Università degli Studi di Torino
Facoltà di Scienze Matematiche Fisiche e Naturali
Corso di Laurea Magistrale in Fisica

Tesi di Laurea Magistrale in Fisica

A new technique to measure the ratio of gg over $q\bar{q}$ induced ZZ production cross sections in the CMS experiment at LHC and its implication for Higgs searches

Candidata:
Martina Machet

Relatore:
Dott. Nicola Amapane

Corelatrici:
Dott.ssa Chiara Mariotti
Dott.ssa Sara Bolognesi

Controrelatrice:
Dott.ssa Stefania Beolè

Anno Accademico 2011-2012

Contents

Introduction

The Standard Model is, up to now, the most successful theory of subatomic elementary particles. Developed in the early 1970s, it provides an elegant mathematical framework which describes how the fundamental constituents of the matter interact between each other, through the electromagnetic, weak and strong forces. Furthermore, it has successfully explained several experimental results and precisely predicted a wide variety of phenomena.
The Standard Model of particle physics predicts the existence of a unique physical Higgs scalar boson associated to the spontaneous electroweak symmetry breaking, whose mass is a free parameter of the theory, and which is regarded as the responsible for the masses of all known elementary particles. In order to confirm this theory it is thus necessary to observe the Higgs boson experimentally.
The "Higgs hunting" is one of the main objectives of the physics program at the Large Hadron Collider (LHC). The Compact Muon Solenoid (CMS) is one of the two largest experiments at LHC and the Higgs search is just the central part of its physics program.
On 4th of July 2012, the discovery of a Higgs boson like particle has been announced by the CMS and ATLAS experiments around a mass of 125 GeV/c^2. The inclusive production of SM Higgs boson followed by the decay $H \to ZZ^{(*)} \to \ell^{\pm}\ell^{\mp}\ell'^{\pm}\ell'^{\mp}$ with $\ell, \ell' = e$ or μ (in short $H \to 4\ell$) has been, as expected, one of the main decay channels for the Higgs boson discovery. In fact, the largest contributors to the overall excess in the combination of the five analyzed decay channels are the $\gamma\gamma$ and ZZ decay modes.

The work presented in this thesis concerns the study of irreducible backgrounds that the search for a Higgs signal in the $H \to 4\ell$ channel has to handle. This kind of background is constituted by four-leptons events from non resonant di-boson production ($q\bar{q} \to ZZ \to 4\ell$ and $gg \to ZZ \to 4\ell$); since these events topology and kinematic are very similar to those of signal events, these processes are considered in the category of *irreducible* backgrounds. An appropriate knowledge of these processes is therefore needed to minimize the systematic uncertainties.
In Chapter 1, after a short introduction to the Standard Model, the attention is focused on Higgs boson properties and the most recent results in the

Higgs search from LEP, Tevatron and LHC. In Chapter 2, after an introduction to the LHC, the CMS detector is described. The overall structure of CMS, consisting of several cylindrical layers closed at both ends by detector disks, allows an excellent lepton reconstruction and particle identification, fundamental for the $H \to ZZ^{(*)} \to 4\ell$ analysis.
In Chapter 3, the $H \to ZZ^{(*)} \to 4\ell$ decay channel analysis is presented. This is the experimental context in which my analysis is included. The 4ℓ final state signal and background processes are described in detail, for a mass range from 110 to 1000 GeV/c^2.
Chapter 4 deals with the irreducible backgrounds of the $H \to ZZ^{(*)} \to 4\ell$ signal. The gluon-gluon induced ZZ background amounts to a non-negligible fraction ($\sim 10\%$) of the total irreducible background at masses above the $2m_Z$ threshold. The aim of my work is to develop for the first time a method to measure this $gg \to ZZ$ fraction. In this chapter, after an introduction on irreducible backgrounds, the study of $q\bar{q} \to ZZ$ and $gg \to ZZ$ angular distributions at generator and reconstruction level, essential to evaluate if the construction of a kinematic discriminant to separate $gg \to ZZ$ from $q\bar{q} \to ZZ$ is possible, is presented.
Finally, in Chapter 5 a new technique to measure the ratio of gg over $q\bar{q}$ induced ZZ production cross sections is presented. After the development of a model which is able to extract, with a fit, the fraction of $gg \to ZZ$ events, the method is tested on a variety of MC samples, in order to study its goodness on varying the statistic, the luminosity and the energy.

Chapter 1

The Standard Model Higgs Boson at LHC

The fundamental components of matter and their interactions are nowadays best described by the Standard Model of Particle Physics (SM) [1] [2], which is based upon two separate quantum field theories, describing the electroweak interaction (Glashow-Weinberg-Salam model or GWS) and the strong interaction (Quantum Chromo-Dynamics or QCD). The SM also predicts the existence of a single physical Higgs scalar boson associated to the spontaneous electroweak symmetry breaking mechanism. The mass m_H of this boson is a free parameter of the theory and it is introduced in order to correctly reproduce the phenomenology of weak interactions. Its introduction in the Lagrangian gives origin to the mass of both fermion and gauge boson fields, in agreement with experimental results. This occurs without explicitly breaking the gauge invariance, thus preserving the renormalisability of the theory.

After a brief introduction of the theoretical framework, in the following Higgs mechanism, theoretical and experimental constraints to the Higgs boson mass and Higgs phenomenology at the LHC collider are described.

1.1 The Standard Model of Elementary Particles

The SM describes the matter as composed by twelve elementary particles, the *fermions*, all having half-integer spin. Fermions can be divided into two main groups, *leptons* and *quarks*, whose classification is given in Table 1.1. Leptons can just interact by electroweak forces, while quarks are subject to both strong and electroweak interactions. Moreover, quarks do not exist as free states, but only as constituents of a wide class of particles, the *hadrons*, such as protons and neutrons.
In the SM, the interactions between elementary particles are mediate by

Table 1.1: *Classification of the three families of fundamental fermions.*

Fermions	1^{st} fam.	2^{nd} fam.	3^{rd} fam.	Charge	Interactions
Quarks	u	c	t	$+\frac{2}{3}$	All
	d	s	b	$-\frac{1}{3}$	
Leptons	e	μ	τ	-1	Weak, E.M.
	ν_e	ν_μ	ν_τ	0	Weak

Table 1.2: *Properties of the three fundamental forces (gravitational interaction is not taken into account).*

	Electromagnetic	Weak	Strong
Quantum	Photon (γ)	$W^{\pm}$, Z	Gluons
Mass [GeV/c^2]	0	80, 90	0
Coupling constant	$\alpha(Q^2=0) \approx \frac{1}{137}$	$\frac{G_F}{(\hbar c)^3} \approx 1.2 \cdot 10^{-5}\ GeV^{-2}$	$\alpha_s(m_Z) \approx 0.1$
Range [cm]	∞	10^{-16}	10^{-13}

bosons, integer-spin particles which are carriers of the fundamental forces. The main characteristics of bosons and of the corresponding interactions are summarised in Table 1.2.
The gravitational interaction is not taken into account, as it is not relevant at the typical mass and distance scales of particle physics.

This complex phenomenology arises from a mathematical formalism according to which the SM is a perturbatively renormalizable quantum field theory (QFT) based on the *local gauge symmetry* of its Lagrangian. According to Noether's theorem, a conservation law corresponds to each of these local invariances, explaining why they are so important. The SM is therefore a local gauge quantum field theory describing three of the four fundamental interactions: electromagnetic, weak and strong interaction. It is based on the symmetry group

$$SU(3)_C \otimes SU(2)_I \otimes U(1)_Y,$$

the direct product of $SU(3)_C$, the color symmetry group upon which Quantum Chromo Dynamics (QCD) is built, the gauge groups of weak isospin, $SU(2)_I$, and hypercharge, $U(1)_Y$. Electromagnetic and weak interactions are unified in the *electroweak* gauge group $SU(2)_I \otimes U(1)_Y$, upon which the Glashow-Weinberg-Salam Model is built.
Despite this simmetry predicts with precision and accuracy the phenomenology of particle interactions, it is broken by the mass terms of the Lagrangian.

A necessary ingredient of the SM is therefore a *symmetry breaking* mechanism that allows to introduce the mass terms in a local gauge invariant Lagrangian.

1.2 The Electroweak Theory

From a historical point of view, the starting point for the study of electroweak interactions is Fermi's theory of muon decay [3], which is based on an effective four-fermion Lagrangian:

$$\mathcal{L} = -\frac{4G_F}{\sqrt{2}}\bar{\nu}_\mu\gamma^\alpha\frac{1-\gamma_5}{2}\mu\bar{e}\gamma_\alpha\frac{1-\gamma_5}{2}\nu_e, \tag{1.1}$$

where G_F is the Fermi coupling constant reported in Table 1.2.
Equation 1.1 represents a "point-like" interaction, with only one vertex and without any intermediate boson exchanged. It is usually referred to as $V-A$ interaction, being formed by a vectorial and an axial component. The term $\frac{1}{2}(1-\gamma_5)$ that appears in it is the negative helicity projector. Only the negative helicity (*left-handed*) component of fermions takes part to this interaction, corresponding to the projection of spin opposite to particle's motion.

Fermi's Lagrangian is not renormalisable and it results in a non-unitary scattering matrix. Both problems of renormalisability and unitarity are overcome, as already said, requiring the weak interaction Lagrangian to be invariant under local transformations generated by the elements of a Lie group (*gauge transformations*). The resulting Lagrangian must in fact reduce to Equation 1.1 in the low energy limit.
A gauge theory for weak interactions is conceived as an extension of the theory of electromagnetic interaction, the Quantum Electro-Dynamics (QED), which is based on the gauge group $U(1)_{EM}$, associated to the conserved quantum number Q (electric charge). In this case, the condition of local invariance under the $U(1)_{EM}$ group leads to the existence of a massless vector boson, the *photon*.
A theory reproducing both the electromagnetic and weak interaction phenomenology is achieved by extending the gauge symmetry to the group $SU(2)_I \otimes U(1)_Y$. In this sense, the weak and electromagnetic interactions are said to be unified. The generator of $SU(2)_I$ is the weak isospin operator and the generator of $U(1)_Y$ is the weak hypercharge Y operator. The corresponding quantum numbers satisfy the Gell-Mann-Nishijima formula

$$Q = I_3 + \frac{Y}{2},$$

where I_3 is the third component of the weak isospin. Fermions can be divided in doublets of negative-helicity (*left-handed*) particles and singlets of positive-

helicity (*right-handed*) particles, as follows:

$$L_L = \begin{pmatrix} \nu_{\ell,L} \\ \ell_L \end{pmatrix},\ \ell_R,\ Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix},\ u_R,\ d_R, \tag{1.2}$$

where $\ell = e, \mu, \tau,\ u = u, c, t$ and $d = d, s, b$. Neutrinos have no right component, as their mass is taken as null. In Table 1.3, $I_3,\ Y$ and Q quantum numbers of all fermions are reported. As well as for QED, the requirement of

Table 1.3: *Isospin (I_3), hypercharge (Y) and electric charge (Q) of all fermions.*

	I_3	Y	Q
$\begin{pmatrix} u_L \\ d_L \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} \frac{1}{3} \\ -\frac{1}{3} \end{pmatrix}$	$\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$
$u_R,\ d_R$	$0,\ 0$	$\frac{4}{3},\ -\frac{2}{3}$	$+\frac{2}{3},\ -\frac{1}{3}$
$\begin{pmatrix} \nu_{\ell,L} \\ \ell_L \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$\begin{pmatrix} -1 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$
ℓ_R	0	-2	-1

local gauge invariance with respect to the $SU(2)_I \otimes U(1)_Y$ group introduces now four massless vector fields (gauge fields), $W_\mu^{1,2,3}$ and B_μ, which couple to fermions with two different coupling constants, g and g'. Note that B_μ does not represent the photon field, because it arises from the $U(1)_Y$ group of hypercharge, instead of $U(1)_{EM}$ group of electric charge. The gauge-invariant Lagrangian for fermion fields can be written as follows:

$$\mathcal{L} = \bar{\Psi}_L \gamma^\mu \Big(i\partial_\mu + g t_a W_\mu^a - \frac{1}{2} g' Y B_\mu \Big) \Psi_L + \bar{\psi}_R \gamma^\mu \Big(i\partial_\mu - \frac{1}{2} g' Y B_\mu \Big) \psi_R, \tag{1.3}$$

where

$$\Psi_L = \begin{pmatrix} \Psi_L^1 \\ \Psi_L^2 \end{pmatrix}$$

and where Ψ_L and Ψ_R are summed over all the possibilities in Equation 1.2. As already stated, $W_\mu^{1,2,3}$ and B_μ do not represent physical fields, which are given instead by linear combinations of the four mentioned fields: the charged bosons W^+ and W^- correspond to:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right), \tag{1.4}$$

while the neutral bosons γ and Z correspond to

$$A_\mu = B_\mu cos\theta_W + W_\mu^3 sin\theta_W \tag{1.5}$$

$$Z_\mu = -B_\mu sin\theta_W + W_\mu^3 cos\theta_W \tag{1.6}$$

obtained by mixing the neutral fields W_μ^3 and B_μ with a rotation defined by the Weinberg angle θ_W. In terms of the fields in Equations 1.4 to 1.6, the interaction term between gauge fields and fermions, taken from the Lagrangian in Equation 1.3, becomes

$$\mathcal{L}_{int} = \frac{1}{2\sqrt{2}} g(J_\alpha^+ W^{(+)\alpha} + J_\alpha^- W^{(-)\alpha}) + \frac{1}{2}\sqrt{g'^2 + g^2} J_\alpha^Z Z^\alpha - e J_\alpha^{EM} A^\alpha, \quad (1.7)$$

where J^{EM} is the electromagnetic current connected to the photon field, while J^+, J^- and J^Z are the three weak isospin currents. It is found that

$$J_\alpha^Z = J_\alpha^3 - 2sin^2\theta_W \cdot J_\alpha^{EM}.$$

A^α can then be identified with the photon field and, requiring the coupling terms to be equal, one obtains

$$g \sin\theta_W = g' \cos\theta_W = e \quad (1.8)$$

which represents the electroweak unification (e is the electron charge). The GWS model thus predicts the existence of two charged gauge fields, which only couple to left-handed fermions, and two neutral gauge fields, which interact with both left- and right-handed components.

1.3 The Higgs mechanism

In order to correctly reproduce the phenomenology of weak interactions, both fermion and gauge boson fields must acquire mass, in agreement with experimental results. Up to this point, however, all particles are considered massless: in the electroweak Lagrangian, in fact, a mass term for the gauge bosons would violate gauge invariance, which is needed to ensure the renormalizability of the theory. Masses are thus introduced with the *Higgs mechanism* [4], which allows fermions and $W^\pm$, Z bosons to be massive, while keeping the photon massless. Such mechanism is accomplished by means of a doublet of complex scalar fields,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix}, \quad (1.9)$$

which is introduced in the electroweak Lagrangian within the term

$$\mathcal{L}_{EWSB} = (D^\mu \phi)^\dagger (D_\mu \phi) + V(\phi^\dagger \phi), \quad (1.10)$$

where $D_\mu = \partial_\mu - igt_a W_\mu^a + \frac{i}{2} g' Y B_\mu$ is the covariant derivative. The Lagrangian in Equation 1.10 is invariant under $SU(2)_I \otimes U(1)_Y$ transformations, since the kinetic part is written in terms of covariant derivatives and

the potential V only depends on the product $\phi^\dagger\phi$. The ϕ field is characterised by the following quantum numbers:

	I_3	Y	Q
$\begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}$	$\begin{pmatrix}\frac{1}{2}\\ -\frac{1}{2}\end{pmatrix}$	$\begin{pmatrix}1\\ 1\end{pmatrix}$	$\begin{pmatrix}1\\ 0\end{pmatrix}$

Writing the potential term as follows (see also Figure 1.1 for a graphical representation)

$$V(\phi^\dagger\phi) = -\mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \tag{1.11}$$

with $\mu^2 < 0$ and $\lambda > 0$, it results to have a minimum for

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}. \tag{1.12}$$

This minimum is not found for a single value of ϕ, but for a manifold of non-zero values. The choice of (ϕ^+, ϕ^0) corresponding to the ground state, i.e. the lowest energy state or vacuum, is arbitrary, and the chosen point is not invariant under rotations in the (ϕ^+, ϕ^0) plane: this is referred to as *spontaneous symmetry breaking*. If one chooses to fix the ground state on the ϕ^0 axis, the vacuum expectation value of the ϕ field is

$$<\phi> = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v\end{pmatrix}, \quad v^2 = -\frac{\mu^2}{\lambda}. \tag{1.13}$$

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

$$\phi = \frac{1}{\sqrt{2}} e^{\frac{i}{v}\phi^a t_a}\begin{pmatrix}0\\ H+v\end{pmatrix}, \quad a = 1, 2, 3,$$

where the three fields ϕ^a and the fourth $\phi^4 = H + v$ are called *Goldstone fields*. Being scalar and massless, they introduce four new degrees of freedom, in addition to the six degrees due to the transverse polarizations of the massless vector bosons $W^\pm$ and Z. The unitary gauge is fixed by the transformation

$$\phi' = e^{-\frac{i}{v}\phi^a t_a}\phi = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ H+v\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ \phi^4\end{pmatrix}.$$

The remaining field, the *Higgs field*, has now a zero expectation value.
Rewriting the Lagrangian in Equation 1.10 with the ϕ field in the unitary gauge, $\mathcal{L}_{EWSB}$ results from the sum of three terms:

$$\mathcal{L}_{EWSB} = \mathcal{L}_H + \mathcal{L}_{HW} + \mathcal{L}_{HZ}, \tag{1.14}$$

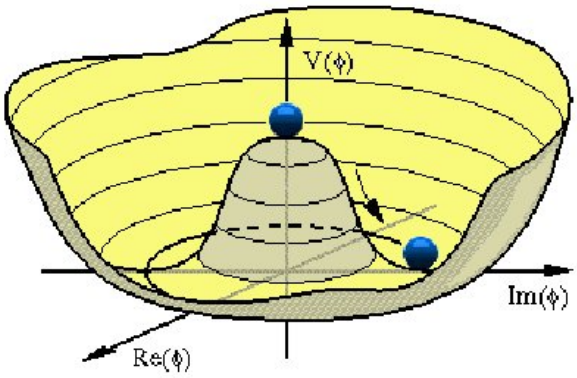

Figure 1.1: *Shape of the Higgs potential of Equation.*

where the three terms can be written as follows, using the approximation $V \sim \mu^2 H^2 + const$ and neglecting higher order terms:

$$\begin{aligned} \mathcal{L}_H &= \frac{1}{2}\partial_\alpha H \partial^\alpha H + \mu^2 H^2 \\ \mathcal{L}_{HW} &= \frac{1}{4}v^2 g^2 W_\alpha W^{\dagger\alpha} + \frac{1}{2}v g^2 H W_\alpha W^{\dagger\alpha} \qquad (1.15) \\ &= m_W^2 W_\alpha W^{\dagger\alpha} + g_{HW} H W_\alpha W^{\dagger\alpha} \\ \mathcal{L}_{HZ} &= \frac{1}{8}v^2(g^2+g'^2) Z_\alpha Z^\alpha + \frac{1}{4}v(g^2+g'^2) H Z_\alpha Z^\alpha \qquad (1.16) \\ &= \frac{1}{2}m_Z^2 Z_\alpha Z^\alpha + \frac{1}{2}g_{HZ} H Z_\alpha Z^\alpha \end{aligned}$$

Equations 1.15 and 1.16 now contain mass terms for fields $W^\pm$ and Z: each of the three gauge bosons has acquired mass and an additional degree of freedom, corresponding to the longitudinal polarization. At the same time, three of the four Goldstone bosons have disappeared from the Lagrangian $\mathcal{L}_{EWSB}$, thus preserving the total number of degrees of freedom: the degrees related to the missing Goldstone bosons have become the longitudinal degrees of the vector bosons. Only the H scalar field is still present and has acquired mass itself: it is the Higgs boson.
Summarizing, the Higgs mechanism is used to introduce the weak boson masses without explicitly breaking the gauge invariance and thus preserving the renormalizability of the theory. When a symmetry is "spontaneously" broken, in fact, it is not properly eliminated: it is rather "hidden" by the choice of the ground state. It can be shown that the minimum of the Higgs field is still invariant under the $U(1)_{EM}$ group. The electromagnetic symmetry is therefore unbroken and the photon does not couple to the Higgs boson and remains massless.

1.3.1 Vector Boson Masses and Couplings

Equations 1.15 and 1.16 show that the masses of vector bosons $W^{\pm}$ and Z are related to the parameter v, characteristic of the $EWSB$, and to the electroweak coupling constants:

$$\begin{cases} m_W = \frac{1}{2}vg \\ m_Z = \frac{1}{2}v\sqrt{g^2+g'^2} \end{cases} \rightarrow \frac{m_W}{m_Z} = \frac{g}{\sqrt{g^2+g'^2}} = cos\theta_W. \tag{1.17}$$

Also the couplings of vector bosons to the Higgs can be obtained from Equations 1.15 and 1.16, and are found to depend on the square of m_W and m_Z:

$$g_{HW} = \frac{1}{2}vg^2 = \frac{2}{v}m_W^2 \tag{1.18}$$

$$g_{HZ} = \frac{1}{2}v(g^2+g'^2) = \frac{2}{v}m_Z^2. \tag{1.19}$$

A relation between the decay ratios of the Higgs boson to a W pair and to a Z pair can be derived from Equations 1.18 and 1.19:

$$\frac{BR(H \rightarrow W^+W^-)}{BR(H \rightarrow ZZ)} = \left(\frac{g_{HW}}{\frac{1}{2}g_{HZ}}\right)^2 = 4\left(\frac{m_W^2}{m_Z^2}\right)^2 \simeq 2.4.$$

Finally, the $EWSB$ energy scale can be determined from the relation between the v parameter and the Fermi constant G_F:

$$v = \left(\frac{1}{\sqrt{2}G_F}\right)^{\frac{1}{2}} \simeq 246\ GeV. \tag{1.20}$$

1.3.2 Fermion Masses and Couplings

The Higgs mechanism is also used to generate the fermion masses, by introducing in the SM Lagrangian an $SU(2)_I \otimes U(1)_Y$ invariant term (called *Yukawa term*) which represents the interaction between the Higgs and the fermion fields. Since ϕ is an isodoublet, while the fermions are divided in left-handed doublet and right-handed singlet, the Yukawa terms (one for each fermion generation) must have the following expression for leptons:

$$\mathcal{L}_\ell = -G_{H\ell} \cdot \bar{l}_\ell \phi \ell_R + \bar{\ell}_R \phi^\dagger l_\ell \,.$$

In the unitary gauge, the first component of ϕ is zero, therefore a mass term will arise from the Yukawa Lagrangian only for the second component of l : this correctly reproduces the fact that neutrino is (approximately) massless.

$$\mathcal{L}_\ell = -\frac{G_{H\ell}}{\sqrt{2}}v\bar{\ell}\ell - \frac{G_{H\ell}}{\sqrt{2}}H\bar{\ell}\ell \,. \tag{1.21}$$

For what concerns the quark fields, the *down* quarks (*d, s, b*) are treated in the same way as leptons; *up* quarks (*u, c, t*), instead, must couple to the charge-conjugate of ϕ

$$\phi^c = -i\tau_2\phi^* = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi^3 - i\phi^4 \\ -\phi^1 + i\phi^4\end{pmatrix}$$

which becomes in the unitary gauge

$$\phi^c = \frac{1}{\sqrt{2}}\begin{pmatrix}\eta + v \\ 0\end{pmatrix}$$

The Yukawa Lagrangian will be therefore

$$\mathcal{L}_Y = -G_{H\ell}\overline{L}_L\phi\ell_R - G_{Hd}\overline{Q}_L\phi d_R - G_{Hu}\overline{Q}_L\phi^c u_R + h.c.. \tag{1.22}$$

From Equation 1.21, the mass of a fermion (apart from neutrinos) and its coupling constant to the Higgs boson are found to be

$$m_f = \frac{G_{Hf}}{\sqrt{2}}v \tag{1.23}$$

$$g_{Hf} = \frac{G_{Hf}}{\sqrt{2}} = \frac{m_f}{v}. \tag{1.24}$$

Being G_{Hf} free parameters, fermions mass cannot be predicted by the theory.

1.4 Higgs Boson Mass

The Higgs boson mass is the only yet unknown free parameter of the SM. The tree-level relation for the Higgs mass reads $m_H^2 = 2\lambda v^2$, thus Higgs mass depends on the parameters v and λ. The former, which is the Higgs expectation value, can be estimated by its relation with the constant G_F of Fermi's theory, while the latter can be determined only measuring the Higgs mass itself, because it is characteristic of the field ϕ. However theoretical indications exist and experimental constraints, from direct and indirect searches at other colliders, narrow the possible range.

1.4.1 Theoretical bounds on m_H

An upper bound on the Higgs mass comes from unitarity arguments. In fact, considering the elastic scattering of longitudinal Z bosons:

$$Z_L Z_L \to Z_L Z_L$$

the unitarity bound on the corresponding amplitude, in the limit $s \ll m_Z^2$, implies

$$m_H < \sqrt{\frac{16\pi}{3}}v \sim 1\ TeV/c^2$$

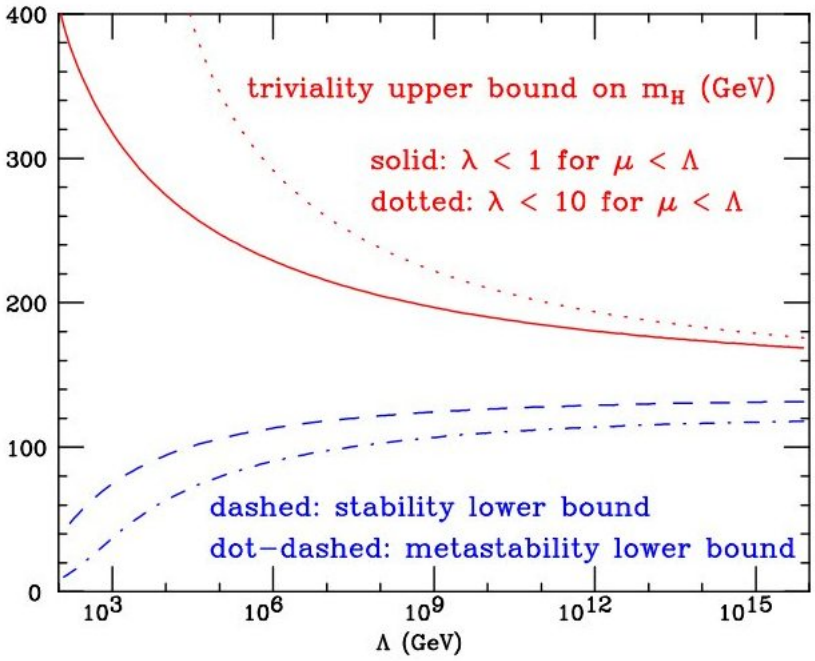

Figure 1.2: *Red line: triviality bound (for different upper limits to λ); blue line: vacuum stability (or metastability) bound on the Higgs boson mass as a function of the new physics (or cut-off) scale Λ [5].*

Considering other scattering processes, such as $Z_L W_L \rightarrow Z_L W_L$, more restrictive bounds ($\sim 800\ GeV/c^2$) can be found.
Moreover constraints to the Higgs boson mass can be found by imposing the energy scale Λ up to which the SM is valid, before the perturbation theory breaks down and non-SM phenomena emerge [5]. An upper limit is obtained requiring that the running quartic coupling of Higgs potential λ remains finite up to the scale Λ (*triviality*). On the other hand, a lower limit is found by requiring that λ remains positive after the inclusion of radiative corrections, at least up to Λ: this implies that the Higgs potential is bounded from below, that is the minimum of such potential is an absolute minimum (*vacuum stability*). A looser constraint is furthermore obtained by requiring such minimum to be local, instead of absolute (*metastability*). These theoretical bounds on the Higgs boson mass as a function of Λ are shown in Figure 1.2. If the validity of the SM is assumed up to the Plank scale ($\Lambda \sim 10^{19}\ GeV$), the allowed Higgs boson mass range is between 130 and 190 GeV/c^2, while for $\Lambda \sim 1\ TeV$ the Higgs mass can be up to 700 GeV/c^2.

1.4.2 Experimental bounds on m_H

First experimental constraints on the Higgs mass have been provided by measurements at LEP [6], SLC [7] and Tevatron [8]. In fact, a lower limit of 114.4 GeV/c^2 (95% C.L.) has been set by direct searches at LEP [9], while more recent results from the Tevatron exclude the mass range from 156 to 177 GeV/c^2 (95% C.L.) [10]. Moreover, since the Higgs boson contributes to radiative corrections, many electroweak observables have a logarithmic dependence on m_H and can therefore be used to constraint its mass. All the precision electroweak measurements performed by the four LEP experiments and by SLD, CDF and DØ [11] [12] have been combined together and fitted, assuming the SM as the correct theory and using the Higgs mass as free parameter. The result of this combination is summarised in Figure 1.3, where $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ is plotted as a function of m_H. The black solid curve represents the result of the fit, while the shaded band is the theoretical uncertainty due to unknown higher order corrections and the yellow area shows the regions excluded by LEP and Tevatron measurements. The indirectly

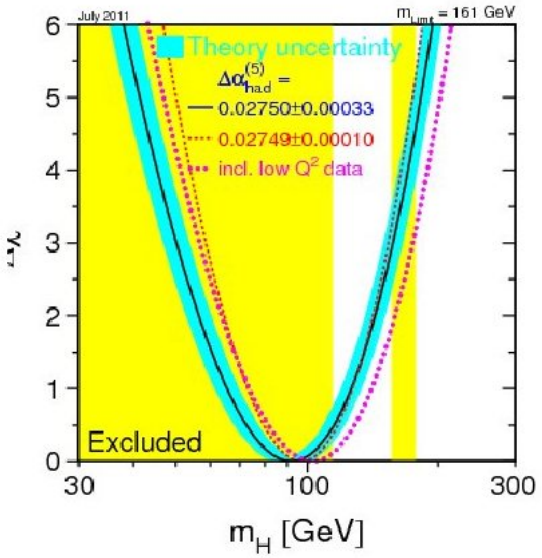

Figure 1.3: $\Delta\chi^2$ *of the fit to the electroweak precision measurements performed at LEP, SLC and Tevatron as a function of the Higgs boson mass (July 2011). The black, solid line represents the result of the fit, and the blue, shaded band is the theoretical error from unknown higher-order corrections. The yellow area represents the regions excluded by LEP-II and Tevatron.*

measured value of the Higgs boson mass, corresponding to the minimum of the curve, is $m_H = 92^{+34}_{-26}\ GeV/c^2$, where the errors represent the experimen-

tal uncertainty at 68% C.L. derived from the black line, thus not taking the theoretical uncertainty into account. An upper limit of 161 GeV/c^2 can also be set (95% $C.L.$) including the theoretical uncertainty. This limit increases to 185 GeV/c^2 including the direct search limit of 114.4 GeV/c^2.
Starting 23 November 2009 a new collider, the Large Hadron Collider, is running (see Chapter 2). One of its main goals is just the Higgs boson discovery, and in July 2012 first results obtained at a centre-of-mass energy of 7 and 8 TeV have been presented, showing the evidence for a new boson at a mass of 125 GeV/c^2 [13].
In the following Higgs phenomenology at the LHC collider and recent LHC results are reported.

1.5 Standard Model Higgs Boson search at LHC

The main goal of the LHC project is certainly the search for the SM Higgs boson in the mass range going from 100 GeV/c^2 to about 1 TeV. In this section the main Higgs production and decay processes are described.
While the Higgs boson mass cannot be predicted by the theory, the Higgs couplings to the fermions and bosons are predicted to be proportional to the corresponding particle masses as in Equations 1.18, 1.19 and 1.24. For this reason the Higgs production and decay processes are dominated by channels involving the coupling of Higgs boson to heavy particles, especially to $W^{\pm}$ and Z bosons and to the third generation of fermions. About the other gauge bosons, the Higgs does not couple to photons and gluons at tree level, but only by one-loop diagram, where the main contribution is given by t loops, for the $gg \to H$ channel, and by W^+W^- and t loops for the $\gamma\gamma \to H$ channel.

1.6 Higgs Boson production

In proton-proton collisions at $\sqrt{s} = 7 - 14\ TeV$, like those at the Large Hadron Collider, the main processes contributing to the Higgs boson production are represented by the Feynman diagrams in Figure 1.4. In Figure 1.5 the Higgs cross sections for the different production mechanisms are shown as a function of m_H for a centre-of-mass energy of 14 TeV, the nominal value at the LHC [14].
Below the main Higgs production mechanisms are described.

1.6.1 Gluon-gluon fusion

The gg fusion is the dominating mechanism for the Higgs boson production at the LHC over the whole mass range. The process is shown in Figure 1.4a, with a t quark-loop as the main contribution. The latest results in

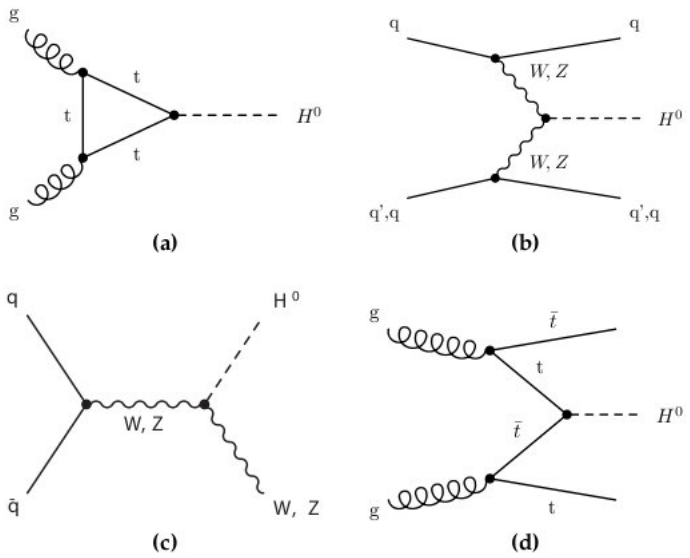

Figure 1.4: *Higgs boson production mechanisms at tree level in proton-proton collisions: (a) gluon-gluon fusion; (b) VV fusion; (c) W and Z associated production (or Higgsstrahlung); (d) $t\bar{t}$ associated production.*

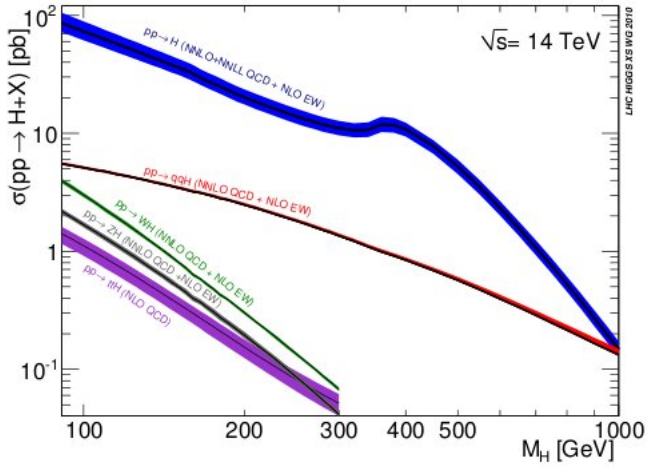

Figure 1.5: *Cross sections for the different Higgs boson production channels, as functions of the Higgs boson mass, at 14 TeV LHC centre-of-mass energy [14].*

the computation of the cross section for this process include next-to-next-to-leading order (NNLO) QCD contributions, complemented with next-to-next-to-leading log (NNLL) resummation, and next-to-leading order (NLO) electroweak corrections. The theoretical uncertainties of the total cross section can be estimated as 20% at NNLO due to the residual scale dependence, the uncertainties of the parton density functions (PDFs), and neglected quark mass effects.

1.6.2 Vector boson fusion

The VV fusion (Figure 1.4b) is the second rate mode and it is proportional to the WWH (ZZH) coupling. It is about one order of magnitude lower than gg-fusion for a large range of m_H values and the two processes become comparable only for Higgs masses of the order of 1 TeV/c^2. However, this channel is very interesting because of its clear experimental signature: the presence of two spectator jets with high invariant mass in the forward region provides a powerful tool to tag the signal events and discriminate the backgrounds, improving in this way the signal to background ratio, despite the low cross section. For this process NNLO cross sections are known with low uncertainties and QCD (NNLO) and EW (NLO) corrections are quite small.

1.6.3 Associated production

The *Higgsstrahlung* process (Figure 1.4c) is lower than those of gg-fusion and VBF, the Higgs boson is produced in association with a $W^{\pm}$ or Z boson, which can be used to tag the event. The QCD corrections are quite large and the NLO cross section is increased by a factor of $1.2 \div 1.4$ with respect to the leading order one. This process is also known at NNLO (QCD) and at NLO (EW).
The last process, illustrated in Figure 1.4d, is the associated production of a Higgs boson with a $t\bar{t}$ pair. In this case the cross section is much lower than gg and VV fusion ones, but the presence of the $t\bar{t}$ pair in the final state can provide a good experimental signature. The higher order corrections increase the cross section by a factor of about 1.2.

1.7 Higgs Boson decay channels

The branching ratios of the different Higgs decay channels are shown in Figure 1.6 as a function of the Higgs mass. Indeed, not just its production but also its decaying modes change dramatically across the possible mass range. Fermion decay modes dominate in the low mass region (up to about 150 GeV/c^2). In particular, the channel $H \to b\bar{b}$ gives the largest contribution, since the b quark is the heaviest fermion available. When the decay channels into vector boson pairs open up, they quickly dominate. A peak

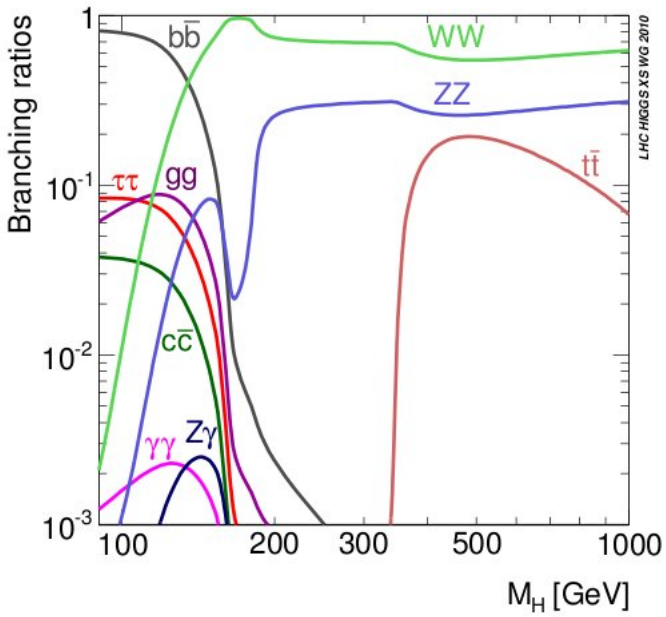

Figure 1.6: *Branching ratio of different Higgs boson decay channels as a function of the Higgs boson mass [14].*

in the $H \to W^+W^-$ channel is visible around 160 GeV/c^2, when the production of two on-shell W bosons becomes possible and the production of a real ZZ pair is still not allowed. At high masses (about 350 GeV/c^2) also $t\bar{t}$ pairs can be produced. As shown in Figure 1.6, the branching ratios change dramatically across the possible mass range, requiring different strategies for the Higgs identification depending on its mass. The most promising decay channels for the Higgs discovery do not only depend on the corresponding branching ratios, but also on the capability to experimentally separate the signal from the background. Fully hadronic events are the most copious final states from Higgs boson decays, but they cannot be easily resolved when merged in QCD background. For this reason topologies with leptons or photons are preferred, even if they have smaller branching ratios.
Such channels are illustrated in the following, depending on the Higgs mass range.

1.7.1 Low Mass Region

Though for $m_H < 135\ GeV/c^2$ the branching ratio is dominated by the Higgs decay into $b\bar{b}$, the process $gg \to H \to b\bar{b}$ is experimentally very difficult to explore because of the background constitued by the di-jet production (more than six order of magnitude higher than the signal). Some results from this channel can be obtained when the Higgs boson is produced in association

with a $t\bar{t}$ or via *Higgsstrahlung*, since in this case the event has a clearer signature, despite its low cross section.
For the Higgs search at low mass, the $H \to \gamma\gamma$, $H \to ZZ$ and $H \to WW$ channels are the most promising.
$H \to \gamma\gamma$, in spite of its lower branching ratio (around 10^{-3}), provides a clean final-state topology with an effective mass peak reconstructed with great precision. This decay channel only suffers from the $q\bar{q} \to \gamma\gamma$ and $Z \to e^+e^-$ backgrounds or jets faking photons. The expected signal rate is at least one order of magnitude smaller than the SM background rate.
Among the ZZ channels that are explored, despite the $\sigma \times BR$ is higher for the $2\ell 2q$ and $2\ell 2\nu$ final states ($\ell, \ell' = e$ or μ), the $H \to ZZ^{(*)} \to 4\ell$ channel is the only one that can play a role in the low mass region. Its clear experimental signature can be reconstructed with high efficiency and very low background contamination. The four-lepton invariant mass resolution with the LHC experiment setups is about 1%. Therefore it can contribute to the upper limit of the Higgs boson cross section exclusion as well as to a hypothetical discovery. The analysis of this channel at CMS is the context in which this thesis is placed and the expected and observed sensitivity over the full Higgs boson mass range will be presented in detail in Chapter 3. From these results it will be possible to infer the important role that this channel plays at low mass, due to the very limited background.
In $H \to W^+W^- \to 2\ell 2\nu$ channel, W^+W^- candidates, with both W bosons decaying leptonically, are selected in final states consisting of two isolated, high-p_T, oppositely-charged leptons (electrons or muons) and large missing transverse energy due to the undetected neutrinos. As the Higgs mass cannot be reconstructed due to neutrinos, this channel is much more powerful for the exclusion than for the discovery.

1.7.2 Intermediate Mass Region

For mass values between 130 $GeV/c^2 \leq m_H \leq 2m_Z$, the Higgs boson decays into $WW^{(*)}$ and $ZZ^{(*)}$ open up and their branching ratios quickly increase. Thus the best channels in this mass region are $H \to WW^{(*)} \to 2l2\nu$ and $H \to ZZ^{(*)} \to 4l$. The branching ratio of $H \to WW^{(*)}$ is higher, because of the stronger coupling of the Higgs to charged current with respect to neutral current. Moreover, this decay mode becomes particularly important in the mass region between $2m_W$ and $2m_Z$, where the Higgs boson can decay into two real W's and not yet into two real Z's, and its branching ratio is close to one. Anyway in this channel, because of the presence of the two ν's in the final state, the Higgs mass cannot be reconstructed. Such measurement can be performed instead when one W decays leptonically and the other one decays in two quarks. But in this case the final state suffers from the high hadronic background.
On the other hand, the decay $H \to ZZ^{(*)} \to 4l$, despite its lower branching

ratio, offers a very clear experimental signature and high signal to background ratio. Furthermore, it allows to reconstruct the Higgs boson mass with high precision. For these reasons this channel seems to be the best candidate for a Higgs discovery in this mass range.

1.7.3 High Mass Region

This region corresponds to mass values above the $2m_Z$ threshold, where the Higgs boson can decay into a real ZZ pair. Though the $H \to ZZ$ partial width is lower than the $H \to WW$ one, a decay into four charged leptons (muons or electrons) is surely the "golden channel" for a high mass Higgs boson discovery. The upper mass limit for detecting the Higgs boson in this decay channel is given by the reduced production rate and the increased width of the Higgs. As an example, less than 200 Higgs particles with $m_H = 700\ GeV/c^2$ will decay in the $H \to ZZ \to 4l$ channel in a year at high luminosity and the large width will increase the difficulty to observe the mass peak.
In order to increase the sensitivity to a heavy Higgs boson production, decay channels with one boson decaying into jets or neutrinos can also be studied. The decay channel $H \to WW \to l\nu_l jj$, where j denotes a jet from a quark in the W decay, has a branching ratio close to 30%, yelding a rate some 50 times higher than the four lepton channel from $H \to ZZ$ decay. The decay channels $H \to ZZ \to l\bar{l}\nu_{l'}\bar{\nu_{l'}}$, which has six times larger branching ratio than the four lepton channel, and $H \to ZZ \to 2\ell 2q$ could also be considered.

1.7.4 Higgs Total Decay Width

The total width of the Higgs boson resonance is shown in Figure 1.7 as a function of m_H. Below the $2m_W$ threshold, the Higgs width is of the order of the MeV, then it rapidly increases, but it remains lower than 1 GeV/c^2 up to $m_H \simeq 200\ GeV/c^2$. The region at low mass is therefore the most challenging one, because the Higgs boson width is dominated by the experimental resolution.
Instead, in the high mass region ($m_H > 2m_Z$), the total Higgs boson width is dominated by the W^+W^- and ZZ partial widths. Summing over the W^+W^- and ZZ channels, the Higgs width in the high mass region can be written as

$$\Gamma(H \to VV) = \frac{3}{32\pi}\frac{m_H^3}{v^2}.$$

From this equation it results that $\Gamma_H \simeq m_H$ for $m_H \simeq 1\ TeV$. When m_H becomes larger than 1 TeV, therefore, it becomes experimentally difficult to separate the Higgs resonance from the VV continuum. Actually, being the resonance width larger than its own mass, the Higgs boson cannot be properly considered as a particle any more. Furthermore, if the Higgs mass is

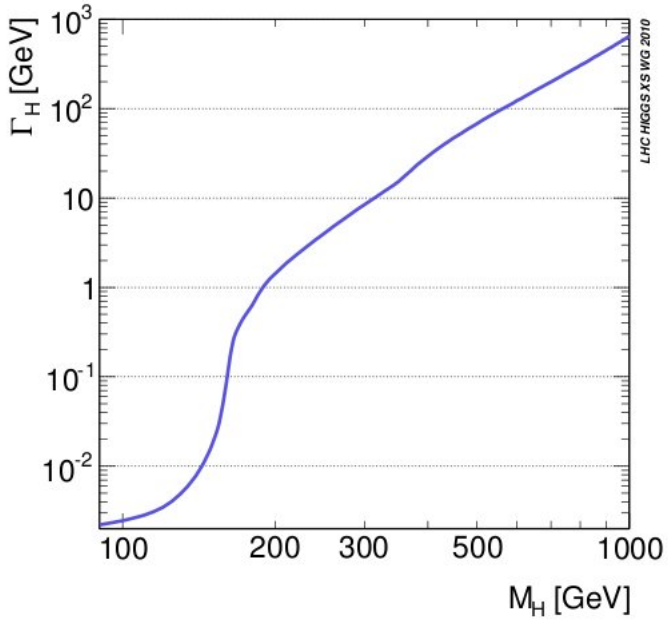

Figure 1.7: *Total decay width of the Higgs boson as a function of its mass.*

above 1 TeV, the SM predictions violate unitarity. All these considerations suggest the TeV as a limit to the Higgs mass: at the TeV scale at least, the Higgs boson must be observed, or new physics must emerge.

1.8 Recent LHC results

On 4th of July 2012, the discovery of a Higgs boson like particle has been announced by the CMS and ATLAS experiments around a mass of 125 GeV/c^2. In this section these recent results from searches for the standard model Higgs boson in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV in the CMS experiment at the LHC, using data samples corresponding to integrated luminosities of up to 5.1 fb^{-1} at 7 TeV and 5.3 fb^{-1} at 8 TeV, are briefly presented [13]. The search has been performed in five decay modes: $\gamma\gamma$, ZZ, WW, $\tau^+\tau^-$ and $b\bar{b}$. From the combination of the five decay channels, an excess of events is observed above the expected background, a local significance of 5.0 standard deviations, at a mass near 125 GeV/c^2, indicating the production of a new particle.
The CL_s is shown in Figure 1.8 as a function of the Higgs boson mass hypothesis. As one can see, the SM Higgs boson is excluded at 95% CL in the range $110 < m_H < 121.5$ GeV/c^2. In the range $121.5 < m_H < 128$ GeV/c^2 a significant excess is seen and the SM Higgs boson cannot be excluded at

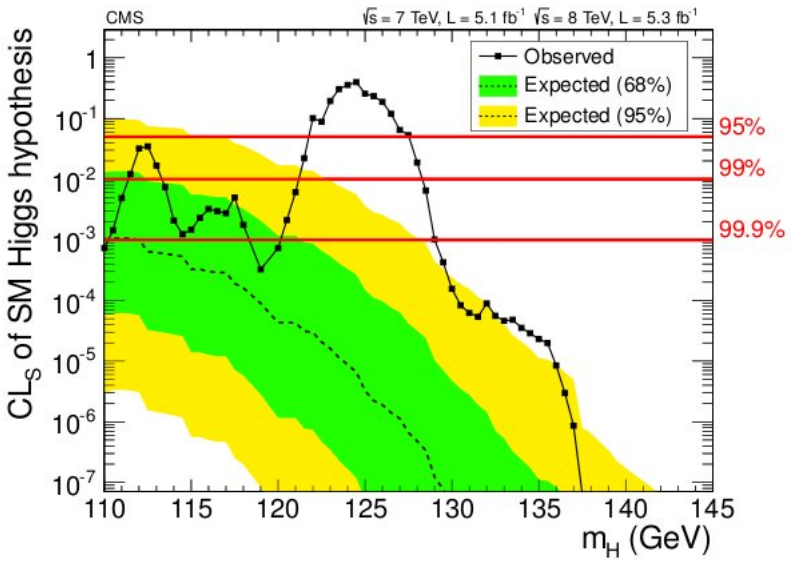

Figure 1.8: *The CL_s values for the SM Higgs boson hypothesis as a function of the Higgs boson mass in the range $110 - 145$ GeV/c^2. The background-only expectations are represented by their median (dashed line) and by the 68% and 95% CL bands.*

95% CL.

The consistency of the observed excess with the background-only hypothesis may be judged from Figure 1.9, which shows a scan of the local p-value for the 7 and 8 TeV data sets and their combination. The 7 and 8 TeV data sets exhibit an excess of 3.2σ and 3.8σ significance, respectively, for a Higgs boson mass of approximately 125 GeV/c^2. In the overall combination the significance is 5.0σ for $m_H = 125.5$ GeV/c^2. Figure 1.10 gives the local p-value for the five decay modes individually and displays the expected overall p-value. From this figure is evident that the largest contributors to the overall excess in the combination are the $\gamma\gamma$ and ZZ decay modes. In fact, they both have very good mass resolution, allowing good localization of the invariant mass of a presumed resonance responsible for the excess. A fit to these signals gives a mass of 125.3 ± 0.4 $(stat.) \pm 0.5$ $(syst.)$ GeV/c^2. The decay to two photons indicates that the new particle is a boson with spin different from one.

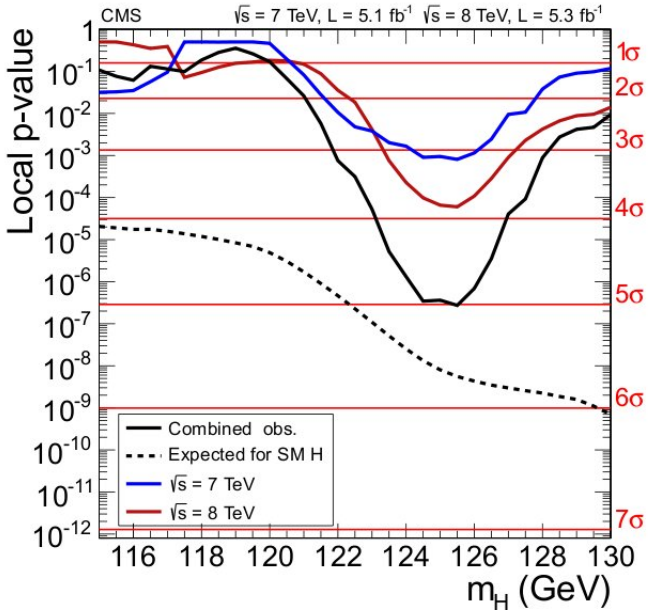

Figure 1.9: *The observed local p-value for 7 TeV and 8 TeV data, and their combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H.*

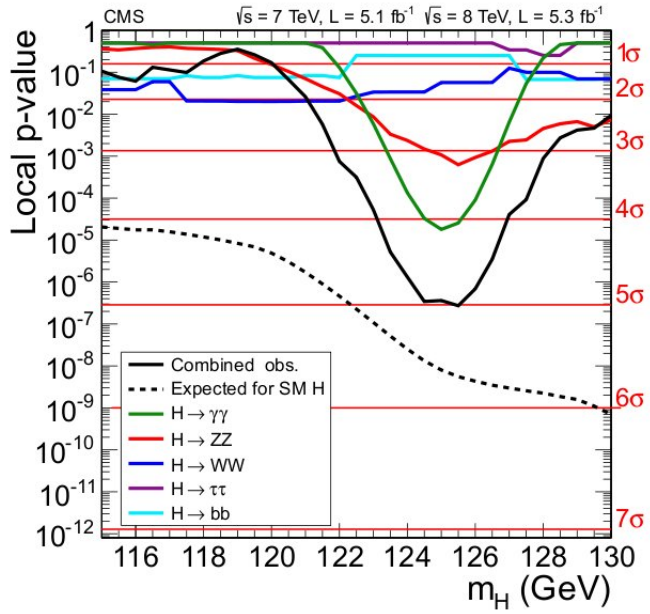

Figure 1.10: *The observed local p-value for the five decay modes and the overall combination as a function of the SM Higgs boson mass. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H.*

Chapter 2

LHC and the CMS Experiment

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [15] is the world's major particle accelerator. Proposed and realized by the European Organization for Nuclear Research (CERN), it was designed to collide protons, as well as lead ions, at an unprecedented energy and rate, in order to answer to some of the most fundamental questions of physics.
On 23 November 2009, the accelerator produced the first proton-proton collisions. After few pilot runs at energies of 450 GeV and 1.18 TeV per beam, the energy was ramped up to 3.5 TeV and, on 30 March 2010, the first collisions at a centre-of-mass energy of 7 TeV, the highest ever reached at a particle collider, were recorded by the four experiments: ALICE, ATLAS, CMS and LHCb. During these years the LHC has worked beyond expectations and the total integrated luminosity delivered in 2011 is 5.73 fb^{-1} (see Figure 2.1). In 2012 the centre-of-mass energy has been incremented up to 8 TeV, thus allowing to reach an integrated luminosity of 21.75 fb^{-1} (see Figure 2.2). In the next years, after a shutdown starting in 2013, the LHC will progressively increase its energy and instantaneous luminosity, reaching eventually the design values of 14 TeV, seven times the highest energy reached so far at Tevatron, and 10^{34} $cm^{-2}s^{-1}$, about two orders of magnitude more than the luminosity of any previous machines.

2.1.1 LHC design and performances

The LHC is placed in the already existent 26.7 km long LEP tunnel, situated at a depth of about 100 m underground at the boundary between Switzerland and France. The main design characteristics of the machine are listed in Table 2.1. Since collisions occur between particles of the same charge, the tunnel contains two adjacent and parallel beam pipes, where proton (or ion) beams travel in opposite directions and intersect in four points, where the main experimental halls are built and detectors are placed (see Figure 2.3).

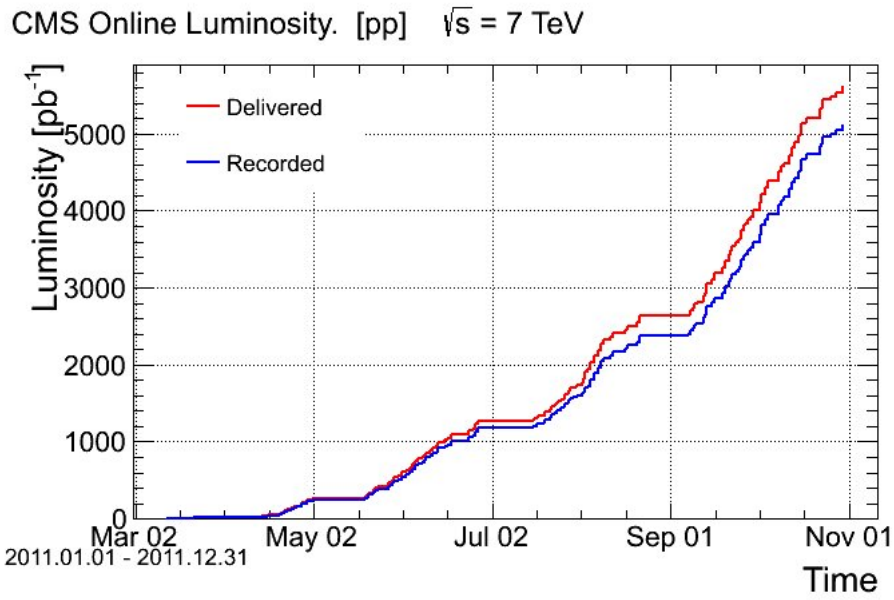

Figure 2.1: *Integrated luminosity delivered by the LHC (in red) and recorded by CMS (in blue) in 2011 proton-proton collisions.*

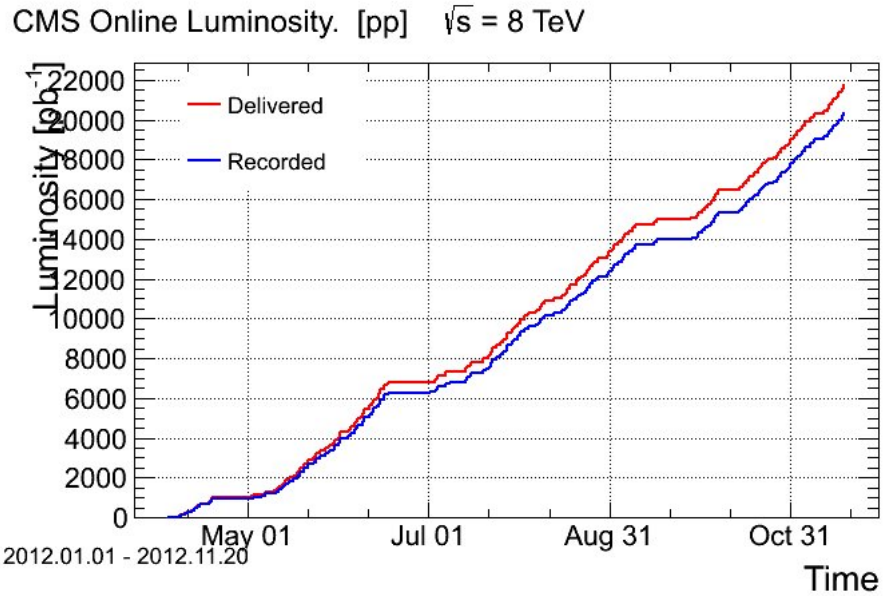

Figure 2.2: *Integrated luminosity delivered by the LHC (in red) and recorded by CMS (in blue) in 2012 proton-proton collisions.*

Table 2.1: *LHC design parameters for p-p and Pb-Pb collisions.*

Parameter	p-p	Pb-Pb
Circumference [km]	26.659	
Beam radius at interaction point [μm]	15	
Dipole peak field [T]	8.3	
Design centre-of-mass energy [TeV]	14	1148
Design Luminosity [$cm^{-2}s^{-1}$]	10^{34}	$2 \cdot 10^{27}$
Luminosity lifetime [h]	10	4.2
Number of particles per bunch	$1.1 \cdot 10^{11}$	$\sim 8 \cdot 10^{7}$
Number of bunches	2808	608
Bunch length [mm]	53	75
Time between collisions [ns]	24.95	$124.75 \cdot 10^{3}$
Bunch crossing rate [MHz]	40.08	0.008

Some 1232 dipole magnets keep the beams on their circular path, while additional 392 quadrupole magnets are used to keep the beams focused, in order to maximize the chances of interaction in the four intersection points, where the two beams cross. In total, over 1600 superconducting magnets are installed. Approximately 96 tonnes of liquid helium is needed to keep the superconducting magnets at their operational temperature of 1.9 K. The field in the magnets increases from 0.53 T to 8.3 T while the protons are accelerated from 450 GeV to 7 TeV. Two of the main experiments, ATLAS and CMS, are designed for a high luminosity regime, in order to catch the rare events of their physics programs. For this reason the beam intensity, together with the beam energy, is a crucial parameter for the LHC.
Assuming a gaussian beam shape, the luminosity at LHC can be written in terms of machine parameters as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F$$

where:

- N_b is the number of particles per bunch;
- n_b is the number of bunches per beam;
- f_{rev} is the revolution frequency;
- γ is the relativistic Lorentz factor;
- ϵ_n is the normalized transverse beam emittance;
- β^* is the optical Beta function at the collision points;

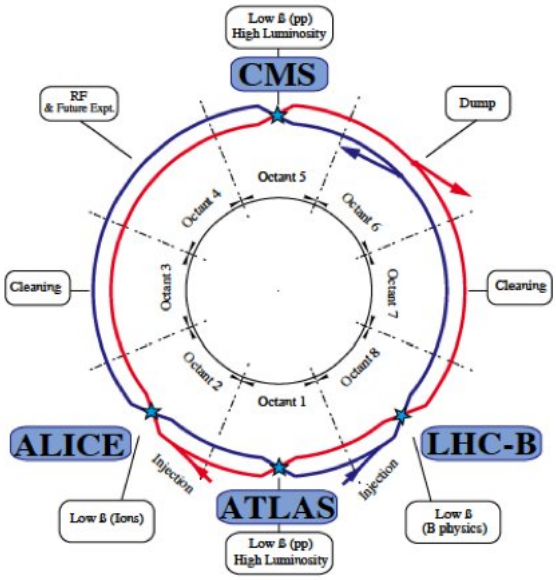

Figure 2.3: *A schematic picture of the LHC layout.*

- F is the geometric luminosity reduction factor due to the crossing angle at the IP (intersection points).

2.2 The CMS Experiment

The Compact Muon Solenoid (CMS) [16] is one of the two general-purpose detectors which operates at LHC. Its main physics goals are the search for the Higgs boson, the search for new physics beyond the SM and precision measurements of already known physics processes. For these reasons an excellent lepton reconstruction and particle identification are required.
The main characteristics of CMS detector are a compact design with a strong magnetic field, which is obtained using a 3.8 T superconducting solenoid, a robust and redundant muon system, a good electromagnetic calorimeter and a high quality central tracking system.
CMS is composed by a cylindrical barrel, with several layers coaxial to the beam axis, closed at both ends by endcap disks orthogonal to the beam direction. Its full length is 21.6 m, the diameter 14.6 m and the total weight about 12.500 t. A three dimensional view of the CMS detector is shown in Figure 2.4.

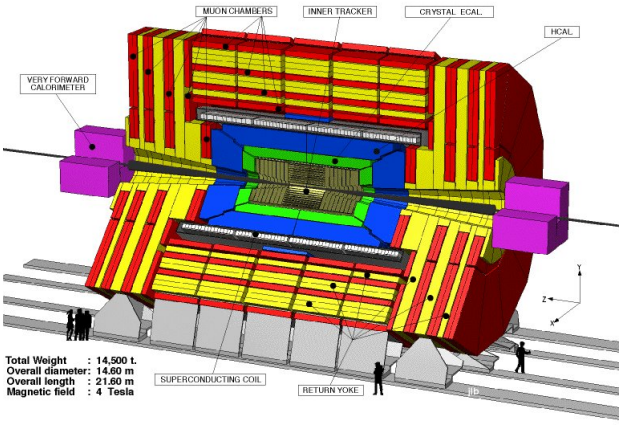

Figure 2.4: *A three dimensional view of the CMS detector.*

2.2.1 Coordinate Conventions

CMS uses a cylindrical coordinate system, whose origin is at the nominal collision point inside the detector. The x axis points to the centre of the LHC ring, the z axis is parallel to the beam and the y axis points upwards. The azimuthal angle Φ is measured in the $x-y$ plane from the x axis, while the polar angle θ is measured from the z axis. Instead of θ, the pseudorapidity η is used, which is defined as:

$$\eta = -\ln \tan \frac{\theta}{2}.$$

The transverse momentum p_T is defined from x,y components of the momentum as:

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

The transverse energy is defined as $E_T = Esin\theta$ and the missing transverse energy is denoted with E_T^{miss}.
Finally, the ΔR parameter is defined as:

$$\Delta R = \sqrt{\Delta\Phi^2 + \Delta\eta^2}.$$

2.2.2 The Tracker

The tracker (see Figure 2.5) [17] is the innermost subdetector and the closest to the interaction point. Its goal is to reconstruct high-p_T charged tracks, in the region $|\eta| < 2.5$, $r < 120\ cm$, $|z| < 270\ cm$, with high efficiency and momentum resolution, to measure their impact parameter and to reconstruct primary and secondary vertices. In order to achieve these aims the tracker

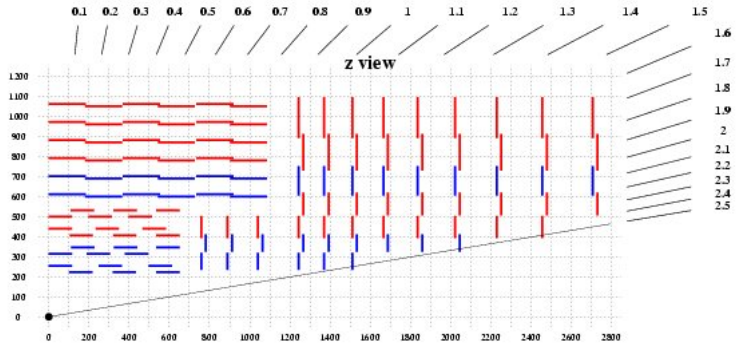

Figure 2.5: *Longitudinal view of one quarter of the silicon tracker, including the pixel detector.*

is completely based on several layers of silicon detectors. Close to the interaction point the first layers, composed by finely segmented pixel detectors, are fundamental for the measurement of the impact parameter and have to cope with a very high particle flux. The rest of the tracker is made up of single-sided and double-sided silicon strip detectors.
In order to limit the radiation damage to silicon sensors due to the high particle flux, both pixel and microstrip detectors have to be kept at a working temperature of $-10\ ^oC$.

Pixel Detector

The pixel detector (see Figure 2.6) provide high-resolution three-dimensional measurements, that are used for charged track reconstruction. Its excellent resolution also allows the measurement of track impact parameters, the identification of b- and τ-jets and the reconstruction of vertices in three dimensions. This detector consists of three barrel layers and two endcap disks for each side. The barrel layers, extending for a total length of 53 cm, are placed at mean radii of 4.4 cm, 7.3 cm and 10.2 cm. The two disks of the endcaps, placed on each side at z=34.5 cm and 46.5 cm, have the inner radius of 6 cm and the outer of 15 cm. Each silicon pixel has a size of 150 μm $\times$ 150 μm and the total area covered with pixels is about 0.92 m^2.
The pixel detector can achieve a spatial resolution of about 15 μm.

Microstrip Detector

The silicon microstrip detector is divided in two main regions. The inner part consists of 4 barrel layers and 3 disks at each side, while the outer part is made of 6 barrel layers and 9 forward disks. The full tracker consists of about 15000 microstrip detectors, covering a radial region between 20 and 120 cm and the pseudorapidity region $|\eta| < 2.5$.

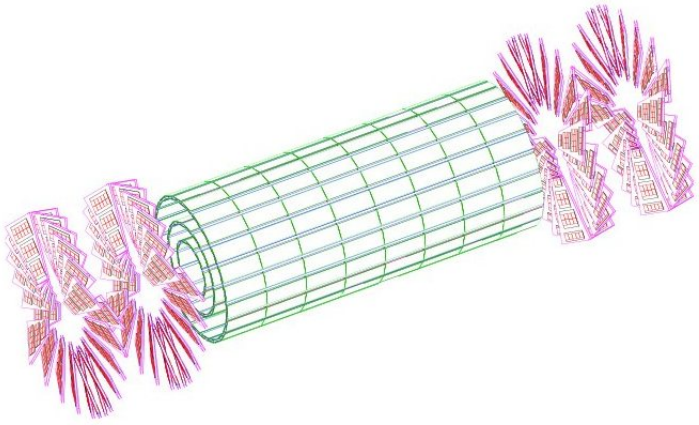

Figure 2.6: *The pixel detector: the barrel section and the two disks of the endcaps are visible.*

Finally the microstrip detector provides a spatial resolution of about 50 μm in the r-ϕ plane and about 500 μm along z.

2.2.3 The Electromagnetic Calorimeter

The goal of the Electromagnetic Calorimeter (ECAL) [18] is the accurate measurement of the position and energy of electrons and photons. For this reason it is made of about 80000 lead tungstate ($PbWO_4$) crystals, which provide an excellent energy resolution. Furthermore, lead tungstate is characterised by a short radiation length ($X_0 = 0.89\ cm$) and a small Molière radius (2.2 cm), thus having a high compactness and a very fine granularity. Moreover, the very short scintillation decay time of these crystals allows to collect about 80% of the light within 25 ns, so that they can be used at the crossing rate of 40 MHz. The length of the crystals is 23 cm in the barrel and 22 cm in the endcaps, with a front face area of $2.2 \times 2.2\ cm^2$ and $2.47 \times 2.47\ cm^2$ respectively. Scintillators light is collected by silicon avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps.
Finally the ECAL barrel covers the central rapidity region ($|\eta| < 1.48$) and the two ECAL endcaps extend the coverage up to $|\eta| = 3$.
The longitudinal view of one quarter of the ECAL is shown in Figure 2.7.

2.2.4 The Hadronic Calorimeter

The hadronic calorimeter (HCAL) [19] plays an essential role measuring the direction and energy of jets, the total transverse energy and the imbalance in the transverse energy (missing E_T). To achieve this goal a high hermeticity

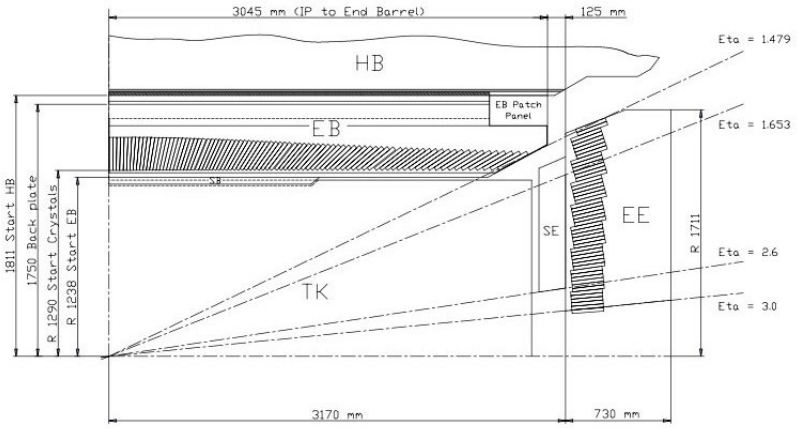

Figure 2.7: *Longitudinal view of one quarter of the CMS ECAL.*

is required. In particular, in addition to the barrel and endcap parts installed inside the magnet, there is a very forward calorimeter (HO) which is placed outside the magnet return yokes, covering the pseudorapidity region $|\eta| < 5.3$. This calorimeter allows the identification of forward jets, which is very important for the rejection of many backgrounds.
The barrel (HB) and endcap (HE) calorimeters, covering the region $|\eta| < 3$, are sampling calorimeters made of plastic scintillators interleaved with brass absorber plates. The read-out system is constituted by wavelength-shifting fibres. The first layer is read out separately, while all others are read out together in towers of $\Delta\eta \times \Delta\Phi = 0.087 \times 0.087$ rad.
The energy resolution (expressed in GeV) is:

- $\frac{\sigma_E}{E} \sim 65\%\sqrt{E} \oplus 5\%$ in the barrel,
- $\frac{\sigma_E}{E} \sim 85\%\sqrt{E} \oplus 5\%$ in the endcaps,
- $\frac{\sigma_E}{E} \sim 100\%\sqrt{E} \oplus 5\%$ in the very forward calorimeter.

2.2.5 The Muon System

The aim of the muon spectrometer [20] is to identify muons and measure, in combination with the inner tracker, their transverse momenta accurately. As a matter of fact, since high-p_T muons provide a clean signature for many processes, the muon system plays an important role in the trigger. The muon spectrometer, placed outside the magnet, is embedded in the iron return yoke, so that the 1.8 T average magnetic field bends the tracks and

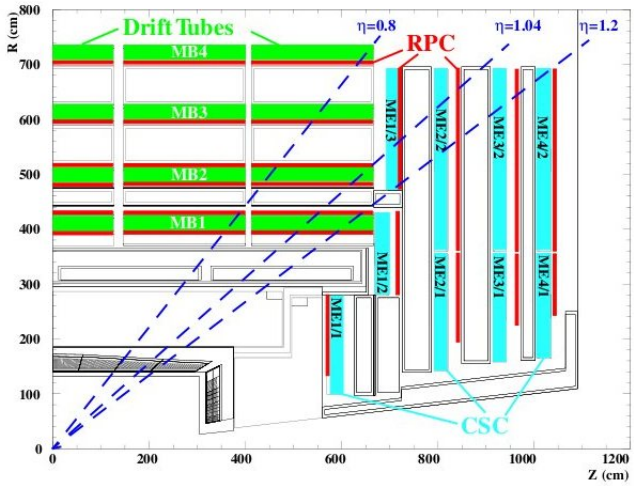

Figure 2.8: *Longitudinal view of one quarter of the muon system.*

allows muon p_T measurements.

The muon system consists of 3 types of gaseous particle detectors (see Figure 2.8):

- Drift Tube (DT) Chambers in the barrel, covering the region ($|\eta| < 1.2$),
- Cathode Strip Chambers (CSC) in the endcaps, covering the region ($0.9 < |\eta| < 2.4$),
- Resistive Plate Chambers (RPC) in both the barrel and the endcaps, covering the region ($|\eta| < 1.6$).

These different technologies are used because of the different particle rates and occupancies, both higher in the endcaps, and the intensity of the residual magnetic field, which is lower in the barrel.

Drift Tubes Chambers

Since in the barrel region the expected occupancy is low ($< 10\ Hz/cm^2$) drift tubes were chosen. The DT segmentation follows that of the iron plates of the yoke, which consists of 5 wheels along the z-axis, each one divided in 12 sectors. Chambers are arranged in 4 *stations* named MB1,...,MB4 as shown in Figure 2.8. Each station consists of 12 chambers, except for MB4 which has 14 chambers.
The basic detector element is a drift tube cell, whose section is shown in

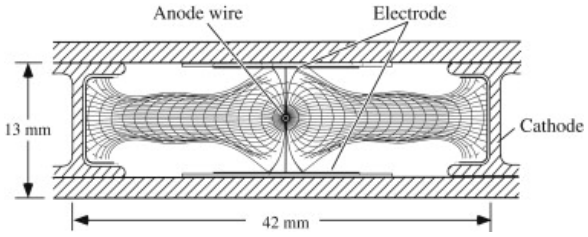

Figure 2.9: *Section of a drift tube cell.*

Figure 2.9. A layer of cells is made of parallel aluminium plates, with cells obtained with perpendicular "I"-shaped aluminium cathodes. The anodes are 50 μm diameter steel wires placed between the cathodes. The distance of the track from the wire is measured by the drift time of electrons; to improve the distance-time linearity, additional field shaping is obtained with two positively-biased insulated strips, glued on the planes in correspondence to the wire.
The internal volume is filled with a gas mixture of Ar (85%) and CO_2 (15%) at atmospheric pressure, which provides good quenching properties and a saturated drift velocity of about 5.6 $cm/\mu s$. A single cell has an efficiency close to 100% and a resolution of about 180 μm.

Cathode Strip Chambers

Because of the large occupancy of the endcap regions, from few Hz/cm^2 to more than 100 Hz/cm^2, and the intense and inhomogeneous magnetic field, cathode strip chambers were adopted in this region.
CSC chambers are arranged in four disks (*stations*) placed between the iron disks of the yoke and named ME1,...,ME4 (see Figure 2.8). The CSCs, multiwire proportional chambers with good spatial and time resolution, are composed of one cathode plane segmented in strips running orthogonal to the wires. An avalanche developed on a wire induces a distributed charge on the cathode plane. The orthogonal orientation of the cathode strips with respect to the wires allows the determination of two coordinates from a single detector plane, as shown in Figure 2.10. Each chamber is formed by 6 trapezoidal layers, with strips in the radial direction for a precise measurement of the azimuthal coordinate Φ. The wires resolution is of the order of about 0.5 cm, while for the strips is of about 50 μm.

Resistive Plate Chambers

Resistive plate chambers are installed both in the barrel and in the endcap regions, in order to add robustness and redundancy to the muon trigger. They have a limited spatial resolution, but an excellent time resolution, of the order of few nanoseconds.

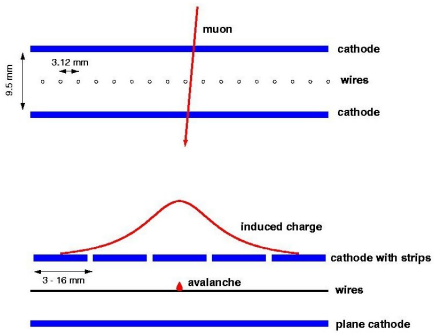

Figure 2.10: *Orthogonal sections of a cathode strip chamber.*

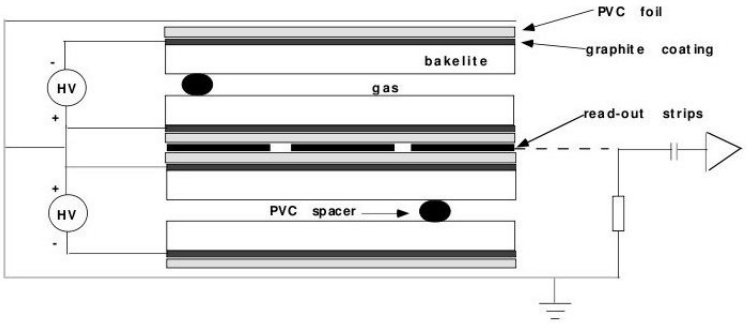

Figure 2.11: *Section of a double gap resistive plate chamber.*

The RPCs used in CMS are composed of 4 bakelite planes forming two coupled gaps 2 mm thick, as shown in Figure 2.11. The gaps are filled with a mixture of 90% $C_2H_2F_4$ (freon) and 5% $i - C_4H_{10}$ (isobutane).
They operate in avalanche mode rather than in the more common streamer mode. This is obtained with a moderate electric field across the gap which allows to sustain higher rate. However the gas multiplication is reduced, and improved electronic amplification is needed.

2.2.6 The Trigger and Data Acquisition System

The huge amount of data delivered by LHC cannot be sustained by any storage system available nowadays. The collisions rate is 40 MHz, but only a rate of about 100 Hz can be stored for offline analysis. For this reason a huge reduction factor is performed by the trigger and the data acquisition systems.

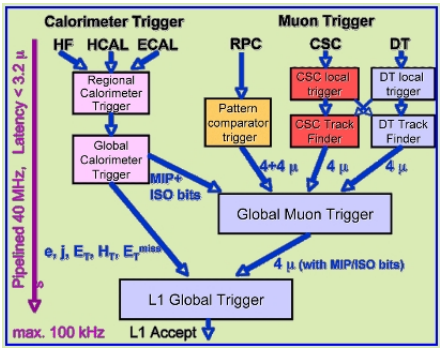

Figure 2.12: *Scheme of the L1 trigger.*

The Trigger System

The event rate at LHC is principally composed of protons interactions with particles of low transverse momentum. A good trigger should provide a large rejection of the less interesting events and maintain at the same time a high efficiency on those which are potentially interesting.
At CMS this goal is achieved in two steps: a Level 1 Trigger (L1) and a High Level Trigger (HLT).

The Level 1 Trigger

The Level 1 Trigger [21], which is at the hardware level, reduces the rate of selected events down to 100kHz for the high luminosity runs. Waiting for the trigger decision, data are stored in pipelines of processing elements. Within a latency period of 3.2 μs, if the L1 accepts the event, the data are moved to be processed by the High Level Trigger. Because of the high bunch crossing rate, the full readout of the detector is not possible: only the calorimetric and muons information are employed. The Calorimeter Trigger identifies, from the shape of the deposited energy, the best four candidates of electrons and photons, central jets, forward jets and so on. The information about these objects is sent to the Global Trigger, together with the measured E_T^{miss}. Since the Muon Trigger is performed separately for each muon detector, the information is then merged and the best four muon candidates are transferred to the Global Trigger. The Global Trigger finally decides to reject or to accept the event for further evaluation by the HLT. This decision is based on algorithm calculations and on the readiness of the sub-detectors and the DAQ. The L1 Trigger electronics is placed partly on the sub-detectors, partly in the underground control room located at a distance of about 20m from the CMS detector site.
In Figure 2.12 is shown a scheme of the L1 Trigger.

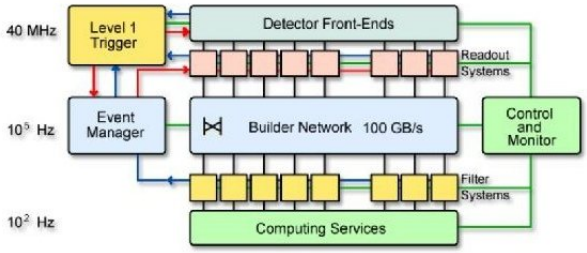

Figure 2.13: *Scheme of the DAQ system.*

The High Level Trigger (HLT)

The HLT [22] is a software system which reduces the output rate down to around 100 Hz. In this trigger only objects in the useful regions are reconstructed, while uninteresting events are rejected as soon as possible. The HLT can perform complex calculations because it has access to the high-resolution data in pipelined memories in the front-end electronics as well as the information from the silicon tracker. Joining the L1 and the HLT schema is therefore possible to distinguish three "virtual trigger" levels: at the first level only the muon system and the calorimeters information is used, in the second level the information of the tracker pixels is added and in the third level the full event information is finally available.

The Data Acquisition (DAQ)

The goal of the CMS Data Acquisition System (see Figure 2.13) is to transport the data from about 650 data sources at the detector side, each of which provides event fragments of about 2kB, to the filter units for processing of complete events. The central DAQ runs online software on about 3000 PC used for buffering and processing event data. The detector is read out, through a builder network with a bandwidth of 100GB/s, by the so called Front-End Drivers (FED), which are located in the underground counting room at a distance of 70m from the detector. Complete events are fed to the event filter systems at a maximum rate of 100kHz. This large rate is due to the choice of building the full event already after the first level trigger instead of building partial events as in traditional multilevel trigger systems. This requires the read-out, assembly and forwarding of the full event data at the nominal level one trigger rate. The total rate of data produced by the online trigger system, which need to be stored for further processing and analysis, is about 230 MB/s.

Chapter 3

Search for the SM Higgs boson in the Decay Channel $H \to ZZ \to 4\ell$

As described in Chapter 1, the Standard Model (SM) of electroweak interactions relies on the existence of the Higgs boson, a scalar particle of mass m_H associated with the field responsible for the spontaneous electroweak symmetry breaking.
The Higgs boson production followed by the decay $H \to ZZ$ is expected to be one of the main discovery channels at the LHC for a wide range of m_H values.
In this chapter the $H \to ZZ \to 4\ell$ analysis, designed to search for a Higgs boson in the mass range $110 < m_H < 1000\ GeV/c^2$, is presented. The data collected at $\sqrt{s} = 7\ TeV$ by CMS in 2011, corresponding to an integrated luminosity of 5.1 fb^{-1}, are combined with new data collected in 2012 at $\sqrt{s} = 8\ TeV$, corresponding to an additional 12.2 fb^{-1} [23, 24].
This search essentially relies on the reconstruction, identification, and isolation of leptons, combined with the use of a discriminant exploiting the production and decay kinematics expected for the signal events.
This thesis work is included in $H \to ZZ \to 4\ell$ analysis, and in particular the two main irreducible backgrounds, $gg \to ZZ$ and $q\bar{q} \to ZZ$, are treated (see Chapter 4). In fact, a new technique to calculate their ratio is developed and presented in Chapter 5.

3.1 General Analysis Strategy

The analysis $H \to ZZ \to 4\ell$ uses data collected with the di-electron and di-muon triggers and selects events with well-reconstructed four leptons: $e^+e^-e^+e^-$, $\mu^+\mu^-\mu^+\mu^-$, $e^+e^-\mu^+\mu^-$.
The main background is constituted by an irreducible contribution from

ZZ^* production via $q\bar{q}$ and gluon-gluon induced processes. Furthermore, the sample of events with reconstructed 4ℓ receives contributions from reducible background $Zb\bar{b}$ and $t\bar{t} \to W^+bW^-\bar{b}$, with W undergoing leptonic decays, where the final states contain two isolated leptons and two b jets possibly giving rise to secondary leptons. Instrumental background is also present, such as Z+jets or WZ + jet(s) where jets are misidentified as leptons.
The analysis aim is to attain the highest possible lepton reconstruction, identification and isolation efficiencies, that are compatible with a quasi-negligible reducible and instrumental background.
An optimal working point can be obtained where the contributions from the reducible and instrumental backgrounds are quasi-eliminated. This is achieved by applying cuts on the maximum allowed energy flow in the isolation cones around leptons and on the maximum impact parameter of lepton tracks with respect to the primary interaction vertex. The statistical analysis of selected events is based on their four-lepton mass ($m_{4\ell}$) distribution. Thus, the Higgs boson signal is expected to manifest itself as a resonance over the continuum $m_{4\ell}$ distribution originated from $pp \to ZZ \to 4\ell$. The peak width for a SM Higgs boson with low mass ($m_H < 250\ GeV/c^2$) is expected to be dominated by the detector resolution, while, for higher masses, Higgs boson's intrinsic width overtakes the detector resolution.
The irreducible backgrounds prediction relies fully on MC. All instrumental uncertainties associated with selecting four prompt leptons (trigger, reconstruction, isolation and impact parameter cuts) are instead derived directly from data. The contribution of reducible backgrounds is evaluated using data driven techniques.

3.2 Datasets

3.2.1 Experimental Data

The data sample used in this analysis was recorded by the CMS experiment during 2011 for the run range from 160431 to 180252 and during 2012 for the run range from 190645 to 203002. The CMS standard selection of runs and luminosity sections is applied, which requires high quality data with a good functioning of the different sub-detectors.
Of the total integrated luminosity, a sample corresponding to $\mathcal{L} = 5.1\ fb^{-1}$ in 2011 at $7\ TeV$ and $\mathcal{L} = 12.2\ fb^{-1}$ in 2012 at $8\ TeV$ is used. The absolute pp luminosity is known with a precision of 2.2% in 2011 and 4.4% in 2012.
Collision events are selected by the trigger system that requires the presence of a pair of electrons or a pair of muons. A cross-trigger requiring an electron and a muon is also used for the 2012 data. The requirements on the transverse momenta for the first and second lepton are 17 and 8 GeV/c respectively. The trigger efficiency within the acceptance of this analysis is

greater than 99% (96%, 98%) in the 4μ ($4e$, $2e2\mu$) channels, for a Higgs boson signal with $m_H > 120\ GeV/c^2$.

3.2.2 Simulated Samples

SM Higgs boson signal samples, as well as samples for a large variety of electroweak and QCD-induced SM background processes, have been obtained using detailed Monte Carlo (MC) simulations. These Monte Carlo samples are used for the optimization of the event selection, the evaluation of acceptance corrections and systematics, and for the background evaluation procedure where measurements in a "background control" region are extrapolated to the "signal" region.

The Higgs boson signal from gluon-gluon fusion ($gg \to H$) is generated with POWHEG [25] at next-to-leading order (NLO) and a dedicated generator for angular correlations [32]. Additional samples of WH, ZH, and $t\bar{t}H$ events are instead generated with PYTHIA [26]. Furthermore, events at generator level are reweighted according to the total cross section $\sigma(pp \to H)$, which contains contributions from gluon fusion up to next-to-next-to-leading order (NNLO) and next-to-next-to-leading log and from the weak-boson fusion contribution computed at NNLO. The total cross section is scaled by the branching fraction $B(H \to 4\ell)$ calculated with PROPHECY4F, which includes NLO QCD and electroweak corrections and Z-Z interference in the $4e$ and 4μ channels [27, 40]. In the matter of SM backgrounds, the contribution from ZZ production via $q\bar{q}$ is generated at NLO with POWHEG, while other diboson processes (WW,WZ) are generated with MADGRAPH [28] with cross sections rescaled to NLO predictions. The $gg \to ZZ$ contribution is generated with GG2ZZ [29]. The $Zb\bar{b}$, $Zc\bar{c}$, $Z\gamma$, and $Z+$ light jets samples are generated with MADGRAPH, as contributions to inclusive Z production, with cross sections rescaled to NNLO prediction for inclusive Z production. The $t\bar{t}$ events are generated at NLO with POWHEG. All generated samples are interfaced with PYTHIA, which takes into account the internal initial-state and final-state radiation effects which can lead to the presence of additional hard photons in an event. All events are processed through a detailed simulation of the CMS detector based on GEANT4 [30] and are reconstructed with the same algorithms that are used for data.

3.3 Physics Objects: electrons, muons and photons

The reconstruction of the SM Higgs boson in the decay chain $H \to ZZ^* \to 4l$ imposes high-performance lepton reconstruction, identification and isolation as well as excellent lepton energy-momentum measurements. The identification of isolated leptons emerging from the event primary vertex cuts

drastically QCD-induced sources of mis-identified leptons or non-prompt leptons coming from hadron decays ("fake"). Furthermore, the precise energy-momentum measurements translate in a precise Higgs boson mass measurement $m_{4\ell}$, the most discriminating observable for the Higgs boson search. With four leptons in the final state, and in view of the modest fraction of the total production cross-section observable in the 4ℓ channels, a very high lepton reconstruction efficiency is fundamental. For Higgs bosons with masses $m_H < 2\ m_Z$, one lepton pair at least couples to an off-shell Z^* boson. The softest lepton in this pair typically has $p_T^\ell < 10\ GeV/c$ for masses $m_H < 140\ GeV/c^2$. Preserving the highest possible reconstruction efficiency while ensuring sufficient discrimination against hadronic jets faking leptons is especially challenging for the reconstruction of leptons at very low p_T^ℓ. In the low p_T range, a full combination of information provided by the tracker and electromagnetic calorimeter (for electrons) or by the tracker and muon spectrometer (for muons) becomes essential for the leptons reconstruction, identification and isolation.

3.3.1 Electron Reconstruction and Identification

Reconstruction

The electron reconstruction combines ECAL and tracker information. Electron candidates are reconstructed from clusters of energy deposits in the ECAL, which are then matched to hits in the silicon tracker. The standard CMS electron reconstruction algorithm is used for this analysis. The energy deposited in the electromagnetic calorimeter (ECAL) is measured in clusters of clusters (superclusters) which collect bremsstrahlung photons emitted in the tracker volume. Trajectories in the tracker volume are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian Sum Filter. Electron candidates are preselected using loose cuts on track-cluster matching observables so to preserve the highest possible efficiency while removing part of the QCD background. The four-momenta for an electron is obtained by taking angles from the associated Gaussian Sum Filter (GSF) track, and the energy from a combination of tracker and ECAL information. The information from the track is measured at the distance-of-closest approach to the beam spot position in the transverse plane. For this physics analysis, the electron candidates are required to have transverse momentum $p_T^e > 7\ GeV/c$ and a reconstructed $|\eta^e| < 2.5$. The reconstruction efficiency for isolated electrons is expected to be above $\approx 90\%$ over the full ECAL acceptance, apart from some narrow "crack" regions. Integrated over the acceptance, the reconstruction efficiency for basic electron objects steeply rises to reach $\approx 90\%$ at $p_T^e = 10\ GeV/c$, and then more slowly to reach a plateau of $\approx 95\%$ for $p_T^e = 30\ GeV/c$.

Identification

In the $H \to ZZ \to 4l$ analysis the electron candidates are required to have p_T larger than 7 GeV/c and a reconstructed $|\eta| < 2.5$. In addition, the purity of the sample of electron candidates is enhanced for the analysis by applying identification requirements on top of the basic collection of reconstructed electron objects. Electrons are identified among the reconstruction candidates and then used, together with the other PF particles, to obtain a consistent description of the event. Their identification relies on a multivariate technique that combines observables sensitive to the amount of bremsstrahlung along the electron trajectory, the geometrical and momentum matching between the electron trajectory and associated clusters, as well as shower-shape observables. The multivariate identification is trained using a Higgs boson Monte Carlo (MC) sample for the signal and a W+1 jet data sample for background, and the working point is optimized using a Z+1 jet data sample.

3.3.2 Muon Reconstruction and Identification

Reconstruction

The muon reconstruction combines the information from both the silicon tracker and the muon spectrometer. Muon tracks are first reconstructed independently in the inner tracker (*tracker track*) and in the muon system (*standalone-muon track*). Based on these objects, two reconstruction approaches are used:

- *Global Muon reconstruction (outside-in).* For each standalone-muon track, a matching tracker track is found by comparing parameters of the two tracks propagated onto a common surface, and a *global-muon track* is fitted combining hits from the tracker track and standalone-muon track, using the Kalman-filter technique [33]. At large transverse momenta, $p_T \gtrsim 200$ GeV/c, the global-muon fit can improve the momentum resolution compared to the tracker-only fit.

- *Tracker Muon reconstruction (inside-out).* All tracker tracks with $p_T > 0.5$ GeV/c and the total momentum $p > 2.5$ GeV/c are considered as possible muon candidates and are extrapolated to the muon system taking into account the magnetic field, the average expected energy losses, and multiple scattering in the detector material. If at least one muon segment (i.e., a short track stub made of DT or CSC hits) matches the extrapolated track, the corresponding tracker track qualifies as a Tracker Muon.

Tracker Muon reconstruction is more efficient than the Global Muon reconstruction at low momenta, $p < 5$ GeV/c, because it requires only a single

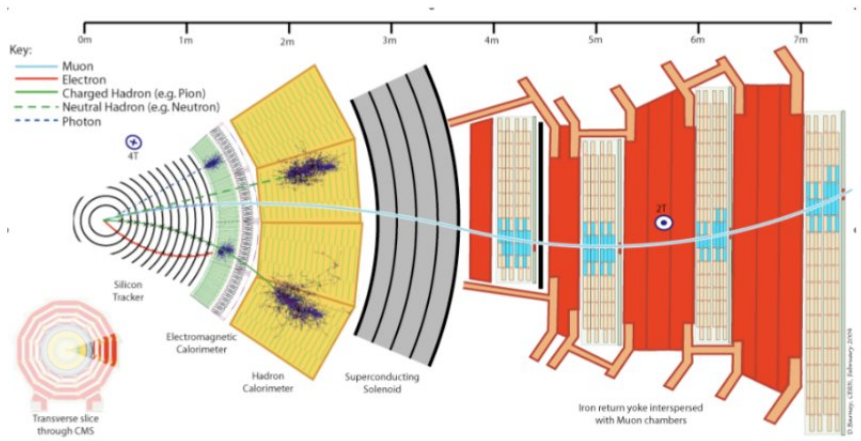

Figure 3.1: *Schematic representation of a reconstructed muon track crossing the CMS detector.*

muon segment in the muon system, whereas Global Muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station and typically requires segments in at least two muon stations.
Thanks to the high tracker-track efficiency and a very high efficiency of reconstructing segments in the muon system, about 99% of muons produced in pp collisions and having sufficiently high momentum are reconstructed either as a Global Muon or a Tracker Muon, and very often as both. Candidates found both by the Global Muon and the Tracker Muon approaches that share the same tracker track are merged into a single candidate. Muons reconstructed only as standalone-muon tracks have worse momentum resolution and less favourable collision muon to cosmic-ray muon ratio than the Global and Tracker Muons and are usually not used in physics analyses.

Identification

A given physics analysis can achieve the desired balance between identification efficiency and purity by applying a selection based on various muon identification variables. For this analysis it has been chosen the *Particle Flow Muon Selection*: the CMS particle-flow event reconstruction combines the information from all subdetectors to identify and reconstruct individually particles produced in the collision. The resulting list of particles is then used to construct higher-level particle-based objects and quantities, such as jets and missing transverse energy. To identify Particle-Flow Muons, a selection is performed on all the muon candidates reconstructed with the standard algorithms described above (Tracker and Global Muons). This selection has been optimized to identify muons in jets with high efficiency,

keeping the misidentification rate from charged hadrons low. Therefore, the Particle-Flow Muon selection has been designed to retain non-isolated muons, including the muons from hadron decays in flight, usually considered as a background in typical muon analyses. This is accomplished by applying selection criteria, which differ in strictness depending on whether the muon candidate is isolated or not, and whether its momentum is compatible with the energy deposition in the calorimeters assigned to the candidate by the particle-flow event reconstruction. The result is that the identification criteria on prompt isolated muons can be relaxed, without having an increase in fake reconstruction probability.

3.3.3 Lepton Impact Parameter and Isolation

The electron or muon pairs from Z decays should originate from the primary vertex. This is ensured by requiring that the significance of the impact parameter to the event vertex, called SIP_{3D}, satisfies $|SIP_{3D} = \frac{IP}{\sigma_{IP}}| < 4$ for each lepton. The IP is the lepton impact parameter in three dimensions at the point of closest approach with respect to the primary interaction vertex, and σ_{IP} the associated uncertainty.

The isolation of individual e or μ leptons is measured relative to their transverse momentum p_T^ℓ, by summing over charged and neutral particles in a cone $\Delta R = \sqrt{(\eta^\ell - \eta^i)^2 + (\Phi^\ell - \Phi^i)^2} < 0.4$ around the lepton direction at the interaction vertex:

$$R_{Iso}^\ell = \Big(\sum p_T^{charged} + MAX \left[0, \sum E_T^{neutral} + \sum E_T^\gamma - \rho \times A_{eff} \right] \Big) / p_T^\ell.$$

The $\sum p_T^{charged}$ is the scalar sum of the transverse momenta of charged hadrons originating from the primary vertex. The primary vertex is chosen as the vertex with the highest sum of p_T^2 of its constituent tracks. The $\sum E_T^{neutral}$ and $\sum E_T^\gamma$ are the scalar sums of the transverse energies for neutral hadrons and photons, respectively. The term $\rho \times A_{eff}$ subtracts an estimate obtained using a "jet area" technique [34] of the transverse energy from neutrals in the isolation cone coming from pileup of additional pp collisions. The transverse energy density ρ is calculated in each event as the median of the neutral-energy distribution around "jets" (any PF jet in the event having $p_T^{jet} > 3\ GeV/c$) with mean effective $\eta - \Phi$ area A_{eff}. A small residual dependence on the number of pileup collisions is absorbed as a correction factor on A_{eff}.
The electrons or muons are considered isolated in the $H \to 4\ell$ analysis if $R_{Iso}^\ell < 0.4$.
Isolation variables are among the most pile-up sensitive variables in this analysis. Pile-up causes the mean energy deposited in the detector to increase, leading to the rise of the mean isolation values. Thus, the efficiency of a cut

on isolation variables strongly depends on pile-up conditions. In order to have a pile-up robust analysis, the isolation variable has to be corrected.

3.3.4 Leptons Measurements: T&P methodology

The efficiencies for reconstruction, identification and trigger for electrons and muons are measured with data based on a selection of events of inclusive single Z production. The well-known tag-and-probe technique [35, 36] combines the requirements of a vertex from a pair of basic objects (e.g. tracks for muons, or clusters of calorimetry cells for electrons) with a tight lepton selection applied on one leg (the "tag"), so to ensure sufficient purity. The other leg (the "probe") is used to measure the efficiency of a given reconstruction algorithm or identification criterion. Since no selection is applied on the probe, its efficiency is unbiased with respect to this analysis selection criteria (Figure 3.2 and 3.3). The efficiency is defined as the ratio of the number of passing probes to the total number of probes before the cut.
It is important to perform such a measurement with the exact electron and muon objects as used in this analysis. By using appropriate definitions for probes, the overall efficiency per lepton can be factorized in a series of terms, that can be measured independently:

$$\epsilon = \epsilon_{RECO|trackorclustering} \times \epsilon_{ID|RECO} \times \epsilon_{ISO|ID} \times \epsilon_{SIP|ISO}$$

where each term represents the efficiency for the probe to pass a given selection or reconstruction step, given that it passes the criteria for the previous one. The efficiencies for selecting electrons in the ECAL barrel (endcaps) varies from about 71% (65%) for $7 < p_T^e < 10$ GeV/c to 82% (73%) at $p_T^e \simeq 10$ GeV/c, and reaches 90% (89%) for $p_T^e \simeq 20$ GeV/c. It drops to about 85% in the transition region, $1.44 < |\eta| < 1.57$, between the ECAL barrel and endcaps. The muons are reconstructed and identified with efficiencies above $\sim 98\%$ in the full $|\eta^\mu| < 2.4$ range.
By applying the method to both data and simulation is possible to derive data to simulation scale factors. These scale factors are later used to either correct the signal efficiency in the simulation (using in this case their uncertainty as systematic error) or to provide systematic uncertainties. Statistical uncertainties on the efficiencies are estimated using Clopper-Pearson confidence intervals [37]. In the computation of the final systematics on the signal efficiency the systematics on the tag-and-probe method are also considered; they are evaluated varying the signal and background modeling.

3.3.5 Photon Observables and FSR recovery

A Z decay into a lepton pair can be accompanied by final state radiation (FSR), $Z \to \ell^+\ell^-\gamma$. If the photon transverse momentum, p_T^γ, is required to exceed 2 GeV/c, about 8% (15%) of the decays into muons (electrons)

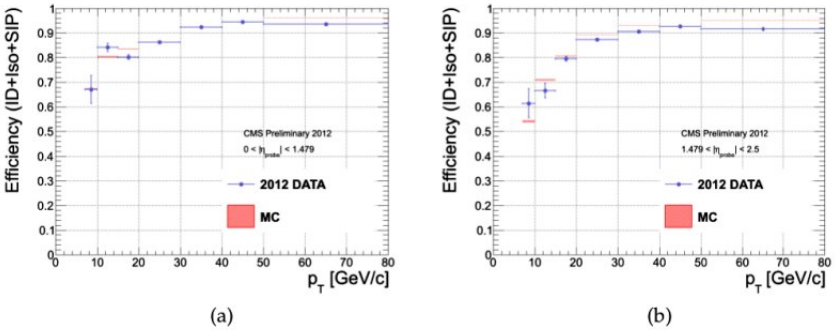

Figure 3.2: *Electron identification+isolation+$|SIP_{3D}|$ efficiencies computed with the tag-and-probe method as a function of the probe p_T in two different η bins: (a) $|\eta| < 1.442$, (b) $1.442 < |\eta| < 2.5$. Results are for 8 TeV data [23].*

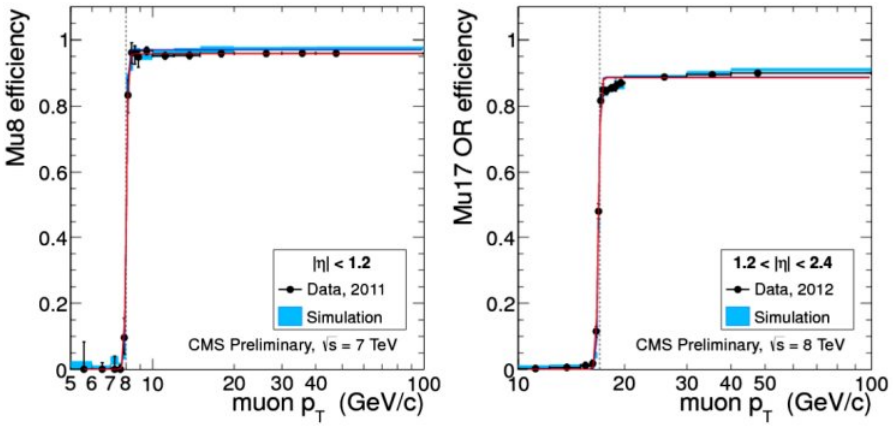

Figure 3.3: *Examples of muon HLT Trigger efficiency, as a function of the muon p_T. Muons are asked to pass ID, Isolation and SIP_{3D} requirements of the analysis [23].*

are affected. As the photon emission is most often collinear with one of the leptons, electron measured energies automatically include the energy of a large fraction of the emitted photons in the associated electromagnetic super-cluster. On the other hand, muon measured momenta do not include the emitted photons. Final state radiation is therefore expected to degrade the Z mass resolution when measured with the sole muon pairs, and in turn degrade the Higgs boson mass resolution when measured with the four leptons momenta, especially in the 4μ and in the $2e2\mu$ final states and, to a lesser extent, in the $4e$ final state. It is also expected to reduce the efficiency of the lepton isolation cut when the emitted photon is in the lepton isolation cone.
Both an excellent Higgs boson mass resolution and a large selection efficiency are essential ingredients in view of the small production cross section in the 4ℓ channels, in particular to discriminate the Higgs boson signal from the background continuum. The purpose of this analysis is to recover the FSR photons with large efficiency and purity, to remove the energy of the recovered photons from the lepton isolation cones, and to measure the mass of the Higgs boson candidate from the momenta of the leptons and the recovered photons.
In practice, photons reconstructed within $|\eta^\gamma| < 2.4$ are possible FSR candidates. To be accepted as FSR, a reconstructed photon must either have $p_T^\gamma > 2\ GeV/c$ and be found within a conical distance $\Delta R < 0.07$ from a selected lepton candidate, or have $p_T^\gamma > 4\ GeV/c$ and be found isolated within the conical distance of $0.07 < \Delta R < 0.5$ around a selected lepton candidate. The photon isolation observable R_{Iso}^γ is obtained by summing over the transverse momenta of charged hadrons, other photons and neutral hadrons identified by the PF reconstruction in a cone of size $\Delta R = 0.3$ around the candidate photon direction, correcting for pileup, and dividing by the photon transverse momentum, p_T^γ. Isolated photons must satisfy $R_{Iso}^\gamma < 1$.
The performance of the FSR selection algorithm has been measured using MC simulation samples, and the rate was verified with single-Z data events. The photons within the acceptance for the FSR selection are measured with an efficiency of $\simeq 50\%$ and with a mean purity of 80%. FSR photons are selected in 5% of single-Z events with muon pairs, and 0.5% of single-Z events with electron pairs. A gain of $\simeq 3\%$ (2%,1%) in efficiency is expected for the selection of $H \to 4\mu$ ($2e2\mu$, $4e$) events in this analysis.

3.4 Event Selection

As previously described, the selection uses well identified and isolated primary leptons. The lepton isolation requirements suppress the $Z + jet$, $Zb\bar{b}$ and $t\bar{t}$ backgrounds. The requirement on the significance of the impact parameter to the event vertex $|SIP_{3D}| < 4$ further suppresses the $Zb\bar{b}$ and

$t\bar{t}$ backgrounds. When building the Z candidates, only the FSR photons associated with the closest lepton and which make the "dressed" lepton-pair mass closer to the nominal Z mass are kept, with a maximum mass $m_{\ell\ell\gamma} < 100\ GeV/c^2$. The analysis requires a Z candidate formed with a pair of leptons of the same flavour and opposite charge ($\ell^+\ell^-$). The pair with an invariant mass closest to the nominal Z mass is denoted m_{Z_1} and retained if it satisfies $40 < m_{Z_1} < 120\ GeV/c^2$. Then, all remaining leptons are considered and a second pair of $\ell^+\ell^-$ is required, with mass denoted m_{Z_2} and satisfying $12 < m_{Z_2} < 120\ GeV/c^2$. The $12\ GeV/c^2$ cut provides an optimal sensitivity for a Higgs boson mass hypothesis in the range $110 < m_H < 160\ GeV/c^2$. If more than one Z_2 candidate satisfies all criteria, the ambiguity is resolved by choosing the leptons of highest p_T. Among the four selected leptons forming Z_1 and Z_2, at least one should have $p_T > 20\ GeV/c$ and another one have $p_T > 10\ GeV/c$. These p_T thresholds ensure that the selected events have leptons on the high-efficiency plateau for the trigger. Furthermore, to protect against leptons originating from hadron decays in jet fragmentation or from the decay of low-mass hadronic resonances, is required that any opposite-charge pair of leptons chosen among the four selected leptons (irrespective of flavour) satisfy $m_{\ell\ell'} > 4\ GeV/c^2$. The phase space for the search of the SM Higgs boson is defined by restricting the mass range to $m_{4\ell} > 100\ GeV/c^2$. A higher minimal threshold on m_{Z_1} and m_{Z_2} could be used for higher m_H values but only with marginal improvement of the sensitivity.
The event yields are found to be in good agreement with the MC background expectation at each step of event selection.

3.5 Background control and systematics

As said in section 3.1, in the $H \to ZZ \to 4\ell$ channel, the background sources include an irreducible four-lepton contribution from direct ZZ production via $q\bar{q}$ and gluon-gluon processes. Reducible contributions arise from $Z + b\bar{b}$ and $t\bar{t}$ production where the final states contain two isolated leptons and two b-quark jets producing secondary leptons. Additional background arises from $Z + jets$ and $WZ + jets$ events where jets are misidentified as leptons. The analysis relies on MC simulation to evaluate the local density ($\Delta N/\Delta m_{4\ell}$) of events expected as a function of the mass $m_{4\ell}$ from the ZZ background. The cross section for ZZ production at NLO is calculated separately for the dominant process, $q\bar{q}$ annihilation, and for gluon-gluon fusion. The theoretical uncertainties are computed as a function of $m_{4\ell}$, varying both the QCD renormalisation and factorization scales and the PDF set, following the PDF4LHC recommendations. The uncertainties for the QCD and PDF scales for each final state are on average 8%.
The number of predicted $ZZ \to 4\ell$ events and their uncertainties after the signal selection are given in Table 3.1.

Table 3.1: *The number of event candidates observed, compared to the mean expected background and signal rates for each final state. For the ZX background, the estimations are based on data. The results are given integrated over the full mass measurement range for the SM-like Higgs boson search from* 100 *to* 1000 GeV/c^2 *and for 2011 and 2012 data combined.*

Channel	$4e$	4μ	$2e2\mu$
ZZ background	53.0 ± 6.3	82.7 ± 8.9	131.1 ± 14.3
Z+X	$7.6^{+6.9}_{-5.2}$	$2.9^{+2.2}_{-1.6}$	$10.1^{+9.9}_{-6.5}$
All backgrounds	$60.7^{+9.3}_{-8.2}$	$85.6^{+9.2}_{-9.1}$	$141.3^{+17.3}_{-15.7}$
$m_H = 125\ GeV/c^2$	2.4 ± 0.4	4.6 ± 0.5	6.0 ± 0.7
$m_H = 126\ GeV/c^2$	2.7 ± 0.4	5.1 ± 0.6	6.6 ± 0.8
$m_H = 200\ GeV/c^2$	15.9 ± 1.9	23.1 ± 2.6	38.5 ± 4.3
$m_H = 350\ GeV/c^2$	9.5 ± 1.2	13.6 ± 1.5	23.2 ± 2.7
$m_H = 500\ GeV/c^2$	3.3 ± 0.4	4.7 ± 0.6	8.1 ± 0.9
$m_H = 200\ GeV/c^2$	0.5 ± 0.1	0.6 ± 0.1	1.1 ± 0.1
Observed	59	95	162

There will be a more accurate treatment of irreducible backgrounds in Chapter 4, since the main goal of this work is to find a method for separating the two irreducible backgrounds ($q\bar{q} \to ZZ$ and $gg \to ZZ$).

To estimate the reducible ($Zb\bar{b}, t\bar{t}$) and instrumental ($Z + lightjets, WZ + jets$) backgrounds, a $Z_1 + X$ background control region, well separated from the signal region, is defined. In addition, a sample $Z_1 + \ell_{reco}$, with only one reconstructed lepton object is defined for the measurement of the lepton misidentification probability, the probability for a reconstructed object to pass the isolation and identification requirements (fake rate). The contamination from WZ in these events is suppressed by requiring the imbalance of the measured energy deposition in the transverse plane to be below 25 GeV. The lepton misidentification probability is compared, and found compatible, with the one derived from MC simulation. The event rates measured in the background control region are extrapolated to the signal region using fake rate measurement.
For the 4ℓ background estimate, two different approaches are used. Both start by relaxing the isolation and identification criteria for two additional reconstructed lepton objects. In the first method the additional pair of leptons is required to have the same charge (to avoid signal contamination) and same flavour ($e^{\pm}e^{\pm}, \mu^{\pm}\mu^{\pm}$), a reconstructed invariant mass $m_{Z_2} > 12\ GeV/c^2$, and $m_{4\ell} > 100\ GeV/c^2$. The expected number of $Z+X$ background events in the

signal region is obtained by taking into account the lepton misidentification probability for each of the two additional leptons. The second approach uses the control region with two opposite-sign leptons failing the isolation and identification criteria. In addition, a control region with three passing and one failing lepton is also used to account for contributions from backgrounds with three prompt leptons and one misidentified lepton. Comparable background counts in the signal region are found within uncertainties from both methods. Results are reported in Table 3.1.

Systematic uncertainties are evaluated from data for trigger (1.5%), and combined lepton reconstruction, identification and isolation efficiencies (varying from 1.2% in the 4μ channel at high masses to about 11% in $4e$ channel at low masses). Systematic uncertainties on energy-momentum calibration (0.4% for muons and 0.2% for electrons) and energy resolution are accounted for by their effects on the reconstructed mass distributions. The effect of the energy resolution uncertainties is taken into account by introducing a 20% uncertainty on the simulated width of the signal mass peak. To validate the level of accuracy with which the absolute mass scale and resolution are known, the analysis uses $Z \to \ell\ell$ and $J/\psi \to \ell\ell$ events. Additional systematic uncertainties arise from limited statistical precision in the reducible background control regions. All reducible and instrumental background sources are derived from control regions, and the comparison of data with the background expectation in the signal region is independent of the uncertainty on the LHC integrated luminosity of the data sample. This uncertainty (4.4%) enters the evaluation of the ZZ background and the calculation of the cross section limit through the normalisation of the signal.

3.6 Kinematic Discriminant (MELA)

3.6.1 Introduction of the methodology

Kinematics of the Higgs decay to ZZ final state has been extensively studied in the literature in application to the studies of the Higgs boson or new exotic boson properties. Recently a complete set of angular observables was introduced [32] and it was suggested that it may also help in background rejection. In this approach, the signal-to-background probability is created using analytical multi-dimension likelihood for an event to be signal or background. Signal and background analytical parametrisations are taken from Refs.[32] and [38], respectively. In the following, the methodology will be introduced in more detail with the analytical MELA approach (Matrix Element Likelihood Approach).
In Figure 3.4 the angles involved in the production and decay chain $ab \to X \to ZZ \to 4\ell$ are illustrated. The full kinematics in the production and decay of an X resonance $ab \to X \to Z_1 Z_2 \to 4\ell$ can be described with the

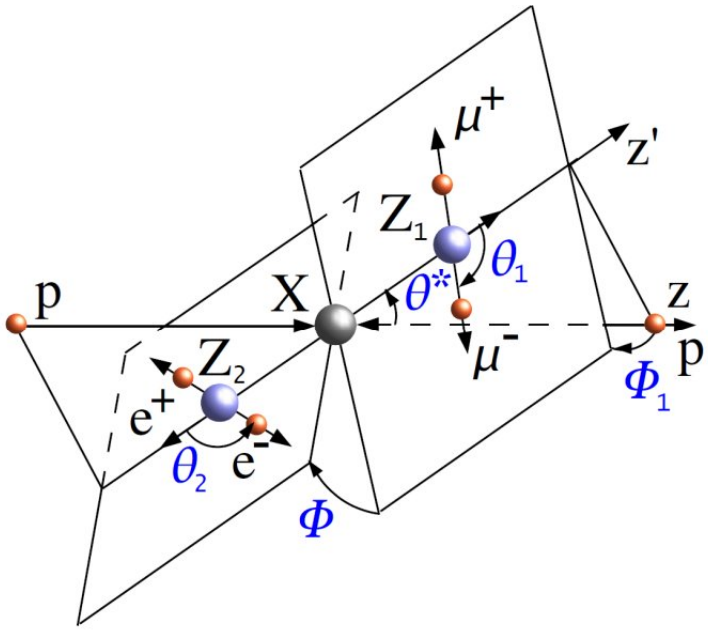

Figure 3.4: *Illustration of a particle X production and decay* $ab \to X \to Z_1 Z_2 \to 4\ell$ *with the two production angles* θ^* *and* Φ_1 *shown in the X rest frame and three decay angles* θ_1, θ_2 *and* Φ *shown in the* P_i *rest frames.*

help of the following 12 observables:

- three resonance masses (including the off-shell cases): $m_{4\ell}, m_1, m_2$;
- five production and decay angles defined in Figure 3.4 as $\vec{\Omega} = \theta^*, \Phi_1, \theta_1, \theta_2, \Phi$;
- longitudinal boost of the resonance, expressed as rapidity Υ;
- transverse momentum of the resonance p_T and its azimuthal angle;
- one arbitrary azimuthal angle Φ^* reflecting the overall orientation of the system.

In the current approach, a kinematic discriminant KD from the seven mass and angular observables, $KD = F(m_1, m_2, \theta^*, \Phi_1, \theta_1, \theta_2, \Phi)$, is constructed, and a 2D shape fit with the two observables $(m_{4\ell}, KD)$ is performed. In Figures 3.5 and 3.6 separation power between signal and background for each individual observable is shown. The KD discriminant combines this power in a single observable using full correlation of all input observables in the most optimal way. In this approach only observables coming from well-understood electro-weak Quantum Mechanics of the processes of either Higgs or continuum ZZ production are kept, while those observables which depend on QCD kinematics, such as Υ and p_T, are not considered. It is important

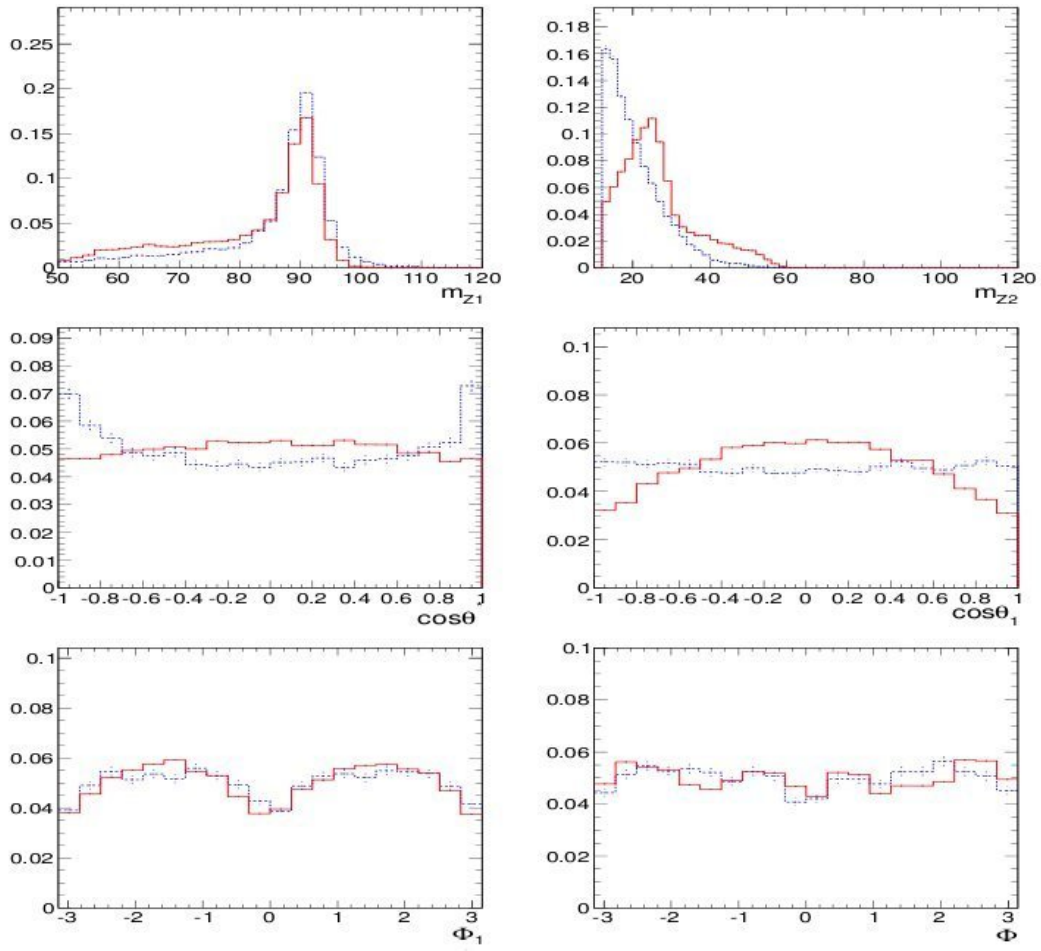

Figure 3.5: *Distribution of Higgs signal events with* $m_H = 120\ GeV/c^2$ *(solid red) and background ZZ events (dashed blue) in the range* $100 < m_{4\ell} < 135\ GeV/c^2$.

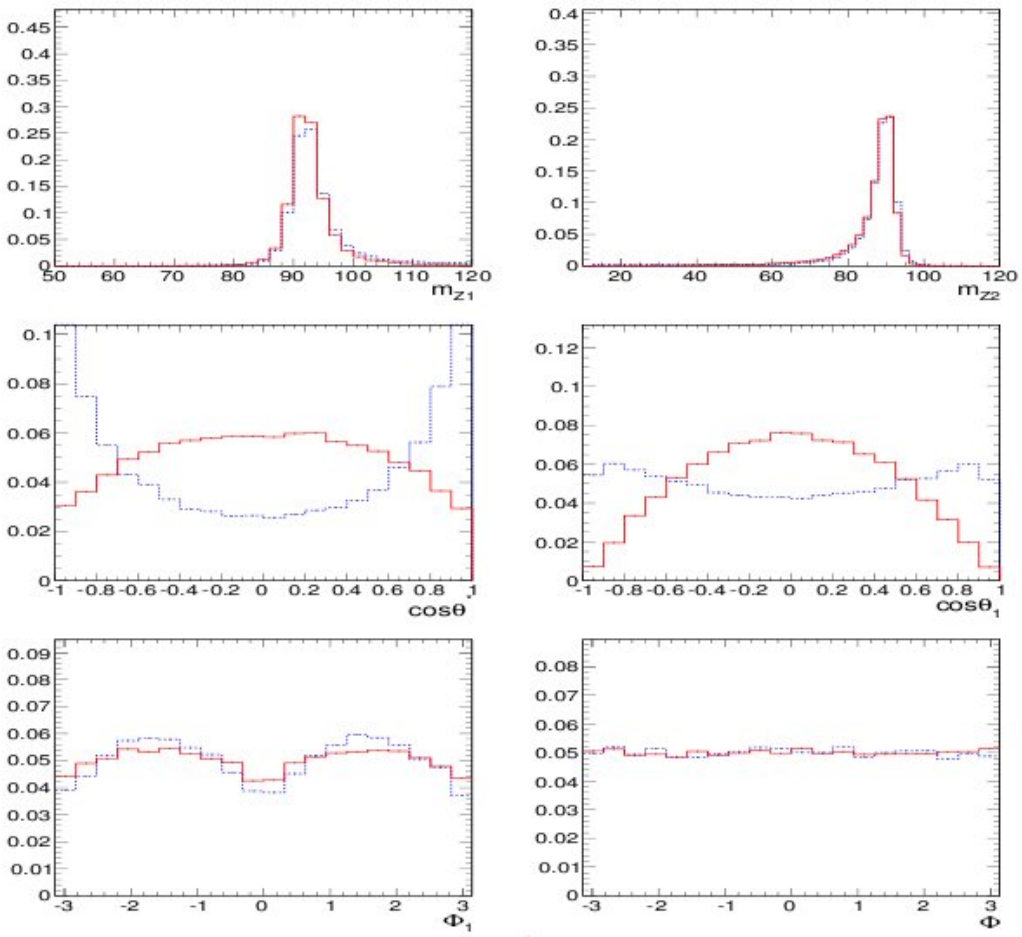

Figure 3.6: *Distribution of Higgs signal events with* $m_H = 400\ GeV/c^2$ *(solid red) and background ZZ events (dashed blue) in the range* $300 < m_{4\ell} < 500\ GeV/c^2$.

to underline that the seven observables used are independent from the Higgs production mechanism, as long as the analysis consider the SM Higgs boson, which is spin zero according to the theory. The same seven observables are also the key input to measuring the new boson properties, such as spin and CP quantum numbers.

3.6.2 Construction of the MELA discriminant

Construction of the kinematic discriminant KD in the MELA approach (Matrix Element Likelihood Approach) relies on probability for an event with a set of observables $(m_{4\ell}, m_1, m_2, \vec{\Omega})$ to come either from signal or background

$$P_{sig}(m_1, m_2, \vec{\Omega}|m_{4\ell})$$

$$P_{bkg}(m_1, m_2, \vec{\Omega}|m_{4\ell})$$

where probabilities are normalized with respect to the seven observables and $m_{4\ell}$ is treated as a conditional parameter. These probabilities are calculated analytically and are quoted in Ref.[32] for signal and in Ref.[38] for continuum ZZ background. Then the discriminant is constructed as follows

$$KD = \frac{P_{sig}}{P_{sig} + P_{bkg}} = \left[1 + \frac{P_{bkg}(m_1, m_2, \vec{\Omega}|m_{4\ell})}{P_{sig}(m_1, m_2, \vec{\Omega}|m_{4\ell})}\right]^{-1} \tag{3.1}$$

This discriminant is continuously distributed between 0 and 1, with signal being closer to 1 and background closer to 0. The signal probability is parametrised as a function of $m_{4\ell}$ instead of m_H. This allows continuous selection of the data-sample independent of the m_H hypothesis. Moreover, both signal and background probabilities are normalized for any given value of $m_{4\ell}$, which removes unnecessary correlation of KD with $m_{4\ell}$ and makes further fit implementation more robust. Parametrisation is performed for ideal distributions, not including the detector effects; this is still an optimal approach because detector acceptance effects are identical for signal and background and would cancel in the ratio in Equation 3.1.
The ideal probability density functions for signal and irreducible background can both be calculated analytically. Projections both of the signal and of the irreducible background probability density function are shown in Figures 3.7 and 3.8. For background below threshold, it is also possible to substitute the analytical parametrisation with the correlated template distribution. This simplified parametrisation was used in the past for events below the $2m_Z$ kinematic threshold while the analytic PDF mentioned above was used above threshold, fixing both Z masses to 91.2 GeV.
The resulting MELA KD distributions for signal and background are shown in Figure 3.9 in three different mass ranges. Good agreement is found between data and background MC. Overall, significant separation between signal and background is evident from the MELA KD distributions.

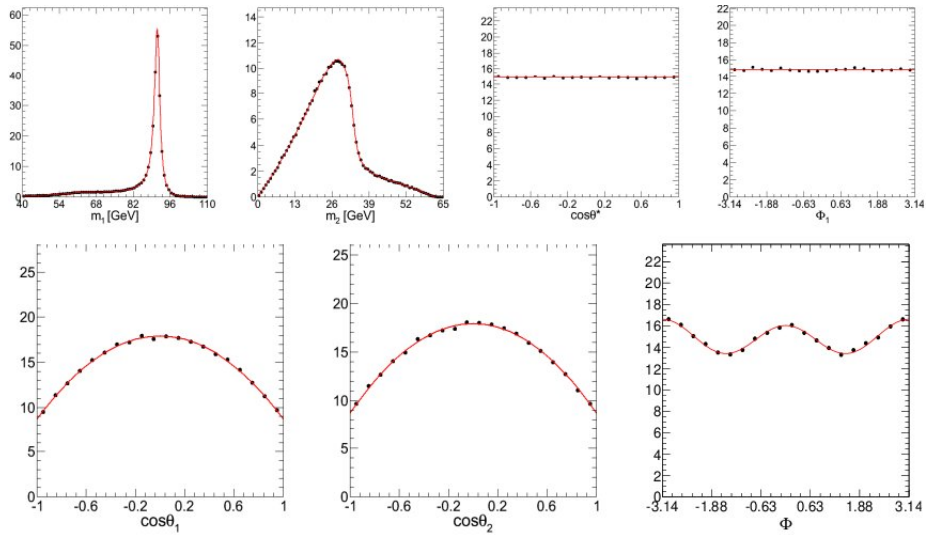

Figure 3.7: *Projections of SM Higgs PDF. Events are generated at leading order through gluon-gluon fusion.*

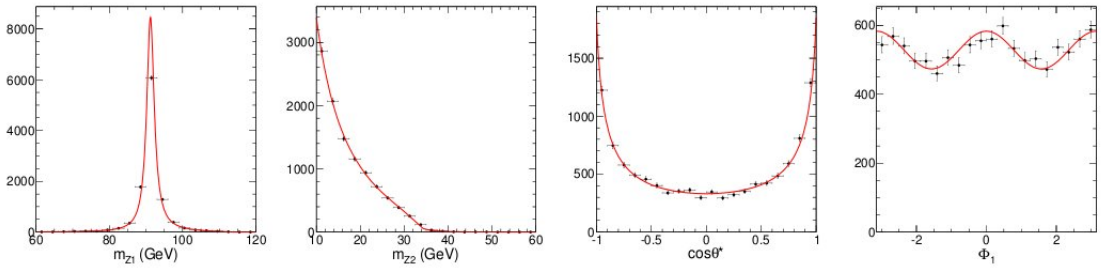

Figure 3.8: *Projections of ZZ continuum PDF. Data is leading order ideal Madgraph MC events which includes both $ZZ \to 4\ell$ and $Z\gamma^* \to 4\ell$ processes.*

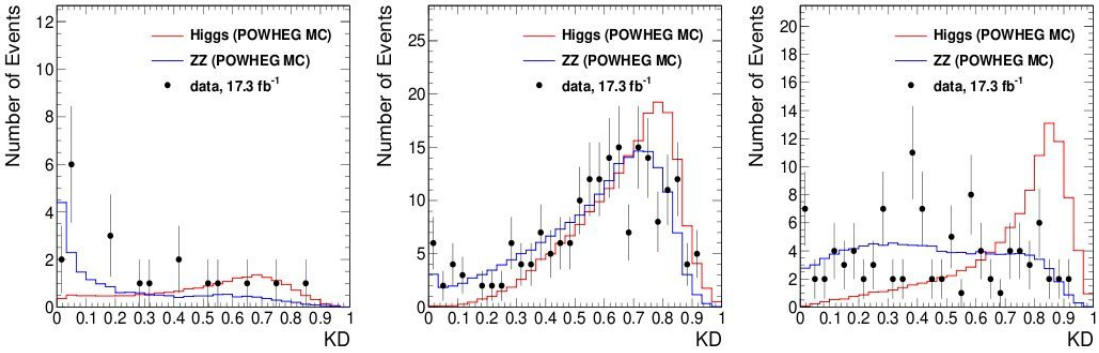

Figure 3.9: *The KD distributions for signal and background in three mass ranges:* $140 < m_{4\ell} < 160$ *GeV*$/c^2$ *(top),* $200 < m_{4\ell} < 300$ *GeV*$/c^2$ *(middle), and* $250 < m_{4\ell} < 450$ *GeV*$/c^2$ *(bottom). The signal (red solid histogram) is shown for* $m_H = 150, 250,$ *and* 350 *GeV*$/c^2$*, respectively. The ZZ continuum background is shown as blue solid histogram. The top plot also shows Z+X background estimated from data control region.*

3.6.3 Parametrisation of the MELA discriminant

While the MELA parametrisation is performed using the ideal distributions in Equation 3.1, the resulting observable KD is a certain number for each event candidate. This observable is parametrised using MC samples (with full CMS simulation and with data-to-MC corrections applied) for the processes well modeled in MC, such as signal and ZZ background, and using data control samples for instrumental and reducible background, such as $Z + X$.

The KD distributions are somewhat affected by interference of identical leptons in the final states $ZZ \to e^+e^-e^+e^-$ and $\mu^+\mu^-\mu^+\mu^-$, but not in $2e2\mu$. This effect is only relevant at low masses, below the ZZ threshold, where at least one of the Z bosons is off-shell. The background simulation of continuum ZZ background already includes interference effects and this is taken into account when KD distributions are parametrised for background. The signal POWHEG+Pythia simulation of $H \to ZZ$ signal does not include such interference effect and Prophecy event generator is used to calculate the difference.

The KD distributions depend on the value of the $m_{4\ell}$, due to kinematics dependence on the mass. Therefore, any further analysis of the data which includes KD must also include its full correlation with $m_{4\ell}$. Below $2m_Z$ threshold, the dominant background is $q\bar{q} \to ZZ$ with a secondary contribution of $Z + X$ background. Above $2m_Z$ threshold, the dominant background is also $q\bar{q} \to ZZ$ with a secondary contribution of $gg \to ZZ$ background. For the signal, $q\bar{q} \to ZZ$ and $gg \to ZZ$ background, the ideal angular and mass distributions come from basic quantum mechanics given the EWK couplings and are modeled well by MC.

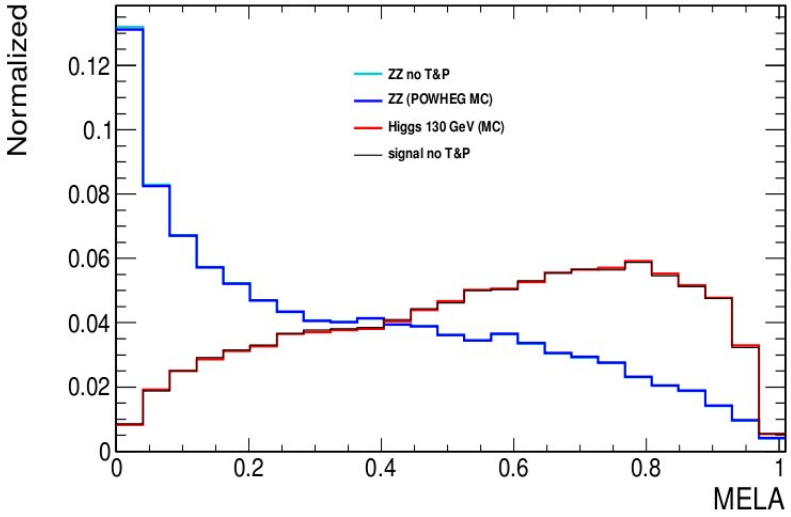

Figure 3.10: *Distribution of the MELA KD for ZZ background and* 130 GeV/c^2 *Higgs signal in MC, for events below* $2m_Z$ *threshold (background peaks towards 0 and signal peaks towards 1). For both signal and background two very close distributions are shown: with and without the Tag-and-Probe corrections for data-MC differences.*

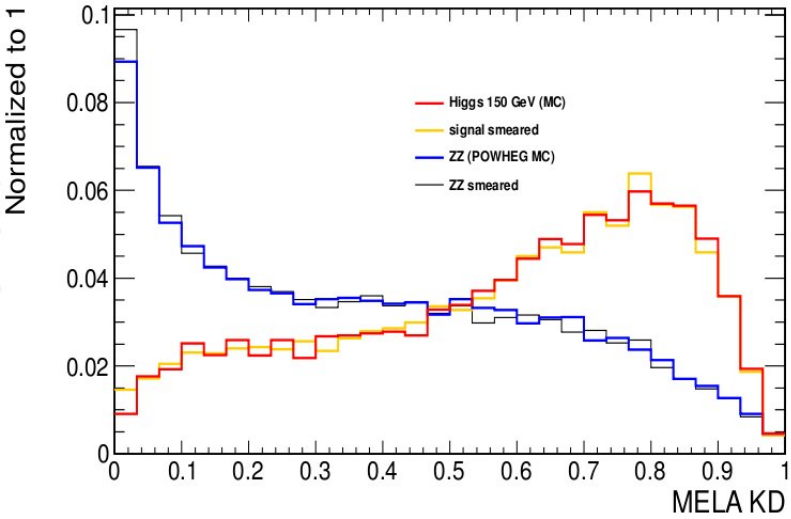

Figure 3.11: *Distribution of the MELA KD for ZZ background and* 150 GeV/c^2 *Higgs signal in MC, for events below* $2m_Z$ *threshold (background peaks towards 0 and signal peaks towards 1). For both signal and background two very close distributions are shown: with and without smearing of electron energy resolution, which is equal to uncertainty on this resolution. Only* $H \to ZZ \to 4e$ *channel is considered here, to show the largest effect.*

Detector effects related to mis-modeling of lepton efficiency and resolutions have very small effects on KD distributions because, as opposed to $m_{4\ell}$, there is no distinct peak but just smearing of broad distribution. In Figure 3.10, an example of the shape variation due to extreme variation of MC efficiency by changing the Tag-and-Probe corrections from those obtained from data to flat ones (typically this variation is larger than the errors) is shown. In Figure 3.11, an example of the shape variation due to extreme variation of MC resolution in the $H \to ZZ \to 4e$ channel is presented. Therefore, such detector effects are ignored in systematic uncertainties for the KD, since they are significantly smaller than statistical effects in the data to be analyzed.

3.7 Results

The reconstructed four-lepton invariant-mass distribution obtained combining the $4e$, 4μ and $2e2\mu$ channels with the baseline selection is shown in Figure 3.12 for data and MC expectations. As one can see, the $Z \to 4\ell$ de-

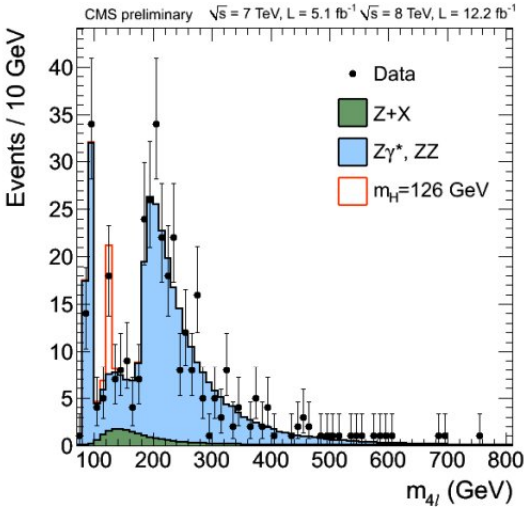

Figure 3.12: *Distribution of the four-lepton reconstructed mass in full mass range for the sum of the* $4e$, 4μ *and* $2e2\mu$ *channels. The distributions are presented as stacked histograms. The measurements are presented for the sum of the data collected at* $\sqrt{s} = 7\ TeV$ *and* $\sqrt{s} = 8\ TeV$. *No event is observed for* $m_{4\ell} > 800\ GeV/c^2$ *[24].*

cays give a clean resonant peak in the four-lepton invariant mass distribution around $m_{4\ell} = m_Z$, as expected. The $Z \to 4\ell$ can be used as a standard candle in the context of the Higgs boson search in the $H \to ZZ \to 4\ell$ decay mode [23], being a crosscheck of the understanding of the four-lepton mass scale, the four-lepton mass resolution, and the overall four-lepton reconstruction efficiency. The measured distribution at higher mass is in agreement with the expectation dominated by the irreducible ZZ background.

The number of candidates observed as well as the estimated background in the signal region are reported in Table 3.1, for the selection in the full mass measurement range for the Higgs boson search.
The expected number of signal events is also given for several SM-like Higgs boson mass hypotheses.

The distributions of the MELA KD versus the four-lepton reconstructed mass $m_{4\ell}$ is shown for the selected events and compared to SM background expectation in Figure 3.13. The distribution of events in the $(m_{4\ell}, KD)$

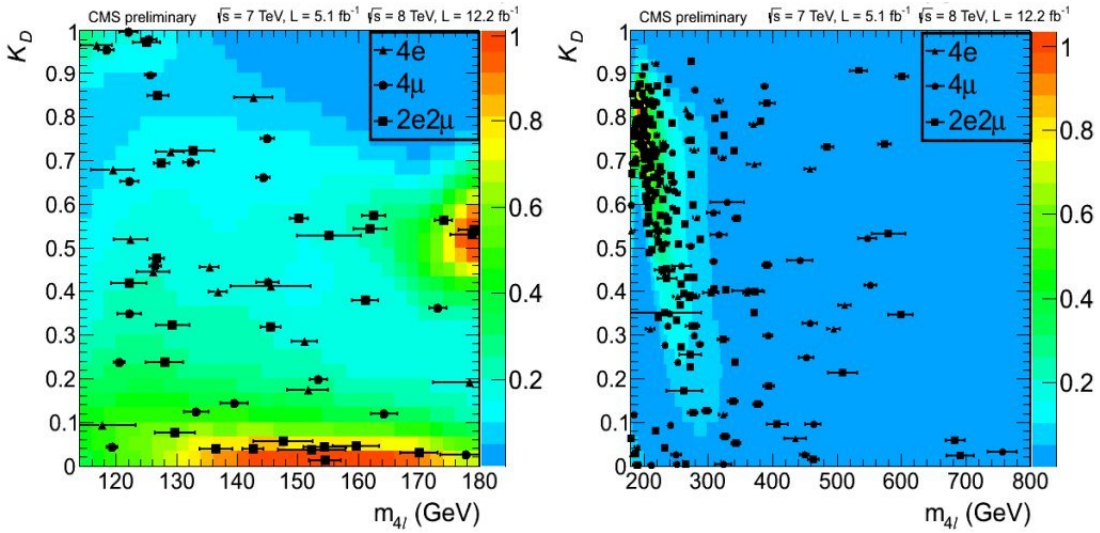

Figure 3.13: *Distribution of the MELA KD versus the four-lepton reconstructed mass $m_{4\ell}$ in the low-mass (left) and high-mass (right) regions. The contours represent the expected relative density of background events. The points show data with measured invariant mass uncertainties. No event is observed for $m_{4\ell} > 800$ GeV/c^2.*

plane agrees well with the SM expectation in the high mass range (Figure 3.13, right).
The probability distribution of $P(m_{4\ell})$ for the background is parametrised with empirical functions using MC simulation for ZZ background and data control regions for $Z + X$ background. The reconstructed signal $m_{4\ell}$ distributions are described with a relativistic Breit-Wigner parametrisation convoluted with a double-sided Crystal-Ball function. The correlated two-dimensional $(m_{4\ell}, KD)$ distribution is described by the one-dimensional probability distribution $P(m_{4\ell})$ multiplied by a two-dimensional template distribution normalised in the KD dimension. This template distribution is obtained from simulation for both signal and ZZ background, accounting for interference effects of identical leptons in the final state. It has been verified that the KD distribution of the $Z + X$ background is consistent with that of

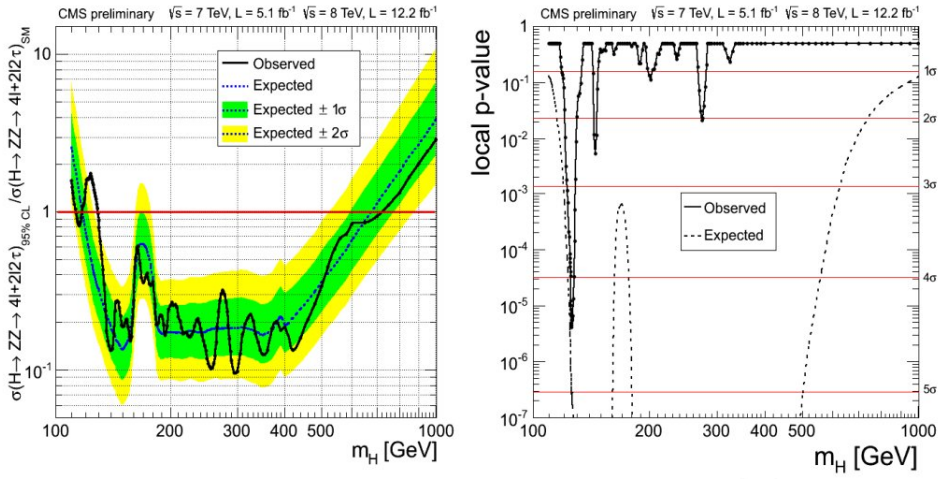

Figure 3.14: *Observed and expected* 95% *CL upper limit (left) on the ratio of the production cross section to the SM expectation. The* 68% *and* 95% *ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. Significance of the local excess (right) with respect to the standard model background expectation as a function of the Higgs boson mass in the full interpretation mass range* $110 - 1000\ GeV/c^2$.

the ZZ background, and any potential small difference is accounted for in the systematic uncertainties.
The measured distributions are compared with the expectation from SM background processes, and exclusion limits at 95% CL on the ratio of the production cross section for the SM-like Higgs boson to the SM expectation are derived. For this, the $(m_{4\ell}, KD)$ distributions of the selected events are split into six categories based on three final states and two running periods (7 and 8 TeV). These events are examined for 187 hypothetical SM-like Higgs boson masses in a range between 110 GeV/c^2 and 1000 GeV/c^2, where the mass steps are optimized to account for the expected width, Γ_H, and resolution for the measurement of m_H. For each mass hypothesis, a simultaneous likelihood fit of the six two-dimensional $(m_{4\ell}, KD)$ distributions is performed. Due to the large number of mass points and availability of simulated signal samples, one must interpolate the mass shapes of signal hypotheses where no simulation exists.

The upper limits on the ratio of the production cross section to the SM expectation obtained from the combination of the 4ℓ channel are shown in Figure 3.14 (left). If, for a particular range of m_H, the normalized upper limit is lower than unity, it means that the allowed cross section maximum value is smaller than the SM prediction. The SM Higgs boson is excluded by the search in the four-lepton channels at 95% CL in the range $113 - 116\ GeV/c^2$

and $129 - 720\ GeV/c^2$. The upper limits in the low-mass region are given in Figure 3.15 (left). The local p-value, representing the significance of local

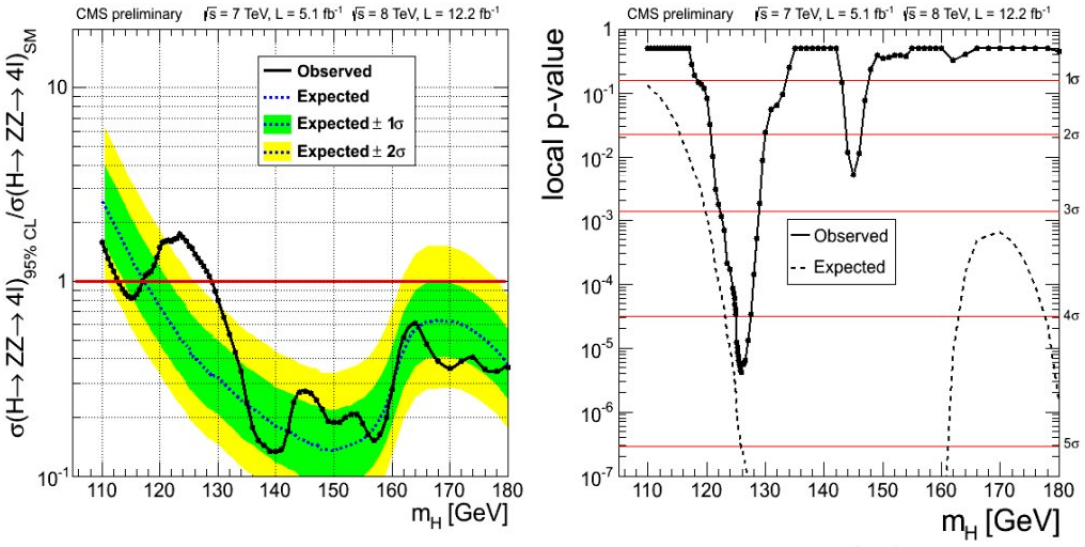

Figure 3.15: *Observed and expected* 95% *CL upper limit (left) on the ratio of the production cross section to the SM expectation, in the low-mass region. The* 68% *and* 95% *ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. Significance of the local excess (right) with respect to the standard model background expectation as a function of the Higgs boson mass. The results are shown for the full data sample in the low-mass region only.*

excesses relative to the background expectation, is shown for the full mass range as a function of m_H in Figure 3.14 (right) and for the low-mass region in Figure 3.15 (right). As one can note, the minimum of the local p-value is reached at low mass around $m_{4\ell} = 125.9\ GeV/c^2$, near the mass of the new boson [31], and corresponds to a local significance of 4.5σ. The number of observed and predicted events in the mass region near the signal, from 110 to 160 GeV/c^2, where the background is expected to be relatively flat, is reported in Table 3.2. The distribution of the four-lepton reconstructed mass for the sum of the $4e$, 4μ and $2e2\mu$ channels, and the distribution of the MELA KD versus the four-lepton reconstructed mass $m_{4\ell}$ are shown in Figure 3.16 in the low mass range. A signal-like clustering of events is apparent at high values of KD, as seen in Figure 3.17 (left), and for $m_H \approx 126\ GeV/c^2$. As an illustration, the reconstructed four-lepton invariant-mass distribution for the 4ℓ is shown in Figure 3.17 (right) for events with $KD > 0.5$ as well as the KD distribution for $m_{4\ell}$.

The signal strength μ, relative to the expectation for the SM Higgs boson, is measured to be $\mu = \frac{\sigma}{\sigma_{SM}} = 0.80^{+0.35}_{-0.28}$ at $126\ GeV/c^2$. The local significance of 3.1σ is reached in the 1D fit without the MELA KD. The average

Table 3.2: *The number of event candidates observed, compared to the mean expected background and signal rates for each final state. For the $Z + X$ background, the estimations are based on data. The results are given integrated the mass range from 110 to 160 GeV/c^2.*

Channel	$4e$	4μ	$2e2\mu$	4ℓ
ZZ background	4.7 ± 0.6	9.6 ± 1.0	12.5 ± 1.4	26.8 ± 1.8
Z+X	$3.4^{+3.0}_{-2.3}$	$1.6^{+1.2}_{-0.9}$	$5.6^{+5.4}_{-3.6}$	$10.6^{+5.3}_{-4.4}$
All backgrounds	$8.0^{+3.1}_{-2.3}$	$11.2^{+1.6}_{-1.4}$	$18.1^{+5.6}_{-3.8}$	$37.3^{+6.6}_{-4.7}$
$m_H = 125\ GeV/c^2$	2.4 ± 0.4	4.6 ± 0.5	5.9 ± 0.7	12.9 ± 0.9
$m_H = 126\ GeV/c^2$	2.7 ± 0.4	5.1 ± 0.6	6.6 ± 0.8	14.4 ± 1.1
Observed	12	16	19	47

expected significance for a SM Higgs boson at this mass is 5.0σ and 4.3σ for the 2D and 1D fits, respectively. Using simulation it was found that the MELA KD distribution for signal at a mass around $m_H = 126\ GeV/c^2$ is similar for a scalar, pseudo-scalar, or a spin-two resonance with the minimal couplings. Therefore the analysis presented is nearly model-indepedent in the low-mass region. In summary, the new boson recently discovered by the ATLAS and CMS experiments is observed in the 4ℓ channel, with a local significance of 4.5 standard deviations above the expected background. The signal strength μ, relative to the expectation for the standard model Higgs boson, is measured to be $\mu = 0.80^{+0.35}_{-0.28}$ at $126\ GeV/c^2$.

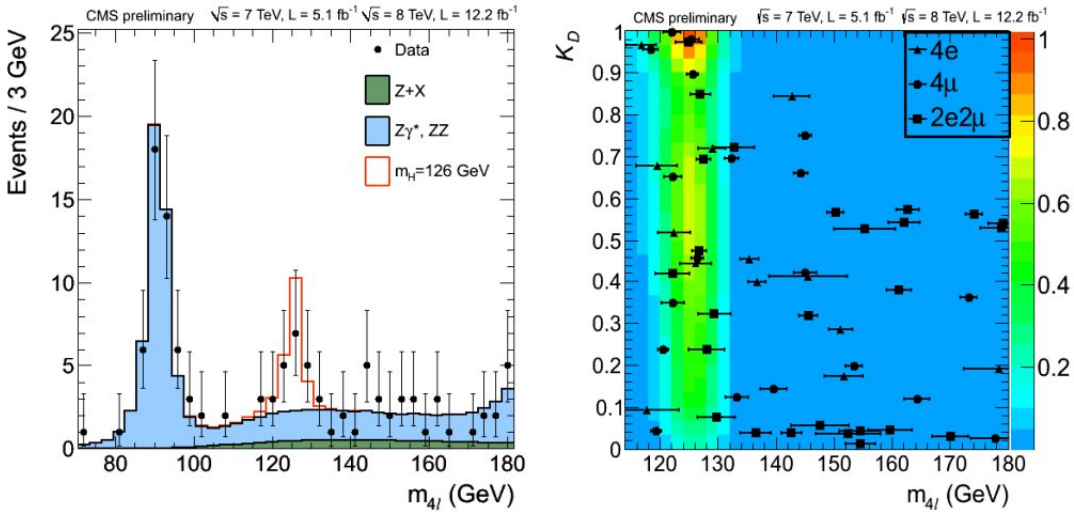

Figure 3.16: *Distribution of the four-lepton reconstructed mass for the sum of the* $4e$, 4μ *and* $2e2\mu$ *channels (left). Points represent the data, shaded histograms represent the background and unshaded histogram the signal expectations. Distribution of the MELA KD versus the four-lepton reconstructed mass* $m_{4\ell}$ *(right) with contours shown for expected relative density of signal events for hypothesis* $m_H = 126\ GeV/c^2$*. The points show data with measured invariant mass uncertainties.*

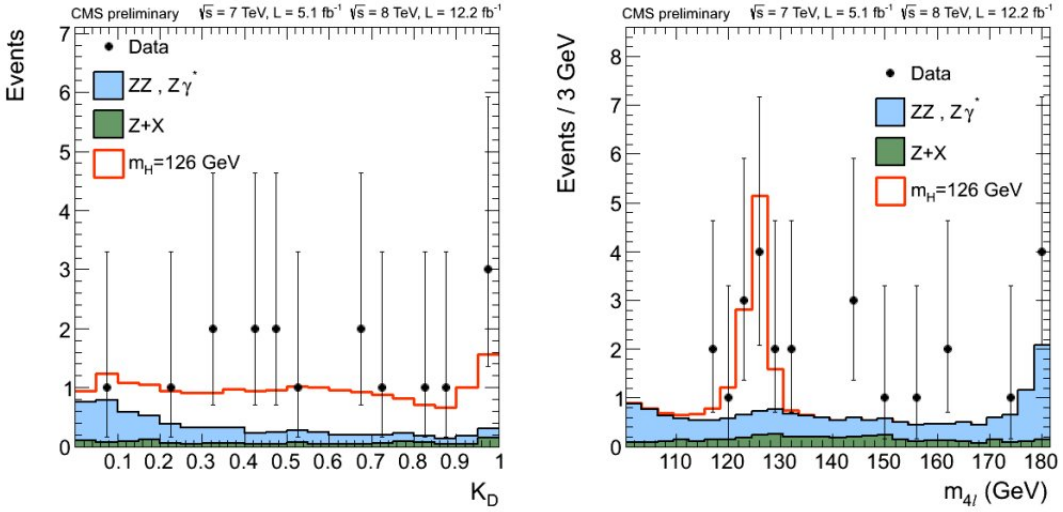

Figure 3.17: *Distribution of the MELA kinematic discriminant for events in the mass region* $121.5 < m_{4\ell} < 130.5\ GeV/c^2$ *(left). Distribution of the four-lepton reconstructed mass for the sum of the* $4e$, 4μ *and* $2e2\mu$ *channels for events with a value* $KD > 0.5$ *of the MELA kinematic discriminant (right). Points represent the data, shaded histograms represent the background and unshaded histogram the signal expectations. The measurements are presented for the sum of the data collected at* $\sqrt{s} = 7\ TeV$ *and* $\sqrt{s} = 8\ TeV$.

Chapter 4

Study of $q\bar{q} \to ZZ$ and $gg \to ZZ$ Backgrounds

In this Chapter the attention is focused on the $q\bar{q} \to ZZ$ and $gg \to ZZ$ processes, irreducible backgrounds of the $H \to ZZ^{(*)} \to 4\ell$ channel.
These backgrounds have a topology and kinematic very similar to those of signal events. An appropriate knowledge of these processes is therefore needed to minimize the systematic uncertainties.

In the following an introduction on irreducible backgrounds, the motivations and the aim of this thesis work, and the study of $q\bar{q} \to ZZ$ and $gg \to ZZ$ angular distributions at generator and reconstruction level are presented.

4.1 Irreducible ZZ background model and uncertainties

Four-leptons events from continuum di-boson production constitute the main source of background for the direct searches for the Higgs boson in the $H \to ZZ \to 4\ell$ channel at LHC. As said before, since the event topology and kinematic are very similar to those of signal events, these processes are considered in the category of *irreducible* backgrounds. The lowest order produc-

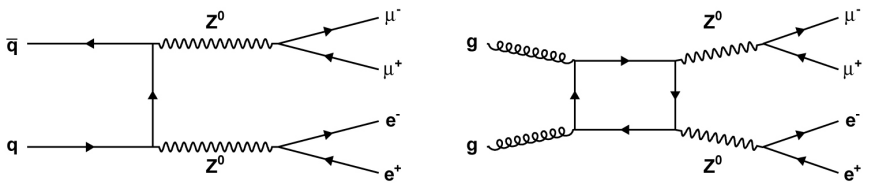

Figure 4.1: *Lowest order diagrams for the $q\bar{q} \to ZZ^*/Z\gamma^*$ process (left) and for the $gg \to ZZ^*/Z\gamma^*$ process (right).*

tion mechanism is the one represented in Figure 4.1 (left), $q\bar{q} \to ZZ^*/Z\gamma^*$. The gluon-gluon induced ZZ background, although technically at order α_s^2 compared to the first order Z-pair production, amounts to a non-negligible fraction of the total irreducible background at masses above the $2m_Z$ threshold. The associated diagram is represented in Figure 4.1 (right).
The main goal of this section is to compare the predictions from different generators for the ZZ process (in particular for the four-lepton final state) [39]. The PDF's used are: CT10, NNPDF2.0, and MSTW2008. The central value of the cross section and the uncertainty originating from the PDF are estimated following the PDF4LHC prescription. In order to estimate the QCD scale uncertainty the CT10 PDF has been used as central value and the QCD scale is varied following the prescription of Ref.[40].
Differential distributions and theoretical uncertainties are calculated for a few sets of cuts:

- Cut 1: $m_{Z_1} > 12\ GeV/c^2$ and $m_{Z_2} > 12\ GeV/c^2$,
- Cut 2: $m_{Z_1} > 50\ GeV/c^2$, $m_{Z_2} > 12\ GeV/c^2$, $p_T(\ell) > 5\ GeV/c$, and $|\eta(\ell)| < 2.5$,
- Cut 3: $60 < m_{Z_1} < 120\ GeV/c^2$ and $60 < m_{Z_2} < 120\ GeV/c^2$.

4.1.1 $q\bar{q} \to ZZ$ generators

MCFM predictions

The MCFM program v6.1 computes the cross section at LO and NLO for the process $q\bar{q} \to \ell\bar{\ell}\ell'\bar{\ell}'$ mediated by the exchange of the two bosons Z, γ^* and their interference, for the doubly-resonant (or t-channel) and singly-resonant (or s-channel) diagrams, and for the process $gg \to \ell\bar{\ell}\ell'\bar{\ell}'$.
Figure 4.2 (left) shows the four-lepton invariant-mass distribution obtained with MCFM at LO for ZZ production, including or not the singly-resonant (or s-channel) contribution. The prediction without the singly-resonant contribution is obtained with MCFM v5.8. The "Cut 1" selection is applied, i.e. $m_{\ell\ell} > 12\ GeV/c^2$. At LO the cross section as a function of $m_{2e2\mu}$ for the process $q\bar{q} \to ZZ \to 2e2\mu$ and for three different set of cuts is shown in Figure 4.2 (right). The black line corresponds to the "Cut 1" selection, i.e. $m_{\ell\ell} > 12\ GeV/c^2$, that is the minimal cut requested in the analyses to subtract the contribution to heavy-flavour resonances decaying into leptons. The blue line is obtained taking into account the acceptance of the detectors, thus asking the leptons to have $p_T > 5\ GeV/c$ and $|\eta| < 2.5$ ("Cut 2" selection). The red line is the differential cross section for the production of two Z bosons, both on shell, i.e. asking for $60 < m_{\ell\ell} < 120\ GeV/c^2$ ("Cut 3" selection).
Figure 4.3 (left) shows the cross section at NLO for the full process $pp \to ZZ^{(*)} \to 2e2\mu$ and the gg contribution separately in the inset, as a function

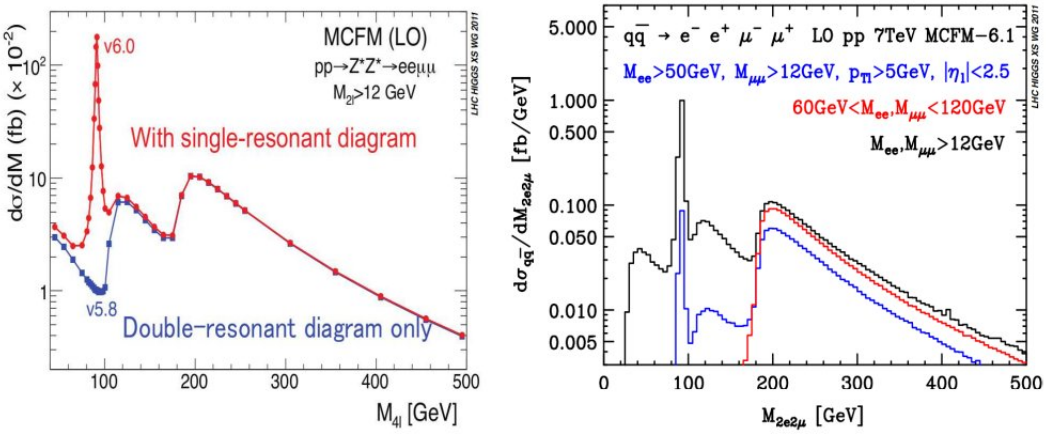

Figure 4.2: *(left) The cross sections for $q\bar{q} \to ZZ^{(*)} \to 2e2\mu$ as a function of $m_{2e2\mu}$ at 7 TeV from MCFM for the full calculation and without the singly-resonant contribution; (right) LO cross sections for $q\bar{q} \to ZZ^{(*)} \to 2e2\mu$ as a function of $m_{2e2\mu}$ at 7 TeV from MCFM at LO [39].*

of the invariant mass of the 4 leptons. As one can see, the peak at 91 GeV/c^2 is dominated by the contribution of the singly-resonant (or s-channel) diagrams, while for $m_{4\ell} > 120$ GeV/c^2 the doubly-resonant diagrams (or t-channel) are essentially the only contribution.
In Figure 4.3 (right) the cross section for $m_{2e2\mu}$ from MCFM v6.1 is shown for three different sets of cuts for the doubly-resonant region only. In the inset the ratio "k" of the NLO and LO predictions is plotted for the "Cut 1" selection. The "k"-factor depends on $m_{4\ell}$ and increases to about 20% at high masses.

POWHEG predictions

In POWHEG BOX a new implementation of the vector-boson pair production process at NLO has been provided. The Z/γ^* interference as well as singly-resonant contributions are properly included, and interference terms arising from identical leptons in the final state are considered.
If the four-lepton-invariant-mass distributions, for the final state 4ℓ and for the three sets of cuts, for POWHEG and for MCFM v6.1 are compared, the two programs show a very good agreement. A full comparison has been also performed between the predictions of both POWHEG at NLO and PYTHIA at LO. This is shown in Figure 4.4 where the distributions are normalised to the the corresponding cross sections. The differences come from the lack of the singly-resonant contribution in PYTHIA and the LO calculations for the doubly-resonant contribution.

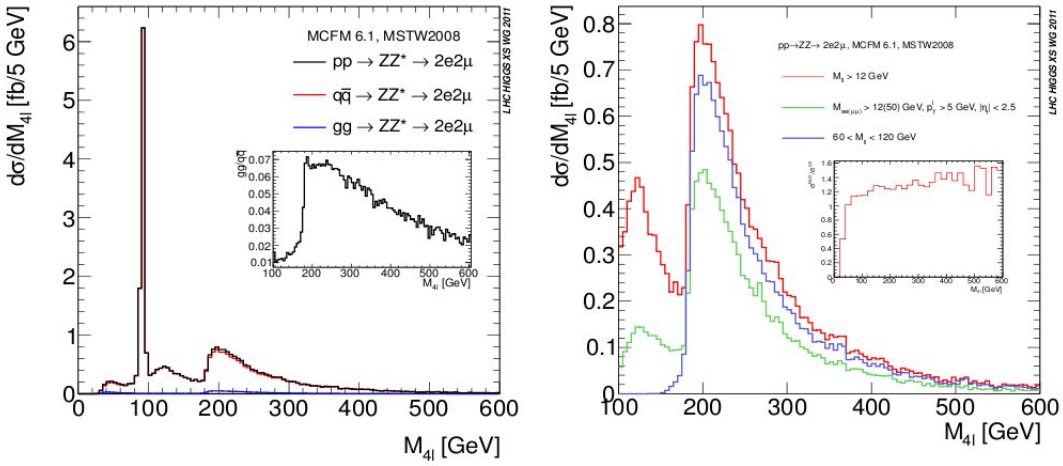

Figure 4.3: *(left) The cross sections for $ZZ^{(*)}$ production as a function of $m_{2e2\mu}$ at 7 TeV from MCFM at NLO; (right) cross sections for $ZZ^{(*)}$ production as a function of $m_{2e2\mu}$ at 7 TeV from MCFM at NLO for three different sets of cuts [39].*

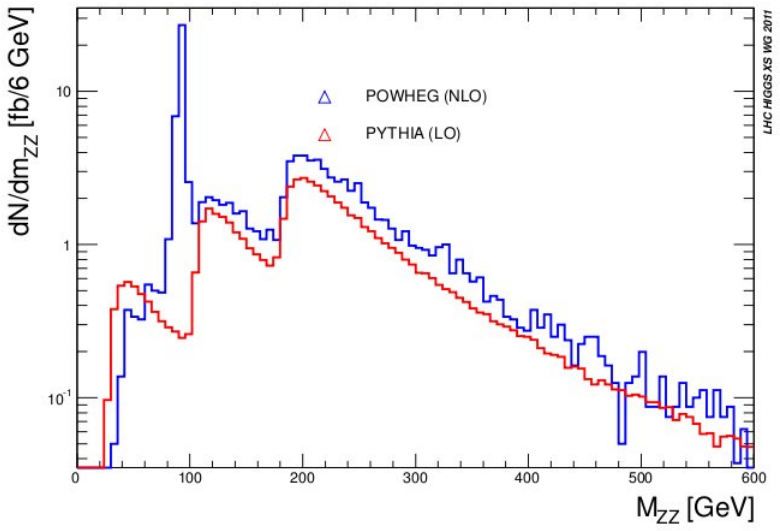

Figure 4.4: *$ZZ \to 4\mu$ invariant-mass distributions at 7 TeV as derived by PYTHIA and POWHEG [39].*

4.1.2 $gg \to ZZ$ generators

The gluon-gluon induced ZZ background, although technically of NNLO compared to the lowest-order Z-pair production, can amount to a substantial fraction of the total irreducible background ($\sim$ 10%). A full NNLO calculation for the ZZ production which would also take these gluon-gluon induced diagrams into account is not available. Therefore the contributions are estimated by using the dedicated tool $gg2ZZ$, which computes $gg \to ZZ$ at LO, which is of order α_s^2, compared to α_s^0 for the LO $q\bar{q} \to ZZ$. The hard scattering $gg \to ZZ^{(*)} \to 4\ell$ events are then showered and hadronized using PYTHIA.
The $gg2ZZ$ tool provides the functionality to compute the cross-section after applying a cut on the minimally generated invariant mass of the same-flavour lepton pairs (which can be interpreted as the Z/γ^* invariant mass), $m_{\ell\ell}^{min} = 12\ GeV/c^2$. This number is computed by using the LO PDF set CTEQ6L1, and the central renormalisation and factorisation scales $\mu_R = \mu_F = m_Z$, where $m_Z = 91.188\ GeV/c^2$ is the nominal Z-boson mass. To estimate the accuracy of this number the renormalisation and factorisation scales were varied in the range $\mu \in [\mu_0/2, 2\mu_0]$; an error of $^{+28\%}_{-20\%}$ is therefore computed. This large uncertainty is expected, since the calculation is only LO, and only at NLO the scale dependencies start to be reduced. Therefore, it is very hard to estimate the accuracy of the convergence of the perturbative series, thus an uncertainty of $\pm 50\%$ on this number is assumed.

4.1.3 Theoretical uncertainties

PDF+α_s and QCD scale uncertainties for $pp \to ZZ \to 2e2\mu$ at NLO and $gg \to ZZ \to 2e2\mu$ are evaluated using MCFM version 6.1. The following cuts are applied to the leptons: $m_{ee} > 12\ GeV/c^2$, $m_{\mu\mu} > 12\ GeV/c^2$, electrons' $p_T > 7\ GeV/c$ and $|\eta| < 2.5$, and muons' $p_T > 5\ GeV/c$ and $|\eta| < 2.4$. No cuts on the minimal ΔR-distance between jets and lepton, and lepton and lepton pairs are applied. The cross sections are calculated inclusively for the number of jets found at NLO.
For the estimation of the PDF+α_s systematic errors, the PDF4LHC prescription is applied. The three PDF sets used are CT10, MSTW08, NNPDF. The four-lepton mass dependent PDF+α_s systematic errors, for both $7\ TeV$ and $8\ TeV$, can be parametrised as follows:

$$ZZ@NLO:\ k(m_{4\ell}) = 1 + 0.0035\sqrt{m_{4\ell}/GeV - 30}, \tag{4.1}$$

$$gg \to ZZ:\ k(m_{4\ell}) = 1 + 0.0066\sqrt{m_{4\ell}/GeV - 10}. \tag{4.2}$$

In Figure 4.5 the difference between the central value of the cross section and the cross section computed with plus and minus 1σ of the total PDF+α_s variation, following the PDF4LHC prescription, is shown with the filled triangle

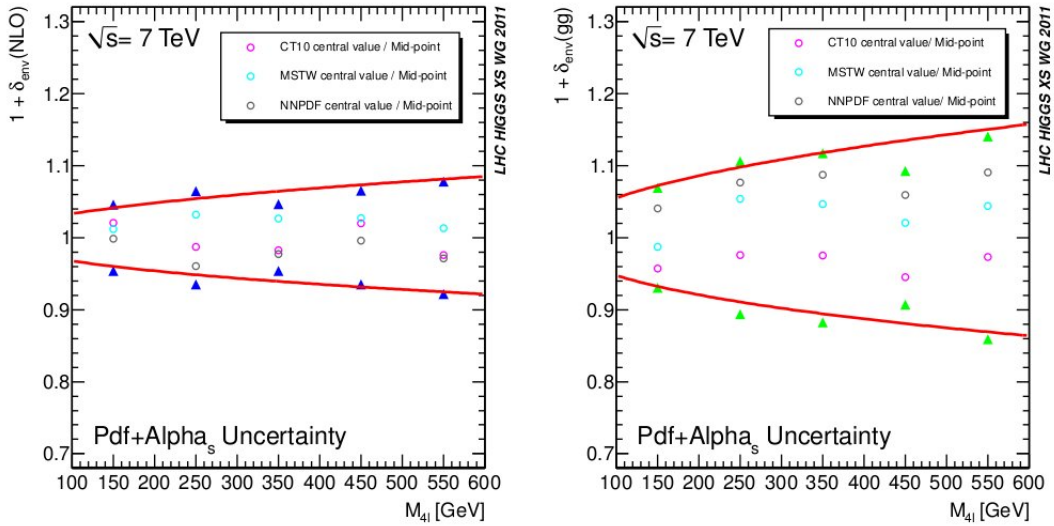

Figure 4.5: *The difference between the central value of the cross section and the cross section computed with 3 different PDF sets and the total PDF+α_s variation (blue markers) varying them by plus and minus 1σ for $q\bar{q} \to ZZ^{(*)} \to 2e2\mu$ (left) and for $gg \to ZZ^{(*)} \to 2e2\mu$ (right) as a function of $m_{2e2\mu}$ at 7 TeV from MCFM [39].*

marker for $q\bar{q} \to ZZ^{(*)} \to 2e2\mu$ (left) and for $gg \to ZZ^{(*)} \to 2e2\mu$ (right) as a function of $m_{2e2\mu}$ at 7 TeV from MCFM. The red line is the parametrisation from Equation 4.1 and Equation 4.2.
For the estimation of QCD scale systematic errors, variations in the differential cross section $d\sigma/dm_{4\ell}$ are calculated and the renormalisation and factorisation scales are changed by a factor of two up and down from their default value $\mu_R = \mu_F = M_Z$. The dependence of the systematic QCD scale errors on the four-lepton invariant mass can be parametrised as follows:

$$ZZ@NLO: \; k(m_{4\ell}) = 1.00 + 0.01\sqrt{(m_{4\ell}/GeV - 20)/13}, \tag{4.3}$$

$$gg \to ZZ: \; k(m_{4\ell}) = 1.04 + 0.10\sqrt{(m_{4\ell}/GeV + 40)/40}. \tag{4.4}$$

In Figure 4.6 the ratio of the cross section computed at the different scale and the cross section computed at the central value of the QCD scale (i.e. at m_Z) is shown as a function of the four-lepton invariant mass. The red lines are the parametrisation from Equation 4.3 and Equation 4.4.

4.1.4 Cross sections for the analysis

In Table 4.1 a summary of the Monte Carlo simulation datasets used for the $H \to ZZ \to 4\ell$ analysis and of ZZ production cross sections obtained from the theory at generator level is presented. In Table 4.2 the ZZ production cross sections used in this thesis work are shown; they have been calculated

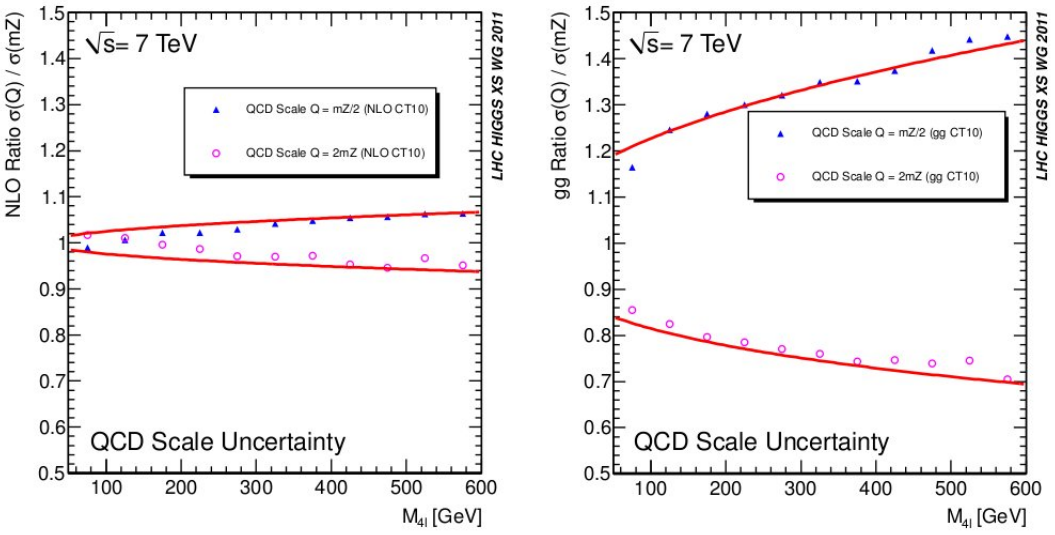

Figure 4.6: *The ratio of the cross sections computed at different scales and the cross section computed at central value of the QCD scale for* $q\bar{q} \to ZZ^{(*)} \to 2e2\mu$ *(left) and for* $gg \to ZZ^{(*)} \to 2e2\mu$ *(right) as a function of* $m_{2e2\mu}$ *at 7 TeV from MCFM [39].*

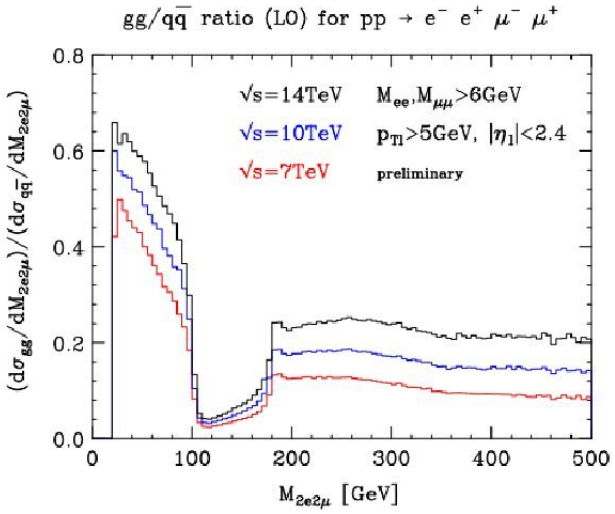

Figure 4.7: $gg \to ZZ$ *contribution relative to* $qq \to ZZ$ *contribution at different centre-of-mass energies.*

Table 4.1: *Monte Carlo simulation datasets used for the irreducible background processes and ZZ production cross sections obtained from the theory at generator level. Z stands for Z, Z^*, γ^*; ℓ means e, μ or τ.*

Process	**MC generator**	$\sigma_{(N)NLO}$ $[fb]$ - 7 TeV	$\sigma_{(N)NLO}$ $[fb]$ - 8 TeV
ZZ continuum			
$q\bar{q} \to ZZ \to 4e(4\mu)(4\tau)$	POWHEG	66.09	76.91
$q\bar{q} \to ZZ \to 2e2\mu$	POWHEG	152	176.7
$q\bar{q} \to ZZ \to 2e(2\mu)2\tau$	POWHEG	152	176.7
$gg \to ZZ \to 2\ell 2\ell'$	gg2ZZ	3.48	12.03
$gg \to ZZ \to 4\ell$	gg2ZZ	1.74	4.8

Table 4.2: *Monte Carlo simulation datasets used for the irreducible background processes and ZZ production cross sections calculated after the reconstruction. σ_{rec} at 7 TeV is calculated for $\mathcal{L} = 5.05$ fb^{-1}, and σ_{rec} at 8 TeV is calculated for $\mathcal{L} = 12.21$ fb^{-1}.*

Process	**MC generator**	σ_{rec} $[fb]$ - 7 TeV	σ_{rec} $[fb]$ - 8 TeV
ZZ continuum			
$q\bar{q} \to ZZ \to 4e$	POWHEG	2.320	2.610
$q\bar{q} \to ZZ \to 4\mu$	POWHEG	3.367	3.741
$q\bar{q} \to ZZ \to 2e2\mu$	POWHEG	5.484	6.262
$gg \to ZZ \to 4e$	gg2ZZ	0.168	0.168
$gg \to ZZ \to 4\mu$	gg2ZZ	0.233	0.228
$gg \to ZZ \to 2e2\mu$	gg2ZZ	0.402	0.506

after the reconstruction. As one can see the gluon-gluon induced ZZ process is $\sim 10\%$ of the total ZZ background. This is evident also in Figure 4.7, which illustrates the $gg \to ZZ$ background predicted contribution relative to the $qq \to ZZ$ process at leading order at different centre-of-mass energies. Furthermore, it can be noted that the $gg \to ZZ$ fraction increases with $\sqrt{s}$, up to $\sim 20\%$ at 14 TeV (LHC design centre-of-mass energy).

4.2 Motivations and aim of the work

The goal of this thesis is to develop a new technique to measure the ratio of gg over $q\bar{q}$ induced ZZ production cross sections. In fact, in the current $H \to ZZ \to 4\ell$ analysis, in the P_{bkg} definition in MELA only $q\bar{q} \to ZZ$ is considered. Because of this approximation a quite high systematic uncertainty is present and MELA is suboptimal in the separation between

$gg \to ZZ$ and other processes (signal or backgrounds). Such a measurement is thus important for two reasons: on the one hand, to know the exact fraction of $gg \to ZZ$ allows to reduce systematic uncertainties on $H \to ZZ$, on the other hand this is an interesting SM measurement per se.

4.3 Study of $q\bar{q} \to ZZ$ and $gg \to ZZ$ angular distributions

As seen in Section 3.6, the angular distributions analysis allows to distinguish between Higgs signal and background. It might be thought to apply the same reasoning to separate the two irreducible backgrounds, $q\bar{q} \to ZZ$ and $gg \to ZZ$. Therefore, it is firstly necessary to do an accurate study of $q\bar{q} \to ZZ$ and $gg \to ZZ$ angular distributions, both at generator and at reconstruction level, in order to see if there is the possibility of developing a specific kinematic discriminant following the MELA example.
This study is performed above 180 GeV/c^2. As a matter of fact, below $2m_Z$ threshold the dominant background is $q\bar{q} \to ZZ$ with a secondary contribution of $Z + X$ background, whereas above $2m_Z$ threshold the dominant background is also $q\bar{q} \to ZZ$ with a secondary contribution of $gg \to ZZ$ background.

4.3.1 Angular distributions at generator level

Monte Carlo samples used for the study at generator level are the same of the $H \to ZZ \to 4\ell$ analysis, for centre-of-mass energies of 7 TeV and 8 TeV, described in Section 3.2.2 and in Table 4.1. Only samples $gg \to ZZ \to 2e2\mu$ and $q\bar{q} \to ZZ \to 2e2\mu$ are used, since there is substantially no difference between the three channels $2e2\mu$, $4e$ and 4μ (see Section 4.3.2), except for interference effects which are the same for $gg \to ZZ$ and $q\bar{q} \to ZZ$.
The generator level study is performed for the following set of cuts:

- $m_{Z_1} > 40\ GeV/c^2$,
- $m_{Z_2} > 12\ GeV/c^2$,
- $m_{ZZ} > 180\ GeV/c^2$,

where Z_1 is the lepton pair with an invariant mass closest to the nominal Z mass. These are the same cuts on $m_{Z_{1,2}}$ and the same choice of the Z_1 applied at the reconstruction level; in this way it is easy to compare generator and reconstruction level in order to see if the detector acceptance and resolution affect the shape of the angular distributions.
In Figures 4.8, 4.9 and 4.10 distributions for $gg \to ZZ$ and $q\bar{q} \to ZZ$, at a centre-of-mass energy of 7 TeV and for the seven observables which describe the $X \to ZZ \to 4\ell$ decay ($m_{Z_1}, m_{Z_2}, \cos\theta_1, \cos\theta_2, \cos\theta^*, \Phi, \Phi_1$), are shown.

The gap at $\sim 91\ GeV/c^2$ in Z_2 mass distribution is due to the choice of the Z_1 as the one with an invariant mass closest to the nominal Z mass. With

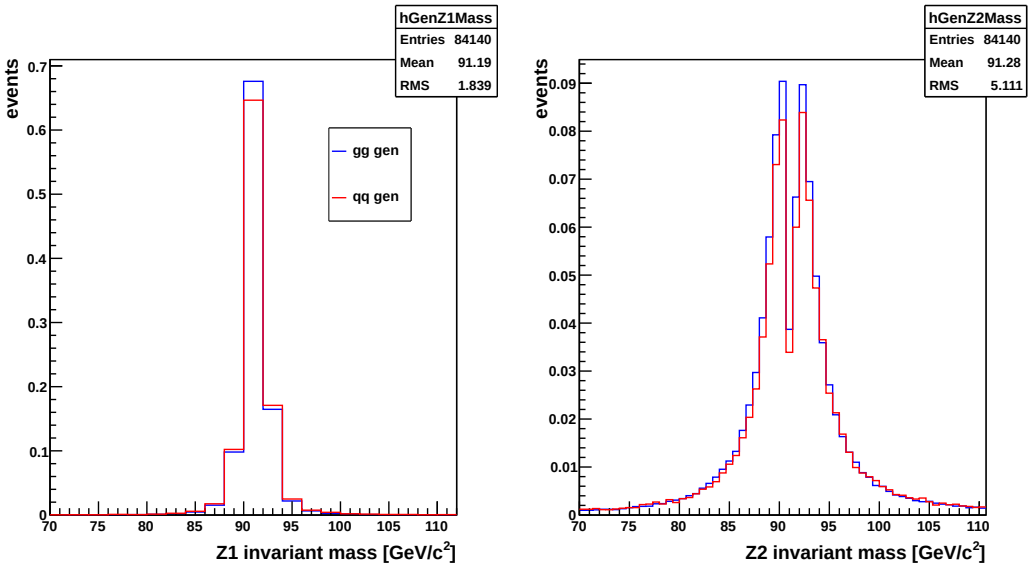

Figure 4.8: *Z_1 and Z_2 invariant mass distributions at generator level for $gg \to ZZ \to 2e2\mu$ and $q\bar{q} \to ZZ \to 2e2\mu$ processes at a centre-of-mass energy of 7 TeV.*

regard to the angular distributions, one can note that the main differences between $gg \to ZZ$ and $q\bar{q} \to ZZ$ processes arise in the angles $\cos\theta_1$ and $\cos\theta_2$.
The same plots are done also for a centre-of-mass energy of 8 TeV, to check if the angular distributions shapes change with the energy. Figures 4.11, 4.12 and 4.13 show the seven observables distributions for $gg \to ZZ$ and $q\bar{q} \to ZZ$ at a centre-of-mass energy of 8 TeV. It is apparent that at a centre-of-mass energy of 8 TeV there are no substantial changes in the distributions shapes with respect to an energy of 7 TeV: the separation power is essentially the same for the two energies.

4.3.2 Angular distributions at reconstruction level

In the study at reconstruction level, performed above $2m_Z$ threshold, the "$H \to ZZ \to 4\ell$ analysis" selection, described in Chapter 3, is applied. The following set of cuts is used:

- $40 < m_{Z_1} < 120\ GeV/c^2$,
- $12 < m_{Z_2} < 120\ GeV/c^2$,

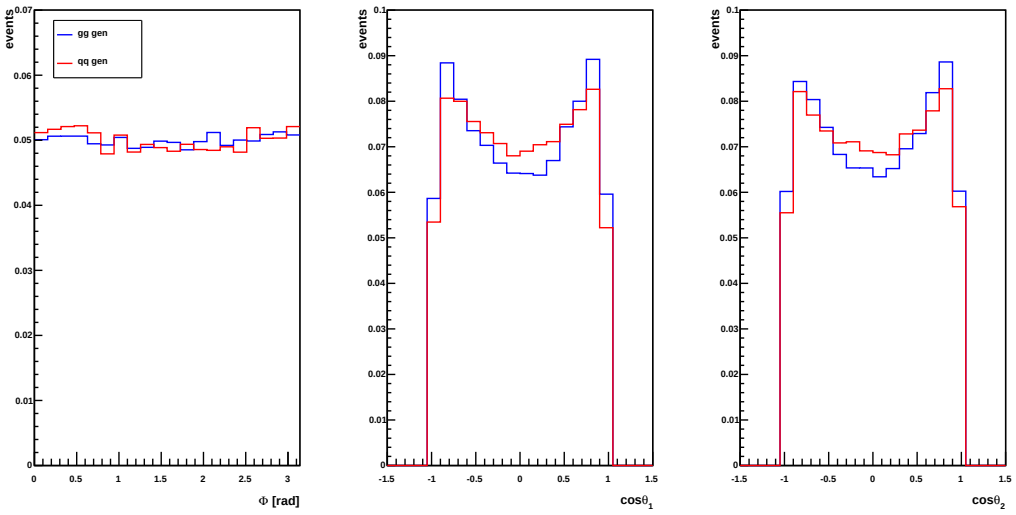

Figure 4.9: Φ, $\cos\theta_1$ *and* $\cos\theta_2$ *distributions at generator level for* $gg \to ZZ \to 2e2\mu$ *and* $q\bar{q} \to ZZ \to 2e2\mu$ *processes at a centre-of-mass energy of* 7 TeV.

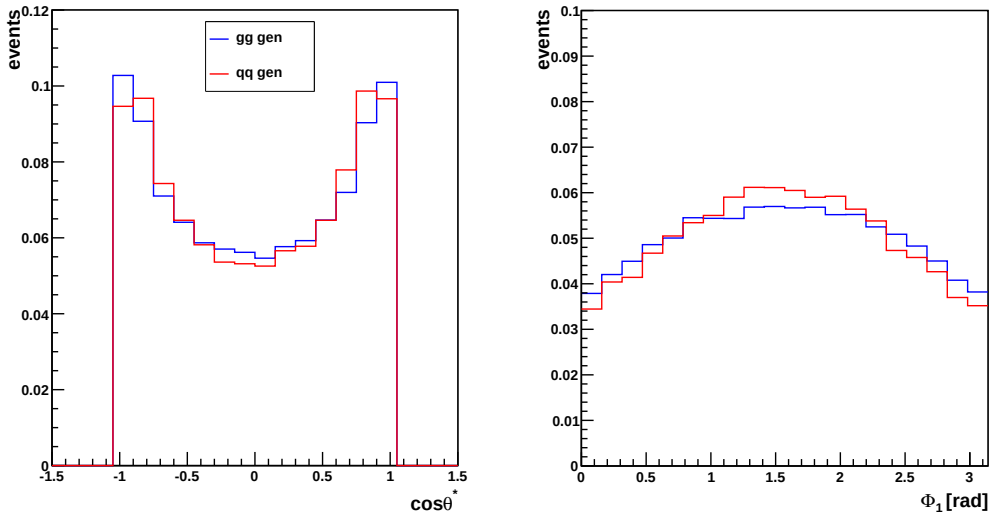

Figure 4.10: $\cos\theta^*$ *and* Φ_1 *distributions at generator level for* $gg \to ZZ \to 2e2\mu$ *and* $q\bar{q} \to ZZ \to 2e2\mu$ *processes at a centre-of-mass energy of* 7 TeV.

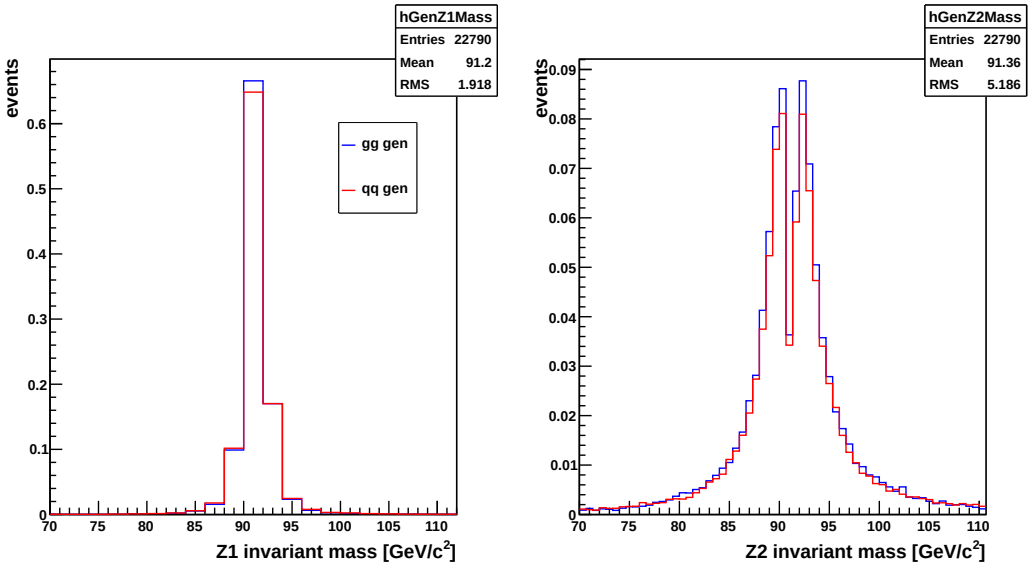

Figure 4.11: Z_1 *and* Z_2 *invariant mass distributions at generator level for* $gg \to ZZ \to 2e2\mu$ *and* $q\bar{q} \to ZZ \to 2e2\mu$ *processes at a centre-of-mass energy of* 8 TeV.

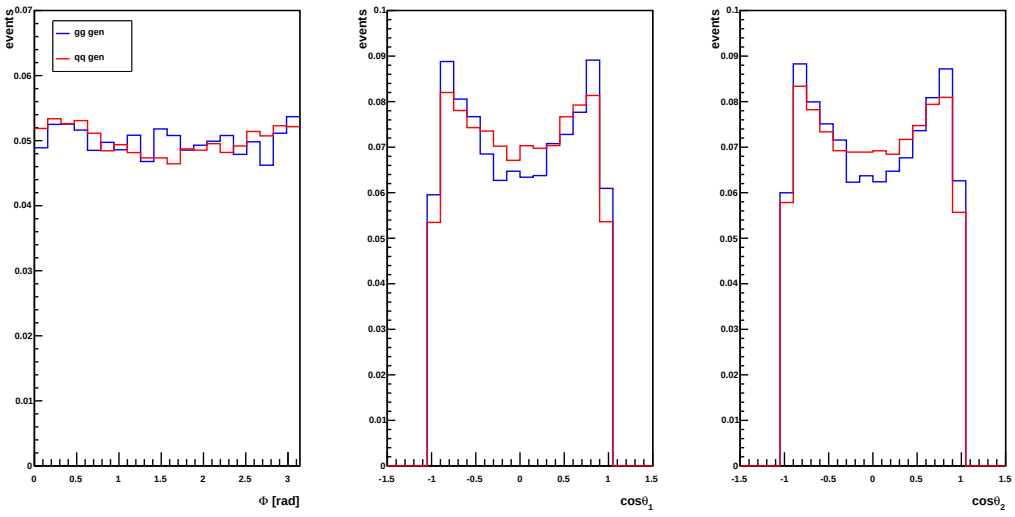

Figure 4.12: Φ, $\cos\theta_1$ *and* $\cos\theta_2$ *distributions at generator level for* $gg \to ZZ \to 2e2\mu$ *and* $q\bar{q} \to ZZ \to 2e2\mu$ *processes at a centre-of-mass energy of* 8 TeV.

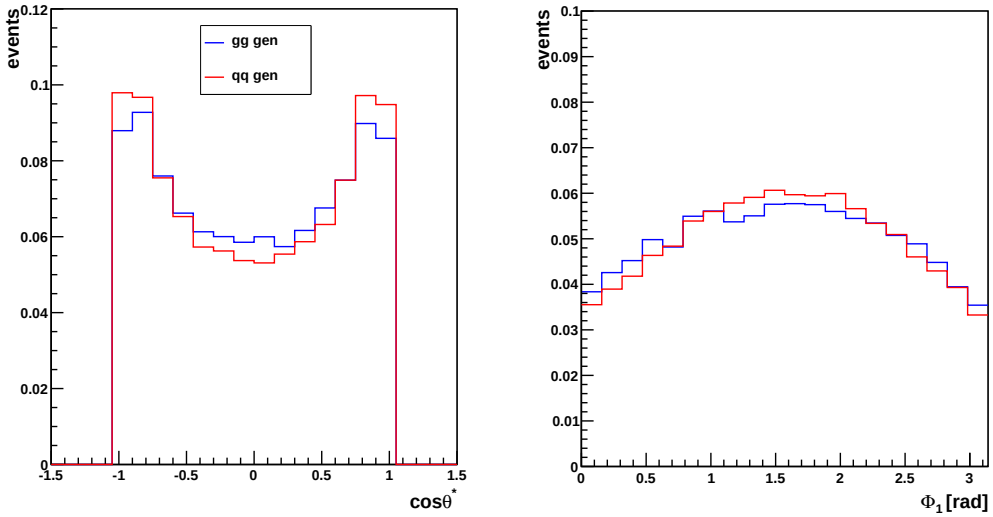

Figure 4.13: $\cos\theta^*$ *and* Φ_1 *distributions at generator level for* $gg \to ZZ \to 2e2\mu$ *and* $q\bar{q} \to ZZ \to 2e2\mu$ *processes at a centre-of-mass energy of* 8 TeV.

- $m_{4\ell} > 100\ GeV/c^2$,
- $p_T^e > 7\ GeV/c$ and $|\eta^e| < 2.5$,
- $p_T^\mu > 5\ GeV/c$ and $|\eta^\mu| < 2.4$,
- $SIP_{3D} < 4$ and $R_{Iso}^\ell < 0.4$.

The study at reconstruction level illustrates the effect of resolution and detector effects on production and decay angles.
As an example, in Figures 4.14 and 4.15 the ratio between reconstructed and generated histograms, for $q\bar{q} \to ZZ$ at a centre-of-mass energy of 7 TeV, is presented. As one can note, the distributions are not perfectly flat, since acceptance effects are present, and $\sim 80\%$ of generated events is reconstructed. It has been verified that the same behaviour is present for $q\bar{q} \to ZZ$ at the energy of 8 TeV and for $gg \to ZZ$ at 7 and 8 TeV.
In Figures 4.16, 4.17 and 4.18 distributions for $gg \to ZZ$ and $q\bar{q} \to ZZ$ at reconstruction level, for the seven observables which describe the $X \to ZZ \to 4\ell$ decay and at a centre-of-mass energy of 7 TeV, are shown. The same plots are presented in Figures 4.19, 4.20 and 4.21 for an energy of 8 TeV.
In Figures 4.16 and 4.19 (right) it is evident that also in this case there is a gap at $\sim 91\ GeV/c^2$, due to the choice of the Z_1 as the one with an invariant mass closest to the nominal Z mass. As expected, the reconstructed shapes being very close to the generated ones, there is not much difference between

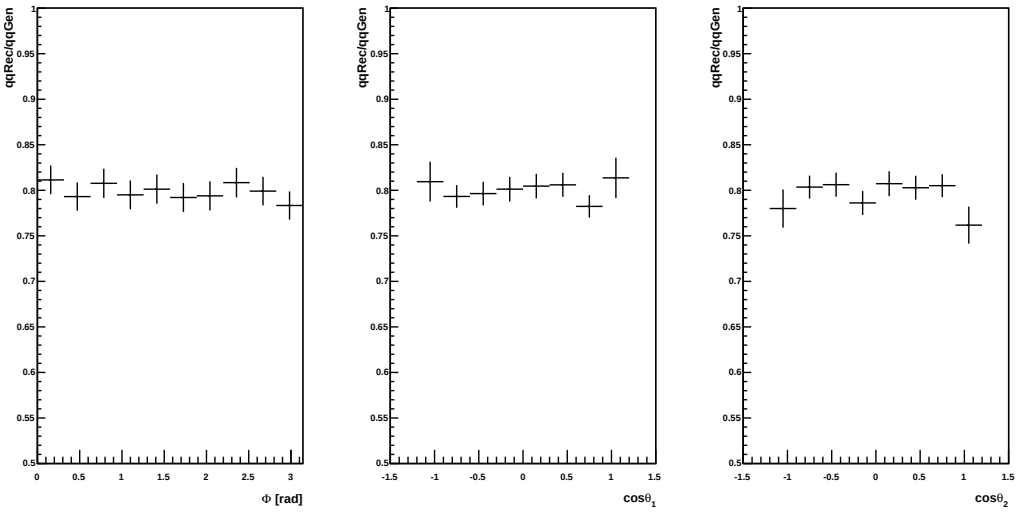

Figure 4.14: *Ratio between reconstructed and generated distributions for* Φ, $\cos\theta_1$ *and* $\cos\theta_2$ *for* $q\bar{q} \to ZZ \to 2e2\mu$ *process at a centre-of-mass energy of* $7\ TeV$.

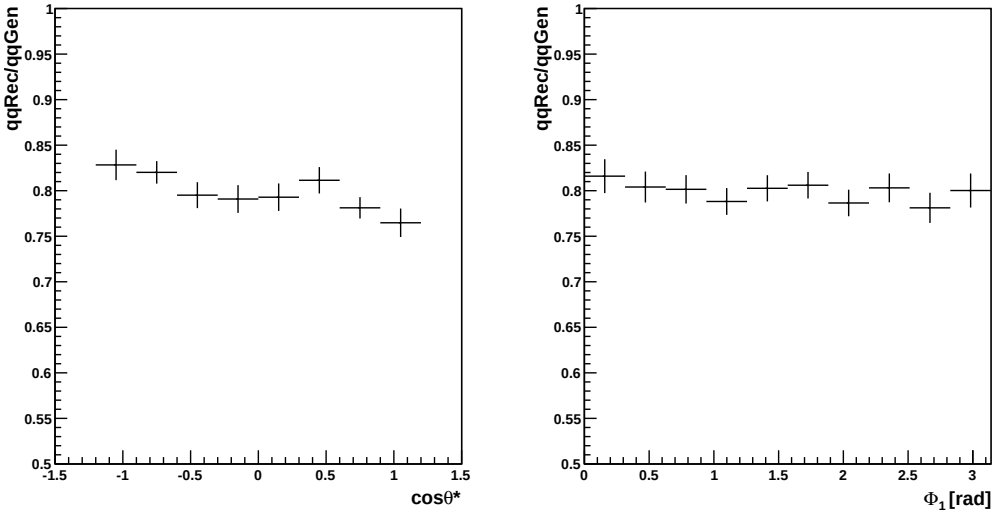

Figure 4.15: *Ratio between reconstructed and generated distributions for* $\cos\theta^*$ *and* Φ_1 *for* $q\bar{q} \to ZZ \to 2e2\mu$ *process at a centre-of-mass energy of* $7\ TeV$.

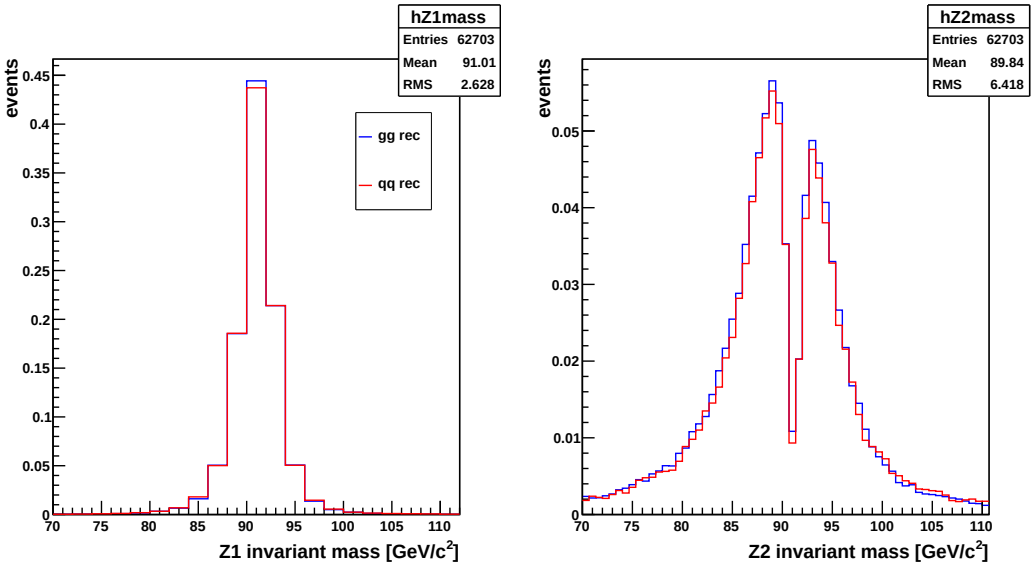

Figure 4.16: *Z_1 and Z_2 invariant mass distributions at reconstruction level for $gg \to ZZ \to 2e2\mu$ and $q\bar{q} \to ZZ \to 2e2\mu$ processes at a centre-of-mass energy of 7 TeV.*

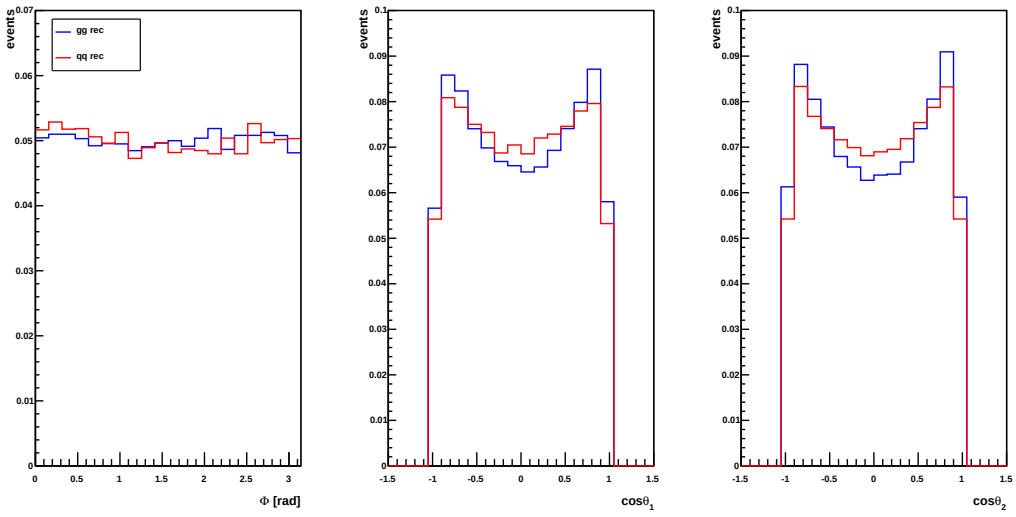

Figure 4.17: *Φ, $\cos\theta_1$ and $\cos\theta_2$ distributions at reconstruction level for $gg \to ZZ \to 2e2\mu$ and $q\bar{q} \to ZZ \to 2e2\mu$ processes at a centre-of-mass energy of 7 TeV.*

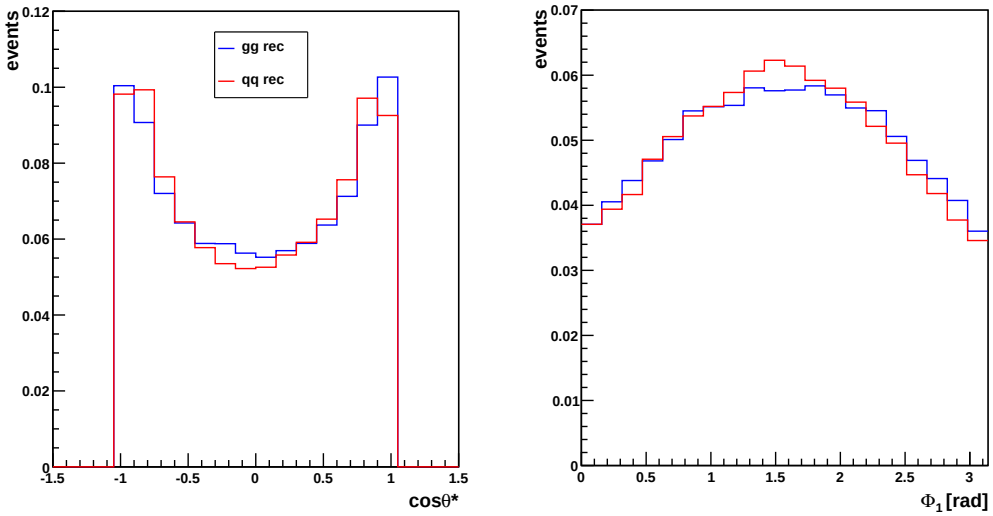

Figure 4.18: $\cos\theta^*$ and Φ_1 distributions at reconstruction level for $gg \to ZZ \to 2e2\mu$ and $q\bar{q} \to ZZ \to 2e2\mu$ processes at a centre-of-mass energy of 7 TeV.

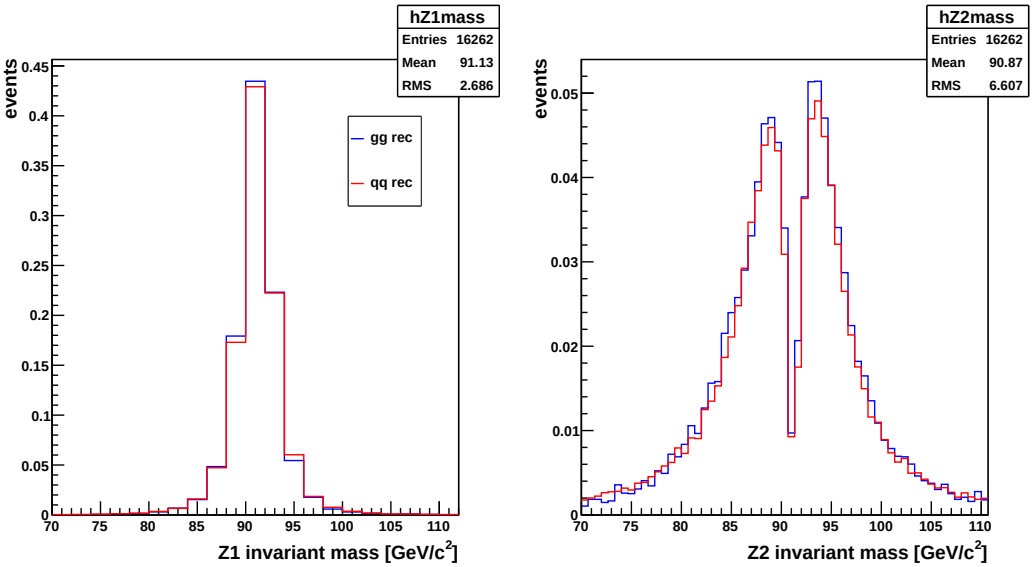

Figure 4.19: Z_1 and Z_2 invariant mass distributions at reconstruction level for $gg \to ZZ \to 2e2\mu$ and $q\bar{q} \to ZZ \to 2e2\mu$ processes at a centre-of-mass energy of 8 TeV.

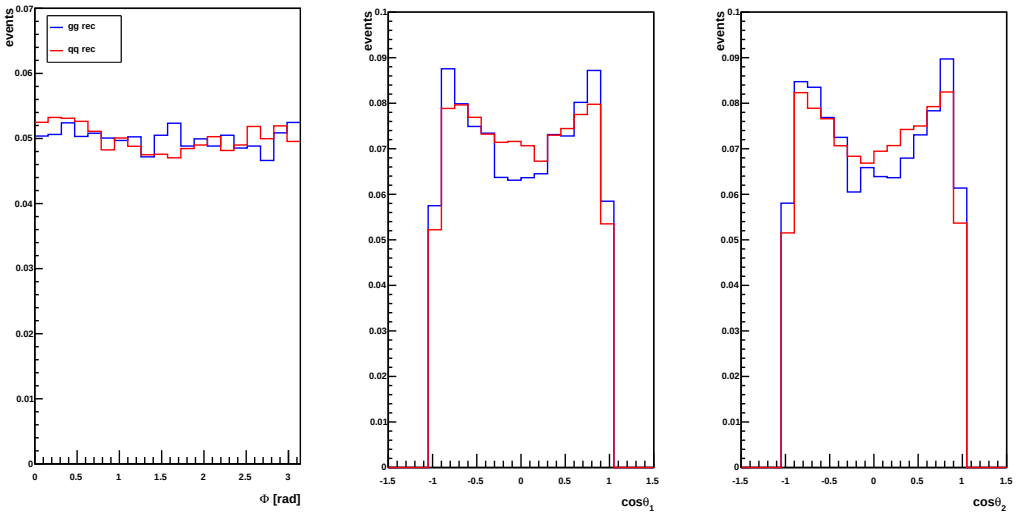

Figure 4.20: Φ, $\cos\theta_1$ *and* $\cos\theta_2$ *distributions at reconstruction level for* $gg \to ZZ \to 2e2\mu$ *and* $q\bar{q} \to ZZ \to 2e2\mu$ *processes at a centre-of-mass energy of* 8 TeV.

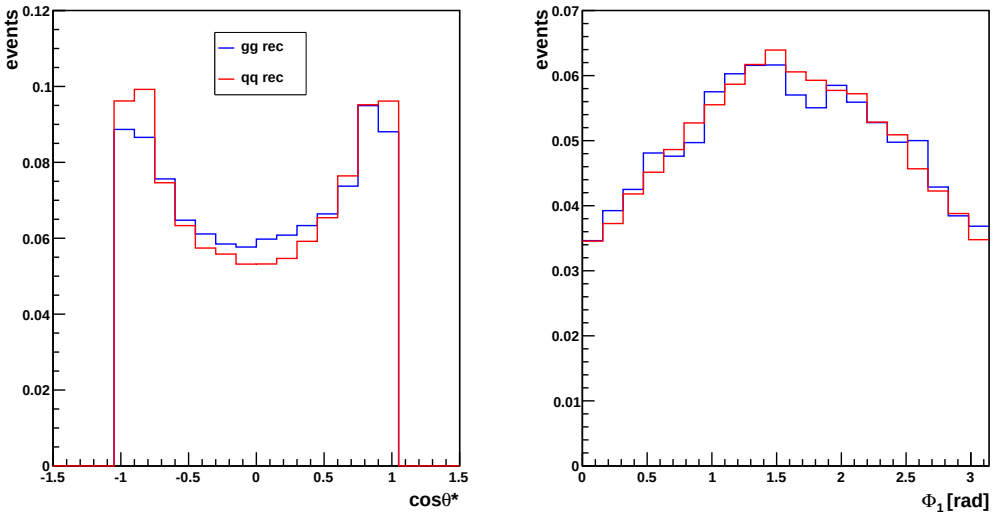

Figure 4.21: $\cos\theta^*$ *and* Φ_1 *distributions at reconstruction level for* $gg \to ZZ \to 2e2\mu$ *and* $q\bar{q} \to ZZ \to 2e2\mu$ *processes at a centre-of-mass energy of* 8 TeV.

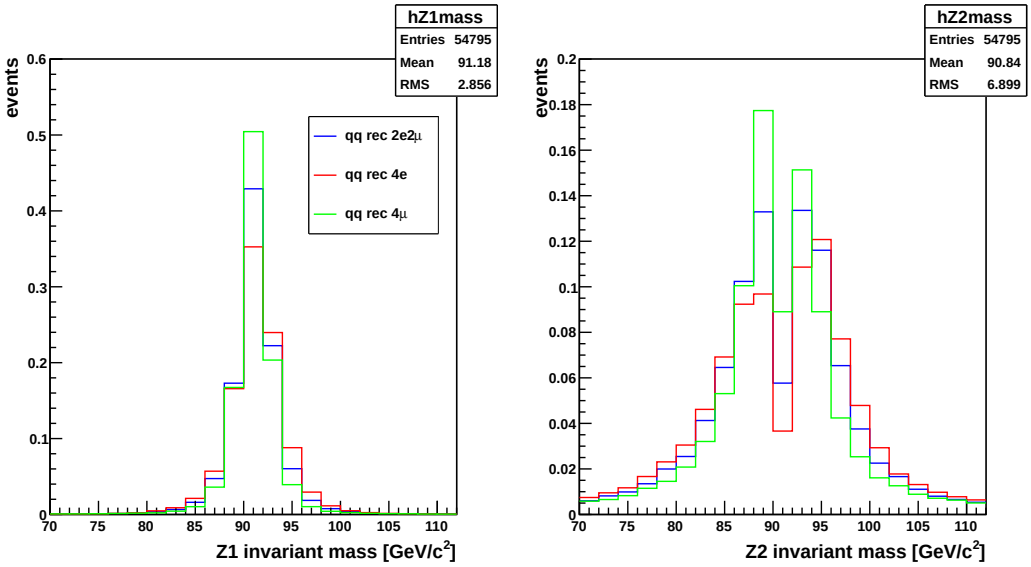

Figure 4.22: *Z_1 and Z_2 invariant mass distributions at reconstruction level for $q\bar{q} \to ZZ \to 2e2\mu(4e)(4\mu)$ processes at a centre-of-mass energy of 8 TeV.*

these plots and those in Section 4.3.1. The separation between $gg \to ZZ$ and $q\bar{q} \to ZZ$ is substantially the same with respect to the generator level and it is more enhanced in $\cos\theta_1$ and $\cos\theta_2$ distributions.

For the studies both at generator and at reconstruction level only the results for $gg \to ZZ \to 2e2\mu$ and for $q\bar{q} \to ZZ \to 2e2\mu$ are presented here. In fact the same study has been performed for $4e$ and 4μ channels, finding angular distributions very similar to those of $2e2\mu$ channel. This is shown in Figures 4.22, 4.23 and 4.24, where angular distributions for the three channels $2e2\mu, 4e$ and 4μ, at reconstruction level and for a centre-of-mass energy of 8 TeV, are reported. As one can see, distributions for the three different channels are similar, the main difference being the Z mass resolution, thus in the second part of this analysis (see Chapter 5) the $2e2\mu, 4e$ and 4μ channels will be merged together, in order to have a larger dataset.

4.3.3 $gg \to ZZ$ and $q\bar{q} \to ZZ$ likelihood distributions with MELA

Before starting with the construction of a new discriminant, it is interesting to verify if MELA has some separation power between the two irreducible backgrounds. In fact, if MELA worked good with $gg \to ZZ$ and $q\bar{q} \to ZZ$ processes, a new discriminant would not be necessary.
The results are shown in Figure 4.25, where the MELA distributions for the

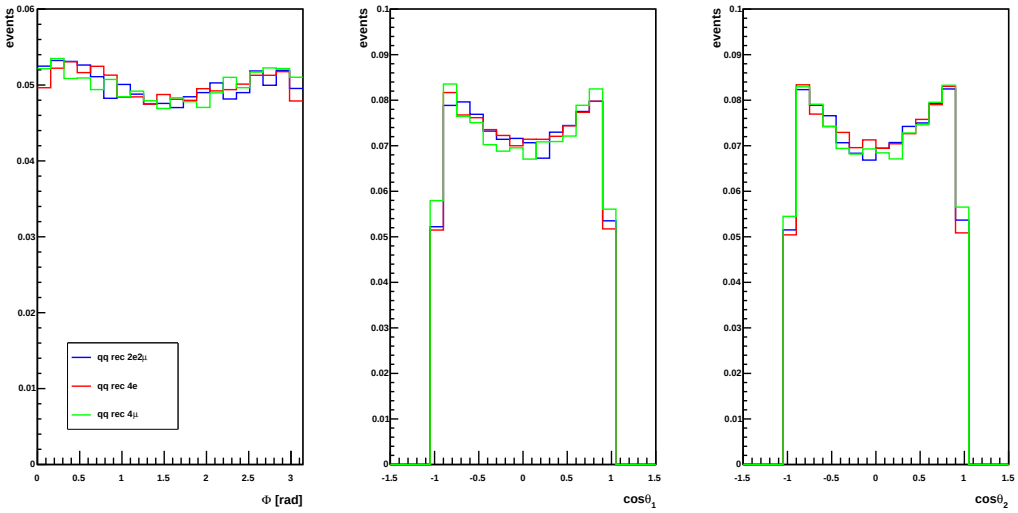

Figure 4.23: Φ, $\cos\theta_1$ *and* $\cos\theta_2$ *distributions at reconstruction level for* $q\bar{q} \to ZZ \to 2e2\mu(4e)(4\mu)$ *processes at a centre-of-mass energy of* $8\ TeV$.

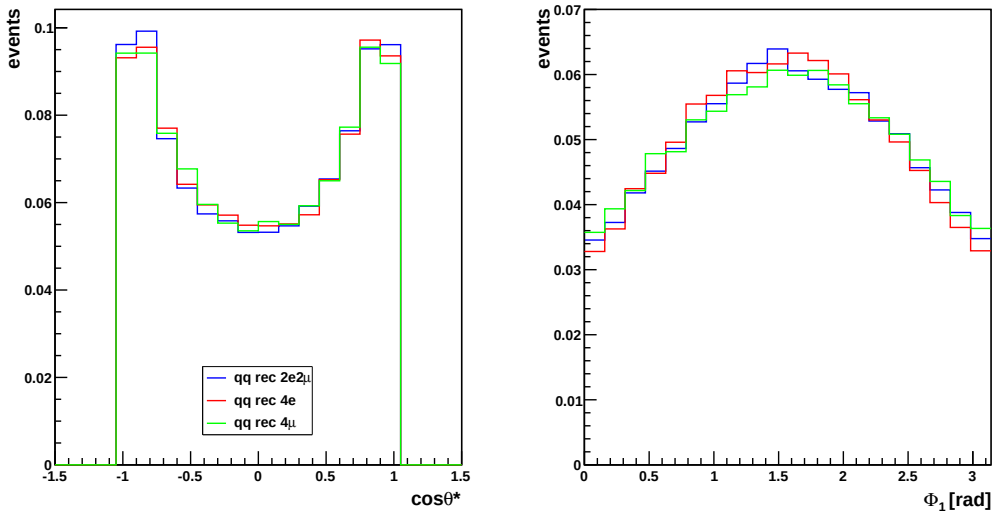

Figure 4.24: $\cos\theta^*$ *and* Φ_1 *distributions at reconstruction level for* $q\bar{q} \to ZZ \to 2e2\mu(4e)(4\mu)$ *processes at a centre-of-mass energy of* $8\ TeV$.

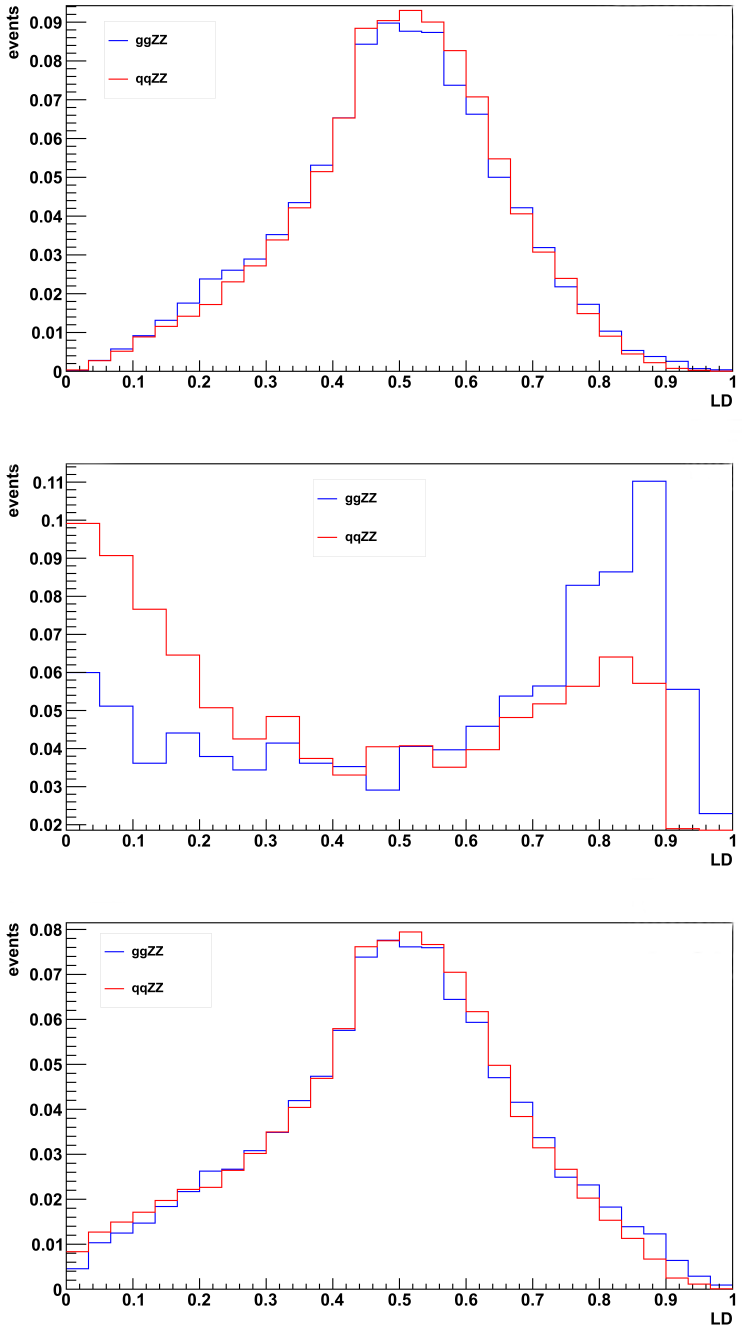

Figure 4.25: *Discriminant variable distributions obtained from MELA, both for $gg \to ZZ$ and $q\bar{q} \to ZZ$, at a centre-of-mass energy of 8 TeV and in three different mass ranges: $180 < m_{ZZ} < 300$ GeV/c^2 (top), $400 < m_{ZZ} < 700$ GeV/c^2 (middle), and $180 < m_{ZZ} < 700$ GeV/c^2 (bottom).*

two processes are presented in three mass ranges: $180 < m_{ZZ} < 300\ GeV/c^2$, $400 < m_{ZZ} < 700\ GeV/c^2$ and $180 < m_{ZZ} < 700\ GeV/c^2$ (full range). It is apparent that the discriminants for $gg \to ZZ$ and $q\bar{q} \to ZZ$ in the full mass range are very similar. In particular, they substantially overlap in the region where there are more events, at $m_{ZZ} \sim 200\ GeV/c^2$, and start to separate only for $m_{ZZ} > 400\ GeV/c^2$. Similar results have been found at a centre-of-mass energy of $7\ TeV$. Thus, the implementation of a new discriminant seems to be necessary in order to distinguish $gg \to ZZ$ from $q\bar{q} \to ZZ$. A further indication of the necessity of a new discriminant comes from the comparison between $gg \to ZZ$ and Higgs angular distributions. In fact, since the hypothetical new discriminant would be constructed in the same way as MELA (see Equations 3.1 and 5.1), but replacing P_{sig} with P_{gg}, if the angular distributions of the Higgs signal and $gg \to ZZ$ was very similar, also MELA and the new discriminant would be close. In Figures 4.26 and 4.27 angular distributions for $gg \to ZZ$ and an Higgs signal with $m_H = 200\ GeV/c^2$, at a centre-of-mass energy of $8\ TeV$, are shown. As one

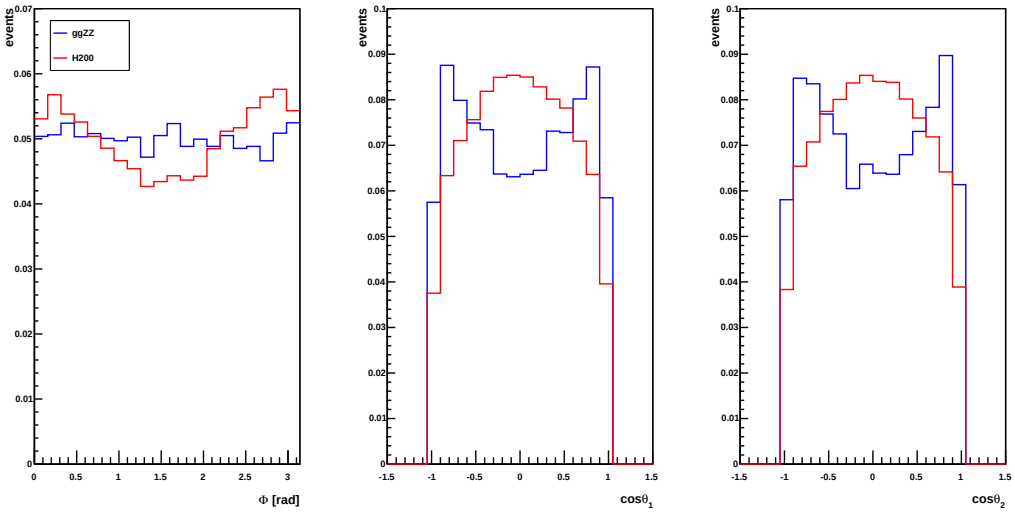

Figure 4.26: Φ, $\cos\theta_1$ *and* $\cos\theta_2$ *distributions at reconstruction level for* $gg \to ZZ$ *and an Higgs signal with* $m_H = 200\ GeV/c^2$ *at a centre-of-mass energy of* $8\ TeV$.

can see, the distributions are very different, in particular in $\cos\theta_1$, $\cos\theta_2$ and $\cos\theta^*$; as a consequence the new discriminant will be different from MELA.

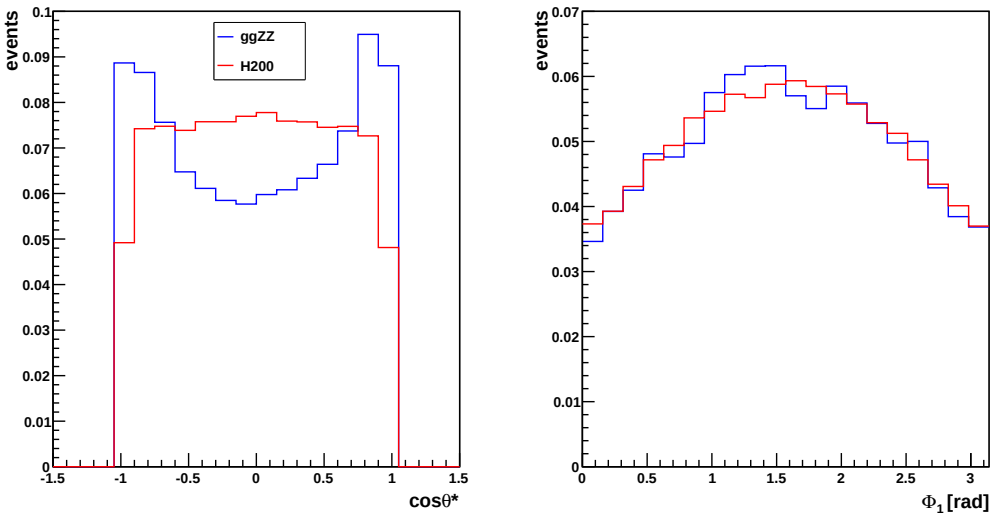

Figure 4.27: $\cos\theta^*$ and Φ_1 distributions at reconstruction level for $gg \to ZZ$ and an Higgs signal with $m_H = 200\ GeV/c^2$ at a centre-of-mass energy of $8\ TeV$.

Chapter 5

New method for ratio of gg over $q\bar{q}$ induced ZZ production cross sections measurement

As has already been said in Section 4.3.3, from the comparison between the discriminant variable distributions obtained with MELA approach for $gg \to ZZ$ and for $q\bar{q} \to ZZ$ processes it turns out that the two likelihoods are very close. In particular, they are nearly the same in the mass range $180 < m_{ZZ} < 300\ GeV/c^2$, where there is the maximum number of events, while they start to separate only above $400\ GeV/c^2$.
In order to exploit in the best possible way the little differences present between $gg \to ZZ$ and $q\bar{q} \to ZZ$ angular distributions it is thus necessary to define a new discriminant apt to distinguish the two irreducible backgrounds.
As said at the end of Section 4.3.2, from now on the three channels $2e2\mu, 4e$ and 4μ will be merged together, since their angular distributions are similar.

5.1 A new kinematic discriminant: ggMELA

5.1.1 Construction of the ggMELA discriminant

In the second part of this thesis a new kinematic discriminant is defined (ggMELA), following the MELA example. The seven mass and angular observables used in the ggMELA construction are the same included in MELA: the two Z invariant masses, m_1 and m_2, and the five production and decay angles, $\vec{\Omega} = \theta^*, \Phi_1, \theta_1, \theta_2, \Phi$.
Construction of the kinematic discriminant ggKD in the ggMELA approach relies on probability for an event with a set of observables $(m_{4\ell}, m_1, m_2, \vec{\Omega})$ to come either from $gg \to ZZ$ or $q\bar{q} \to ZZ$

$$P_{gg}(m_1, m_2, \vec{\Omega}|m_{4\ell})$$

$$P_{q\bar{q}}(m_1, m_2, \vec{\Omega}|m_{4\ell})$$

where probabilities are normalized with respect to the seven observables and $m_{4\ell}$ is treated as a conditional parameter. These probabilities are calculated analytically for $q\bar{q} \to ZZ$ and with a template for $gg \to ZZ$ (see Figures 5.1 and 5.2). Then the discriminant is constructed as follows

$$ggKD = \frac{P_{gg}}{P_{gg} + P_{q\bar{q}}} = \left[1 + \frac{P_{q\bar{q}}(m_1, m_2, \vec{\Omega}|m_{4\ell})}{P_{gg}(m_1, m_2, \vec{\Omega}|m_{4\ell})}\right]^{-1} \tag{5.1}$$

In Figures 5.1 and 5.2 2D templates for the five angles as a function of m_{ZZ}, for $gg \to ZZ$ at a centre-of-mass energy of 7 TeV, are shown as an example.

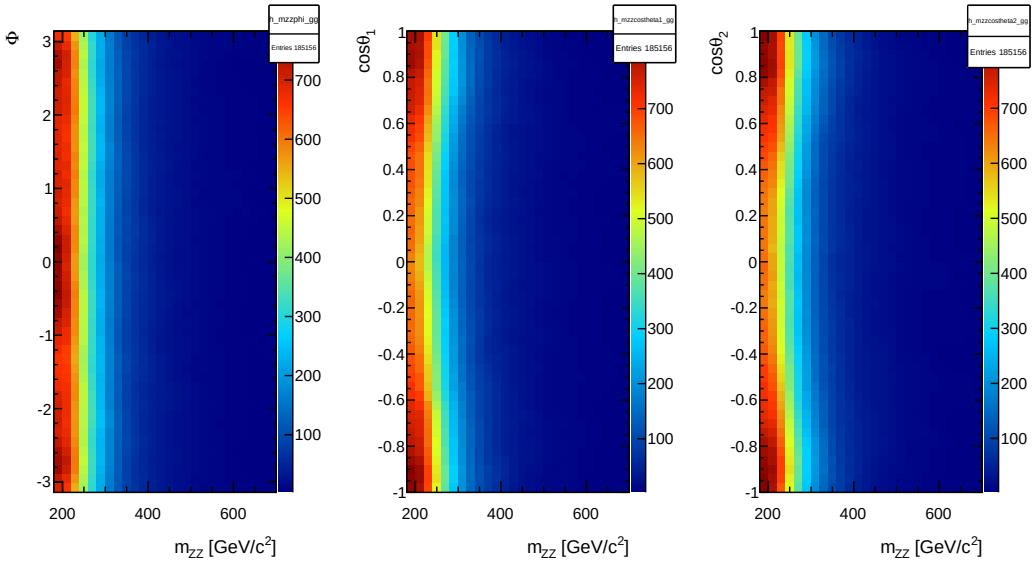

Figure 5.1: Φ, $\cos\theta_1$ *and* $\cos\theta_2$ *distributions as a function of* m_{ZZ} *at reconstruction level for* $gg \to ZZ$ *process at a centre-of-mass energy of* 7 TeV. *The three channels* $2e2\mu, 4e$ *and* 4μ *are merged together.*

5.1.2 ggMELA likelihoods

From ggMELA a discriminant variable distribution both for $gg \to ZZ$ and $q\bar{q} \to ZZ$ is obtained. In Figures 5.3 and 5.4 the two likelihoods for $gg \to ZZ$ and $q\bar{q} \to ZZ$ are reported for 7 and 8 TeV, in three different mass ranges: $180 < m_{ZZ} < 300$ GeV/c^2, $300 < m_{ZZ} < 500$ GeV/c^2 and $500 < m_{ZZ} < 700$ GeV/c^2. From these plots it is evident that the discrimination effect increases with m_{ZZ}.
Furthermore, in Figure 5.5 the same discriminant distributions are shown in the full mass range $180 < m_{ZZ} < 700$ GeV/c^2 for a centre-of-mass energy of

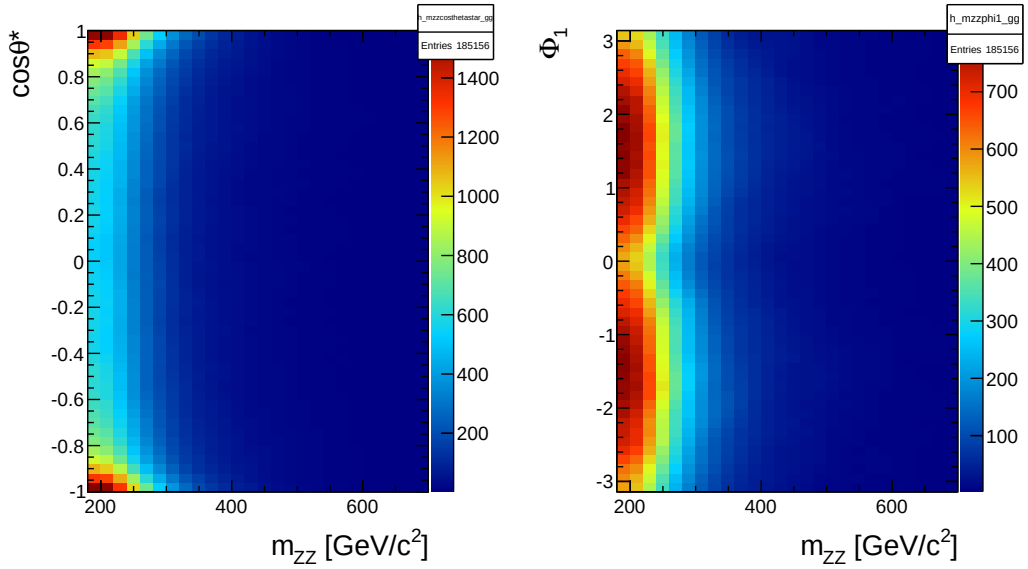

Figure 5.2: $\cos\theta^*$ *and* Φ_1 *distributions as a function of* m_{ZZ} *at reconstruction level for* $gg \to ZZ$ *process at a centre-of-mass energy of* 7 TeV. *The three channels* $2e2\mu, 4e$ *and* 4μ *are merged together.*

7 (top) and 8 TeV (bottom). As one can see, the separation power between the two irreducible backgrounds is very similar at the two energies.
In Figure 5.6 discriminant distributions for data and MC samples at a centre-of-mass energy of 8 TeV and for a luminosity of 12.2 fb^{-1} are compared. The red and green (stacked) histograms represent the MC simulations of $gg \to ZZ$ and $q\bar{q} \to ZZ$ backgrounds, points are data with statistical errors. As one can see, good agreement is found between data and MC.
In Figures 5.3 and 5.4 is apparent that $gg \to ZZ$ and $q\bar{q} \to ZZ$ likelihoods are separated. A more precise quantification of the discriminant power of ggMELA can be done with a ROC curve.
The ROC curve is constructed calculating the efficiency for $gg \to ZZ$ (ϵ_{gg}) and $q\bar{q} \to ZZ$ ($\epsilon_{q\bar{q}}$) as a function of a cut on the likelihood. In case of no separation, the ROC curve is simply a diagonal line. Comparing the ROC curve for MELA and ggMELA on the two irreducible backgrounds (see Figure 5.7), it can be seen that ggMELA has a better separation power than MELA, as expected. Furthermore, it is visible that ggMELA has a slightly better discriminant power at 8 TeV.
These ROC curves are in the full mass range $180 < m_{ZZ} < 700\ GeV/c^2$. In a limited range $300 < m_{ZZ} < 700\ GeV/c^2$ one could obtain better ROC curves, but losing most part of the events. The separation power becomes very good at high masses, so in this case it is less evident.
As one can see in Figures 5.3 and 5.4, for values of ggMELA > 0.7 the $q\bar{q} \to ZZ$ discriminant variable distribution is suppressed with respect to the

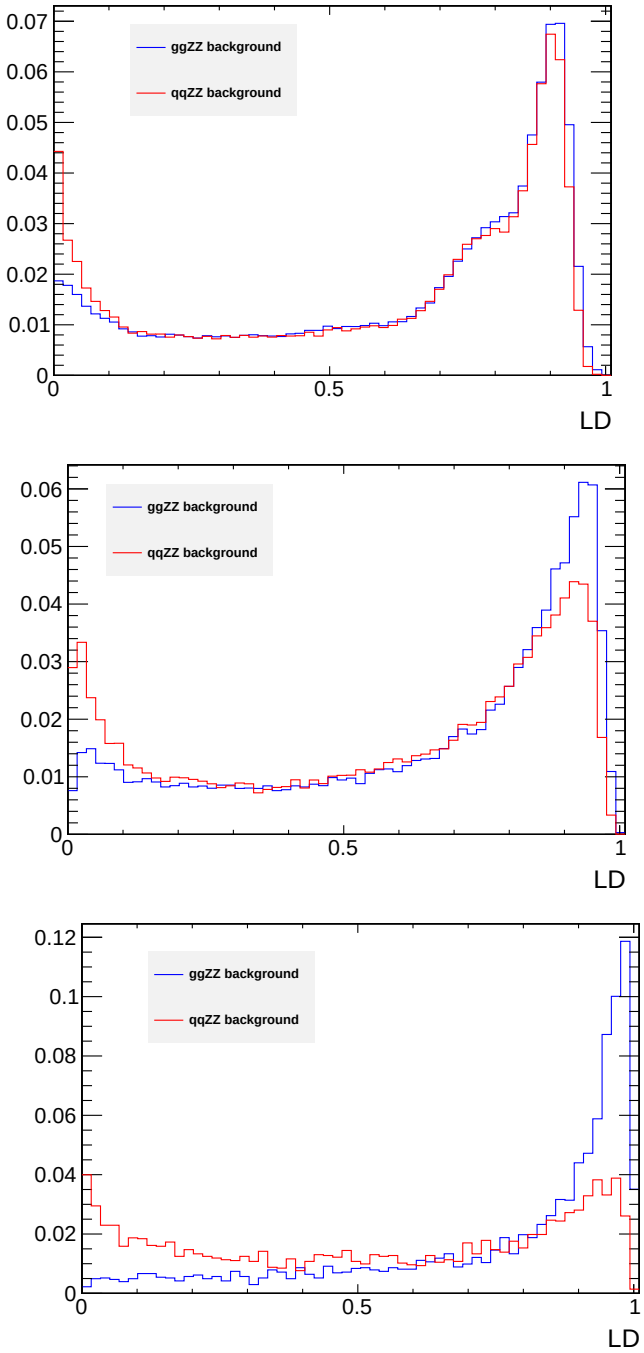

Figure 5.3: *Discriminant distributions obtained from ggMELA, both for* $gg \to ZZ$ *and* $q\bar{q} \to ZZ$*, at a centre-of-mass energy of* 7 TeV *and in three different mass ranges:* $180 < m_{ZZ} < 300$ GeV/c^2 *(top),* $300 < m_{ZZ} < 500$ GeV/c^2 *(middle), and* $500 < m_{ZZ} < 700$ GeV/c^2 *(bottom).*

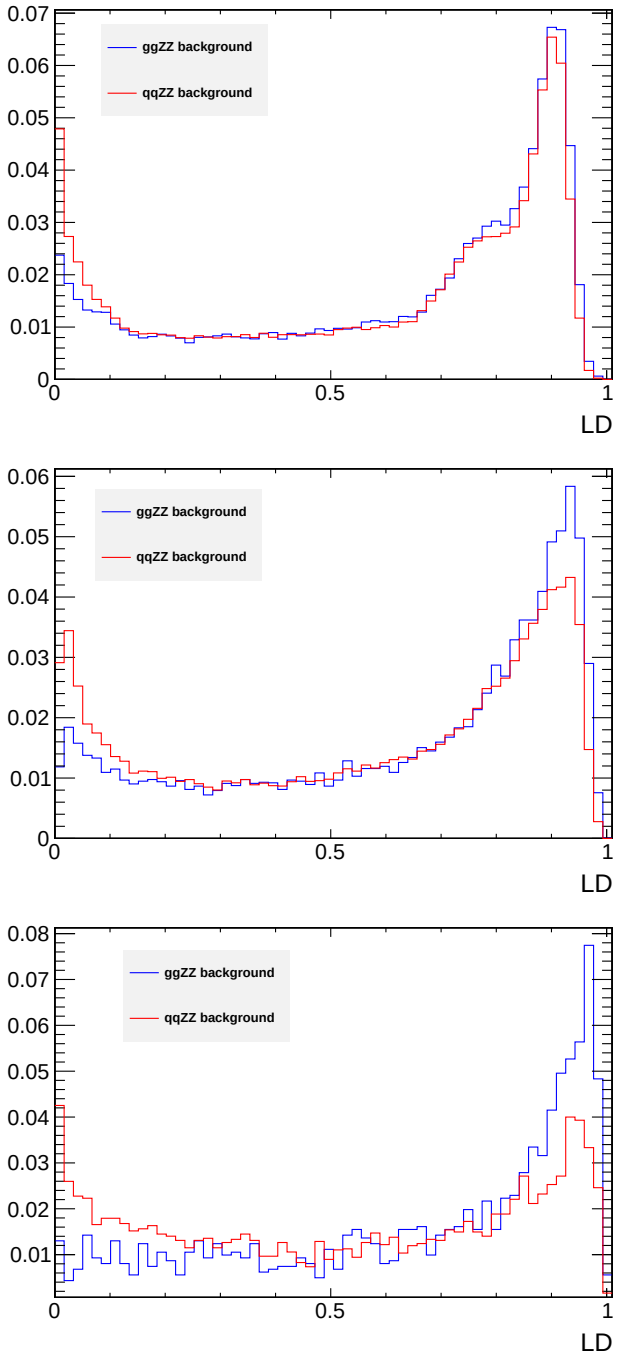

Figure 5.4: *Discriminant distributions obtained from ggMELA, both for* $gg \to ZZ$ *and* $q\bar{q} \to ZZ$*, at a centre-of-mass energy of* $8\ TeV$ *and in three different mass ranges:* $180 < m_{ZZ} < 300\ GeV/c^2$ *(top),* $300 < m_{ZZ} < 500\ GeV/c^2$ *(middle), and* $500 < m_{ZZ} < 700\ GeV/c^2$ *(bottom).*

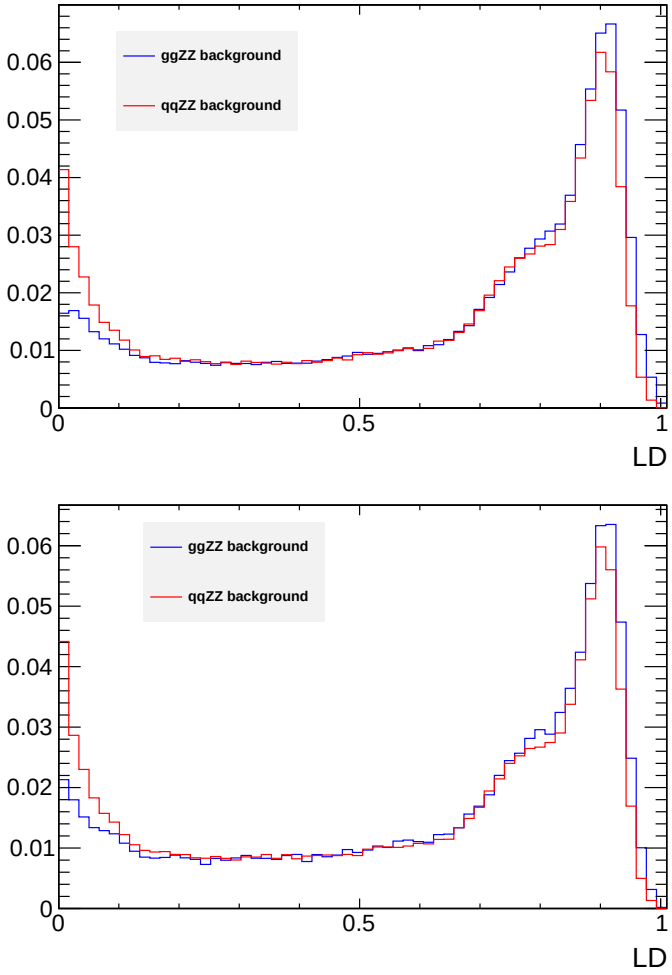

Figure 5.5: *Discriminant distributions obtained from ggMELA, both for* $gg \to ZZ$ *and* $q\bar{q} \to ZZ$*, at a centre-of-mass energy of* $7\ TeV$ *(top) and* $8\ TeV$ *(bottom), in the full mass range:* $180 < m_{ZZ} < 700\ GeV/c^2$.

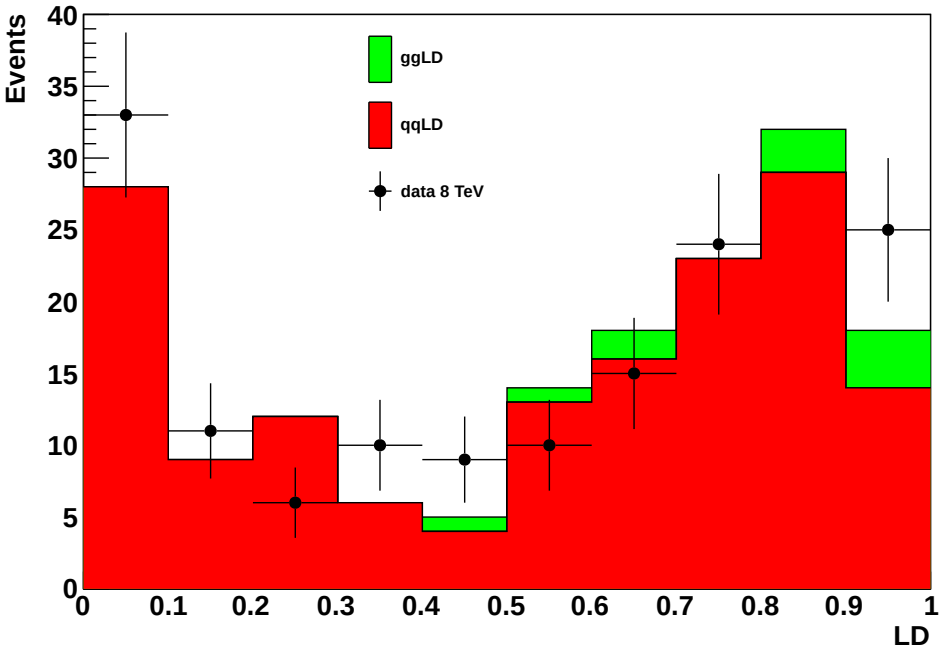

Figure 5.6: *Comparison between data and MC for the discriminant variable distribution at 8 TeV: points represent data, and histograms represent the $gg \to ZZ$ and $q\bar{q} \to ZZ$ background contributions, obtained from MC.*

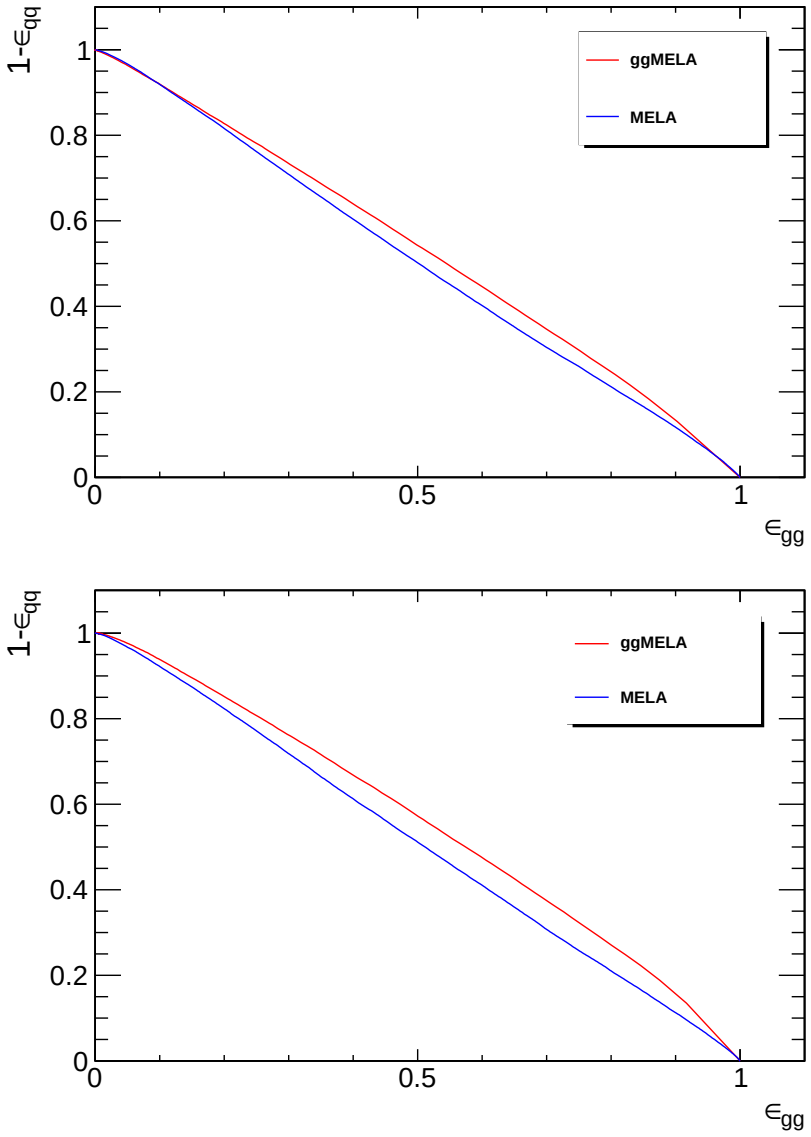

Figure 5.7: *ROC curves obtained from MELA and from ggMELA kinematic discriminants for a centre-of-mass energy of* 7 *(top) and* 8 *TeV (bottom).*

$gg \to ZZ$ one. As a consequence it is expected that, plotting $m_{4\ell}$ with the cut ggMELA > 0.7, the $q\bar{q} \to ZZ$ events are more suppressed than the $gg \to ZZ$ ones. In Figure 5.8 the $m_{4\ell}$ distributions for $gg \to ZZ$ and $q\bar{q} \to ZZ$ MC and for data, without (top) and with (bottom) the cut on ggMELA, are shown for a centre-of-mass energy of 8 TeV and a luminosity of 12.2 fb^{-1}. With the cut ggMELA > 0.7 the $q\bar{q} \to ZZ$ events are reduced by 60%, while the $gg \to ZZ$ events are reduced by 45%. Therefore, the $q\bar{q} \to ZZ$ contribution is more suppressed than the $gg \to ZZ$ one, as expected.

5.2 A model to calculate the $gg \to ZZ$ contribution

After the ggMELA implementation, the final step is the construction of a model which is able to fit MC samples or data and to extract the $gg \to ZZ$ fraction of events f_{gg}. This procedure starts creating 2D templates of discriminant variable obtained from ggMELA as a function of m_{ZZ} (in the full mass range $180 < m_{ZZ} < 700\ GeV/c^2$); this is done both for $gg \to ZZ$ and $q\bar{q} \to ZZ$ at a centre-of-mass energy of 7 and 8 TeV. These templates are shown in Figures 5.9 and 5.10.
At this point it is possible to construct the model according to the formula:

$$M(x,y) = f_{gg} \cdot ggZZ(x,y) + (1 - f_{gg}) \cdot qqZZ(x,y), \tag{5.2}$$

where

- $ggZZ(x,y)$ is the probability density for the $gg \to ZZ$ process,
- $qqZZ(x,y)$ is the probability density for the $q\bar{q} \to ZZ$ process,
- f_{gg} is the $gg \to ZZ$ contribution, free parameter of the model,
- x corresponds to m_{ZZ}, while y corresponds to the discriminant variable.

5.3 Fitting MC samples and extraction of $gg \to ZZ$ fraction

A set of MC samples, with a certain percentage of $gg \to ZZ$ and $q\bar{q} \to ZZ$ events, has been created and fitted with the model. In Table 5.1 the results of this check are reported, both for a centre-of-mass energy of 7 and 8 TeV. It can be seen that the model is able to extract the correct fraction of $gg \to ZZ$ with a small relative error (ratio between the fit statistic uncertainty and f_{gg} extracted). Furthermore, the value of f_{gg} extracted from the fit is more and more close to the real one, and the relative error is more and more small as the fraction of $gg \to ZZ$ increases. In Figure 5.11, as

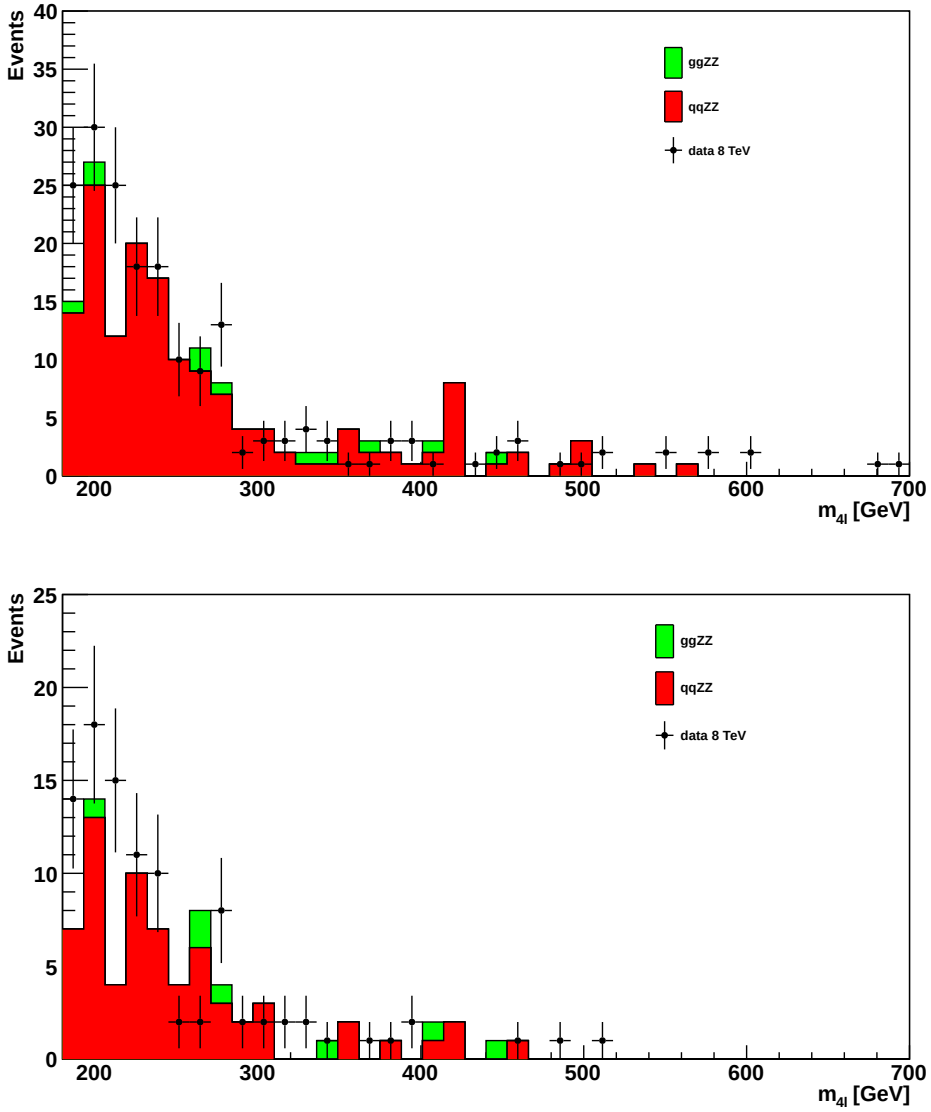

Figure 5.8: *Comparison between data and MC for the $m_{4\ell}$ distribution, without (top) and with (bottom) the cut ggMELA > 0.7, for a centre-of-mass energy of 8 TeV and a luminosity of 12.2 fb^{-1}. Points represent data, and histograms represent the $gg \to ZZ$ and $q\bar{q} \to ZZ$ background contributions, obtained from MC.*

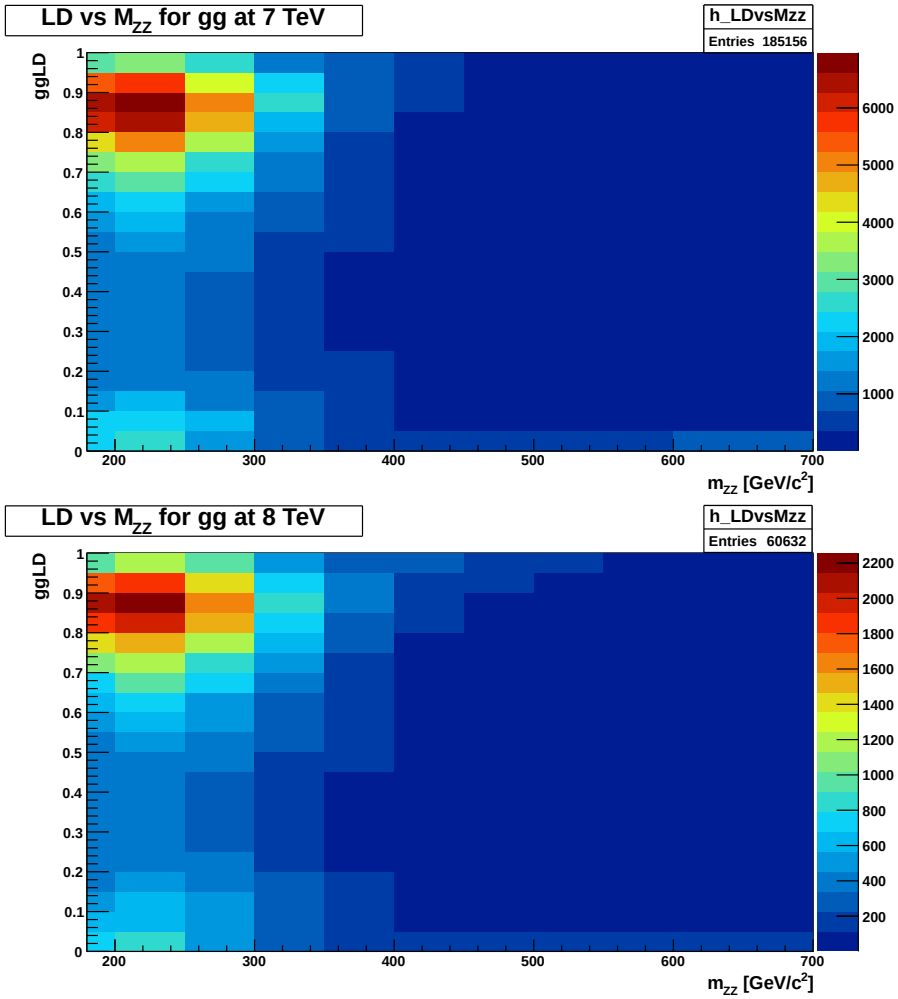

Figure 5.9: *2D templates of ggMELA discriminant as a function of m_{ZZ}, in the full mass range $180 < m_{ZZ} < 700\ GeV/c^2$, at a centre-of-mass energy of 7 (top) and 8 TeV (bottom), for the $gg \to ZZ$ process.*

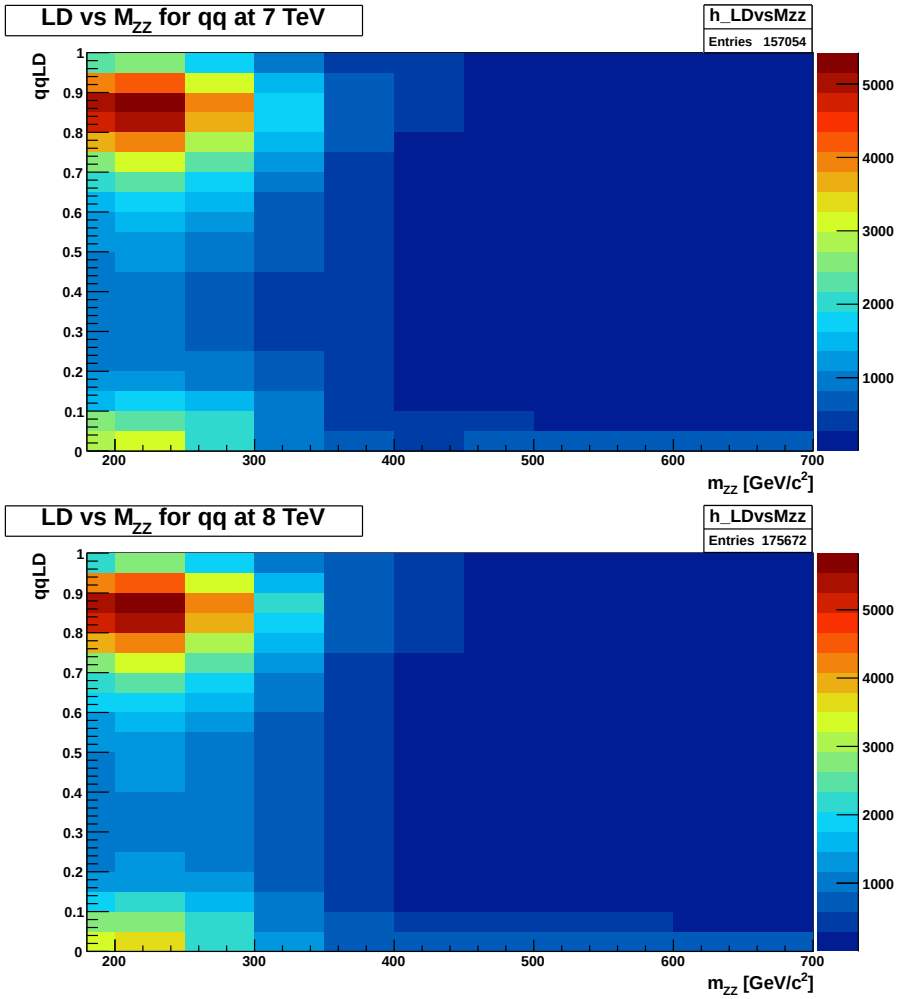

Figure 5.10: *2D templates of ggMELA discriminant as a function of m_{ZZ}, in the full mass range $180 < m_{ZZ} < 700\ GeV/c^2$, at a centre-of-mass energy of 7 (top) and 8 TeV (bottom), for the $q\bar{q} \to ZZ$ process.*

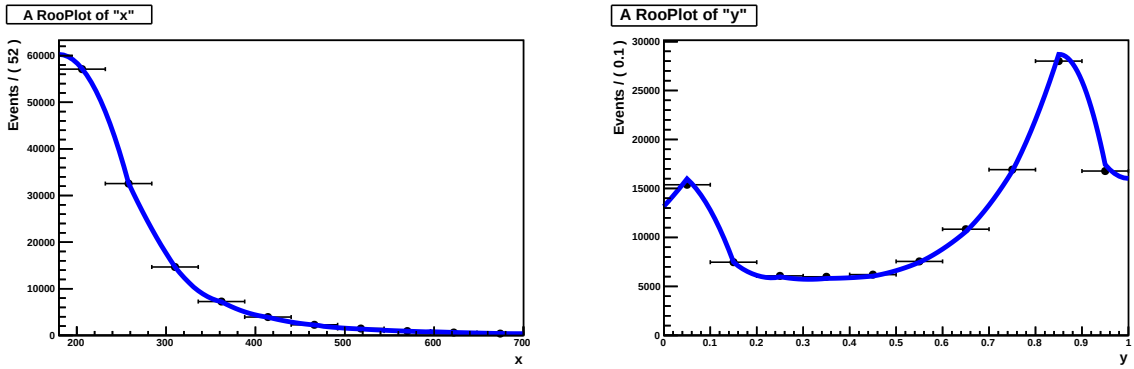

Figure 5.11: *Fit with the model (blue line) at 8 TeV on MC sample (black points) with 50% gg and 50% $q\bar{q}$ composition. x projection corresponds to m_{ZZ}, while y projection corresponds to the discriminant variable.*

an example, the fit at 8 TeV on MC sample with 50% gg and 50% $q\bar{q}$ composition is shown. It is important to underline that the samples which are fitted are done with the same events used to do the model. Fitting the same events from which the model is obtained, statistic fluctuations in the distributions are substantially ignored. Systematic uncertainties on distributions in the MC are ignored too. For these reasons it is more difficult for the fit to converge at the correct value.
Furthermore, from a study with MC samples with the same $gg \to ZZ$ composition ($\sim$ 7%, that is the $gg \to ZZ$ fraction expected from data) but different luminosities, it is evident the decrease of the relative error with an increasing statistic. In Figure 5.12 the relative error has been plotted as a function of the luminosity, both for a centre-of-mass energy of 7 (top) and 8 TeV (bottom).
Finally, a fit on data has been done, both for 7 and 8 TeV. Data collected at $\sqrt{s} = 7\ TeV$, corresponding to an integrated luminosity of 5.1 fb^{-1}, and data collected at $\sqrt{s} = 8\ TeV$, corresponding to an integrated luminosity of 12.2 fb^{-1}, have been used. In the following the obtained results:

- 7 TeV: $f_{gg} = 0.7 \pm 0.8$
- 8 TeV: $f_{gg} = 0.0 \pm 0.9$

As one can see, with this amount of data is not possible to obtain a useful measurement of f_{gg}, since the range $0 - 1$ is covered by the statistic uncertainty. It is expected that the use of this model at an energy of 14 TeV (the LHC design centre-of-mass energy) can give better results. In fact, at this energy both the ZZ cross-section and the $gg \to ZZ$ fraction are higher, allowing the model to work better.

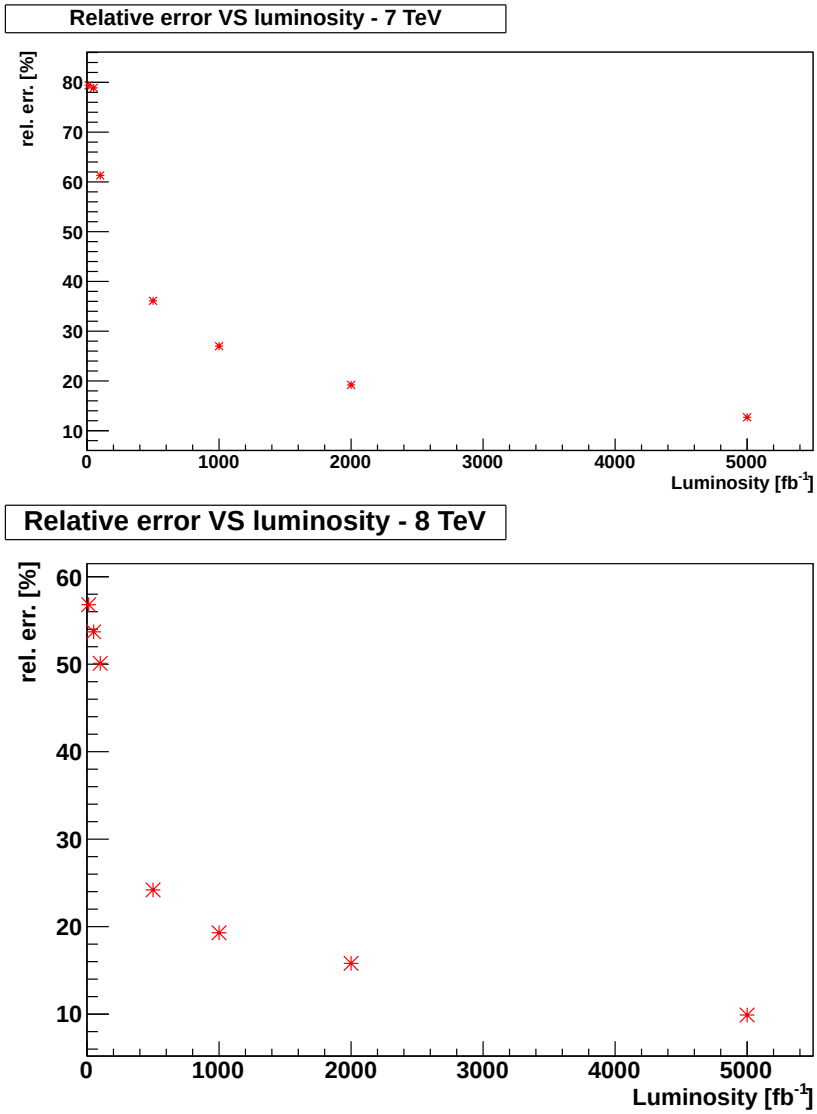

Figure 5.12: *Relative error as a function of luminosity, both for 7 and 8 TeV model.*

Table 5.1: *Results of the fits, with the models at 7 and 8 TeV obtained from Equation 5.2, on MC samples with different composition in $gg \to ZZ$ and $q\bar{q} \to ZZ$.*

Sample composition (% and N_{ev})	f_{gg} **-** 7 TeV	**rel. err.**	f_{gg} **-** 8 TeV	**rel. err.**
$10\%gg - 90\%q\bar{q}$ 7 TeV: 16000 gg − 144000 $q\bar{q}$ 8 TeV: 12120 gg − 109080 $q\bar{q}$	0.086 ± 0.017	19.8%	0.113 ± 0.009	7.9%
$50\%gg - 50\%q\bar{q}$ 7 TeV: 157000 gg − 157000 $q\bar{q}$ 8 TeV: 60600 gg − 60600 $q\bar{q}$	0.475 ± 0.012	2.5%	0.498 ± 0.011	2.2%
$75\%gg - 25\%q\bar{q}$ 7 TeV: 150000 gg − 50000 $q\bar{q}$ 8 TeV: 60000 gg − 20000 $q\bar{q}$	0.749 ± 0.014	1.9%	0.772 ± 0.014	1.8%
$100\%gg - 0\%q\bar{q}$ 7 TeV: 185000 gg − 0 $q\bar{q}$ 8 TeV: 60632 gg − 0 $q\bar{q}$	1.000 ± 0.003	0.3%	0.999 ± 0.004	0.4%
$0\%gg - 100\%q\bar{q}$ 7 TeV: 0 gg − 157000 $q\bar{q}$ 8 TeV: 0 gg − 109080 $q\bar{q}$	0.000 ± 0.005		0.000 ± 0.001	

Conclusions

The work presented in this thesis was carried out within the CMS Torino group, working in the search for the Higgs boson in the decay channel $H \to ZZ^{(*)} \to 4\ell$.
In this thesis, I presented a study of the irreducible backgrounds, that the search for a Higgs signal in the $H \to ZZ^{(*)} \to 4\ell$ channel has to cope with. Exploiting the slight differences present between $gg \to ZZ$ and $q\bar{q} \to ZZ$ angular distributions, a new kinematic discriminant, ggMELA, was developed in order to separate the two irreducible backgrounds. ggMELA was then used to construct a model able to extract the $gg \to ZZ$ fraction (f_{gg}), fitting a sample of reconstructed ZZ candidates.
This model was tested on different MC samples, finding that its performance improves at the increase of the statistic and of the $gg \to ZZ$ fraction. As a consequence it is expected that the use of the method presented in this work at an energy of 14 TeV (the LHC design centre-of-mass energy) can give very good results. In fact, at a centre-of-mass energy of 14 TeV both the ZZ cross-section and the $gg \to ZZ$ fraction are higher. Therefore, at this energy the extracted fraction f_{gg} should be very accurate, with a relative error smaller than 10%.
Finally, data collected at $\sqrt{s} = 7\ TeV$, corresponding to an integrated luminosity of 5.1 fb^{-1}, and data collected at $\sqrt{s} = 8\ TeV$, corresponding to an additional 12.2 fb^{-1}, were fitted. Nevertheless, this amount of data was found not to be sufficient to derive f_{gg}.

Bibliography

[1] F. Halzen, A. Martin, *Quarks and leptons: an introductory course in modern Particle Physics*, Wiley, New York USA (1984).

[2] A. Salam, "Elementary Particle Physics," *N. Svartholm ed., Almquvist and Wiksell, Stockholm*, 1968.

[3] E. Fermi, Trends to a theory of Beta radiation, Nuovo Cimento, vol. 11, pp. 1-19, 1934.

[4] S. Dawson, "Introduction to electroweak symmetry breaking," 1998.hep-ph/9901280.

[5] G. Ridolfi, "Search for the Higgs boson: Theoretical perspectives," 2001. arXiv:0106300 [hep-ph].

[6] P. Proudlock, "LEP200 Design report," Tech. Rep. SL-PC-Tech-Note-91-87, CERN, Geneva, Jun 1991.

[7] *SLC design handbook. Stanford Linear Collider. Design report.* Stanford,CA: SLAC, 1984.

[8] Tevatron I Group, "Design Report Tevatron 1 Project," 1984. FERMILAB-DESIGN-1984-01.

[9] The LEP Working Group for Higgs boson searches, "Search for the standard model Higgs boson at LEP," *Phys. Lett.*, vol. B565, pp. 61–75, 2003.

[10] CDF and D⊘ Collaborations, "Combined CDF and D⊘ Upper Limits on Standard Model Higgs-Boson Production with up to $8.6 fb^{-1}$ of Data." 2011. arXiv:1107.5518v2.

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

[12] Tevatron Electroweak Working Group. http://tevewwg.fnal.gov/.

[13] The CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, CERN-PH-EP/2012-220, 2012/07/31.

[14] 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.

[15] "LHC Design Report." CERN 2004-003, 2004.

[16] The CMS Collaboration, "Technical Proposal", CERN/LHCC 94-38.

[17] The CMS Collaboration, "The Tracker Project Technical Design Report", CERN/LHCC 98-6 and CMS TDR 5 (1998).

[18] The CMS Collaboration, "The Electromagnetic Calorimeter Project Technical Design Report", CERN/LHCC 97-33 and CMS TDR 4 (1997).

[19] The CMS Collaboration, "The Hadronic Calorimeter Technical Design Report", CERN/LHCC 97-31 and CMS TDR 2 (1997).

[20] The CMS Collaboration, "The Muon Project Technical Design Report", CERN/LHCC 97-32 and CMS TDR 2 (1997).

[21] The CMS Collaboration, "The TriDAS Project Technical Design Report", Volume I: The Level-1 Trigger, CERN/LHCC 2000-038, CMS TDR 6.1 (2000).

[22] The CMS Collaboration, "The TriDAS Project Technical Design Report", Volume II: Data Acquisition and High-Level Trigger, CERN/LHCC 02-26 and CMS TDR 6.1 (2002).

[23] CMS NOTE AN-12-367, CMS Draft Note, "Updated results on the new boson discovered in the search for the standard model Higgs boson in the $H \to ZZ \to 4\ell$ channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV.". (2012/11/15)

[24] CMS PAS HIG-12-041, CMS Physics Analysis Summary, "Updated results on the new boson discovered in the search for the standard model Higgs boson in the $H \to ZZ \to 4\ell$ channel in pp collisions at $\sqrt{s} = 7$ and 8 TeV.". (2012/11/09)

[25] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", JHEP 11 (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[26] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", JHEP 05 (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[27] A. Bredenstein et al., "Precise predictions for the Higgs-boson decay H $\to$ WW/ZZ $\to$ 4 leptons", Phys. Rev. D 74 (2006) 013004, doi:10.1103/PhysRevD.74.013004, arXiv:hep-ph/0604011.

[28] J. Alwall et al., "MadGraph/MadEvent v4: The New Web Generation", JHEP 09 (2007) 028, doi:10.1088/1126-6708/2007/09/028, arXiv:0706.2334.

[29] T. Binoth, N. Kauer, and P. Mertsch, "Gluon-induced QCD corrections to $pp \to ZZ \to \ell\bar{\ell}\ell'\bar{\ell}'$", in Proceedings of the XVI Int. Workshop on Deep-Inelastic Scattering and Related Topics (DIS'07). 2008. arXiv:0807.0024. doi:10.3360/dis.2008.142.

[30] J. Allison et al., "Geant4 developments and applications", IEEE Trans. Nucl. Sci. 53 (2006) 270, doi:10.1109/TNS.2006.869826.

[31] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", Phys. Lett. B 716 (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[32] Y. Gao et al., "Spin determination of single-produced resonances at hadron colliders", Phys. Rev. D 81 (2010) 075022, doi:10.1103/PhysRevD.81.075022, arXiv:1001.3396.

[33] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", Nucl. Instrum. Meth. A 262 (1987) 444, doi:10.1016/0168-9002(87)90887-4.

[34] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", Phys. Lett. B 659 (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[35] CMS Collaboration, "MUO-10-004: Performance of CMS muon reconstruction in pp collision events at sqrts = 7 TeV", technical report, (2010). In preparation to JINST.

[36] G. Daskalakis et al., "Measuring Electron Efficiencies at CMS with Early Data", CMS Analysis Note 2007/019 (2007).

[37] K. Nakamura et al. (Particle Data Group), "The Review of Particle Physics", J. Phys. G 37 (2010) 075021.

[38] J. S. Gainer, K. Kumar, I. Low et al., "Improving the sensitivity of Higgs boson searches in the golden channel", JHEP 1111 (2011) 027, doi:10.1007/JHEP11(2011)027, arXiv:1108.2274.

[39] 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.

[40] 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)

Acknowledgements

Grazie di cuore a...

GRUPPO CMS TORINO, perché è formato da persone fantastiche, pronte ad aiutarti e a consigliarti in qualsiasi occasione;
CHIARA, perché sei la miglior corelatrice che si possa desiderare, anzi, sei più di una corelatrice, sei un'amica. Grazie per il tuo sorriso e i tuoi incoraggiamenti, grazie per aver sempre trovato il tempo per seguire il mio lavoro nonostante i tuoi numerosi impegni, grazie per l'entusiasmo che hai e che trasmetti; sono poche le persone che ti spiegano la fisica con gli occhi che brillano, e tu sei una di quelle;
SARA, perché anche tu sei stata una corelatrice stupenda, sempre disponibile e piena di entusiasmo. Grazie per tutto quello che mi hai insegnato, grazie per esserci stata anche quando non stavi bene. Spero di avere ancora il piacere di lavorare con te;
MARIO, per tutto quello che mi hai insegnato durante l'estate 2011, facendomi appassionare a questo lavoro. Grazie per i preziosi consigli che mi hai dato in questi mesi ginevrini, per la tua disponibilità, per la tua serietà, la dedizione e la passione che mi hai trasmesso. Grazie di avere quello humour sottile che mi ha fatto ridere e sorridere tante volte. E grazie per il Dragone da 1700 W...
Ah...ti prometto che prima o poi mi leggerò la guida di C++...
NICOLA, relatore preciso e attento. Prendersi la responsabilità di seguire contemporaneamente 3 laureandi non è facile. Tu lo hai fatto, e nonostante fossi sommerso dal lavoro sei sempre stato disponibile per qualsiasi domanda o dubbio. Grazie per l'attenzione con cui mi hai corretto la tesi e per i consigli che mi hai dato;
NICOLO', per le magnifiche chiacchierate di fisica e non che ci siamo fatti, grazie per la tua innata capacità di rendere meraviglioso e interessante tutto ciò di cui parli. Grazie per il tuo sorriso e il tuo buonumore;
CRIS, per le risposte sempre pronte alle mie domande e per la passione con cui svolgi il tuo lavoro;
STEFANO, per i preziosi consigli;
NADIA e SILVIA, per la vostra infinita disponibilità;
PROF.SSA ROMERO, per aver più volte letto e corretto i miei progetti di

ricerca volti ad ottenere borse di studio;
PROF. COSTA, per avermi spinta a fare il summer student al CERN nell'estate 2011. Grazie a quell'esperienza ho conosciuto Chiara e Mario, e a quel punto non ho più avuto dubbi su quali sarebbero state le persone con cui avrei svolto la mia tesi magistrale;
LINDA, perché sei un'amica fantastica, perché in questi 2 anni di magistrale abbiamo condiviso praticamente tutto. Grazie per la tua voglia di vivere e di provare esperienze nuove, grazie per le innumerevoli risate che ci siamo fatte (comprese quelle in ufficio...), grazie per essermi stata vicino anche nei momenti più delicati;
GIANLU, GIULIA, LINDA, VIO, EMA, per essere stati dei coinquilini stupendi: senza di voi la Indian House non sarebbe stata la stessa!! Gianlu e Vio grazie per il vostro "pòco pòco", ormai non posso farne a meno!
MAX, LUCA, ALBERTO, per essere entrati a far parte della Indian House Family, per le numerose e divertenti serate-film fatte insieme;
TIZIANO, per gli n moduli che mi hai firmato. E grazie per quel "Ragazzi, ho la pelle d'oca!" detto in ufficio quel giorno di giugno; hai fatto venire la pelle d'oca anche a noi :-)
DILIP, ebbene sì, ti ringrazio perché ci hai insegnato cosa vuol dire avere spirito di squadra. Grazie per avermi fatto apprezzare il profumo della candeggina come mai prima d'ora, e grazie per avermi insegnato che anche un fisico certe domande non se le deve porre;
ALICE, DANIELA, DARIA, per i bellissimi anni di università passati assieme, perché ci siete sempre state, sia nei momenti belli sia in quelli difficili. Vi voglio bene;
SIMONA, per aver condiviso 5 meravigliosi anni al mitico n.2 di Via Sacchi;
AMICI TORINESI, perché i 5 anni di università con voi sono stati unici. Il Corso di Studi in Fisica è bello non solo per quello che si studia ma anche per le persone che lo frequentano;
AMICI VALDOSTANI, bandisti e non, perché ogni volta che tornavo in Valle durante questi mesi, anche solo per pochi giorni, potevo sempre contare sulla vostra allegria!
LUDOVICO, perché difficilmente potrei sperare di avere un fratello migliore;
MAMMA E PAPA', perché senza di voi non sarei qui a scrivere questa tesi. Grazie per avermi sempre supportato, sia moralmente sia economicamente;

Infine grazie a lui, al BOSONE DI HIGGS, che ha fatto capolino poche settimane dopo il nostro arrivo al CERN, permettendoci di vivere momenti indimenticabili!