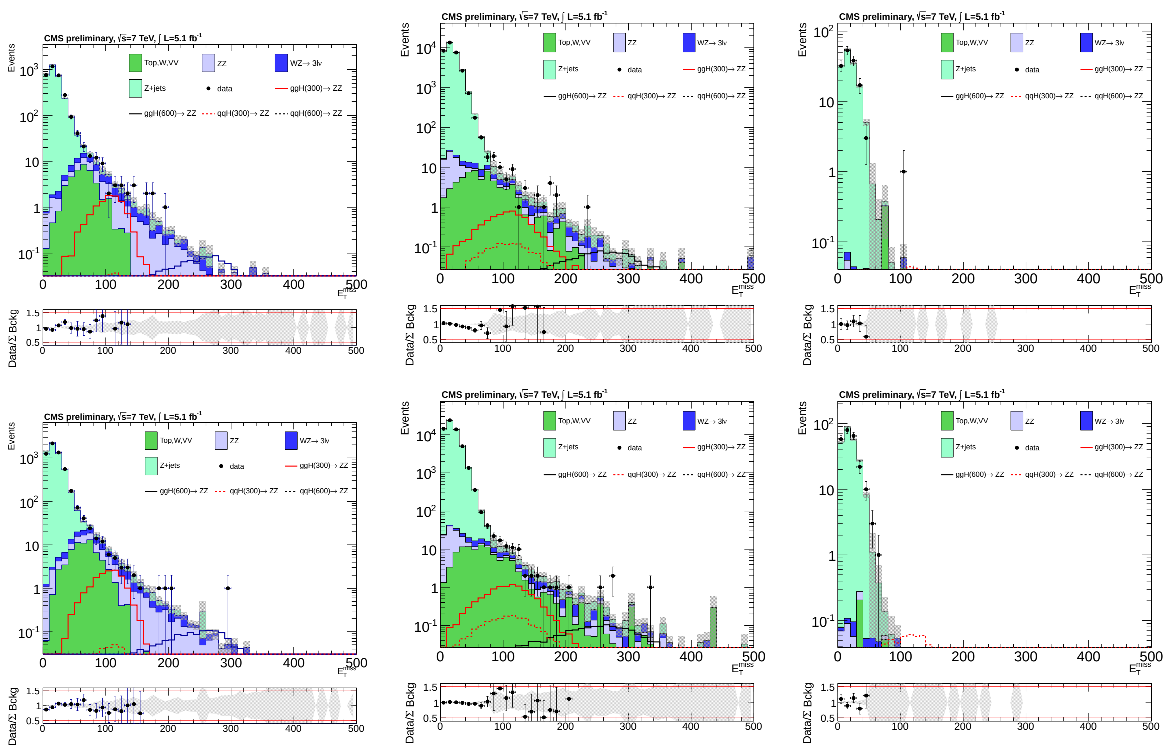


Figure B.7: From *left* to *right*: E_T^{miss} spectrum in di-electron (*top*) and di-muon (*bottom*) events with 0, at least 1 or VBF event categories.

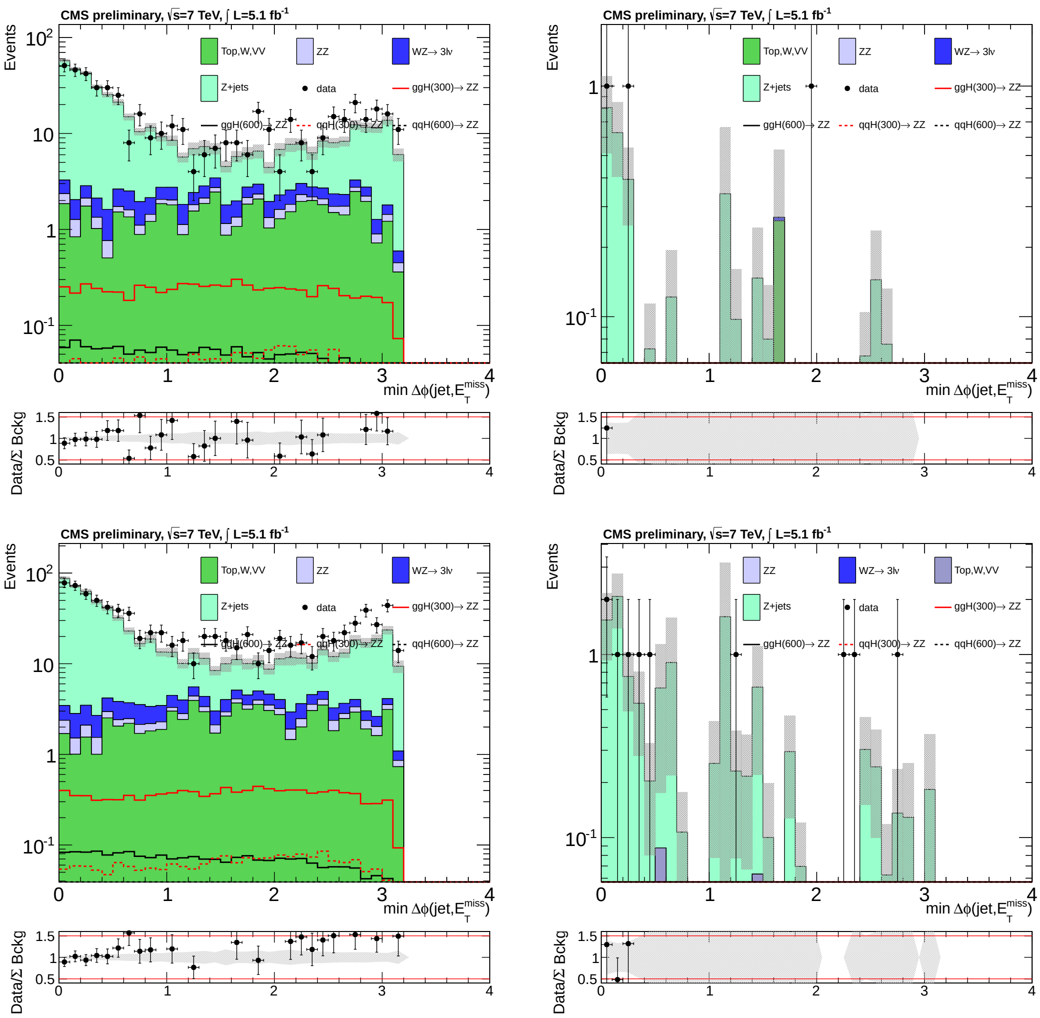


Figure B.8: From (*left*) to (*right*): min $\Delta\phi(\text{jet}, E_{\mathrm{T}}^{miss})$ variable in di-electron (*top*) and di-muon (*bottom*) events with $E_T^{miss} > 50$ GeV and with at least 1 or VBF event categories.

Figure B.9 shows the M_{T} distributions in the different event categories. A very good agreement is found overall.

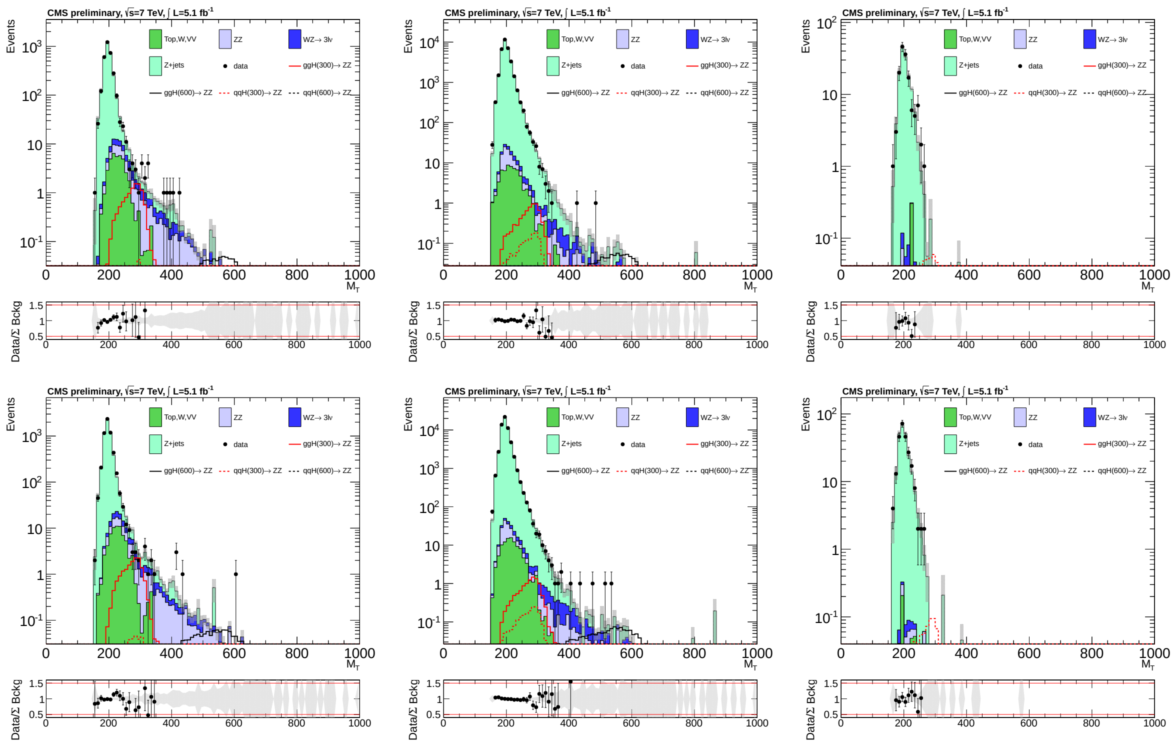


Figure B.9: From (*left*) to (*right*): M_{T} spectrum in di-electron (*top*) and di-muon (*bottom*) events with 0, at least 1 or VBF event categories.

Appendix C

C.1 Data-MC validation plots for BDT input variables for $\sqrt{s} = 8$ TeV

Figure C.1 shows the distribution for the min $\Delta\phi(\text{jet}, E_{\mathrm{T}}^{miss})$ variable for each ee and $\mu\mu$ channel after E_T^{miss} >50 GeV. Figure C.2 shows the M_{T} distributions in the different categories. A good agreement is found between data and background prediction. Figure C.3 shows the di-jet invariant mass M_{jj} for leading jets for ee and $\mu\mu$ channels. Similarly, Figure C.4 and Figure C.5 show $\Delta\eta$(jj) and $\Delta\phi$(jj) respectively. In general, a good agreement is found between data and background prediction.

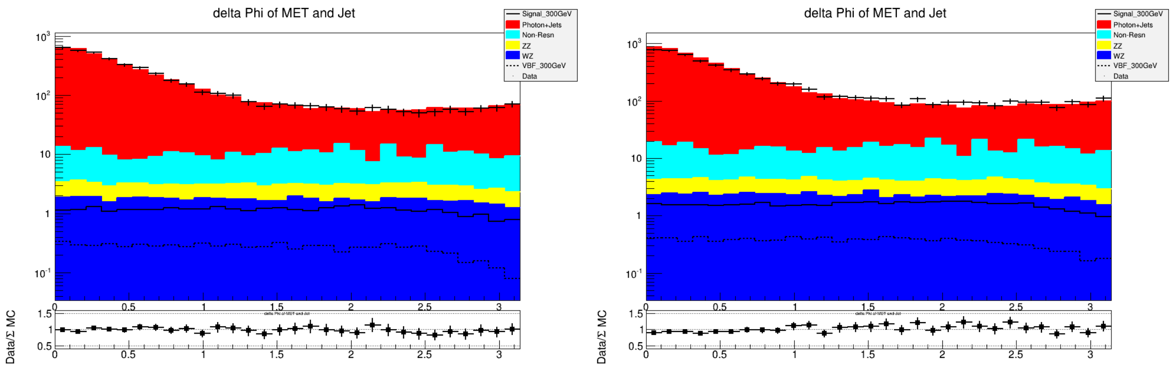


Figure C.1: min $\Delta\phi(\text{jet}, E_{\mathrm{T}}^{miss})$ variable for di-electron (*left*) and di-muon (*right*) channels with E_T^{miss} >50 GeV with at least 1 jets.

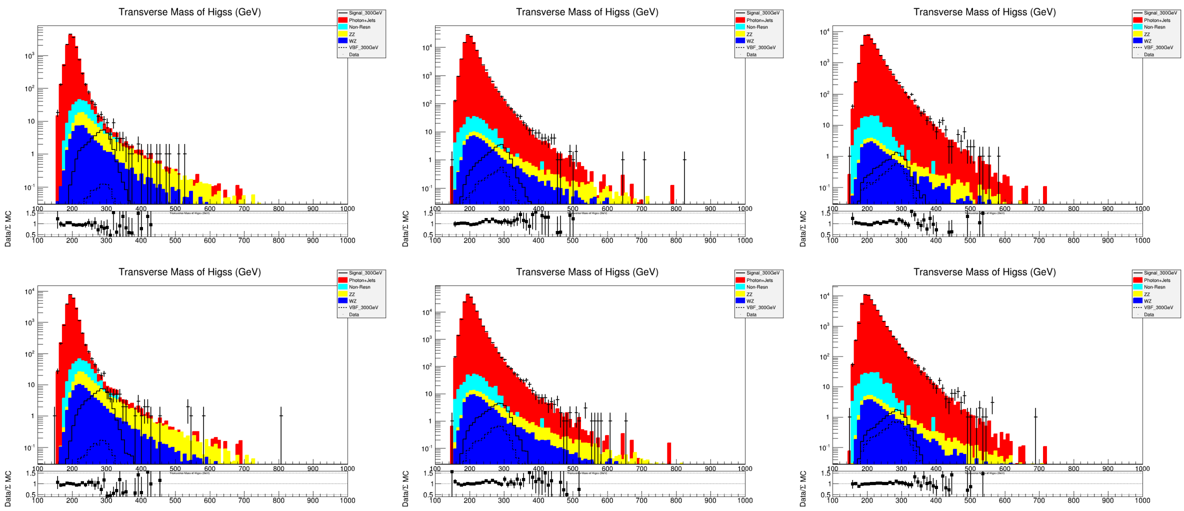


Figure C.2: From (*left* to *right*): M_{T} spectrum in di-electron (*top*) and di-muon (*bottom*) events with 0, 1 and =>2 jets.

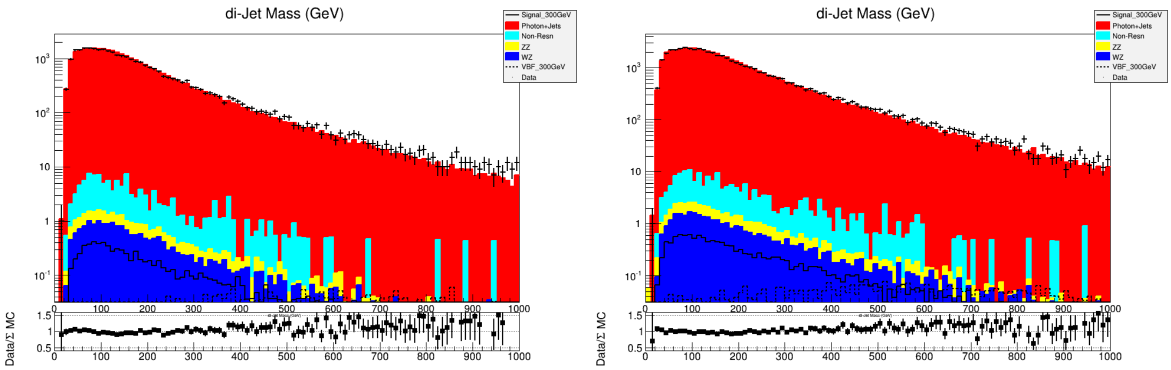


Figure C.3: Di-jet invariant mass distributions are shown for di-electron (di-muon) channel in the *left* (*right*).

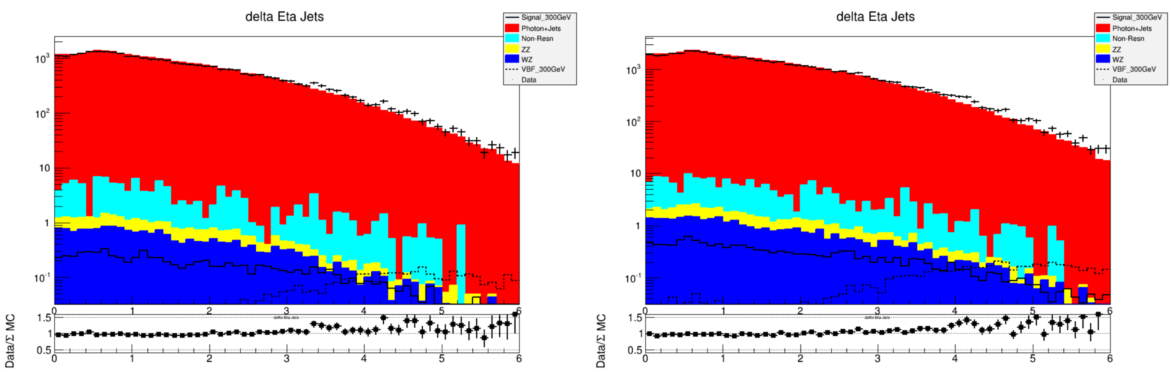


Figure C.4: Distributions of pseudorapidity gap ($\Delta\eta$) of two leading jets are shown for di-electron (di-muon) channel in the *left* (*right*).

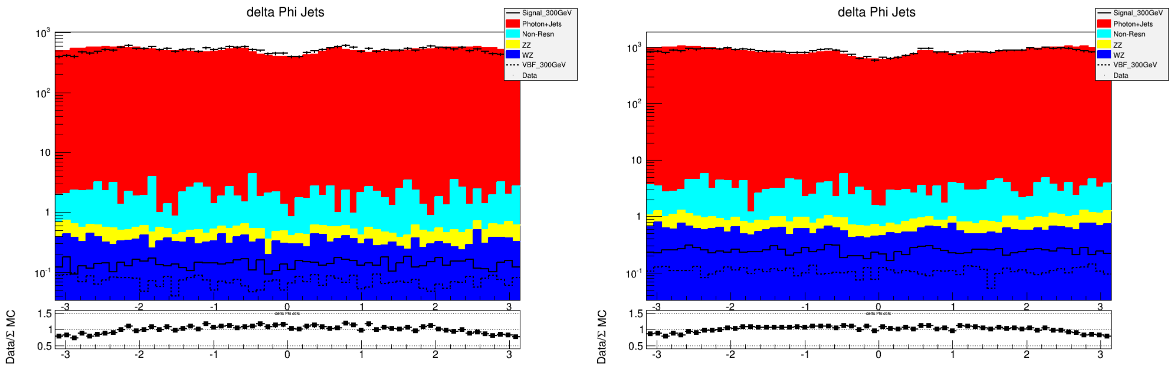


Figure C.5: Distributions of azimuthal gap ($\Delta\phi$) of two leading jets are shown for di-electron (di-muon) channel in the *left* (*right*).

C.2 BDT Input variable distributions after training

This section shows the input variable distributions for $\sqrt{s}$ = 8TeV for M_H =400 GeV, 800 GeV and 1000 GeV. For the masses 200 GeV and 600 GeV, distributions are already shown in Figures 5.6 to 5.9.

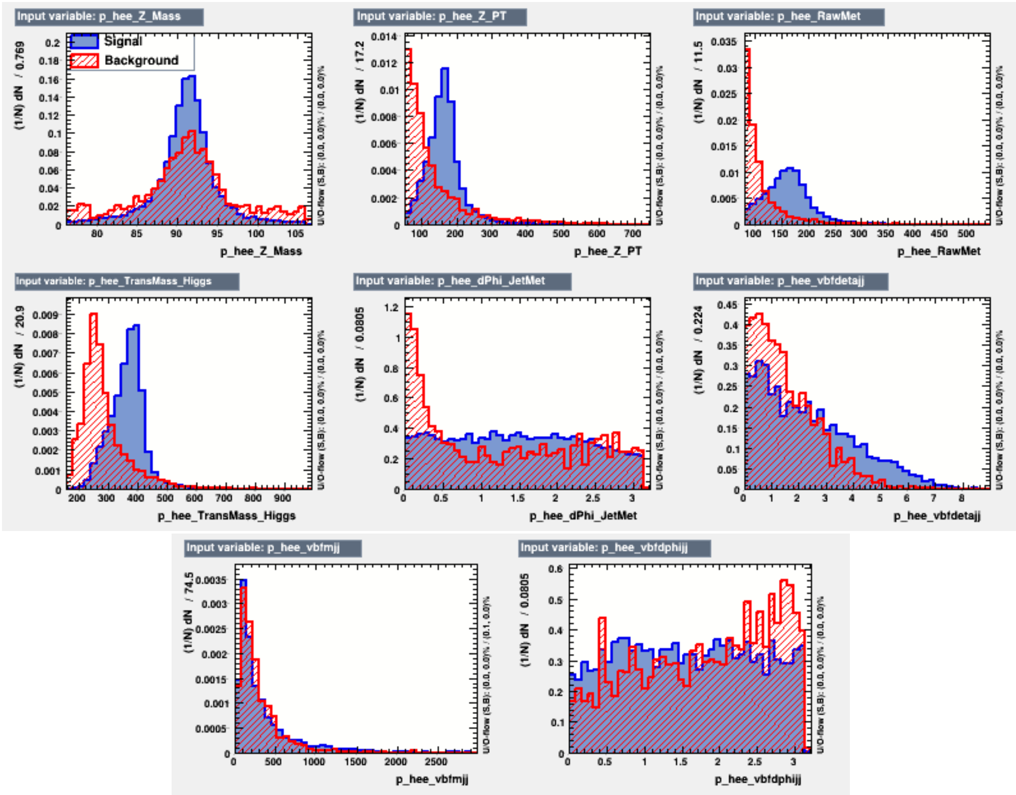


Figure C.6: BDT Input variables distribution for di-electron channel for M_H= 400 GeV after training. Blue is the signal and Red is the combined background.

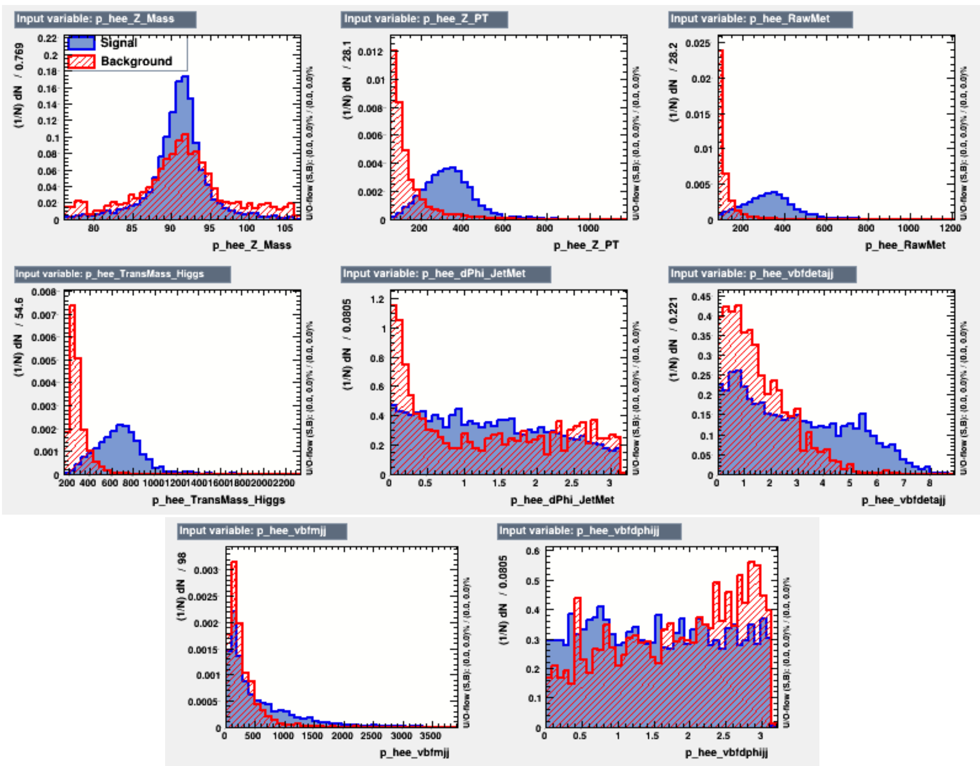


Figure C.7: BDT Input variables distribution for di-electron channel for M_H= 800 GeV after training. Blue is the signal and Red is the combined background.

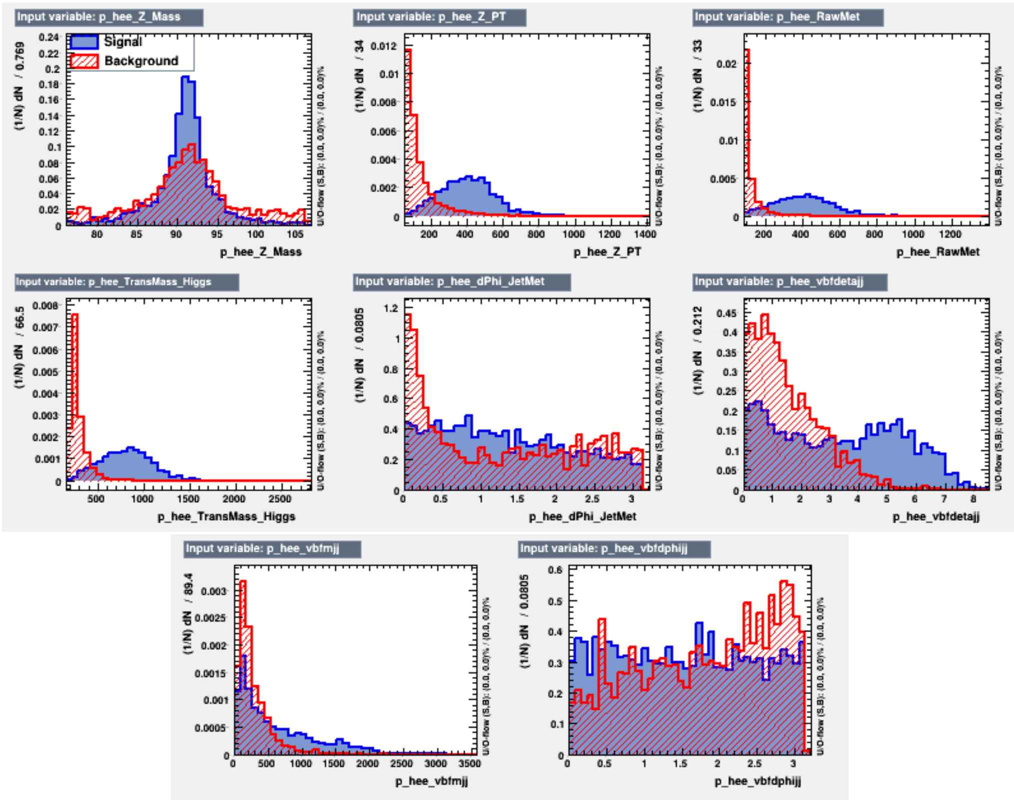


Figure C.8: BDT Input variables distribution for di-electron channel for M_H= 1000 GeV after training. Blue is the signal and Red is the combined background.

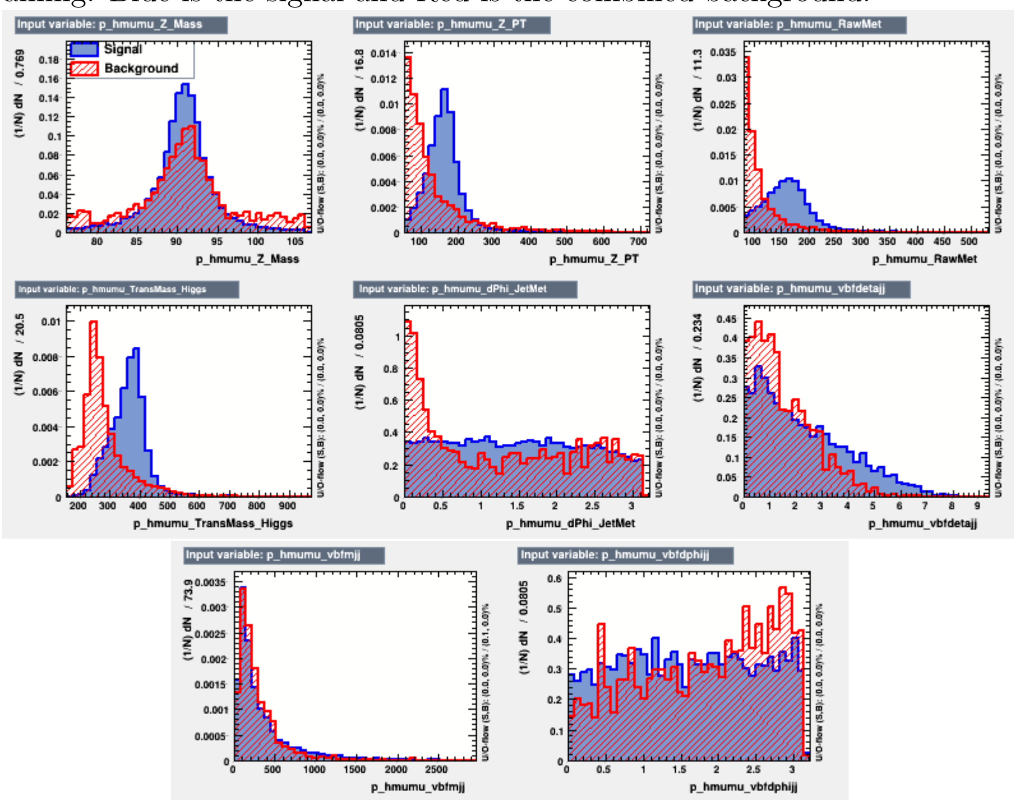


Figure C.9: BDT Input variables distribution for di-muon channel for M_H= 400 GeV after training. Blue is the signal and Red is the combined background.

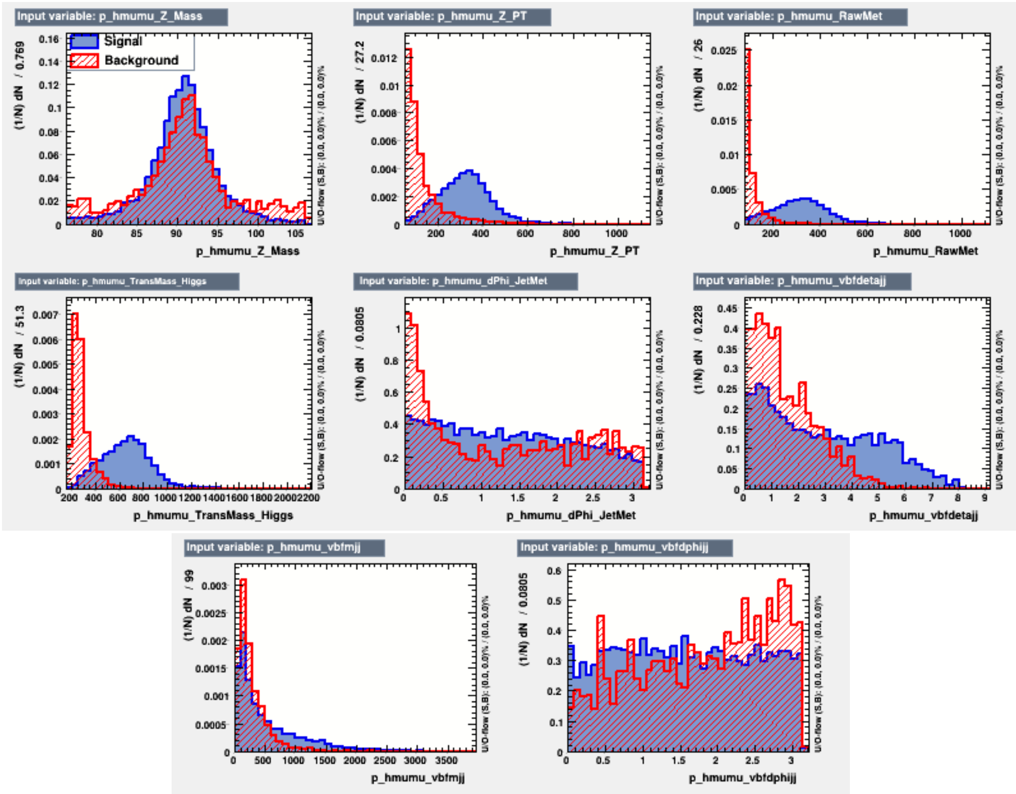


Figure C.10: BDT Input variables distribution for di-muon channel for M_H= 800 GeV after training. Blue is the signal and Red is the combined background.

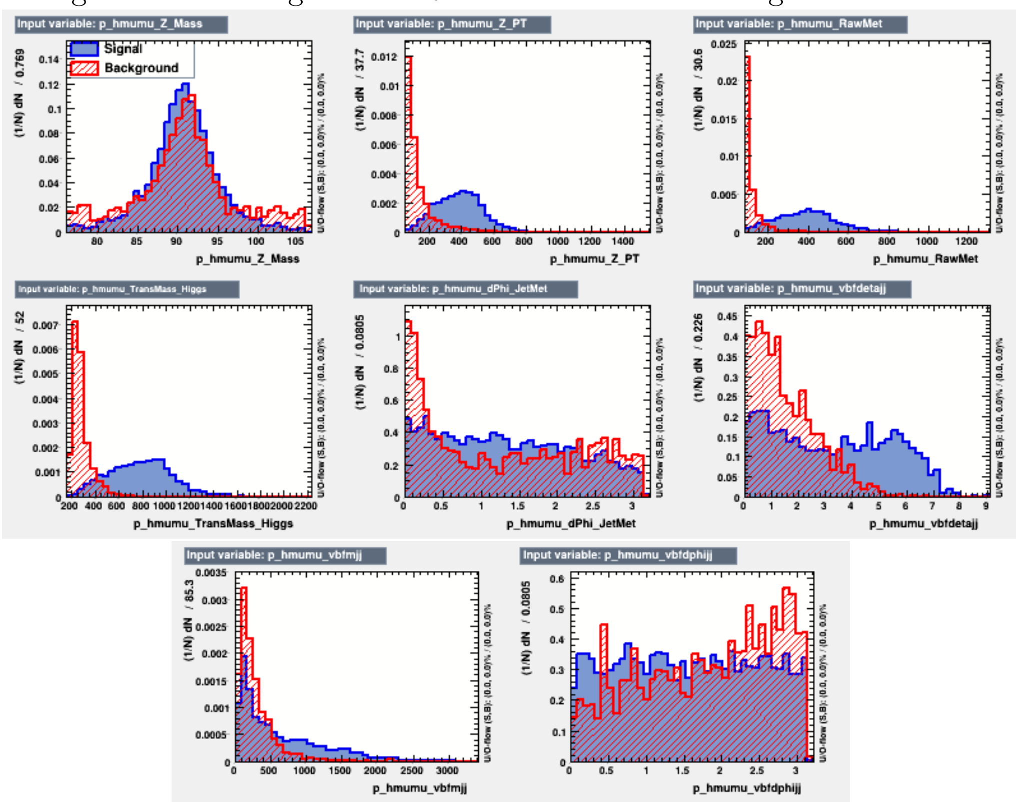


Figure C.11: BDT Input variables distribution for di-muon channel for M_H= 1000 GeV after training. Blue is the signal and Red is the combined background.

C.3 $\sqrt{s} = 7$ TeV BDT final output distributions

Figure C.12 and C.13 show the final BDT output distributions for ee channel at various M_H points and Figure C.14 and C.15 shows the same for $\mu\mu$ channel for $\sqrt{s} = 7$ TeV. It can be observed that the background composition changes as we go to higher masses. $Z + jets$ is the dominating background at lower masses and at higher masses diboson processes becomes dominating.

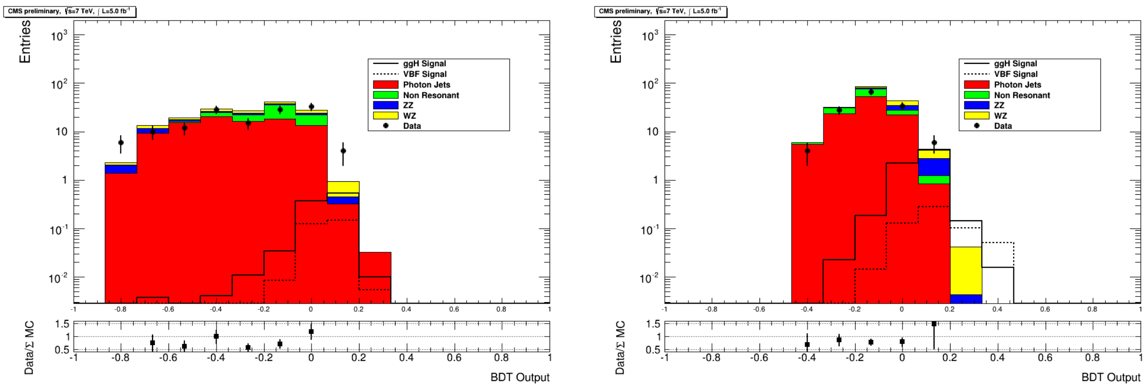


Figure C.12: Final BDT Output distributions for di-electron channel for $M_H = 200$ GeV (*left*) and for $M_H = 400$ GeV (*right*) for $\sqrt{s} = 7$ TeV.

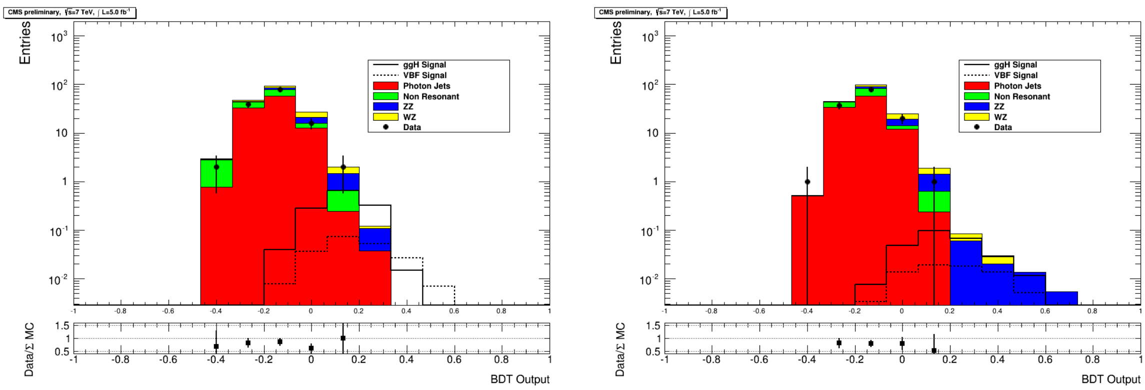


Figure C.13: Final BDT Output distributions for di-electron channel for M_H= 600 GeV (*left*) and for M_H= 800 GeV (*right*) for $\sqrt{s} = 7$ TeV.

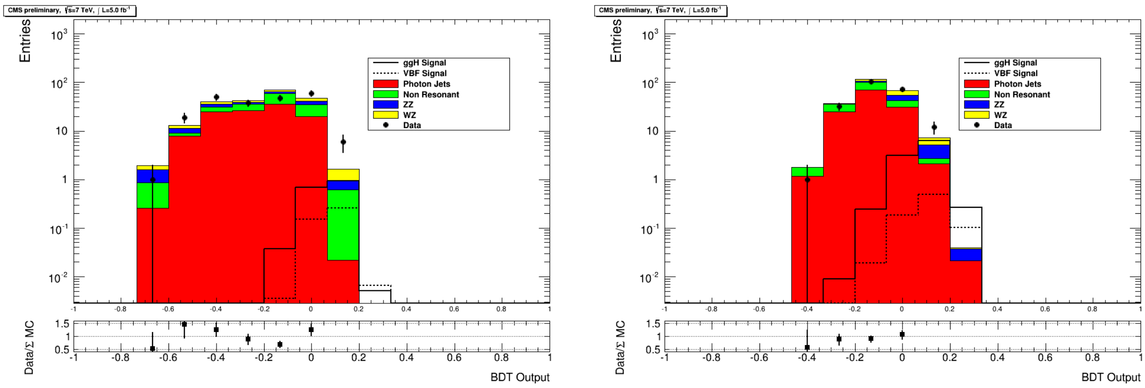


Figure C.14: Final BDT Output distributions for di-muon channel for M_H= 200 GeV (*left*) and for M_H= 400 GeV (*right*) for $\sqrt{s}$ = 7 TeV.

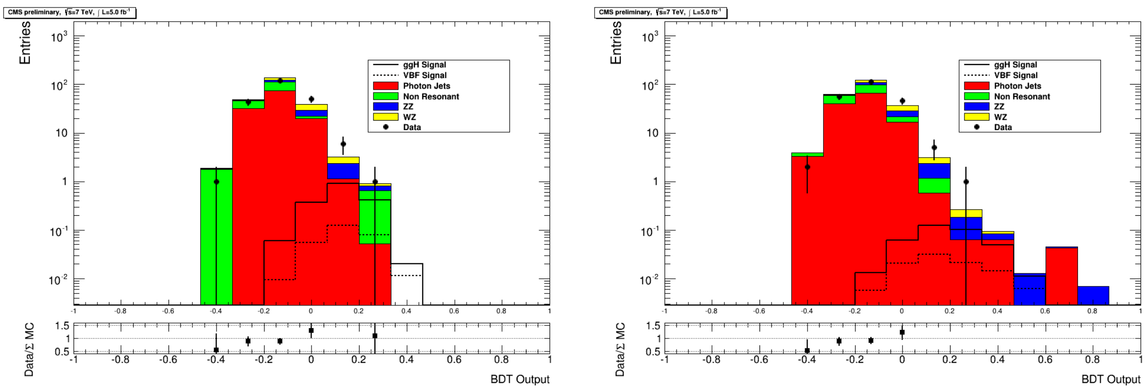


Figure C.15: Final BDT Output distributions for di-muon channel for M_H= 600 GeV (*left*) and for M_H= 800 GeV (*right*) for $\sqrt{s}$ = 7 TeV.