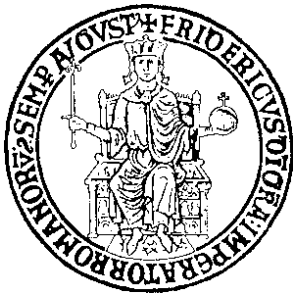
Università degli Studi di Napoli "Federico II"

Facoltà di Scienze Matematiche Fisiche e Naturali



Dottorato di Ricerca
in Fisica Fondamentale ed Applicata
XXV ciclo

Search for a Higgs boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ channel with the CMS experiment at the LHC

Relatori
Ch.mo Prof. C. Sciacca
Dott. L. Lista
Dott. F. Fabozzi

Candidata
Annapaola de Cosa

To my family

Contents

Introduction

The Standard Model (SM) of Particle Physics is the theory that describes three of the four fundamental interactions among elementary particles. Its great success is due to the confirmation of many of its theoretical predictions with measurements performed by experiments at Large Electron-Positron collider (LEP), Standford Linear Collider (SLC) and Tevatron. However, the experiments performed at these colliders did not confirm the existence of the Higgs boson, the particle predicted by the mechanism of spontaneous symmetry breaking, responsible of the generation of masses of all fundamental particles.

The Large Hadron Collider (LHC) has been designed and built with the purpose to explore energies never probed in a laboratory in order to search for the Higgs boson and for new physics beyond the SM, but 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 (c.m.e.) up to 14 TeV. It has been running since March 2010 at the c.m.e. $\sqrt{s} = 7$ TeV. In 2012 the energy of the two proton beams has been increased up to a c.m.e of 8 TeV. The Compact Muon Solenoid (CSM) is one of the two general-purpose experiments, together with ATLAS, built around one of the LHC interaction points.

This thesis describes the Ph.D. research activity that I have carried out over the last three years within the CMS collaboration. The main goal of my work has been the search for a Higgs boson decaying through two Z bosons into pairs of leptons and jets, $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q}\overline{\mathrm{q}}$, in the mass range [200, 600] GeV with data collected by the CMS experiment at the LHC. The $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q}\overline{\mathrm{q}}$ channel is one of the most relevant channel for searching the Higgs boson in the mass range $m_\mathrm{H} > 2m_\mathrm{Z}$, due to its high branching ratio. An integrated luminosity of 4.9 fb^{-1} collision data produced at $\sqrt{s} = 7$ TeV and 5.1 fb^{-1} produced at $\sqrt{s} = 8$ TeV have been analyzed.

During the first part of my Ph.D activity I worked on the development of the official CMS software framework for data analysis, PAT (Physics Analysis Toolkit). In particular, I have been one of the main developer of the

CMS ConfigEditor, a Graphical User Interface (GUI) for browsing and editing the files used to build the analysis workflow and I implemented the major features of a general tool for producing a reduced data format ("ntuples") perfectly integrated within the framework of the CMS experiment. Thanks to this work, I received the Achievement Award in the Offline project of the CMS experiment at CERN. These tools are widely used within the CMS collaboration in many analyses. I used all these tools to build the software infrastructure needed to perform the analysis of the $H \to ZZ \to \ell^+\ell^-q\overline{q}$ channel.

I focused my analysis work on the development of an effective analysis strategy aimed to reconstruct and select Higgs boson candidates from a pair of oppositely charged high-p_T leptons and a pair of hard jets. A challenging aspect of this analysis is the definition of a strategy devoted to the control of a huge background and enhance the sensitivity of the analysis itself. A powerful handle against background is the resonant feature of both the dilepton and the dijet systems as well as of the diboson system. Further background suppression is achieved considering spin correlations in signal decay channel, which reflect in the angular distribution of decay products for the signal. This correlation is absent in background.

The analysis strategy is based on the classification of signal-like events in three exclusive channels based on the jet-flavour tagging information. The combination of the three classes of events, which are characterized by a different signal-background composition, allows to improve the sensitivity of the analysis. In particular, the class of events with two jets originating from b quarks is the one less contaminated by background. The statistical analysis is performed on the reconstructed diboson invariant mass spectrum. I was responsible for selecting the final list of events used to set upper limits on the Higgs boson production cross section. As contact person between the $H \to ZZ \to \ell^+\ell^-q\overline{q}$ group and the CMS Higgs Combination Group, I was also responsible for the integration of the results of the $H \to ZZ \to \ell^+\ell^-q\overline{q}$ channel into the combination of all CMS Higgs searches.

Chapter 1 gives a brief introduction to the Standard Model with particular attention to the Higgs sector. Chapter 2 provides an overview of the LHC and of the CMS experiment along with the reconstruction of particle candidates in the detector. The analysis strategy is described in details in Chapter 3 including reconstruction, selection and background estimation from data, but also systematic uncertainties affecting the signal extraction. Chapter 4, finally, describes the signal modeling and the statistical method used to determine upper limits on the Higgs boson production cross section, along with the results of the analysis performed on the 7 TeV and 8 TeV data collected by CMS in 2011 and 2012. The last part of the chapter is

dedicated to the observation of a new boson in the decay channels $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to 4\ell$ with the CMS detector.

Over the past three years I carried out my research activity mainly at CERN firstly as Doctoral student and then as CERN-INFN associate. The work presented in this thesis has been produced thanks to the support of INFN (Istituto Nazionale di Fisica Nucleare) and CERN.

Chapter 1

The Standard Model

The Standard Model (SM) of Particle Physics, introduced in the late 1960's, provides so far the best description of the phenomenology of high energy particle interactions. The mechanism of spontaneous electroweak symmetry breaking (EWSB), already proposed in the mid-sixties, was coupled to the SM in order to complete the picture of the theory, explaining the generation of the weak vector boson masses by means of coupling to the Higgs boson.

The high-precision measurement carried out in the pre-LHC era at LEP, SLC and Tevatron [1, 2] tested the Standard Model confirming the correctness of its description of the electroweak and strong interactions at the probed energies. However, no evidence of Higgs boson was found, this induced the need for a new collider.

One of the major goal of high-energy collider physics is to search for the Higgs boson in all the allowed range of its mass. The Large Hadron Collider (LHC) at CERN has been designed to probe energies up to the TeV scale and give a clear scientific response regarding the existence of the Higgs boson.

In Section 1.1 an overview of the Standard Model and of the mechanism of the EWSB will be given. Section 1.2 summarizes the theoretical and experimental constraints on Higgs boson properties. A description of the Higgs production mechanisms at hadron colliders and its decay modes will be provided in Section 1.3.

1.1 The Standard Model

Elementary particles interact by means of four fundamental interactions: gravity, weak interaction, electromagnetic interaction and strong interaction.

Fundamental matter particles are fermions (particles with half spin integer) of two types – six *leptons* and six *quarks*. Both leptons and quarks are

grouped in three doublets each – associated to an isospin quantum number – and organized into three families:

- Lepton doublets are constituted each of a charged particle interacting both electromagnetically and weakly – the electron e, the muon μ and the tau τ – and by a neutral particle interacting only weakly – the electron neutrino ν_e, the muon neutrino ν_μ, the tau neutrino ν_τ.

- Quark doublets are composed each of a particle of charge $+\frac{2}{3}$ – up (u), charm (c), top (t) – and a particle of charge $-\frac{1}{3}$ – down (d), strange (s), bottom (b). The quarks interact via all three interactions: strong, electromagnetic and weak. The quarks are the costituents of hadrons, and are confined into them by the strong interaction.

Table 1.1: Leptons and quarks generations.

	Families			**Charge**
Leptons	$\binom{\nu_e}{e}$	$\binom{\nu_\mu}{\mu}$	$\binom{\nu_\tau}{\tau}$	0 -1
Quarks	$\binom{u}{d}$	$\binom{c}{s}$	$\binom{t}{b}$	$+\frac{2}{3}$ $-\frac{1}{3}$

The SM describes three of the four fundamental interactions (electromagnetic, weak and strong) in terms of gauge theory.

Particle interactions arise from a *gauge principle*, i.e. they are a consequence of the invariance of the theory under a local gauge transformation.

Quantum Electrodynamics (QED), which describes electromagnetic interactions, derives from the imposition of gauge invariance under the symmetry group $U(1)_{em}$ involving the introduction of a Lagrangian term that couples the fermion field to the electromagnetic field. A generalization of this gauge invariance also underlies the theories of weak interactions, generated by the symmetry group $SU(2)_L$ and strong interactions (the *Quantum Chromodynamics*, QCD), generated by the symmetry group $SU(3)_c$.

Quarks and leptons interact in terms of exchange of quanta (*gauge vector bosons*, integer spin) associated to the particular type of force. These field quanta are identified as the photon γ, the $W^{\pm}$, the Z and the *gluons*:

- The photon is the mediator of electromagnetic interaction coupling to

charged particles by a running coupling constant[1] α, which shows an increase with energy involved. The photon does not carry charge and does not couple to neutral particles, like neutrinos.

- The W^+, W^- and Z mediate the weak interactions, but unlike the photon, the weak interaction is of short range and therefore the mediators are massive. It is noteworthy that while the photon doesn't carry charge, weak current are both charged and neutral.

- The strong force is mediated by gluons associated to eight generators of the group SU(3). They couple with quarks and gluons. Unlike photon, the gluon carries strong charge, the *color*, and gluon self-coupling is possible . The running coupling constant α_s for strong interactions decreases at high energy.

The electromagnetic and weak interactions were unified successfully under a theory developed by Glashow[3], Weinberg[4] and Salam[5]: the Electroweak (EW) Theory generated by the symmetry group $SU(2)_L \otimes U(1)_Y$.

The presence of massive particles implies the introduction of a mass term in the Lagrangian which causes breaking of gauge invariance. In the context of this theory, such issue is overcome by including the mechanism of spontaneous symmetry breaking in the electroweak theory itself.

The mechanism was proposed in 1964 by Higgs, Brout, Englert, Guralnik, Hagen and Kibble [6, 7, 8, 9, 10]. It predicts the existence of a new scalar particle, the *Higgs boson*. The coupling to the Higgs field generates the vector boson masses.

The first experimental confirmation of the Standard Model of electroweak interactions was the observation of neutral weak currents predicted by the theory. The discovery was made in 1973 at CERN using the Gargamelle bubble chamber exposed to a neutrino beam[11] It was followed by a direct confirmation occurred with the realization of a proton-antiproton collider at CERN with the discover of the W and Z bosons[12], the mediators of weak force, by the collaborations UA1 and UA2 in 1983.

---

[1] α is the electromagnetic coupling constant which runs with the momentum transfer due to a phenomenon called *polarization of the vacuum*. This running constant it is defined by:
$\alpha(Q^2) = \frac{\alpha}{[1-(\frac{\alpha}{3\pi})ln(\frac{Q^2}{Am^2})]}$, for $Q^2 \gg m^2$ and with Q^2=-q^2, where q is the transferred momentum and A=exp$\frac{5}{3}$.

1.1.1 The Glashow–Weinberg–Salam theory

Electromagnetic and weak interactions were described together in a unified way by Glashow basing on the symmetry group $SU(2)_L \otimes U(1)_Y$[3]: the *Electroweack theory*. Building on Glashow's work, Weinberg[4] and Salam[5], independently, introduced the Higgs mechanism of the spontaneous symmetry breaking in the theory by adding to the Lagrangian a scalar potential that generates the vector boson and fermion masses in a gauge invariant way.

Looking at the leptonic transitions associated with the weak charged current, a particular structure of the lepton pairs emerges: they are regarded as *doublets* under the weak $SU(2)_L$ group, and the quarks show a similar structure. Since the generators of fundamental representation of $SU(2)_L$ are the Pauli matrices, it is possible to adopt the same formalism as for the angular momentum. In the weak $SU(2)_L$ group a *weak isospin* quantum number T is associated to the leptonic fields:

$$T = \frac{1}{2}: \quad \begin{cases} T_3 = +1/2 \\ T_3 = -1/2 \end{cases} \quad L_e = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad L_\mu = \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad L_\tau = \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L , \tag{1.1}$$

where $e_L = \frac{1}{2}(1-\gamma_5)e$ is the left-handed component of the leptonic field. The same holds for μ and τ. At present, there is no evidence for any weak interactions coupling to right-handed field components, and it is a basic assumption of the electroweak theory that all 'R' components are singlets under $SU(2)_L$ group[2]:

$$T = 0: \qquad e_R \qquad \mu_R \qquad \tau_R \qquad \nu_{e\ R} \qquad \nu_{\mu\ R} \qquad \nu_{\tau\ R} . \tag{1.2}$$

For the quark fields the doublets are:

$$T = \frac{1}{2}: \qquad \begin{cases} T_3 = +1/2 \\ T_3 = -1/2 \end{cases} \qquad \begin{pmatrix} u_L \\ d'_L \end{pmatrix} \qquad \begin{pmatrix} c_L \\ s'_L \end{pmatrix} \qquad \begin{pmatrix} t_L \\ b'_L \end{pmatrix} \tag{1.3}$$

It is noteworthy that d', s', b' are superpositions of strong interaction eigenstates d, s, b given by the CKM matrix (Cabibbo 1963[13], Kobayashi and Maskawa 1973[14]).

Right quark components are also singlets as well as lepton ones:

$$T = 0: \qquad u_R \qquad d_R \qquad s_R \qquad c_R \qquad b_R \qquad t_R . \tag{1.4}$$

[2]The right-handed neutrinos are not involved in the interaction, since their quantum numbers are null.

The standard EW Lagrangian can be split into two parts: one involving fermions and gauge bosons, $\mathcal{L}_{\text{symm}}$, and one involving Higgs boson couplings, $\mathcal{L}_{\text{Higgs}}$:

$$\mathcal{L} = \mathcal{L}_{\text{symm}} + \mathcal{L}_{\text{Higgs}} \ . \tag{1.5}$$

Gauging an abelian theory consists of extending the global invariance of the Lagrangian density under $U(1)$ transformation group to a local invariance. According to Yang and Mills[15], this procedure can be applied to every continuous transformation group. The invariance of the EW Lagrangian under local $SU(2)_L \otimes U(1)_Y$ transformation is recovered by replacing the ordinary derivative by the covariant derivative through the introduction of four gauge fields: W_μ^i $(i = 1, 2, 3)$ associated to $SU(2)_L$, and B_μ associated to $U(1)_Y$. Therefore in the Dirac Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\partial\!\!\!/ - m)\psi \tag{1.6}$$

$\partial\!\!\!/$ has, therefore, to be replace by the covariant derivative $D\!\!\!/$:

$$\mathcal{L} = \bar{\psi}(iD\!\!\!/ - m)\psi \ . \tag{1.7}$$

where the covariant derivative is:

$$D_\mu = \partial_\mu + ig\frac{\boldsymbol{\tau}}{2}\mathbf{W}_\mu + i\frac{g'}{2}Y_W B_\mu \ . \tag{1.8}$$

The Dirac field can be expressed in terms of chirality eigenstate components[3] ψ_L and ψ_R. Right and left-handed components have different quantum numbers, hence, the mass term, which couples them in the Lagrangian, involves a breaking of gauge invariance:

$$\bar{\psi}(\mathrm{i}D\!\!\!/ - m)\psi = \bar{\psi}_L \mathrm{i}D\!\!\!/\psi_L + \bar{\psi}_R \mathrm{i}D\!\!\!/\psi_R - m(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L) \ . \tag{1.9}$$

Neglecting the mass term of fermions and bosons the local invariance of the theory is preserved, and right and left components are treated independently. Fermion and boson masses will be introduced in a gauge invariant way through the Higgs mechanism.

[3]The left-handed component ψ_L is given by the operation of the projection operator P_L on the field, and ψ_R by the projector P_R:

$$P_{\mathrm{L}} = \left(\frac{1-\gamma^5}{2}\right) \qquad P_{\mathrm{R}} = \left(\frac{1+\gamma^5}{2}\right) \ .$$

The fermionic Lagrangian describes the interaction of massless fermionic fields among them and with the gauge fields:

$$\mathcal{L}_{\text{fermionic}} = \sum_f \bar{\psi} i \not{D} \psi \ . \tag{1.10}$$

$\mathcal{L}_{\text{fermionic}}$ results to be invariant under rotations in weak isospin space:

$$SU(2)_L : \qquad \begin{cases} L_f \to e^{-ig\frac{\boldsymbol{\tau}}{2}\cdot\boldsymbol{\alpha}(x)} L_f \\ R_f \to R_f \end{cases} \ . \tag{1.11}$$

The right-handed fermionic components do not interact with gauge fields introduced within the Lagrangian by the covariant derivative, but they are sensitive to a local $U(1)_Y$ transformation:

$$U(1)_Y : \quad \begin{cases} L_f \to e^{-ig'\frac{Y}{2}\cdot\boldsymbol{\beta}(x)} L_f \\ R_f \to e^{-ig'\frac{Y}{2}\cdot\boldsymbol{\beta}(x)} R_f \end{cases} \ . \tag{1.12}$$

The abelian field $U(1)_Y$ is associated to a *weak ipercharge* Y, as well as the non abelian $SU(2)_L$ is associated to the *weak isospin* T. The relation between T and Y was extablished by the Gell-Mann-Nishijima relation:

$$Q = T_3 + \frac{Y}{2} \ , \tag{1.13}$$

where Q is the electric charge in units of e. Two different coupling constant g and g', were introduced correspondly respectively to $SU(2)_L$ and $U(1)_Y$, since the transformations under the two groups are independent.

In order to have the Lagrangian (1.7) invariant, the transformations of fermionic fields (1.11) and (1.12) must be coupled to the gauge fields transformations:

$$W^i_\mu(x) \to W'^i_\mu(x) = W^i_\mu(x) - \partial_\mu \alpha^i(x) - g\epsilon^{ijk}\alpha_j(x) W_{k\mu}(x) \ , \tag{1.14}$$

$$B_\mu(x) \to B'_\mu(x) = B_\mu(x) - \partial_\mu \beta(x) \ . \tag{1.15}$$

The transformation for the gauge fields W^i_μ is more complex than that for the B_μ field, due to the non-abelian properties of the $SU(2)_L$ group. The additional term in the (1.14) is responsible for the self-interactions of $SU(2)_L$ gauge fields and also for the inclusion of the last term in strength field tensor (1.32).

In correspondence to this invariance, four currents are preserved for $SU(2)_L$ and one for $U(1)_Y$: two charged currents and two neutral currents. Coming

back to the covariant derivative, it can be written as below:

$$iD_\mu = i\partial_\mu - \underbrace{g(W_\mu^1 T^1 + W_\mu^2 T^2)}_{\text{charged current contribution}} - \underbrace{gT^3 W_\mu^3 - g'\frac{Y}{2}B_\mu}_{\text{neutral current contribution}} \quad . \tag{1.16}$$

It is clear that R components couple also to neutral currents. The second term describes the electromagnetic field interaction A_μ and a new field Z_μ, both neutrals. It's convenient to perform an orthogonal transformation:

$$W_3^\mu = \sin\theta_W A^\mu + \cos\theta_W Z^\mu \ , \tag{1.17a}$$
$$B^\mu = \cos\theta_W A^\mu - \sin\theta_W Z^\mu \ , \tag{1.17b}$$

to produce the physical vector fields for $\mathrm{W}^\pm$, Z and photon:

$$Z^\mu = -\sin\theta_W B^\mu + \cos\theta_W W_3^\mu \ , \tag{1.18}$$

$$A^\mu = \cos\theta_W B^\mu + \sin\theta_W W_3^\mu \ , \tag{1.19}$$

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(\mathrm{W}_\mu^1 \pm i\mathrm{W}_\mu^2) \ , \tag{1.20}$$

where the mixing angle θ_W is known as the Weinberg angle, defined as:

$$\cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}} \qquad \sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}} \tag{1.21}$$

The interaction term of the Lagrangian due to the coupling with W_μ^3 and B_μ fields can be rewritten in terms of physical field Z_μ and A_μ according to (1.18) and (1.19) through the Weinberg angle. According to the Gell–Mann–Nishijima relation, the electromagnetic current can be expressed as the sum of weak isospin current and weak hypercharge current:

$$J_\mu^{em} = J_\mu^W + \frac{J_\mu^Y}{2} \ . \tag{1.22}$$

The gauge field A_μ is the electromagnetic field coupled to electric charges through the term $J_{em}^\mu A_\mu$. This yields a relation among the electromagnetic charge, the two coupling constants, g and g', and the Weinberg angle:

$$e = g\sin\theta_W = g'\cos\theta_W \ . \tag{1.23}$$

The fermionic Lagrangian term is now complete; it splits in a free Lagrangian term and an interaction part with gauge vector bosons:

$$\mathcal{L}_{\text{fermionic}} = \mathcal{L}_{\text{free}} + \mathcal{L}_{\text{int}} \ , \tag{1.24}$$

where $\mathcal{L}_{\text{free}}$ is given by:

$$\mathcal{L}_{\text{free}} = \sum_f \bar{\psi} i \not{\partial} \psi \ . \tag{1.25}$$

The interaction term is:

$$\mathcal{L}_{\text{int}} = \mathcal{L}_{\text{w}}^{\text{charged}} + \mathcal{L}_{\text{w}}^{\text{neutral}} + \mathcal{L}_{\text{int}}^{\text{e.m}} \ , \tag{1.26}$$

where the first term is:

$$\mathcal{L}_{\text{w}}^{\text{charged}} = -\frac{g}{2\sqrt{2}}[J^{\mu+} W_{\mu}^{-} + J^{\mu-} W_{\mu}^{+}] \ , \tag{1.27}$$

the second one is given by:

$$\mathcal{L}_{\text{w}}^{\text{neutral}} = -\frac{g}{\cos\theta_W}[J_3^{\mu} - \sin^2\theta_W J_{em}^{\mu}] Z_{\mu} \ , \tag{1.28}$$

and the last term can be expressed as:

$$\mathcal{L}_{\text{int}}^{\text{e.m}} = -e J_{em}^{\mu} A_{\mu} \ . \tag{1.29}$$

In order to complete the dynamics it is needed to build the kinematic terms for vector bosons. A strength field tensor, $B_{\mu\nu}$, related to the gauge field B_{μ} is defined as:

$$B_{\mu\nu} = \partial_{\mu} B_{\nu} - \partial_{\nu} B_{\mu} \ . \tag{1.30}$$

In analogy to electromagnetic case, the kinematic term is:

$$\mathcal{L}_B = -\frac{1}{4} B_{\mu\nu} B^{\mu\nu} \ . \tag{1.31}$$

The strength field tensor associated to gauge fields $\mathbf{W}_{\mu}$ is given by the equation:

$$F_{\mu\nu}^{\alpha} = \partial_{\mu} W_{\nu}^{\alpha} - \partial_{\nu} W_{\mu}^{\alpha} - g\epsilon^{\alpha\beta\gamma} W_{\mu}^{\beta} W_{\nu}^{\gamma} \ , \tag{1.32}$$

and the correspondly Lagrangian density is:

$$\mathcal{L}_W = -\frac{1}{4} \mathbf{F}_{\mu\nu} \mathbf{F}^{\mu\nu} \ . \tag{1.33}$$

The $SU(2)_L \otimes U(1)_Y$ gauge invariant Lagrangian is:

$$\mathcal{L}_{\text{symm}} = \sum_f \bar{\psi} \text{i} \not{D} \psi - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} \mathbf{F}_{\mu\nu} \mathbf{F}^{\mu\nu} \tag{1.34}$$

where the sum is done over all fundamental fermions, f.

1.1.2 Spontaneously broken local $SU(2)_L \otimes U(1)_Y$ symmetry

In order to preserve the Lagrangian invariance, fermionic and bosonic masses have been neglected, but the theory has to describe the fact that $W^{\pm}$ and Z have masses around 100 GeVIt is therefore necessary to introduce a new term that coherently represents the masses of the particles, and at the same time, preserves the gauge principle. The spontaneous symmetry breaking mechanism allows to generate such term by postulating the existence of a new scalar field doublet. The Higgs boson field is written as an $SU(2)_L$ doublet with two scalar components:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \\ \frac{1}{\sqrt{2}}(\phi_3 + i\phi_4) \end{pmatrix} \tag{1.35}$$

The Lagrangian for such boson includes a potential which is responsible of the symmetry breaking mechanism :

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \Phi)^\dagger D_\mu \Phi - V(\Phi) = (D^\mu \Phi)^\dagger D_\mu \Phi - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 \tag{1.36}$$

By requiring that $\mu^2 < 0$ and $\lambda > 0$, the non trivial solution, the minimum of the potential is not unique anymore, but it is located on a continuous ring on a complex plane, as seen in Figure 1.1. An easier form of scalar potential can be chosen applying a gauge isotopic spin transformation:

$$\Phi = \begin{pmatrix} 0 \\ \phi(x) \end{pmatrix} \tag{1.37}$$

with ϕ(x) real. It can be seen that the degree of freedom in the Higgs field that are removed by this transformation can be absorbed into three longitudinal degrees of freedom of the new massive bosons.

More possible vacuum expectation values (v.e.v.) are possible choises. The choice suggested by Weinberg was:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \tag{1.38}$$

where

$$v = \sqrt{-\mu^2/\lambda}\ . \tag{1.39}$$

Considering fluctuations around the minimum value (1.38):

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}\ , \tag{1.40}$$

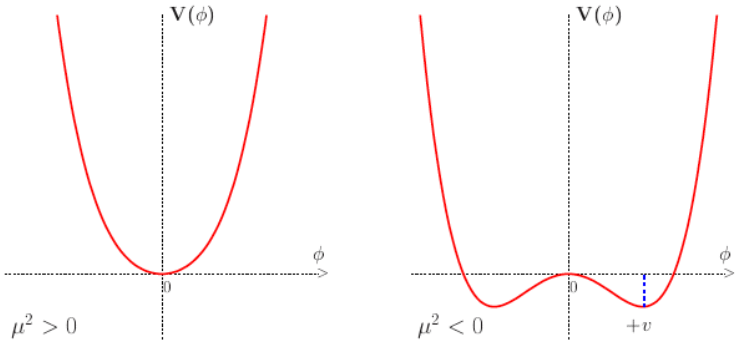


Figure 1.1: The potential V of the scalar field ϕ in the case of $\mu^2 > 0$, on the left. On the right the Higgs potential described from the last two terms of equation (1.36) with $\mu^2 < 0$ and $\lambda > 0$. The potential minimum is not unique, in the latter case, but there is an infinite number of minimum points located on a continuous ring.

$v/\sqrt{2}$ is the Higgs vacuum expectation value and $H(x)$ is the fluctuation around the minimum. After the breaking of the symmetry three bosons acquire mass ($W^{\pm}$, Z), and one remains massless(γ).

The covariant derivative (1.8) acts on Φ, so one obtains:

$$\begin{aligned}(D^{\mu}\Phi)^{\dagger}D_{\mu}\Phi = &\frac{1}{2}\partial^{\mu}H\partial_{\mu}H + \\ &\frac{1}{8}(v+H)^2 g^2 (W^{1\mu} + iW^{2\mu})(W^1_{\mu} - iW^2_{\mu}) + \\ &\frac{1}{8}(v+H)^2 (g'B^{\mu} - gW^{3\mu})(g'B_{\mu} - gW^3_{\mu})\ .\end{aligned} \tag{1.41}$$

The Lagrangian of the sector consisting of the gauge fields and the Higgs is:

$$\begin{aligned}\mathcal{L}_{G\Phi} &= \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Higgs}} \\ &= -\frac{1}{4}W^{\mu\nu}_i W^i_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + (D^{\mu}\Phi)^{\dagger}D_{\mu}\Phi - \mu^2\Phi^{\dagger}\Phi - \lambda(\Phi^{\dagger}\Phi)^2\ .\end{aligned} \tag{1.42}$$

Replacing W^{μ}_3 and B^{μ} by the physical fields Z^{μ} and A^{μ}, from (1.18) and from (1.19), and neglecting the quadratic terms, (1.42) can be rewritten as:

$$\mathcal{L}_{G\Phi} = \frac{1}{2}\partial^\mu H \partial_\mu H - \mu^2 H^2 \tag{1.43}$$

$$-\frac{1}{4}(\partial_\mu W_{1\nu} - \partial_\nu W_{1\mu})(\partial_\mu W_1^\nu - \partial_\nu W_1^\mu) + \frac{1}{8}g^2 v^2 W_{1\mu} W_1^\mu \tag{1.44}$$

$$-\frac{1}{4}(\partial_\mu W_{2\nu} - \partial_\nu W_{2\mu})(\partial_\mu W_2^\nu - \partial_\nu W_2^\mu) + \frac{1}{8}g^2 v^2 W_{2\mu} W_2^\mu \tag{1.45}$$

$$-\frac{1}{4}(\partial_\mu Z_\nu - \partial_\nu Z_\mu)(\partial_\mu Z^\nu - \partial_\nu Z^\mu) + \frac{1}{8}(g^2 + g'^2)v^2 Z_\mu Z^\mu \tag{1.46}$$

$$-\frac{1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial_\mu A^\nu - \partial_\nu A^\mu) \tag{1.47}$$

where from (1.43) the mass of the Higgs boson is

$$m_{\mathrm{H}} = \sqrt{2}\mu = \sqrt{2\lambda}v \ . \tag{1.48}$$

The second term in (1.44) and (1.45) has exactly the form of a mass term for the W_1 and W_2 fields, since the W mass can be defined as:

$$m_{\mathrm{W}} = \frac{1}{2}gv \ , \tag{1.49}$$

while (1.46) gives the mass of the Z as

$$m_{\mathrm{Z}} = \frac{1}{2}v\sqrt{g^2 + g'^2} = \frac{m_{\mathrm{W}}}{\cos\theta_W} \tag{1.50}$$

From (1.47) it is clear that the photon remains massless, the symmetry $U(1)_{em}$ remains unbroken under the gauge subgroup generated by the electric charge.

Equations (1.49) and (1.50) bind the boson masses to θ_W and the parameter v, which is related to the Fermi constant G_F by the relation[4]:

$$v^2 = \frac{1}{\sqrt{2}G_F} \backsimeq (246\ \mathrm{GeV})^2 \ . \tag{1.51}$$

A theoretical prediction was given about Z and W boson masses as function of the three parameters $\alpha = e^2/4\pi$, G_F and $\sin^2\theta_W$. The latter was initially measured by neutrino-electron scattering and, later, with higher precision at the e^+ e^- collider LEP. Theoretical predictions of boson masses revealed a perfect agreement with experimental data of the SppS collider and LEP collaborations at CERN[12][16], and also with CDF and D0 at Fermilab[17], confirming the predictive power of the GWS model.

[4]The relation between v and G_F comes out imposing the equality between transition amplitude according to Fermi theory and the form of Weinberg Salam model. Hence v is determined directly by the Fermi constant.

1.1.3 Fermionic masses

In order to complete the EW Lagrangian, another term has to be added: $\mathcal{L}_{\text{Yukawa}}$, that introduces fermionic masses. As explained before, a mass term such as that in (1.9) breaks the gauge invariance, therefore it has to be removed, and fermion masses, such as boson masses, must be included through the alternative mechanism seen in the previous paragraph.

It is interesting to see how it is possible to produce a gauge invariant mass term by introducing a Yukawa coupling between fermions and the Higgs field, the same doublet (1.40) that gave raise to the Z and W boson mass:

$$\mathcal{L}_{\text{Yukawa}} = -\sum_f g_f(\bar{\psi}_L^f \Phi \psi_R^f + \bar{\psi}_R^f \Phi^\dagger \psi_L^f) \ . \tag{1.52}$$

Consider the electron, for example, and hypothesize such a coupling between the electron-type $SU(2)_L$ doublet in (1.1), the weak electron singlet in (1.4) and the Higgs doublet in Φ (1.40). This leads to the Yukawa term for the electron:

$$\mathcal{L}^e_{\text{Yukawa}} = -g_e(\bar{L}_e \Phi R_e + \bar{R}_e \Phi^\dagger L_e) \tag{1.53}$$

The two $SU(2)_L$ doublets L_e and Φ are dotted together to form an $SU(2)_L$ scalar, which multiplies the $SU(2)_L$ scalar right-handed component. In this way, (1.53) is invariant under $SU(2)_L \otimes U(1)_Y$ transformation preserving the gauge invariance of such term. Inserting the vacuum value of the field Φ and considering the fluctuations about the v.e.v. (1.40) into (1.53), a Dirac mass term is obtained together with the coupling between the Higgs boson and the electron which is proportional to the electron mass:

$$\mathcal{L}^e_{\text{Yukawa}} = -g_e \frac{v}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) - g_e \frac{H}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) \ . \tag{1.54}$$

This term is a Dirac mass term as in (1.9), and it allows to identify the electron mass:

$$m\mathrm{e}^- = g_e \frac{v}{\sqrt{2}} \ . \tag{1.55}$$

The mass term for the electron, and in general for all fermions, are proportional to the Yukawa coupling. The second term, the Higgs boson coupling to fermions, is proportional to the fermions' mass:

$$g_{\text{H ee}} = i\frac{m_e}{v} \ . \tag{1.56}$$

This approach be applied to the *up* component of the $SU(2)_L$, i.e. for the $T_W^3 = +\frac{1}{2}$ component such as neutrinos and quarks u, c, and t. The coupling

with the charge conjugate of Φ has to be considered as well:

$$\Phi_C = i\tau_2\Phi^* = \begin{pmatrix} \bar{\Phi}^0 \\ -\Phi^- \end{pmatrix} \ , \tag{1.57}$$

and the corresponding vacuum expectation value:

$$\Phi_C = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H(x) \\ 0 \end{pmatrix} \ . \tag{1.58}$$

Taking this into account, the Lagrangian with quark mass terms can be written as:

$$\begin{aligned} \mathcal{L}^{\text{quarks}}_{\text{Yukawa}} &= \frac{g_q}{\sqrt{2}} - \left[(u_{i,L}, d_{i,L}) \begin{pmatrix} v + H \\ 0 \end{pmatrix} u_{i,R} + u_{i,R}(v + H, 0) \begin{pmatrix} u_{i,L} \\ d_{i,L} \end{pmatrix} \right] \\ &= \frac{g_q}{\sqrt{2}}(v + H)(u_{i,L}u_{i,R} + u_{i,R}u_{i,L}) \end{aligned} \ , \tag{1.59}$$

where $u_i = (u, c, t)$, $d_i = (d, s, b)$. The mass terms have the same form as (1.55), but with a different coupling constant, depending on the quark type, or in general fermion type.

It is clear that the Higgs mechanism solves the problem of the non-invariance of the Lagrangian in the presence of a mass term with the introduction of the Higgs field by means of the spontaneously broken symmetry mechanism. In this way it is possible to give mass to the vector bosons Z and W, but also to all fermions[5]. Nevertheless, the Yukawa coupling doesn't come from a gauge principle and, hence, is not possible to predict the Higgs mass, that is a free parameter of the theory, since the constant λ is undefined. Also the fermion masses that depend on Higgs-fermion Yukawa couplings g_f are undefined, although they are well determined from experimental data. This means that the Yukawa coupling is a phenomenological model and not a predictive theory.

1.2 The Higgs boson

The Higgs mass value depends not only on v (equation 1.48), the minimum value of Higgs potential, but also on the field self-coupling constant λ. The latter has to be determined experimentally, therefore a prediction of the Higgs mass is not possible. Nevertheless, bounds on m_{H} are placed by theoretical arguments as well as by indirect and direct searches.

[5]In this thesis work, the terms relative to neutrino masses and their coupling to the Higgs field are not considered.

1.2.1 Theoretical constraints on the Higgs boson mass

Theoretical constraints on m_{H} [2] emerge from the requirement of unitary and evaluations about perturbativity of the Higgs self-coupling, but also from the requirement of vacuum state stability. These bounds depend on an energy cut-off Λ_C, above which new physics is expected.

Like the other coupling constants of renormalizable theories, also λ runs with energy. Looking at the one-loop radiative corrections to the Higgs boson quartic coupling (Figure 1.2) and considering only the Higgs contribution, $\lambda(Q^2)$ shows a logarithmic dependence on the squared energy scale Q^2:

$$\lambda(Q^2) = \frac{\lambda(v^2)}{1 - \frac{3}{4\pi^2}\lambda(v^2)\log\frac{Q^2}{v^2}} \tag{1.60}$$

The energy cut-off Λ_C, below which the self-coupling λ remains finite, depends on m_{H} and is:

$$\Lambda_C = v\,\exp\Big(\frac{2\pi^2}{3\lambda}\Big) = v\,\exp\Big(\frac{4\pi^2 v^2}{3m_{\mathrm{H}}^2}\Big) \tag{1.61}$$

This translates into a upper bound on m_{H}: the upper curve in Figure 1.3.

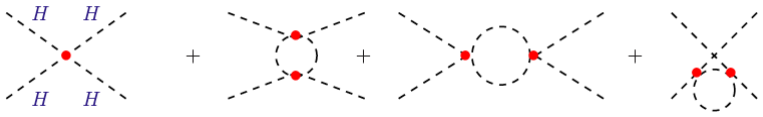


Figure 1.2: Higgs boson quartic coupling and one loop radiative corrections.

For small values of λ, the top quark contribution could determine a negative value of $\lambda(Q^2)$. In order to have vacuum state stability, it is necessary to keep a positive value of $\lambda(Q^2)$; therefore, m_{H} is constrained to be larger than a certain value which is proportional to squared top mass[18] and depends on the cut-off scale value Λ_C. This constraint corresponds to the lower curve in Figure 1.3.

These limits imply that if the SM is valid up to the Grand Unification scale, i.e. the energy scale at which a unification of strong, electromagnetic and weak interactions is expected, $\Lambda_{\mathrm{GUT}} \backsim 10^{16} GeV$, the Higgs mass should belong to the range:

$$130\ \mathrm{GeV}/c^2 \lesssim m_{\mathrm{H}} \lesssim 180\ \mathrm{GeV}/c^2 \tag{1.62}$$

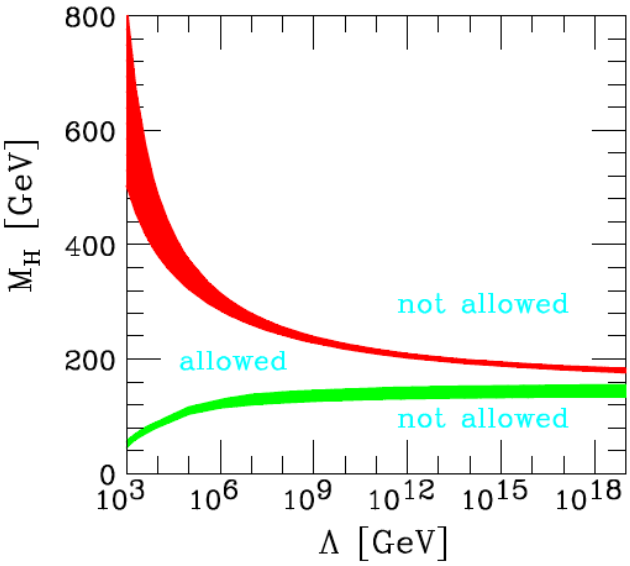


Figure 1.3: The upper limit and the lower limit to the Higgs boson mass as a function of the cut-off scale Λ_C. The cut-off scale Λ is the scale at which the SM is no longer able to describe the known physics, and new physics is therefore expected.

Although the theory cannot predict the exact value of Higgs boson mass, it imposes some constraints and, therefore, limits the range in which it lies. It must be taken into account that these limits depend on the energy scale Λ_C above which the Standard Model has to be replaced by an other theory describing a new phenomenology.

1.2.2 Experimental constraints on Higgs boson mass

Bounds on the Higgs mass are also provided by direct searches of Higgs bosons and, indirectly, by precision measurements carried out during the last decades.

Precision electroweak measurements are sensitive to radiative corrections which depend logarithmically on m_{H}. This allows to estimate a confidence interval for m_{H} from a combined fit of the electroweak parameters. The SM is assumed to be the correct theory, the Higgs boson mass is left as a free parameter.

For instance, the radiatively-corrected value for the mass of $\mathrm{W}^{\pm}$ can be expressed as:

$$m_{\mathrm{W}}^2 = \frac{\pi\alpha/\sqrt{2}G_F}{\sin^2\theta_W(1-\Delta r)} \tag{1.63}$$

where Δr includes the radiative corrections related to α, G_F, m_{Z}. Δr depends on the running coupling constant $\alpha(m_{\mathrm{Z}})$, quadratically on top quark mass (due to the loop $t-b$ correction) and logarithmically in $\mathrm{m_H}$. Experiments at Tevatron measured the top mass[18]. Therefore, radiative corrections to the $\mathrm{W}^{\pm}$ mass, and also to the other parameter contemplated by the SM, depends only on m_{H}.

Measurements from the four LEP collaborations, ALEPH, DELPHI, L3 and OPAL and from the SLD experiment at the Stanford Linear Collider, are collected and a global fit is performed on them [16].

Taking into account all the precision electroweak data one obtains, by minimizing $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ as function of m_{H} (Figure 1.4), the value of the most likely SM Higgs boson mass to be:

$$m_{\mathrm{H}} = 94^{+29}_{-24}\ \mathrm{GeV}/c^2 \tag{1.64}$$

with an upper limit at the 95 % Confidence Level on the SM of m_{H}:

$$m_{\mathrm{H}} < 152\ \mathrm{GeV}/c^2 \tag{1.65}$$

A top quark mass of 173 ± 0.9 Gev [18] and a $\mathrm{W}^{\pm}$ boson mass of 80.385 ± 0.015 GeV [16] were used.

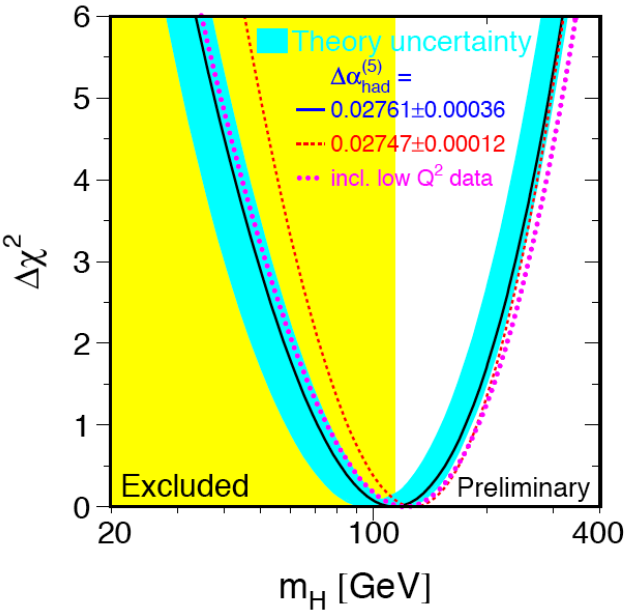


Figure 1.4: The χ^2 of the fit to the electro-weak precision data as function of m_{H}. The continuos line results from high energy precision measurement. The blue band takes into account theoretical uncertainties, caused by neglecting higher order corrections.

At LEP the direct search for the Higgs bosons at centre of mass (c.m.) energies up to 209 GeV has been focused on the analysis of the Higgs-stralhung production process, where the electron-positron pair annihilates into a virtual Z which then splits into a Higgs particle and a real Z. This process has the highest cross section.

The combination of results from all four collaborations at LEP of the direct search for the Higgs boson production carried to a lower limit at 95% of confidence level to the Higgs mass [19]:

$$m_{\mathrm{H}} > 114 \text{ GeV}/c^2 \tag{1.66}$$

Results on direct searches for a SM Higgs Boson in $\mathrm{p\overline{p}}$ collisions at the Fermilab Tevatron accelerator, with 10.0 fb^{-1} of data collected at $\sqrt{s}$ = 1.96 TeV, presented on July 2012, exclude at 95% C.L. the presence of a SM-like Higgs boson in the mass range m_{H} between 100 and 103 GeV/c^2 and between 147 and 180 GeV/c^2 [20] (Figure 1.5). CDF and DØ experiments performed a search for the Higgs boson in the following channels:

$$\mathrm{H} \to \mathrm{b\overline{b}} \ , \ \mathrm{H} \to \mathrm{W^+W^-} \ , \ \mathrm{H} \to \mathrm{ZZ} \ , \ \mathrm{H} \to \tau^-\tau^+ \ , \ \mathrm{H} \to \gamma\gamma \ .$$

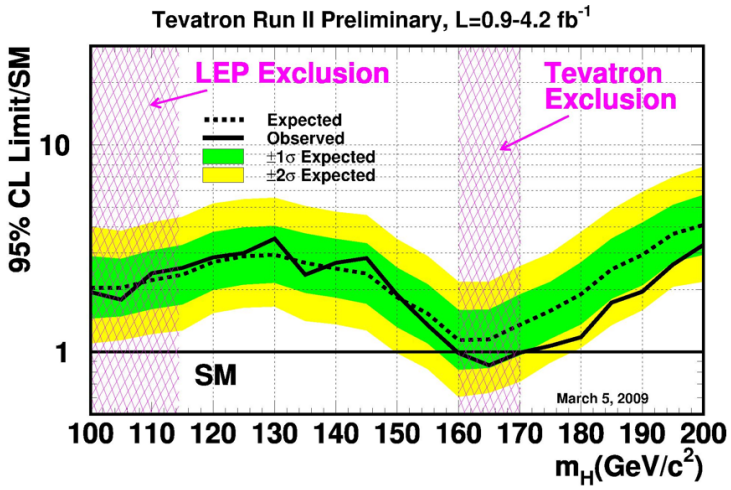


Figure 1.5: Observed and expected 95 % C.L. upper limits on the ratios of the Higgs production cross section and the SM cross section, as a function of the Higgs boson mass for the combined CDF and DØ analyses. Both experiments performed dedicated searches in different channels.

1.3 The Higgs boson at the LHC

Searches for Higgs bosons are ongoing at the Large Hadron Collider. The pp collider has been conceived to probe energies up to the TeV scale and perform searches with much higher sensitivity than the Tevatron. The LHC will be detailed in the following chapter. Here the expected scenario for the production and detection of a Higgs boson at the LHC will be discussed.

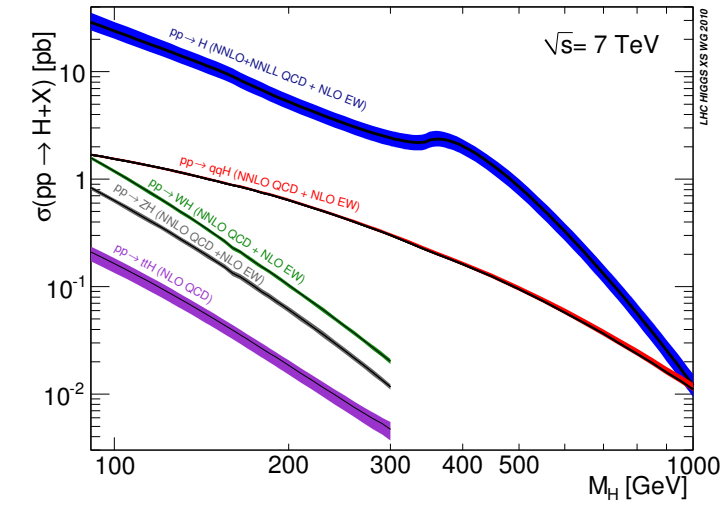
1.3.1 Higgs boson production

The most important SM Higgs boson production mechanisms at the LHC are:

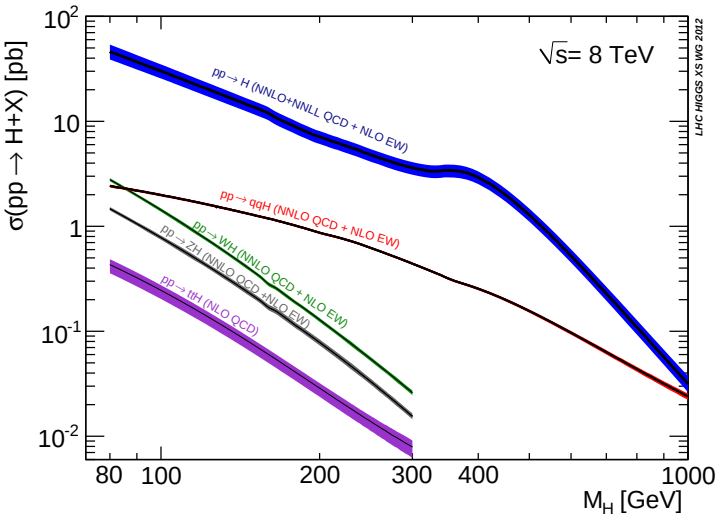
- the gluon-gluon fusion ($gg \to H$)
- the vector-boson fusion ($qq' \to qq'H$)
- the vector-boson associated production, *Higgsstralhung* ($q\bar{q} \to WH/ZH$)
- the top-quark associated production ($q\bar{q}/gg \to t\bar{t}H$)

The cross sections of these processes depend on the Higgs boson mass as well as on the pp collision centre-of-mass (c.m.) energy. In general all production cross sections decrease with increasing the Higgs boson mass. Raising the c.m. energy reflects in an increase in Higgs production cross sections over all the mass range. Figure 1.7 shows the dependence of Higgs production cross sections on m_H and compares the $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV scenarios.

At LHC, the gluon fusion is the dominating Higgs production mechanism over the entire Higgs mass spectrum; a gg pair annihilates into a Higgs boson through a quark loop. The diagram is shown in Figure 1.7(a). The main contribution is given by the top loop because of the strong coupling of Higgs boson to the top quark. The value of the cross section increases of a factor ~ 2 considering nex-to-leading orders (NLO) with respect to the leading order (LO) cross section. This production process is very interesting not only because of its relatively larger cross section in the entire mass range, but also because it is very sensitive to an hypothetical fourth generation of quarks, being the Higgs coupling proportional to the fermion mass. The second contribution to the Higgs boson production is the vector-boson fusion mechanism (Figure 1.7(b)): two vector bosons, radiated off quarks, couple to produce a Higgs boson. The cross section for this process is about one order of magnitude lower than for the gluon fusion one in a wide range of m_H values, even if they become almost comparable for very high Higgs boson mass values. This process has a very clear signature being characterized by



(a)



(b)

Figure 1.6: Higgs production cross sections as a function of the Higgs boson mass for $\sqrt{s}$ =7 TeV and $\sqrt{s}$ =8TeV scenarios [21, 22]

the hadronization of two spectator quarks with high invariant mass in the forward region. In the case of *Higgsstralhung* (Figure 1.7(c)) and of Higgs production associated to a $t\bar{t}$ pair (Figure 1.7(c)), the Higgs boson is radiated off a gauge boson or a $t\bar{t}$ pair respectively. These processes have cross sections several orders of magnitude lower than the previous two. Nevertheless, their signature is very specific and provides a clear tag of the Higgs production.

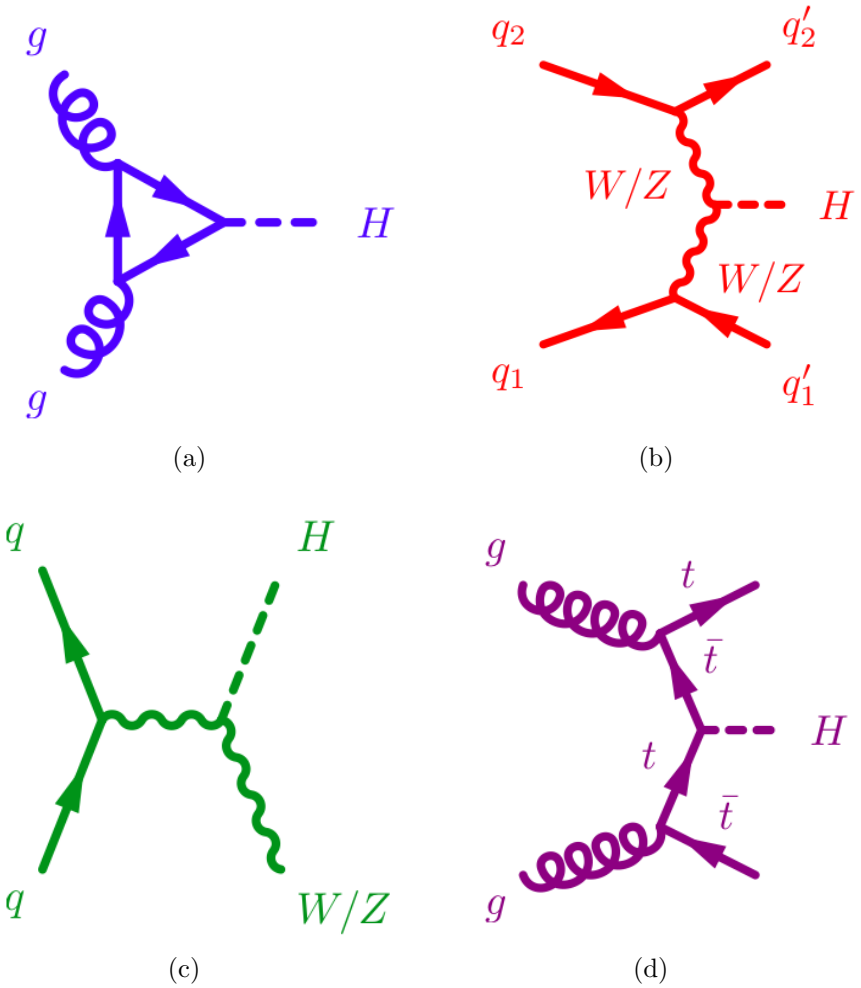


Figure 1.7: Higgs boson production mechanisms at tree level in proton-proton collisions: (a) gluon-gluon fusion; (b) VV fusion; (c) W and Z associated production (*Higgsstralhung*); (d) $t\bar{t}$ associated production.

1.3.2 Higgs boson decay

The Higgs search at LHC is performed in a range of m_H values that goes from 115 GeV up to 1 TeV. Figure 1.8 illustrates the inclusive SM Higgs production cross section times the branching ratio for the main channels that can be explored at the LHC. The branching ratios (BRs) of the decay channels vary significantly with m_H and the mass range can be divided into two region: low and high mass.

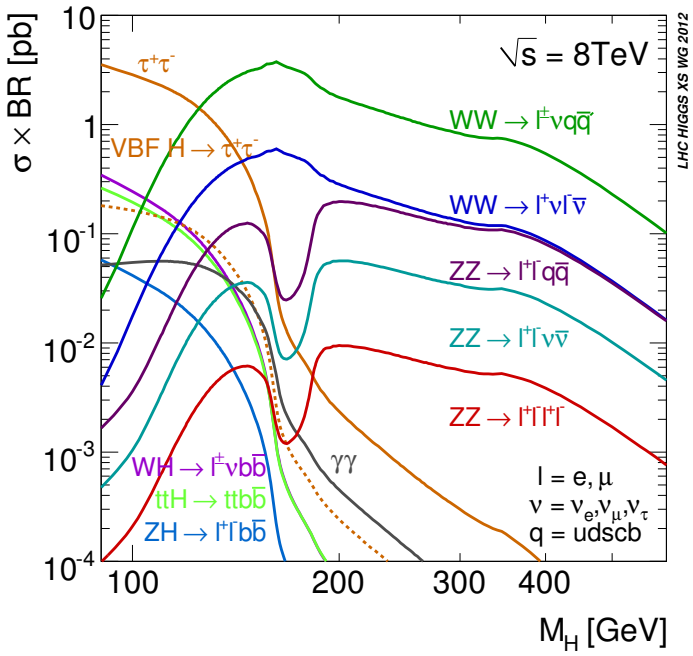


Figure 1.8: Inclusive Standard Model Higgs boson production cross section times branching ratio at $\sqrt{s} = 8$ TeV.

In the low mass region, up to $m_H \sim 150$ GeV, the channel whith 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. An inclusive search for $H \to b\bar{b}$ is not feasible. In spite of its low cross section the channel with a Higgs boson produced via Higgsstralhung 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. Background sources for this channel are photons coming from $q\overline{q} \to \gamma\gamma$, $\mathrm{Z} \to \mathrm{e^-e^+}$ and from jets. The sensitivity of this channel has a strong dependence on invariant mass resolution dependent on the detector performances.

As m_{H} rises the Higgs decays into $\mathrm{WW}^{(*)}$ and into $\mathrm{ZZ}^{(*)}$ open up. The $\mathrm{H} \to \mathrm{W^+W^-}$ cross section peaks when the production of a pair of on-shell W bosons is kinematically allowed, i.e. around $\sim 2m\mathrm{W}$. The $\mathrm{H} \to \mathrm{ZZ}$ cross section has a maximum around $\sim 2\mathrm{m}_Z$, when the production of two real Z bosons becomes possible.

The most interesting channels in this region are: $\mathrm{H} \to \mathrm{WW}^{(*)} \to 2\ell 2\nu$ and $\mathrm{H} \to \mathrm{ZZ}^{(*)} \to 4\ell$. 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 of the channel $\mathrm{H} \to \mathrm{WW}^{(*)} \to 2\ell 2\nu$ is two isolated high p_{T} leptons with a small opening angle in the transverse plane, undetected neutrinos and no signal of quark hadronization in the central region of the detector. The dominant backgrounds are the continuum $\mathrm{W^+W^-}$ and $\mathrm{t\overline{t}}$ productions. The missing energy in the detector is due to the neutrinos and it does not allowed the full reconstruction of Higgs boson, resulting in a broad peak in the W W invariant mass.

The decay mode $\mathrm{H} \to \mathrm{ZZ}^* \to 4\ell$ has a lower branching fraction with respect to WW decay, but its final state is very clean and almost background free. It is considered a discovery channel. Differently from the decay into two W bosons, this decay channel allows a precise reconstruction of the Higgs boson thanks to the presence of isolated charged leptons with high transverse momentum. The main sources of backgrounds are Z + jets, $\mathrm{t\overline{t}}$ and $\mathrm{ZZ}^{(*)}$.

Other final states of $\mathrm{H} \to \mathrm{ZZ}$ decay such as with two leptons and two quarks and two leptons and two neutrinos are mostly relevant for a Higgs boson mass above the ZZ doubly resonant peak where the background contribution can be significantly reduced. Moreover they benefit from a branching fraction larger than the golden channel, in particular the $\mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\overline{q}}$ has a BR about 20 times the $\mathrm{ZZ} \to \ell^+\ell^-\ell^+\ell^-$ decay BR. They provide an important contribution to the combined sensitivity to the Higgs boson exclusion in the mass range over $2m_{\mathrm{W}}$.

The total width of the Higgs boson resonance is a function of m_{H}. Below the $2m_{\mathrm{W}}$ threshold the width is dominated by experimental resolution and is of the order of the MeV, over that threshold it becomes larger (Figure 1.9).

At high values of Higgs boson mass, the width $\Gamma(\mathrm{H} \to VV)$ is proportional

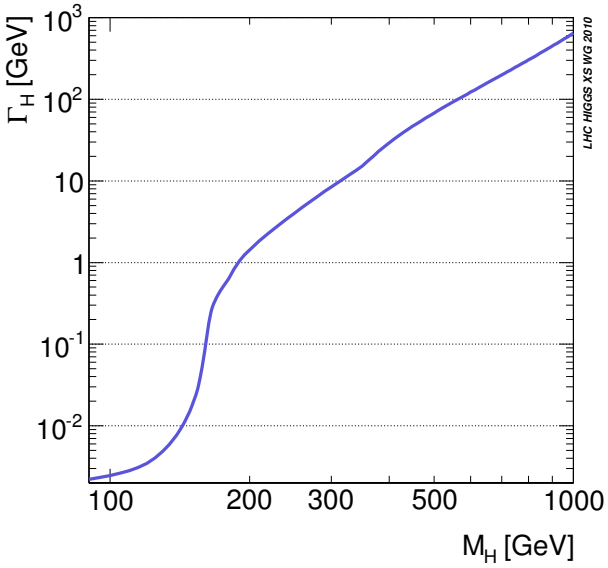


Figure 1.9: Higgs boson total decay width as a function of the Higgs boson mass, m_{H}.

to m_{H}:

$$\Gamma(\mathrm{H} \to VV) = \frac{3}{32\pi} \frac{m_{\mathrm{H}}^3}{v^2} \ . \tag{1.67}$$

As m_{H} increases, it becomes more problematic to separate the Higgs signal peak from the background.

Chapter 2

The Large Hadron Collider and the CMS experiment

2.1 The LHC collider

The Large Hadron Collider (LHC) [23] is an accelerator and collider, of 27 Km circumference, built at CERN and located in the already existing LEP tunnel on the the Swiss-French border, at about 100 m depth underground.

It has been designed to produce proton-proton collisions up to a centre-of-mass energy of 14 TeV.

The choice to project a pp collider, rather than an e^+e^- one, is mainly due to the need to reach very high energy without great loss due to synchroton radiation, which is proportional to the fourth inverse power of mass of accelerated particles.

The protons are provided by a hydrogen source removing electrons. They are then accelerated up to 50 MeV by the LINAC (LINear ACcelerator) and put in the Proton Synchroton Booster (PSB) to be accelerated up to 1.4 GeV. Protons are grouped in bunches and farther accelerated up to 26 GeV in the Proton Synchroton (PS) with the correct bunch-bunch spacing, that should get values of 25 ns as by design, and then they are injected into the Super Proton Synchroton (SPS) that accelerates them up to 450 GeV. Finally, they are transferred into the two rings of LHC. Here proton bunches circulate guided by high field superconducting magnets (up to 8.4 T) that are cooled by a huge cryogenics system.

Figure 2.1 shows an overview of CERN accelerator complex.

Proton beams collide in four interaction points (IP), at which four detectors are installed: **CMS** (Compact Muon Solenoid)[24], **ATLAS** (A Toroidal LHC ApparatuS)[25], **LHCb** (Large Hadron Collider beauty experiment)[26],

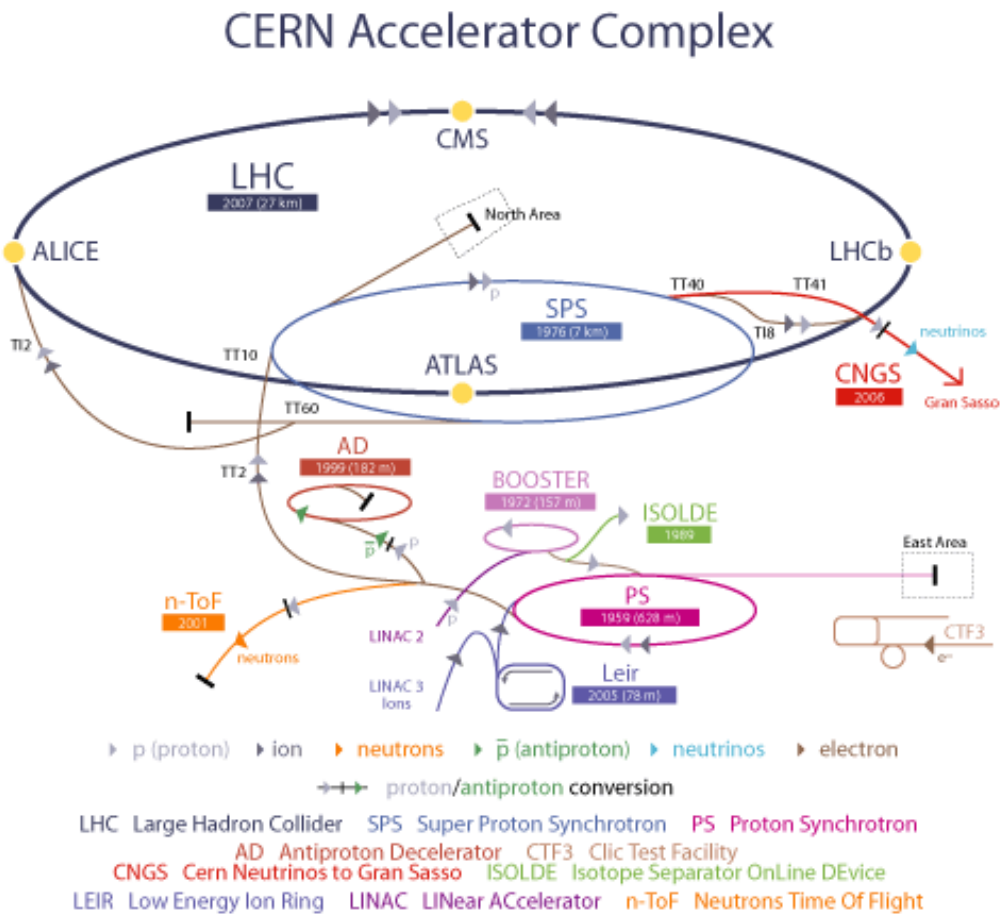


Figure 2.1: Overview of the CERN accelerator system.

and **ALICE** (A Large Ion Collider Experiment)[27].

One of the most important parameter of an accelerator is the instantaneous luminosity $\mathcal{L}$ which is related to the total cross section σ of the two beam collisions and to the event rate R by the relation:

$$R = \mathcal{L}\sigma$$

The integrated luminosity $L = \int \mathcal{L}dt$ is a measurement of the number of events produced in certain time interval per cross section unit. For a given data acquisition time, larger luminosity corresponds to a larger number of events produced, making possible the investigation of processes with small cross sections, such as the Higgs boson production.

The design luminosity for LHC is $\mathcal{L} = 10^{34} cm^{-2} s^{-1}$, leading to around 1 billion proton-proton interaction per second. It depends on machine parameters according to the relation:

$$\mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n \beta} F$$

where γ is the Lorentz factor of the accelerated protons, f is the beam revolution frequency, k_B is the number of bunches per beam, N_p is the number of protons per bunch, ϵ_n is the normalized transverse emittance[1], β is the betatron function at the IP, and F a the reduction factor due to the crossing angle. In Table 2.1 the main parameters for the LHC machine are listed.

Table 2.1: The machine parameters relevant for LHC.

Parameter		**p p**	**Heavy Ions**	
Energy per nucleon	E	7	2.76	TeV
Dipole field at 7 TeV	B	8.33	8.33	T
Design Luminosity	$\mathcal{L}$	10^{34}	10^{27}	$cm^{-}2s^{-}1$
Bunch separation		25	100	ns
Number of bunches	k_B	2808	592	
Number of particles per bunch	N_p	1.15 x 10^{11}	7.0 x 10^7	
β value at IP	β	0.55	0.5	m
RMS beam radius at IP	σ	16.7	15.9	μm
Luminosity lifetime	τ_L	15	6	hr
Number of collisions/crossing	n_c	~ 20	-	

[1]The design value of the emittance is 3.75 μ m

The LHC startup was in September 2008, but due to an accident caused by a failure in an interconnection between two magnets, the operation was stopped to restart in March 2010 and continued in the following years. LHC has been running at $\sqrt{s} = 7$ TeV in 2010 and 2011 and at $\sqrt{s} = 8$ TeV in 2012. Figure 2.2 shows the cumulative luminosity delivered by the machine to CMS versus time over the three years of data taking.

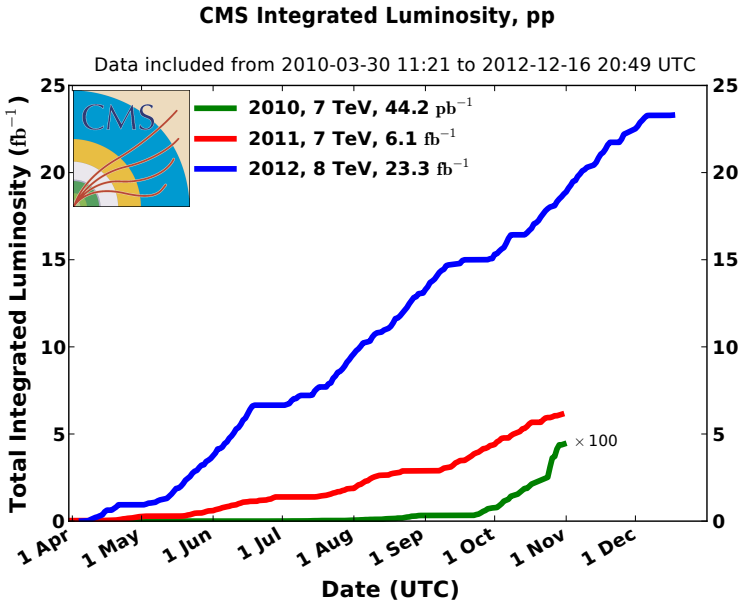


Figure 2.2: Cumulative luminosity versus time delivered to CMS by LHC during stable beams for pp collisions. This is shown for 2010 (green), 2011 (red) and 2012 (blue) data-taking periods.

2.2 The CMS detector

The Compact Muon Solenoid (CMS) [24, 28] experiment is one of the two general purpose experiments which take data at the LHC. Its aim is to investigate a wide range of physics: from the search for the Higgs boson to searches for new physics beyond the Standard Model, to the precision measurements of already known particles and phenomena. CMS has been conceived in order to supply its scientific program, and its related features are:

- A superconducting solenoidal magnet generates a magnetic field of

3.8 TeV, allowing a compact design of the detector and ensuring, together with the Muon System sub-detector, a very good muon momentum resolution and also dimuon invariant mass resolution ($\sim$ 1% at 100 GeV) with high capability to determine unambiguosly the charge of muons up to $p \sim 1$ TeV;

- The inner track system guarantees a high track reconstruction efficiency and a good momentum resolution of charged particles. Pixel detectors close to the impact region provide an efficient triggering and offline τ and b tagging.

- Electromagnetic calorimeter assures a good energy resolution as well as a good photon pairs and electron pairs invariant mass resolution ($\sim$ 1% at 100 GeV). It also provides a correct measurement of the direction of photons and the identification of primary vertices.

- A highly hermetic hadronic calorimeter system is capable to deliver good performance in missing transverse energy reconstruction.

These features will be presented and explained in more details in the following sections.

The CMS detector has been designed also taking into account some relevant problems such as the high rate of data to record but also the large amount of multiple interactions overlapping in the same event.

The total pp cross section at $\sqrt{s} = 14$ TeV is about 100 mb, thus, at the design luminosity CMS will observe an event rate of 10^9 inelastic events/s. The trigger and the online selection process must be able to reduce the number of events, to store and analyze, up to about 100 events/s. The short time between two bunch-crossing (25 ns) requires very fast readout and trigger systems.

Products of different collisions within the same bunch or even different bunches can *pile up* on the same event, thus making information on the event uncorrect. In order to avoid a superimposition of events coming from two following bunches, the readout electronics must be very fast. The effect of pile up can be reduced also by using high granularity detectors with good time resolution. This implies a great complexity because a high number of electronic channels is needed and therefore a good synchronization among them.

Moreover the detector is subject to radiation damage due to the large flux of particles coming from the interaction region and it must be high resistant to it. Another source of radiation damage is the *beam-halo* which provides a not negligible rate of particles hitting the sub-detectors.

Over the last three years of data-taking, CMS has shown excellent performance with a very high efficiency of data recording. Figure 2.3 shows the integrated luminosity delivered to (blue), and recorded by CMS (orange) during stable beams and for pp collisions at 8 TeV in 2012.

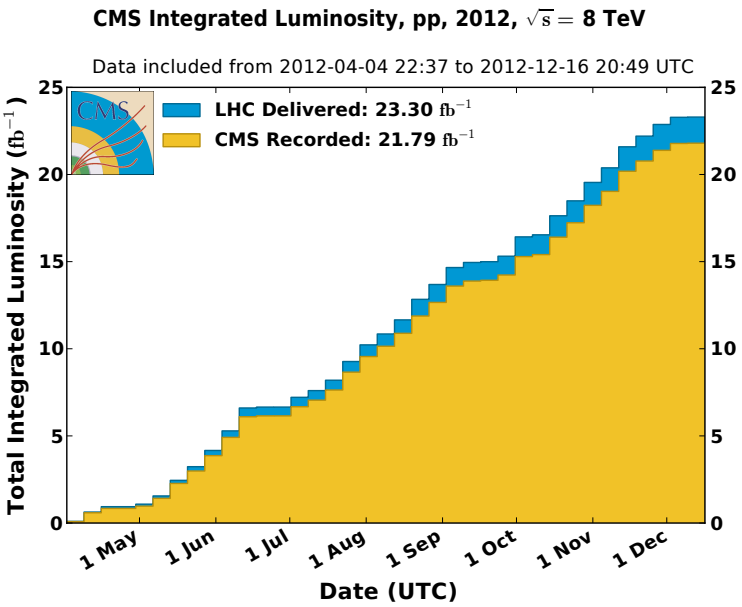


Figure 2.3: Total cumulative luminosity versus week delivered to (blue), and recorded by CMS (orange) during stable beams and for pp collisions at 8 TeV centre-of-mass energy in 2012.

The coordinate system

The CMS detector has of a cylindrical longitudinal structure, symmetrical with respect to the beam line direction. It is composed of a central region developed parallel to the symmetry axis, called *barrel*, and of two disks orthogonal to the beam pipe, the *endcaps*. An overview of the CMS detector with all its sub-detectors is illustrated in Figure 2.4.

The coordinate system adopted by CMS is shown in Figure 2.4 close to the detector. It has the origin centred at the nominal interaction point of the experiment. The z-axis points along beam direction. The x-axis and the y-axis belong to the plane orthogonal to the beam direction, the former points at the centre of the LHC, while the latter points vertically upward. The CMS detector has a cylindrical symmetry around the z-axis, and for this reason a

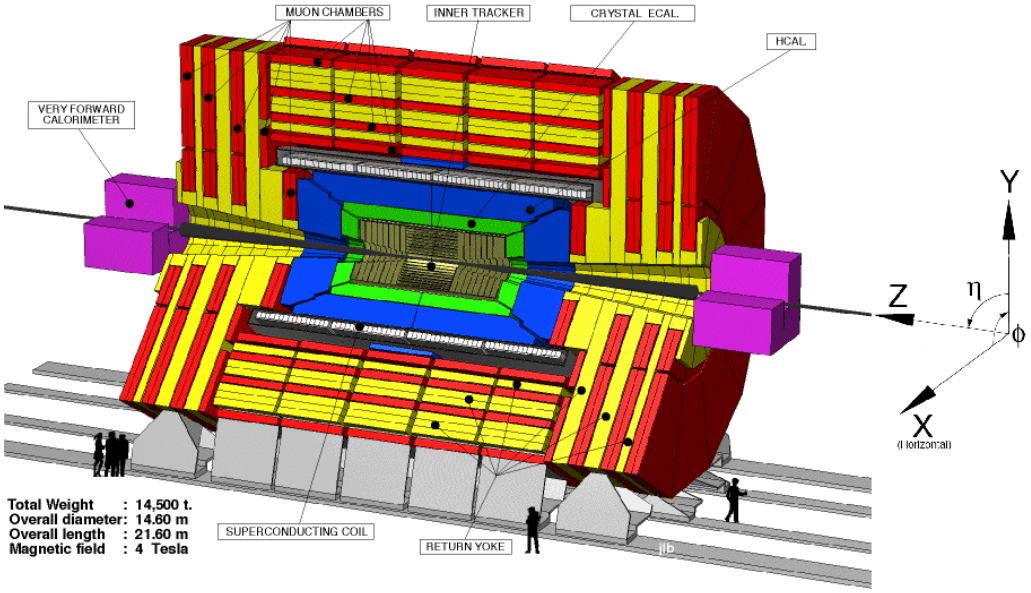


Figure 2.4: An overview of the CMS detector. On the right the CMS coordinate frame.

more convenient coordinate system is given by the radius r, i.e. the distance from z-axis, the azimuthal angle ϕ defined from the x-axis in the $x-y$ plane and the polar angle from z-axis in the $y-z$ plane. For hadron colliders the polar angle θ is more conveniently replaced by the pseudorapidity η defined as $\eta = -\ln(\tan(\theta/2))$, which is invariant under Lorentz transformations.

2.2.1 The magnet

The CMS magnet [29] is a large superconducting solenoid designed to reach a 4 T field. A high bending power is required to reach a good p_{T} resolution and to distinguish unambiguously the sign for muons with momentum up ~ 1 TeV. The magnet parameters are given in Table 2.2.

Table 2.2: CMS superconducting solenoid parameters

Field	4 T
Inner Bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA
Stored energy	2.7 Gj
Hoop stress	64 atm

The magnetic coil surrounds the two calorimeters (electromagnetic and hadronic) and the tracking system. The goal is to achieve a muon resolution about 10% p_{T} for 1 TeV muons. The p_{T} resolution scales with 1/B, where B is the strength of the magnetic field[2].

2.2.2 The tracking system

The inner tracking system of CMS [30] is designed to provide a precise and efficient measurement of the trajectories of charged particles coming from LHC collisions, as well as a precise reconstruction of primary and secondary vertices. It is placed in the inner part of the apparatus, completely embedded in the 4 T magnetic field generated by the solenoid. It extends for a length

[2] The resolution in p_T improves with the increase of magnetic field intensity according to this relation:

$$\left.\frac{\sigma_{p_T}}{p_T}\right|_{geom} = \frac{\sigma_s}{s} = \frac{8}{0.3BL^2} p_T \sigma_s \ .$$

of 5.8 m and for a diameter of 2.5 m and it is centred around the interaction point. The high number of particles traversing the tracker for each bunch-crossing (up to 25 ns) requires high granularity in order to identify and reconstruct the trajectories and a fast response to assign them to the correct bunch-crossing. Indeed the high density of the readout electronics needs an efficient cooling. The large particle flux in the impact region is also an intense source of radiation, hence the detector has to be resistant to radiation damage. All these features are provided by the silicon technologies. For this reason the tracker has been designed entirely as a silicon detector. As already mentioned, at design LHC luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$, about 1000 particles per bunch-crossing hit the tracker. This means that at a radius of 4 cm, the hit rate density is 1 MHz/mm^2, 60 kHz/mm^2 at a radius of 22 cm and 3 kHz/mm^2 at a radius of 115 cm. The tracking system is therefore divided into three regions:

- *Pixel* detectors are placed closest to interaction vertex where the particle flux is higher: three cylindrical barrel layers at radii of 4.4, 7.3 and 10.2 cm and two disks on each endcap. The size of each pixel is of 100 x 150 μm^2 giving an occupancy of about 1%;

- The radial region between 20 cm and 55 cm, where the particle flux is lower than the region closest to the interaction vertex, is covered by a *silicon microstrip* tracker with a minimum cell size of 10 cm x 80 μm. leading to an occupancy of about $2 \div 3$%;

- In the outermost region, that extends from 55 cm to 116 cm a system of *larger-pitch silicon strips* is adopted since the particle flux is sufficiently low to allow their use. With a maximum cell size of 25 cm$\times$180 mm, keeping an occupancy around 1%;

Figure 2.5 shows a schematic drawing of the CMS tracker.

The Pixel system (Figure 2.6) provides precise measurements of trajectory points in $r - \phi$ and z with a spatial resolution about 10 μm for the $r - \phi$ measurement and about 20 μm for the z measurement. It is very important to reconstruct secondary vertices from b and τ decay, and also to provide a seed track for outer track reconstruction. The Pixel detector covers the pseudorapidity region $|\eta| < 2.5$. The barrel detector is made of three layers at radii of 4.4, 7.3 and 10.2 cm, while the forward pixel detector is composed of two disks for each endcap, extending from 6 to 15 cm in radius and placed at $|z| = 34.5$ cm and 46.5 cm.

The intermediate and outermost regions of tracking system, between 20 and 116 cm, are occupied by the silicon strip tracker (Figure 2.7).

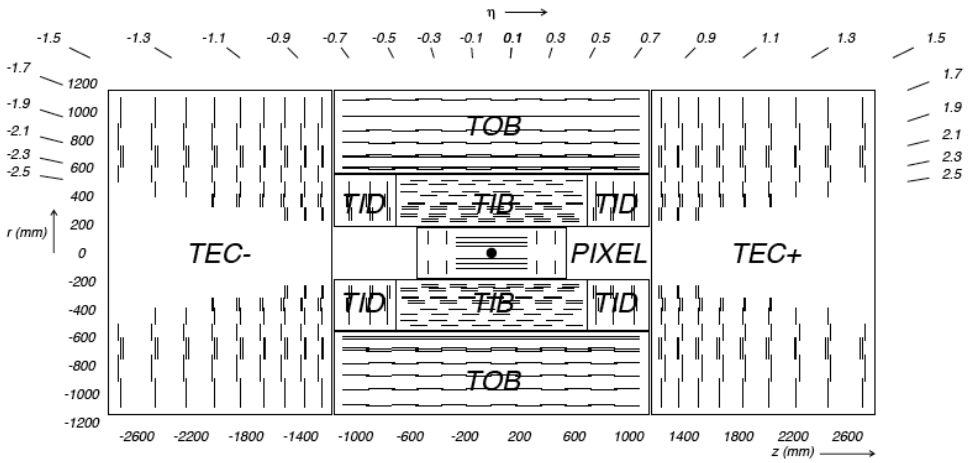


Figure 2.5: Transverse section of the CMS tracking system. It is made of an inner Pixel detector and a outer Strip detector divided into 4 sub-detectors: Tracker Inner Barrel (TIB), Tracker Inner Disks (TID), Tracker Outer Barrel (TOB) and Tracker EndCaps (TEC+ and TEC-). Each line represents a detector module, while double lines indicate back-to-back modules which deliver stereo hits.

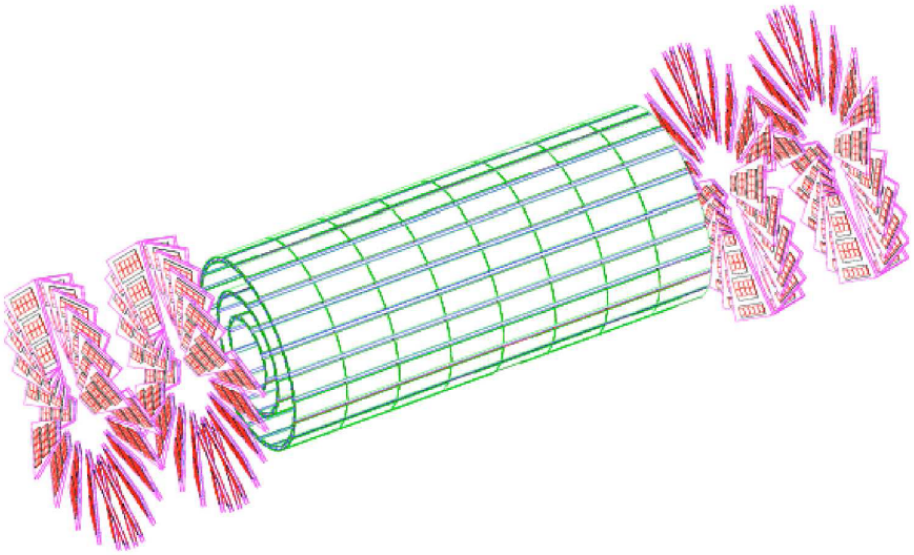


Figure 2.6: The three layers of barrel section of the Pixel detector (green) and the four disks of the endcap (red).

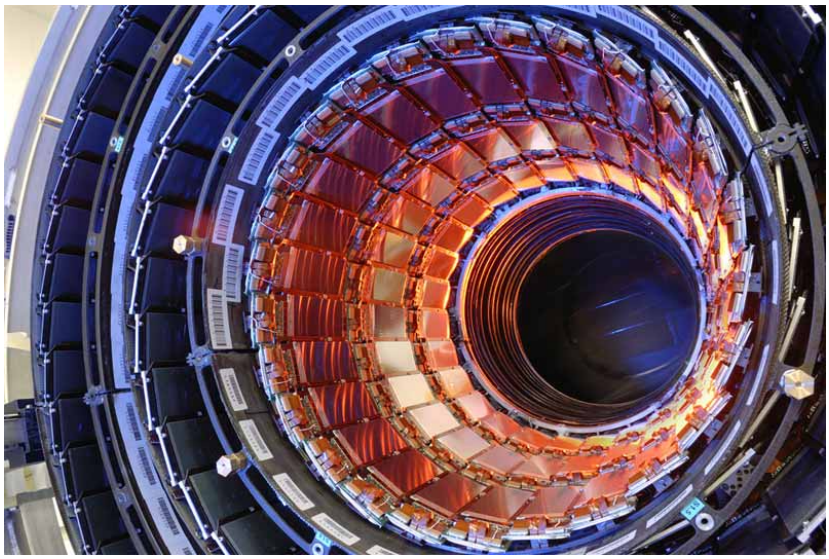


Figure 2.7: Overview of the Strip detector.

The silicon microstrip detector is divided in two regions: the inner and the outer regions. The inner region consists of 4 barrel layers, the Tracker Inner Barrel (TIB), and 3 disks on each side, the Tracker Inner Disks (TIDs). The outer system is made of 6 barrel layers, the Tracker Outer Barrel (TOB) and 9 disks for each endcap, the Tracker EndCaps (TEC+ and TEC-). It covers the pseudorapidity region $|\eta| < 2.5$. All the four regions are provided with both single and double sided microstrip modules. The strips are oriented along the z direction in the barrel and along the r coordinate in the endacps. The microstrip detector is designed to provide a spacial resolution of about $40 \div 60$ μm in the $r - \phi$ plane and about 500 μm along z, with an occupancy lower than 1%.

2.2.3 The electromagnetic calorimeter

The Electromagnetic calorimeter of CMS (ECAL) [31] is a hermetic homogeneous calorimeter made of 61200 lead tungstate ($PbWO_4$) crystals mounted in a central barrel and of 7324 crystals in each of the two endcaps. The system is completed with a preshower (designed to reject π^0) inserted in front of the endcaps.

The ECAL is critical to identify the decay in two photons of the Higgs boson. This requirement demands a very good energy resolution, provided by the homogeneous calorimeter design. The lead tungstate crystals have high density (8.28 g/cm^3), short radiation length (0.89 cm) and short Moliére radius (2.2 cm) resulting in a fine granularity and in a compact structure. An interesting feature of these crystals is their scintillation decay time, which is of the same order of magnitude as the bunch-crossing time, about 80% of the light is emitted in 25 ns. However their light yield is very low (30 photons/MeV) and the dependence on the temperature is strong. This requires the system must to maintain to a constant temperature with high precision. A cooling system keeps the temperature of crystals and photodetectors stable within ± 0.05°C at 18°C.

The barrel section of the ECAL (EB) covers the region at $|\eta| < 1.479$ with 61200 crystals of dimensions: 22×22 mm^2 at the front face, 26×26 mm^2 at the rear face, and a length of 230 mm, corresponding to 25.8 X_0.

In the barrel the light signal coming from scintillators is converted into an electric signal by avalanche photodiodes which are highly resistant to radiation, able to operate in the 4 T magnetic field and have high gain and high quantum efficiency. In the endcaps vacuum photodiodes are used, which have lower quantum efficiency and internal gain, but a larger detection surface.

A preshower device covers the region $1.653 < |\eta| < 2.6$. It is made of two

planes of silicon strips which lie behind disks of lead absorber at depths of 2 X_0 and 3 X_0. The aim of the preshower detector is to identify neutral pions in endcap region, but also to help with the position determination of electrons and photons.

The energy resolution of CMS ECAL can be parametrized as function of the energy (Figure 2.8) as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.1}$$

where S is a stochastic term, N a noise term and C a constant term. Different contributions have a role in each term. The fluctuation in the number of produced and collected electrons is included in stochastic term, while the noise term is an overall term due to electronic noise and to pile-up events. The constant term is related to the calibration of the calorimeter.

In test beam the energy resolution has been found to be [32]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \tag{2.2}$$

2.2.4 The hadron calorimeter

The CMS Hadron Calorimeter (HCAL) [33] has been designed to cover a wide range of physics processes with different signatures in final states, particularly those involving hadron jets and neutrinos, but also exotic particles resulting in missing transverse energy (MET). In order to have a good jet four-momentum and MET measurement, the HCAL must have good energy resolution, provide a good containment, good transverse granularity and hermeticity. A strong condition for the HCAL is its location: inside the magnet coil and surrounding the ECAL (Figure 2.9). The HCAL extends from radius 1.77 m to the inner surface of the magnet at radius 2.95 m. In order to absorb the hadronic shower, a brass absorber has been chosen because of its short interaction length. Moreover this material is non-magnetic and suitable to be placed inside the magnetic field.

An outer calorimeter is placed outside the solenoid complementing the barrel calorimeter, and acts as a *tail catcher.* A forward hadron calorimeter extends the pseudorapidity coverage up to $|\eta| = 5.2$ using Cherenkov technology.

The barrel hadron calorimeter (HB) is a sampling calorimeter placed inside the magnet coil and surrounding the ECAL barrel. It covers the pseudorapidity range $|\eta| < 1.3$. The plastic scintillator is divided into 32 η sectors resulting in a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$.

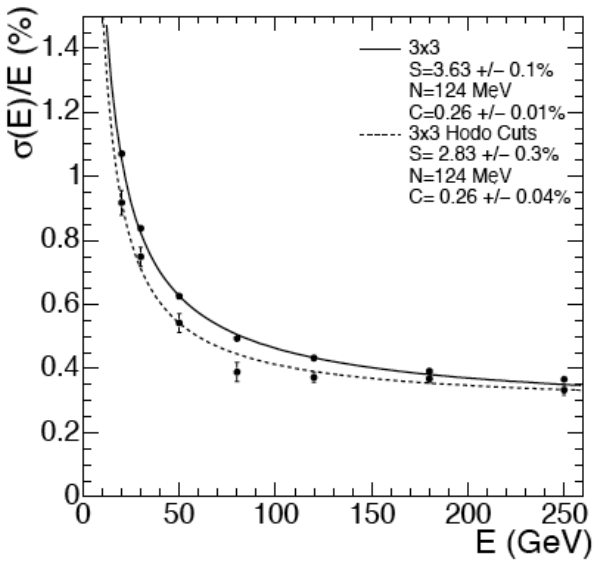


Figure 2.8: Relative ECAL energy resolution, σ_E/E, as a function of the electron energy as measured from a beam test. The energy was measured in an array of 3×3 crystals with electrons.

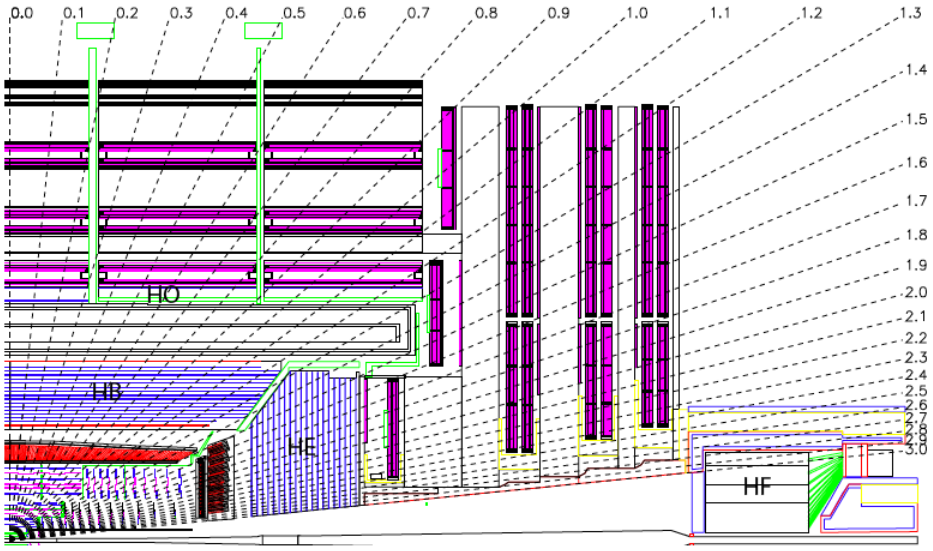


Figure 2.9: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. HB ad HE are placed surronding the ECAL and inside the solenoid. HO and HF are installed outside the magnet coil.

The HB baseline active material is Kuraray SCSN81 plastic scintillator and has a thickness of 3.7 mm. The main reasons which have led to choose it are its long-term stability and moderate radiation hardness. The first layer, layer 0, is made of 9 mm thick Bicron BC408. The light is collected in plastic scintillator tiles, readout through embedded wavelength shifting fibres and converted into an electric signal by multichannel hybrid photodiodes (HPDs).

The endcap part (HE) extends in a pseudorapidity interval $1.3 < |\eta| < 3$, a region which collects about 34% of the particles produced in the final states. It consists of 14 η towers with a segmentation in ϕ of 5° for the 5 towers at lower η, and of 10° for the 8 innermost towers. The total number of HE towers is 2304. The design of the two HE is driven by the need to avoid "dead" zones rather than to achieve high resolution. In addition, the jet energy resolution in the HE is limited by the pile up, the magnetic field effects and also by the parton fragmentation. As for the HB, the absorber material is the brass.

The trapezoidal scintillators adopted are 3.7 mm thick of SCSN81 for all the layers except the layer 0 for which a 9 mm thick Bicron BC408 is used. The scintillation light is collected also in this case by wavelength shifting fibres and readout by the multipixel hybrid photodiodes (HPDs), due to their very low sensitivity to magnetic field and also to their large dynamical range.

The outer calorimeter (HO) is placed in the pseudorapidity central region, in order to provide sufficient containment for hadron showers together with EB and HB, and it extends outside the magnet coil. The solenoid is used as the additional absorber with a thickness equal to $1.4/\sin\theta\ \lambda_I$ and allows to identify late starting showers and to measure the shower energy deposited after HB. The pseudorapidity region covered by the HO is $|\eta| < 1.3$. Acting as "tail catcher" the HO improves the MET resolution of the calorimeter.

The forward hadron calorimeter (HF) provides a coverage of the pseudorapidity range $3.0 < \eta < 5.0$ by steel/quartz fibre. The front face is located 11.2 m away from the interaction point, outside the solenoid.

In order to evaluate the performance of the three parts of HCAL (HB, HE and HF) it is usual to look at the jet energy resolution and at the missing transverse energy resolution. The granularity of these three parts has been chosen to make the jet energy resolution quite uniform , as a function of the transverse energy E_{T}. Figure 2.10 shows the transverse energy resolution as function of E_{T}. The missing transverse energy resolution is given by $\sigma_{MET} \sim 1.25\sqrt{\sum E_{\mathrm{T}}}$, where $\sum E_{\mathrm{T}}$ is the jet transverse energy, without considering clustering corrections.

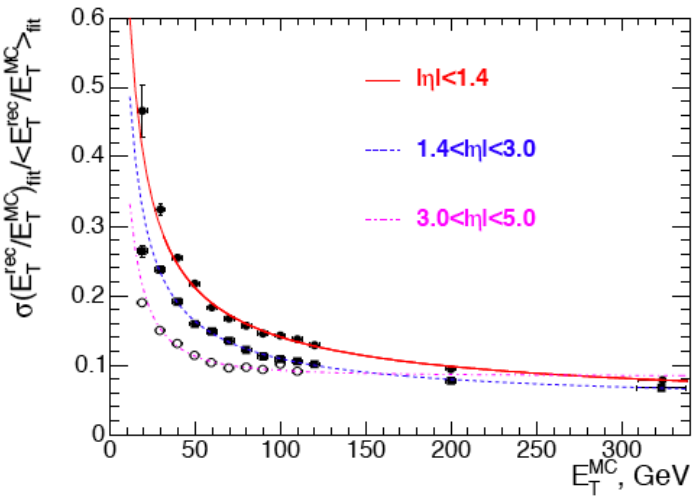


Figure 2.10: The jet transverse energy resolution, in the HCAL, as a function of the simulated jet transverse energy for barrel, endcap and forward region.

2.2.5 The muon system

The muon system [34] provides a precise muon momentum measurement, but also a time measurement of the bunch-crossing, and also works as trigger for events involving muons.

Momentum measurement, in the muon system, is determined by the muon bending angle at the exit of the 4 T coil, considering the interaction point as the origin of the muon. Up to p_T values of 200 GeV the resolution of the muon system is dominated by multiple scattering and the best resolution is given rather by the silicon tracker.

The CMS muon system is designed using three kinds of gaseous particle detectors for muon identification: drift tubes chambers (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC). As the tracker and the calorimeters, it is made of a cylindrical barrel and two endcaps. The first two provide an excellent spatial resolution, while the RPCs have a very good timing. Figure 2.11 shows the longitudinal view of the sub-detector.

The barrel extends up to $|\eta| < 1.2$ and is based on DT chambers. The drift cells (Figure 2.12) consist of a stainless steel anode wire placed between two parallel alluminium layers. The efficiency of single chamber lies around 99.8% with a spatial resolution of $\sim 180\ \mu m$. The device is organized into 4 stations

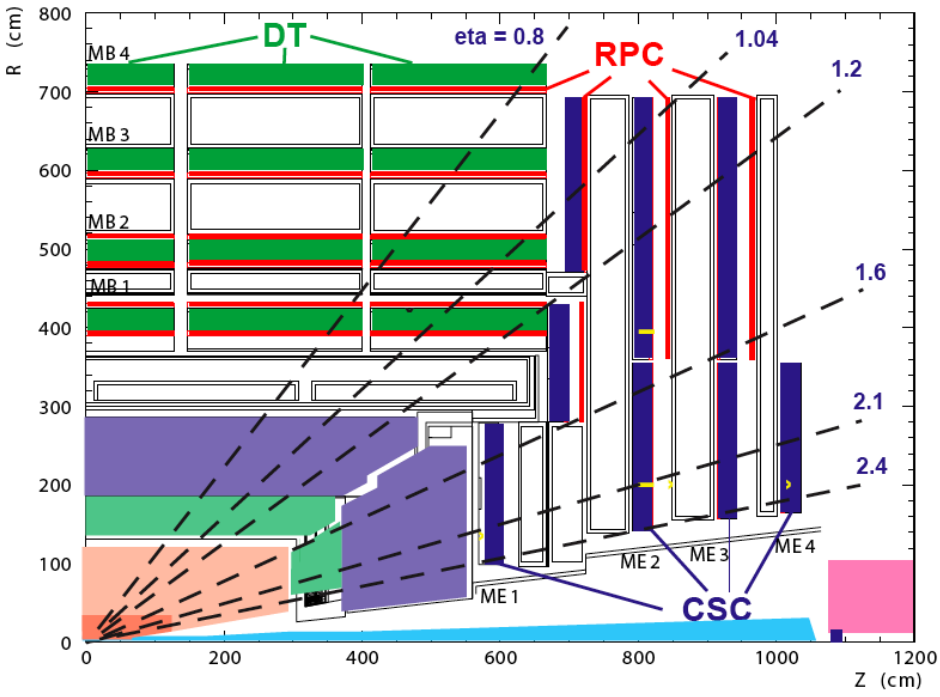


Figure 2.11: Longitudinal section of the quarter of the CMS muon system, with DT chambers in the barrel, CSCs in the endcap and RPCs coupled to both DT chambers and CSC.

(MB1, MB2, MB3 and MB4) interspersed with layers of the flux return plates as shown in Figure 2.13. The first 3 stations contain 8 chambers for the measurement of the muon coordinate in the $r - \phi$ plane, and 4 chambers which provide a z-measurement. The last station does not contain a z-measuring plane. Along the longitudinal direction the muon system barrel is divided into 5 wheels which are subdivided into 12 sectors each covering a 30° azimuthal angle. The design chosen for the barrel provides a single point resolution of $\sim 200\mu m$ and a ϕ precision better than $\sim 100\mu m$ in position and ~ 1 mrad in direction [35].

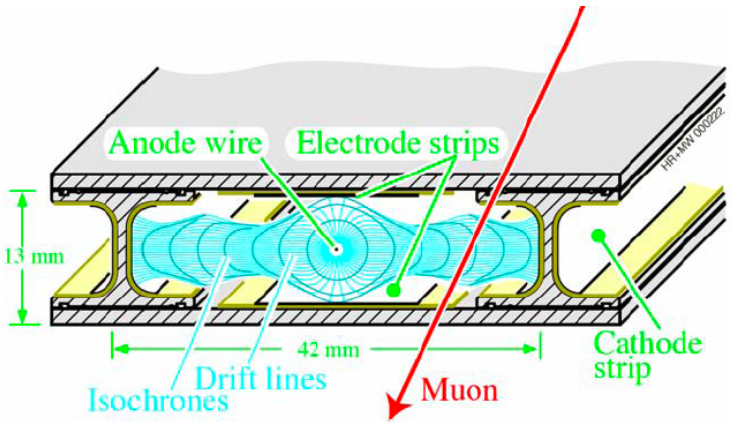


Figure 2.12: Schematic view of a drift cell. The electrons, coming from the gas ionization due to the passage of a muon, drift towards the wire anode. The voltages applied to the electrodes are +3600V for wires, +1800V for strips, and −1200V for cathodes.

DT chambers in stations MB1 and MB2 are installed between two RPCs and in the stations MB2, MB3 they are coupled to one RPC. High p_T muons traversing the muon system cross up to 6 RPCs and 4 DT chambers providing in the DT system up to 44 measured points from which the muon track candidate can be built.

The two endcap regions of CMS are characterized by a non-uniform magnetic field and large background levels. The muon system is constructed, in these regions, with CSCs, which are multi-wire proportional chambers, with fast response time, fine segmentation and radiation resistance. The covered pseudorapidity range is $0.9 < |\eta| < 2.4$. Each endcap is divided into 4 stations of CSCs of trapezoidal shape and is installed perpendicularly to the beam line. Each CSC (Figure 2.14) consists of 6 gas gaps. Each gap has a

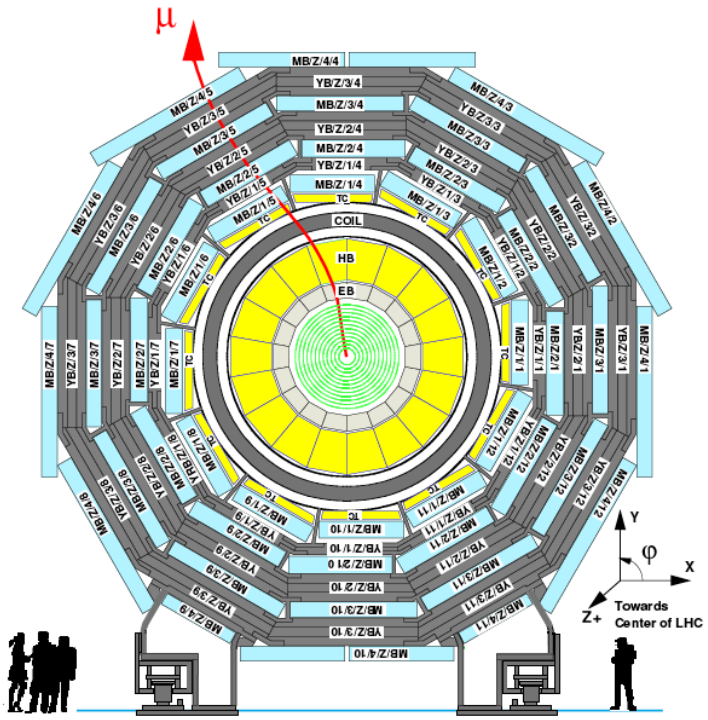


Figure 2.13: Layout of the CMS barrel muon DT chambers in one of the 5 wheels. Each wheel is divided into 12 sectors, each covering a 30° azimuthal angle, and consisting of 4 stations.

plane of radial cathode strips, providing an $r - \phi$ measurement, and a plane of anode wires placed orthogonally to the cathode plane and read out in order to give an η measurement, as well as a beam-crossing time of a muon. Cathode strips allow to achieve a spatial resolution tipically $\sim 200\mu m$, and an angular resolution in ϕ of 10 mrad [36]. In order to improve the time and p_{T} resolution and to resolve ambiguities due to multiple hits in a chamber, RPCs are installed also in the endcap. An RPC is coupled to each of the first three stations.

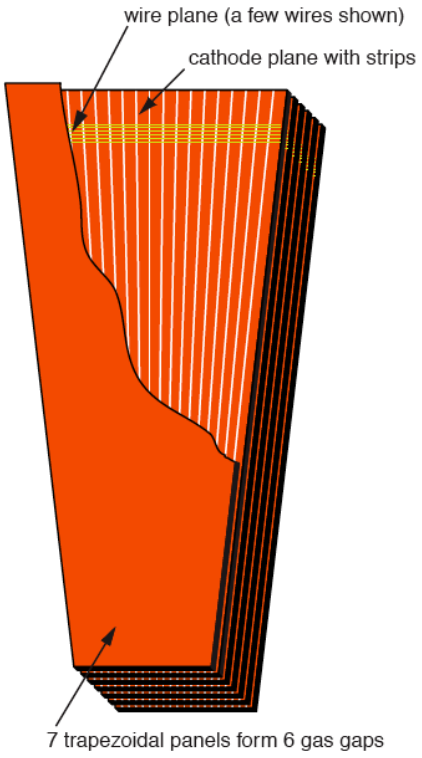


Figure 2.14: View of the CSC made of 7 parallel trapezoidal panels forming 6 gaps.

RPCs [37] provide spatial informations with a time resolution comparable to scintillators: the tagging time of an ionizing event is shorter than 25 ns. It allows to know which bunch-crossing the event belongs to. For this reason a dedicated muon trigger is based on RPCs.

The RPCs consist of 2 gaps formed by four bakelite electrodes, which are covered by graphite in order to distribute uniformly the high voltage over

the surface. Since the RPCs work in avalanche mode the gas gain is low, and the signal has to be amplified by the readout electronics. Figure 2.15 shows a schematic view of an RPC.

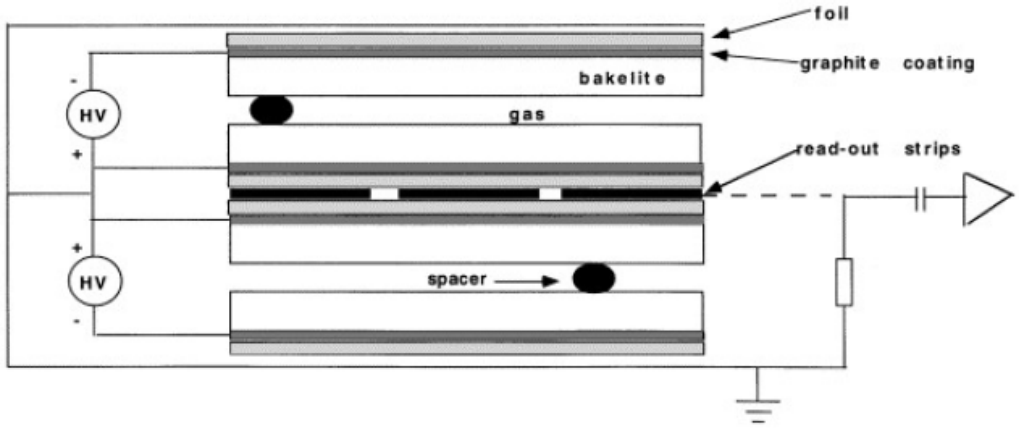


Figure 2.15: Layout of the RPC double-gap structure.

2.2.6 The trigger

At LHC design luminosity, for a bunch-crossing rate of 40 MHz, a very high rate of interactions is expected ($\sim 10^9$ interactions/s). Each event has a size about 1MB. It is impossible to store and process such a large amount of data. A selection has to be made on the events in order to reduce the selected event rate to about 100 Hz, according to what is allowed by the limits on the storage capacity. This goal is achieved by the trigger system in two steps: the Level-1 (L1) Trigger [38] and the High-Level Trigger (HLT) [39].

Data readout from the front-end electronics must reach the service cavern that houses the L1 Trigger system and return back to the front-end electronics and provides a decision about taking or discarding the data from a particular bunch-crossing. It takes about 3.2 μs to perform a decision.

The L1 Trigger must reduce the events rate to 100 KHz, mantaining at the same time a high efficiency on interesting events. The time in which the Trigger has to take a decision is too short to consider information from all raw data; then, it works involving only the calorimeters (Calorimeter Trigger) and the muon system (Muon Trigger), as well as correlations among

informations of both two systems (Global Trigger). The presence of "trigger objects" such as electrons, photons, muons and jets, satisfying specific requirements on E_{T} and p_{T}, is the base for L1 Trigger decisions. The triggered objects pass to the subsequent Data Acquisition system (DAQ) and HLT for further reconstruction and selection steps. The performance of the L1-Trigger measured during the CMS commissioning with cosmic muons and with first LHC beams is described in Ref. [40]

The HLT aims at further reducing the event rate to about 100 Hz, using more detailed informations than the L1 trigger, and more sophisticated reconstruction algorithms. The data coming from the readout buffers are transferred to processors, each running the HLT software code to produce a smaller output rate for mass storage. The idea on which the HLT is based, is to reconstruct only objects in the region of interest and discard other objects. Many virtual trigger levels are used. A "local reconstruction" is made initially using the full information of the muon system and the calorimeters. In a second step also the information of the tracker hits is taken into account and added. The last step is the use of the full event informations (calorimeters, muon system and tracker).

2.3 Object reconstruction at CMS

2.3.1 Electron and muon reconstruction

Muons cross the detector and interact with the silicon tracker through ionization. The loss of energy by bremsstrahlung is negligible except for those muons produced with very high transverse momentum ($p_{\mathrm{T}} > 100$ GeV) and these particle are capable to reach the outer part of the CMS detector. Muons are reconstructed in the inner silicon tracker and in the external muon chambers [41, 42]. The first step of reconstruction happens in the muon spectrometer collecting position measurements from DT, CSC and RPC sub-detectors. Hits in DT and CSC chambers are matched to form segments, which are then collected and matched to generate seeds used as a starting point for the track fit of DT, CSC and RPC hits. The reconstructed track in the muon spectrometer is called *Stand-alone muon*. As in the muon system, track reconstruction is performed in the inner tracker as well. Seed are built using two or three consecutive hits in the pixel and/or in the strip detector. The pattern recognition is performed starting from these seeds and proceedings layer by layer, with an iterative technique based on the Kalman filter technique [43]. At the end of this algorithm a fit is performed. The identified tracker tracks are then combined with a given stand-alone muon track in

order to construct a global track: *global muon*. A global fit is performed for each pair of track in the tracker and in the muon system. If more than one track matching the stand-alone track is found, then the one with the best χ^2 is chosen. The reconstruction ends with the association of energy deposits in the calorimeters to the global tracks. Muons which are reconstructed in the inner tracker only and simply matched to a segment in either DT or CSC are called *tracker muons*.

Electron reconstruction [44, 45] is based on the combination of ECAL and tracker information. The reconstruction algorithm starts by measuring the energy deposited in ECAL in groups of one or more associated clusters, which are assembled in *superclusters* (SCs). Superclusters are reconstructed with an algorithm which takes account of their characteristic narrow width in the η coordinate and their characteristic spread in ϕ due to the bending in the magnetic field of electrons radiating in the tracker material. After SCs reconstruction, the algorithm proceeds matching them to track seeds identified as pairs or triplets of hits in the inner tracker layers. Electron trajectories are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian Sum Filter (GSF) [44]. In order to solve ambiguous cases where several tracks are reconstructed due to the conversion of radiated photons in the tracker material, a preselection is applied. This preselection is based on the matching between the GSF track and the supercluster in η and ϕ. Together with "ECAL driven" reconstruction, at CMS, electron reconstruction is performed also in a "tracker driven" approach, i.e. based on tracks of electrons reconstructed in the tracker only. A preselection is applied on these tracks, based on a multivariate analysis. If electrons reconstructed with "ECAL driven" approach do not satisfy the specific preselection, but they pass the multivariate preselection, then they are kept. More detailed information about the electron reconstruction and results from the commissioning with the first collision data can be found in Ref. [45]

2.3.2 Jet reconstruction

Quarks and gluons cannot be directly detected but they rather appear in detector as "jets" of particles. Jets are the result of hadronization of the parent parton. Jets reconstruction is based on the clustering of particle candidates reconstructed using the PF approach, described in the following section. The identified particles are clustered according to the anti $-K_t$ algorithm [46] with distance parameter $R = 0.5$.

The anti $-K_t$ algorithm satisfies a very important property of a good jet algorithm, the infrared safety [47], which means that the number of jets must be sensitive to:

- *Soft radiation*: the addition of a infinitely soft particle to the list of jet constituents must not affect jet reconstruction;

- *Collinear radiation*: if a particle of momentum p is split into two collinear particles each carrying half a momentum of the original particle, the result of clustering must not be affected;

The anti$-K_t$ algorithm works clustering particles proceeding by computing the distance d_{ij} between entities (particles and pseudojets) and d_i between the i^{th} particle and the beam:

$$d_{ij} = min\ (p_{\mathrm{T},i}^{2k}, p_{\mathrm{T},j}^{2k})\ \frac{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}{R^2}, \tag{2.3}$$

and

$$d_i = p_{\mathrm{T},i}^{2k}, \tag{2.4}$$

where $k = -1$ and R is the algorithm radius parameter.
The minimum between d_{ij} and d_i is calculated, if it is a d_{ij}, the i^{th} and j^{th} objects are merged summing up their momenta and then they are removed from the list and a new iteration is performed, otherwise the object is removed from the list and called *jet*. This procedure continues until only jets objects are left. The meaning of R is to constraint particles in the jet to have distance smaller than R resulting in a conical jet of radius equal to the distance parameter.

2.3.3 Particle Flow reconstruction

The Particle Flow (PF) event reconstruction algorithm [48, 49, 50] aims at reconstruction of all stable particles in a given pp collision - muons, electrons, photons, charged hadrons and neutral hadrons - combining all CMS subdetectors to determine their direction, energy and type. The design of the CMS detector is well suited for the Particle Flow technique. The presence of an internal large silicon tracker immersed in an uniform 3.8 T magnetic field allows to reconstruct charged particles with high efficiency and to provide a precise measurement of the transverse momentum down to p_{T} of 150 MeV. Together with the excellent tracking system and the large magnetic field, the high granularity of the ECAL is a key element for the PF reconstruction, allowing the reconstruction of photons and electrons with high energy resolution.
Firstly the PF algorithm proceeds with the identification of fundamental *elements* as reconstructed in the sub-detectors: charged-particles, calorimeter clusters and muon tracks. These elements are then connected to each other

by making use of link algorithms identifying *blocks* of elements which are topologically compatible, e.g. charged-particle track is linked to a calorimeter cluster if the extrapolated position from the track to the calorimeter is within the cluster boundaries. From the *blocks* particle flow particle candidates (PF Candidates) are fully reconstructed and identified in the following order. Muons, electrons, charged hadrons, photons and neutral hadrons are reconstructed in this order according to the following procedures:

- *Muons*: a global muon, reconstructed from the combination of a track in the tracker and a track in the muon system, gives rise to a PF muon. After the identification, the corresponding track is removed from the block;

- *Electrons*: the link between a charged-particle track (refitted with the Gaussian-Sum Filter [44]) and one or more ECAL clusters identifies PF electrons, The corresponding track and ECAL clusters are removed from further processing;

- *Charged hadrons*: The remaining tracks give rise to PF charged hadrons and the momentum of the particle is taken directly from the track momentum. Tracks can be linked to ECAL and HCAL clusters if they are not identifies as electrons, and the momentum is redefined taking into account information from calorimeters;

- *Photons and Neutral hadrons*: ECAL clusters not compatible with charged-tracks give rise to PF photons, while unaccounted HCAL deposits are interpreted as PF neutral hadrons.

Once the list of PF Candidates is defined PF Jets can be reconstructed using the clustering jet algorithm described in Section 2.3.2.

Typically the larger fraction of jet energy, about the 65%, is carried by charged particles, the 25% by photons and the 10% by neutral hadrons. The 90% of the jet energy can be reconstructed with good precision, thanks to the high resolution of the tracker and of the electromagnetic calorimeter, while just the 10% is affected by the poor HCAL resolution. The combination of tracks and calorimeter clusters is a key point of the PF algorithm because it allows to get a very high efficiency for PF jets comparing with the sole use of information from calorimeters.

Compared to the traditional calorimeter-based jet reconstruction (Calo-Jets), PF reconstruction shows better jet reconstruction performance at CMS. In Figures 2.16 and 2.17 the response[3] of PF Jets and Calo-Jets are

[3]The response is defined as the mean value of the Gaussian function fitted to the distribution of the ratio $(p_{\mathrm{T}}^{rec} - p_{\mathrm{T}}^{gen})/p_{\mathrm{T}}^{gen}$

compared as function of jet pseudorapidity and transverse momentum: PF Jets are characterized by a better response throughout the detector. The comparison of jet resolution for the two techniques as function of jet p_{T} is shown in Figure 2.18. Also for what concerns jet resolution, PF offers a higher improvement especially at low and medium value of p_{T}.

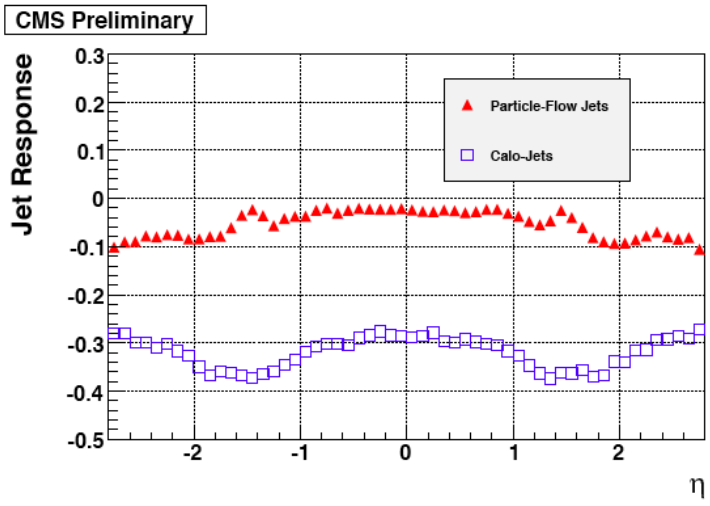
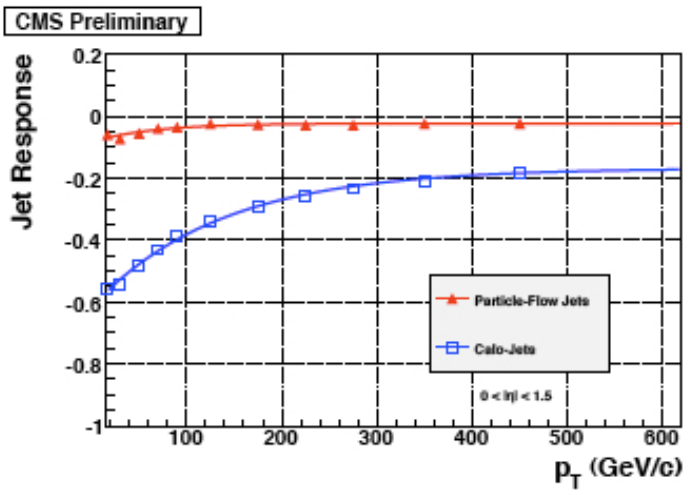


Figure 2.16: Jet Response as as a function of η, for the two jet reconstruction algorithms: PF Jets (red) and CaloJets (blue) [48].

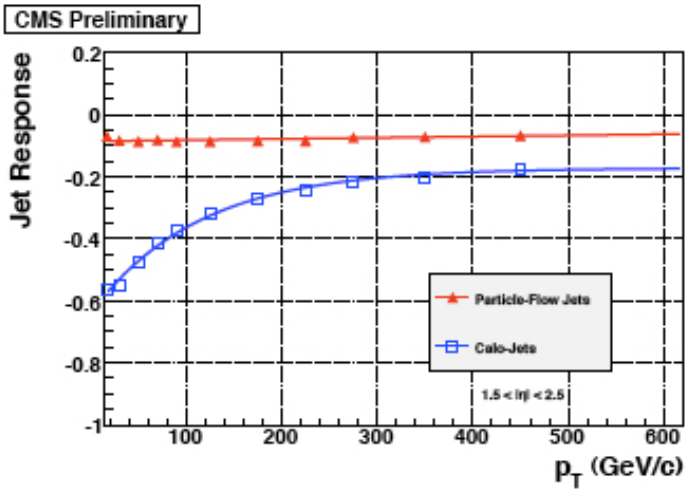
PF Missing Transverse Energy (PF MET) is reconstructed at the end of event reconstruction: it consists in forming the transverse momentum-vector sum over all reconstructed PF Candidates in the event and then taking the opposite of this vectorial sum. The missing transverse energy is the modulus of this vector.

2.4 Analysis Tools

The analysis workflow for the CMS experiment is based on the concept of modular analysis. The configuration of data flow and data analysis is build assembling modules together, each module accomplishes a particular task. The collection of software developed for CMS, referred as CMSSW, provides the infrastructure to process events and analyze them to produce physic results. The Physics Analysis Tools (PAT) [51, 52] is a set of tools conceived to facilitate the production of analysis code. They are common to different kind of analysis and can be configured and customized in order to fit specific user's physics issues.



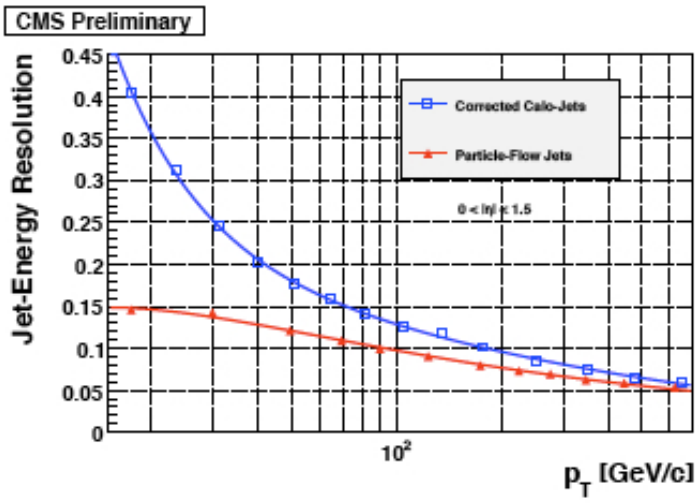
(a) Barrel



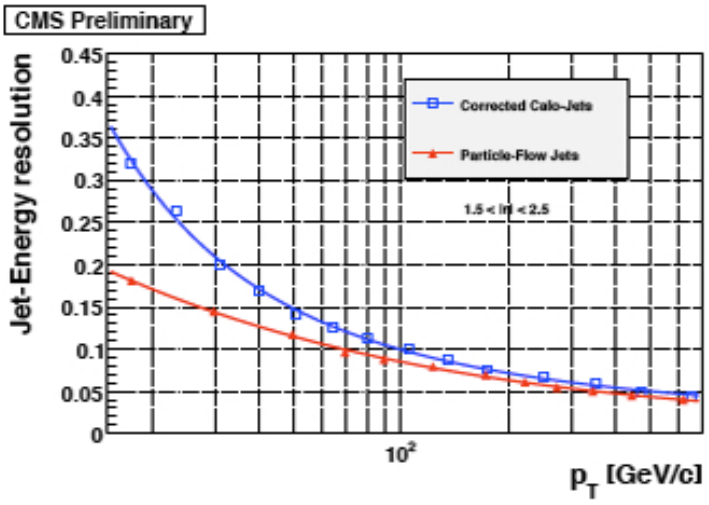
(b) Endcaps

Figure 2.17: Jet Response as a function of p_{T} integrated over all p_{T} range below 750 GeV, in the barrel (a)) and in the endcaps (b)), for the two jet reconstruction algorithms: PF Jets (red) and CaloJets (blue). The response curves are fit with exponential functions of p_{T}. The result of the fits is superimposed [48].

Examples of PAT tools are the *removeMCmatching* and the *AddJetCollection*. The former can be included to remove the request for the matching of reconstructed particles to generated particles, which would not be available in case of collision data, while the latter can be used to produce a further jet



(a) Barrel



(b) Endcaps

Figure 2.18: Jet Energy resolutions as function of p_{T} for corrected CaloJets (blue) and for PF Jets (red) in the barrel (a) and in the endcaps (b). The resolution curves are fit to the sum of a stochastic term, a noise term and a constant term [48].

collection, for instance a collection of jets produced with another clustering algorithm with respect to the default one.

The Python scripting language is adopted to define the job configuration that drives the analysis workflow. It can be challenging for users managing so many modules to produce their own analysis workflow. PAT tools can be

steered and configured using the ConfigEditor, a Graphical User Interface (GUI). PAT tools are integrated in the GUI thanks to an object-oriented organization of their architecture. Details on the infrastructure of PAT tools designed to allow their integration within the GUI can be found in Ref.[53].

2.4.1 The ConfigEditor

The ConfigEditor allows to browse and create analysis configuration files without writing code and making use of PAT tools.

The main window appear as in Figure 2.19. The configuration sequence is visualized in the *Tree View* box in the left column, the same sequence is shown in the central column in the *Connection Structure* box. In this section the job structure is shown graphically, each module is visible and can be selected. In the *Property View* box (on the right) the user can get information about the selected module, checking which are the dependencies among this and the other modules, and visualizing what is the parameter setting for the module itself and change it according to user' s needs.

PAT tools are provided as plugins to import in the Python configuration and can be accessed by the ConfigEditor and included in the analysis workflow.

The three steps to include and configure a PAT tool in the analysis job are:

- Import a standard configuration, or a user configuration. Clicking on *Import Configuration* in the menu bar of the GUI it is possible to choose the Python configuration file to start from among those available;
- Customize the configuration file by clicking on *Apply tool* on the menu bar and selecting in the box the chosen tool. A list of parameters to set along with a description of the tool itself is provided.
- Configure the tool by replacing values in the *Parameter View* box.

The resulting user configuration file appears in real time in the top left corner, in this way it can be checked immediately.

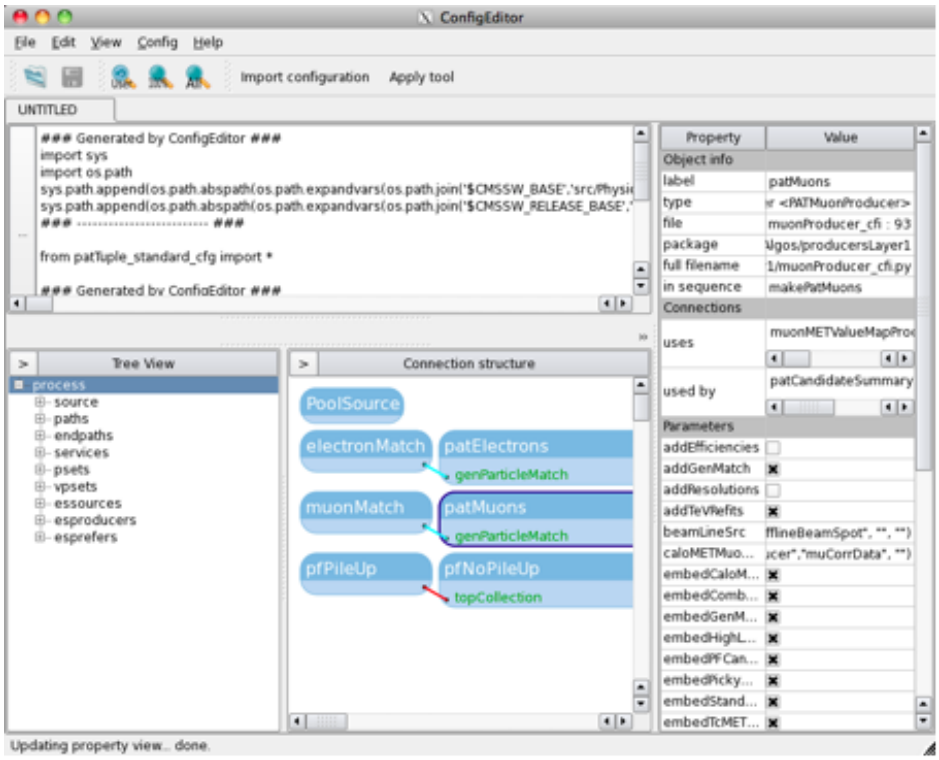


Figure 2.19: The configuration sequence is visualized by a Tree View in the left column, the same sequence is shown in the central part in the Connection Structure by a graphical representation. The Property View Box on the right gives information about the selected module: in which file it is implemented and in which package user can find it, its connections with other modules and finally the list of its configuration parameters.

Chapter 3

Search for a Higgs boson in the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^- q\overline{q}$ channel

In order to achieve high sensitivity over the entire search range of Higgs masses, different analyses, investigating all possible decay channels and their final states, need to be combined. This effort is crucial in the mass region beyond $m_{\mathrm{H}} \sim 2\mathrm{m_Z}$, which is not explored at other colliders such as the Large Electron-Positron collider (LEP) and Tevatron.

Among the possible ZZ decays, the semileptonic channel $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\overline{q}}$ (ℓ = e, μ), with a Z decaying to a lepton pair and the other Z decaying to a quark pair (Figure 3.1), combines large signal rate and a final state that can be fully reconstructed. The branching ratio of this channel is the largest among all Higgs decay modes considered at the CMS, and it is more than 20 times larger than the leptonic one. The kinematic of the event can be fully reconstructed because all four objects in the final state are detectable: two leptons ℓ and two jets originated from the hadronization of q and $\overline{q}$.

Although the decay rate for $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\overline{q}}$ is so copious with respect to the four-lepton channel, background contamination for this channel is much larger, due the very high rate at which jets are produced at the LHC.

All processes with two high transverse momentum and oppositely charged leptons (electrons or muons) in the final state associated with two hard jets constitute a source of background for this analysis. The signal signature is characterized by a resonant dilepton pair and a dijet systems originated from the decay of the two Z. The requirement of a lepton pair and a jet pair both compatible with Z bosons reduces the amount of background.

The production of Z bosons associated with QCD jets represents the most significant contribution to background. Such kind of events are difficult to discern from signal. Nevertheless, there are some features of signal that can be exploited to further reject Z + jets events. Signal-like candidates

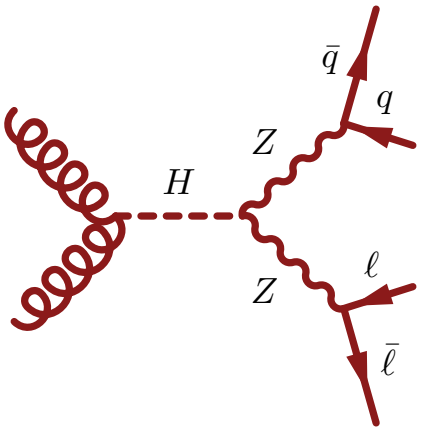


Figure 3.1: Decay diagram of a Higgs boson produced via gluon gluon fusion into two Z bosons and subsequently into a pair of leptons and a pair of jets.

reconstructed from this process are not resonant in diboson spectrum and, moreover, the origin of jets from Z decay is democratically distributed among u, d, s, c and b quarks, while background jets are mostly produced from hadronization of u and d quarks and gluons.

Even events involving top quarks can reproduce signal-like signature. The semileptonic decay of $\mathrm{t\bar{t}}$ to leptons of the same flavour can be identified as Higgs candidates:

$$\mathrm{t\bar{t}} \rightarrow (\mathrm{W}^+ \rightarrow \ell^+\nu)\mathrm{b}\ (\mathrm{W}^- \rightarrow \ell^-\overline{\nu})\overline{\mathrm{b}}$$

These events are characterized by the presence of missing transverse energy in the final state while leptons and jets do not tend to have invariant mass close to the Z mass. Therefore a constraint on reconstructed invariant mass of the lepton pair ($m_{\ell\ell}$) and jet pair (m_{jj}) contributes significantly to reject such kind of background events, along with a requirement on the missing transverse energy

Vector bosons pairs (ZZ, ZW and WW) are a source of background if leptons and jets are present in the decay. W decay contributes as well, because the resolution on jet four-momentum reconstruction is not good enough to discriminate between $\mathrm{Z} \rightarrow \mathrm{q\bar{q}}$ and $\mathrm{W} \rightarrow \mathrm{q\bar{q}}$. Even if the topology of these background events is very similar to the signal, the diboson system is not resonant and the cross section of these processes is small enough to allow to consider them of minor importance.

A further background suppression can be achieved considering spin correlations of the signal decay channel: a spin-0 particle (the Higgs boson) decaying into a pair of spin-1 particles (the Z bosons), which decay to fermions. The

spin correlation reflects in the angular distribution of the decay products which is absent in the background.

A search for a SM-like Higgs boson in the $H \to ZZ \to \ell^+\ell^-q\overline{q}$ channel is the subject of this thesis. The analysis has been performed on pp collision data collected by the CMS experiment during 2011 and 2012, respectively at a centre-of-mass energy $\sqrt{s} = 7$ and 8 TeV.

The strategy adopted to select and analyze the two datasets is almost identical except few minor changes. The analysis performed on the 7 TeV dataset was approved and published in the early 2012 [54]. The most recent analysis, performed on the 8 TeV dataset, includes some improvements with respect to the previous one.At the time in which this thesis has been written, the analysis of the 8 TeV dataset is not already officially approved by the CMS collaboration. Since the two versions of the analysis do not differ substantially, only the most recent one will be presented in details in the following sections. At the end of this Chapter differences among the two will be detailed.

3.1 Datasets

The analysis presented in this work is performed on pp collision data at centre-of-mass (c.m.) energy $\sqrt{s} = 8$ TeV, collected during the 2012 data-taking period by the CMS experiment. The dataset corresponds to an integrated luminosity of 5.1 fb^{-1}. The analysis of this channel has been performed also on data collected during 2011, at a c.m.e. $\sqrt{s} = 7$ TeV. An integrated luminosity of 4.9 fb^{-1} was analyzed, and the results of this analysis are reported in section 3.8. Blocks of data for each data acquisition are selected according to the standard CMS selection: only data recorded under good conditions of the different sub-detectors are certified as "good". Only data declared good are used in this analysis.

The analysis relies on Primary Datasets (PDs) centrally produced. The definition of PDs is based on the trigger decision. Events are stored in a PD if they fire the OR of HLT trigger paths that identify the specific dataset. Triggers characterized by the presence of at least two leptons define the *DoubleMu* and *DoubleElectron* PD, two muons in the first case and two electrons in the second case.

In each of these datasets there is at least one un-prescaled trigger with looser requirements than the offline selection, described in the following sections and applied to select signal candidates. Dimuon and dielectron events from the *DoubleMu* and *DoubleElectron* PD, which satisfy the un prescaled trigger with the lowest-threshold on the lepton transverse momentum for the specific

dataset, are considered.

Dimuon events are required to fire at least one of two HLT paths, named: HLT_Mu17_Mu8 and HLT_Mu17_TkMu8.

Both paths require the presence of two muon candidates, reconstructed at the HLT level. In particular, the HLT_Mu17_Mu8 path requires two *global muons*, i.e. reconstructed from the association of a track in the tracker and a track in the muon system; with $p_{\mathrm{T}} > 17$ GeV and 8 GeV, respectively. The HLT_Mu17_TkMu8, instead, path requires just one global muon and with p_{T} greater than 17 GeV and one tracker muon [42] with p_{T} larger than 8 GeV.

The double electron trigger is the HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_ Ele8_ CaloIdT_ -TrkIdVL_CaloIsoVL_TrkIsoVL path.

This HLT trigger path requires the presence of two HLT electron objects in the event with transverse energy, E_{T}, greater than 17 and 8 GeV, respectively. In order to reduce the rate of fake electrons, candidates are selected only if they pass loose requirements on the electromagnetic calorimeter shower shape and very loose isolation requirements [55].

3.2 Simulated samples

Monte Carlo (MC) simulated samples from the official CMS production are used in this analysis in order to study properties of the SM Higgs boson signal and of the relevant background processes.

The dominant background processes, Z + jets and $\mathrm{t\bar{t}}$, have been generated with MadGraph, a Next-to-Leading-Order (LO) matrix element generator [56]. The background from SM diboson events is simulated with Pythia generator (background simulated samples used in the analysis are listed in Table A.1 in Appendix A).

The $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\bar{q}}$ signal MC samples ($\ell = \mathrm{e}, \mu, \tau$) are generated with the PowHeg [57, 58, 59] event generator which contains NLO calculations and correctly describes spin correlations for the Higgs decay chain (simulated signal samples used in the analysis are listed in Table A.2 in Appendix A). The SM Higgs cross sections and the $\mathrm{H} \to \mathrm{ZZ}$ branching fractions are provided by the LHC Higgs Cross Section Working Group [21, 22], while $\mathrm{BR}(\mathrm{Z} \to \ell^+\ell^-)$ and $\mathrm{BR}(\mathrm{Z} \to \mathrm{q\bar{q}})$ are taken from the Particle Data Group (PDG) [2].

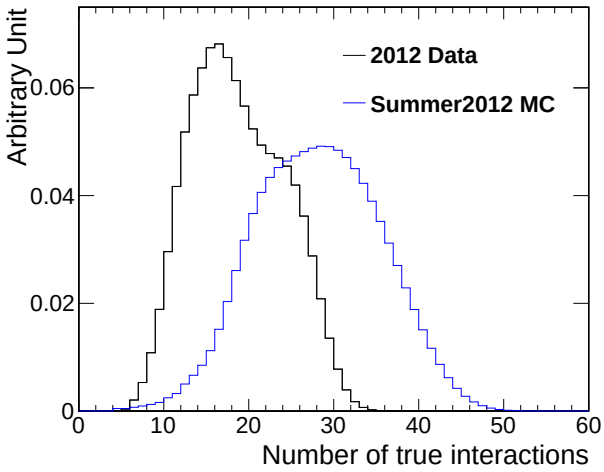


Figure 3.2: Number of true interactions in 2012 data and simulated samples (Summer2012 MC).

3.3 Pile-up reweighting

The probability to have multiple pp interactions overlapping the event of interest is not negligible at high luminosity conditions. The presence of secondary interactions is know as *pile up*. Simulated data have been generated taking into account pile up conditions different from ones observed in data. The distribution of the number of vertexes in data and simulation are shown in Figure 3.2. In order to overcome the mis-match, scale factors are used to reweight the simulated events. The distribution of reconstructed primary vertexes for data and simulation before (left) and after (right) the re-weighting is reported in Figures 3.3 and 3.4 for both the electron and the muon channel (up and down). In the following sections pile-up reweighting is applied to all simulated samples to match the data.

3.4 Object identification and selection

The reconstruction of the Higgs decay products ($Z \to e^- e^+$, $Z \to \mu^- \mu^+$ and $Z \to q\overline{q}$) is performed looking for pairs of leptons and jets in the event compatible with the signal signature, this means high-p_T and oppositely-charged leptons (muons or electrons) and hard jets. Lepton and jet candidates are selected if they satisfy specific kinematic and quality criteria.

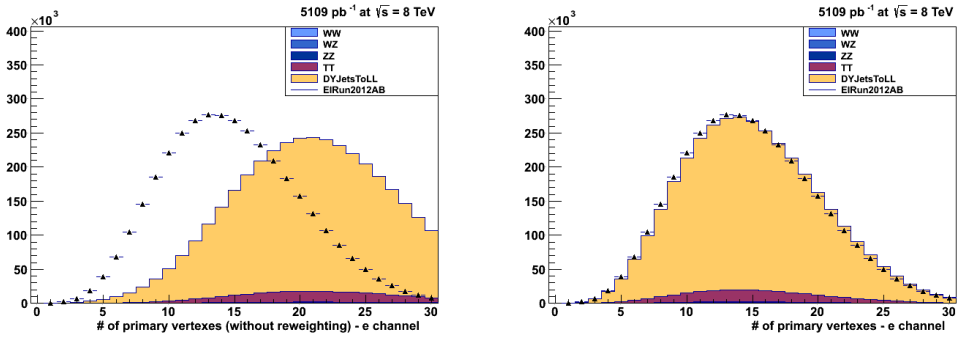


Figure 3.3: Number of reconstructed primary vertexes before (left) and after (right) re-weighting the simulated samples for the electron channel. Pre-selection is applied on both 8 TeV data (dots) and simulated background samples (filled histograms).

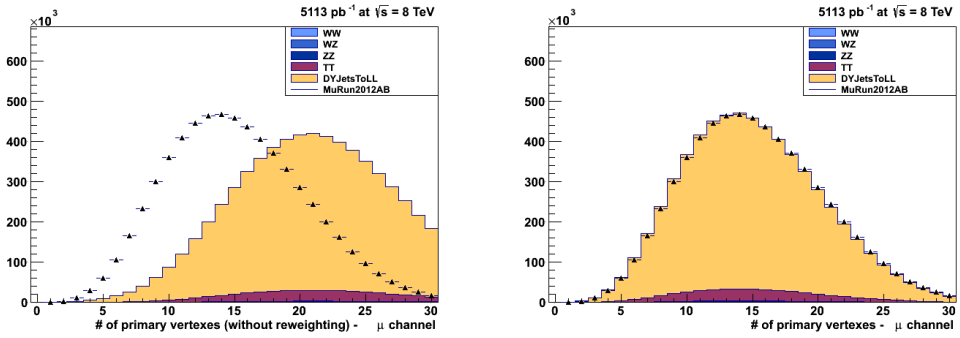


Figure 3.4: Number of reconstructed primary vertexes before (left) and after (right) re-weighting the simulated samples for the muon channel. Pre-selection is applied on both 8 TeV data (dots) and simulated background samples (filled histograms).

3.4.1 Muons

In the analysis only muons reconstructed as global muons are considered. Moreover they are required to be recognized as muons also by the Particle Flow (PF) algorithm, to ensure a good quality of the reconstruction. Further quality requirements are applied:

- Normalized χ^2 of the global muon track fit <10;
- At least one muon chamber hit included in the global muon track fit;
- The global muon track must have muon segments in at least two muon stations;

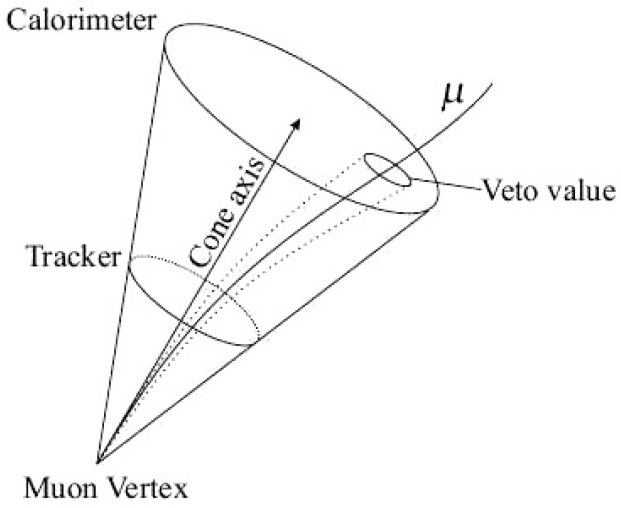


Figure 3.5: The isolation cone is built around the particle trajectory propagated from the tracker to the calorimeters. The contribution from the particle itself, contained within the veto cone, is subtracted in the variable definition.

- The tracker track has to be reconstructed from at least 5 tracker layers with hits;
- At least one hit must be present in the pixel detector;
- The muon track reconstructed in the tracker must have a distance to the primary vertex smaller than 2 mm in the transverse plane and smaller than 5 mm in the longitudinal direction;

An important source of background are muons produced in jets as result of QCD processes. It is possible to distinguish between muons produced in jets and muons coming from heavy particle decays (Z or W) selecting isolated muons, i.e. muons that do not have nearby particles, and whose energy deposit in its vicinity is below a certain threshold.

The variable used to discriminate isolated muons is the PF isolation, I_{PF}, defined as the sum of the transverse momentum p_{T}, or the transverse energy E_{T}, of reconstructed PF particles computed in a cone of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ around the lepton direction (Figure 3.5). Different kind of reconstructed PF particles contribute to the isolation variable: I_{ch}, charged hadrons originated from primary vertex, I_{nh}, neutral hadrons and I_{ph}, photons.

Pile up causes a mean energy deposited in the detector which is not due to particles of the event and that contaminates the energy deposition measured

Table 3.1: Muon identification requirements.

Variable	Requirement
isGlobalMuon	True
isPFMuon	True
$\chi^2/ndof$ (global fit)	< 10
Muon chamber hits in global fit	> 0
Muon stations with muon segments	> 1
d_{xy} (from tracker, wrt primary vertex)	< 2mm
d_z (from tracker, wrt primary vertex)	< 5mm
Valid pixel hits (tracker track)	> 0
Tracker layers with hits	> 5
$I_{PF,\,corr}/p_{\mathrm{T}}$	< 0.12

in given cone. As the instantaneous luminosity rises the pile up increases, thus the mean value of isolation increases as well. The isolation variable is therefore sensitive to the pile up. In this analysis an isolation variable corrected for the pile up effect is used in order to ensure its robustness on the number of pile up interactions. The effect is reduced in the tracker thanks to the requirement on the tracks within the cone to be originated from the primary vertex (I_{ch}^{noPU}). The method applied for controlling the pile up effect on calorimeter isolation uses the energy density in the event as calculated with the FastJet [60, 61] algorithm. The mean pile up contribution within the isolation cone is given by:

$$\rho \cdot A_{\mathrm{eff}} \tag{3.1}$$

where ρ is the mean energy density in the event and A_{eff} is the area of the cone in the (η, ϕ) space corrected by a factor which take into account the correlation between the isolation I and the density ρ. The corrected PF isolation definition is:

$$I_{PF} = I_{ch}^{noPU} + (I_{nh} + I_{ph} - \rho \cdot A_{\mathrm{eff}}) \tag{3.2}$$

The isolation cone for muons is defined using $\Delta R < 0.4$ and the requirement on the relative PF isolation, I_{PF}/p_{T}, is 0.12. Identification and isolation criteria for muons are summarized in Table 3.1.

Further kinematic requirements are imposed:

- $p_{\mathrm{T},1} > 40$ GeV, $p_{\mathrm{T},2} > 20$ GeV
- $|\eta| < 2.4$

Where 1 and 2 are the leading and subleading muons.

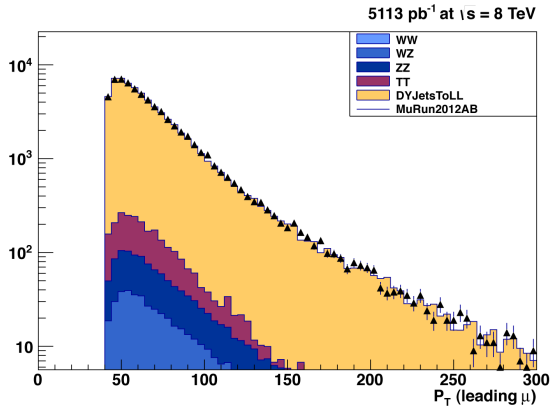


Figure 3.6: Comparison between 8 TeV data (dots) and simulation (filled histograms) for the distribution of the transverse momentum of the leading muon.

3.4.2 Electrons

Electron candidates are reconstructed from the association of an ECAL supercluster [44, 45] (SC) to track seeds found in the pixel detectors. In order to ensure a good electron reconstruction the $|\eta|$ of the electron supercluster must be inside the ECAL acceptance volume, $|\eta| < 2.5$, and outside the ECAL barrel-endcap overlap region, $1.4442 < |\eta| < 1.566$. As muon candidates, electron candidates are required to have high transverse momentum: p_{T} of the leading and sub-leading leptons are required to be greater than 40 GeV and 20 GeV. Furthermore electrons must have opposite charge and satisfy proper identification criteria, photon conversion rejection criteria and isolation selection as listed in Table 3.2.

The electron identification variables found to be the most discriminating are:

- $\Delta\eta_{trk,SC}$ and $\Delta\phi_{trk,SC}$: the measurements of the spacial matching between the track and the supercluster in η and ϕ respectively;
- $\sigma_{i\eta,i\eta}$: a variable related to the calorimeter shower shape, measuring the width of the ECAL supercluster along the η direction computed for all the crystals in the 5×5 block of crystals centred on the highest energy crystal of the seed supercluster;
- H/E: the ratio between the energy deposit recorded in the HCAL tower behind the electromagnetic seed supercluster and the energy of the seed supercluster.

Table 3.2: Electron identification requirements.

Variable	Barrel	Endcap
$\Delta\eta_{trk,SC}$	< 0.007	< 0.009
$\Delta\phi_{trk,SC}$	< 0.15	< 0.1
$\sigma_{i\eta,i\eta}$	< 0.01	< 0.03
H/E	< 0.12	< 0.10
d_0 (wrt primary vertex)	< 0.2mm	< 0.2mm
d_z (wrt primary vertex)	< 2mm	< 2mm
$\lvert 1/E - 1/p \rvert$	< 0.05	< 0.05
$I_{PF,\ corr}/p_{\mathrm{T}}$	< 0.15	< 0.15
Missing hits	≤ 1	≤ 1
Conversion vertex fit prob.	$< 10^{-6}$	$< 10^{-6}$

The rate of high-energy photons converting to electrons ($\gamma \rightarrow \mathrm{e^+e^-}$) and crossing the material in front of the electromagnetic calorimeter is high and, in order to avoid selecting electrons from photon conversion, a dedicated set of cuts is used:

- d_{xy} and d_z are the transverse and longitudinal impact parameters computed with respect to the reconstructed vertex;
- $|1/E - 1/p|$, the absolute value of the difference of the inverse of energy named in the ECAL and inverse of momentum p named in the tracker;
- The number of *missing hits* in the back-propagation of the track to the beam line;
- The conversion-vertex fit probability.

The same algorithm used to define the isolation of muon candidates is applied to electron objects as well. The size of the cone for electrons is $\Delta R = 0.3$ and the requirement applied to consider an electron as isolated is $I_{PF}/p_{\mathrm{T}} < 0.15$.

3.4.3 Lepton identification efficiency

Lepton identification efficiencies (mentioned in the previous paragraph and summarized in Tables 3.1 and 3.2) are evaluated using the tag and probe technique [62]. This method uses a pure sample of $\mathrm{Z} \rightarrow \ell^+\ell^-$ events and requires the reconstruction of a dilepton system compatible with a Z boson

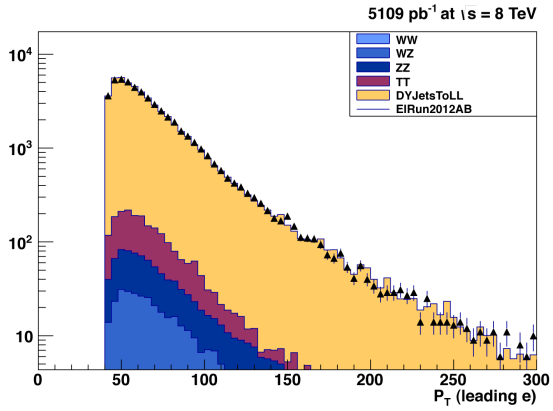


Figure 3.7: Comparison between 8 TeV data (dots) and simulation (filled histograms) for the distribution of the transverse momentum of the leading electron.

decay hypothesis and, therefore, with invariant mass falling in the range [60, 120] GeV. One of the two leptons is required to pass the full selection criteria and to match with the leg of the trigger with tighter request on the p_{T}, and called a *tag*. The other lepton, called a *probe*, is selected with criteria which depend on efficiency being measured. The selected dilepton systems are categorized in two exclusive samples according to whether the probe has passed or not the selection criteria under investigation. Due to the presence of background, a fit is performed to the invariant mass distribution of the dilepton system to obtain the signal yields in the two categories. The measured efficiency is measured as the relative signal yield in subsamples with passing or failing probes. This procedure is performed separately in different bins of probe pseudo-rapidity (η) and transverse momentum (p_{T}). Thus, the efficiency is obtained as function of p_{T} and η of the probe lepton.
Identification efficiency is not perfectly reproduced in simulation, therefore, data/ simulation scale factors are deduced by dividing efficiencies in data to the one obtained from simulation using exactly the same procedure described above. Tables 3.3 and 3.4 report the scale factors for electron and muon selection requirements. Simulated distributions are corrected by these scale factors.

3.4.4 Jets

The two quarks in the final state cannot be detected directly: they hadronize and the products of the hadronization process are identified and reconstructed as jets. PF jets are reconstructed with the anti-k_t algorithm [46]

Table 3.3: Data to simulation scale factors for electron identification selection.

p_T	$0.0 < \|\eta\| < 0.8$	$0.8 < \|\eta\| < 1.442$	$1.556 < \|\eta\| < 2.0$	$2.0 < \|\eta\| < 2.5$
20÷30	1.017 ± 0.004	0.998 ± 0.005	1.009 ± 0.008	1.102 ± 0.010
30÷40	1.019 ± 0.002	1.008 ± 0.002	1.010 ± 0.004	1.066 ± 0.005
40÷50	1.016 ± 0.001	1.002 ± 0.002	1.009 ± 0.003	1.040 ± 0.004
50÷200	1.005 ± 0.002	0.993 ± 0.003	1.003 ± 0.005	1.019 ± 0.007

Table 3.4: Data to simulation scale factors for muon identification selection.

p_T	$0.0 < \|\eta\| < 0.8$	$0.8 < \|\eta\| < 2.1$	$2.1 < \|\eta\| < 2.4$
20.00÷40.00	1.00425 ± 0.00042	1.00740 ± 0.00046	1.0216 ± 0.0014
40.00÷100.00	1.00119 ± 0.00039	1.00425 ± 0.00039	1.0140 ± 0.0014

with radius parameter set to $R = 0.5$. Jets are required to be inside the tracker acceptance, $|\eta| < 2.4$, thus allowing high reconstruction efficiency and precise energy measurements using PF techniques. Jet-energy corrections are applied to data and simulation as explained in [63]. Pile up (PU) energy is accounted for using the FastJet algorithm at the L1-correction level.

The amount of jets which originate from PU interactions can be reduced taking into account the fact that they do not originate from primary vertex. The variable β is defined as the sum of transverse momenta carried by charged particles in the jet, normalized to the total sum of transverse momenta of all charged particles in the jet. Only jets with $\beta \geq 0.2$ are selected. Figure 3.8 show the comparison between 8 TeV data and background expectation.

All jets in the event passing this selection are combined to reconstruct $\mathrm{Z} \to \mathrm{q\overline{q}}$ candidates from jet-jet pairs. In order to reject fake candidates made by low-p_{T} jets from QCD background, both jets must have $p_{\mathrm{T}} \geq 30$ GeV.

Kinematic Fit to decay chain

The search for a Higgs boson in this channel is performed analyzing the invariant mass spectrum of the diboson system looking for a resonant structure. The presence of signal is expected to appear as a peak centred at the Higgs boson mass in the distribution of the reconstructed $m_{\ell\ell\ jj}$.

The width of the Higgs mass peak depends on the intrinsic width of the

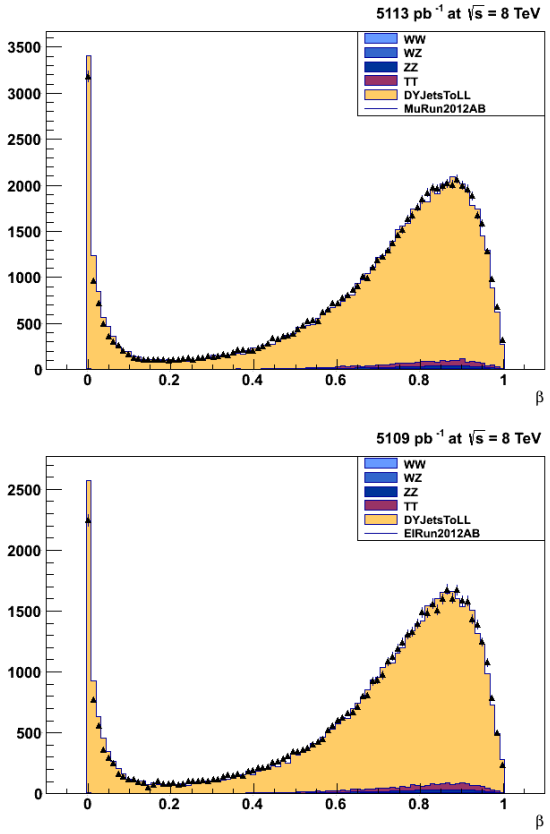


Figure 3.8: Comparison between 8 TeV data (dots) and simulation (filled histograms) for the distribution of variable β in the muon (left) and electron (right) channel.

signal and on the effect of detector resolution.

Jet energy resolution is the dominant cause of the spread at the dijet invariant mass, m_{jj}, and consequently in diboson invariant mass m_{ZZ} . The two variables are highly correlated, as can be seen in Figure 3.10, left side.

However, the jet energy resolution can be corrected exploiting an additional information on signal: jets are stemmed from a Z boson and the dijet invariant mass should correspond to the Z. This correction is expected to improve the resolution on the Higgs invariant mass. For what concerns the background, this assumption introduces a constraint which is not correlated to the underlying physics process, therefore the effect on the background is just to shuffle randomly the events in the diboson invariant mass spectrum.

In order to scale the dijet four-momentum to the Z boson mass, a kinematic fit of the two jets is performed. The fit takes as input the two jets

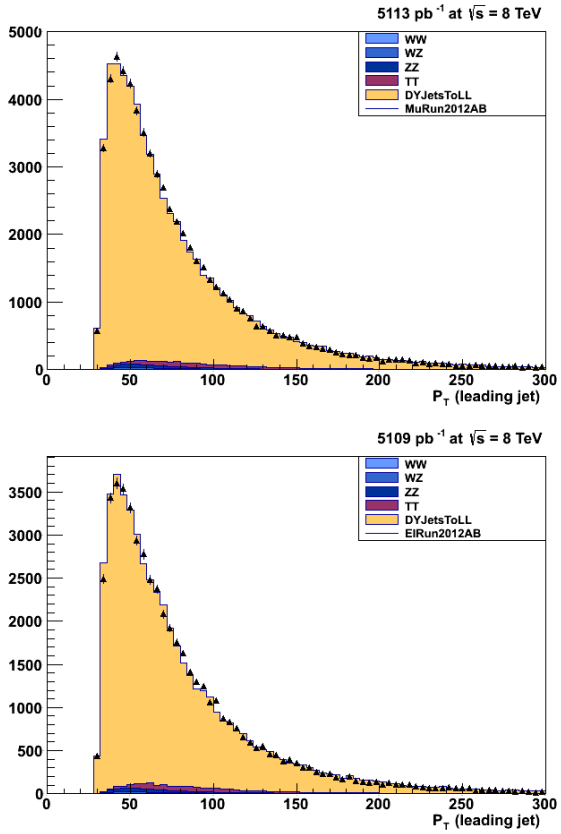


Figure 3.9: Comparison between 8 TeV data (dots) and simulation (filled histograms) for the distribution of the leading-jet transverse momentum in the muon (top) and electron (bottom) channel.

four-momenta and takes into account the prior knowledge on jet transverse momentum and position resolutions that are function of p_{T} and η. Indeed, jets with higher energies are reconstructed with better resolution with respect to jets with lower energy and the uncertainty on jet reconstruction is worse in specific detector regions. The jet four-momenta are then modified according to their resolution to constraint the dijet mass to the value of the Z boson mass. This procedure significantly improves the signal invariant mass resolution and remove the m_{jj}– m_{ZZ} correlation (Figure 3.10, right side) because it minimizes the dependence of the diboson invariant mass on the jet resolusions. This feature facilitates the definition of signal and sideband regions in the m_{jj} spectrum, because in this case it is possible to identify signal region by means of a rectangular cut (Section 3.5).

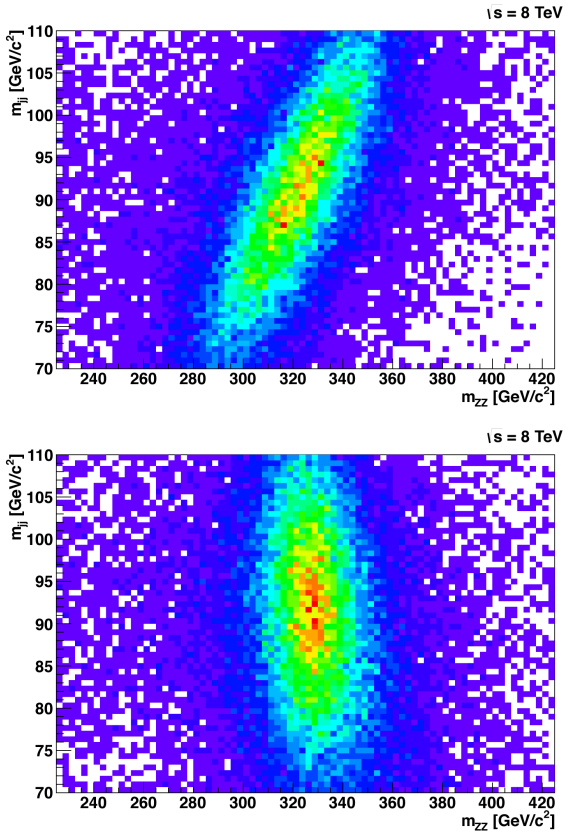


Figure 3.10: Kinematic fit results. dijet invariant mass vs. diboson invariant mass for Higgs candidates (simulated signal, $m_H = 325$ GeV) afterpreselection. Top: before kinematic fit; bottom: after kinematic fit.

3.4.5 Identification of b jets

The identification of the jet parton type provides a powerful tool for background discrimination. Jets in signal events are produced in hadronic decays of a Z boson, and therefore originate from the hadronization of quarks. The flavor of quarks in Z decays is almost equally distributed among the five types of quarks, d, u, s, c, b, with some preference given to the down-type quarks because of preferential electroweak couplings to the Z. The dominant background is represented by events containing a leptonically-decaying Z boson produced in association with high-p_{T} jets. In this process gluon radiation and light-quark hadronization are expected to play a major role. After gluons, the u and d quarks dominate the jet production associated with the Z. This is due to the fact that u and d are valence parton of the protons, therefore, the main handles to discriminate signal from background are the

relatively large contribution of heavy flavor quarks and the absence of gluons in signal events. The strategy used in this work to enhance the sensitivity of the analysis takes advantage of this feature by performing identification of the b flavor.

The discrimination of heavy-flavour jets from jets originating from gluons (g) and light-flavour quarks (u, d, s) can be obtained considering the properties of the bottom and the charm quarks, such as their hard fragmentation functions and the relatively large mass and long lifetime of the heavy flavour hadrons. Good performance of the CMS b-jet identification, *b tagging* is due to the precise charged particle tracking system and robust lepton identification of the experiment.

To identify heavy-flavour jets in the events the CMS Jet Probability (JP) tagging algorithm [64] is used. The JP tagger exploits the long lifetime of beauty hadrons and is based on the compatibility of tracks associated with the jet to come from the primary vertex. These probabilities are combined to provide jet probability. If the probability to come from the primary vertex is low, then the jet is likely to be a b jet.

The JP algorithm yields a single discriminator value for each jet (the distribution of the JP discriminator for data and background expectation is shown in Figure 3.11). To discriminate jets coming from heavy-flavour quarks and light-flavour quarks, two working points corresponding to two different thresholds on the JP discriminator are used: loose (L) and medium (M) working points which corresponds to a nominal misidentification probability for light-flavour jets of 10% and 1% respectively. From this point in this paper tagging criteria will be labelled appending the letter relative to the working point to the acronym of the tagger: JPL and JPM.

The tagging requirement is used to classify Higgs boson candidates according to the flavour of the jets in the final state. This classification, detailed in Section 3.5.2, allows to enhance the sensitivity to signal.

3.5 Final selection

3.5.1 Higgs candidates reconstruction

Leptons and jets in the event are selected if they satisfy the requirements described in Section 3.1, and are than combined to reconstruct leptonic and hadronic Z bosons. For each event, the selected dijet and dilepton pairs are finally assembled to reconstruct Higgs candidates. The event display of a Higgs boson candidate decaying to a pair of electrons and two jets reconstructed by the CMS experiment is reported in Figure 3.12, it gives a

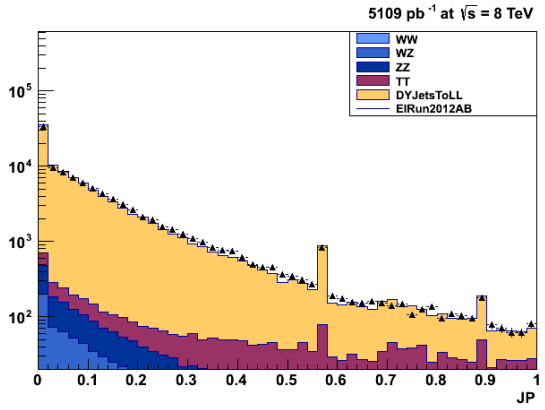


Figure 3.11: Comparison between 8 TeV data (dots) and simulation (filled histograms) for the distribution of the JP discriminator.

good idea of how a Higgs candidate can be reconstructed in the detector.

Events are identified as signal event candidates if the dilepton invariant mass spectrum, $m_{\ell\ell}$, is in the range 70÷110 GeV. The dilepton invariant mass for the selected $Z \to \ell^+\ell^-$ candidates is shown in Fig. 3.13. In the rest of this work, the entire selection procedure described above, including the requirement on $m_{\ell\ell}$, will be referred as *preselection* of Higgs boson candidates.

The dominant background for this channel is represented by the production of a Z boson decaying leptonically in association with jets. The Z is real and a hard cut on the dilepton invariant mass is not a powerful handle against such kind of events. Indeeed, a more tight cut on $m_{\ell\ell}$ would imply a reduction of signal yield without improving background rejection. The distribution of the dijet invariant mass m_{jj} for candidates reconstructed in the electron and muon channels is shown in Fig. 3.14. A relevant help in background rejection is given by setting a more stringent requirement on m_{jj}. Only signal candidates which have a pair of jets with m_{jj} in the range 75÷105 GeV are selected as signal candidates. The definition of the interval is compatible with the dijet mass resolution. All events passing the requirement on m_{jj} are considered as signal-like candidates, while those not satisfying this criterium are used for populating sidebands, as will be described in the Section 3.6.

3.5.2 Categorization

The preselected events are split into three exclusive categories according to the number of jets identified as b jet with the JP tagging algorithm. Events are then classified in the following categories:

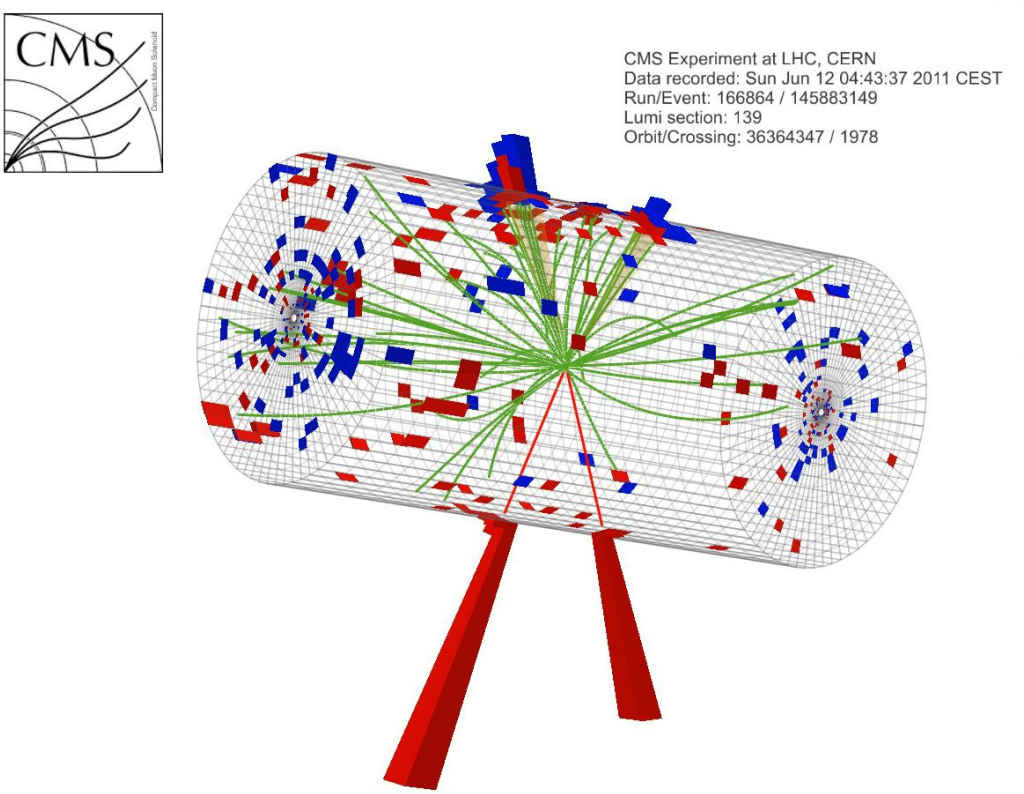


Figure 3.12: Event display of a typical Higgs event candidate reconstructed with the CMS detector with two electrons and two jets in the final state, $m_{ZZ} = 580$ GeV.

- *2 b-tag*: events with at least two jets identified as originating from b quark hadronization;
- *1 b-tag*: events in which one and only jet is identified as b jet.
- *0 b-tag*: events failing the requirements for falling in the previous categories are classified as no-tag events.

The three categories are built in order to be mutually exclusive. Events falling in the 2 b-tag category are selected if at least one jet passes the medium requirements for the JP tagger and the other jet the loose one. All the events failing the selection requirements for the first category and that have at least one jet loose-tagged are placed in the second category, the 1 b-tag. The last category, the 0 b-tag, is populated from all events which do not satisfy previous criteria.

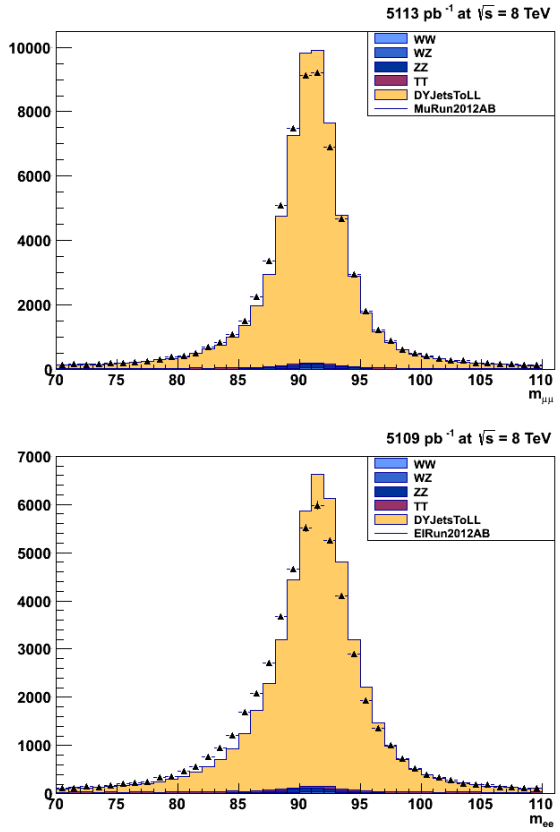


Figure 3.13: Dilepton invariant mass distribution of $Z \to \ell\ell$ candidates after preselection. Top: muon channel. Bottom: electron channel. 8 TeV data (dots) are compared with background expectation (filled histograms).

The signal-background composition of these three categories is quite different. In particular the 2 b-tag category is characterized by high purity, but low signal yield. The major backgrounds in this case are Z + heavy-flavour jets and $t\bar{t}$ processes. The 0 b-tag category has the highest signal yield, but it is also dominated by background, mainly coming from Z + light-flavour jets.

A difference between the b-tagging performances measured in data and the ones predicted by simulation is observed. In order to correct this effect, b-tagging scale factors, SF, are defined to correct the efficiency of the b-tag algorithm (the probability not to recognize a b-jet, even if it stemmed from a heavy-flavour quark) and the light-flavours/gluon jet mis-identification rate (the probability to mis-identificate a light-flavour or gluon jet as b jet). Efficiency and misidentification rate, labelled respectively SF_{hf} and SF_{lf}, are

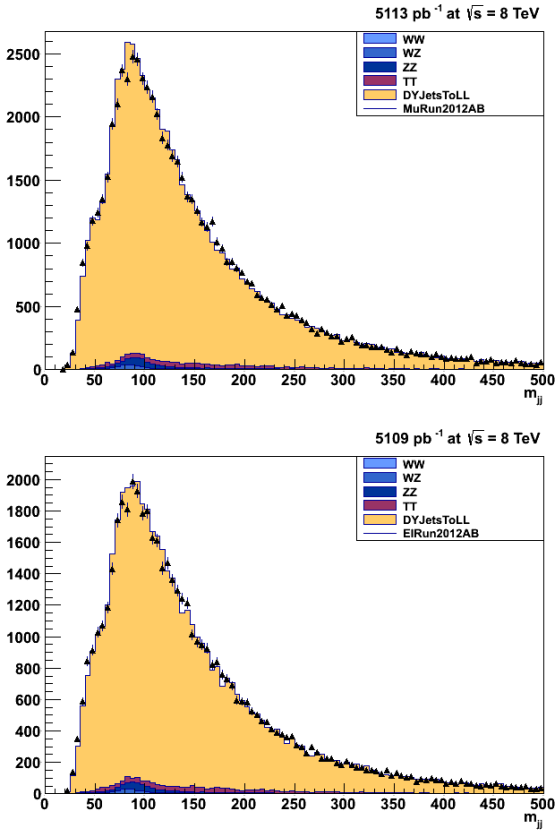


Figure 3.14: Dilepton invariant mass distribution of $Z \to jj$ candidates after preselection. Top: muon channel. Bottom: electron channel. 8 TeV data (dots) are compared with background expectation (filled histograms).

defined as the ratio between tagging efficiencies for high-flavour and light-flavour jets in data and simulation. In general these factors, as the relative efficiency measurements, are functions of the transverse momentum and pseudorapidity of a jet [64]. The SF_{hf} scale factors are in general smaller than one, this means that the efficiency measured for data is smaller than prediction, while SF_{lf} scale factors are greater than one, meaning that the probability to tag a light quark jet as b jet in data is larger than in the simulation.

The discrepancy in b tagging performances between data and simulation has to be corrected, and at the same time it must be taken into account that the classification of the events is based on jet flavour: both tagged and not-tagged jets are used in the analysis. In order to consider this scenario the correction is performed on stochastic base. Random numbers are generated

and used to degrade or upgrade the tagged jet in the simulation on a jet-by-jet base, altering the outcome of the b tagging decision according to the comparison of the random number and the percentage of jets that need to be upgraded and the one of jets that must be downgraded. In this way, jets which are matched to heavy quarks have a non zero probability to be downgraded to a lower b-tag category. This means a medium-tagged jet can be labelled as loose-tagged, or a loose-tagged jet can be considered as a not-b jet. Viceversa, light-quark and gluon jets can be promoted to a higher b-jet category according to the mis-tag rate scale factors.

The result of this process is the migration of the events from one category to an other, according to the outcome of the algorithm, and, therefore, a change in the population of each b-tag category.

3.5.3 Angular analysis

The $H \to ZZ \to \ell^+\ell^- q\overline{q}$ kinematics can be further exploited to discriminate it against the Z + jets background. The production of a heavy Higgs boson translates in a boosted kinematics of Z bosons. The transverse momentum of each Z boson tends to increase as the mass of the mother particle increases, and the ΔR spacing between the two jets decreases as the boost increases. This does not happen for background events in which the two jets are uncorrelated. These variables are correlated with the reconstructed diboson invariant mass.

Another feature that characterizes the $H \to ZZ \to \ell^+\ell^- q\overline{q}$ signal is the well defined spin correlation that affects angular distribution of decay products. It has been shown in Ref.s [65, 66] that, if the final state is fully reconstructed and the mass of the two spin-1 particles are fixed, five angular observables fully describe the kinematics in the decay $2 \to 1 \to 2 \to 4$ as in $ab \to X \to ZZ \to 2\ell 2j$, these observables are independent on the three invariant masses of the X and the two Z bosons. The above orthogonal observables are largely uncorrelated and are more attractive to be used in event selection rather than raw kinematic observables.

Figure 3.15 illustrates the angular distribution in the production and decay chain $pp \to H \to ZZ \to \ell^+\ell^- q\overline{q}$. The five angles that describe the angular distribution of decay products are three helicity angles θ_1, θ_2, and Φ, and two production angles θ^* and Φ_1, as shown in Fig. 3.15. More details can be found in [65, 67], where a parameterization of both signal and background distributions has been derived and implemented.

Here θ_i are the angle between the direction of the ℓ^- or q from the $Z \to \ell^+\ell^-$ or $Z \to q\overline{q}$ and the direction opposite the X in the Z rest frame, and Φ is the angle between the decay planes of the two Z systems. The two Z bosons are

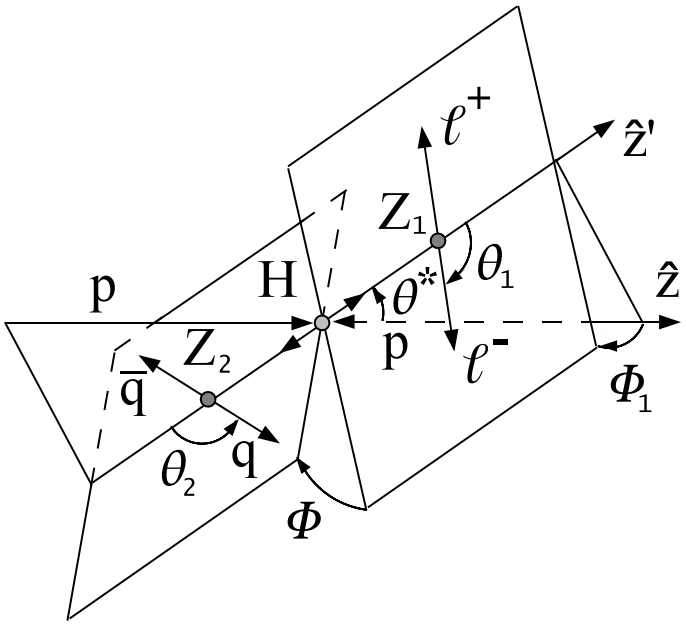


Figure 3.15: Diagram depicting the decay H $\rightarrow$ ZZ $\rightarrow \ell^+\ell^-\mathrm{q}\overline{\mathrm{q}}$ and the 5 angles which describe such a decay.

distinguished by their decay type or, in case their daughters are the same type of particles, such as in the case of the two quarks, by an arbitrary convention. The production angle θ^* is defined as the angle between the parton collision axis z and the decay axis in the rest frame. The fifth angle can be defined as Φ_1, the angle between the production plane and the Z_1 decay plane.

The spin correlation of the signal is the absent in the background and the angular distributions for signal and background are expected to be different. A comparison of angular distribution in data and simulation for electron and muon events, in background-dominated samples, can be found in Figures. 3.16 and 3.17, where good agreement is observed.

The difference in the final state angular distribution of signal and background events can be exploited to compute a likelihood discriminant (LD) in oder to perform a selection based on the angular these information. The likelihood discriminant is given by the following ratio

$$LD = \frac{P_{sig}}{P_{sig} + P_{bkg}}. \tag{3.3}$$

P_{sig} and P_{bkg} are the probability densities of the five helicity angles for signal and background respectively. The signal is most likely to have LD values close to one while the background is most likely to have values closer to zero.

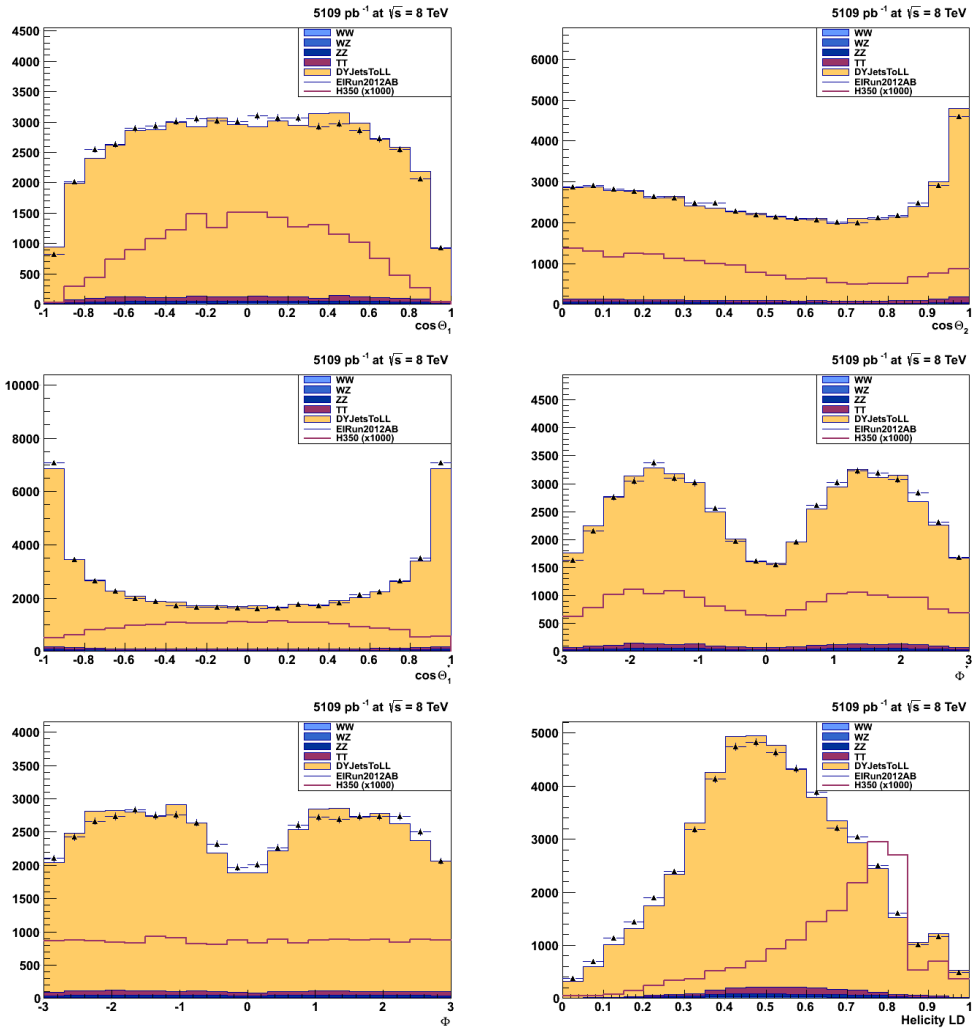


Figure 3.16: Five angular distributions of $\cos\theta_1, \cos\theta_2, \cos\theta^*, \Phi$, Φ_1 and the angular likelihood discriminant for 2012 electron data (dots) and simulated samples (histograms). Red line indicates the expected distribution for a Higgs boson with mass 350 GeV, multiplied by a factor of 1000 for illustration. Preselection is applied.

Events are then selected by requiring LD to be above a certain threshold. The helicity angles, in spite of kinematic variables, are largely decoupled from the mass variables, and by applying selections based on the helicity likelihood discriminant one can better preserve the shape of the ZZ invariant mass distribution of background with respect to tight kinematic cuts. The method for obtaining such a likelihood discriminant is described below.

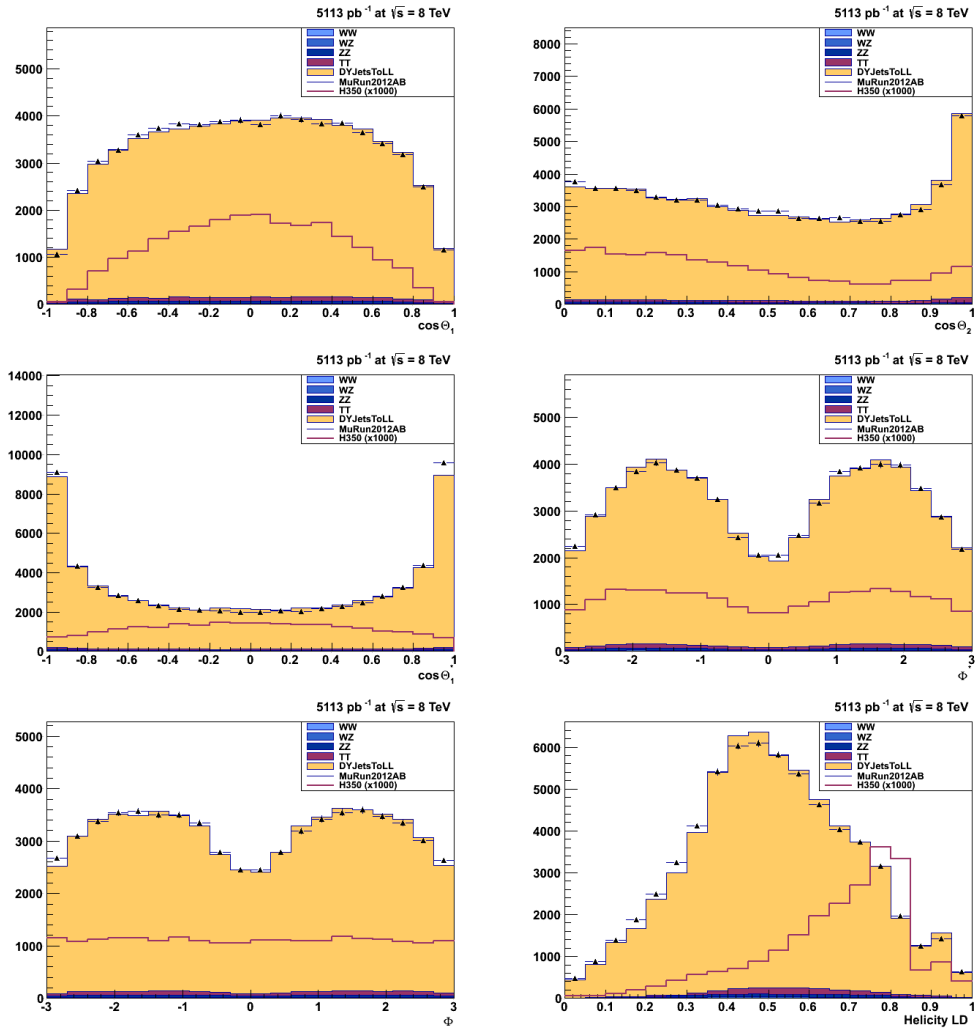


Figure 3.17: Five angular distributions of $\cos\theta_1, \cos\theta_2, \cos\theta^*, \Phi$, Φ_1 and the angular likelihood discriminant for 2012 muon data (dots) and simulated samples (histograms). Red line indicates the expected distribution for a Higgs boson with mass 350 GeV, multiplied by a factor of 1000 for illustration. Preselection is applied.

The probability distribution function for signal is taken to be a product of the ideal, fully correlated, distribution which is derived in Ref. [65] and a set of four one-dimensional acceptance functions.

$$\begin{aligned}\mathcal{P}_{\text{sig}} =& \mathcal{P}_{\text{ideal}}(\theta^*, \theta_1, \theta_2, \Phi, \Phi_1; m_{\text{ZZ}}) \cdot \mathcal{G}_{\theta^*}(\theta^*; m_{\text{ZZ}}) \cdot \mathcal{G}_{\theta_1}(\theta_1; m_{\text{ZZ}}) \\ & \cdot \mathcal{G}_{\theta_2}(\theta_2; m_{\text{ZZ}}) \cdot \mathcal{G}_{\Phi_1}(\Phi_1; m_{\text{ZZ}}) \end{aligned} \tag{3.4}$$

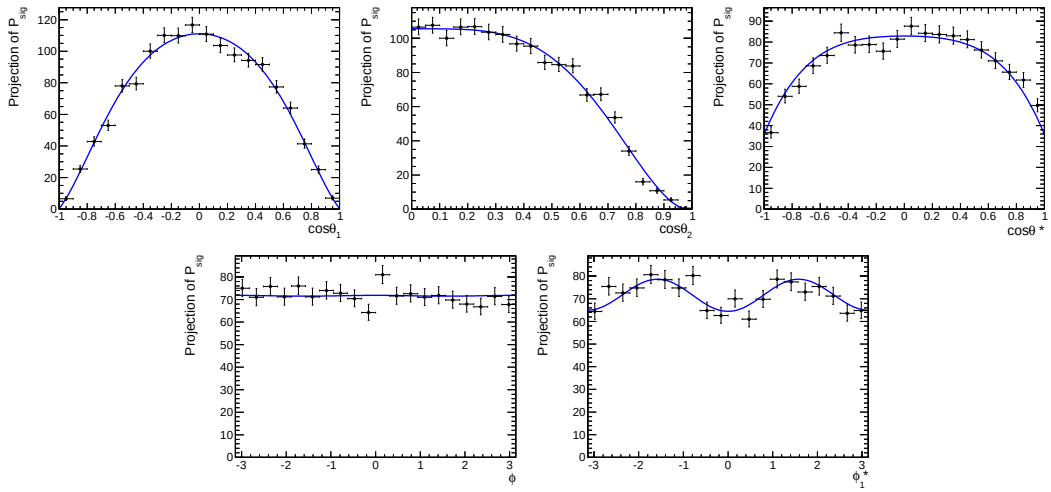


Figure 3.18: Distributions of $\cos\theta_1, \cos\theta_2, \cos\theta^*, \Phi$, and Φ_1 for a 500 GeV Higgs boson.

The four acceptance functions, $\mathcal{G}_{\theta^*}, \mathcal{G}_{\theta_2}, \mathcal{G}_{\theta_2}$, and $\mathcal{G}_{\Phi_1}$, have been obtained empirically from fits to 2011 simulation at $\sqrt{s} = 7$ TeV, in the work presented in the Ref. [68]. For the analyzed mass range, not significant difference is expected in the simulation at 8 TeV, therefore, the study performed on 2011 MC is used for 2012 analysis as well.

Distributions of the five angular variables for a Higgs boson signal of $m_{\mathrm{H}} = 500$ GeV are reported in Figure 3.18.

The ideal function, $\mathcal{P}_{\mathrm{ideal}}$, depends on the ZZ invariant mass, and the parameters of the four acceptance functions have been reparameterized in terms of m_{ZZ} only. This was done by fitting eight different simulated samples, each corresponding to a different Higgs mass hypothesis. The resulting parameters have been then fitted with either a linear or quadratic function of m_{ZZ} .

The probability distribution function for the background was approximated with a product of five one-dimensional functions:

$$\begin{aligned}\mathcal{P}_{\mathrm{bkg}}(\theta^*, \theta_1, \theta_2, \Phi, \Phi_1; m_{\mathrm{ZZ}}) =& \mathcal{P}_{\theta^*}(\theta^*; m_{\mathrm{ZZ}}) \cdot \mathcal{P}_{\theta_1}(\theta_1; m_{\mathrm{ZZ}}) \cdot \mathcal{P}_{\theta_2}(\theta_2; m_{\mathrm{ZZ}}) \\ & \cdot \mathcal{P}_{\Phi}(\Phi; m_{\mathrm{ZZ}}) \cdot \mathcal{P}_{\Phi_1}(\Phi_1; m_{\mathrm{ZZ}}) \end{aligned} \quad (3.5)$$

All functions were obtained empirically from fits to simulation [68] in 2011. As for the signal, the background shapes are not expected to change from

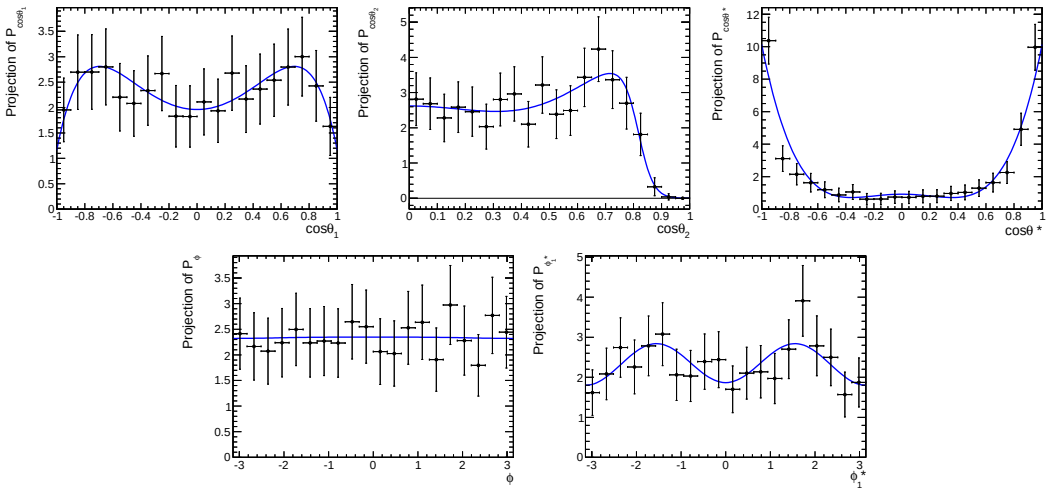


Figure 3.19: Distributions of $\cos\theta_1, \cos\theta_2, \cos\theta^*, \Phi$, and Φ_1 for background around 500 GeV.

2011 to 2012, therefore ones obtained from 2011 simulation have been used for 2012 analysis as well. Projections of $\mathcal{P}_{\text{bkg}}$ over the five variables can be found in Figure 3.19.

Similar to the case of $\mathcal{P}_{\text{sig}}$, the parameters of $\mathcal{P}_{\text{bkg}}$ were fit using either linear or quadratic functions of m_{ZZ} .

$\mathcal{P}_{\text{sig}}$ and $\mathcal{P}_{\text{bkg}}$ were combined as in equation (). The discriminant defined in equation 3.3 is a function of the five angles and parametrized by the reconstructed invariant mass for the event under study. The distribution for the likelihood discriminant is shown in Figure 3.16 for the electron channel and in Figure 3.17 for the muon channel, for both background (solid histograms) and signal (red line) simulation after the preselection and, therefore, before the classification into b-tag categories. There is a good agreement in the likelihood discriminant, LD, distribution between data and background simulation, as it is expected based on agreement of variables entering the LD calculation.

3.5.4 Missing transverse energy

Without further selection, in the 2 b-tag category the dominant background originates from $\text{t}\bar{\text{t}}$ events which contains two b-quark jets. Since the signature of $\text{H} \to \text{ZZ} \to \ell^+\ell^-\text{q}\bar{\text{q}}$ channel has also 2 high-p_{T} leptons in the final state, the semileptonic t-quark decay (with a pair of same flavour leptons) mainly contributes to background:

$$\mathrm{t\bar{t}} \rightarrow (\mathrm{W}^{+} \rightarrow \ell^{+}\nu)\mathrm{b}\ (\mathrm{W}^{-} \rightarrow \ell^{-}\overline{\nu})\overline{\mathrm{b}}$$

In order to reduce the contribution from this background source, it is useful to exploit a characteristic feature this kind of process: the presence of neutrinos in the final state, which are not detectable directly through the detector.

For each event, the total momentum in the transverse plane with respect to the beam line, must be conserved. Therefore the total transverse momentum must be balanced among all the particles produced in that collision. An imbalance is most likely a sign of weakly interacting particles escaping the detector, such as neutrinos. Therefore it is a clear indication of SM processes such as leptonic W decay or t quark decay, whereas signal events have no neutrinos in the final state.

The amount of energy resulting in an imbalance to the total transverse momentum is called Missing Transverse Energy, MET, determined as the vectorial sum of transverse energy:

$$\mathrm{MET} = \not{E}_{\mathrm{T}} = -\sum \overrightarrow{E}_{\mathrm{T_i}} \tag{3.6}$$

The MET significance is a quantity used to measure the probability that the measured MET is consistent with a with a balanced event with the effect of the finite resolution of the detector, which affects the reconstruction of PF candidates. The adopted variable is the logarithm of the ratio of two likelihood functions built on the hypothesis of true MET equal to the value measured with the PF algorithm and a null hypothesis which corresponds to the case in which true MET is equal to zero [69]:

$$2\ln\lambda(\not{E}_{\mathrm{T}}) = 2\ln\frac{\mathcal{L}(\not{E}_{\mathrm{T}}=\not{E}_{\mathrm{T}}^{meas})}{\mathcal{L}(\not{E}_{\mathrm{T}}=0)} \tag{3.7}$$

The likelihood ratio, λ, provides a measure of how genuine is the missing transverse energy measured in the event and can be used to discriminate "real" MET from the one caused by detector effects.

Because of the absence of neutrinos, signal events are not expected to be source of MET, while true MET arises from background events. The distribution of the $2\ln\lambda(\not{E}_{\mathrm{T}})$ for simulated background and data is shown in Figure 3.20. A loose requirement is applied on the MET significance in the 2 b-tag category: it must be less than 10. The requirement of this variable is placed in the tail of the distribution for signal, which makes the disagreement between data and background prediction from simulation to introduce a systematic uncertainty negligible compared to other sources.

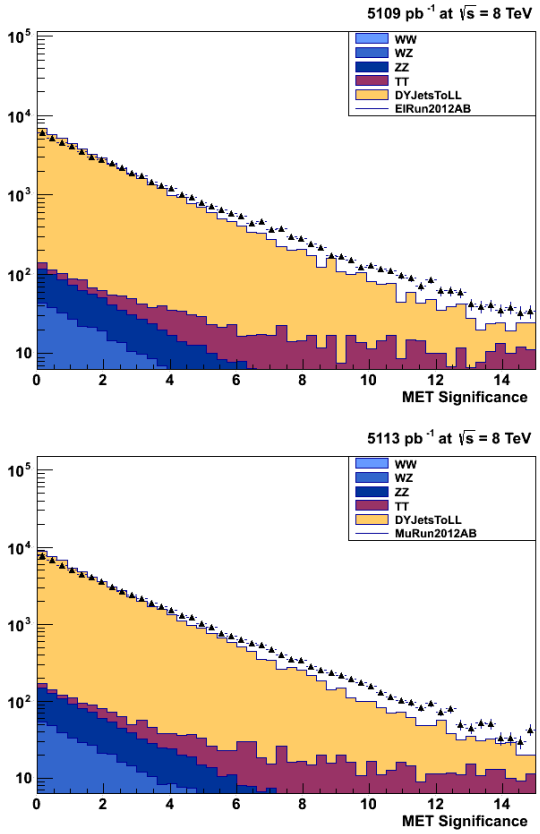


Figure 3.20: PF MET significance distributions. Comparison between data (dots) and simulation (filled histograms) for electron (top) and muon (bottom) channels. A requirement of PF MET significance less than 10 is demanded in the 2 b-tag category.

3.5.5 Summary of final selection

The composition of background is not the same in each category and in order to maximize the effectiveness of background rejection, the selection is optimized separately in each category. While the angular discrimination is applied in all of them, each with a different threshold, a cut on the MET significance is used only in the 2 b-tag category.

The kinematics depends on the Higgs boson mass, and the present analysis searches for a Higgs boson in the whole accessible Higgs mass range. In order to suppress background in an efficient way, different selections for different Higgs mass hypotheses should be applied. In order to simplify the selection and to assure the independence of the selection criteria on the mass of the searched particle, kinematics cuts which depend on the reconstructed

Table 3.5: Summary of optimized kinematic and topological selection requirements.

	preselection		
$p_{\mathrm{T}}(\ell^{\pm})$	lowest $p_{\mathrm{T}} > 20$ GeV, highest $p_{\mathrm{T}} > 40$ GeV		
p_{T}(jets)	> 30 GeV		
$\lvert\eta\rvert(\ell^{\pm})$	$\mathrm{e}^{\pm} < 2.5$, $\mu^{\pm} < 2.4$		
$\lvert\eta\rvert$(jets)	< 2.4		
	0 b-tag	1 b-tag	2 b-tag
b-tag	none	1L	1L & 1M
angular LD	$> 0.55 + 0.00025 \times m_{\mathrm{ZZ}}$	$> 0.302 + 0.000656 \times m_{\mathrm{ZZ}}$	> 0.5
$2ln\lambda(\not{E}_{\mathrm{T}})$	none	none	< 10
m_{jj}	$\in$ [75, 105] GeV		
$m_{\ell\ell}$	$\in$ [70, 110] GeV		
m_{ZZ}	$\in$ [18, 1000] GeV		

dijet+dilepton invariant mass, m_{ZZ} are applied. This strategy simplifies in a relevant way the analysis chain, because it implies observed and simulated background distributions of m_{ZZ} , for the selected candidates, are unique, do not change m_{H}. This feature implies also that the computation of the likelihood functions for background simulation and observed data must be done just once, independently on the specific m_{H} under study. This point will be discussed again in the last chapter, where the method used to compute upper limits on the Higgs production cross section will be explained.

The requirement on the angular LD is therefore optimized as a linear function of the reconstructed mass m_{ZZ} . This dependence varies with the b-tag category and is summarized in Table 3.5.

The optimization procedure was accomplished looking for the selection threshold which minimizes the exclusion upper limit and was performed on 2011 simulation for the three categories. In the 2 b-tag category a constant threshold is fixed to a value of 0.5.

Table 3.5 summarizes the entire selection from kinematic requirements to the categorization and angular discrimination, and Figure 3.21 shows the workflow of the analysis from the reconstruction of Higgs candidates from the four objects in the final state) to the final selection.

In general, the presence of more than two jets in signal events is not excluded, and it causes the reconstruction of more than one Higgs boson candidate in the event from the combination of the same objects. In order to avoid the reconstruction, and therefore the analysis, of more than one Higgs boson

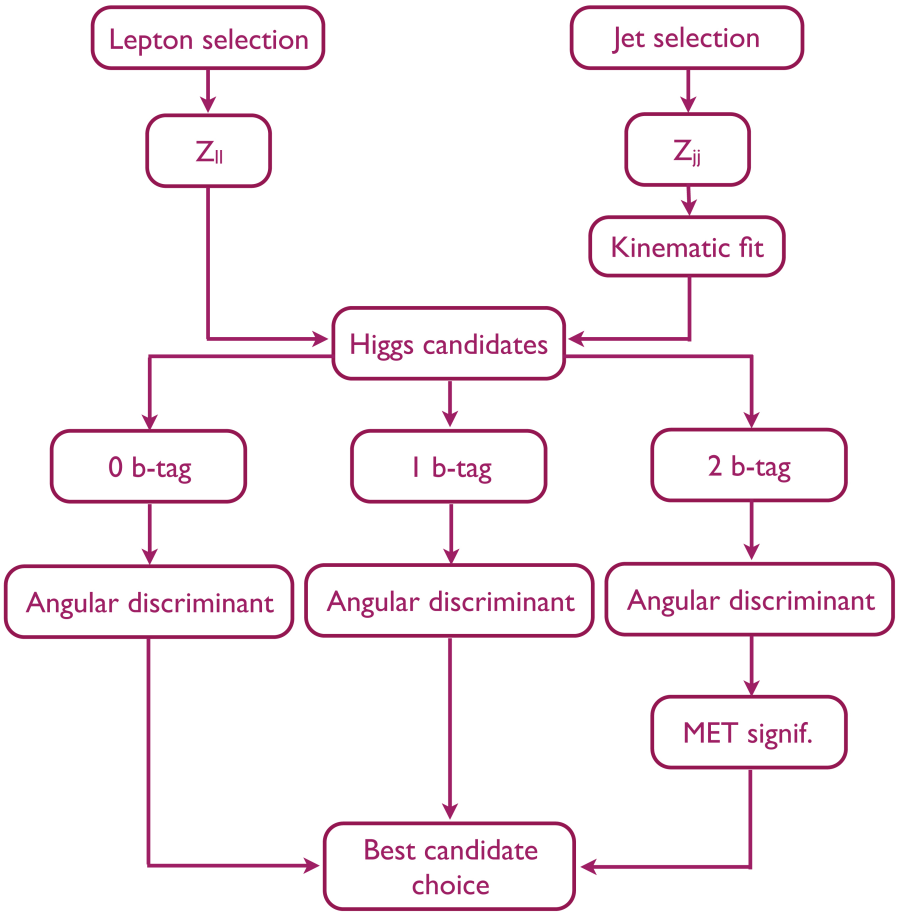


Figure 3.21: Workflow of the analysis. Z bosons reconstruction from 2 selected leptons and 2 selected jets is followed by the Higgs boson candidate reconstruction. An improvement in m_{ZZ} resolution is achieved applying a kinematic fit. Signal candidates are then classified in three exclusive categories according to the jet flavour in the final state. Further signal-background discrimination is achieved applying a cut on the angular discriminant. In the 2 b-tag category a requirement on the MET significance helps in rejection of $t\bar{t}$ events.

candidate per event, a best-candidate choice is performed to decide what is the candidate which is mostly like a signal-candidate, the Higgs candidate with the leptonic Z mass closest to the nominal Z boson mass value [2], is chosen. For each category the Higgs candidate with dilepton invariant mass

closer to the nominal Z boson mass [2] is chosen as the best candidate This is equivalent to minimize the function:

$$\chi = |m_{ll} - m_Z|$$

Once the ambiguity among multiple candidates in the same category is solved, a check is performed on the other categories, in order to have at the end just one candidate per event. The priority is given to the purest category and follows this order: 2 b tag, 1 b tag, 0 b tag.

If there are no candidates satisfying the requirements in the signal region for the event of interest, the same procedure is applied to candidate events selected in sidebands. The priority rules are the same.

The expected yields corresponding to 5 fb^{-1} of data of signal and background events in the full ZZ invariant mass range [0, 1000] GeV are listed in Table 3.6 for the electron channel and in Table 3.7 for the muon channel. Further discrimination of the signal and background will be achieved with the likelihood fit performed on the m_{ZZ} invariant mass distributions (Chapter 4)

3.5.6 Blind analysis

The final statistical analysis exploits the resonant feature of the signal in the dijet-dilepton invariant mass system. Therefore a further background discrimination is achieved using as discriminating variable the m_{ZZ} distribution. After the entire selection chain is applied on selected events, six exclusive distributions of this variables are isolated: three quark-flavour channels times two leptonic channels (μ and e). These distributions do not rely on the hypothetical Higgs boson mass, therefore both the simulated background distribution and the observed data distribution do not depend on the signal hypotheses under study. They are displayed in Figure 3.22 in the sideband region. To avoid a possible bias of the experimenter in the process of developing the analysis strategy and optimizing the selection requirements, a *blind analysis* is performed: data in the signal region, i.e. with m_{jj} within the range [75, 105] GeV were examined until the event selection criteria were settled. Indeed, the analysis strategy was decided looking at the simulated events. Only when data distributions in signal-depleted regions defined by the sidebands of m_{jj} distribution, confirmed the robustness of the analysis, data in signal region were inspected.

Table 3.6: List of expected background and signal yields in the electron channel with 5 $fb^{-1}$of data after all selection and within the ZZ invariant mass range [0,1000]

	0 b-tag yields	1 b-tag yields	2 b-tag yields
Background	1951.83±34.43	1386.68±30.51	94.14±8.67
200 GeV	6.64±0.32	7.47±0.37	1.55±0.16
210 GeV	8.77±0.40	7.96±0.40	1.62±0.16
220 GeV	12.10±0.46	9.22±0.39	2.78±0.23
230 GeV	13.13±0.46	9.95±0.41	2.69±0.21
250 GeV	15.65±0.49	10.24±0.38	2.83±0.19
275 GeV	17.98±0.49	11.61±0.40	3.61±0.22
275 GeV	17.98±0.49	11.61±0.40	3.61±0.22
300 GeV	19.25±0.47	11.45±0.36	3.88±0.20
325 GeV	20.03±0.46	12.14±0.36	4.90±0.24
350 GeV	22.53±0.49	13.30±0.38	5.44±0.25
375 GeV	21.56±0.47	13.32±0.37	5.57±0.24
400 GeV	19.53±0.41	11.05±0.31	5.10±0.22
450 GeV	13.38±0.30	8.10±0.24	3.68±0.16
500 GeV	8.48±0.18	4.52±0.13	2.11±0.09
550 GeV	5.17±0.13	2.98±0.10	1.45±0.07
600 GeV	3.27±0.08	1.77±0.06	0.96±0.04

3.6 Background determination from data

The dominant background to a potential Higgs boson signal stems from Z events produced in association with jets (Z+jets) and from $t\bar{t}$ events. Di-boson events are a less relevant background source, contributing less than 10% to the total. Jets from the dominant backgrounds originate from the hadronization of both light and heavy quarks in Z+jets events, and mainly in the hadronization of b quarks in the $t\bar{t}$ process.

In order to estimate the expected background yields, the analysis does not rely fully on the simulation, rather the background prediction is extracted from data events selected in control regions, where the signal is expected to be negligible. The distribution of the dijet invariant mass, m_{jj}, provides a handle to define such control region, since signal events cluster around the true Z boson mass, unlike background events, which show a broad spectrum. Events in a m_{jj} region close to m_{Z}, 75 GeV $< m_{jj} <$105 GeV, are placed in the *signal region* and are those of interest for the final analysis results, while events passing exactly the same selection but lying in a different m_{jj} region

Table 3.7: List of expected background and signal yields in the muon channel with 5 fb^{-1}of data after all selection and within the ZZ invariant mass range [0,1000]

	0 b-tag yields	1 b-tag yields	2 b-tag yields
Background	2607.14±44.59	1818.05±38.83	125.31±7.66
200 GeV	8.12±0.39	9.69±0.45	2.37±0.22
210 GeV	10.90±0.46	10.42±0.47	2.06±0.19
220 GeV	14.80±0.52	12.04±0.48	2.90±0.22
230 GeV	16.50±0.53	12.63±0.47	3.77±0.25
250 GeV	19.23±0.53	13.54±0.47	4.14±0.26
275 GeV	23.25±0.58	14.20±0.45	4.50±0.24
275 GeV	23.25±0.58	14.20±0.45	4.50±0.24
300 GeV	23.94±0.53	15.45±0.44	5.22±0.24
325 GeV	25.27±0.53	15.58±0.42	5.76±0.26
350 GeV	27.85±0.54	17.13±0.44	6.73±0.28
375 GeV	28.14±0.54	16.30±0.41	6.60±0.27
400 GeV	25.36±0.48	13.82±0.35	6.57±0.25
450 GeV	17.69±0.36	10.01±0.27	4.46±0.18
500 GeV	10.90±0.21	5.49±0.14	2.80±0.11
550 GeV	6.62±0.15	3.69±0.11	1.82±0.08
600 GeV	4.22±0.09	2.12±0.06	1.21±0.05

are kept and placed in what is referred as *sideband region*: 60 GeV $<m_{jj}<$ 75 GeV and 105 GeV $<m_{jj}<$130 GeV.

This region presents a four-fermion invariant mass distribution, m_{ZZ} , similar to that expected for the events selected in the signal region of the analysis, but the signal contribution is notably reduced. Data events selected in the sidebands provide a good prediction of the background in the signal region, $N_{\mathrm{bkg}}(m_{\mathrm{ZZ}})$, both in shape and normalization. The number of background events expected in the signal region is extrapolated from the distribution of observed data in the control region, $N_{\mathrm{sb}}(m_{\mathrm{ZZ}})$, computing the scale factor $\alpha(m_{\mathrm{ZZ}})$ on simulation:

$$N_{\mathrm{bkg}}(m_{\mathrm{ZZ}}) = N_{\mathrm{sb}}(m_{\mathrm{ZZ}}) \times \frac{N_{\mathrm{bkg}}^{\mathrm{MC}}(m_{\mathrm{ZZ}})}{N_{\mathrm{sb}}^{\mathrm{MC}}(m_{\mathrm{ZZ}})} = N_{\mathrm{sb}}(m_{\mathrm{ZZ}}) \times \alpha(m_{\mathrm{ZZ}}). \tag{3.8}$$

The function $\alpha(m_{\mathrm{ZZ}})$, determined from simulated events, accounts for the small kinematic differences expected between yield and shape of the diboson

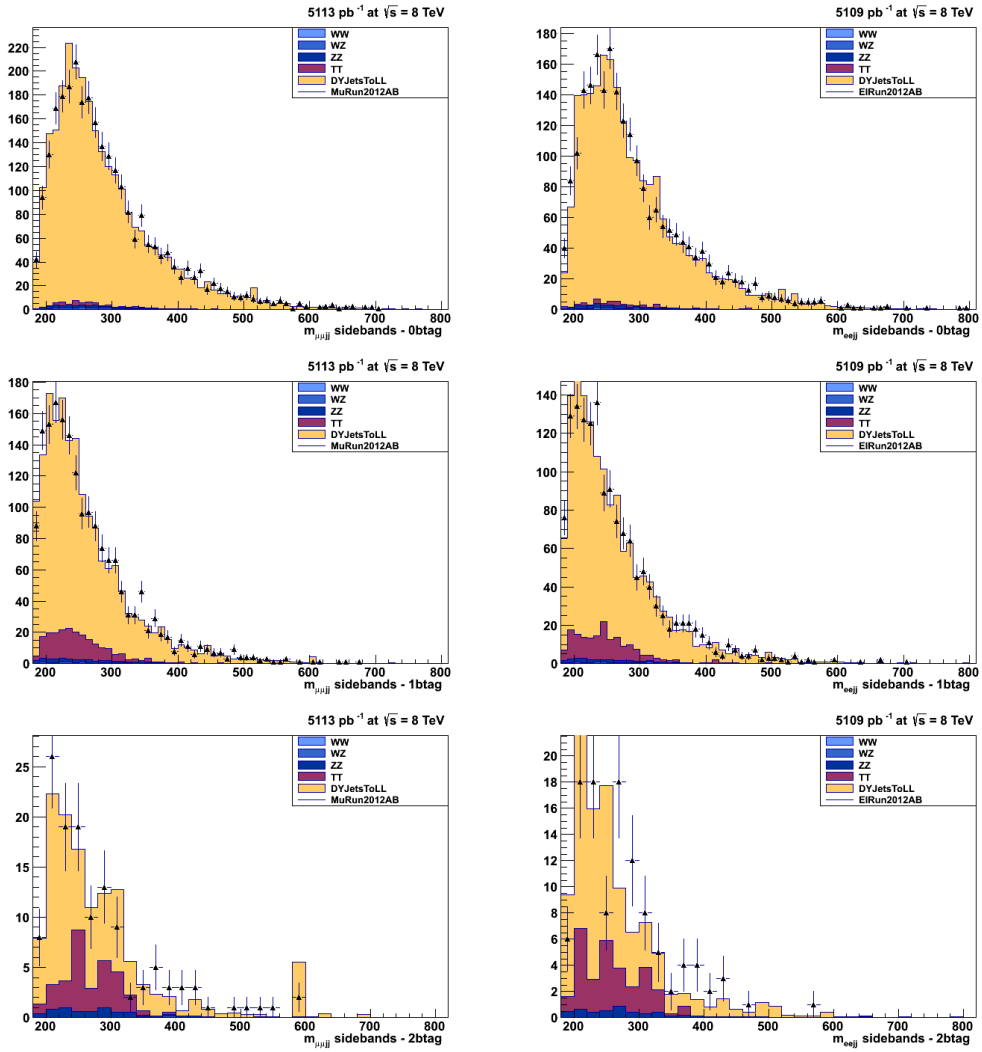


Figure 3.22: The m_{ZZ} distribution in the control region defined in m_{jj} distribution in the three b-tag categories from top to bottom: 0 b-tag (top), 1 b-tag (middle), and 2 b-tag (bottom). Left-side plots refer to events in the muon channel and those on the right side to events in the electron channel. Dots with error bars show data after final selection, histograms show background prediction with the dominant contributions shown separately.

invariant mass distribution as obtained in the signal or in the sideband regions. The background estimated this way includes contributions from the main sources: Z+jets, $t\bar{t}$ and diboson processes, which means that $\alpha(m_{ZZ})$

contains simulation-based corrections for all those background processes. The determination of a proper estimate of the $\alpha(m_{ZZ})$ factor, is affected by the small amount of data and simulated backgrounds. For this reason the $\alpha(m_{ZZ})$ factor is estimated as the ratio of two functions instead of two binned distributions. The distribution of m_{ZZ} for background events arises from random combination of jets and it is not related to the decay of a potential heavy particle, such as a Higgs boson. $N(m_{ZZ})$ is expected to decay exponentially at high mass. A modified Moyal function well reproduces the background shape, both the exponential tail at high mass and the sharp rise at low mass caused by the kinematic selection. The function is:

$$f(x) = \frac{N}{\sqrt{2\pi\sigma_1\sigma_2}} \exp\left(-\frac{1}{2}\mathrm{e}^{-\frac{x-\mu}{\sigma_1}} - \frac{x-\mu}{2\sigma_2}\right), \tag{3.9}$$

where for readability m_{ZZ} is denoted by x and the exponential function by *exp*.

The shapes and normalizations of the m_{ZZ} distributions of background in sidebands and signal regions are obtained performing a fit to the binned distributions of events with the modified Moyal functions, for each b-tag category. The results of the fits are shown in Figure 3.23 for the three b-tag categories, along with the uncertainties from the fits. The function $\alpha(m_{ZZ})$ is, therefore, calculated as the ratio of the two analytical fit functions. This procedure allows to correctly estimate the uncertainty on $\alpha(m_{ZZ})$ and on the background prediction in the signal region by simply propagating the uncertainties on the fit parameters to the background distributions in the sideband and signal region.

The uncertainties $\sigma(m_{ZZ})$ of the $N(m_{ZZ})$ functions are extracted from the variance of the fitted functions:

$$\sigma(x) = \sqrt{\mathrm{Var}[f(x)]} \tag{3.10}$$

The $\alpha(m_{ZZ})$ functions are displayed in Figure 3.24. Once $\alpha(m_{ZZ})$ has been estimated, the background prediction in the signal region is extracted from the data distribution for sidebands corrected by $\alpha(m_{ZZ})$. Figure 3.25 shows the background predictions for the three b-tag categories.

3.7 Systematic uncertainties

The possible sources of systematic uncertainties associated with this measurement and affecting signal normalization are summarized in table 3.8. Effects from lepton reconstruction, jet energy scale and resolution, pile up, MET

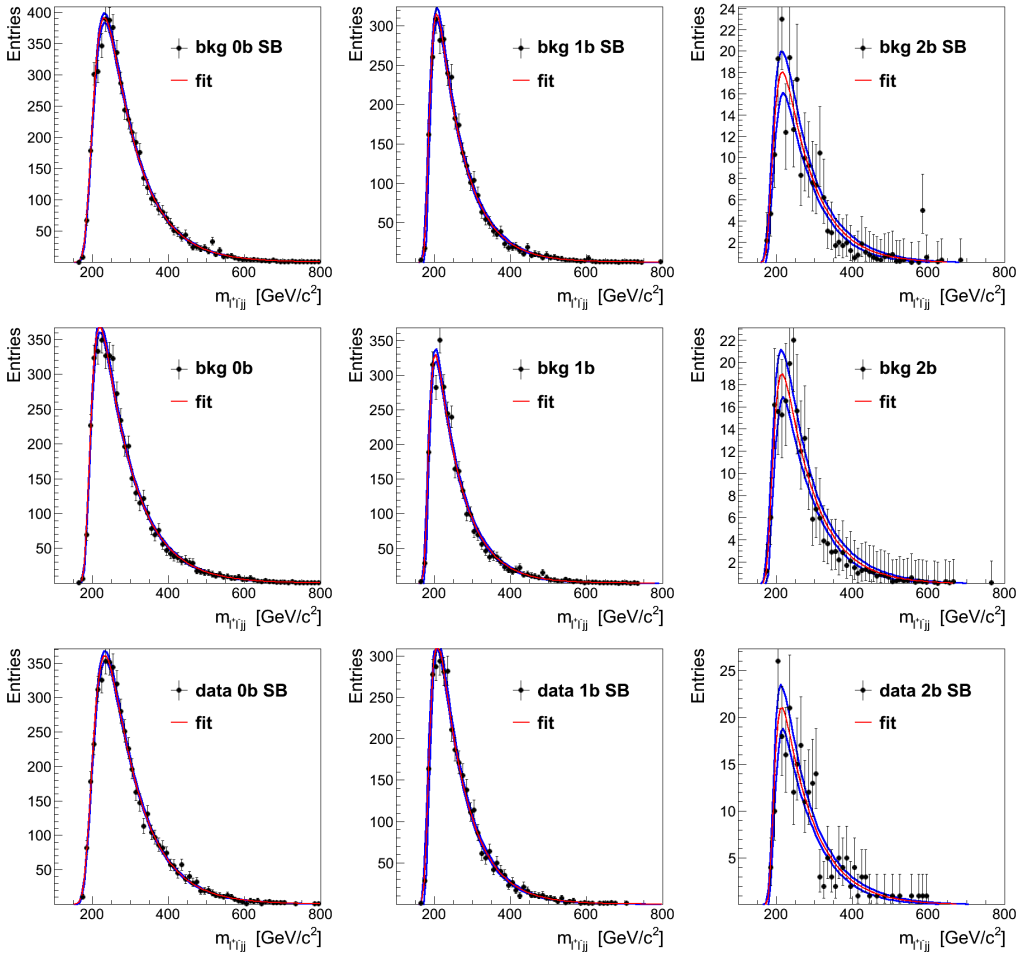
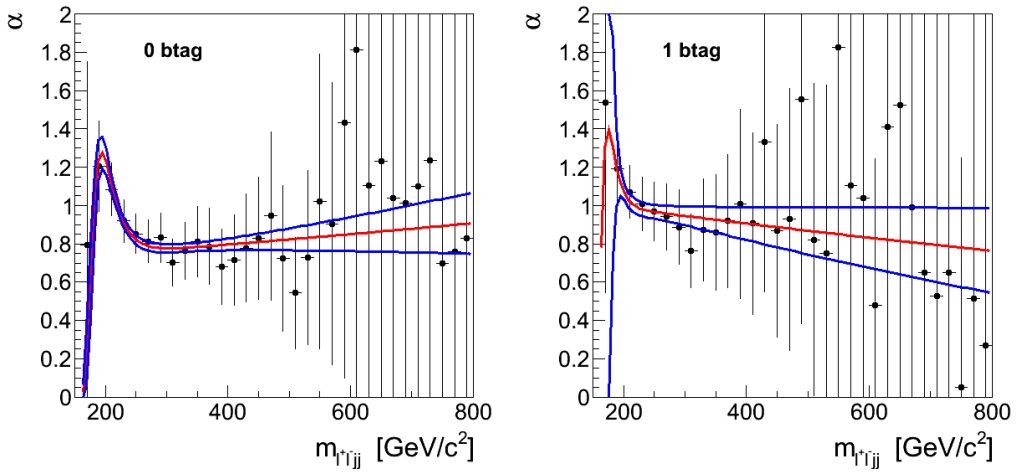


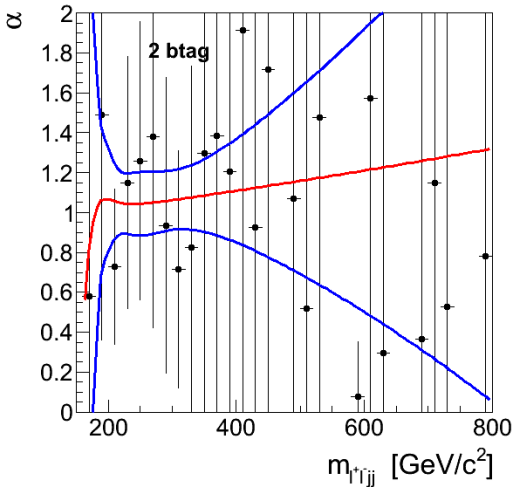
Figure 3.23: Results of the fits to the m_{ZZ} binned distributions for the three b-tag categories, 0 b-tag (left), 1 b-tag (centre), 2 b-tag (right). The three distributions refer to the background in the sideband region (upper) and in the signal region (middle), and also to the data in the sideband region (bottom). The outer solid lines (blue) correspond to the $\pm 1\sigma$ uncertainty on the fit, calculated using the covariance matrix.

significance, heavy-quark flavour tagging, Higgs boson production mechanism, cross section and branching fractions, and LHC luminosity [70] are considered.



(a) 0 b-tag

(b) 1 b-tag



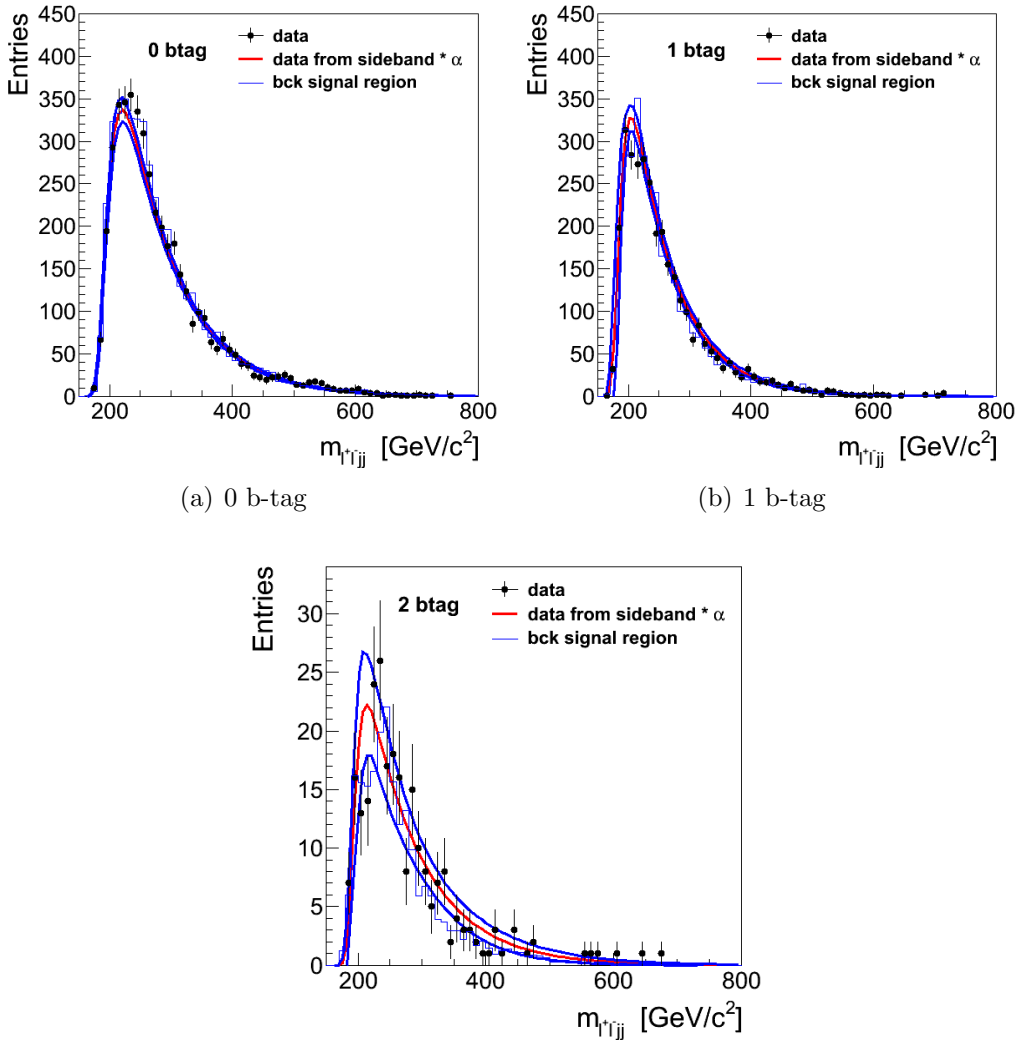
(c) 2 b-tag

Figure 3.24: $\alpha(x)$ functions (inner line in red) together with their uncertainties (outer lines in blue), for the three b-tag categories. For comparison, the ratio of binned distributions is shown as dots with error bars. No fit is performed to these points.

3.7.1 Lepton energy scale, resolution, selection, and trigger

Trigger and identification criteria used to select leptons are common among several H $\rightarrow$ ZZ analyses. A common study has been done based on tag and probe approach, when one lepton from an inclusive sample of Z decays

serves as a tag and efficiency for the other lepton is calculated. In particular, recent studies [71] indicate a systematic uncertainty of 1.0% due to trig-



(a) 0 b-tag

(b) 1 b-tag

(c) 2 b-tag

Figure 3.25: Reconstructed invariant mass distribution, $m_{\ell\ell jj}$, of data events (dots with error bars) in the signal region, for the three b-tag categories: 0 b-tag (a), 1 b-tag (b) and 2 b-tag (c). The background prediction from the α function is depicted as the inner red line and its uncertainty represented by the outer thick blue lines. For comparison, the background prediction in the signal region from the simulation is shown as a thin blue histogram. No fit is performed to the distributions.

ger, 0.5(3.3)% due to muon (electron) identification, 0.2(0.8)% due to muon (electron) isolation, independently on the mass hypothesis, 1.0(2.0)% due to muon (electron) momentum/energy scale.

3.7.2 Jet energy scale and resolution

The main uncertainty on jet reconstruction comes from jet energy scale (JES) uncertainty, while the uncertainty on the resolution has a negligible effect on the total uncertainty. The uncertainty on JES is evaluated shifting the reconstructed p_{T} of jets by $\pm 1\sigma$ of the measured JES uncertainty. This implies a modification in the jet p_{T} spectrum, and hence, in the dijet invariant mass distribution and consequently an effect on the reconstructed Higgs boson mass is expected. The effect on m_{ZZ} results to be lower with respect to the expectation because of the kinematic fit. Results of a detailed study as a function of Higgs mass hypothesis and different categories, is provided in Table 3.9. The effect of jet resolution uncertainty on the signal was evaluated by applying an additional smearing to jets and comparing to the same sample without additional smearing. As the background is evaluated directly from the data, the jet energy resolution is not expected to have a significant effect.

Table 3.8: Summary of systematic uncertainties on signal normalization. Most sources give multiplicative errors on the cross-section measurement, except for the expected Higgs boson production cross section, which is relevant for the measurement of the ratio to the SM expectation.

Source	0 b-tag	1 b-tag	2 b-tag
muon reco	2.7%		
electron reco	4.5%		
jet reco	1–8%		
pile up	1–2%		
MET	–	–	3–4%
b-tagging	2–7%	3–5%	10–11%
acceptance (PDF)	2-4%		
acceptance (WBF)	1%		
theoretical uncertainty (lineshape)	0-3%		
Higgs cross section	13–18%		
luminosity	4.4%		

Table 3.9: Variations in efficiency due to Jet Energy Scale Uncertainties for several signal samples and categories

m_{H} [GeV]	Category	Nominal	JES+	JES-
200	ele	0.0196	0.0203	0.0184
	mu	0.0255	0.0264	0.0239
	0 b-tag	0.0186	0.0237	0.0184
	1 b-tag	0.0223	0.0220	0.0200
	2 b-tag	0.0044	0.0044	0.0040
300	ele	0.0537	0.0550	0.0525
	mu	0.0678	0.0694	0.0663
	0 b-tag	0.0561	0.0573	0.0548
	1 b-tag	0.0502	0.0514	0.0490
	2 b-tag	0.0153	0.0156	0.0149
400	ele	0.0726	0.0735	0.0707
	mu	0.0903	0.0913	0.0877
	0 b-tag	0.0732	0.0741	0.0691
	1 b-tag	0.0651	0.0659	0.0646
	2 b-tag	0.0245	0.0247	0.0247
525	ele	0.0796	0.0797	0.0793
	mu	0.0951	0.0951	0.0947
	0 b-tag	0.0802	0.0804	0.0799
	1 b-tag	0.0660	0.0660	0.0657
	2 b-tag	0.0285	0.0284	0.0284
600	ele	0.0732	0.0774	0.0764
	mu	0.0871	0.0922	0.0909
	0 b-tag	0.0759	0.0759	0.0793
	1 b-tag	0.0576	0.0610	0.0599
	2 b-tag	0.0268	0.0281	0.0280

3.7.3 Pile up reweighing

As mentioned in Section 3.3, the number of true interactions per bunch crossing in the simulated samples was reweighed to match the distributions in data. The uncertainty on the measurement of the amount of pileup in data is a source of systematics for this analysis. This uncertainty is studied by estimating the number of true interactions in data with different values of minimum-bias cross section as input, as recommended by the CMS pileup group [72]: a $\pm 5\%$ variation of the nominal value 69.40 mb is considered. The two distributions and the central value are compared. The number of true interactions is reweighted in the simulation to match the two shifted

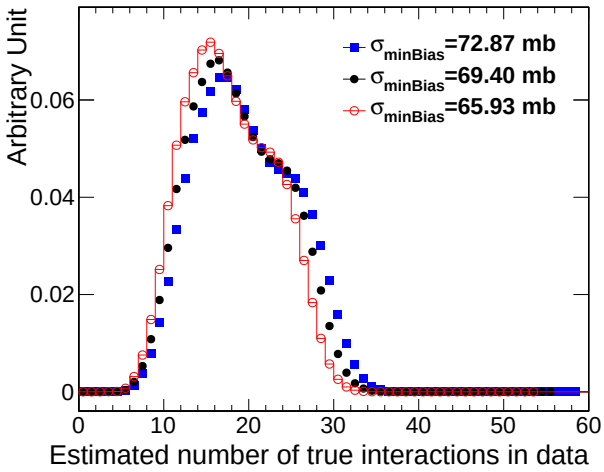


Figure 3.26: The estimated number of true interactions in 2012 data, assuming different values of minimum-bias cross section. The central value is 69.4 mb (solid circles).

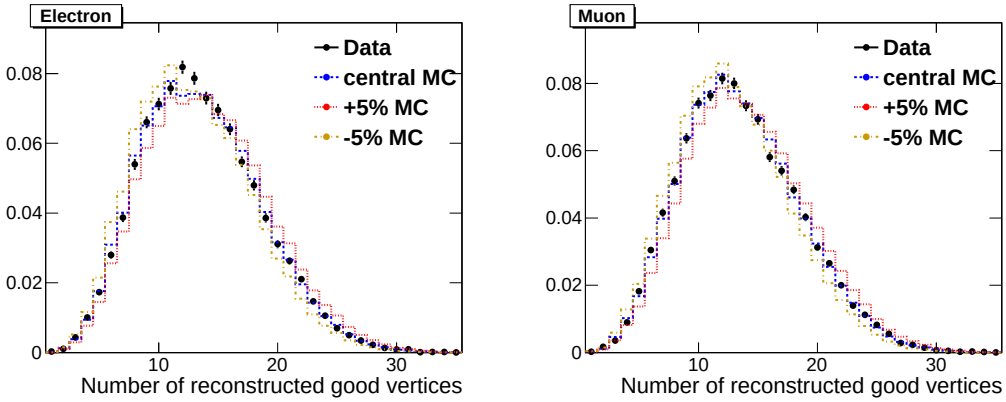


Figure 3.27: The reconstructed number of good vertexes in data (points with error bars) and in reweighed simulation (various lines), from the electron channel (left) and the muon channel (right). The number of true interactions in simulation has been reweighed to match the estimated distributions in data, assuming three different values of minimum-bias cross section.

distributions in data in Figure 3.26 and compute again the signal efficiency. This leads to a change of $\sim 1 \div 2\%$ in the signal efficiency for $m_{\mathrm{H}} < 600$ GeV, approximately independent on lepton channel and b-tag category, as shown in Table 3.10.

Table 3.10: The relative systematic uncertainty in percentage (%) from PU reweighing.

Electrons			
m_{H} [GeV]	$N_{\mathrm{btag}} = 0$	$N_{\mathrm{btag}} = 1$	$N_{\mathrm{btag}} = 2$
200	$^{+1.57}_{-1.91}$	$^{+0.34}_{-0.14}$	$^{+0.01}_{+0.08}$
300	$^{+1.04}_{-0.93}$	$^{+0.99}_{-0.77}$	$^{+1.07}_{-1.01}$
400	$^{+0.38}_{-0.37}$	$^{+0.26}_{-0.31}$	$^{+1.64}_{-1.50}$
525	$^{+0.13}_{-0.15}$	$^{+1.07}_{-1.06}$	$^{-0.11}_{-0.06}$
600	$^{+1.03}_{-0.99}$	$^{+0.95}_{-0.96}$	$^{+0.16}_{-0.27}$
700	$^{+0.39}_{-0.39}$	$^{+0.22}_{-0.23}$	$^{+0.51}_{-0.60}$
800	$^{+0.98}_{-0.83}$	$^{+0.05}_{-0.22}$	$^{+0.97}_{-1.25}$
900	$^{+0.61}_{-0.69}$	$^{+1.59}_{-1.39}$	$^{+0.70}_{-0.58}$
1000	$^{+0.99}_{-1.10}$	$^{+0.35}_{-0.53}$	$^{+0.24}_{-0.45}$

Muons			
m_{H} [GeV]	$N_{\mathrm{btag}} = 0$	$N_{\mathrm{btag}} = 1$	$N_{\mathrm{btag}} = 2$
200	$^{-0.08}_{-0.31}$	$^{+1.06}_{-1.05}$	$^{+1.15}_{-1.26}$
300	$^{+0.76}_{-0.66}$	$^{+1.20}_{-1.11}$	$^{+0.37}_{-0.40}$
400	$^{+0.46}_{-0.46}$	$^{+0.31}_{-0.39}$	$^{+2.31}_{-2.04}$
525	$^{+0.39}_{-0.40}$	$^{+0.63}_{-0.52}$	$^{+0.25}_{-0.45}$
600	$^{+0.94}_{-0.87}$	$^{+0.69}_{-0.78}$	$^{+0.28}_{-0.61}$
700	$^{+0.68}_{-0.76}$	$^{+0.20}_{-0.28}$	$^{+0.95}_{-1.19}$
800	$^{+0.62}_{-0.70}$	$^{+1.42}_{-1.25}$	$^{+1.14}_{-1.35}$
900	$^{+0.67}_{-0.65}$	$^{+0.17}_{-0.25}$	$^{+2.21}_{-1.94}$
1000	$^{+1.19}_{-1.00}$	$^{+1.10}_{-0.94}$	$^{+3.09}_{-2.76}$

3.7.4 Heavy-quark flavour tagging uncertainty

Data-simulation scale factors (SF_{hf}) have been used to correct simulation prediction to consider the different efficiency of b-tagging in data and simulation. Likewise, a mistag rate scale factor (SF_{lf}) for light quarks misreconstructed as b jets has been used as well in this analysis. To study the systematic effects of b tagging, both the SF_{hf} and SF_{lf} were simultaneously varied up and down by the uncertainty related to each kind of SF. The study was performed separately for the muon and electron channels, calculating the effect for simulated signal simulation.

Tables 3.11 and 3.12 give the b-tagging systematic uncertainty for the signal,

Table 3.11: Systematic uncertainty on the signal in the electron channel.

m_{H}	0 b-tag		1 b-tag		2 b-tag	
	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$
200	0.95	1.02	1.08	1.01	1.08	0.87
250	0.97	1.03	1.05	0.94	1.06	0.96
300	0.97	1.01	1.02	0.99	1.11	0.91
350	0.97	1.03	1.02	0.98	1.09	0.92
400	0.97	1.02	1.03	1.00	1.08	0.91
425	0.98	1.04	1.02	0.96	1.08	0.94
525	0.96	1.03	1.04	0.98	1.07	0.91
600	0.97	1.03	1.01	0.96	1.09	0.91

Table 3.12: Systematic uncertainty on the signal in the muon channel.

m_{H}	0 b-tag		1 b-tag		2 b-tag	
	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$	$\frac{SF_{Up}}{SF_{Central}}$	$\frac{SF_{Down}}{SF_{Central}}$
200	0.96	1.03	1.04	1.00	1.05	0.82
250	0.97	1.03	1.01	0.99	1.12	0.83
300	0.98	1.03	1.01	0.98	1.09	0.90
350	0.97	1.02	1.00	0.98	1.13	0.92
400	0.97	1.03	1.02	1.01	1.07	0.86
425	0.98	1.03	1.01	0.99	1.10	0.91
525	0.97	1.03	1.01	0.97	1.11	0.95
600	0.97	1.02	1.04	1.01	1.06	0.89

for muons and electrons. The systematic effect is computed as the ratio of the number of tagged jets with a SF varied by plus and minus its uncertainty to the number of tagged jets with the nominal SF. The uncertainty is reported for the cases where both jets are tagged and it is function of the Higgs mass.

3.7.5 Uncertainties on background prediction

Uncertainty on background is considered separately and is one of the dominant effects on the exclusion limits. Systematic uncertainties on background affect the expectation of the shape and yield extrapolated from the sidebands as described in section 3.6. Since most of the sources of systematic uncertain-

ties affecting the simulation predictions cancel their effect in the computed ratio to obtain the estimated $\alpha(m_{ZZ})$, the main uncertainty is actually coming from the parameterization of the final sideband distribution.

The errors on the fitted parameters are properly propagated into the uncertainty on the shape. The full uncertainty is then taken into account in the statistical treatment of the result (Chapter 4).

3.7.6 Missing transverse energy uncertainty

MET affects directly only the 2 b-tag category. The dominant effects are from the knowledge of the rest of the event, such as jet energy reconstruction and pileup. Therefore, both of the above subsections cover MET uncertainty to a large extent. In signal as well as in Z + jets background it is not expected to have real MET, while it is expected for $t\bar{t}$ background events. The uncertainty is computed subtracting $t\bar{t}$ constribution predicted by simulation from both data and simulated events, applying the requirement on the MET significance and evaluating the difference in efficiency in data and simulation. This difference gives an estimation of the uncertainty which is not larger than the inefficiency of the requirement on the MET significance, about 3%. In order to be conservative, the latter value is taken as uncertainty.

3.7.7 Signal production mechanism uncertainties

The expected kinematics of the Higgs production is subject to uncertainties due to limited knowledge of the underlying parton distribution functions (PDFs), missing higher-order corrections in perturbation theory and the contribution to the VBF. These uncertainties are propagated as uncertainty on the selection acceptance and efficiency. Their additional effect on the Higgs boson production cross section is discussed in a separate section (Section 3.7.8).

PDF uncertainties are evaluated according to the PDF4LHC recommendations, by evaluating the selection efficiency for the PDF sets CT10 [73], MSTW2008NLO [74] and NNPDF2.1 [75] and their error sets. Table 3.13 summarizes the resulting acceptance uncertainties. The envelope of the various PDF sets is used as total uncertainty, as recommended and amounts to 2÷4% without strong dependence on b-tag category. The uncertainty noticeably increases for very high Higgs masses.

The VBF contribution to the Higgs boson production mechanism is around 10% of gluon fusion. This analysis has been tuned using gluon fusion based simulation while a real signal contains a mixture of events produced by gluon fusion and VBF. The uncertainty that originates from this point, summarized

Table 3.13: Summary of systematic uncertainties on the signal acceptance following PDF4LHC recommendations.

PDF	$m_{\mathrm{H}} = 200$ GeV			$m_{\mathrm{H}} = 600$ GeV		
	0 b-tag	1 b-tag	2 b-tag	0 b-tag	1 b-tag	2 b-tag
CT10	$^{+1.8\%}_{-2.3\%}$	$^{+1.9\%}_{-2.4\%}$	$^{+2.1\%}_{-2.1\%}$	$^{+1.3\%}_{-2.0\%}$	$^{+1.4\%}_{-2.2\%}$	$^{+2.1\%}_{-3.5\%}$
all categories		$^{+1.9\%}_{-2.3\%}$			$^{+1.5\%}_{-2.4\%}$	
MSTW2008NLO	$^{1.1\%}_{-0.1\%}$	$^{+1.2\%}_{-0.1\%}$	$^{+1.4\%}_{-0.4\%}$	$^{+1.2\%}_{+0.4\%}$	$^{+1.2\%}_{+0.5\%}$	$^{+2.0\%}_{+0.9\%}$
all categories		$^{1.2\%}_{-0.1\%}$			$^{+1.3\%}_{-0.5\%}$	
NNPDF2.1	$^{+2.3\%}_{+0.7\%}$	$^{+2.5\%}_{+0.7\%}$	$^{+2.7\%}_{+0.7\%}$	$^{+2.1\%}_{+1.0\%}$	$^{+2.0\%}_{+1.0\%}$	$^{+3.5\%}_{+1.7\%}$
all categories		$^{+2.4\%}_{+0.7\%}$			$^{+2.3\%}_{+1.1\%}$	
Total	$^{+2.3\%}_{-2.3\%}$	$^{+2.5\%}_{-2.4\%}$	$^{+2.7\%}_{-2.1\%}$	$^{+2.1\%}_{-2.0\%}$	$^{+2.0\%}_{-2.2\%}$	$^{+3.5\%}_{-3.5\%}$
all categories		$^{+2.4\%}_{-2.3\%}$			$^{+2.3\%}_{-2.4\%}$	

Table 3.14: Summary of systematic uncertainties due to VBF.

m_{H}	$m_{\mathrm{H}} = 200$ GeV			$m_{\mathrm{H}} = 400$ GeV			$m_{\mathrm{H}} = 600$ GeV		
	0 b-tag	1 b-tag	2 b-tag	0 b-tag	1 b-tag	2 b-tag	0 b-tag	1 b-tag	2 b-tag
Δ_{eff}	20%	15%	6%	-2%	4%	4%	-8%	6%	-0.3%
total	7%			11%			13%		
unc.	2.4%	1.8%	0.7%	-0.16 %	0.3 %	0.3 %	-1.4 %	1 %	0%
total	2%			0.1%			-1.4%		

in Table 3.14, has to be taken into account. The difference in signal acceptance between the two production mechanisms in simulation is estimated and multiplied by the expected fraction of VBF production, leading to a global uncertainty on the production cross section.

Additional uncertainties arise from the Higgs signal shape and have been evaluated as described in [76]. Due to the mass-dependence of the signal efficiency on selection, the total signal efficiency is affected by the signal shape (line shape). Efficiency curves for the nominal and alternative line shapes are shown in Figure 3.28. The uncertainty is negligible below 400 GeV and rises to $\sim 3\%$ at 600 GeV, with only small dependence on b-tag category.

Additionally the line-shape used in the limit extraction procedure is extracted again with the alternative line-shape models (see Figure 3.29). The tail caused by mismatched jets is not affected at all as it is a random mixture

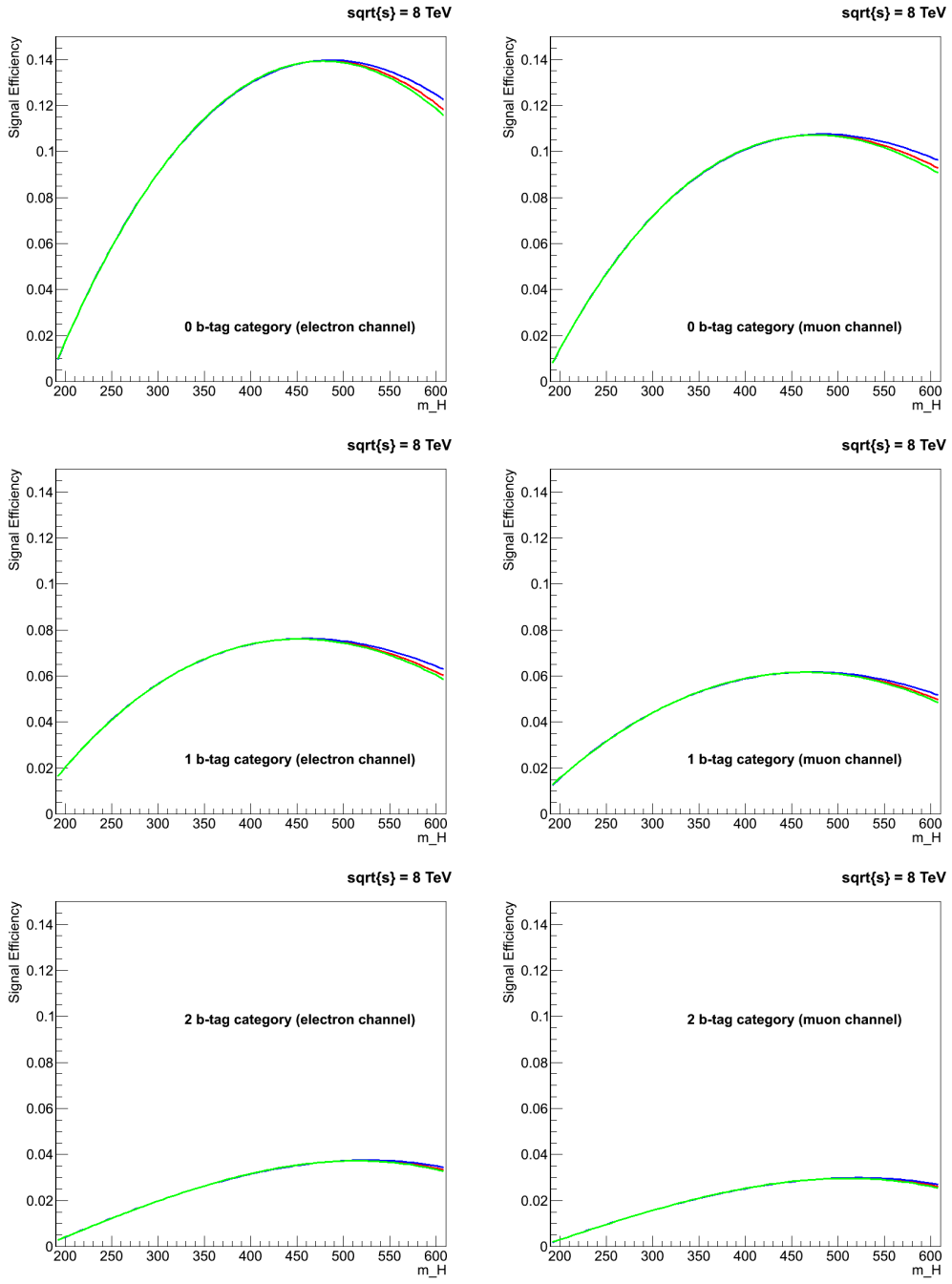


Figure 3.28: Signal selection efficiency for the nominal line shape (read) and alternative shapes (green/blue) for electrons (left) and muons (right) for 0, 1, 2 b-tag categories (top, centre, bottom).

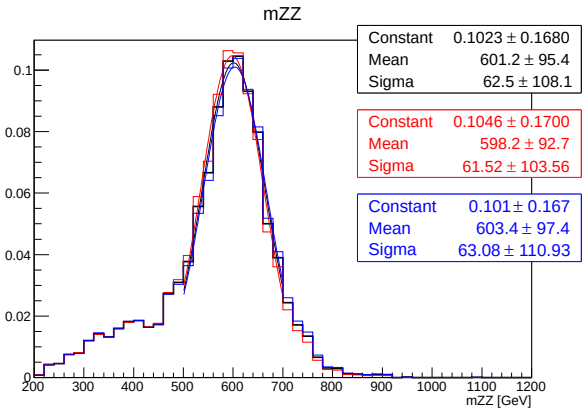


Figure 3.29: Reconstructed Higgs boson signal for m_{H} =600 GeV (area normalized) with the nominal lineshape (black) and alternative models (blue/red). Gaussian fits to the core of the distribution are overlaid.

of events, averaging out any shift from this uncertainty. The core of the signal distribution is only weakly affected by this effect. In the worst case (the highest mass we consider), the peak-position shifts by $\sim$ 2 GeV (compared to a width of 60 GeV) and the width changes by $\sim$ 1 GeV. Due the negligible effect of this uncertainty, it is not propagated further.

3.7.8 Higgs cross section and branching fractions

The Higgs boson production cross-section uncertainty depends on production mechanism, either gluon fusion or VBF. The gluon fusion mechanism dominates between the two and drives the total uncertainty. Gluon fusion and VBF uncertainties are considered separately and for each mass point according to Yellow Report prescription in [21]. The total weighted uncertainty is in the range 13.4÷18.0%. This uncertainty is relevant only for the measurement of the ratio to SM expectation r, while it does not affect the absolute cross-section measurement.

3.7.9 Luminosity uncertainty

The uncertainty on the measured integrated luminosity corresponds to a 4.4% [70] as from the latest official recommendation for the 2012 data samples.

3.8 Analysis of the 7 TeV data sample

The search for a Higgs boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ channel has been performed also with a sample of pp collisions at a c.m energy $\sqrt{s} = 7$ TeV corresponding to an integrated luminosity of $\mathcal{L} = 4.9$ fb^{-1} recorded by the CMS experiment during 2011.

The selection of events and the reconstruction of signal-like candidates are the same as the analysis workflow presented in the previous sections and performed on the 8 TeV dataset, except for few details of technical kind:

- Isolation criteria suggested for the analysis of 2011 datataking were different from the ones used in 2012;
- An additional discriminant variable is used to reject background events: quark-gluon discriminant;

The isolation variable used for 2011 analysis is not based on particles reconstructed with the PF algorithm, but rather directly on tracks reconstructed in the tracker and energy deposited in the calorimeters:

$$I = I_{tkr} + I_{ECAL} + I_{HCAL} \tag{3.11}$$

where I_{tkr} is the sum of transverse momentum of all the tracks in tracker within the isolation cone($I_{tkr} = \sum p^i_{\mathrm{T}}$), I_{ECAL} and I_{HCAL} are the the sum of deposits of energy in the electromagnetic and hadronic calorimeters ($I_{CAL} = \sum E^i_{\mathrm{T}}$), again within the isolation cone. The identification criteria for 2011 analysis was optimized for leptons with $p_{\mathrm{T}} > 20\ GeV$. In 2012 the selection has been reoptimized in order to assure good quality of lepton reconstruction also for leptons with lower p_{T}.

In both signal and Z + jets background events, leptons originate from a Z boson decay, for this reason a different lepton selection does not help improving background rejection. Moreover the working point of the lepton identification has been chosen in order to give the same signal efficiency in the two analyses. The resulting effect of the different selection criteria is negligible.

For what concerns the second point, events are categorized in three exclusive categories according to the number of b jets in the final state both in 2011 and 2012 analyses. The 0 b-tag category is dominated by Z + jets events, with jets originated mainly from light-flavour quarks and gluons. In order to reduce the amount of background in this category, events with jets originated from gluons, which are certainly not signal jets, can be isolated and excluded from further analysis. Jets consistent with being initiated by gluons can be

identified performing a discrimination based on three measured quantities: the number of charged hadronic particle tracks, the number of photons and neutral hadrons and the variable PTD = $\sqrt{\sum p_{\mathrm{T}}^2/(\sum p_{\mathrm{T}})^2}$, where the sum is extended over all jet constituents. Gluon jets, differently from quark jets, are softer and are characterized by an higher number of constituents. A quark-gluon likelihood discriminant (LD_{q-g}) is constructed from the aforementioned three observables. A requirement on quark-gluon discriminant is used to isolate gluon jets: $LD_{q-g} > 0.10$ While for 2011 the systematics uncertainty on LD_{q-g} was provided by the CMS group responsible of jet reconstruction and performance, for the 2012 analysis, this measurement was not ready at the time this thesis was written. In order to avoid not to take into account correctly the systematic uncertainty on this variable, it was decided not to make use of it performing the analysis on the 8 TeV dataset.

A slightly different statistical approach has been used to perform the final analysis of selected events in 2011 and 2012 datasets. A shape analysis has been performed on the first set of data. Even if it is planned in the near future to apply the same kind of statistical analysis in the 8 TeV dataset, a cut and count approach has been used to set exclusion limit on the Higgs boson production cross section in the analysis of 2012 dataset, due to a lack of time. The two approaches will be detailed in Section 4.2 along with the final results.

Chapter 4

Statistical interpretation of results

This Chapter aims at describing the statistical analysis performed on the events selected according to the strategy described in Chapter 3 and reports the results of the study on data collected in 2011 at 7 TeV and in 2012 at 8 TeV, for an integrated luminosity corresponding to 4.9 fb^{-1} and 5.1 fb^{-1} respectively. In Section 4.1 the signal modeling adopted in this search will be described. Section 4.2 will introduce the statistical method used to set upper limits on the Higgs production cross section and the final results for both 7 TeV and 8 TeV analyses will be given. The results of the analysis performed on pp collision data collected at 7 TeV were published in the early 2012 [54]. Here an update of the published results will be presented, which includes improvements to the analysis developed after the publication. In Section 4.3 the results of the combination of Higgs searches at CMS and the observation of a new boson will be presented.

4.1 Signal shape parametrization

The final variable used in this analysis to distinguish the signal on top of background is the diboson invariant mass,m_{ZZ} , reconstructed from 4 objects in the event: two leptons and two jets. The m_{ZZ} variable is significantly different for events containing a real Higgs (signal events) or just coming from random combination of leptons and jets produced in the event (background events). While the background expectation is obtained from data as described in Section 3.6, the signal shape and normalization are obtained from simulation. The signal distribution depends on Higgs boson mass and it requires a more detailed and mass-dependent treatment. For each available

simulated signal sample the distributions of events in the reconstructed m_{ZZ} spectrum, are obtained for the six classes of events: three b-tag times two lepton flavour categories.

Signal mass distribution for nominal m_H beyond 400 GeV is not correctly modeled by simulation. Since the Higgs width is too large to approximate it as narrow (at 400 GeV $\Lambda_H > 70$ GeV). The problem has been discussed in details in Ref. [77, 78] and a more correct approach to describe the Higgs invariant-mass distribution, known as Complex Pole Scheme (CPS), has been proposed. Thus, the total Higgs production cross-section has been recomputed by the Higgs Cross-Section Working Group to include corrections due to CPS for large m_H values. In the published analysis, CPS effects were included in the cross section calculation, but neglected for the signal shape (covered by an appropriate uncertainty). In order to consider the correct shape of the signal, the simulated signal samples (both 7 TeV and 8 TeV simulated samples) have been reweighted to follow the CPS prescription. Moreover, a paper recently published [79] shows that, at high Higgs mass, the interference between the Higgs signal and the $gg \to ZZ$ background becomes large. The effect of interference has been shown to be constructive below the Higgs mass peak and destructive above. It has therefore a negligible effect on the total cross-section ($\sim 1 \div 2\%$) but it biases the ZZ invariant mass distribution. This further effect is taken into account in the reweighting procedure. In the updated 2011 analysis this effect has been considered properly.

The modeling of reweighted signal distributions is done by splitting the accepted events in two sets:

- **Matched events**: events in which the four reconstructed objects used to build the Higgs candidates are correctly identified from the original underlying objects at the generator level;
- **Unmatched events**: events in which jets used for the reconstruction do not match the quarks originated in the Higgs boson decay.

Since the events of the second type are not reconstructed from the signal decay products, they are not expected to give the right quadrimomentum of the generated Higgs. It is clear that the distribution of the reconstructed ZZ mass for the two sets of events is not the same, being the first one an actual resonance and the second one a combinatorial background, although still related to the actual Higgs mass since most of the objects used in the mass reconstructions are actually the correct ones.

The parametrization of the total distribution shape can be treated as the result of two different components, parametrized independently and com-

bined to give the final one. Matched events are parameterized by a Double Crystal-Ball function [80] (i.e. a Gaussian core with powerlaw tails on both sides). A single Crystal-Ball function has been used to parametrize the unmacthed subsample. The addition of the two functions gives a very good parameterization of the overall shape, as shown in Figure 4.1, for some m_{H} points simulated at 8 TeV, where the expected distributions are compared to parametrization.

The same procedure is used to extract signal shape expectation from 7 TeV simulations. This represents a minor update with respect to the published analysis where the distribution of unmatched events was fitted with an additional triangle function, now removed because not useful any longer. Figure 4.2 shows some examples of the fits performed to parametrize the signal shape. As can be seen in the plots, the signal lineshape is well described by the parameterization.

The search for a Higgs boson is performed at 73 mass hypotheses between 200 and 600 GeV. The mass steps in this range are optimized to account for the expected width, Γ_{H}, and resolution for measurement of m_{H} [81]. Nevertheless, due to limited computing resources, the generation of signal samples for each mass point in which this analysis is intended to perform a search, was not possible.

The fitted parameterizations for the available simulated samples (Table A.2 in Appendix A) are used to produce the interpolated parameterization for all 73 mass points.

The signal efficiency of the selection described in Chapter 3 is evaluated as the ratio between the number of selected events in each of the six channels under study and the total number of generated events in the generated samples. The signal efficiency as a function of the m_{H} is fitted to a polinomium in order to be estimated also for all the points as shown in Figure 4.3.

4.2 Statistical analysis

The statistical analysis is based on the invariant mass of the Higgs boson candidate. Six independent channels are considered: the six analysis categories aforementioned. The observed m_{ZZ} distributions on 5 fb^{-1} of 2012 data for both muon and electron channels and for the three b-tag categories are shown in Figure 4.4.

The expected and observed upper limits to the SM production cross section, for 73 mass points in the range 200-600 GeV, are determined as function of the Higgs boson mass, m_{H}, with the modified frequentist method CL_s [82]. Systematic uncertainties affecting both signal and background predictions

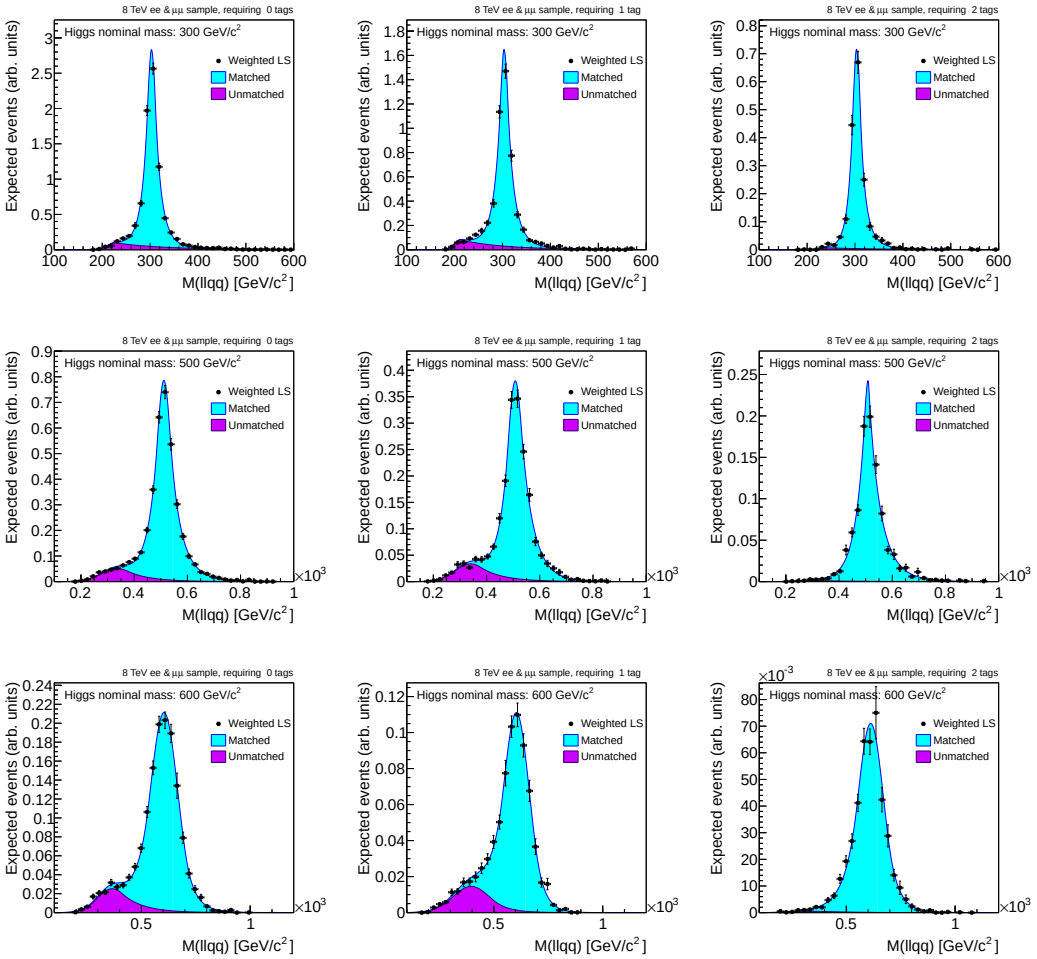


Figure 4.1: Reconstructed m_{ZZ} distribution of the selected signal events (2012 analysis) in the three b-tag categories, 0 b-tag (left), 1 b-tag (centre), 2 b-tag (right) and for m_{H} equal to 300, 500 and 600 GeV (from top to bottom). Electron and muon channels are combined. Distributions are fitted with a sum of a Double Crystal-Ball and a Crystal-Ball function representing the matched and un-matched components respectively. The results of the fits are superimposed to the distribution and the fitted components are shown with distinctive colours.

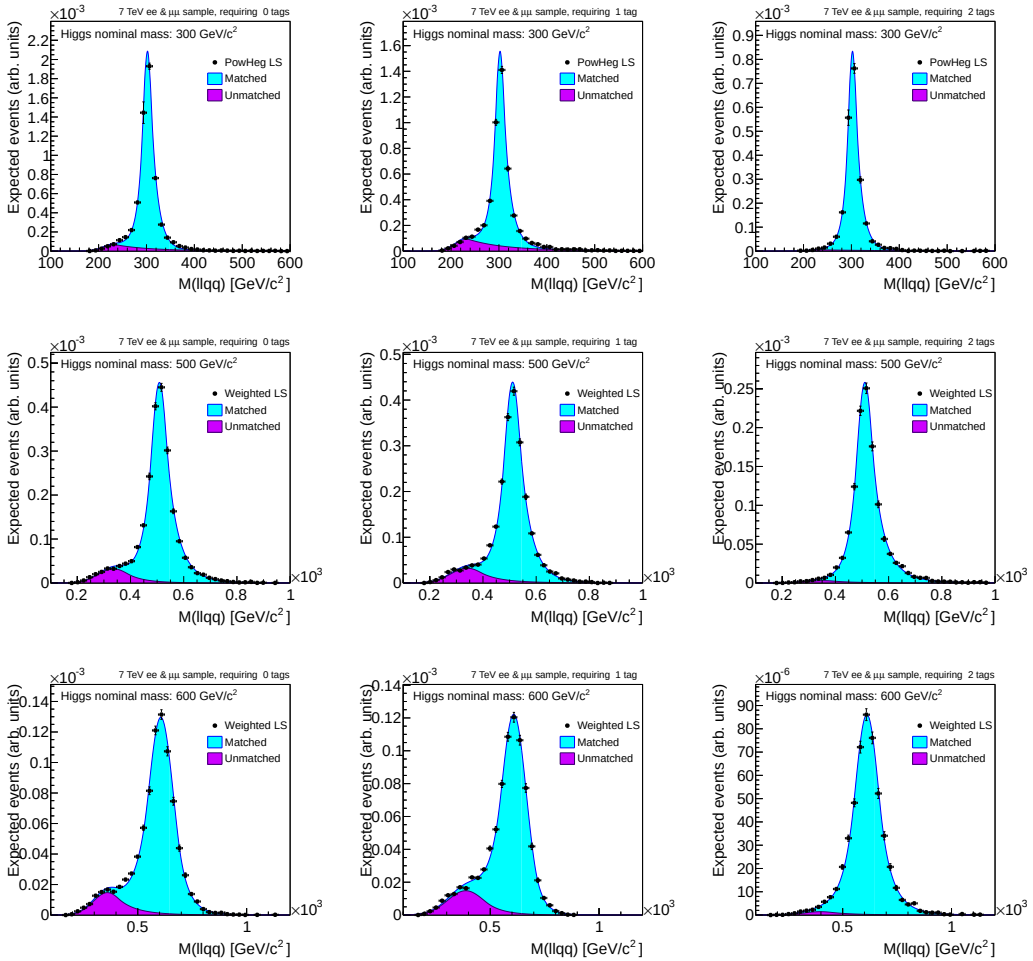


Figure 4.2: Reconstructed m_{ZZ} distribution of the selected signal events (2011 analysis) in the three b-tag categories, 0 b-tag (left), 1 b-tag (centre), 2 b-tag (right) and for m_{H} equal to 300, 500 and 600 GeV (from top to bottom). Electron and muon channels are combined. Distributions are fitted with a sum of a Double Crystal-Ball and a Crystal-Ball function representing the matched and un-matched components respectively. The results of the fits are superimposed to the distribution and the fitted components are shown with distinctive colours.

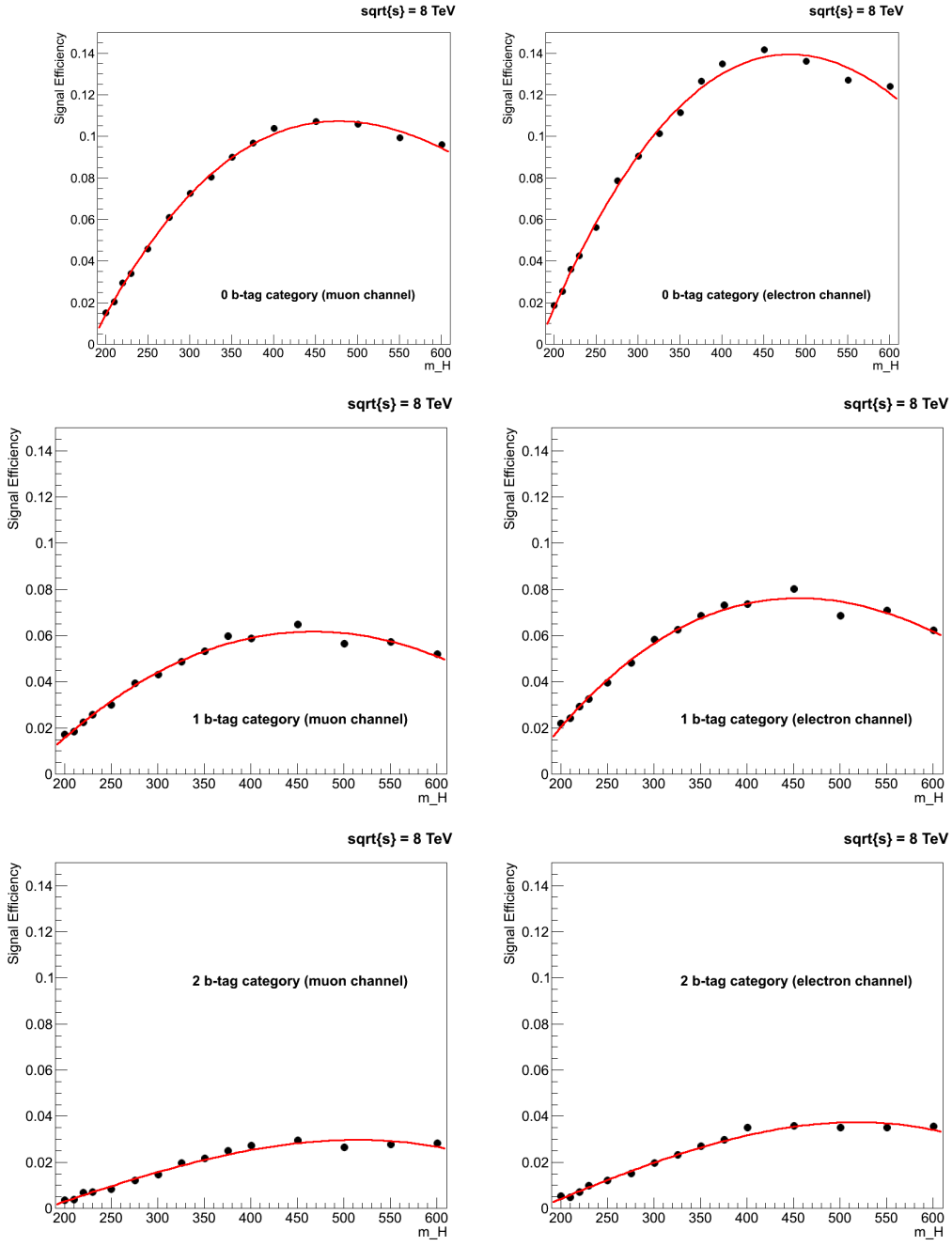


Figure 4.3: Result of the polinomial fit performed to get the parametrization of signal efficiency as a function of m_{H} in 0 b-tag (top), 1 b-tag (middle), 2 b-tag (bottom) categories and in the muon (left) and electron (right) channels. The efficiency decrease at high m_{H} values is due to the high boost of the two jets, which makes difficult to reconstruct them as two separate entities. In such cases the couple of jets is reconstructed as a unique jet, and this feature does not match the selection criteria.

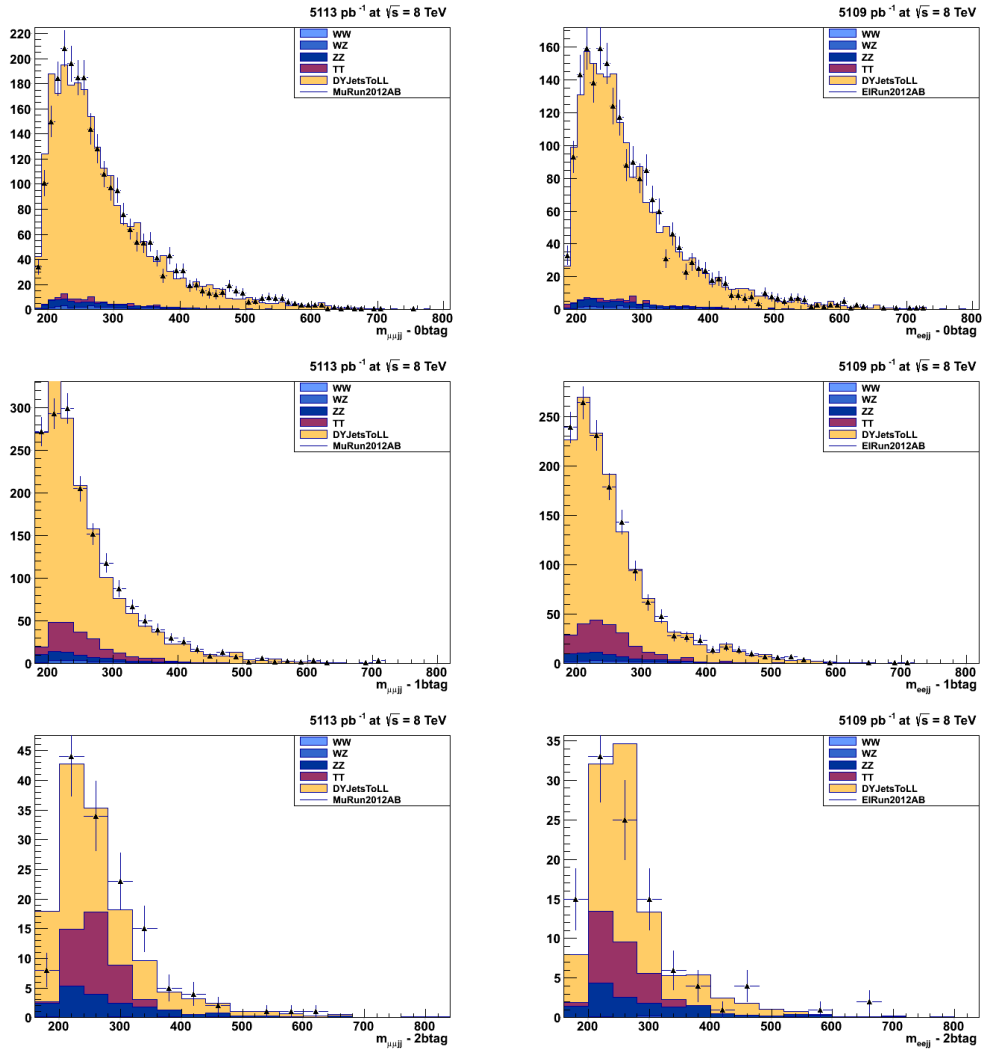


Figure 4.4: The m_{ZZ} distribution in the signal region in the three b-tag categories: 0 b-tag (top), 1 b-tag (middle), and 2 b-tag (bottom). Left-side plots refer to events in the muon channel and those on the right side to events in the electron channel. Points with error bars show data after final selection, histograms show prediction with the dominant contributions shown separately.

are incorporated in the analysis via nuisance parameters θ. Signal and background expectations, become functions of the this set of parameters: $s(\theta)$ and $b(\theta)$. Upper limits are set on the signal strength modifier $\mu = \sigma/\sigma_{SM}$. This

choice takes into account the possibility of modifications from new physics beyond the SM could affect the production rates or create new decay channels reducing the branching fractions in the modes searched for.

The test statistic used to discriminate signal + background from the background-only hypothesis is a likelihood ratio Q, that according to the Neyman-Pearson lemma is the most powerful discriminator. The profile likelihood ratio is defined as:

$$q_\mu = -2\ln\frac{\mathcal{L}(obs \mid \mu \cdot s + b, \hat{\theta}_\mu)}{\mathcal{L}(obs \mid \hat{\mu} \cdot s + b, \hat{\theta})} \tag{4.1}$$

$\hat{\theta}_\mu$ is the set of nuisance parameters that maximizes the likelihood at the numerator, for a given value of μ, while $\hat{\theta}$ and $\hat{\mu}$ are, respectively, the set of nuisances and the value of μ which maximize the likelihood at the denominator. The confidence in the signal + background hypothesis, $\mathrm{CL_{s+b}}$, is defined by the probability to observe a value of the test statistics equal to or larger than the value observed in the experiment:

$$\mathrm{CL_{s+b}} = \mathrm{P_{s+b}}(q_\mu \geq q_\mu^{obs})$$

The probability to observe a value of the test statistic equal to or larger than the value observed in data under the background only hypothesis is $\mathrm{CL_b}$:

$$\mathrm{CL_b} = \mathrm{P_b}(q_\mu \geq q_\mu^{obs})$$

Then, the quantity $\mathrm{CL_s}$ is calculated as the ratio of these two probabilities:

$$\mathrm{CL_s} = \frac{\mathrm{CL_{s+b}}}{\mathrm{CL_b}}$$

Under high statistic regime the probability density functions, p.d.f, of the test statistic q_μ, for signal + background and background-only hypotheses, are well defined by an analytical formula. Indeed, in the asymptotic regime, according to the Wilks theorem, q_μ is expected to have a well defined shape as described in [83, 81]. Therefore, it is possible to get the expected limit with error bands without generating pseudo-experiments.

The tool used to extract the limits uses the RooStats [84].

The likelihood function is defined as the product of likelihoods in each channel (channels are six in this specific case) times the product of the p.d.f. for the measurement associated to the nuisance parameters:

$$L(obs, \tilde{\theta}|\mu, \theta) = \prod_{c=1}^{N_{ch}} L_c(obs_c|\mu, \theta) \times \prod_{i=1}^{N_\theta^c} p_i^c(\tilde{\theta}_i^c|\theta_i^c) \tag{4.2}$$

In the analysis of the 7 TeV dataset, a shape analysis is performed and an extended unbinned shape technique is used. The p.d.f. of signal expectation ($f_s(x)$, $x = m_{\mathrm{ZZ}}$) is extracted from a fit to MC simulation according to the method described in Section 4.1, while the background p.d.f ($f_b(x)$, $x = m_{\mathrm{ZZ}}$) is extracted from data in control regions according to the procedure detailed in Section 3.6. These p.d.f.s are used afterwards to formulate the likelihood for this system.

In oder to take into account the effects of systematic uncertainties affecting the background and signal shapes, for each uncertainty three alternative shapes are given as input to the tool: the nominal shape and two more shapes obtained shifting up and down the nuisance parameter related to the uncertainty by 1σ. The tool performs a morphing of the shapes. The source of systematic uncertainties affecting signal and background normalizations, s and b, are taken into account in the determination of the limit assuming log-normal distributions of the corresponding nuisance parameters.

The first term of the likelihood function is a Poissonian evaluated for the observed number of events n given the expected signal and background yields, $s(\theta)$ and $b(\theta)$. The second term is constituted by the signal and background models:

$$L(obs|\mu,\theta) = Poisson(n, \mu \cdot s(\theta) + b(\theta)) \times \prod_{k=1}^{N_{evt}} f(x_k|\mu,\theta)\,, \tag{4.3}$$

where $f(x_k|\mu,\theta)$ is the composition of signal and background models:

$$f(x_k|\mu,\theta) = \frac{\mu \cdot s(\theta)}{\mu \cdot s(\theta) + b(\theta)} f_s(x,\theta) + \frac{b(\theta)}{\mu \cdot s(\theta) + b(\theta)} f_b(x,\theta) \tag{4.4}$$

For the analysis of the 8 TeV dataset the limit on Higgs boson production cross section is determined from the event count only without use of the shape information. The yields of events selected within a window in the reconstructed mass m_{ZZ} (defined as [-6%, +10%] of the value of the mass point m_{H}), for observed data and signal and background expectations, are used to determine cross section limits. Background expectation is extracted from data by applying the procedure described in Section 3.6, and computing the integral of the shape extracted from data in the signal mass window.

In case of a counting experiment the likelihood function assumes a simplified expression and is defined as a Poissonian term evaluated on the number of observed events, n_{obs} and with expected yield given by the sum of the expected signal and background yields, $s(\theta)$ and $b(\theta)$:

$$L(obs|\mu,\theta) = Poisson(n_{obs}, \mu \cdot s(\theta) + b(\theta))\,. \tag{4.5}$$

4.2.1 Results

The observed and expected limits on the ratio of the SM Higgs boson production cross section to the SM expectation are performed between 200 and 600 GeV for the 7 TeV and 8 TeV analysis using the modified frequentist method. Observed limits are indicated by a solid-dotted line and for comparison the expected limits (dashed line) are shown as well, together with the 68% and 95% probability ranges around the expected limits. The SM Higgs boson is said to be excluded at 95% Confidence Level, C.L., when limit on μ drops below one for CL_s = 1 - C.L. = 0.05.

For what concerns the 7 TeV analysis, a further update with respect to the published results has been included [85]. Modified expected yields have been computed because the original analysis used the preliminary integrated luminosity of $4.6 \pm 0.2 pb^{-1}$. This parameter has been replaced by the final number of $4.9 \pm 0.1 pb^{-1}$. New shapes (as from Section 4.1) and yields have been, therefore, used to compute the limit related to the results of the 2011 analysis with the modified frequentist method. Results for the combination of the six channels for update 7 TeV analysis and the 8 TeV analysis are shown in Figures 4.5 and 4.6.

This search alone, with only 4.9 fb^{-1} of collision data collected at 7 TeV, excludes the existence of a resonance with properties as those expected for a SM Higgs boson in the mass range between 340 and 400 GeV at 95% C.L. The analysis performed on 5.1 fb^{-1} of 8 TeV data improves this result enlarging the range of exclusion to 295 and 460 GeV at 95% C.L.

4.3 Observation of a SM Higgs-like boson with the CMS experiment

The searches for a SM Higgs boson carried out in pp collisions at $\sqrt{s} = 7$ TeV in 2011 and at $\sqrt{s} = 8$ TeV in 2012 is based on five main decay modes: $H \to \gamma\gamma$, $H \to ZZ$, $H \to WW$, $H \to bb$ and $H \to \tau\tau$. They have been combined using the same methodology described in the previous section. The results of the analysis of the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ channel have been integrated in the combination of all the channels contributing to set limits in the high mass range. The search sensitivity for a given Higgs boson mass depends on the cross section of the specific production mechanism, but also on the branching fractions of the decay modes playing a role at the considered mass point, and also on the mass resolution and the signal-to-background ratio. On the 4^{th} of July 2012, the observation of a new boson with mass close to 125 GeV was announced by both CMS [86] and ATLAS [87] experiments.

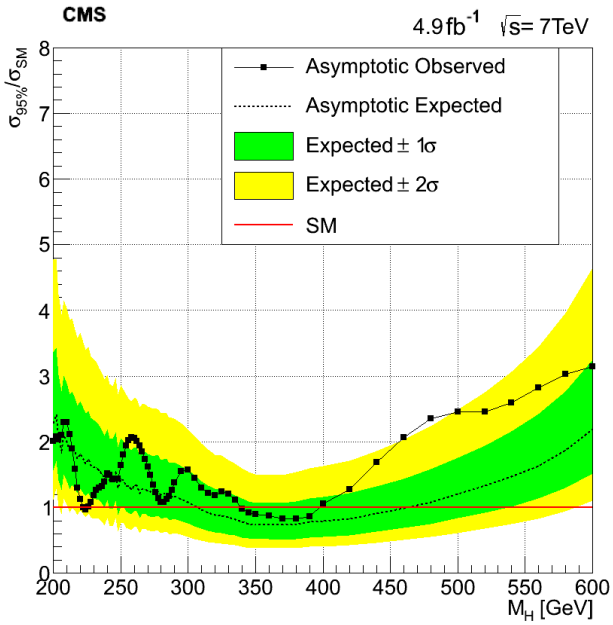


Figure 4.5: The 95% C.L. limits on the Higgs boson production cross section times the branching fractions of $H \to ZZ \to \ell^+\ell^-q\bar{q}$ channel, relative to the SM expectation ($\mu = \sigma/\sigma_{SM}$). Observed (black dots) limits are obtained for the analysis of the 2011 dataset corresponding to an integrated luminosity of 4.9 fb^{-1} at 7TeV [85]. The background-only expectations is indicated by the dashed line. The green and yellow bands indicate the ranged that are expected to contain 68% and 95% of all observed excursions from the expected limit, respectively.

In this region the $H \to \gamma\gamma$ and $H \to ZZ \to \ell^+\ell^-\ell^+\ell^-$ channels play a leading role due to the excellent mass resolution of the reconstructed final states, while for high m_H values the sensitivity is driven by $H \to WW$ and $H \to ZZ$ decay modes.

$H \to ZZ^{(*)} \to 4\ell$

The analysis of the $H \to ZZ^{(*)} \to 4\ell$ channel is focused on the search for a peak in the invariant mass distribution of four-lepton system. It is detailed in the public CMS Physics Analysis Summary [88]. The background is small, the signature of this channel is very clear and helps in background rejection: 4 isolated leptons. Signal from background separation is achieved using a

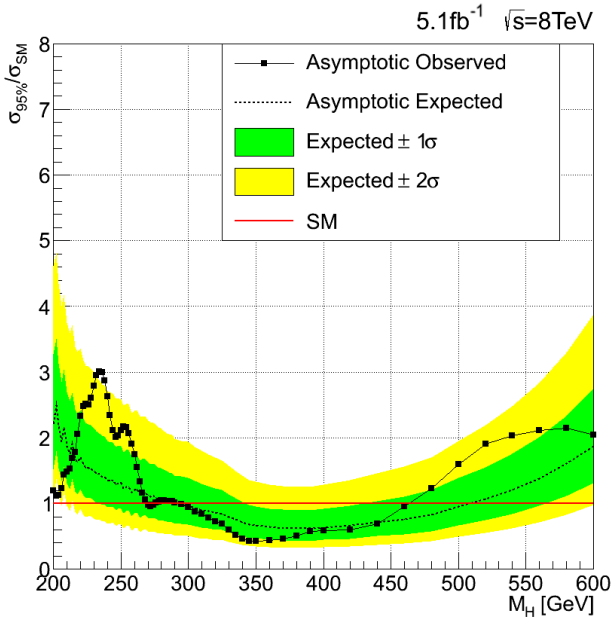


Figure 4.6: The 95% C.L. limits on the Higgs boson production cross section times the branching fractions of $H \to ZZ \to \ell^+\ell^- q\bar{q}$ channel, relative to the SM expectation ($\mu = \sigma/\sigma_{SM}$). Observed (black dots) limits are obtained for the analysis of the 2012 dataset corresponding to an integrated luminosity of 5.1 fb^{-1} at 8TeV. The background-only expectations is indicated by the dashed line. The green and yellow bands indicate the ranged that are expected to contain 68% and 95% of all observed excursions from the expected limit, respectively.

likelihood discriminant built considering the probabilities for the invariant mass of the dilepton pairs and the distribution of five angles that defines completely the kinematics of the four-lepton system in their centre-of-mass frame. Three subchannels are considered separately because of different resolutions and background rates: 4μ, 4e, 2μ2e. The dominating background for this channel are the irreducible ZZ background, that is estimated from simulation, and the reducible and small Z + jets background. The analysis uses pp collision data recorded by CMS and corresponding to an integrated luminosity of 5.1 fb^{-1} at $\sqrt{s} = 7$ TeV and of 19.6 fb^{-1} at $\sqrt{s} = 8$ TeV. The new boson is observed in this channel with a local significance of 4.5 standard deviations. A measurement of its mass gives 126.2 $\pm$ 0.6(stat) $\pm$ 0.2(syst), and a measurement of the signal strength $\mu = \sigma/\sigma_{SM}$, relative to the SM ex-

pectation, is equal to $0.8^{+0.35}_{-0.28}$ at 126 GeV. Figure 4.7 shows the distribution of the four-lepton reconstructed mass in the range [70, 180] GeV. It includes the events from all subchannels. Dots represent data, while background expectation is indicated by filled histograms and signal expectation is indicated by a red-line. The p-value of a measurement tells what is the probability that the observed excess is due to a background fluctuation, quantifying how much it is compatible with background, Figure 4.8 shows the observed p-values for a SM Higgs boson decaying via ZZ to 4ℓ.

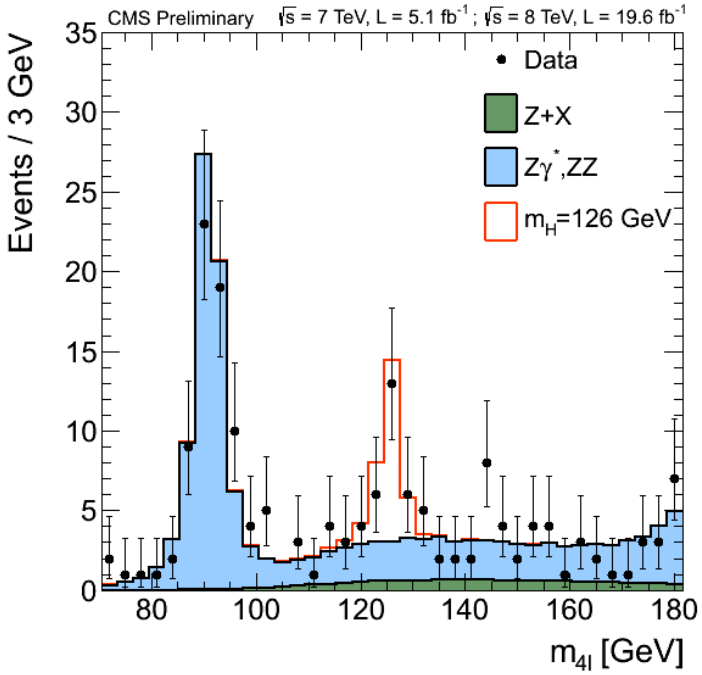


Figure 4.7: Distribution of the 4ℓ reconstructed mass in the range [70, 180] GeV. Dots represent data ($\sqrt{s} = 7$ TeV + $\sqrt{s} = 8$ TeV), while filled histograms represent background expectation and red-line histogram represent a $m_{\mathrm{H}} = 126$ GeV signal expectation [89].

$\mathbf{H \to \gamma\gamma}$

The $\mathrm{H} \to \gamma\gamma$ analysis, described in detail in [90], is focused on the search for a narrow resonance on a smoothly falling continuous background distribution in the diphoton invariant mass spectrum (Figure 4.9). To achieve the best sensitivity, the events are separated into classes based on their mass

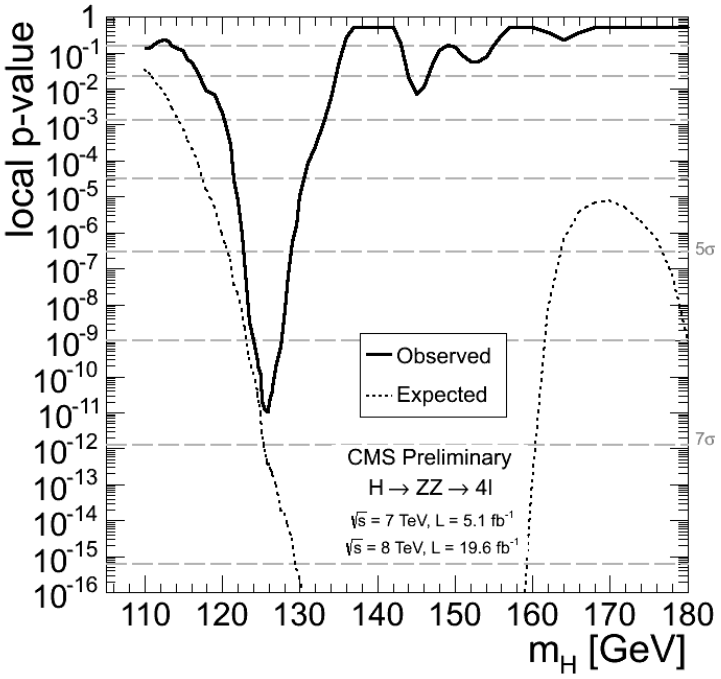


Figure 4.8: P-value of the local excess observed (black line) in the analysis of 4ℓ invariant mass spectrum. SM expectation, for the hypothesis of a Higgs boson of mass equal to 126 GeV, is shown as well (dashed line). P-value is converted in significance and reported on he right y-axis [89].

resolution and signal to background ratio. The classes of events are analyzed separately and the results are then treated simultaneously in the final statistical treatment. The analysis makes use of Multi-Variate Analysis (MVA) techniques both for photon identification and event classification.

The background is estimated from polinomial fits to the observed diphoton mass distributions in each of the event classes. The signal extraction is performed by fitting the diphoton mass spectrum.

The analysis is performed using 7 TeV and 8 TeV datasets for an integrated luminosity of 5.1 fb^{-1} and 19.6 fb^{-1}, respectively. Limits are set on the cross section of the Standard Model Higgs boson decaying to two photons. An excess of events is observed over the SM background for a Higgs boson mass hypothesis of 125 GeV, with a local significance of 3.2 σ, where a local significance of 4.2 σ is expected from a SM Higgs boson (Figure 4.10). The best-fit signal strength, σ/σ_{SM}, is 0.78 ± 0.27 at $m_{\mathrm{H}} = 125$ GeV, and the mass

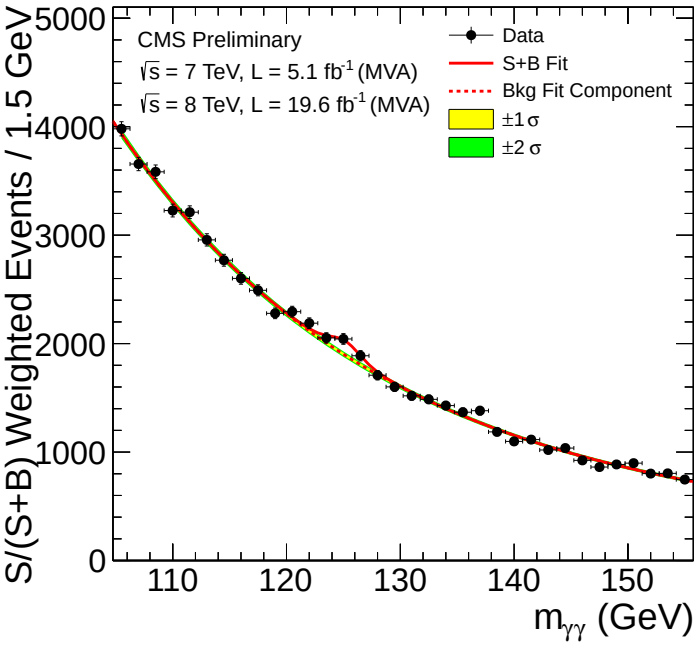


Figure 4.9: Diphoton invariant mass distribution of the selected events for the mass fit MVA analysis on the 7 and 8 TeV datasets combined (dots). Each event is weighted by S/(S+B) value of the category it belongs to. The lines represent the fitted background + signal model (red solid line) and the sole background fit component (red dashed line). The coloured bands represent the $\pm$ 1 and $\pm$ 2 standard deviation uncertainties in the background estimate [91].

is fitted to be 125.4 $\pm$ 0.5 (stat) $\pm$ 0.6 (syst). The result of this channel alone, as well as the $H \to ZZ^{(*)} \to 4\ell$ channel, represents an evidence of a new state. In addition a cut-based analysis is performed as cross check. The two analyses show results which are compatible among themselves and with a Higgs boson with a mass of 124.5 GeV.

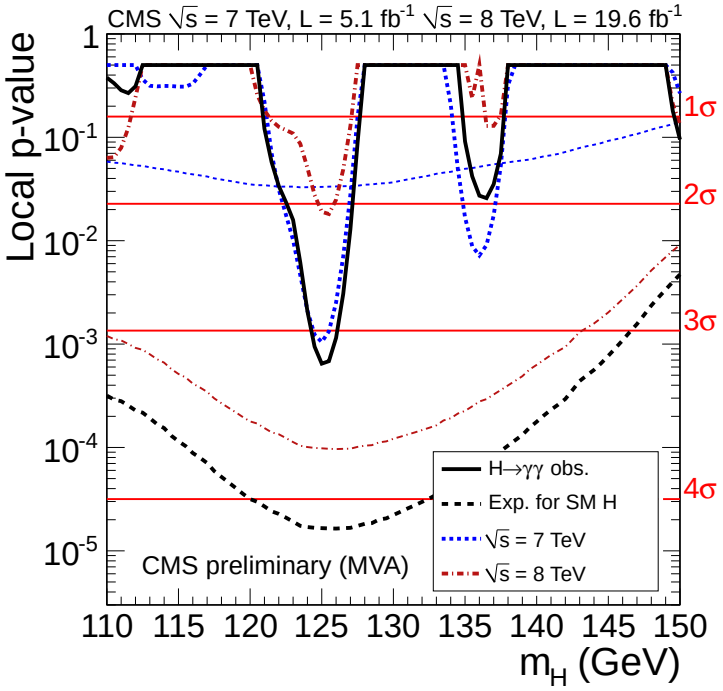


Figure 4.10: Observed local p-values as a function of m_H resulting from the MVA $H \to \gamma\gamma$ analysis for the the 7 and 8 TeV datasets separately (dashed blue line for the 7 TeV analysis and dashed red line for the 8 TeV analysis), and for the combination of the two datasets (black line). The SM expectation for a Higgs boson mass of 126 GeV is illustrated as well (smooth dashed lines) On the right y-axis the corresponding significance is reported [91].

4.3.1 Results

The mass range explored by the combination of Higgs searches at CMS is 110-1000 GeV. The results obtained from the combination exclude the presence of a SM Higgs boson at 95% C.L. in the mass range up to 700 GeV (Figure 4.11), except for a small window around $m_H = 125$ GeV, where a new boson has been observed with a statistical significance of 6.9 σ [92, 93] (Figure 4.12).

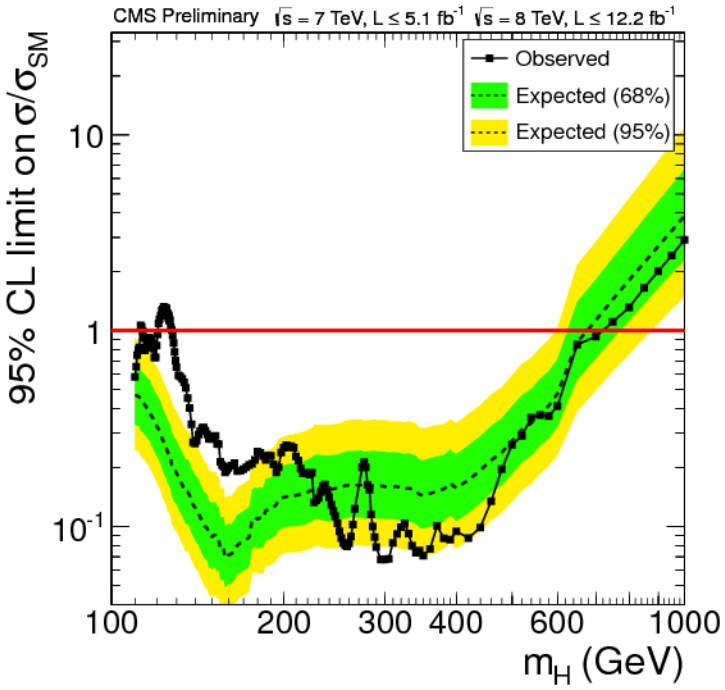


Figure 4.11: The 95% C.L. upper limit on the cross section ratio σ/σ_{SM} for the SM Higgs boson hypothesis as a function of m_{H} for the combination of 5 decay modes: $\mathrm{H} \to \gamma\gamma$, $\mathrm{H} \to \mathrm{ZZ}$, $\mathrm{H} \to \mathrm{WW}$, $\mathrm{H} \to \mathrm{bb}$ and $\mathrm{H} \to \tau\tau$. The observed values are shown by the solid line. The dashed line indicates the expected limits for the background-only hpothesis. The green and yellow bands indicate the ranged that are expected to contain 68% and 95% of all observed excursions from the expected limit, respectively [94].

The very good resolution of $\gamma\gamma$ and 4ℓ channels allows to localize with high precision the observed resonance. A fit for a common mass performed assuming the signal in both channels is due to a same state with unique mass m_X. The result of the fit is shown in Figure 4.13, where the solid line is obtained with all the nuisance parameters profiled and includes both statistical and systematic errors. The final mass measurement is $m_X = 125.8 \pm 0.4(\mathrm{stat}) \pm 0.4(\mathrm{syst})$ GeV.

A compatibility test to the SM prediction is given by the best fit value for the common signal strength modifier $\mu = \sigma/\sigma_{SM}$. Figure 4.14 shows a scan of the best fit value of μ versus the different mass hypotheses. The observed $\hat{\mu}$ value for a Higgs boson mass of 125.8 GeV is found to be 0.88 ± 0.21.

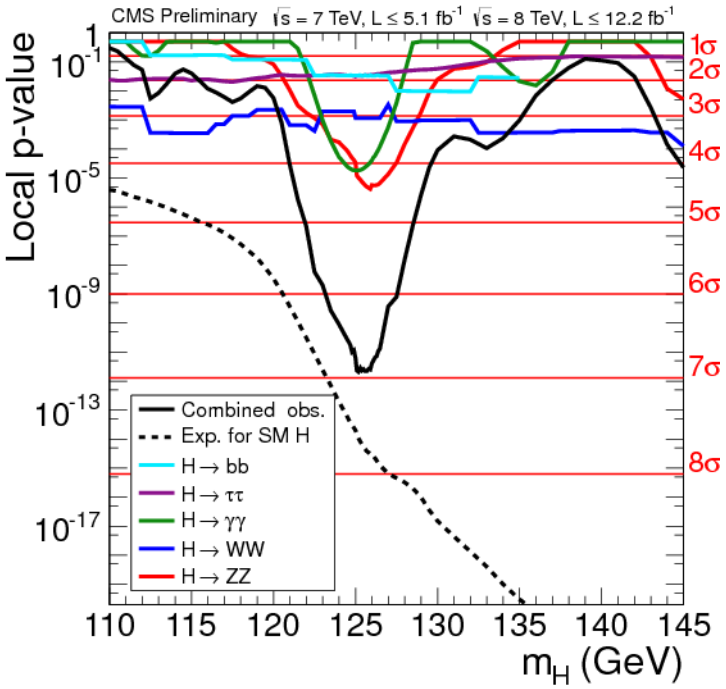


Figure 4.12: The observed local p-values for each of the five decay modes ($H \to \gamma\gamma$, $H \to ZZ$, $H \to WW$, $H \to bb$ and $H \to \tau\tau$) and the overall combination as a function of the SM Higgs boson mass. The dashed line shows the expected p-value for a signal of mass m_H [94].

It is, therefore, possible to state that the observed excess is consistent with the hypothesis of a SM Higgs boson within one standard deviation.

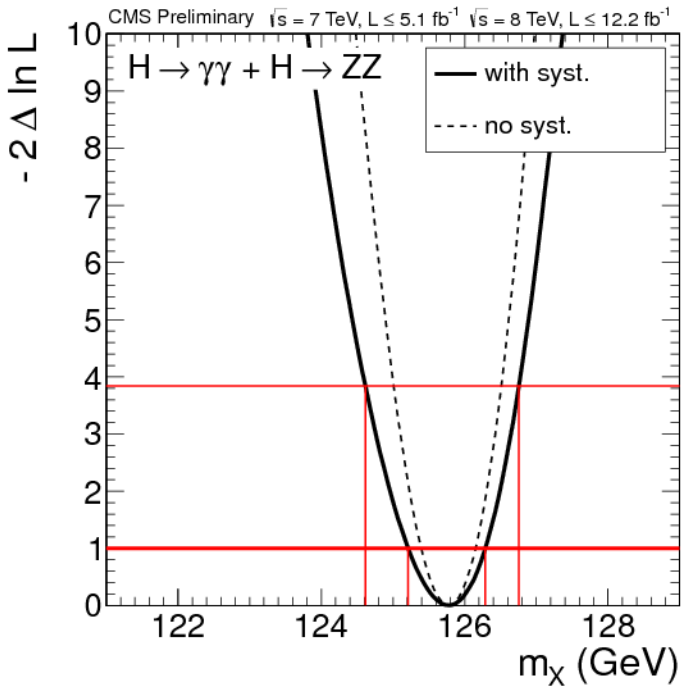


Figure 4.13: 1D-scan of the test statistic $q(m_X)$ versus hypothesized boson mass m_X for the combination of the $\gamma\gamma$ and 4ℓ final states. The solid line includes both statistical and systematic uncertainties. The dashed line includes only statistical uncertainties. The crossings with the thick (thin) horizontal lines define the 68% (95%) CL interval for the measured mass [94].

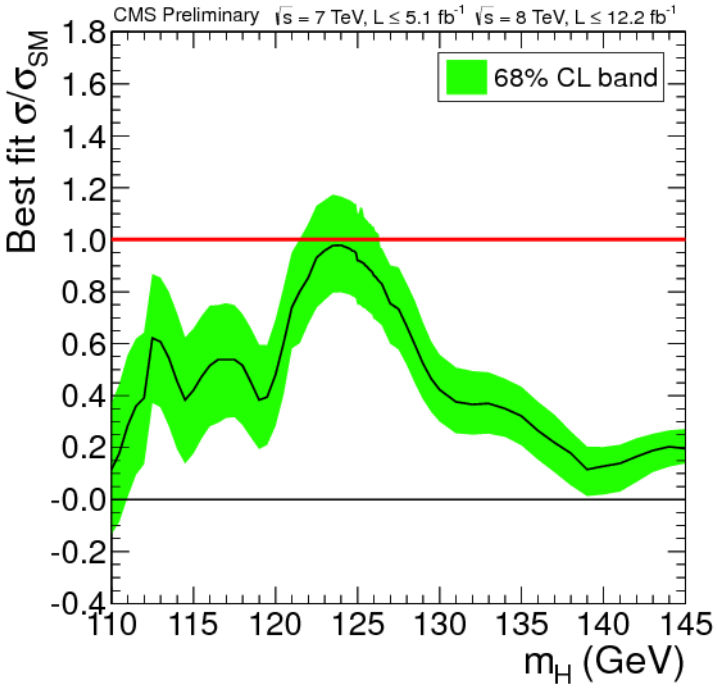


Figure 4.14: The observed best-fit signal strength $\hat{\mu} = \sigma/\sigma_{SM}$ as a function of the SM Higgs boson mass m_{H} in the range 110-145 GeV. The bands correspond to the $\pm 1\sigma$ uncertainties on the $\hat{\mu}$ values [94].

Conclusions

A search for a Standard Model Higgs boson in the $H \to ZZ \to \ell^+\ell^-q\overline{q}$ decay channel has been performed. This decay mode is particularly relevant at high mass because, even if its final state is not as clear as the $H \to ZZ^{(*)} \to 4\ell$ channel, its branching ratio is about a factor 20 greater than the 4ℓ channel. It contributes significantly in a region which has never been probed in any laboratory before the LHC: from 200 to 600 GeV.

The kinematic properties of signal events for this channel can be fully reconstructed because all of the four objects in the final state, two leptons, ℓ, and two jets originated from the hadronization of q and $\overline{q}$, can be detected and identified. Lepton pair as well as jet pair are the decay products of two Z bosons, for this reason the invariant mass of the two systems is expected to be resonant around the nominal Z boson mass [2]. The main sources of background for this channel are represented by the production of a Z boson in association to QCD jets as well as the production of a $t\overline{t}$ pair decaying semileptonically, and the production of diboson events such as ZZ, ZW and WW.

The reconstructed Higgs boson candidates are split in three categories according to the number of jets identified as stemmed from b quarks. This classification allows to isolate a sample of events with better signal-to-background ratio (the category with two b-tagged jets) and to exploit as much as possible information from data combining the results from the three classes which are characterized by a different background composition. A further background suppression is achieved by exploiting the characteristic angular distribution of the Higgs decay products, which reflects the spin correlations among the Higgs boson (spin 0), the two Z bosons (spin 1) and the fermions in the final state (spin 1/2). This correlation is not present in background. Two lepton flavour channels are considered: muonic and electronic channels.

The background expectation is extracted with a data-driven procedure from control samples in data in order to reduce the dependence on the simulation.

The final statistical analysis is performed on the reconstructed ZZ invariant mass, which is expected to be peaked for signal events around the Higgs boson

mass. The modified frequentist method, CL_s, has been used to determine upper limits on the Higgs boson production cross section.

The analyzed data samples are those collected by the CMS experiment in 2011 for an integrated luminosity corresponding to 4.9 fb^{-1} at $\sqrt{s} = 7$ TeV and those collected in 2012 at 8 TeV corresponding to a total luminosity of 5.1 fb^{-1}.

No evidence for a Standard Model Higgs boson has been found and upper limits on the Higgs boson production cross section have been determined in the mass range [200, 600] GeV. With the 7 TeV analysis alone, it has been possible to exclude a Standard Model Higgs boson at 95% CL in the mass range between 340 GeV and 400 GeV. The analysis on 7 TeV dataset has been approved and published in the early 2012 [54]. The exclusion has been confirmed by the analysis of 8 TeV dataset which has allowed to enlarge the range of exclusion between 295 and 460 GeV.

In the next future this analysis will be extended to the low mass region, down to $m_H = 120$ GeV and to the high-mass region up to 1 TeV. Moreover, a dedicated analysis is being developed for the Vector Boson Fusion production in this channel. The entire data sample collected during 2012 ($\sim$20 fb^{-1}) will be processed and analyzed with the same strategy described in this thesis. Concerning the statistical treatment a parametric shape analysis will be performed on the entire 8 TeV dataset instead of using a counting approach. This is expected to improve the sensitivity of the analysis. Finally, the results from 7 and 8 TeV will be combined. A publication of the results of the analysis of the complete dataset is planned by the end of Summer 2013.

Acknowledgements

I would like to thank Luca and Francesco, for their supervision and for sharing their expertise. I am grateful for all the support they provided me. It was a pleasure to collaborate with them.

I am grateful to Prof. Sciacca for the opportunity to join the CMS collaboration and work in the CERN environment.

I also wish to thank my referees, Prof. Santorelli and Dr. Della Volpe, their suggestions allowed to improve this work.

Thanks to all the members of the *2l2q* group, for all the work done together, and to Giovanni and Mingshui, for all the answers to my questions.

Thanks to Benedikt for having welcomed me at CERN and for his constant encouragement.

A thank to all the friends who were with me in the most difficult moments but also in the most pleasant ones, over the last three years: Ale, Titti, Gianni, Orso, Michele, Salvatore, Anna, Roberto, Luca, Nic and *Casa Briganti*!

Grazie Kika, Claudio, Ferdi and Tommy. Voi siete la mia forza.

Appendix A

Simulated samples

Table A.1: Background simulated samples from official CMS production (Summer12) used in the analysis. The equivalent luminosity of the processed events for each sample is computed using the (N)NLO cross section in the 3^{rd} column.

Process	Name	σ [pb]	lumi [fb^{-1}]
Z +1 jet (exclusive)	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	660.6	36.4
Z +2 jet (exclusive)	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	215.1	10.9
Z +3 jet (exclusive)	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	65.79	167.4
Z +4 jet (exclusive)	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/ Summer12-PU_S7_START52_V9-v1/AODSIM	27.59	232.0
$t\bar{t}$	/TTJets_TuneZ2star_8TeV-madgraph-tauola/ Summer12-PU_S7_START52_V5-v1/AODSIM	225.197	29.9
ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12-PU_S7_START50_V15-v1/AODSIM	17.654	555
WZ	/WZ_TuneZ2star_8TeV_pythia6_tauola/ Summer12-PU_S7_START50_V15-v1/AODSIM	22.88	437
WW	/WW_TuneZ2star_8TeV_pythia6_tauola/ Summer12-PU_S7_START50_V15-v1/AODSIM	57.1097	175

Table A.2: Signal samples ($H \to ZZ \to \ell^+\ell^- q\bar{q}$, ℓ = e, μ, τ) simulated with POWHEG.

m_H (GeV)	Name	σ [pb]
200	/GluGluToHToZZTo2L2Q_M-200_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.2566
210	/GluGluToHToZZTo2L2Q_M-210_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.2538
220	/GluGluToHToZZTo2L2Q_M-220_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.2416
230	/GluGluToHToZZTo2L2Q_M-230_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.2278
250	/GluGluToHToZZTo2L2Q_M-250_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.2022
275	/GluGluToHToZZTo2L2Q_M-275_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.1751
300	/GluGluToHToZZTo2L2Q_M-300_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.1563
325	/GluGluToHToZZTo2L2Q_M-325_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.1478
350	/GluGluToHToZZTo2L2Q_M-350_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.1482
375	/GluGluToHToZZTo2L2Q_M-375_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.1360
400	/GluGluToHToZZTo2L2Q_M-400_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.1111
425	/GluGluToHToZZTo2L2Q_M-425_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.0914
450	/GluGluToHToZZTo2L2Q_M-450_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.7311
475	/GluGluToHToZZTo2L2Q_M-475_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.6
500	/GluGluToHToZZTo2L2Q_M-500_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.4719
525	/GluGluToHToZZTo2L2Q_M-525_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.0380
550	/GluGluToHToZZTo2L2Q_M-550_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.0305
575	/GluGluToHToZZTo2L2Q_M-575_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.025
600	/GluGluToHToZZTo2L2Q_M-600_8TeV-powheg-pythia6/ Summer12-PU_S7_START52_V9-v1/AODSIM	0.0201

Bibliography

[1] Precision electroweak measurements and constraints on the standard model. Technical Report arXiv:0911.2604. CERN-PH-EP-2009-023. FERMILAB-TM-2446-E. LEPEWWG-2009-01. ALEPH-2009-001-PHYSICS-2009-001. CDF-Note-9979. D0-Note-6005. DELPHI-DELPHI-2009-PHYS-951. L3-Note-2836. OPAL-PR430. SLAC-PUB-13830, Nov 2009. Also available from `http://arxiv.org/abs/0911.2604`.

[2] J. et al. Beringer. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012. Also available from `http://link.aps.org/doi/10.1103/PhysRevD.86.010001`.

[3] Sheldon Glashow. Partial symmetries of weak interactions. *Nucl. Phys.*, 22:579, 1961. Also available from `http://www.sciencedirect.com/science/article/pii/0029558261904692`.

[4] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264, 1967. Also available from `http://prl.aps.org/abstract/PRL/v19/i21/p1264_1`.

[5] Abdus Salam. Weak and electromagnetic interactions. In W. Svartholm, editor, *Elementary Particle Theory*, page 367, Stockholm, 1968. Almquist and Wiksell. Also available from `www.springerlink.com/index/K82L50T478430475.pdf`.

[6] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. Also available from `http://link.aps.org/doi/10.1103/PhysRevLett.13.321`.

[7] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. Also available from `http://link.aps.org/doi/10.1103/PhysRevLett.13.508`.

[8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964. Also available from http://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[9] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966. Also available from http://link.aps.org/doi/10.1103/PhysRev.145.1156.

[10] T. W. B. Kibble. Symmetry breaking in non-Abelian gauge theories. *Phys. Rev.*, 155:1554–1561, 1967. Also available from http://link.aps.org/doi/10.1103/PhysRev.155.1554.

[11] F.J. Hasert et al. Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment. *Phys.Lett.*, B46:138–140, 1973. Also available from http://www.sciencedirect.com/science/article/pii/0550321374900388.

[12] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV. *Phys.Lett.*, B122:103–116, 1983. Also available from http://www.sciencedirect.com/science/article/pii/0370269383911772.

[13] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963. Also available from http://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[14] Makoto Kobayashi and Toshihide Maskawa. *cp*-violation in the renormalizable theory of weak interaction. *Progress of Theoretical Physics*, 49(2):652–657, 1973. Also available from http://ptp.ipap.jp/link?PTP/49/652/.

[15] Chen-Ning Yang and Robert L. Mills. Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Phys.Rev.*, 96:191–195, 1954. Also available from http://prola.aps.org/abstract/PR/v96/i1/p191_1.

[16] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, and the SLD Electroweak and Heavy Flavour Groups. Precision electroweak measurements and constraints on the Standard Model. Also available from http://lepewwg.web.cern.ch/LEPEWWG, 2010.

[17] Tevatron Electroweak Working Group. 2012 Update of the Combination of CDF and D0 Results for the Mass of the W Boson. 2012. Also available from http://arxiv.org/abs/1204.0042.

[18] T. Aaltonen et al. Combination of the top-quark mass measurements from the Tevatron collider. *Phys.Rev.*, D86:092003, 2012. Also available from http://arxiv.org/abs/1207.1069.

[19] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys.Lett.*, B565:61–75, 2003. Also available from http://www.sciencedirect.com/science/article/pii/S0370269303006142.

[20] Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data. 2011. Also available from http://arxiv.org/abs/1107.5518.

[21] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011. Also available from http://arxiv.org/abs/1101.0593.

[22] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, CERN, Geneva, 2012. Also available from http://arxiv.org/abs/1201.3084.

[23] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008. Also available from http://iopscience.iop.org/1748-0221/3/08/S08001.

[24] CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. Also available from http://iopscience.iop.org/1748-0221/3/08/S08004/.

[25] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. Also available from http://iopscience.iop.org/1748-0221/3/08/S08003.

[26] The LHCb Collaboration. The lhcb detector at the lhc. *J. Instrum.*, 3(LHCb-DP-2008-001. CERN-LHCb-DP-2008-001):S08005, 2008. Also available from http://iopscience.iop.org/1748-0221/3/08/S08005.

[27] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008. Also available from http://iopscience.iop.org/1748-0221/3/08/S08002.

[28] G L et al. Bayatian. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006. Also available from http://cds.cern.ch/record/922757?ln=en.

[29] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. Also available from http://cds.cern.ch/record/331056?ln=en.

[30] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. Also available from http://cds.cern.ch/record/368412?ln=en.

[31] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. Also available from http://cds.cern.ch/record/349375?ln=en.

[32] P. Adzic et al. Energy resolution of the barrel of the cms electromagnetic calorimeter. Technical Report CMS-NOTE-2006-148, CERN, Geneva, Nov 2006. Also available from http://cds.cern.ch/record/1009081.

[33] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. Also available from http://cds.cern.ch/record/357153?ln=en.

[34] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. Also available from http://cds.cern.ch/record/343814?ln=en.

[35] CMS Collaboration. Performance of the CMS Drift Tube Chambers with Cosmic Rays. *JINST*, 5:T03015, 2010. Also available from http://iopscience.iop.org/1748-0221/5/03/T03015.

[36] CMS Collaboration. Performance of the CMS Cathode Strip Chambers with Cosmic Rays. *JINST*, 5:T03018, 2010. Also available from http://iopscience.iop.org/1748-0221/5/03/T03018.

[37] D et al. Piccolo. Resistive plate chambers performance with cosmic rays in the cms experiment. Technical Report CMS-CR-2009-196. 1-3,

CERN, Geneva, Jul 2009. Also available from http://cds.cern.ch/record/1204548?ln=en.

[38] G L et al. Bayatyan. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical Design Report CMS. Also available from http://cds.cern.ch/record/706847?ln=en.

[39] W. Adam et al. The CMS high level trigger. *Eur.Phys.J.*, C46:605–667, 2006. Also available from http://arxiv.org/abs/hep-ex/0512077.

[40] CMS Collaboration. Performance of the cms level-1 trigger during commissioning with cosmic ray muons and lhc beams. *J. Instrum.*, 5(arXiv:0911.5422. CMS-CFT-09-013):T03002. 49 p, Dec 2009. Also available from http://arxiv.org/abs/0911.5422.

[41] G. Abbiendi et al. Muon Reconstruction in the CMS Detector. CMS Analysis Note CMS AN-2008-097, 2008.

[42] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012. Also available from http://arxiv.org/abs/1206.4071.

[43] R Frhwirth. Application of kalman filtering to track and vertex fitting. *Nucl. Instrum. Methods Phys. Res., A*, 262(HEPHY-PUB-503):444. 19 p, Jun 1987. Also available from http://cds.cern.ch/record/178627.

[44] CMS Collaboration. Electron reconstruction and identification at sqrt(s) = 7 tev. Technical Report CMS-PAS-EGM-10-004, CERN, Geneva, 2010. Also available from http://cds.cern.ch/record/1299116?ln=en.

[45] CMS Collaboration. Electron commissioning results at $\sqrt{s}$=7 tev. Technical report, CERN, Mar 2011. Also available from http://cds.cern.ch/record/1360227?ln=en.

[46] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 0804:063, 2008. Also available from http://arxiv.org/abs/0802.1189.

[47] I. Bertram, G. Landsberg, J. Linnemann, R. Partridge, M. Paterno, and H.B. Prosper. A recipe for the construction of confidence limits. Technical Report TM-2104, Fermilab, 2000. Also available from http://lss.fnal.gov/archive/test-tm/2000/fermilab-tm-2104.pdf.

[48] Particle-flow event reconstruction in cms and performance for jets, taus, and met. Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, Apr 2009.

[49] CMS Collaboration. Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7 tev. Technical Report CMS-PAS-PFT-10-002, CERN, Geneva, 2010. Also available from `http://cds.cern.ch/record/1279341?ln=en`.

[50] CMS Collaboration. Particle-flow commissioning with muons and electrons from j/psi and w events at 7 tev. Technical Report CMS-PAS-PFT-10-003, CERN, 2010. Geneva, 2010. Also available from `http://cds.cern.ch/record/1279347?ln=en`.

[51] F. Fabozzi, C.D. Jones, B. Hegner, and L. Lista. Physics analysis tools for the cms experiment at lhc. *Nuclear Science, IEEE Transactions on*, 55(6):3539–3543, 2008.

[52] W. Adam, V. Adler, B. Hegner, L. Lista, S. Lowette, et al. PAT: The CMS physics analysis toolkit. *J.Phys.Conf.Ser.*, 219:032017, 2010. Also available from `http://iopscience.iop.org/1742-6596/219/3/032017`.

[53] A. de Cosa. CMS configuration editor: GUI based application for user analysis job. *J.Phys.Conf.Ser.*, 331:072046, 2011. Also available from `http://iopscience.iop.org/1742-6596/331/7/072046`.

[54] CMS Collaboration. Search for a Higgs boson in the decay channel H to ZZ(*) to q qbar ℓ^- l+ in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1204:036, 2012. Also available from `http://arxiv.org/abs/1202.1416`.

[55] CMS Collaboration. Trigger strategies for Higgs searches in 2011. CMS Analysis Note CMS-AN-2011-065, 2011.

[56] J. Alwall, P Demin, S. de Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D.L. Rainwater, and T. Stelzer. Madgraph/madevent v4: the new web generation. *JHEP*, 09:028, 2007. Also available from `http://arxiv.org/abs/0706.2334`.

[57] P. Nason. A new method for combining nlo qcd with shower monte carlo algorithms. *JHEP*, 11:040, 2004. Also available from `http://arxiv.org/abs/hep-ph/0409146`.

[58] S. Frixione, P. Nason, and C. Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. *JHEP*, 11:070, 2007. Also available from `http://m.iopscience.iop.org/1126-6708/2007/11/070`.

[59] S. Alioli, P. Nason, C. Oleari, and E. Re. Nlo vector-boson production matched with shower in powheg. *JHEP*, 07:06, 2008. Also available from `http://iopscience.iop.org/1126-6708/2008/07/060/pdf/jhep072008060.pdf`.

[60] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur.Phys.J.*, C72:1896, 2012. Also available from `http://link.springer.com/content/pdf/10.1140/epjc/s10052-012-1896-2`.

[61] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys.Lett.*, B659:119–126, 2008. Also available from `http://arxiv.org/abs/0707.1378`.

[62] Vardan Khachatryan et al. Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1101:080, 2011. Also available from `http://arxiv.org/abs/1012.2466`.

[63] CMS Collaboration. Jet energy corrections determination at 7 tev. CMS Physics Analysis Summary CMS-PAS-JME-10-010, 2010. Also available from `http://cds.cern.ch/record/1308178?ln=en`.

[64] CMS Collaboration. b-jet identification in the cms experiment. CMS Physics Analysis Summary CMS-PAS-BTV-11-004, 2011. Also available as `http://cdsweb.cern.ch/record/1427247`.

[65] Yanyan Gao, Andrei V. Gritsan, Zijin Guo, Kirill Melnikov, Markus Schulze, et al. Spin determination of single-produced resonances at hadron colliders. *Phys.Rev.*, D81:075022, 2010. Also available from `http://arxiv.org/abs/1001.3396`.

[66] A. De Rujula, Joseph Lykken, Maurizio Pierini, Christopher Rogan, and Maria Spiropulu. Higgs look-alikes at the LHC. *Phys.Rev.*, D82:013003, 2010. Also available from `http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=21410014`.

[67] A. Bonato, A.V. Gritsan, Z.J. Guo, N.V. Tran, A. Whitbeck. Angular analysis of resonances $pp \to x \to zz$. CMS Analysis Note CMS-AN-2010-351, 2010.

[68] S. Bolognesi, A. Bonato, D. Del Re, A.V. Gritsan, M. Manelli, M. Mozer, F. Pandolfi, N.V Tran, and A. Whitbeck. Search for the standard model higgs boson in the decay channel $h \to zz \to 2\ell 2q$ at cms. *CMS Note*, CMS-AN-2011/388, 2011.

[69] CMS Collaboration. Missing transverse energy performance of the CMS detector. *JINST*, 6:P09001, 2011. Also available from `http://arxiv.org/abs/1106.5048`.

[70] CMS Collaboration. Cms luminosity based on pixel cluster counting - summer 2012 update. CMS Physics Analysis Summary CMS-PAS-LUM-12-001, 2012. Also available from `http://cdsweb.cern.ch/record/1482193`.

[71] L. Borrello et al. "search for the standard model higgs boson in the decay channel $h \to zz \to 2l2b$". CMS Analysis Note CMS-AN-2011-399, 2011.

[72] Michael Hildreth. Estimating systematic errors due to pileup modeling. twiki, 2012. `https://twiki.cern.ch/twiki/bin/view/CMS/PileupSystematicErrors`.

[73] Hung-Liang Lai, Marco Guzzi, Joey Huston, Zhao Li, Pavel M. Nadolsky, et al. New parton distributions for collider physics. *Phys.Rev.*, D82:074024, 2010. Also available from `http://arxiv.org/abs/1007.2241`.

[74] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur.Phys.J.*, C63:189–285, 2009. Also available from `http://arxiv.org/abs/0901.0002`.

[75] Richard D. Ball, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, et al. Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology. *Nucl.Phys.*, B849:296–363, 2011. Also available from `http://arxiv.org/abs/1101.1300`.

[76] CMS Collaboration. Discovery of a new boson in the search for the standard model higgs boson in the hzz4l channel in pp collisions at sqrts=7 and 8 tev. CMS Note 2012/367, 2012. Also available from `http://arxiv.org/abs/1206.3824`.

[77] Stefano Goria, Giampiero Passarino, and Dario Rosco. The Higgs Boson Lineshape. *Nucl.Phys.*, B864:530–579, 2012. Also available from `http://arxiv.org/abs/1112.5517`.

[78] Nikolas Kauer and Giampiero Passarino. Inadequacy of zero-width approximation for a light Higgs boson signal. *JHEP*, 1208:116, 2012. Also available from `http://arxiv.org/abs/1206.4803`.

[79] Giampiero Passarino. Higgs Interference Effects in gg $\rightarrow$ ZZ and their Uncertainty. *JHEP*, 1208:146, 2012. Also available from `http://arxiv.org/abs/1206.3824`.

[80] J.E. Gaiser. *Charmonium Spectroscopy from Radiative Decays of the J/ψ and ψ'*. PhD thesis, Stanford University, 1982. Also available from `http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-255.pdf`.

[81] ATLAS and CMS Collaborations. Procedure for the LHC Higgs boson search combination in Summer 2011. *ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005*, 2011. Also available from `http://cdsweb.cern.ch/record/1363354`.

[82] A. L. Read. Modified Frequentist Analysis of Search Results (The CLs Method). *CERN OPEN*, 2000-205, 2000. Also available from `http://www.sciencedirect.com/science/article/pii/0168900292907945`.

[83] G. Cowan, K. Cranmer, E. Gross, O. Vitells. Asymptotic formulae for likelihood based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. Also available from `http://arxiv.org/abs/1007.1727`.

[84] L. Moneta, K. Belasco, K.S. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf. The RooStats Project. In *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. PoS(ACAT2010)057. Also available from `http://pos.sissa.it/archive/conferences/093/057/ACAT2010_057.pdf`.

[85] CMS Collaboration. Search for a standard-model-like higgs boson with a mass of up to 1 tev at the lhc. (arXiv:1304.0213. CMS-HIG-12-034. CERN-PH-EP-2013-050), Mar 2013. Comments: Submitted to Eur. Phys. J. C.

[86] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett. B*, 716(arXiv:1207.7235. CMS-HIG-12-028. CERN-PH-EP-2012-220):30–61. 59 p, Jul 2012. Also available from `http://arxiv.org/abs/1207.7235`.

[87] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.

Phys.Lett., B716:1–29, 2012. Also available from http://cds.cern.ch/record/1471031?ln=en.

[88] CMS Collaboration. Properties of the higgs-like boson in the decay h to zz to 4l in pp collisions at sqrt s =7 and 8 tev. Technical Report CMS-PAS-HIG-13-002, CERN, Geneva, 2013. Also available from http://cds.cern.ch/record/1523767?ln=en.

[89] CMS Collaboration. Properties of the higgs-like boson in the decay h→zz→4ℓ in pp collisions at $\sqrt{s}$ = 7 and 8 tev. 2013. Available from https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig13002TWiki.

[90] CMS Collaboration. Updated measurements of the higgs boson at 125 gev in the two photon decay channel. Technical Report CMS-PAS-HIG-13-001, CERN, Geneva, 2013. Also available from http://cds.cern.ch/record/1530524/files/HIG-13-001-pas.pdf.

[91] CMS Collaboration. Updated measurements of the higgs-like boson at 125 gev in the two photon decay channel. 2013. Available from https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig13001TWiki.

[92] CMS Collaboration. Observation of a new boson with a mass near 125 gev. Technical Report CMS-PAS-HIG-12-020, CERN, Geneva, 2012. Also available from http://cds.cern.ch/record/1460438?ln=en.

[93] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Phys. Lett. B*, 716(arXiv:1207.7235. CMS-HIG-12-028. CERN-PH-EP-2012-220):30–61. 59 p, Jul 2012. Also available from http://arxiv.org/abs/1207.7235.

[94] CMS Collaboration. Combination of standard model higgs boson searches and measurements of the properties of the new boson with a mass near 125 gev. 2012. Available from https://twiki.cern.ch/twiki/bin/view/CMSPublic/Hig12045TWiki.