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

Tesi di Laurea Magistrale in Fisica

Measurement of the $Zb\bar{b}$, $t\bar{t}$ and $Z + light\ jets$ Backgrounds in the CMS Experiment at the LHC in the Search of the Higgs Boson in the Final State $H \to ZZ \to 4\ell$

Relatore
Dr. Nicola Amapane

Correlatori
Dr.ssa Chiara Mariotti

Dr. Mario Pelliccioni

Controrelatore
Prof. Massimo Masera

Candidata
Linda Finco

Anno accademico 2011-2012

Contents

Introduction

The Standard Model of Particle Physics is, up to now, one of the best-tested physics theories of modern Physics. Developed in the early 1970s, it has successfully explained several experimental results and precisely predicted a wide variety of phenomena. This theory explains how the twelve elementary particles and three of the fourth fundamental forces are related to each other.
The Standard Model predicts the existence of the Higgs scalar boson, that represents the key to the origin of the mass of fundamental particles. In order to confirm this theory it is thus necessary to observe the Higgs boson experimentally, but since the Higgs boson mass is not yet known, physicists have to look for it over a wide range of possible values.
The "Higgs hunting" is a central purpose of the physics program at the Large Hadrons Collider (LHC), the world's major particle accelerator. The Compact Muon Solenoid (CMS) is one of the four main experiments at LHC and the Higgs search is a core part of its physics program.
The Higgs boson production followed by the decay $H \to ZZ \to \ell^{\pm}\ell^{\mp}\ell'^{\pm}\ell'^{\mp}$ with ℓ, $\ell' = e$ or μ (in short $H \to 4\ell$) is expected to be one of the main decay channels for the Higgs boson discovery.
Even if finding the Higgs boson is certainly an important step to confirm the Standard Model validity, it is not conclusive, because this theory, despite its effectiveness at describing the phenomena within its domain, is nevertheless incomplete. For this reason, CMS has a rich program for physics beyond the Standard Model.

The work presented in this thesis concerns the study of reducible and instrumental backgrounds that the search for a Higgs signal in the $H \to 4\ell$ channel has to cope with. Reducible backgrounds derive from the $Zb\bar{b}$ and $t\bar{t}$ processes, with W bosons undergoing leptonic decays and secondary leptons produced within b-jets. Instrumental background contributions come from $Z + jets$ or $WZ + jet(s)$ where jets are misidentified as leptons. Reducible and instrumental backgrounds are studied relying primarily on the measurements of the final state leptons.
After a brief overview of the Standard Model that describes the theoretical framework of this analysis, in Chapter 1 the attention is focused on

Higgs boson properties. Theoretical and experimental mass constraints are discussed and the Higgs phenomenology and search at the LHC are presented. 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.
In Chapter 3, the $H \to ZZ^{(\star)} \to 4\ell$ decay channel analysis is presented. This is the experimental context in which my analysis is focused. The 4ℓ final state signal and background processes are described in detail, for a range of mass from 110 to 600 GeV/c^2. An excess of events is observed above the expected background at a mass near 125 GeV/c^2, signaling the evidence of a new particle.
Chapter 4 deals with the reducible and instrumental backgrounds of the $H \to ZZ^{(\star)} \to 4\ell$ signal. Even if these processes are drastically reduced by the analysis selection cuts, a small contribution is present in the final sample. This requires a detailed study of their composition. My work deals in the reconstruction and selection of the $Zb\bar{b}$ and $t\bar{t}$ candidates and the discrimination of these two contributions. Also, I present for the first time a measurement of the $Zb\bar{b} \to 4\ell$ and $t\bar{t} \to 4\ell$ cross sections at 7 and 8 TeV.
Finally, in Chapter 5 a new data driven procedure to separately measure instrumental and reducible backgrounds that remain after the first steps of the $H \to ZZ^{(\star)} \to 4\ell$ selection is presented. In order to study the two different contributions, I define a control sample and, fitting distributions related to the impact parameter of leptons, I estimate the number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events in the signal region of the $H \to 4\ell$ analysis.

Chapter 1

The Standard Model Higgs Boson

The Standard Model (SM) of elementary particles 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 presence 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 sections theoretical and experimental constraints to the Higgs boson mass are presented. The Higgs phenomenology at the LHC collider is 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*, organized in three families, together with their anti-particles (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 *hadrons*, a class of particles such as protons and neutrons.
In the SM, the interactions between elementary particles are mediate by *bosons*, integer-spin particles which are carriers of the fundamental forces. The main characteristics of bosons and of the corresponding interactions are summarized in Table 1.2.
The gravitational interaction is not taken into account, as it is not relevant

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}

at the typical mass and distance scales of particle physics.

This 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. By the Noether's theorem, a conservation law corresponds to each of these local invariances, explaining why they are so important. Therefore the SM is a local gauge quantum field theory describing three of the four fundamental interactions: electromagnetic, weak interaction and strong interaction. It is based on the symmetry group

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

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, $SU(2)_L$, and hypercharge, $U(1)_Y$. Electromagnetic and weak interactions are unified in the *electroweak* gauge group $SU(2)_L \otimes U(1)_Y$, upon which the Glashow-Weinberg-Salam Model is built.
Despite this symmetry 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 [1], 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 generators 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

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

requirement of 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:

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

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:

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

while the neutral bosons γ and Z correspond to

$$A_\mu = B_\mu cos\theta_W + W_\mu^3 sin\theta_W \quad (1.5)$$

$$Z_\mu = -B_\mu sin\theta_W + W_\mu^3 cos\theta_W \quad (1.6)$$

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

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

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

$$g\ sin\theta_W = g' cos\theta_W = e \tag{1.8}$$

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 a mass term for the gauge bosons would violate gauge invariance, which is needed to ensure the renormalisability of the theory. Explicit mass terms for fermions, instead, would not violate gauge invariance, but in the GWS model the Lagrangian is also required to preserve the invariance under chirality transformations, and this is achieved only with massless fermions. Masses are thus introduced with the *Higgs mechanism* [2], 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,

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

which is introduced in the electroweak Lagrangian within the term

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

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 characterized 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*. It is therefore chosen arbitrarily a minimum along the ϕ^0 axis:

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

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 1.10 with the ϕ field in the unitary gauge, $\mathcal{L}_{EWSB}$ results from the sum of three terms:

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

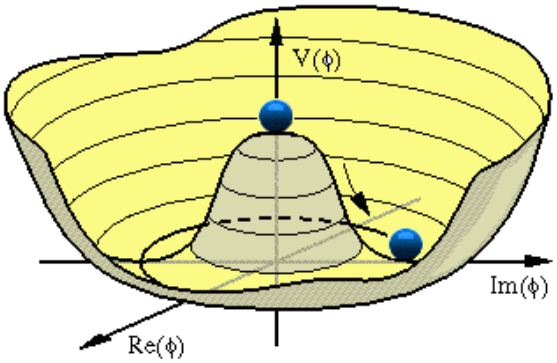


Figure 1.1: *Shape of the Higgs potential.*

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 field.

Summarizing, the Higgs mechanism is used to introduce the weak boson masses, without explicitly breaking the gauge invariance, thus preserving the renormalisability of the theory. When a symmetry is "spontaneously" broken 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. Hence, the electromagnetic symmetry is 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:

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

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:

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

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_ℓ: this correctly reproduces the fact that neutrino is (approximately) massless.

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

As far as the quark fields are concerned, 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

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

Therefore, the Yukawa Lagrangian is

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

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, the mass of the fermions cannot be predicted by the theory.

1.4 Higgs Boson Mass

As written above, the Higgs boson mass cannot be predicted by the theory. One can show in fact that

$$m_H = \sqrt{2\lambda}v,$$

where v is proportional to the vacuum expectation value and λ is the running coupling constant of the Higgs potential. While the former can be estimated by its relation with the constant G_F of Fermi's theory, the latter is characteristic of the Higgs field and can only be determined by measuring the mass itself. However, theoretical arguments and experimental constraints impose lower and upper bounds on the Higgs mass.

1.4.1 Theoretical limits

An upper bound on the Higgs mass comes from unitarity of the scattering matrix. Considering the elastic scattering of longitudinally polarized Z

bosons ($Z_L Z_L \rightarrow Z_L Z_L$) the unitarity bound on the corresponding amplitude implies, in the limit $s \ll m_Z^2$ (where s is the center-of-mass energy),

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

More restrictive bounds of about 800 GeV/c^2 can be found considering other scattering processes, such as

$$Z_L W_L \rightarrow Z_L W_L,$$

if the perturbative regime is applicable. Otherwise, for masses $\sim 1.4 \; TeV/c^2$, the decay width of the SM Higgs into a pair of gauge bosons becomes approximately equal to m_H and the Higgs can no longer be considered a particle. Therefore for large m_H the perturbative approach is not valid anymore and not-perturbative techniques are required.

1.4.2 Experimental constraints

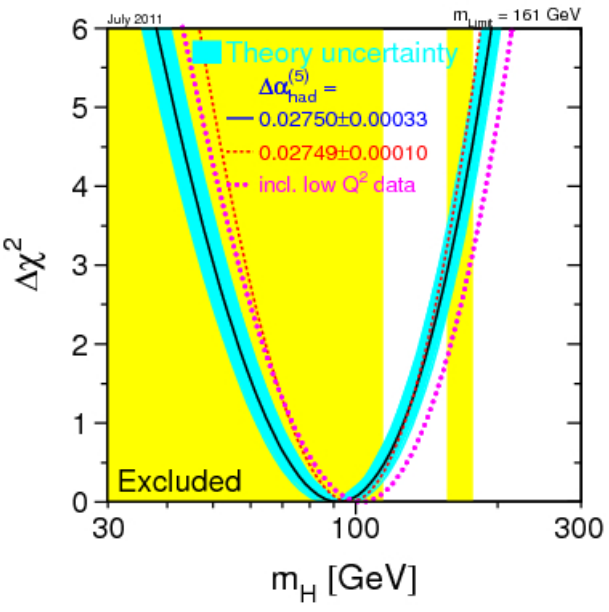


Figure 1.2: $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ *of the fit to the electroweak precision measurements performed at LEP, SLC and Tevatron as a function of the Higgs boson mass [3].*

Bounds on the Higgs mass are provided by measurements at LEP, SLC and Tevatron. LEP experiments exclude the Higgs boson for $m_H \lesssim 114.4\ GeV/c^2$ at 95% confident level (C.L.), setting a lower limit, while results from the Tevatron exclude the mass ranges from 156 to 177 GeV/c^2 and from 100 to 108 GeV/c^2 (95% C.L.).
Moreover, constraints on the Higgs boson mass can be extracted indirectly from the measurement of other electroweak observables, which have a logarithmic dependence on m_H through the radiative corrections. All the precision electroweak measurements performed by the four LEP experiments and by SLD, CDF and $D\varnothing$ have been combined together and fitted, assuming the SM as the correct theory and using the Higgs mass as free parameter [3]. In electroweak fits measured parameters are allowed to vary within their errors and hence indirectly place limits on the range that the mass of the Higgs boson could take to remain compatible with the measurements. The result of this procedure is summarized in Figure 1.2, where the $\Delta\chi^2$ of the fit, defined as $\Delta\chi^2 = \chi^2 - \chi^2_{min}$, is plotted as a function of m_H. The result of the fit is represented by the black curve, while the blue band shows the theoretical uncertainty due to unknown higher order corrections. The yellow area is the regions excluded by LEP-II and Tevatron measurements. The indirectly measured value of the Higgs boson mass, corresponding to the minimum of the curve in Figure 1.2, is $m_H = 92^{+34}_{-26}\ GeV/c^2$, where the errors represent the experimental uncertainty at 68% C.L. derived from the black line. No theoretical uncertainties are taken into account. An upper limit of 161 GeV/c^2 can also be set at 95% C.L., including the theoretical uncertainty. If one includes the direct search limit of 114.4 GeV/c^2, this value increases to 185 GeV/c^2. These results are model-dependent, as the loop corrections take into account only contributions from known physics and are thus well-grounded only within the SM theory. Direct searches for the SM Higgs bosons are performed by the ATLAS and the CMS experiments at the LHC, using data collected in 2010 and 2011. Searches for the SM Higgs boson in several production and decay channels (see Sections 1.5.1 and 1.5.2) are carried by the two collaborations and the combined results exclude the SM Higgs boson in the mass range 127˜600 GeV/c^2 at 95% CL.

1.5 Higgs Phenomenology at the LHC

The Higgs phenomenology at LHC is different than in previous colliders, due mainly to the larger center-of-mass energy of the pp collisions.
When two protons collide at high energies, the interaction involves their constituents (quarks and gluons), carrying only a fraction x of the total momentum of the particle. The distributions of the x variable for different constituents are called *parton density functions* (PDF) and depend on the

exchanged four-momentum during the interaction. Since particle production at the LHC reaches the mass range of some TeV/c^2, a Feynman x region where the gluon density is much larger than the quark density is explored. For Higgs physics the LHC can be thus thought as a "gluon-collider". The experiments at the LHC are searching for the Higgs boson within a mass range going from 100 GeV/c^2 to about 1 TeV/c^2 .

While the Higgs boson mass is not fixed by the theory, its couplings to fermions and bosons are predicted to be proportional to the corresponding particle masses (for fermions) or squared masses (for bosons). For this reason, the Higgs boson production and decay are dominated by channels involving heavy particles: $W^{\pm}$, Z bosons and the third generation of fermions, in particular.

In this section, the main Higgs production processes and decay channels are described analyzing each mass region.

1.5.1 Higgs Production

The cross section of the main processes contributing to the Higgs boson production at a proton-proton collider is shown in Figure 1.3, for center-of-mass energies of 7 and 14 TeV. The former is the energy provided by the LHC during the 2010-2011 runs, the latter is the LHC design energy that will be gradually reached in the next years. As it is shown, the total production cross section at 7 TeV is up to one order of magnitude lower than at 14 TeV [4].

Gluon-Gluon Fusion

The *gluon-gluon fusion* process is the dominant Higgs production mechanism at the LHC over the whole accessible mass range. The diagram at leading order (LO) is presented in Figure 1.4 with a t quark-loop, that is the main contribution. When combined with the decay channels $H \to \gamma\gamma$, $H \to WW$ and $H \to ZZ$, this production mechanism is one of the most important for Higgs boson searches at the LHC.
The dynamics of the gluon-fusion mechanism is controlled by strong interactions. Detailed studies of the effect of QCD radiative corrections are necessary to obtain accurate theoretical predictions. In QCD perturbation theory, the LO contribution to the gluon-fusion cross section is proportional to α^2, the QCD coupling constant, and it is

$$\hat{\sigma}(gg \to H) = \frac{8\pi^2\Gamma(H \to gg)}{N_g^2 m_H}\delta(\hat{s} - m_H^2),$$

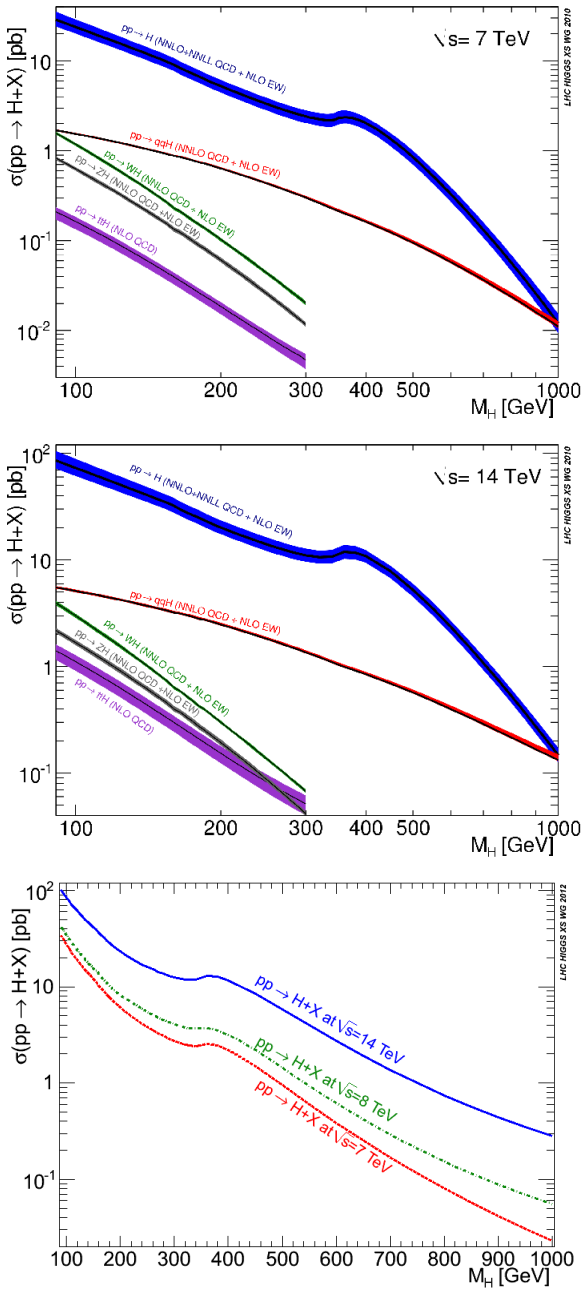


Figure 1.3: *Cross sections for the different Higgs boson production channels, as functions of the Higgs boson mass, at 7 TeV and 14 TeV LHC centre-of-mass energy and their comparison, considering the 8 TeV case too [6].*

where the number of gluons is $N_g = 8$, $\hat{s} = x_1 x_2 s$ is the squared energy of the gluon pair and

$$\Gamma(H \to gg) = \frac{\alpha_S^2 G_F m_H^3}{36\pi^3\sqrt{2}} \left| \sum_q I\left(\frac{m_q^2}{m_H^2}\right) \right|^2,$$

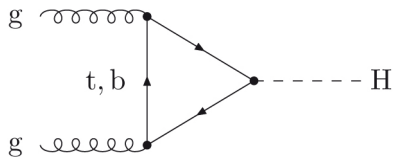


Figure 1.4: *Feynman diagram contributing to $gg \rightarrow H$ at leading order.*

in which m_q is the quark mass and

$$I(z) = 3\Big[2z + 2z(1-4z)\Big(sin^{-1}\frac{1}{2\sqrt{z}}\Big)^2\Big]$$

is a form factor.

The lowest order cross section has large corrections from higher order QCD diagrams. The increase in cross section from higher order diagrams is conventionally defined as the *K*-factor

$$K = \frac{\sigma_{NLO}}{\sigma_{LO}},$$

where LO (NLO) refer to leading (next-to-leading) order results. The *K*-factor for gluon-gluon fusion is ~ 1.7. The QCD radiative corrections to this process at next-to-leading order (NLO) increase the LO cross section by about $80-100\%$ at the LHC. The next-to-next-to-leading order (NNLO) corrections have been computed only in the large-m_t limit, leading to an additional increase of the cross section of about 25%.
The value of the cross section including the K-factor has two main uncertainty contributions. The first depends on the choice of the parton density functions and the second is due to the uncalculated higher-order QCD radiative corrections.
Electroweak (EW) corrections have been computed at NLO and turn out to be small [4].

Vector Boson Fusion

The *vector boson fusion* (VBF, Figure 1.5) is the second contribution to the Higgs boson production cross section. It produces a Higgs particle with two jets, in which the first couples to a weak boson that links two quark lines. Its cross section 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 masses of the order of 1 TeV/c^2 . Nevertheless, this channel is very interesting because of its clear experimental signature: the presence of two spectator jets with high energy provides a powerful tool to

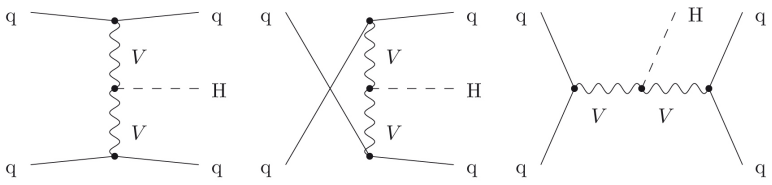


Figure 1.5: *Topologies of t-, u- and s-channel contributions for electroweak Higgs boson production, $qq \rightarrow qqH$ at LO, where q denotes any quark or antiquark and V stands for W and Z boson.*

tag the signal events and discriminate the backgrounds. The hard jet pairs have in fact a strong tendency to be forward–backward directed in contrast to other jet-production mechanisms, offering a clear signature. The signal to background ratio thus improves, despite the low cross section.
LO, NLO and NNLO cross sections for this process are known with small uncertainties and the higher order QCD corrections are quite small. Moreover EW corrections have been computed at NLO.

Vector Boson Associated Production

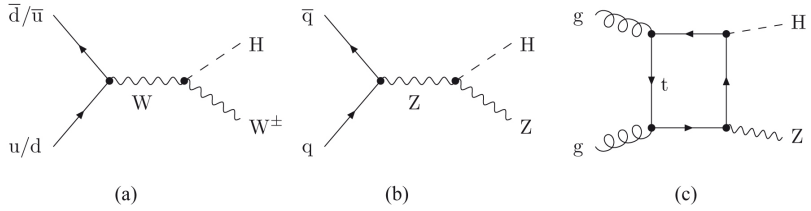


Figure 1.6: *(a), (b) LO diagrams for the partonic processes $pp \rightarrow VH$; (c) diagram contributing to the $gg \rightarrow HZ$ channel.*

In the *Higgsstrahlung* process (Figure 1.6) the Higgs boson is produced in association with a $W^{\pm}$ or Z boson, which can be used to tag the event. This process cross section is several order of magnitude lower than gluon-gluon and VBF production modes.
This process has been considered mainly by exploiting two decay modes, $H \rightarrow W^+W^-$ and $H \rightarrow b\bar{b}$. While the former is interesting because it could contribute to the measurement of the Higgs boson coupling to W bosons, the latter decay mode may contribute to the discovery of a low-mass Higgs boson and may allow to measure the coupling of the Higgs boson to b quarks. The experimental sensitivity to $H \rightarrow W^+W^-$ is highest for Higgs boson masses above about 160 GeV/c^2, while the $H \rightarrow b\bar{b}$ decay modes are investigated for the low Higgs boson mass region, below about 130 GeV/c^2.
The cross section for this process is known at the NNLO QCD and NLO

EW level. The inclusion of these available contributions increases the LO cross section by about $20 - 25\%$.

$t\bar{t}$ Associated Production

The last process, illustrated in Figure 1.7, is the associated production of a Higgs boson with a $t\bar{t}$ pair. Also for this process, the cross section is orders of magnitude lower than those of gluon and vector boson fusion. The presence of the $t\bar{t}$ pair in the final state can provide a good experimental signature and relevant information on the top–Higgs Yukawa coupling. The LO cross section of this production channel is plagued by large theoretical uncertainties due to the strong dependence on the renormalization scale of the strong coupling constant and on the factorization scales of the parton density functions inside the proton, respectively. For this cross section, NLO QCD calculations are available.

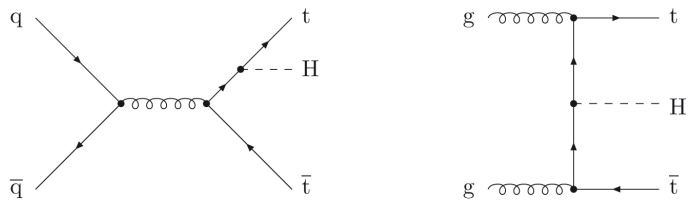


Figure 1.7: *Examples of LO Feynman diagrams for the partonic processes $q\bar{q}$, $gg \to Ht\bar{t}$.*

1.5.2 Higgs Decays

The partial decay width for a general process $a \to b_1 + b_2$ can be written as

$$\Gamma(a \to b_1 + b_2) = \frac{(2\pi)^4}{2m_a} \int \frac{d^3\vec{p}_1}{(2\pi)^3 2E_1} \frac{d^3\vec{p}_2}{(2\pi)^3 2E_2} \delta^4(p_a - p_1 - p_2)|M|^2,$$

where p_a is the four-momentum of the particle a, $p_i = (\vec{p}_i, E_i)$ is the four-momentum of the ith-particle produced in the decay and $|M|^2$ is the squared modulus of the process matrix element.
There are three classes of Higgs boson decays: into massless gauge bosons ($H \to \gamma\gamma$), massive gauge bosons ($H \to VV$) and massive fermions ($H \to f\bar{f}$).
The $H \to \gamma\gamma$ partial width is dominated by top and W boson vector loop

contributions and its expression is

$$\Gamma(H \to \gamma\gamma) = \frac{G_F \alpha_W^2 m_H^3}{128\sqrt{2}\pi^3} \left| \sum_f N_c Q_f^2 A_{1/2}^H(\tau_f) + A_1^H(\tau_W) \right|^2,$$

with

$$\tau_i = \frac{m_H^2}{4m_i^2},$$

$$A_{1/2}^H(\tau) = 2[\tau + (\tau - 1)f(\tau)]\tau^{-2} \ (top),$$

$$A_1^H(\tau) = -[2\tau^2 + 3\tau + 3(2\tau - 1)f(\tau)]\tau^{-2} \ (W),$$

where

$$f(\tau) = \begin{cases} arcsin^2\sqrt{\tau} & \tau \leq 1 \\ -\frac{1}{4}\left[log\frac{1+\sqrt{1-\tau^{-1}}}{1-\sqrt{1-\tau^{-1}}} - i\pi\right]^2 & \tau > 1 \end{cases}$$

It is thus proportional to m_H^3.

The $H \to W^+W^-$ and $H \to ZZ$ partial widths can be written as follows

$$\Gamma(H \to W^+W^-) = \frac{g^2}{64\pi}\frac{m_H^3}{m_W^2}\sqrt{1 - x_W}\left(1 - x_W + \frac{3}{4}x_W^2\right)$$

$$\Gamma(H \to ZZ) = \frac{g^2}{128\pi}\frac{m_H^3}{m_Z^2}\sqrt{1 - x_Z}\left(1 - x_Z + \frac{3}{4}x_Z^2\right),$$

where

$$x_W = \frac{4m_W^2}{m_H^2}, \quad x_Z = \frac{4m_Z^2}{m_H^2}.$$

Finally, the partial width of the $H \to f\bar{f}$ process is

$$\Gamma(H \to f\bar{f}) = \frac{G_F N_c}{4\sqrt{2}\pi} m_H m_f^2 \beta_f^3,$$

where

$$\beta_f = 1 - \frac{4m_f^2}{m_H^2}.$$

A strong phase space suppression at threshold $\beta_f \to 0$ can be noted.

The total width of the Higgs boson resonance is given by the sum of the partial widths of all possible decay channels and it is shown in Figure 1.8 as a function of m_H . Below the $2m_W$ threshold, the Higgs boson width is of the order of the MeV/c^2. Then it rapidly increases, but remains below 1 GeV/c^2 up to $m_H \sim 200$ GeV/c^2. In the high mass region, $m_H > 2m_Z$,

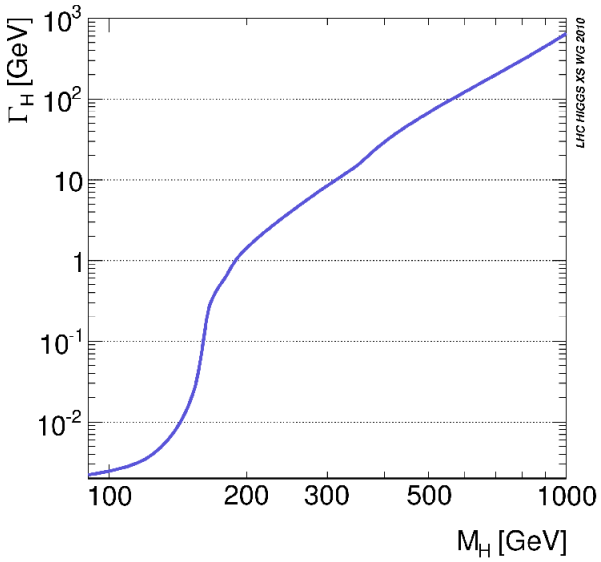


Figure 1.8: *Total decay width of the Higgs boson as a function of its mass [6].*

the total Higgs boson width is dominated by the W^+W^- and ZZ partial widths.
As the mass grows, x_W and x_Z tend to 0 and the width grows proportionally to m_H^3. Summing over the W^+W^- and ZZ channels, the Higgs resonance width in the high mass region can be written as

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

where v is the vacuum energy value ($\sim 264\ GeV$).
From this equation, it can be noted that $\Gamma_H \simeq m_H$ for $m_H \simeq 1.4\ TeV$. If m_H is larger than about 1 TeV/c^2, therefore, it becomes experimentally very problematic 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 a particle anymore. Different strategy for the Higgs boson identification can be developed depending on its mass, as decay modes change dramatically across the mass range. In Figure 1.9 the branching ratios (BRs) are shown as a function of the Higgs boson mass. Fermion decay modes dominate the BR 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 become dominant. A peak in the $H \to W^+W^-$ decay 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, above 350 GeV/c^2, also $t\bar{t}$ pairs can be produced [4] [5].

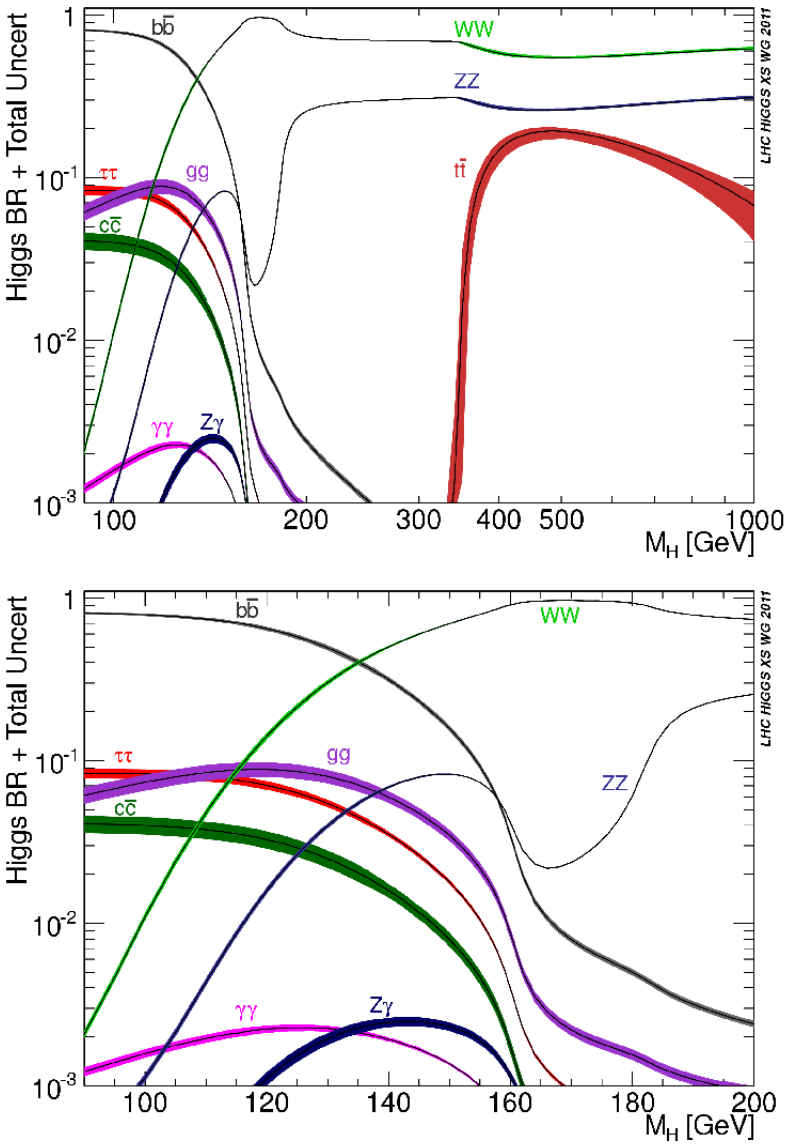


Figure 1.9: *Branching ratios of several Higgs boson decay channels as a function of the Higgs boson mass [6].*

1.6 Higgs Searches at the LHC

The most promising decay channels for the Higgs boson exclusion or discovery do not only depend on the corresponding branching ratios, but also on the capability of experimentally detecting the signal while rejecting backgrounds. To understand the sensitivity of each channel is useful to multiply production cross section by the branching ratio. These values

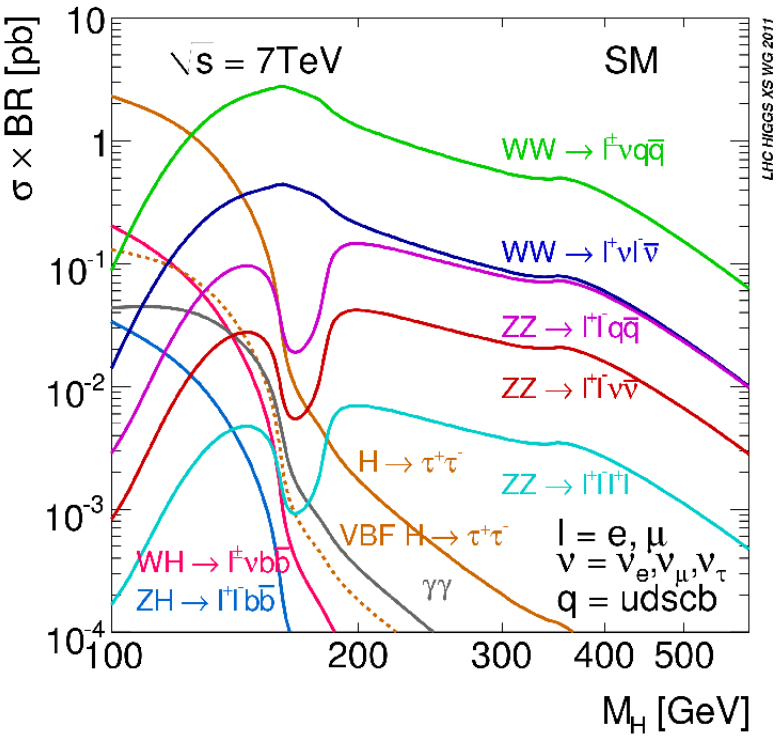


Figure 1.10: *Standard Model Higgs boson production cross section times branching ratio as a function of the Higgs boson mass [6].*

are compared in Figure 1.10 for the main channels that can be explored at the LHC experiments on the basis of the final state reconstruction efficiency and the signal-over-background ratio. The strategy for the search of the SM Higgs boson in the 100 GeV/c^2 - 1 TeV/c^2 mass range, can be described dividing this range into three mass regions. For each of these regions the most important channels are presented.

1.6.1 Low Mass Region ($m_H \leq 135\ GeV/c^2$)

$H \to b\bar{b}$

For $m_H < 135\ GeV/c^2$, the branching ratio favors the $b\bar{b}$ decay. The $b\bar{b}$ decay is explored in association with a W or a Z boson (Higgsstrahlung) and the following final states are examined:

- $H \to b\bar{b},\ W \to \mu\nu$;
- $H \to b\bar{b},\ W \to e\nu$;
- $H \to b\bar{b},\ Z \to \mu\mu$;

- $H \to b\bar{b},\ Z \to ee$;
- $H \to b\bar{b},\ Z \to \nu\nu$.

The event has a clear signature despite the lower cross section $\times$ BR. However, the process $gg \to H \to b\bar{b}$ is experimentally very difficult to explore because of the enormous background from di-jet production (with a cross section more than six orders of magnitude larger than the signal one). The $t\bar{t}$ associated production (Figure 1.7) allows anyway some results. The following final state are considered:

- the fully leptonic decay
$$H \to b\bar{b},\ t \to W(\ell\nu)b,\ \bar{t} \to W(\ell\nu)\bar{b};$$
- the semi-leptonic decay
$$H \to b\bar{b},\ t \to W(q\bar{q})b,\ \bar{t} \to W(\ell\nu)\bar{b};$$
- the fully hadronic decay
$$H \to b\bar{b},\ t \to W(q\bar{q})b,\ \bar{t} \to W(q\bar{q})\bar{b}.$$

This signal has a clear signature due to the presence of a high-momentum lepton from one of the two W decays, missing energy and four b-tagged jets. Because of the presence in the signal of jets, a wide category of multi-jet backgrounds contributes.

$H \to \tau\tau$

Similar results are obtained exploring the $H \to \tau\tau$ decay. Also in this case, the Higgs boson signal produced via the gluon-gluon fusion process is overwhelmed by the Drell-Yan production of τ pairs in the mass region of interest. To overcome this, the analysis relies on the VBF production of Higgs bosons. The two forward "tagging" jets from the incoming quarks which radiate the vector bosons provide means for distinguishing the Higgs boson production from SM background processes. Three independent τ pair final states where one or both taus decay leptonically are studied: $e\tau_h$, $\mu\tau_h$, $e\mu$, where the symbol τ_h is used to indicate a reconstructed hadronic decay of a τ.

$H \to \gamma\gamma$

For the Higgs boson search at low mass, the $H \to \gamma\gamma$ channel is one of the more promising. Despite its small BR, $\sim 0.2\%$ in the low mass region, the $H \to \gamma\gamma$ provides a clean final-state topology with an effective mass peak that is reconstructed with great precision. The dominant sources of background are the irreducible direct di-photon production and the reducible $pp \to \gamma + jet$ and $pp \to jet + jet$.

$H \to W^+W^-$

From Figure 1.9 it is clear that, despite not favored by the BR, the $H \to W^+W^- \to 2\ell 2\nu$ channel can play a role even in this low mass region. 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.

$H \to ZZ$

Among the ZZ channels that are explored, despite the $\sigma \times BR$ (Figure 1.9) is higher for the $2l2q$ and $2\ell 2\nu$ final states (ℓ, $\ell' = e$ or μ), the $H \to ZZ^{(\star)} \to 4\ell$ 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 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

1.6.2 Intermediate Mass Region ($140 \leq m_H \leq 180\ GeV/c^2$)

For mass values between $140 \leq m_H \leq 180\ GeV/c^2$, the Higgs boson decays into $WW^\star$ and $ZZ^\star$ become possible and their branching ratios rapidly increase. The preferred channels in this mass range are thus $H \to WW^\star \to 2\ell 2\nu$ and $H \to ZZ^\star \to 4\ell$, where just one vector boson is on-shell.

$H \to WW^*$

The branching ratio of $H \to WW^\star \to 2\ell 2\nu$ is larger, because of the stronger coupling of the Higgs boson to charged currents than to neutral currents. Moreover, this decay mode becomes particularly important in the mass region between $2m_W$ and $2m_Z$, where the Higgs particle can decay into two on-shell W bosons and not yet into two on-shell Z bosons and its branching ratio is close to unity. This directly reflects in a reduced sensitivity of the 4ℓ channel for the Higgs boson exclusion around $m_H = 160\ GeV/c^2$.
In this channel it is not possible to reconstruct the Higgs invariant mass, because of the presence of the two neutrinos. The experimental signature

of these events is very clear: the final state presents 2 isolated high p_T leptons pointing to the primary vertex, high missing energy and limited hadronic activity.

$H \to ZZ^*$

Despite its lower branching ratio, the $H \to ZZ^* \to 4\ell$ channel is the favored for the Higgs discovery in this mass region. It offers a very clear experimental signature and high signal to background ratio. Furthermore, it allows to reconstruct the Higgs boson mass with high precision. Therefore, this channel seems to be the best signature for a Higgs boson discovery in this mass range.

1.6.3 High Mass Region ($m_H \leq 180\ GeV/c^2$)

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, the $H \to ZZ$ channels play a major role. The 4ℓ decay is the "golden channel" for a high mass Higgs boson discovery. Moreover for masses higher then 250 GeV/c^2 the $H \to ZZ \to 2\ell 2\nu$ and $H \to ZZ \to 2\ell 2q$ can provide an important contribution to the combined sensitivity to the Higgs boson exclusion.

Recently, a new particle has been discovered by the CMS and ATLAS experiments at LHC around a mass of 125 GeV/c^2, with a significance of about 5σ (it will be discussed in detail in Chapter 3, focusing on CMS results). In order to determine if it is or not the Higgs boson, measuring the new boson properties (such as spin and CP quantum number) is necessary.

Chapter 2

The CMS Experiment at LHC

The present chapter gives a short description of the Large Hadron Collider and focuses on the Compact Muon Solenoid (CMS) experiment, one of the four main detectors installed on the LHC beamline. Together with the experiment hardware setup, the CMS reconstruction software and the main CMS trigger features are also described.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [7] is the world's major particle accelerator, designed to collide opposing particle beams of either protons and lead ions. It is built in order to find an answer to the most fundamental questions of physics, like understanding the electroweak symmetry breaking for which the Higgs mechanism is presumed to be responsible. It is an unprecedented machine in terms of energy, luminosity, size, complexity of experiments, cost and involvement of human resources.

The collider is placed in a circular tunnel 27 *Km* long, situated at a depth

Figure 2.1: *View of the CERN site: the LHC tunnel, with the four experiments located at each interaction point.*

of about 100 m underground, on the French-Swiss border. This tunnel was built between 1983 and 1988 for the Large Electron-Positron Collider (LEP) and it contains two adjacent and parallel beam pipes, where proton (or ion) beams travel in opposite directions around the ring. Since collisions occur between particles of the same charge, the two acceleration cavities are required to have different magnetic configurations. The two beams cross in four different points in correspondence to the four main experimental detectors: ATLAS, ALICE, LHCb and CMS (see Figure 2.1). The LHC is designed to collide two beams of protons, each one with a nominal energy of 7 TeV, for a total center-of-mass energy of 14 TeV. It is currently running at a beam energy up to 4 TeV (i.e. center-of-mass energy of 8 TeV). Moreover, for the ion physics program, shorter running periods with lead nuclei collisions are performed.

One of the most important characteristic of a collider is its luminosity $\mathcal{L}$, that ties event rate to the cross section of a process:

$$\frac{dN_i}{dt} = \sigma_i \mathcal{L}(t)$$

(where the "i" subscript indicates a general process). Assuming a Gaussian beam shape, the luminosity at LHC depends only on the accelerating machine parameters and it can be written as:

$$\mathcal{L} = f \frac{n_b N_1 N_2}{4\pi\sigma_x\sigma_y}$$

where f is the revolution frequency of the n_b proton bunches per beam, N_1 and N_2 numbers of protons in the colliding bunches, σ_x and σ_y the beam profiles in horizontal and vertical directions at the interaction point.
The designed peak of instantaneous luminosity for ATLAS and CMS is $10^{34}\ cm^{-2}s^{-1}$, but before reaching this value the LHC will have lower luminosity runs. The total integrated luminosity delivered by LHC in 2011 to CMS is 5.73 fb^{-1}, 5.22 fb^{-1} of which have been recorded by the experiment (see Figure 2.2). In Figure 2.3 is reported the luminosity delivered by LHC and recorded by CMS in 2012, up to the end of August.
The main parameters of the LHC are summarized in table 2.1.

2.2 The Compact Muon Solenoid Experiment

CMS [8] is one of the four main experiments at LHC. Its primary goals are the study of the electroweak symmetry breaking mechanism linked to the Higgs mechanism, the search for physics Beyond the Standard Model (BSM) and the precision measurements of already known physics processes. In order to fulfill these goals, excellent lepton reconstruction and

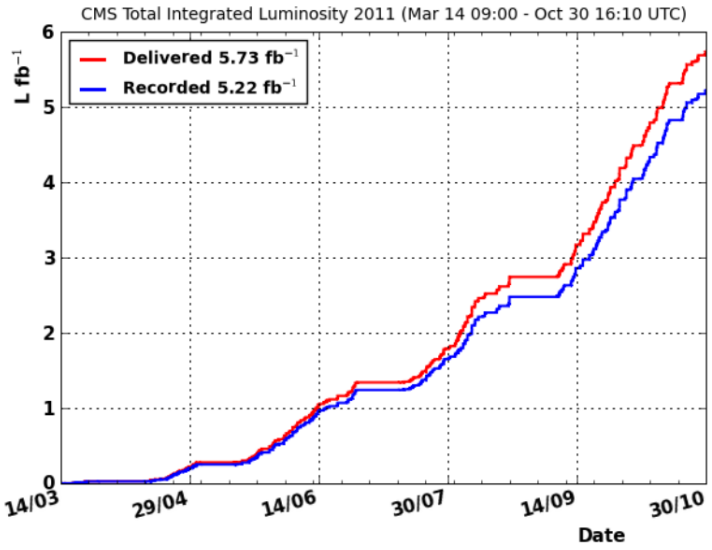


Figure 2.2: *Integrated luminosity delivered by the LHC (in red) and recorded by CMS (in blue) in 2011 proton-proton collisions.*

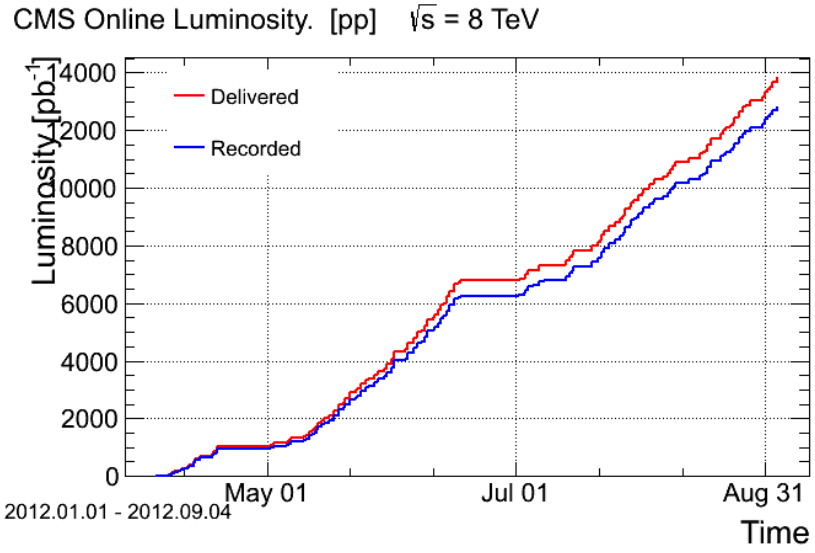


Figure 2.3: *Integrated luminosity delivered by the LHC (in red) and recorded by CMS (in blue) in 2012 proton-proton collisions.*

particle identification are required. A particular magnetic field configuration is necessary to provide a large bending power and to have thus good resolution in measuring charge particles of high momentum. The main

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

Parameter	$p-p$	$Pb-Pb$
Circumference $[km]$	26.659	
Beam radius at interaction point $[\mu m]$	15	
Dipole peak field $[T]$	8.3	
Design center-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

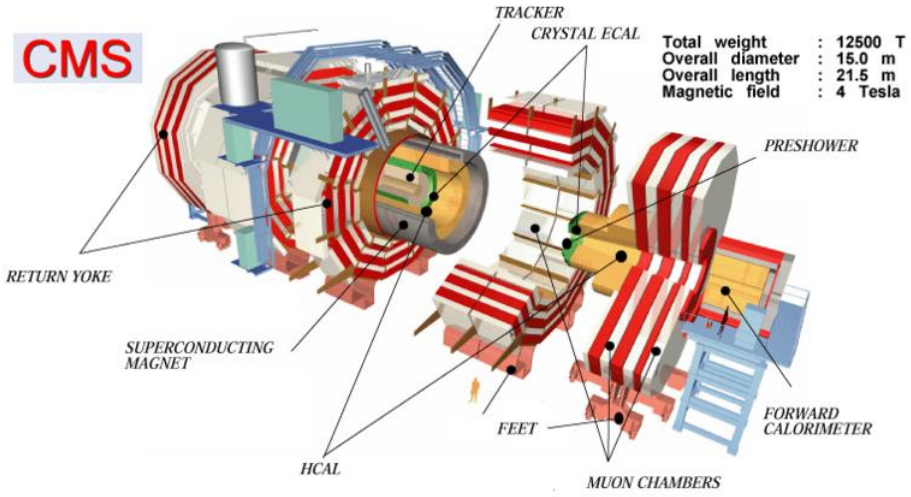


Figure 2.4: *Schematic picture of CMS experiment at the LHC.*

feature of the CMS detector is the 3.8 T superconducting solenoid, which allows a compact design with a strong magnetic field, besides a robust and redundant muon system, a fully active scintillating crystals-based electromagnetic calorimeter and a high-quality tracking system.
The overall structure of CMS consists of several cylindrical layers coaxial to the beam axis (the *barrel layers*), closed at both ends by detector disks orthogonal to the beam direction (the *endcaps*), to ensure optimal hermeticity. The dimensions of the overall detector are fairly remarkable: the total length is 21.5 m, the diameter 15 m and the weight about 12500 tons. A

schematic view of the CMS detector is shown in Figure 2.4.

2.2.1 Definition of Kinematic Variables

The CMS experiment uses a cylindrical coordinate system, centered at the interaction point. The direction of the z axis is chosen along the beam and it is referred as *longitudinal*. The x axis is horizontal and pointing towards the center of the LHC ring and the y axis is vertical and pointing upwards. The $x-y$ plane, orthogonal to the beam line, is called *transverse plane*. According to these definitions, the momentum of a particle can be divided in two components: the *longitudinal momentum* p_z and the *transverse momentum* p_T, defined as

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

The *rapidity* of a particle of energy E is defined as

$$y = \frac{1}{2}\, ln \frac{E + p_z}{E - p_z}$$

and for high-energy particles it can be approximated by the *pseudorapidity*

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

which only depends on the polar angle θ of the particle momentum, measured with respect to the z axis. The azimuthal angle, measured in the $x-y$ plane from the x axis, is called ϕ.

In the following, every component of the CMS detector is briefly presented, focusing on the performances and on their impact on physics analysis.

2.2.2 The Tracker

The Tracker (Figure 2.5) is the CMS innermost subdetector and the closest to the interaction point. Its main purpose is the reconstruction of charged tracks and vertices, together with the measurement of charged particles p_T, that are important tools for identifying signal events and rejecting background.
It extends in the region of $|\eta| < 2.5$, $r < 120\ cm$, $|z| < 270\ cm$ and it is completely based on semiconductor detectors made of silicon, that cover a total surface of 210 m^2.
Because of its tasks, the major physics requirements for the CMS Tracker

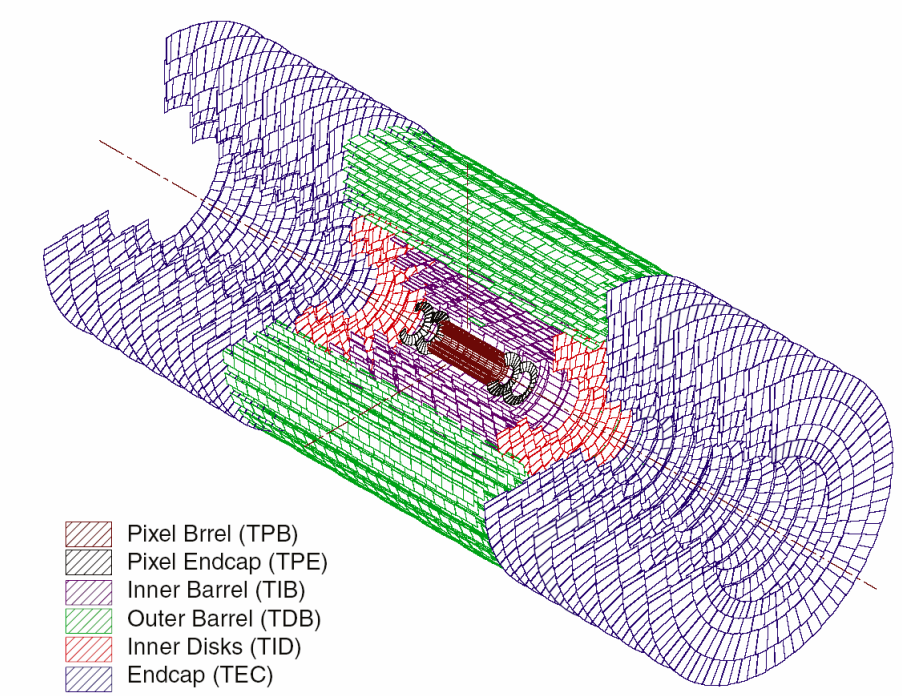


Figure 2.5: *View of the silicon tracker, including the pixel detector.*

are an efficient reconstruction of isolated lepton tracks (in all the pseudo-rapidity coverage is close to 100%), a good lepton momentum resolution (for $|\eta| < 2$, $\sigma(p_T)/p_T < 4\%$ for single muons in a large range of p_T) and a good reconstruction of interaction vertices.
For optimal pattern recognition, the tracker is designed to fulfill two basic properties: low cell occupancy and large hit redundancy. The low occupancy is obtained by high granularity detectors, mainly those closest to the interaction point (because of higher particle fluxes), and a fast primary charge collection, obtained using thin detectors and reducing the silicon bulks. The redundancy is guaranteed by the overall design which provides many measured points per track (on average 12-14, see Figure 2.6). All these characteristics allow for a high tracking efficiency and a low rate of fake tracks, i.e. reconstructed tracks that do not correspond to any real track.
Moreover, one of the major constraints of the Tracker design is to reduce the amount of material crossed by particles, to minimize electron bremsstrahlung and hadronic interactions.

The inner tracking system of the CMS detector has a concentric layout. Closest to the interaction point ($r \leq 10\ cm$), pixel detectors are placed (see

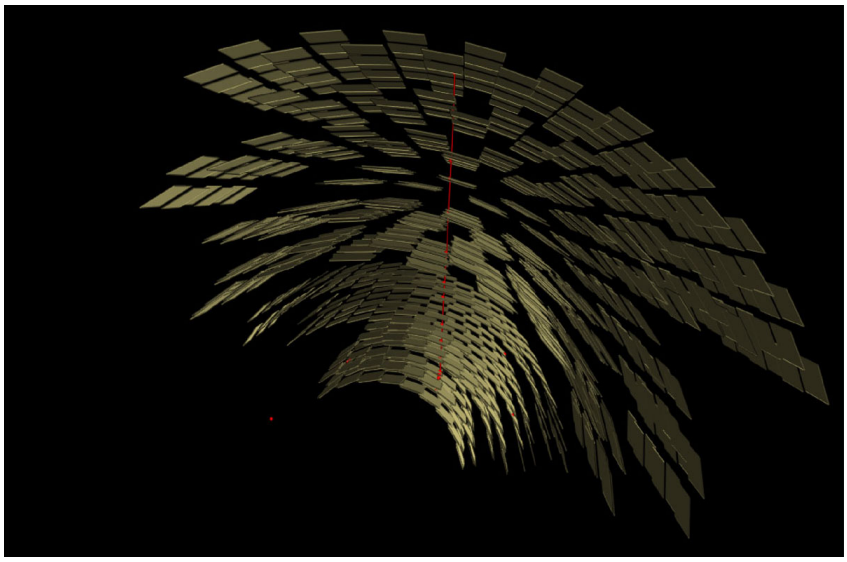


Figure 2.6: *Example of a particle track into the tracker.*

Figure 2.5), because of the high particle density. They consist of three barrel layers and two endcap disks for each side. In the region $20 < r < 55$ *cm* the particle flux decreases enough to allow the use of silicon microstrips (made of 4 barrel layers and 3 disks at each side), while in the outermost region of the tracker ($r > 55$ *cm*) larger pitch microstrips are placed (consisting of 6 barrel layers and 9 disks for each endcap).

2.2.3 The Electromagnetic Calorimeter

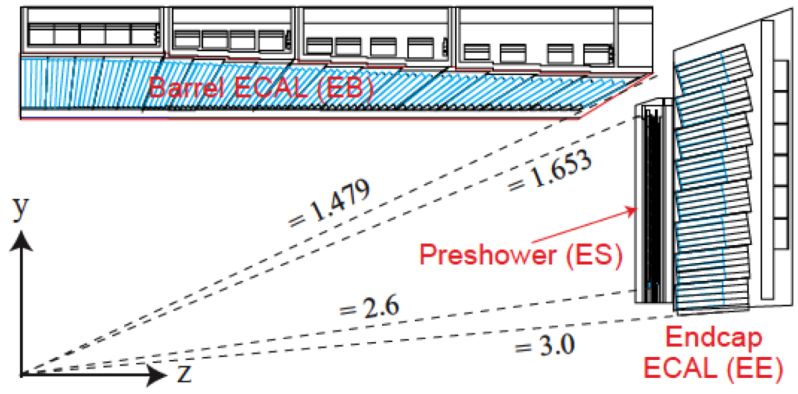


Figure 2.7: *Longitudinal view of one quarter of the CMS ECAL.*

The primary aim of the Electromagnetic Calorimeter (ECAL) is the identification and measurement of the energy of electrons and photons. The physics process that dictates the strictest requirements on its performance is the low mass Higgs decay into two photons $H \to \gamma\gamma$, a very promising channel for Higgs Searches (see Section 1.6.1). The goal is 1% resolution on the di-photon invariant mass and the natural choice to reach this goal is a homogeneous calorimeter.
ECAL is made of 74 848 lead tungstate ($PbWO_4$) crystals, chosen because of their excellent energy resolution. It is a transparent material characterized by a high density (8.28 g/cm^3), a short radiation length ($X_0 = 0.89\ cm$) and a small Molière radius (2.2 cm). These features allow a very compact shape and a fine granularity, necessary because of the high particle density produced at the LHC. Moreover, these crystals have a trapezoidal shape and they are characterized by a very short scintillation decay time, which permits to collect about 80% of the light within 25 ns, so that they can be used at the crossing rate of 40 MHz.
The ECAL barrel covers the central rapidity region ($|\eta| < 1.48$) and the two ECAL endcaps extend the coverage up to $|\eta| = 3$ (Figure 2.7).

2.2.4 The Hadron Calorimeter

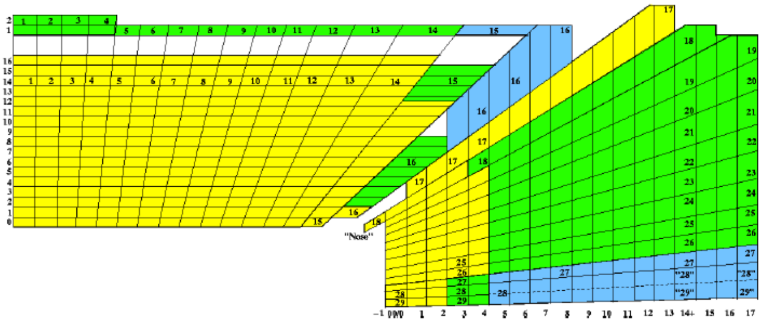


Figure 2.8: *$r - z$ scheme of Hadron Barrel and Hadron Endcaps.*

The Hadron Calorimeter (HCAL) surrounds ECAL and its design is strongly influenced by the choice of the magnet parameters and by the fact that most of it is placed inside the magnet coil. This detector plays an essential role in the identification and measurement of hadrons by estimating the energy and the direction of jets. It is also used for neutrino detection by measuring the missing transverse energy of the event. For this reason, one of the main design requirements is a high hermeticity. In particular, the HCAL angular coverage must include the very forward region, since the identification of forward jets is very important for the rejection of many backgrounds and the evidence of BSM signatures. The HCAL is a

sampling calorimeter, that meaning it finds particles position, energy and arrival time using alternating layers of "absorber" and "scintillator" materials. In particular, brass was chosen as absorber material, thanks to its non magnetic behavior and its quite short interaction length ($\lambda_I \sim 151\ mm$). To maximize the amount of absorber before the magnet, the space devoted to the active medium is minimized. The active part is made of scintillator tiles coupled with wavelength shifting fibers and clear fibers carrying the light to the readout system.
The Hadronic Calorimeter can be divided in four parts, that permit a good segmentation, a moderate energy resolution and a full angular coverage. The *barrel hadronic calorimeter* (HB) surrounds the electromagnetic calorimeter and covers the central pseudorapidity region up to $|\eta| = 1.3$. The endcap regions are covered up to $|\eta| = 3$ by the two *endcap hadron calorimeters* (HE). The HB and HE are located inside the solenoid magnet. To satisfy the hermeticity requirements, *forward calorimeters* are placed outside the magnet yoke, 11 m far from the interaction point, extending the pseudorapidity coverage up to $|\eta| = 5$. Finally, an array of scintillators located outside the magnet, which is referred to as the *outer hadronic calorimeter* (HO), is used to improve the central shower containment.

2.2.5 The Magnet

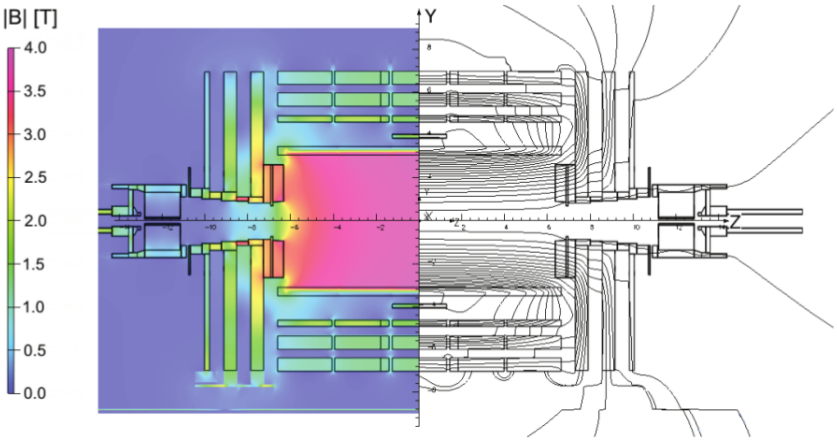


Figure 2.9: *Layout of the magnetic field of CMS.*

In order to achieve a good momentum resolution for momenta up to 1 TeV/c, CMS needs a strong solenoidal magnetic field. The CMS magnet is the central device around which the experiment is built and its dimensions limited the size of the total apparatus.

Its purpose is to bend the paths of particles emerging from high-energy collisions. The higher particle momentum is, the less its trajectory is curved by the magnetic field. A higher strength field, combined with high-precision position measurement in the tracker and muon system, gives accurate measurement of momentum.
The CMS magnet is a 13 m long superconducting solenoid, the largest ever built. It is able to generate a uniform magnetic field of 4 T in the inner region, storing about 2.5 GJ of energy (Figure 2.9). It operates with a superconducting technology at a temperature of 4 K, ensured by a helium cooling system. The magnet is then contained in a huge vacuum cylinder, which isolates it from the external environment. Outside, an iron structure constitutes the iron yoke, necessary to bridge the return magnetic flux, which otherwise would get lost, disturbing the surrounding environment.

2.2.6 The Muon System

Because muons can penetrate several meters of iron without being stopped, they are detected by muon chambers, placed in the outermost part of the experiment. The muon detection system is therefore set outside the magnetic coil and it has multiple tasks: triggering on muons, identifying them and assisting the tracker in measuring their momentum.
After being detected by the inner tracker, muons are detected again in the muon system. This redundancy allows a sensible improvement in the resolution of high-momentum particles, for which spatial resolution dominates, while for lower momenta (up to p_T values of 200 GeV/c) the inner tracker is sufficient. This is due to multiple scattering in the material before the chambers. The trajectory beyond the return yoke can be then extrapolated back to the beam-line, thanks to the compensation of the bend before and after the coil. The minimum value of the muon transverse momentum required to reach the system is 3 GeV/c.

The spectrometer is composed by three independent subsystem, i.e. three different types of gaseous particle detectors: *Drift Tube* (DT) *chambers* and *Cathode Strip Chambers* (CSC) are used in the regions of $|\eta| < 1.2$ and $0.9 < |\eta| < 2.4$ respectively, and are complemented by a system of *Resistive Plate Chambers* (RPC) covering the range of $|\eta| < 1.6$. The detector can thus be divided into three regions, referred to as *barrel* ($|\eta| < 0.9$), *overlap* ($0.9 < |\eta| < 1.2$) and *endcap* ($|\eta| > 1.2$). The reason for these different technologies lies in the different particle rates and occupancies, both higher in the endcaps, and in the intensity of the stray magnetic field, which is lower in the barrel. Thanks to the fast response they provide, all the muon detectors are also used within the first level of the trigger system.

In the barrel, the expected occupancy is low, allowing for the use of drift

tubes as detection element. In this region the detector stations are interleaved with the iron plates of the yoke. The chamber segmentation follows that of the iron yoke, consisting of five wheels along the z axis, each one divided into 12 azimuthal sectors. As one can see in Figure 2.10, in each sector the chambers are organized in four stations (labeled $MB1$, $MB2$, $MB3$ and $MB4$). There are 12 chambers in each of the 3 inner layers; in the 4th layer the top and bottom sectors host 2 chambers each, thus leading to a total of 14 chambers per wheel in this outermost layer. The DTs provide a precise track measurement in the bending plane; their maximum drift time is about 400 ns, with a time resolution of about 1 ns. Each chamber has a resolution of about 100 μm in $r\phi$ and 1 $mrad$ in ϕ. Each DT chamber has 1 or 2 RPCs coupled to it before installation, depending on the station. RPCs are a double-gap bakelite chambers operating in avalanche mode, with gap width of 2 mm.

In each of the endcaps, the cathode strip chambers are arranged in 4 disks perpendicular to the beam and in concentric rings, 3 rings in the innermost station and 2 in the others. Each CSC has a trapezoidal shape and is made of 6 gas gaps (7 layers) with planes of cathode strips in the radial directions and anode wires almost perpendicular to the strips. Most CSCs are overlapped in ϕ in order to avoid gaps in acceptance. The ionization of a charged particle passing through the planes causes the charge to be collected on the anode wire and the image charge on the cathode strips, thus allowing to get (r, z, ϕ) hits in each layer.

2.3 The CMS Trigger System

At the instantaneous luminosity expected in proton-proton collisions at LHC, the interaction rate ($\sim$ 40 MHz) produces 10^9 interactions/sec, but only a small part of them can be written to archival media. Such a high luminosity causes several interactions overlap in the same bunch crossing, effect known as *pile-up*, as well as the overlap in the detector of signals from different bunch crossings, due to the limited speed of detector response and read-out. There are therefore several technical difficulties in handling, storing and processing this huge amount of data and a selection of events written to permanent storage is thus necessary. As a consequence, one of the most important and difficult aspects of the experiment is the trigger, the real-time selection and recording of useful events.
The data size per event is $\sim$ 1 MB and just a rate of 300 Hz can be stored for off-line analysis. In order to fulfill this data reduction, the CMS trigger is divided in two levels: the *Level-1 Trigger* (L1), mainly based on hardware, and the *High Level Trigger* (HLT), software-based.
The following sections briefly describe the design of the L1-Trigger and

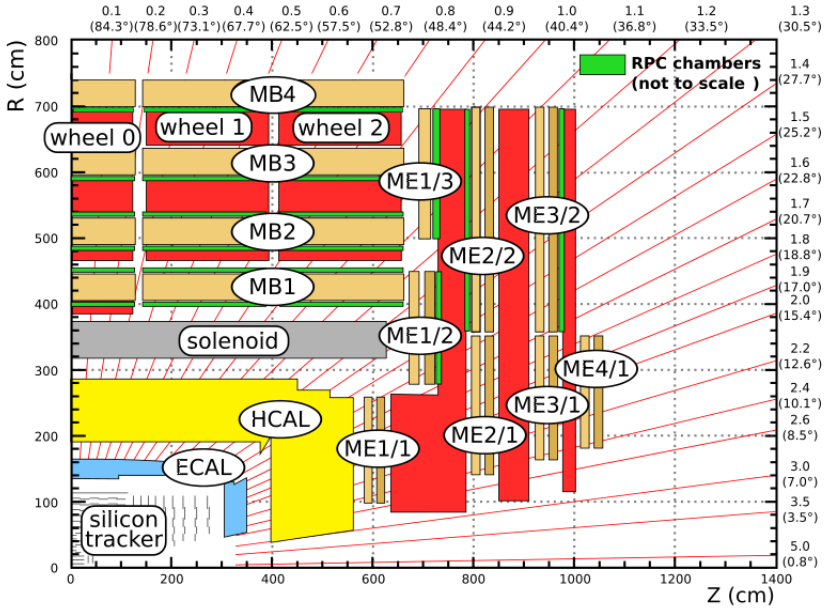


Figure 2.10: *An $r - z$ cross-section of a quadrant of the CMS detector. The locations of the various muon stations and the steel disks (red areas) are shown: four DTs stations are labeled MB ("muon barrel") and CSCs are labeled ME ("muon endcap"). RPCs (in green) are both the barrel and the endcaps of CMS.*

the general structure of the HLT.

2.3.1 The Level-1 Trigger

The Level-1 trigger is based on custom-made hardware and its task is to take an accept-reject decision for each bunch crossing, every 25 ns, reducing the rate of selected events down to 100 kHz. The Level-1 decision must be fast and it is therefore taken within a fixed time of 3.2 μs. This time includes also the transit of data from front-end electronics to Level-1 Trigger processors and back and the time available for calculations can thus be no more than 1 μs.
Because of the bunch crossing high rate and slowness of the tracker algorithms, the Level-1 Trigger uses only coarsely segmented data from calorimeters and muon detectors, while all the high-resolution data is stored in pipelines of processing elements, waiting for the trigger decision. If the L1 accepts the event, the data are moved to be processed by the High Level Trigger.
The calorimeter and the muon subdetectors provide objects called "trigger primitives" in the form of local energy deposits in calorimeter trigger towers (identifying electron, photon and jet candidates) and track segments

or hits in muon chambers (identifying muon candidate). These objects are than ranked and stored and they form the basis for trigger decisions taken by the final L1 stage, the *Global L1 Trigger*, according to programmable algorithms. Finally, the *Trigger Control System* determines if the subdetectors are ready to read out the event and if the data acquisition (DAQ) system is ready to receive it.

2.3.2 The High Level Trigger

The High Level Trigger accepts all events that pass the L1 Trigger filter, using detector data transferred from the pipelines to the front-end readout buffers. The reconstruction and selection in the HLT takes place in steps and uninteresting events are rejected as quickly as possible. The basic idea of the HLT trigger software is to reconstruct those part of each physics object that can be used for selection, while minimizing the overall CPU usage. The first selection step in the HLT process, called *Level-2 Trigger*, has the full Level-1 rate as input and uses only information from the calorimeter and muon detectors. On the other hand, *Level-3 Trigger*, the second step of the selection, includes the reconstruction of full tracks in the tracker. In order to minimize the CPU time required by the HLT, a key feature of the algorithms is to reconstruct the information in the CMS detector only partially, taking the accept-reject decision on the reconstruction of quantities in a limited region of the detector. The HLT software system reduces the event rate from 100 kHz to 300 Hz.

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 associated with the field responsible for the spontaneous electroweak symmetry breaking. The mass m_H of this scalar boson is a free parameter of the theory.
The Higgs boson production followed by the decay $H \to ZZ \to \ell^{\pm}\ell^{\mp}\ell'^{\pm}\ell'^{\mp}$ with ℓ, $\ell' = e$ or μ (in short $H \to 4\ell$) is expected to be one of the main decay channels for the Higgs boson discovery or exclusion over a wide range of m_H values.

3.1 The $ZZ \to 4\ell$ Final State

In the following 4ℓ final state signal and background processes are described in detail.

3.1.1 Signal

In the analysis of the $H \to 4\ell$ channel, all Higgs boson production mechanisms (see Section 1.5.1) are considered as part of the signal. The main diagrams associated to the signal are those described in Figure 3.1. The final states considered are: $2\mu 2e$, 4μ and $4e$. The first has the biggest BR while the second is the cleanest one. The main concern of the last channel is the presence, for low Higgs masses, of very soft electrons, well below the range for which the reconstruction will be best controlled via single Z and W measurements.

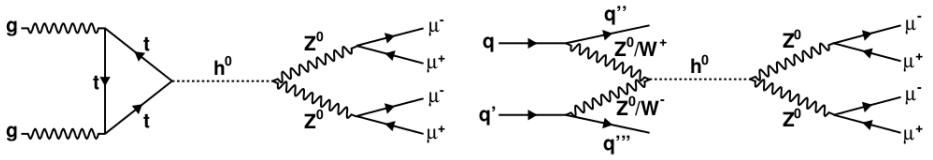


Figure 3.1: *Lowest order signal diagrams: with the Higgs boson produced via gluon-gluon fusion (left) and via vector boson fusion (right).*

The signal most important characteristic is that the four leptons originate from the decay chain of a single particle, the Higgs boson. Their invariant mass thus peaks at the Higgs mass, while for the other backgrounds this does not happen. Signal events can be also identified for the presence of two Z bosons in the intermediate state. Depending on the Higgs mass, the two bosons could be either real or virtual (off mass shell) and the fraction of virtual Z decreasing with the increase of the Higgs mass.

3.1.2 Backgrounds

The sources of background for the $H \to ZZ^{(\star)} \to 4\ell$ channel are events with four high p_T leptons in the final state, mainly coming from $ZZ^{(\star)}$, $Zb\bar{b}$, $t\bar{t}$ and $Z + light\ jets$ events.

Irreducible Background

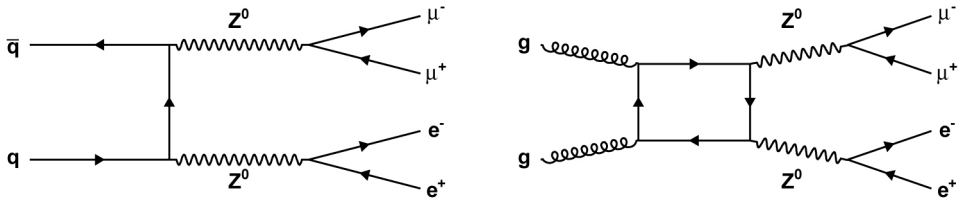


Figure 3.2: *Lowest order diagrams for the $qq \to ZZ^{(\star)}/Z\gamma^{(\star)}$ process (left) and for the $gg \to ZZ^{(\star)}/Z\gamma^{(\star)}$ process (right).*

Four-lepton events from non resonant di-boson production constitute the main source of background events. It belongs to the *irreducible background* category, as the event topology and kinematic is very similar to those of signal events. The lowest order production mechanism is the one represented in Figure 3.2 (left), $qq \to ZZ^{(\star)}/Z\gamma^{(\star)}$. The gluon-induced ZZ background, although technically of NNLO 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 3.2 (right).

Reducible Backgrounds

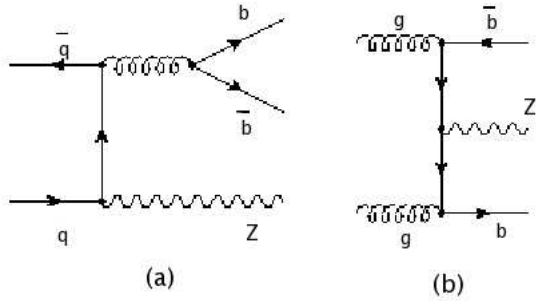


Figure 3.3: *Lowest order diagrams for the $qq/gg \to Zb\bar{b}$ process.*

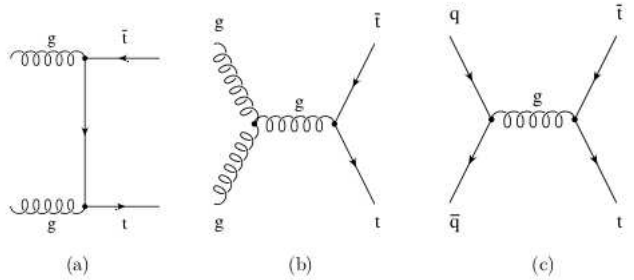


Figure 3.4: *Lowest order diagrams for the $qq/gg \to t\bar{t}$ process.*

The reducible backgrounds have very pronounced kinematic and topological differences with respect to the signal. Therefore by applying appropriate kinematic and topological cuts, these backgrounds can be effectively suppressed. The largest reducible background processes are $Zb\bar{b}$ (and $Zc\bar{c}$), with $Z \to \ell^+\ell^-$ (Figure 3.3), and the production of top quark pairs in the decay $t\bar{t} \to W^+bW^-\bar{b} \to \ell^+\ell^-\nu\bar{\nu}b\bar{b}$ (Figures 3.4 and 3.5). B

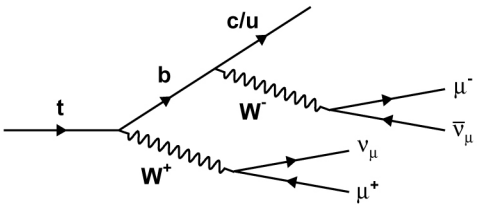


Figure 3.5: *Most probable decay chain with which a quark top can produce two opposite sign muons.*

mesons can decay semi-leptonically in three different ways:

- direct decay $b \to \ell$ with $BR \sim$ 10.7%;
- cascade decay $b \to c \to \ell$ with $BR \sim$ 8%;
- "wrong sign" cascade decay $b \to \bar{c} \to \ell$ with $BR \sim$ 1.6%.

Thus these processes with two b decaying leptonically can lead to 4ℓ final state events. Since $Zb\bar{b}$ and $t\bar{t}$ events are constituted by two leptons coming from the Z decay and other two originate from cascade decays of hadrons, this background presents two non-isolated leptons. It can be thus easily separated from leptons originating by W and Z decays, as it will be discussed in Section 3.5.2.

Instrumental Background

The category of instrumental backgrounds is finally used to indicate background events with final state leptons from mis-identification of other particles, such as QCD multi-jets and $Z/W + light\ jets$ processes where leptons mainly come from jets faking leptons. More precisely this is the general case for electrons, while reconstructed muons in these processes, in addition to those from the Z and W decay, mainly come from decay in flight of light primary hadrons.

3.2 Datasets

3.2.1 Experimental Data

The analysis is designed for a Higgs boson in the mass range $110 < m_H < 600\ GeV/c^2$. It uses the data collected at $\sqrt{s} = 7\ TeV$ by CMS in 2011, corresponding to an integrated luminosity of 5.05 fb^{-1}, combined with data collected in 2012 at $\sqrt{s} = 8\ TeV$, corresponding to an additional 5.26 fb^{-1}. The CMS standard selection of events is applied, which requires high quality data with a good performance of the different sub-detectors. Thus, similar detector operation conditions are imposed in the validation of the data to be used for the analysis of the $4e$, 4μ and $2e2\mu$ channels. The absolute pp luminosity is known with a precision of 2.2% in 2011 and 4.4% in 2012 [9].
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, in order to recover few percent of inefficiency in the $2e2\mu$ channel at low Higgs boson masses. The requirements on the transverse energy (transverse momenta) for the first and second lepton are 17 and 8 GeV 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 $m_H > 120\ GeV/c^2$ [9].

3.2.2 Simulated Samples

SM Higgs boson signal samples and samples for background processes have been obtained using Monte Carlo (MC) simulations. All datasets were subject to full reconstruction. The signal and background samples have been used for the optimization of the event selection strategy, prior to the analysis of the experimental data. They are further used in this analysis for the comparison with the measurements, 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 fusion ($gg \to H$) and vector boson fusion ($qq \to qqH$) is generated with POWHEG at next-to-leading order (NLO) precision. Additional samples of WH, ZH and ttH events are generated with PYTHIA. Events at generator level are re-weighted 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 precision. The total cross section is scaled by the branching fraction $\mathcal{B}(H \to 4\ell)$ calculated with PROPHECY4F, which includes NLO QCD and electroweak corrections and all interference effects at NLO, in particular effects specific to the $4e$ and 4μ channels, due to the presence of identical leptons in the final states. These effects are only relevant at low masses, below the ZZ threshold, where at least one of the Z bosons is off-shell and therefore the phase-space for interference increases [10]. The SM background contribution from ZZ production via qq is generated at NLO with POWHEG, while other di-boson processes (WW and WZ) are generated with MADGRAPH with cross sections rescaled to NLO predictions. The $gg \to ZZ$ contribution is generated with GG2ZZ. 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. The generation 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. For leading-order generators, the default set of parton distribution functions (PDF) used to produce these samples is CTEQ6L, while CT10 is used for NLO generators. All generated samples are interfaced with PYTHIA for the final hadronization. All events are processed through a detailed simulation of the CMS detector, based on GEANT4, and are reconstructed with the same algorithms that are used for data.

3.3 Physics Objects

A complete reconstruction of the individual particles emerging from each collision event is obtained via a particle-flow (PF) technique, [10] using the information from all CMS sub-detectors in order to identify and reconstruct individual particles in the collision event [11]. These particles are classified into mutually exclusive categories: charged hadrons, neutral hadrons, photons, muons and electrons. In this section leptons and photons categories are examined.

3.3.1 Leptons

The reconstruction of the SM Higgs boson in the decay chain $H \to ZZ^{(\star)} \to 4\ell$ imposes high-performance lepton reconstruction, identification, isolation and 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"). The precise momentum measurement leads to a precise measurement of $m_{4\ell}$, the most discriminating observable for the Higgs boson search.

Electron Reconstruction and Identification

The electron reconstruction combines ECAL and tracker information. Electron candidates are reconstructed from clusters of energy deposits in ECAL, which are then matched to tracks reconstructed in the silicon tracker. The track-cluster matching is performed either "outside-in" from energy cluster measurements, or "inside-out" from track reconstruction. Trajectories in the tracker volume are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian sum filter.
For the physics analysis, electron candidates are required to have a transverse momentum p_T^e larger than 7 GeV/c and a reconstructed geometrical acceptance $|\eta_e| < 2.5$. Integrated over the acceptance, the reconstruction efficiency for basic electron objects steeply rises to reach 90% at $p_T^e = 10$ GeV/c and then more slowly to reach a plateau of 95% for $p_T^e > 30$ GeV/c.

Electron 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

sample for the signal and a $W + 1jet$ data sample for background and it is optimized using a $Z + 1jet$ data sample.

Muon Reconstruction and Identification

The robust detection of muons over the full acceptance of the CMS detector in a condition of very high rate and background is obtained with a muon system that allows an efficient and pure identification of muons and with the inner tracker that provides a very precise measurement of their properties (as seen in Chapter 2).
Muons are required to have $p_T^{\mu} > 5\ GeV/c$ and $|\eta_{\mu}| < 2.4$. Their tracks are first reconstructed independently in the inner tracker (*tracker track*) and in the muon system (*standalone-muon track*) and then the information are combined, using two different approaches. The matching between the inner and outer tracks is initiated either "outside-in" (*Global Muon reconstruction*), starting from a track in the muon system, or "inside-out" (*Tracker Muon reconstruction*), starting from a track in the silicon tracker. The second method is more efficient than the *Global Muon reconstruction* at low momenta ($p \leq 5\ GeV/c$), because it requires only a single muon segment in the muon system, whereas *Global Muon reconstruction* is designed to have high efficiency for muons penetrating through more than one muon station and typically requires segments in at least two muon stations.

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 the *Particle-Flow Muon selection* is chosen, which combines the information from all sub-detectors to identify and reconstruct individually particles produced in the collision. To identify particle-flow muons, a selection is performed on all the muon candidates reconstructed by applying minimal requirements on the track components in the muon system.

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 leptons is measured relative to their transverse momentum p_T^{ℓ}, by summing over charged and neutral particles in a cone

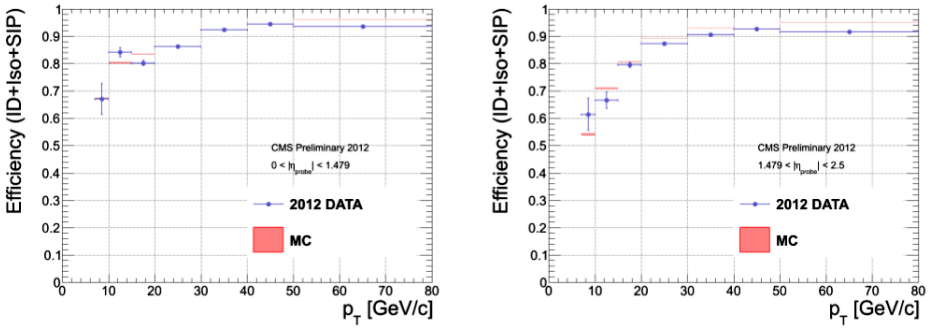
$\Delta R = \sqrt{(\eta^\ell - \eta^i)^2 + (\phi^\ell - \phi^i)^2} < 0.4$ around the lepton direction at the interaction vertex (where i indicates the ith-particle in the cone):

$$R^\ell_{Iso} \equiv \left(\sum p_T^{charged} + MAX\Big[0, \sum E_T^{neutral} + \sum E_T^{\gamma} - \rho \times A_{eff}\Big] \right) / p^\ell_T.$$

The $\sum p_T^{charged}$ is the scalar sum of the transverse momenta of charged hadrons originating from the primary vertex, without consider the lepton momentum it self. 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 latter excludes photons that are candidates for final-state radiation (FSR) from the lepton (see below). 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 a shift in the isolation distribution for higher 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 using the term $\rho \times A_{eff}$, where ρ is the energy density. A_{eff} is the *effective area* defined as the ratio between the slope of the average isolation and ρ as a function of number of vertices. This term subtracts an estimate obtained using a "jet area" technique of the transverse energy from neutrals in the isolation cone coming from pile-up 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 $p_T^{jet} > 3\ GeV$). Effective areas A_{eff} increase with the η of the lepton and this is mainly due to increasing pile-up at large pseudorapidity and with the fact that the geometrical isolation cone gets truncated at the end of the tracker acceptance. The electrons or muons are considered isolated in the $H \to 4\ell$ analysis if $R_{Iso} < 0.4$.

Leptons Efficiency

The efficiencies for reconstruction, identification and trigger for electrons and muons is measured with a data driven method based on a selection of events of inclusive single Z production. The *tag-and-probe technique* [11] 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 (Figures 3.6 and 3.7). The efficiency is defined as the ratio of the number of passing probes to the total number of probes before the cut.



(a) (b)

Figure 3.6: *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 [10].*

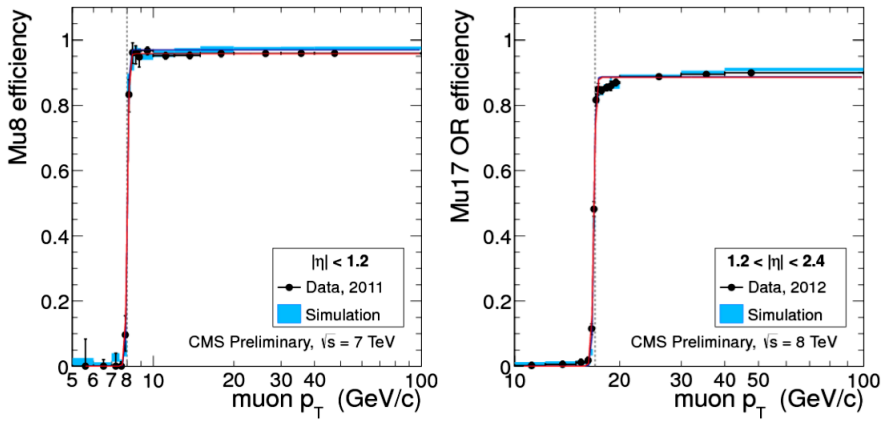


Figure 3.7: *Examples of muon HLT Trigger efficiency, as a function of the muon p_T. Muons are asked to pass ID, Isolation and SIP_{3D} requirements of the analysis [10].*

By using appropriate definitions for probes, the overall efficiency per lepton can be factorized in a series of terms, that can be measured independently:

$$\epsilon = \epsilon_{RECO|track\ or\ clustering} \times \epsilon_{ID|RECO} \times \epsilon_{ISO|ID} \times \epsilon_{SIP|ISO},$$

where each terms 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 clustering efficiency for electrons is assumed to be 100% efficient which is checked by MC to be the case within few per mil in the p_T range of interest for this analysis. The efficiency to reconstruct a muon in the inner tracker has been measured separately and found to be 99% or higher in the whole tracker acceptance, in good agreement with the

expectations from simulations [10]. By applying the method to both data and simulation one can derive data to simulation scale factors, used to either correct the signal efficiency in the simulation or to provide systematic uncertainties.

3.3.2 Photons

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) 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, absorbed by the calorimeter. 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 3 channels, in particular to discriminate the Higgs boson signal from the background continuum. Therefore it is necessary 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 addition to being collinear with the leptons, final state radiation also tends to favor low energy photon emission.

Photon Reconstruction and Identification

Photons are reconstructed and identified using the particle-flow reconstruction with a specific clustering algorithm. The determination of the photons energy and direction is monitored in the data with $\pi^0 \to \gamma\gamma$ decays and is shown to be accurate, reliable and in agreement with the predictions from simulation.

Photon Isolation

The photon isolation R^{γ}_{Iso} is determined from charged and neutral hadrons and from photons identified by the particle-flow reconstruction in a cone of size $\Delta R = 0.3$ around the candidate photon direction. The absolute photon isolation is defined as the sum of the transverse momenta of all these iso-deposits. To discriminate against photons produced in pile-up interactions, an additional isolation deposit is defined. It corresponds to the sum of charged particles from the vertices other than the primary vertex. The isolation observable corrected for pile-up is then divided by the photon transverse momentum p^{γ}_{T}. Isolated photons must satisfy $R_{Iso} < 1$.

Photons within $|\eta^{\gamma}| < 2.4$ are possible FSR candidates. To be accepted as FSR, a reconstructed photon must either have $p^{\gamma}_{T} > 2$ GeV/c and be found within a conical distance $\Delta R < 0.07$ from a selected lepton candidate, or have $p^{\gamma}_{T} > 4$ GeV/c and be found isolated within the conical distance of $0.07 < \Delta R < 0.5$ around a selected lepton candidate. 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 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 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

An optimal analysis in the 4ℓ channel (which includes the 4μ, $4e$, $2\mu 2e$ final states) must preserve the highest possible reconstruction efficiency for the Higgs boson signal while eliminating the contributions from the reducible and instrumental backgrounds. This is achieved focusing the attention on the signature of leptons. The analysis aims at the highest possible lepton reconstruction and selection, identification and isolation efficiencies, that are compatible with a negligible reducible and instrumental background (see Section 3.1.2), in a transverse momentum and pseudorapidity acceptance of $p^{e}_{T} > 7$ GeV/c and $|\eta^{e}| < 2.5$ for electrons and $p^{\mu}_{T} > 5$ GeV/c and $|\eta^{\mu}| < 2.4$ for muons. The analysis strategy relies on a simple sequence of cuts for the lepton identification and isolation, kinematic selection and specific background suppression requirements, as well as on methods relying on data for the evaluation of systematic uncertainties. The first "skimming", common to all 4ℓ channels, occurs imposing some *trigger requirements*. Then *preselection cuts* are applied, in order to choose the best combination of four leptons. Finally, a set of *kinematic requirements*

is introduced, to kill low mass resonances as well as multiple hadron decays. All these selection cuts allows for a drastic reduction of instrumental background rates.

The first "skimming" applied on the trigger selected events requires:

- at least one good primary vertex (PV) with high number of degree of freedom ($N_{PV} > 4$), collisions restricted along the z axis ($z_{PV} < 24\ cm$) and small radius of the PV ($r_{PV} < 2\ cm$)
- at least two reconstructed lepton candidates, either all electron basic track-supercluster object or global muon object, or a trigger muon object;
- $p_{T,1} > 20\ GeV/c$ and $p_{T,2} > 10\ GeV/c$ for electron and muon objects;
- an invariant mass $M > 40\ GeV/c^2$ between two same flavor leptons;

The selection steps act on two flavors of lepton: *loose leptons* and *good leptons*. The *loose leptons* are electrons within the geometrical acceptance of $|\eta^e| < 2.5$, with $p_T^e > 7\ GeV/c$ and having 0 or 1 expected missing inner hits, or muons (global or tracker) satisfying $|\eta^\mu| < 2.4$ and $p_T^\mu > 5\ GeV/c$. Both electrons and muons should satisfy loose requirements on the transverse ($d_{xy} < 0.5\ cm$) and longitudinal ($d_z < 1\ cm$) impact parameter with respect to the primary vertex. In addition, it is required that $\Delta R > 0.02$ between the leptons. The *loose leptons* are used in the estimation of reducible background. The *good leptons* are loose leptons on which additional criteria are imposed. Electrons should pass the electron identification criteria (as described above) and muons should meet the *Particle-Flow Muons* requirements. They both must have $R_{Iso} < 0.4$, that suppresses the $Z + jets$, $Zb\bar{b}$ and $t\bar{t}$ backgrounds. Moreover, the significance of the impact parameter to the event vertex, SIP_{3D}, is required to satisfy $|SIP_{3D} = \frac{IP}{\sigma_{IP}}| < 4$ for each lepton, as already mentioned. This requirement further suppresses $Zb\bar{b}$ and $t\bar{t}$ backgrounds.

The preselection is applied to events that have fired the relevant electron and muon triggers and consists in the following requirements:

1. *First Z*: a pair of good lepton candidates of opposite charge and matching flavor (e^+e^-, $\mu^+\mu^-$) with reconstructed mass $m_{1,2}$ closest to the nominal Z boson mass is retained and denoted Z_1. The selected pair should satisfy $40 < m_{Z_1} < 120\ GeV/c^2$.
2. *Four or more leptons and a matching pair*: another good lepton pair with same flavor and opposite charge.

3. *Choice of the "best* 4ℓ*" and* Z_1, Z_2 *assignments*: retain a second lepton pair, denoted Z_2, among all the remaining $\ell^+\ell^-$ combinations. If more than one Z_2 combination satisfies all the criteria, the one built from leptons of highest p_T^l is chosen. The selected pair should satisfy $12 < m_{Z_2} < 120\ GeV/c^2$. At this stage, it is required that any two leptons from the four selected have $p_{T,i} > 20\ GeV/c$ and $p_{T,j} > 10\ GeV/c$.

4. *QCD suppression*: the reconstruction mass of opposite-sign and same-flavor lepton pairs must satisfy $m_{\ell\ell} > 4\ GeV/c^2$, in order to further protect against leptons originating from hadron decays in jet fragmentation or from the decay of low-mass hadronic resonances.

The first step ensures that the leptons in the preselected events are on the high efficiency plateau for the trigger. Control samples for the $Z + jets$, $Zb\bar{b}/c\bar{c}$ and $t\bar{t}$ backgrounds are obtained as subsets of the event sample that remains after the first step. The second step allows for control of the three-lepton event rates which include WZ di-boson production events. The first four steps are designed to reduce the contribution of the instrumental backgrounds from QCD multi-jets and $Z + jets$, while preserving the maximal signal efficiency and the phase space for the evaluation of background systematics. By reducing the number of jets mis-identified as leptons, fewer combinatorial ambiguities arise when assigning the leptons to candidate Z bosons. The choice of the best combination of four leptons completes the preselection.

Low mass resonances as well as multiple hadron decays are killed applying some kinematic cuts. Three sets are therefore introduced to allow studies of the s-channel contribution or to maximize the sensitivity in different ranges of Higgs boson mass hypothesis:

- $Z \to 4\ell$ *phase space analysis*: it is defined by requiring $70 < m_{4\ell} < 110\ GeV/c^2$, $40 < m_{Z_1} < 120\ GeV/c^2$ and $4 < m_{Z_2} < 120\ GeV/c^2$.

- *Baseline Higgs phase space analysis*: it is defined by requiring $m_{4\ell} > 110\ GeV/c^2$, $40 < m_{Z_1} < 120\ GeV/c^2$ and $12 < m_{Z_2} < 120\ GeV/c^2$. This provides a best sensitivity for masses $m_H < 130\ GeV/c^2$.

- *High-mass phase space analysis*: it is defined by requiring $60 < m_{Z_1} < 120\ GeV/c^2$ and $60 < m_{Z_2} < 120\ GeV/c^2$. It is used to measure the ZZ cross-section.

Note that events satisfying the high-mass selection are a subset of those satisfying the baseline selection. The enlarge phase space of the baseline selection for the Higgs boson signal is needed at very low masses given

the very small cross section×branching ratio, at the price of a larger background. The increased acceptance for the signal becomes small for mass above $\sim 130\ GeV/c^2$ where reducing the phase space to better suppress the background becomes advantageous. For Higgs boson masses above $\sim 2 \times m_Z$, one can restrict the phase space of the pair of Z bosons with the high mass selection, without significant loss of acceptance for signal, with the benefit of a slight reduction of the $ZZ^{(\star)}$ background.

The signal detection efficiencies from MC for a 4ℓ system within the geometrical acceptance is evaluated to be rising from about 45%/72%/54% at $m_H = 190\ GeV/c^2$ to about 59%/82%/71% at $m_H = 400\ GeV/c^2$ for the $4e/4\mu/2e2\mu$ channels.

The event yields are found to be in good agreement with the MC background expectation at each step of event selection (Figure 3.8 and 3.9).

3.5 Background Control and Systematics

The total number of signal-like events surviving the baseline selection is relatively small for the current integrated luminosity. This doesn't allow a precise evaluation of the background in a relevant narrow signal-like mass window from the measurement of nearby side-bands only. Thus one has to rely on other methods, based on experimental data, for the control of the background and the evaluation of associated systematic uncertainties. The background is dominated by the $ZZ^{(\star)}$ continuum and only a small contamination remains from the reducible and instrumental backgrounds. The typical procedure to evaluate background from data consists of choosing a wide background control region outside the signal phase space. This region gets populated by relaxing the event selection and verifying that the event rates change according to the expectation from simulation. If a specific background contribution has to be determined, the corresponding control region must be chosen carefully since any of the other reducible backgrounds might rapidly become dominant if the event selection is relaxed, thus making the extrapolation to the signal phase space difficult.

3.5.1 Evaluation of $ZZ^{(\star)}$ continuum

The expected rate of $ZZ^{(\star)}$ continuum in any given mass range $[m_1, m_2]$ is obtained directly from the absolute rate predicted by the MC simulation:

$$N^{ZZ}_{expect}[m_1, m_2] = \int_{m_1}^{m_2} \left(\frac{dN}{dm_{4\ell}} \right)_{MC} dm_{4\ell}$$

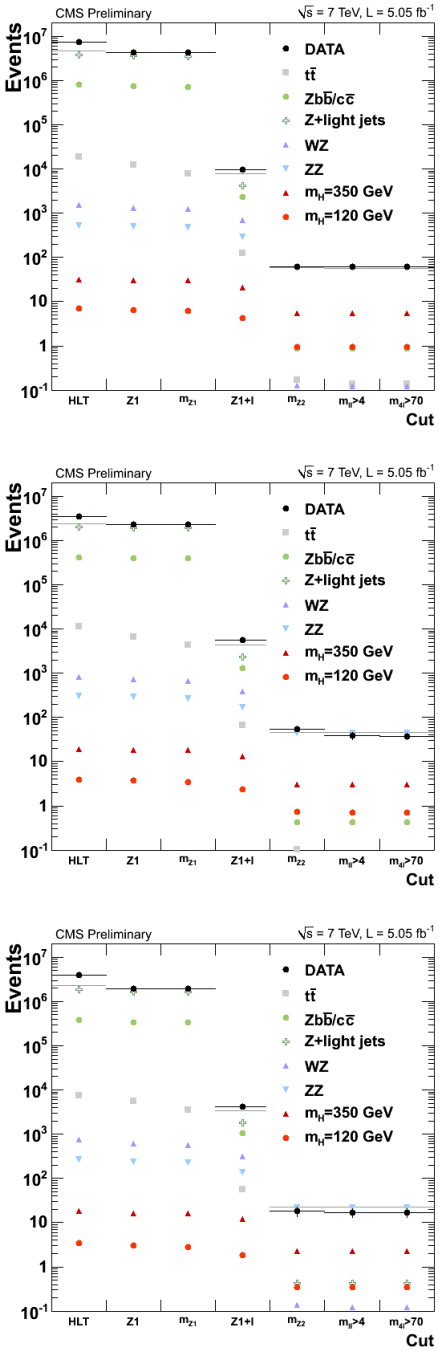


Figure 3.8: *Reduction of events in function of the selection steps, for the 2e2μ, 4e and 4μ channels (7 TeV) [10].*

where $\frac{dN}{dm_{4\ell}}$ is the local density of events expected as function of the mass $m_{4\ell}$ from the $ZZ^{(\star)}$ background.

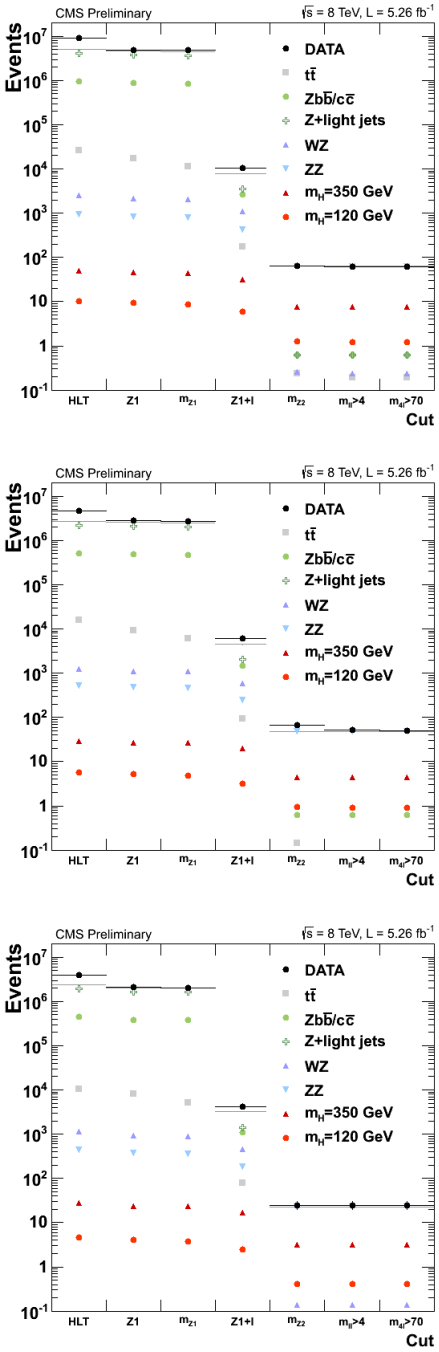


Figure 3.9: *Reduction of events in function of the selection steps, for the 2e2μ, 4e and 4μ channels (8 TeV) [10].*

When used for comparisons with data, the method is affected by the full

systematics uncertainties on the pp integrated luminosity and the theoretical errors as well as systematic errors on acceptance within analysis cuts. The cross section for $ZZ^{(\star)}$ production at NLO is modeled using a functional form very similar to that of the signal and it is calculated with MCFM. This includes the dominant process of qq annihilation, as well as gluon fusion. The theoretical uncertainties are computed as a function of $m_{4\ell}$, varying both the QCD renormalization and factorization scales and the PDF used. 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.

3.5.2 Inclusive Instrumental and Reducible Backgrounds Estimation

The method here presented allows for an inclusive measurement of all the main instrumental and reducible backgrounds, that remain after the first step of the selection. The reducible backgrounds for the $H \to ZZ \to 4\ell$ analysis are processes which contain one or more non-prompt leptons in the four-lepton final state. The main sources of non-prompt leptons are non-isolated electrons and muons coming from decays of heavy-flavor mesons, mis-reconstructed jets (usually originating from light-flavor quarks) and electrons from γ conversions.
The control sample used to evaluate background from data is obtained as subsets of the events that satisfy the *First Z* step of preselection (see Section 3.4), requiring an additional pair of reconstructed leptons of same sign (to avoid signal contamination) and same flavor (SS-SF: $e^{\pm}e^{\pm}$, $\mu^{\pm}\mu^{\pm}$). The SS-SF leptons are requested to pass SIP_{3D} cut while no identification or isolation requirements are imposed. The reconstructed four-lepton invariant mass has to satisfy $m_{4\ell} > 100\ GeV/c^2$ and at least three of the four combinations of $\ell\ell$ pairs must satisfy $m_{\ell\ell} > 12\ GeV/c^2$ (as in the best candidate choice). The SS-SF leptons invariant mass m_{Z_2} is required to satisfy the baseline, intermediate-mass or the high-mass selections.
From this set of events the inclusive number of reducible background in the signal region is estimated by measuring the probability for the two additional leptons to pass the isolation and identification analysis cuts, obtained from a "fake rate measurement". The fake rate is intended to estimate the probability that a "fake-lepton" can be selected as signal-like lepton. The expression "fake leptons" indicates all the reconstructed leptons that are not coming from prompt isolated lepton, such as any jet mis-reconstructed as a lepton and any lepton originating from a heavy meson decay. In a similar way, any electron originating from a photon conversion will be considered "fake electron".
The fake rate measurement is obtained from an independent sample of Z_1 plus exactly one lepton (called "fakeable" object) and where contami-

nation from WZ events is suppressed by requiring the imbalance on the measured energy deposition in the transverse plane to be below 25 GeV. The "fakeable" object is defined as an electron or muon with relaxed ID requirement and passing $|SIP_{3D}| < 4$. In this way, a "fakeable" electron is a reconstructed electron without additional ID requirements and a "fakeable" muon is a reconstructed muon satisfying the *Global Muon* or *Tracker Muon* reconstruction without additional ID requirements. The fake rate is computed as the ratio of "fakeable" object passing the identification and isolation criteria over the total number of "fakeable" objects and it is calculated for each lepton flavor as a function of p_T, for barrel and endcaps separately and for various isolation cuts.

Starting from the control sample previously described and using the fake rate measurement, the final reducible background prediction in the signal region is given by the following expression:

$$N_{expect}^{Z+X} = N^{data} \times \left(\frac{OS}{SS}\right)_{MC} \times \epsilon_1(p_T, \eta) \times \epsilon_2(p_T, \eta),$$

where N_{expect}^{Z+X} is the number of events in the control region, $\left(\frac{OS}{SS}\right)_{MC}$ is a correction factor between opposite sign and same sign control samples and $\epsilon_i(p_T, \eta)$ is the fake rate probability for each of additional pair of loose leptons in function of p_T and η.

With the 5.05 fb^{-1} of collected integrated luminosity at 7 TeV and 5.26 fb^{-1} at 8 TeV, the number of events from $Z + X$ expected and the relative systematics and statistical errors in the signal region in a mass range from $m_1 = 110\ GeV/c^2$ to $m_2 = 160\ GeV/c^2$ is listed in table 3.2.

For the 4ℓ reducible background estimate, a different approach is also implemented. This second method uses the control region with two opposite-sign leptons failing the isolation and identification criteria and using the mis-identification probability to extrapolate to the signal region. 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 mis-identified lepton. Comparable background counts in the signal region are found within uncertainties from both methods.

3.6 Kinematic Discriminant (MELA)

3.6.1 Introduction of the Methodology

A complete set of angular observables was recently introduced to study kinematics of the Higgs decay to ZZ final state and it may also help in

background rejection. In this approach, the signal-to-background probability is calculated using an analytical or empirical multi-dimension likelihood for an event to be signal or background. In Figure 3.10 the angular

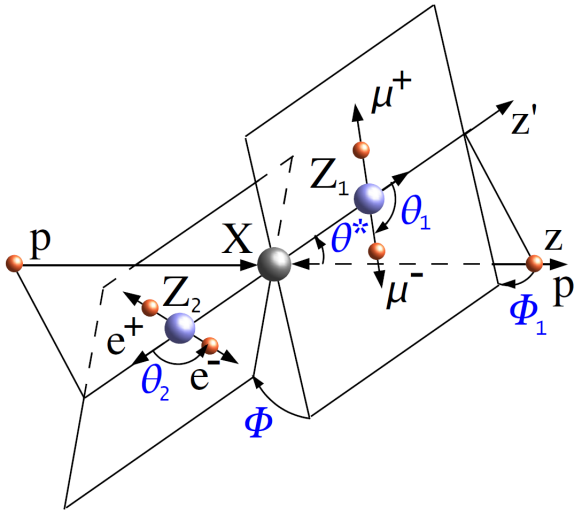


Figure 3.10: *Illustration of a particle X production and decay* $ab \to X \to Z_1 Z_2 \to 4\ell$ *[10].*

distribution in the production and decay chain $ab \to X \to ZZ \to 4\ell$ is illustrated. The angular distribution can be expressed as a function of three helicity angles θ_1, θ_2, and Φ and two production angles $\theta^\star$ and Φ_1. 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 help of the following 12 observables (reflecting all degrees of freedom with four lepton momenta):

- three resonance masses (including the off-shell cases): $m_{4\ell}$, m_1, m_2;
- five production and decay angles defined as $\Omega = \{\theta^\star, \Phi_1, \theta_1, \theta_2, \Phi\}$ in Figure 3.10;
- longitudinal boost of the resonance, expressed as rapidity Y;
- transverse momentum of the resonance p_T and its azimuthal angle;
- one arbitrary azimuthal angle $\Phi^\star$ reflecting the overall orientation of the system.

The Matrix Element Likelihood Approach (MELA) constructs and uses a kinematic discriminant KD from the seven mass and angular observables

$$KD = F\{m_1, m_2, \theta^\star, \Phi_1, \theta_1, \theta_2, \Phi\}$$

and performs a two dimensional fit over the two observables $(m_{4\ell}, KD)$. In Figure 3.11 the distributions of signal and background are shown for each individual observable. The KD discriminant combines these observables in a single observable using full correlation of all inputs in the most optimal way. In this approach the variables which depend on QCD kinematics, such as Y, p_T and its azimuthal angle are removed from consideration, and only observables coming from well-understood electro-weak Quantum Mechanics of the processes of either Higgs or continuum ZZ production are kept.
These seven observables are independent from the Higgs production mechanism, as long as the SM Higgs boson is considered, and they are also the key input to measure the new boson properties, such as spin and CP quantum numbers, should a new boson be discovered.

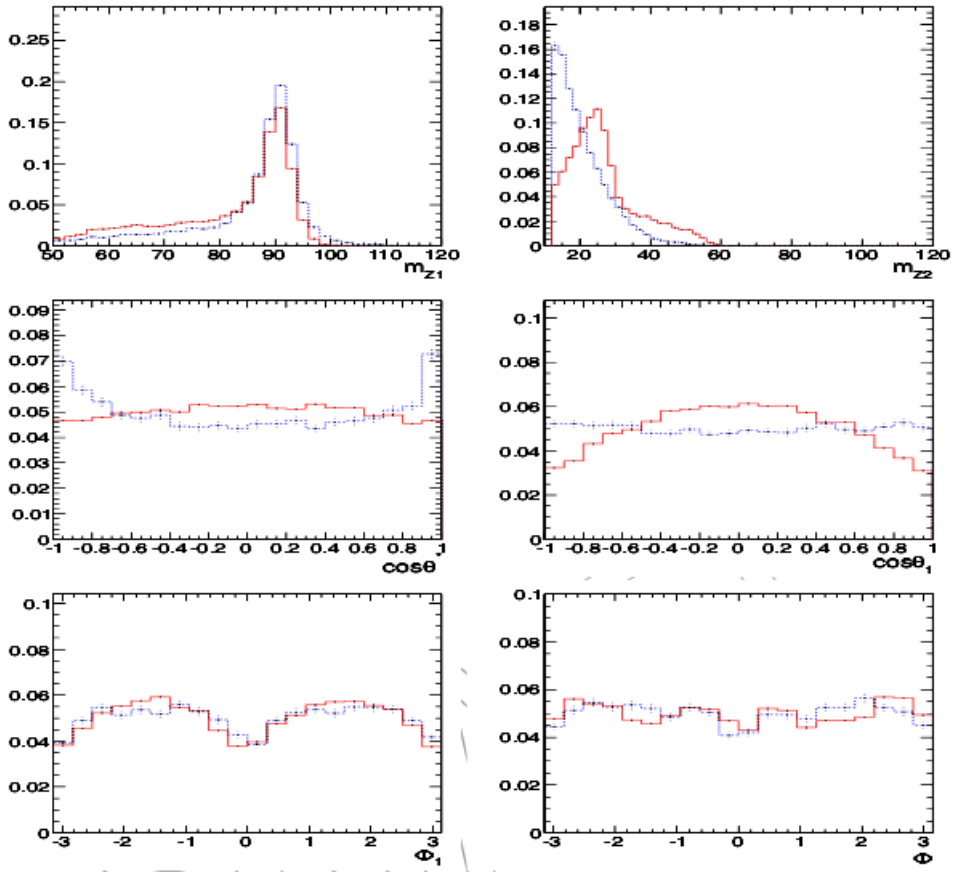


Figure 3.11: *Distribution of Higgs signal events with $m_H = 120$ GeV/c^2 (solid red) and background ZZ events (dashed blue) in the range $100 < m_{4\ell} < 135$ GeV/c^2 for the m_1, m_2 and angular observables [10].*

3.6.2 Construction of the MELA Discriminant

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

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

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

where probabilities are normalized with respect to the seven observables while treating $m_{4\ell}$ as a conditional parameter. The discriminant is constructed as follows

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

Note that this discriminant is continuously distributed between 0 and 1, with signal being closer to 1 and background closer to 0. The signal probability is parameterized as a function of $m_{4\ell}$ instead of m_H in order to allow continuous selection of the data-sample independent of the m_H hypothesis. 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 the further fit implementation more robust. Parameterization is performed not including the detector effects because they are identical for signal and background and would cancel in the ratio.
The case above or below the $2m_Z$ threshold are considered. When $m_{4\ell}$ is above the threshold, the two Z bosons are on-shell and no separation is provided by m_1 and m_2, therefore leaving only the five angles in the parametrization. When $m_{4\ell}$ is below the $2m_Z$ threshold, background has significant contributions from the $qq \to ZZ^{(\star)}/Z\gamma^{(\star)}$ processes and instead of the analytical parametrization it is tabulated in a correlated template distribution using the POWHEG simulation at generator level.

The resulting MELA KD distributions for signal and background are shown in Figure 3.12 in two different mass ranges. Good agreement is found between data and background MC. Overall, significant separation between signal and background is evident.

3.7 Results

The reconstructed four-lepton invariant mass distribution obtained combining the 4μ, $4e$ and $2e2\mu$ channels with the baseline selection is shown in Figure 3.13 for data. It is compared to expectations from the SM background processes.

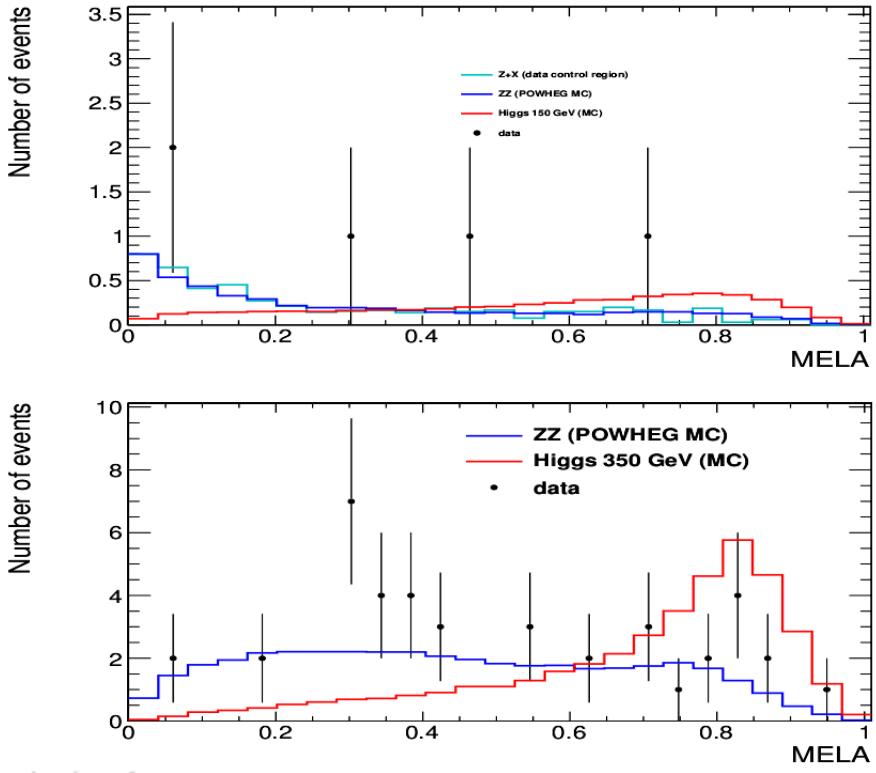


Figure 3.12: *KD distributions for signal and background in two mass ranges:* $140 < m_{4\ell} < 160$ *GeV*$/c^2$ *(top) and* $250 < m_{4\ell} < 450$ *GeV*$/c^2$ *(bottom). The signal (red solid histogram) is shown for* $m_H = 150$ *and* 350 *GeV*$/c^2$*, respectively. The* ZZ *continuum background is shown as blue solid histogram. The top plot also shows* $Z + X$ *background estimated from data control region [10].*

As one can see, the $Z \to 4\ell$ decays 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 [10], 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 $Z + X$ reducible background distribution is obtained putting together information from data-driven method and knowledge on shape taken from the MC samples, as previously described in Section 3.5.2. Its distribution shows that the reducible and instrumental backgrounds contributions are very small. The ZZ and signal normalization and shapes are directly taken from MC samples.
The observed distribution is found to be compatible with the expectation

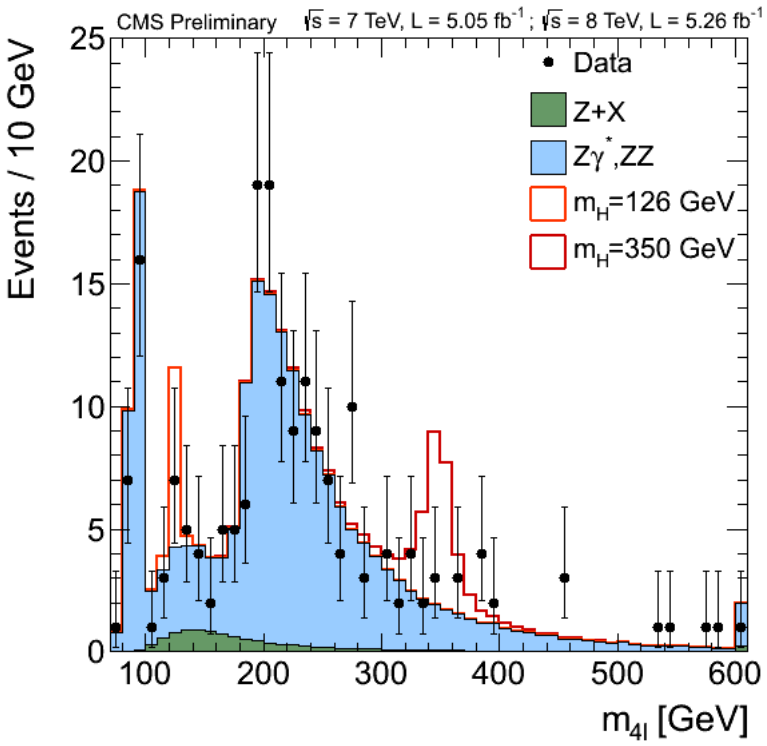


Figure 3.13: *Distribution of the four-lepton reconstructed mass in full mass range for the sum of the 4e, 4μ and 2e2μ channels. The distributions are presented as stacked histograms. The measurements are presented for the sum of the data collected at* $\sqrt{s} =$ *7 TeV and* $\sqrt{s} = 8$ *TeV [11].*

from SM continuum production of $ZZ^{(\star)}$ pairs. 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. One hundred and seventy-two candidates are selected. Of these events, forty-seven are found in the 4μ channel, thirty-two in the $4e$ channel and ninety-three in the $2e2\mu$ channel.

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 3.14. The distribution of events in the $(m_{4\ell}, KD)$ plane agrees well with the SM expectation in the high mass range. As one can see from Figure 3.14, for values of the MELA $KD > 0.5$, three events are observed in the range $520 < m_{4\ell} < 600$ GeV/c^2 and a clustering of candidates is apparent in the low mass range. 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

Table 3.1: *The number of event candidates observed, compared to the mean expected background and signal rates for each final state. The results are given integrated over the full mass measurement range for the Higgs boson search from* 100 *to* 800 GeV/c^2 *[11].*

Channel	$4e$	4μ	$2e2\mu$	4ℓ
ZZ background	29.3 ± 3.4	49.0 ± 5.1	75.5 ± 8.0	153.7 ± 1.5
$Z + X$ background	$3.0^{+2.7}_{-1.9}$	$2.2^{+1.6}_{-1.3}$	$5.0^{+4.0}_{-3.0}$	$10.2^{+11.3}_{-10.8}$
All backgrounds	$32.3^{+4.4}_{-3.9}$	$51.2^{+5.3}_{-5.3}$	$80.5^{+9.0}_{-8.6}$	$163.9^{+11.3}_{-10.8}$
$m_H = 200\ GeV/c^2$	8.3 ± 2.0	13.3 ± 2.7	21.6 ± 4.5	43.2 ± 5.6
$m_H = 350\ GeV/c^2$	4.8 ± 1.2	7.5 ± 1.6	12.7 ± 2.9	24.9 ± 3.5
$m_H = 500\ GeV/c^2$	1.7 ± 0.8	2.6 ± 1.2	4.4 ± 2.0	8.7 ± 2.4
Observed	32	47	93	172

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

Channel	$4e$	4μ	$2e2\mu$	4ℓ
ZZ background	2.7 ± 0.3	5.7 ± 0.6	7.2 ± 0.8	15.5 ± 1.0
$Z + X$ background	$1.2^{+1.1}_{-0.8}$	$0.9^{+0.7}_{-0.6}$	$2.3^{+1.8}_{-1.4}$	$4.4^{+2.2}_{-1.7}$
All backgrounds	$3.9^{+1.1}_{-0.8}$	$6.6^{+0.9}_{-0.8}$	$9.5^{+2.0}_{-1.6}$	$19.9^{+2.4}_{-2.0}$
$m_H = 120\ GeV/c^2$	0.8 ± 0.2	1.6 ± 0.3	1.9 ± 0.5	4.4 ± 0.6
$m_H = 126\ GeV/c^2$	1.5 ± 0.5	3.0 ± 0.6	3.8 ± 0.9	8.3 ± 1.2
$m_H = 130\ GeV/c^2$	2.1 ± 0.7	4.1 ± 0.8	5.4 ± 1.3	11.6 ± 1.6
Observed	32	47	93	172

the 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 183 hypothetical Higgs boson masses in a range between 110 GeV/c^2 and 600 GeV/c^2. The choice of the intervals in between Higgs mass hypotheses is driven by either detector resolution of the resonance or its natural width depending on which is larger. As seen in Section 1.5.2 (Figure 1.8) in the low mass region, the Higgs boson width is of the order of the MeV/c^2 and it is smaller than the experimental mass resolution. Then it rapidly increases and in the high mass region, $m_H > 2m_Z$, the total Higgs boson width is dominated by the W^+W^- and ZZ partial widths. 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

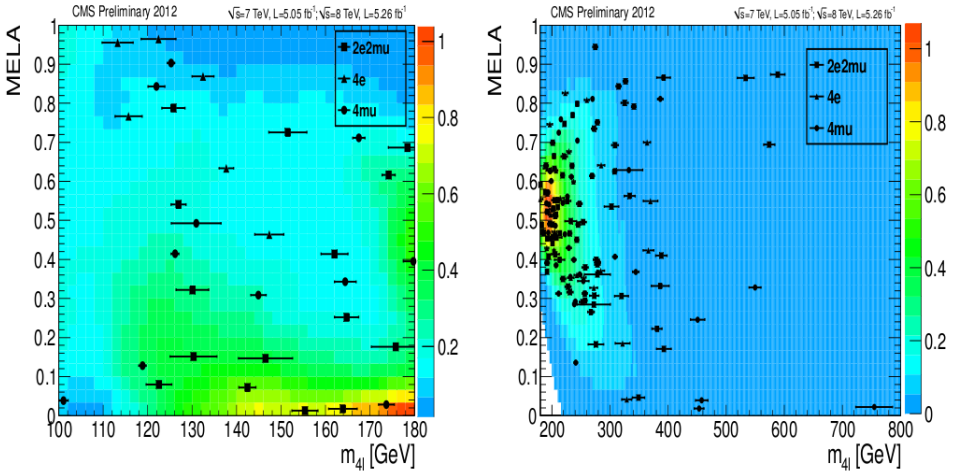


Figure 3.14: *Distribution of the MELA KD versus the four-lepton reconstructed mass $m_{4\ell}$ in the low-mass (left) and full-mass (right) regions [11].*

must interpolate the mass shapes of signal hypotheses where no simulation exists. As a cross-check, one-dimensional $m_{4\ell}$ distribution is studied and found consistent, but systematically higher expected limits.

As already delineated, the probability distribution of $\mathcal{P}(m_{4\ell})$ for the background is parametrized 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 parametrization convoluted with a Crystal-Ball function. 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 the ZZ background and any potential small difference is accounted for in the systematic uncertainties.

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 3.15 (left). If, for a particular range of m_H, the normalized upper limit is lower than unity, it means that the measured cross section maximum value is smaller than the SM prediction. The considered mass region is thus excluded within the SM, because if the Higgs boson were present at that mass values, an excess of events would be visible. On the contrary, if the normalized upper limit is larger than unity, no prediction are possible. The SM Higgs boson is excluded by the search in the four-lepton channels at 95% CL in the range $131 - 525$ GeV/c^2, except for the small window

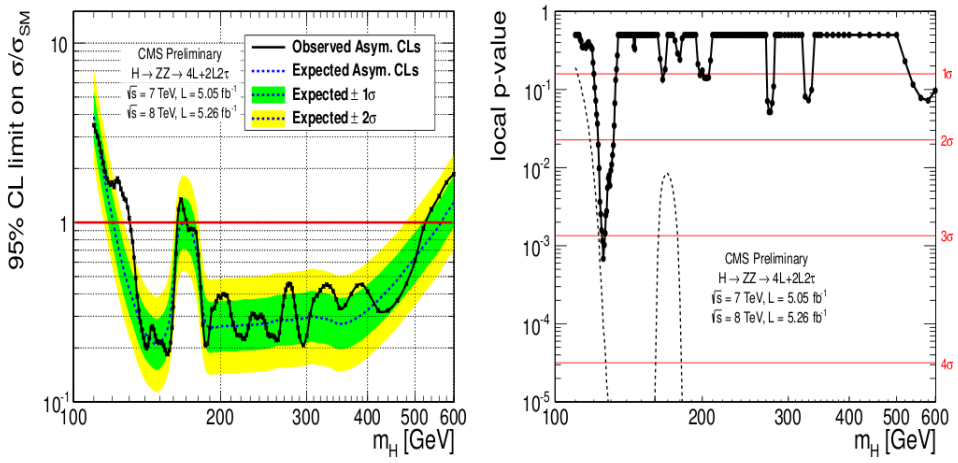


Figure 3.15: *Observed and expected* 95% *CL upper limit (left) on the ratio of the production cross section to the SM expectation. Significance of the local excess (right) with respect to the Standard Model background expectation as a function of the Higgs boson mass in the full interpretation mass range* $110 - 600\ GeV/c^2$ *[11].*

$(162 - 172\ GeV/c^2)$, where the branching ratio for the $H \to ZZ$ decay is disfavored and the sensitivity is therefore lower. The upper limits in the low-mass region are given in Figure 3.16.

The local p-value, representing the significance of local fluctuations with respect to the Standard Model expectation, is shown for the full mass range as a function of the Higgs mass in Figure 3.15 (right). As one can note, the minimum of the local p-value is reached at low mass around $m_{4\ell} = 125.5\ GeV/c^2$, corresponding to an excess of events in the low mass range in the 4ℓ channel. The number of candidates observed as well as the estimated background in the signal region (from 110 to 160 GeV/c^2) are reported in Table 3.2, for the selection in the low mass range. A relatively flat background is expected in this mass range.

In Figure 3.17 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 the low mass range). A signal-like clustering of events is apparent at high values of KD and for $m_H \sim 125\ GeV/c^2$.

The local p-value in the low mass region is shown in Figure 3.18. The minimum is reached for the Higgs boson mass hypothesis of 125.5 GeV/c^2 and corresponds to a local significance of 3.2σ (2.2σ in the 1D fit without the MELA KD). The average expected significance for a Standard Model Higgs boson at this mass is 3.8σ (3.2σ) for the 2D (1D) fit. To extract from the 4ℓ measurements a signal strength modifier $\mu = \frac{\sigma}{\sigma_{SM}}$ for a SM Higgs

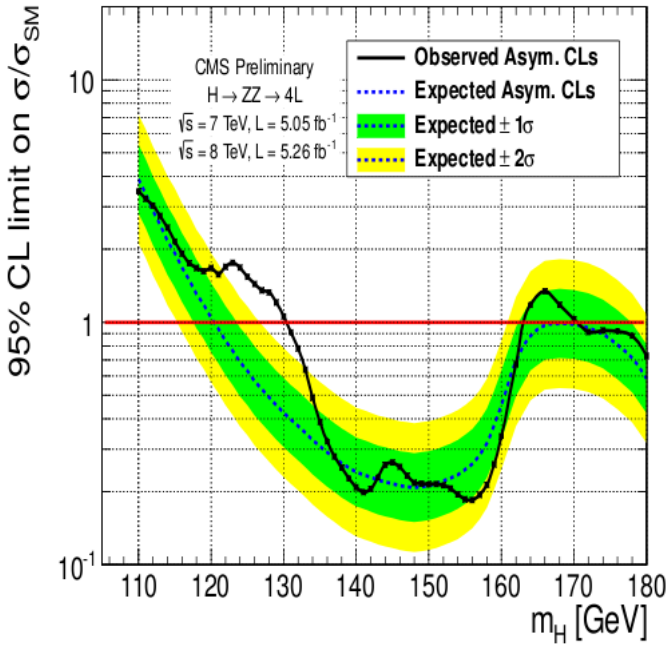


Figure 3.16: *Observed and expected* 95% *CL upper limit (left) on the ratio of the production cross section to the SM expectation, in the low mass region [11].*

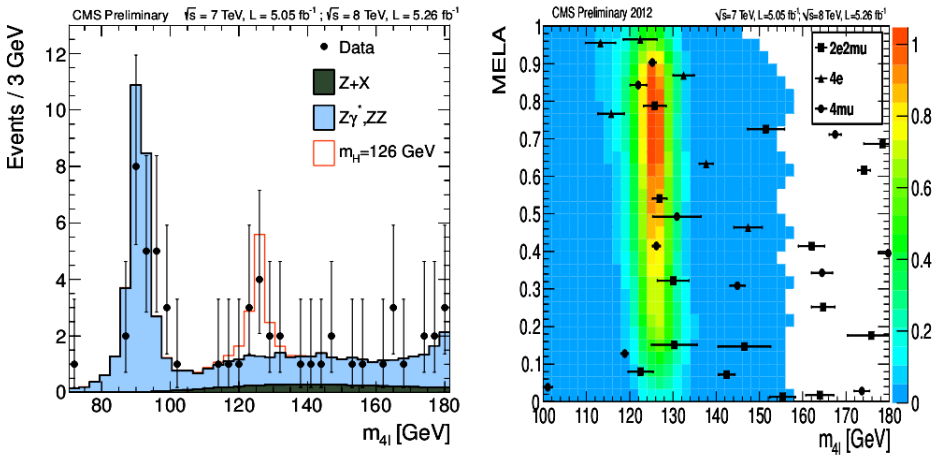


Figure 3.17: *Distribution of the four-lepton reconstructed mass for the sum of the* 4*e*, 4μ, *and* 2*e*2μ *channels (left). Distribution of the MELA KD versus the four-lepton reconstructed mass* $m_{4\ell}$ *(right) with contours shown for expected relative density of signal events for hypothesis* $m_H = 126\ GeV/c^2$ *[11].*

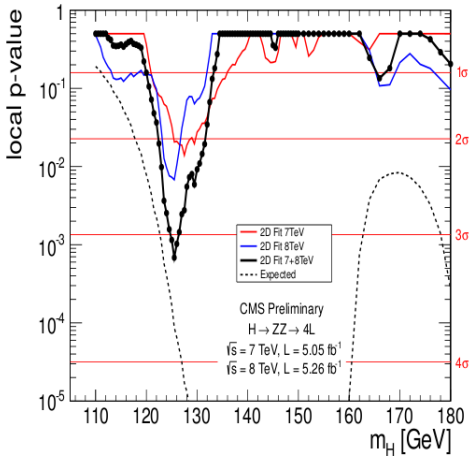


Figure 3.18: *Significance of the local excess with respect to the Standard Model background expectation as a function of the Higgs boson mass. The results are shown for the full data sample and for $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV samples separately in the low-mass region only [11].*

boson signal and a most probable mass, a likelihood scan on data in the 2D space of m_H vs μ is performed, finding a global minimum located at $m_H = 125.3 \pm 0.6\ GeV/c^2$ and $\mu = 0.7$.

Using simulation it was found that the MELA KD distribution for signal at mass around $m_H = 125\ GeV/c^2$ is similar for a scalar, pseudo-scalar or a spin-two resonance with the minimal couplings. In order to study the discrimination between the pseudo-scalar and scalar hypotheses, a modified MELA discriminant is used, with the pseudo-scalar signal hypothesis probability in place of background probability. The expected separation between the two hypotheses with the present data sample is 1.6σ and additional 25 fb^{-1} are needed to reach the 3σ expected separation.

Chapter 4

Study of $Zb\bar{b}$ and $t\bar{t}$ Backgrounds

In this Chapter the attention is focused on the $Zb\bar{b} \to 4\ell$ and $t\bar{t} \to 4\ell$ processes, reducible backgrounds of the $H \to ZZ^{(\star)} \to 4\ell$ signal. Even if they are drastically reduced by the analysis selection cuts, a small amount of these events is present in the final sample. An appropriate strategy is therefore needed to control the contamination of these processes.

In the following the reconstruction and selection of the $Zb\bar{b}$ and $t\bar{t}$ candidates, the discrimination of these two kind of background and the measurement of the $Zb\bar{b} \to 4\ell$ and $t\bar{t} \to 4\ell$ cross sections are presented.

4.1 Monte Carlo and Data Samples

Monte Carlo and data samples are the same of the "$H \to ZZ \to 4\ell$ analysis", described in Chapter 3. The only difference is the luminosity of 11.341 fb^{-1} at 8 TeV, incremented in the last two months.

SM Higgs boson signal samples, as well as samples for a large variety of electroweak and QCD-induced SM background processes, are obtained using detailed Monte Carlo (MC) simulations. Generated events are processed through the full GEANT detector simulation [12], trigger emulation and event reconstruction chain. Pile-up contributions and the misalignment of the muon and inner tracker system are also considered, reproducing a realistic scenario for the reconstruction.

As seen in Section 3.2, MC samples are used for several purposes: the optimization of the event selection strategy prior to the analysis of exper-

imental data, the comparisons with the measurements on data, the evaluation of acceptance corrections and systematic uncertainties and the background evaluation procedure. In this case background "*control regions*" are used and then extrapolated to the "*signal region*" (see Section 4.2).

In the following, a more detailed description with respect to Section 3.2 is presented for the MC samples used in this analysis.

Signal: $H \to ZZ^{(\star)} \to 4\ell$

The Higgs boson samples used in the current analysis are generated with POWHEG [13]. which incorporates NLO gluon fusion ($gg \to H$) and weak-boson fusion $qq \to qqH$. The CTEQ6M parton distribution function (PDF) is used for generation. Additional samples with WH, ZH and $t\bar{t}H$ associated production are generated with PYTHIA. The Higgs boson is forced to decay into two Z bosons, which are allowed to be off-shell, and both Z bosons are forced to decay via $Z \to 2\ell$. Generator level events are reweighted according to the total cross section $\sigma(pp \to H)$ which comprises the gluon fusion contribution up to NNLO and NNLL and the weak-boson fusion contribution at NNLO [4]. The total cross section is scaled by the branching ratio $BR(H \to 4\ell)$.

Background: $q\bar{q} \to ZZ^{(\star)} \to 4\ell$

For the current analysis POWHEG samples including $q\bar{q} \to ZZ^{(\star)} \to 4\ell$ are produced. They include the complete NLO simulation, interfaced to PYTHIA for showering, hadronization, decays and the modelling of the underlying event.

Background: $gg \to ZZ^{(\star)} \to 4\ell$

The gluon-induced ZZ background, although technically of NNLO 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. A full NNLO calculation for the ZZ production which would also take these gluon-induced diagrams into account is not available. Therefore the contributions are estimated by using the dedicated tool GG2ZZ [14], which computes the $gg \to ZZ$ at LO, which is of order α_s^2, compared to α_s for the LO $q\bar{q} \to ZZ$. The hard scattering $gg \to ZZ^{(\star)} \to 4\ell$ events are then showered and hadronized using PYTHIA.

Background: $Z + jets \rightarrow 2\ell + jets$

$Z + jets \rightarrow 2\ell + jets$ samples are generated with MadGraph, with a total of $\sim$ 40 M events representing an equivalent integrated luminosity well above $\mathcal{O}(10)$ fb^{-1}. Both light ($q = d$, u, s) and heavy-flavor ($q = c$, b) jets are included in the sample. A generation cut on two-lepton invariant mass of $m_{\ell\ell} > 50$ GeV/c^2 is imposed in the simulation. The total cross section at NNLO is 3048 (3503.7) pb at 7 (8) TeV. To separate the contribution from heavy-flavor jets (from now on referred to as the $Zb\bar{b}$ sample) the MadGraph $Z + jets$ sample was partitioned in $Z + light\ jets$ and $Z + heavy\ flavor\ jets$ using a filter selecting events with two $b - jets$ or two $c - jets$ in the final state.

Background: $t\bar{t} \rightarrow 2\ell 2\nu 2b$

A $t\bar{t} \rightarrow 2\ell 2\nu 2b$ sample is generated with the powheg event generator using cteq6m. The theoretical NLO cross-section for the process is $\sigma_{NLO}(pp \rightarrow t\bar{t} \rightarrow 2\ell 2\nu 2b) = 17.32\ (23.64)\ pb$ at 7 (8) TeV. A sample of about 10 million events corresponding to an integrated luminosity of more than 600 fb^{-1} is simulated [15].

4.2 Definition of the Control Region

The typical procedure to evaluate background from data consists of choosing a wide background *control region* (CR) outside the signal phase space. A control region is a kinematic region where the signal and the background are well discriminated, since signal is removed by the CR cuts and background can thus be characterized. The control region gets populated by relaxing the event selection and verifying that the event rates change according to the expectation from simulation. This is shown in Figure 4.1: in order to enhance the background with respect to the signal, some of selection criteria applied for the signal region can be inverted to study the background distribution. Then, using MC simulation, the number of background events in the signal region can be estimated.
If a specific background contribution has to be determined, the corresponding control region must be chosen carefully since any of the other reducible backgrounds might rapidly become dominant if the event selection is relaxed, thus making the extrapolation to the signal phase space difficult.

In order to increase the number of $Zb\bar{b}$ and $t\bar{t}$ background events for a more precise measurement, a specific control region is defined. This control region requires in each event a four-lepton candidate with:

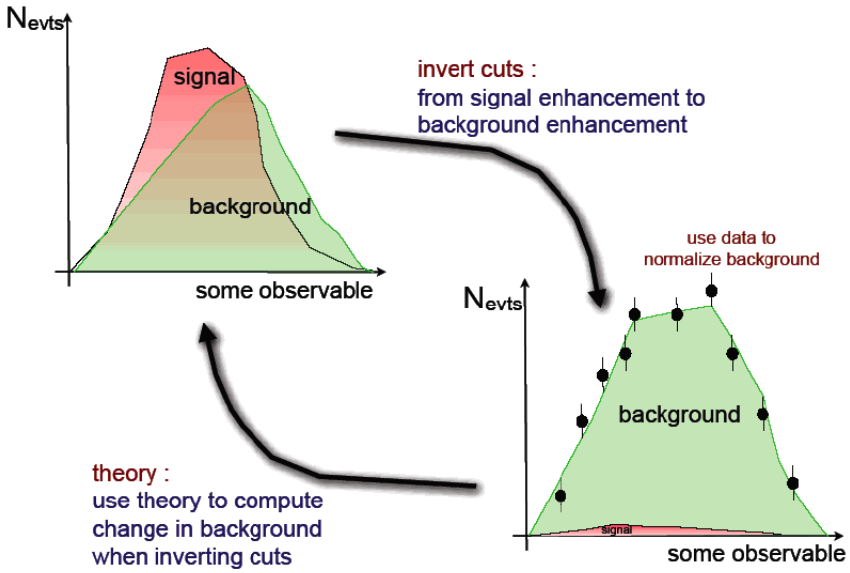


Figure 4.1: *Summary of the control region procedure.*

- a pair of *good lepton candidates* (see Section 3.4) of opposite charge and matching flavor (e^+e^-, $\mu^+\mu^-$) with reconstructed mass $m_{1,2}$ closest to the nominal Z boson mass within the event; this pair is required to satisfy $40 < m_{1,2} < 120\ GeV/c^2$;
- a pair of *loose lepton candidates* (see Section 3.4), without any condition on flavor, sign and charge.

If in an event more than one candidate passes the selection, the one with the largest scalar sum of the p_T of the *loose leptons* is retained.
Because of these requirements, $Zb\bar{b}$ and $t\bar{t}$ samples are quite enlarged.

In Figures 4.2, 4.3, 4.4 and 4.5 the distributions of some significant observables are presented, showing the different contribution of signal (*Higgs*), reducible ($Zb\bar{b}$ and $t\bar{t}$), instrumental ($Z + light\ jets$) and irreducible ($qq/gg - ZZ$) backgrounds, from MC simulation. As far as signal is concerned, the sample with $m_H = 125\ GeV/c^2$ is considered.

The Figure 4.2 shows the distribution of the *good leptons* invariant mass m_{Z_1}: as one can see, the $t\bar{t}$ contribution is the only non-resonant distribution, due to the fact that it does not present any Z boson in the final state. Looking at the m_{Z_2} distribution (Figure 4.3), that is the invariant mass distribution of the loose lepton pair, in order to suppress irreducible backgrounds one can cut the distribution at high values. To fulfill the same purpose, as well as to reduce signal contribution, one can cut at low SIP_{3D} and R_{Iso} too. B hadrons, i.e. hadrons containing at least one

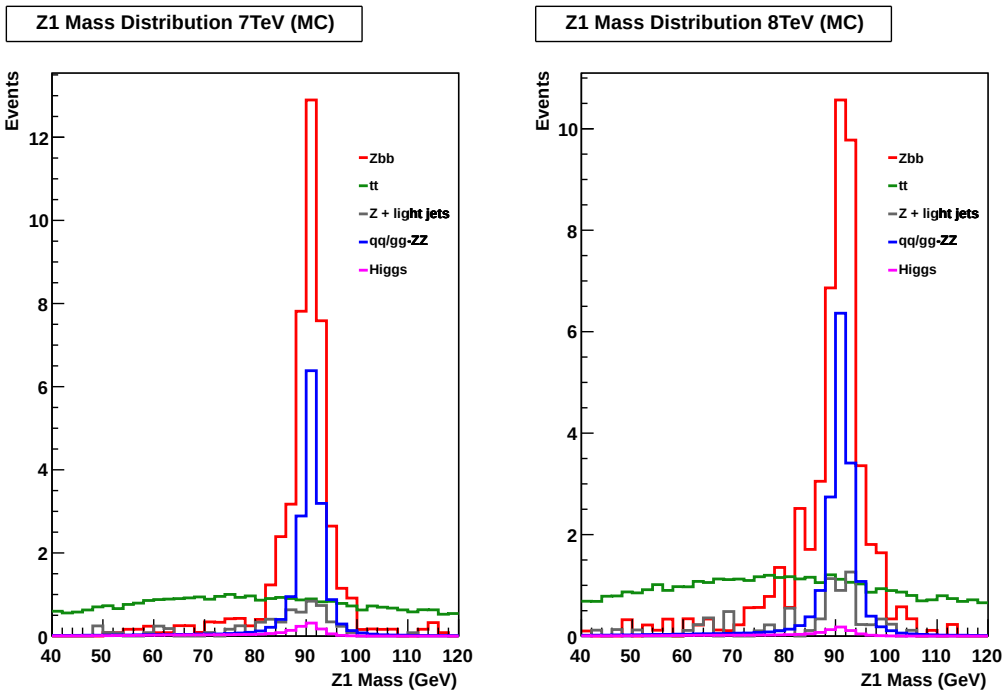


Figure 4.2: *m_{Z_1} distributions of signal and backgrounds simulation samples for 7 (left) and 8 TeV (right): no cut applied.*

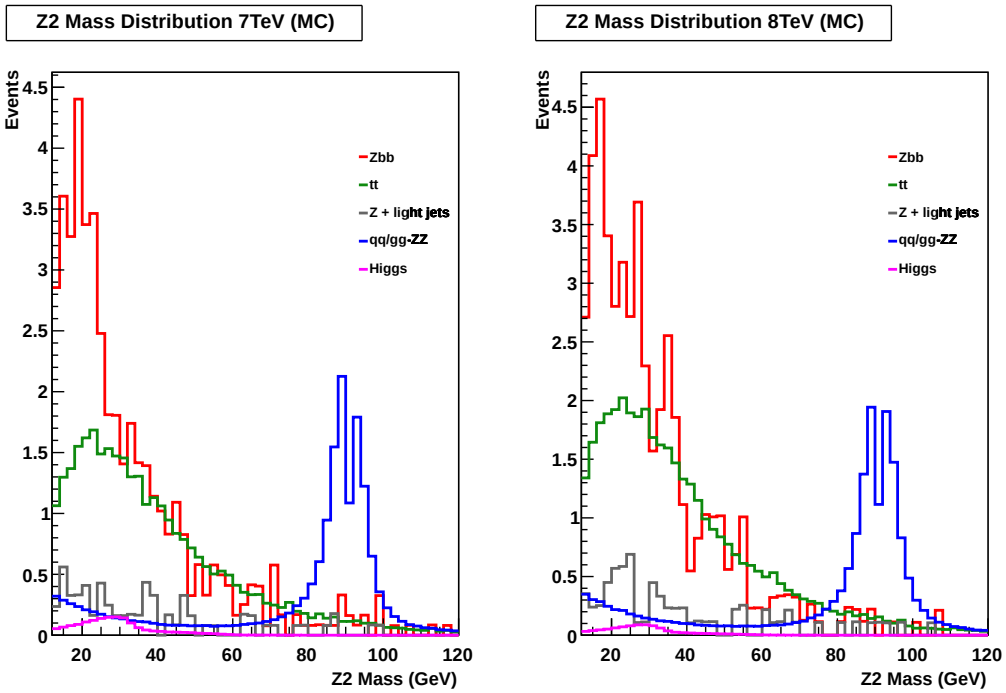


Figure 4.3: *m_{Z_2} distributions of signal and backgrounds simulation samples for 7 (left) and 8 TeV (right): no cut applied.*

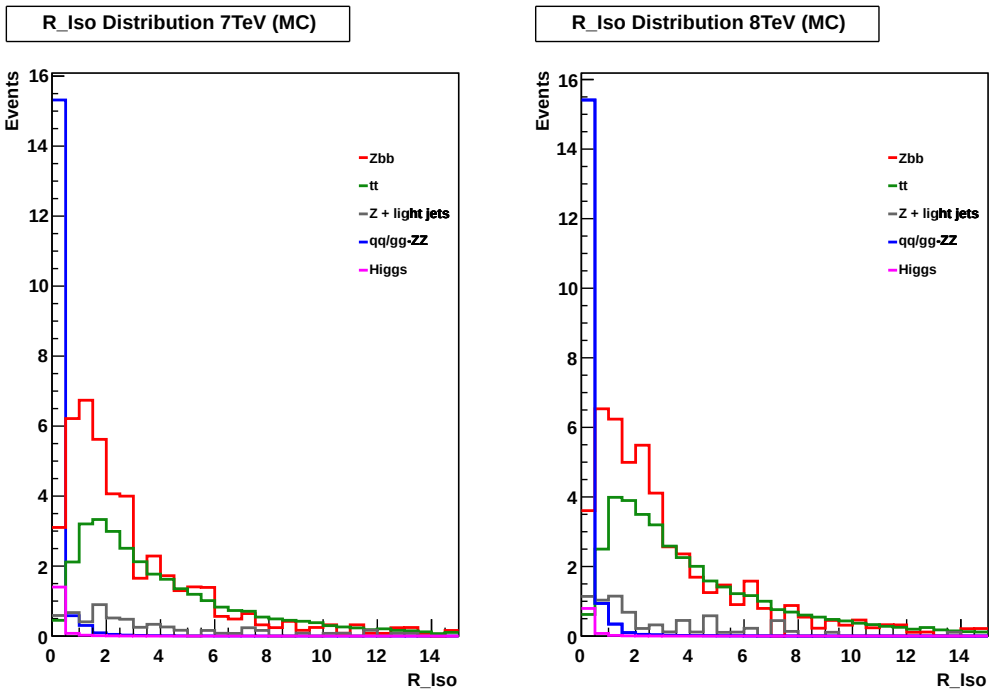


Figure 4.4: *R_{Iso} distributions of signal and backgrounds simulation samples for 7 (left) and 8 TeV (right): no cut applied.*

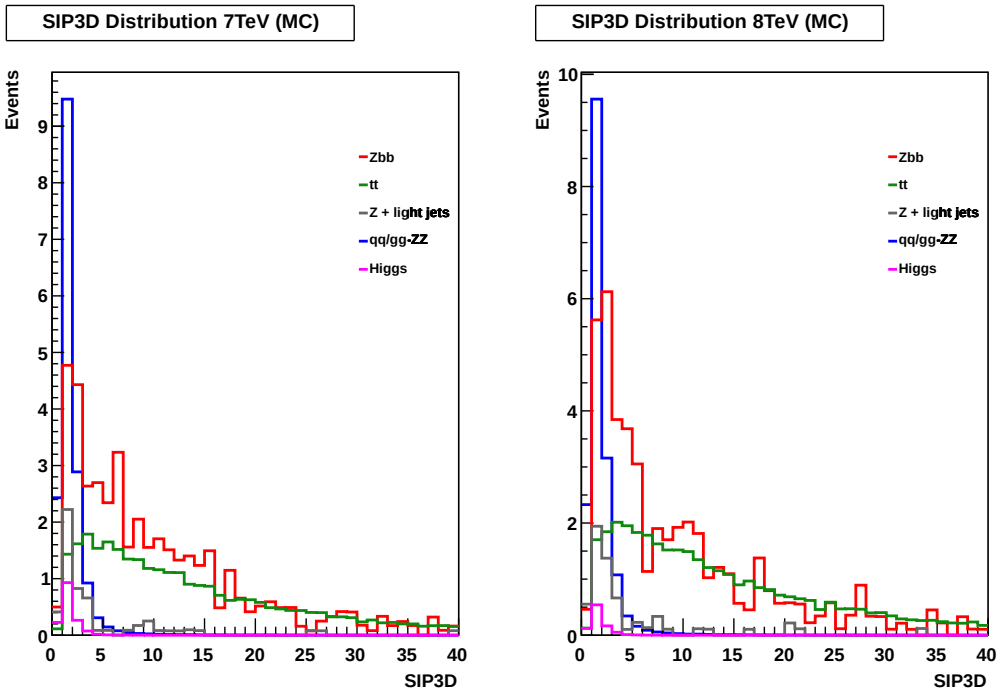


Figure 4.5: *SIP_{3D} distributions of signal and backgrounds simulation samples for 7 (left) and 8 TeV (right): no cut applied.*

b quark, have a quite large mass ($m_b \simeq 4.5\ GeV/c^2$) and a long life time ($\tau_B = 1.564 \pm 0.014$). Since $c\tau_B = 462\ \mu m$, B hadrons fly for few hundreds of micrometers, before decaying. This means that the decaying vertex of a B hadron is far from the primary vertex of the collision and the impact parameter of particles coming from a b decay is thus large. The impact parameter IP in three dimensions is the distance from the point of closest approach of the particle trajectory to the primary interaction vertex. This quantity is used to define the SIP_{3D} observable as

$$SIP_{3D} = \frac{IP}{\sigma_{IP}},$$

where σ_{IP} is the impact parameter uncertainty. Therefore, events with all four leptons coming from the primary vertex are excluded, cutting away low SIP_{3D} values. The highest SIP_{3D} value of the four leptons of each event is reported in Figure 4.5.
The R_{Iso} value is instead linked to the lepton (ℓ) isolation and its expression is

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

The $\sum p_T^{charged}$ is the scalar sum of the transverse momenta of charged hadrons originating from the primary vertex, without the lepton p_T. 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 latter excludes photons that are candidates for final-state radiation (FSR) from the lepton. The isolation variable has to be corrected using the term $\rho \times A_{eff}$, where ρ is the energy density. A_{eff} is the effective area defined as the ratio between the slope of the average isolation and ρ as a function of number of vertices. The R_{Iso} value of the least isolated lepton of each event is reported in the plot (see Figure 4.4). Since in $Zb\bar{b} \to 4\ell$ and $t\bar{t} \to 4\ell$ events two of the four final state leptons come from the hadronization of b quarks, these events are characterized by large R_{Iso} values.

Due to all these considerations, the final control region used for measuring $Zb\bar{b}$ and $t\bar{t}$ backgrounds is defined adding two more cuts, which are:

- $m_{Z_2} < 60\ GeV/c^2$;
- $SIP_{3D} > 4$.

As one can see from Figures 4.6, 4.7 and 4.8, these cuts ensure a negligible contribution of the other backgrounds in the control region and define a signal-free phase-space.
No cut on the isolation variable R_{Iso} is applied because the SIP_{3D} cut is sufficient to define the control region.

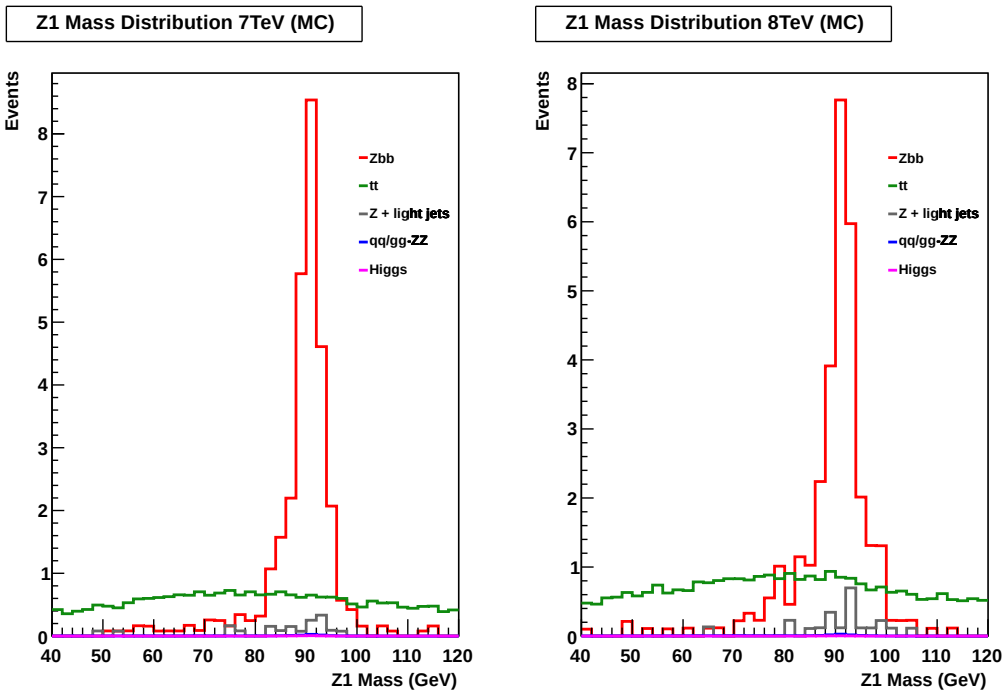


Figure 4.6: *m_{Z_1} distributions of signal and backgrounds simulation samples for 7 (left) and 8 TeV (right): m_{Z_2} and SIP_{3D} cuts applied.*

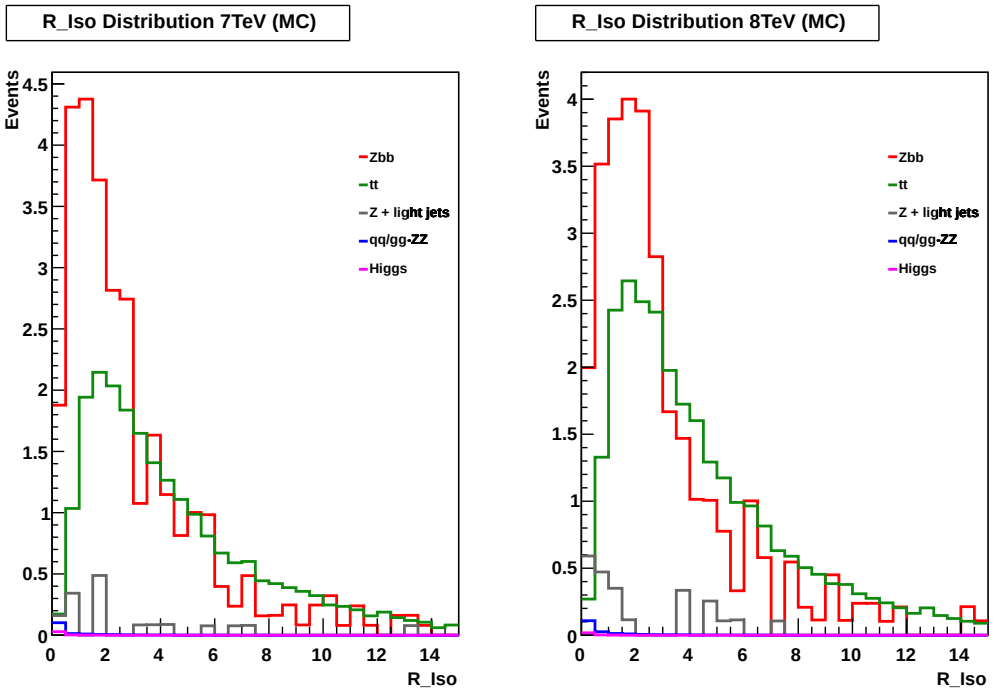


Figure 4.7: *R_{Iso} distributions of signal and backgrounds simulation samples for 7 (left) and 8 TeV (right): m_{Z_2} and SIP_{3D} cuts applied.*

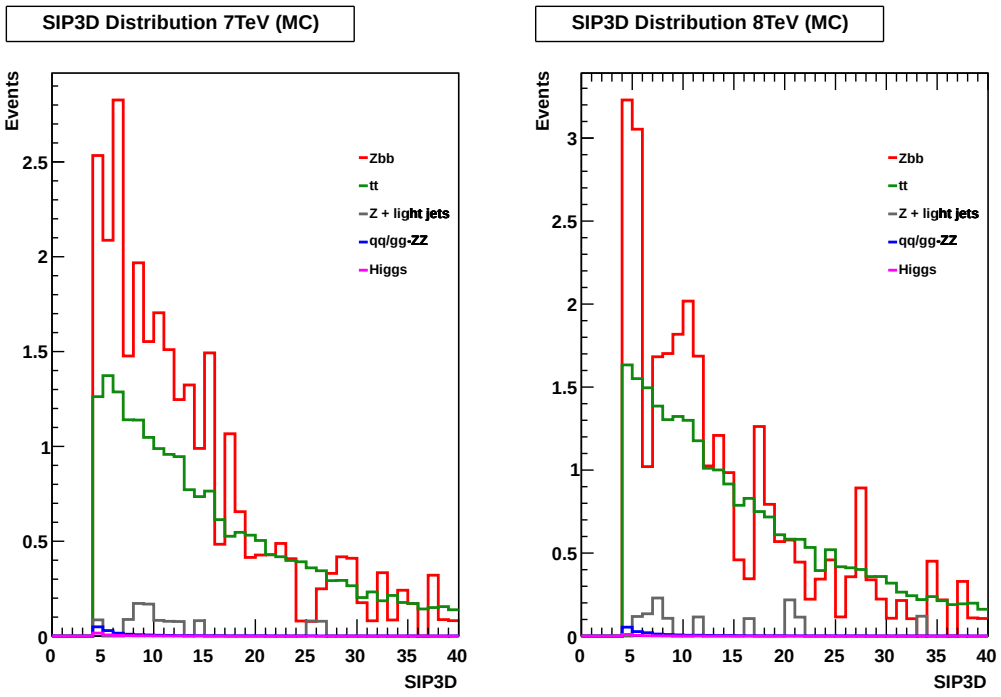


Figure 4.8: *SIP_{3D} distributions of signal and backgrounds simulation samples for 7 (left) and 8 TeV (right): m_{Z_2} and SIP_{3D} cuts applied.*

4.3 Separation of $Zb\bar{b}$ and $t\bar{t}$ Backgrounds

4.3.1 Fitting the m_{Z_1} Distribution

After applying the control region selection cuts, the invariant mass distribution of the on-mass shell Z boson m_{Z_1} (Figure 4.6) shows two main contributions that can be easily disentangled: the Z mass peak of $Zb\bar{b}$ events and the non-resonant $t\bar{t}$ background. In order to separate and measure these two distributions a fit on both Monte Carlo simulation and data is performed.
As far as Monte Carlo events are concerned, the $Zb\bar{b}$ and the $t\bar{t}$ contributions are fitted with two different PDFs. The $Zb\bar{b}$ shape is fitted using the convolution of a Breit-Wigner probability density function, to describe the theoretical resonance line shape, and a Crystal Ball function, to account for the detector resolution. The Breit-Wigner probability distribution is defined as

$$BW(m;\Gamma,m_0) = \frac{1}{N_{BW}}\frac{1}{\Gamma^2/4+(m-m_0)^2}, \tag{4.1}$$

where N_{BW} is the normalizing constant and the parameters m_0 and Γ correspond to the mass and width of the particle [16]. These two quantities

are not free parameters of the fit, but they are fixed to the Z boson nominal values [17]:

$$m_0 = 91.1876\ GeV/c^2$$
$$\Gamma = 2.4952\ GeV/c^2.$$

The Crystal Ball function consists of a Gaussian core portion with a power-law describing the left-hand tail:

$$CB(m;\alpha,n,\bar{m},\sigma) = N_{CB}\cdot\begin{cases} exp\left(-\frac{(m-\bar{m})^2}{2\sigma^2}\right) & \frac{m-\bar{m}}{\sigma} > -\alpha \\ A\cdot\left(B-\frac{m-\bar{m}}{\sigma}\right)^{-n} & \frac{m-\bar{m}}{\sigma} \leq -\alpha. \end{cases} \tag{4.2}$$

The prominence of the left-hand tail is defined by the power n and the parameter α determines where the splicing of the tails and the core is made. Parameters A and B are not independent and they are defined as

$$A = \left(\frac{n}{|\alpha|}\right)^n \cdot exp\left(-\frac{|\alpha|^2}{2}\right)$$

$$B = \frac{n}{|\alpha|} - |\alpha|$$

by requiring the continuity of the function itself and its first derivatives. N_{CB} is the normalizing constant. Since α and n are strongly correlated, in order to have a stable fit one of them is fixed. The value $n = 2$ is chosen based on studies on simulation.

The function formed by the convolution of the two distributions described above is defined as

$$F(m;\alpha,\bar{m},\sigma) = BW(m)\otimes CB(m;\alpha,\bar{m},\sigma) = \int_{-\infty}^{+\infty} BW(m)CB(m-m';\alpha,\bar{m},\sigma)\,dm'.$$

The $t\bar{t}$ distribution is fitted using a first degree Chebychev polynomial, defined for the nth-degree by the identity

$$T_n(cos\theta) = cos(n\theta). \tag{4.3}$$

The first degree polynomial is trivial, but the roofit [18] implementation guarantees that Chebychev polynomial PDFs are defined positive, which is an essential characteristic for PDFs.

For each performed fit, the free parameters are thus: the mean $\bar{m}$, the standard deviation σ and the α parameter of the Crystal Ball distribution, the Chebychev polynomial coefficient (called a_0) and the normalization constants $N_{Zb\bar{b}}$ and $N_{t\bar{t}}$ of the two background contributions.

The fitted m_{Z_1} distribution is the sum of all contributions after the selection cuts, not only $Zb\bar{b}$ and $t\bar{t}$. The Higgs signal and the $qq/gg - ZZ$ backgrounds are negligible, while $Z + light\ jets$ is not. From the MC truth, it is verified that its contribution is counted as $Zb\bar{b}$ events.
All the MC contributions are weighted considering theoretical cross sections and multiplied by the data integrated luminosity. A binned fit is performed because, since the large number of events, it is expected that dividing the sample in bins does not produce significant statistical fluctuations.
The fit is performed for $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$ separately. Results are shown in Figures 4.9 and 4.10. As it can be seen, the width of the Z_1 mass distribution is larger in the case of the electron pair, with respect to the muon final state, due to the radiative losses. The $Zb\bar{b}$ and $t\bar{t}$ cross sections differ for the 7 and 8 TeV. For this reason they are fitted separately.
The number of $Zb\bar{b}$ and $t\bar{t}$ events and the fitted parameters obtained from the MC simulation are reported in Tables 4.1 and 4.2, where only statistical errors are shown.
Comparing the number of $Zb\bar{b}$ and $t\bar{t}$ events obtained from the fit with

Table 4.1: *Fit parameters obtained from the 7 TeV MC simulation, for $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$.*

Parameter	$\mu^+\mu^-$	e^+e^-
$N_{Zb\bar{b}}$	91 ± 12	83 ± 12
$N_{t\bar{t}}$	58 ± 11	43 ± 10
$\bar{m}$	0.21 ± 0.52	0.49 ± 0.86
σ	1.08 ± 0.67	1.84 ± 0.97
α	1.05 ± 0.57	0.90 ± 0.46
a_0	-0.15 ± 0.26	0.03 ± 0.31

Table 4.2: *Fit parameters obtained from the 8 TeV MC simulation, for $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$.*

Parameter	$\mu^+\mu^-$	e^+e^-
$N_{Zb\bar{b}}$	231 ± 20	210 ± 22
$N_{t\bar{t}}$	155 ± 18	107 ± 19
$\bar{m}$	1.57 ± 2.25	1.36 ± 0.63
σ	0.27 ± 6.26	3.68 ± 0.68
α	0.32 ± 1.72	0.97 ± 0.23
a_0	-0.05 ± 0.16	-0.05 ± 0.21

the MC truth (see Tables 4.3 and 4.4), it can be seen that the fit procedure correctly estimates the $Zb\bar{b}$ and $t\bar{t}$ contributions, if considering that $Z + light\ jets$ events are counted as $Zb\bar{b}$ ones.

As far as experimental data are concerned, the same PDFs are used for the fit procedure. However, because of the low number of events, a *simultaneous fit* is performed. It is a fit in which the shape parameters of the probability distributions ($\bar{m}$, σ, α and a_0) are calculated, putting together 7 and 8 TeV data, while the normalization of the two backgrounds is obtained separating the contributions at different energy. The assumption is that the shapes of $Zb\bar{b}$ and $t\bar{t}$ are the same independently of the center-of-mass energy, while the relative normalization is not. Due to the low number of events, fits on data are unbinned.
The results of the simultaneous fit are shown in Figures 4.11 and 4.12. The number of $Zb\bar{b}$ and $t\bar{t}$ events and the parameters obtained fitting experimental data are reported in Table 4.5, where only statistical errors are shown.

4.3.2 Data/Monte Carlo Comparison

With the obtained fit results one can try to check if the fit on MC simulation represents well data or if a correction factor of the estimated number of events in data over MC is necessary. In order to fulfill this purpose the ratios of data over MC events are calculated ($R = N_{data}/N_{MC}$ values), for $Zb\bar{b}$ and $t\bar{t}$ contribution and for $Z_1 \to e^+e^-$ and $Z_1 \to \mu^+\mu^-$ separately.

Table 4.3: *MC truth for the number of $Zb\bar{b}$, $Z + light\ jets$ and $t\bar{t}$ at 7 TeV, for $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$.*

7 TeV	$\mu^+\mu^-$	e^+e^-
$N_{Zb\bar{b}}$	84	68
$N_{Z+light\ jets}$	3	5
$N_{t\bar{t}}$	61	52

Table 4.4: *MC truth for the number of $Zb\bar{b}$, $Z + light\ jets$ and $t\bar{t}$ at 8 TeV, for $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$.*

8 TeV	$\mu^+\mu^-$	e^+e^-
$N_{Zb\bar{b}}$	190	165
$N_{Z+light\ jets}$	15	13
$N_{t\bar{t}}$	178	138

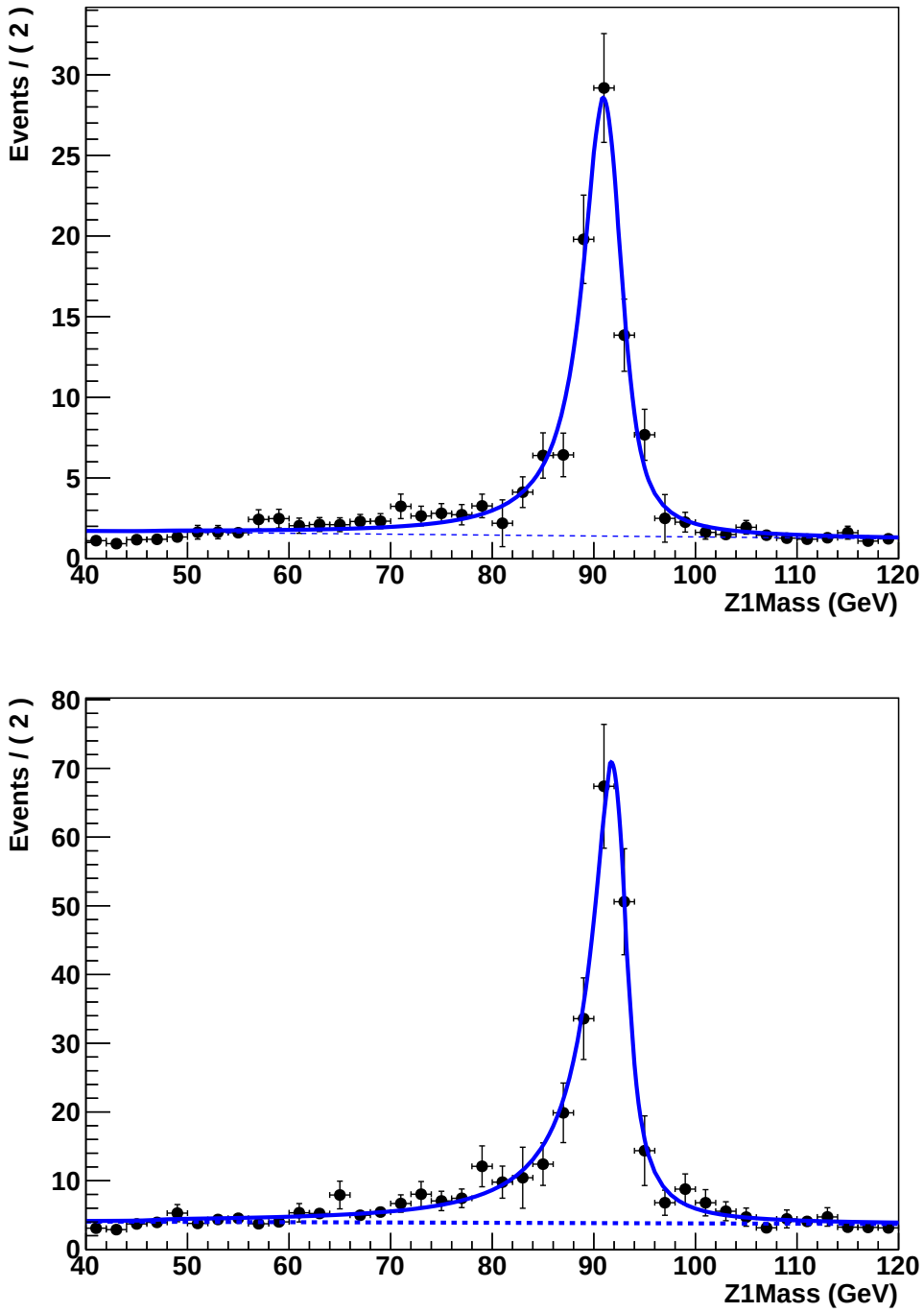


Figure 4.9: *Fit of m_{Z_1} distribution on MC simulation for $Z_1 \to \mu^+\mu^-$ at 7 (top) and 8 (bottom) TeV: points represent the MC events, the continuous line represents the best fit result and the dashed line is the $t\bar{t}$ contribution.*

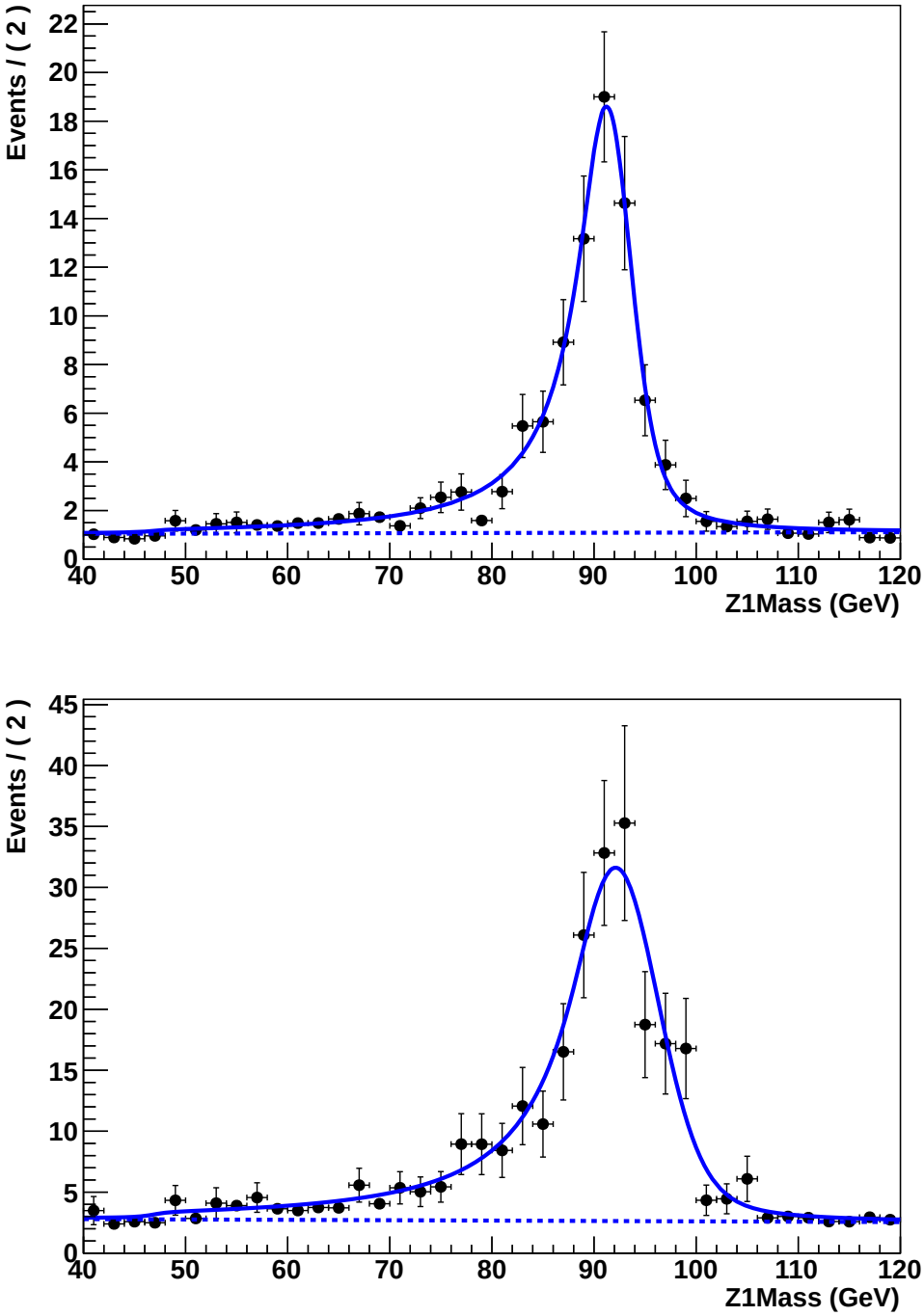


Figure 4.10: *Fit of m_{Z_1} distribution on MC simulation for $Z_1 \to e^+e^-$ at 7 (top) and 8 (bottom) TeV: points represent the MC events, the continuous line represents the best fit result and the dashed line is the $t\bar{t}$ contribution.*

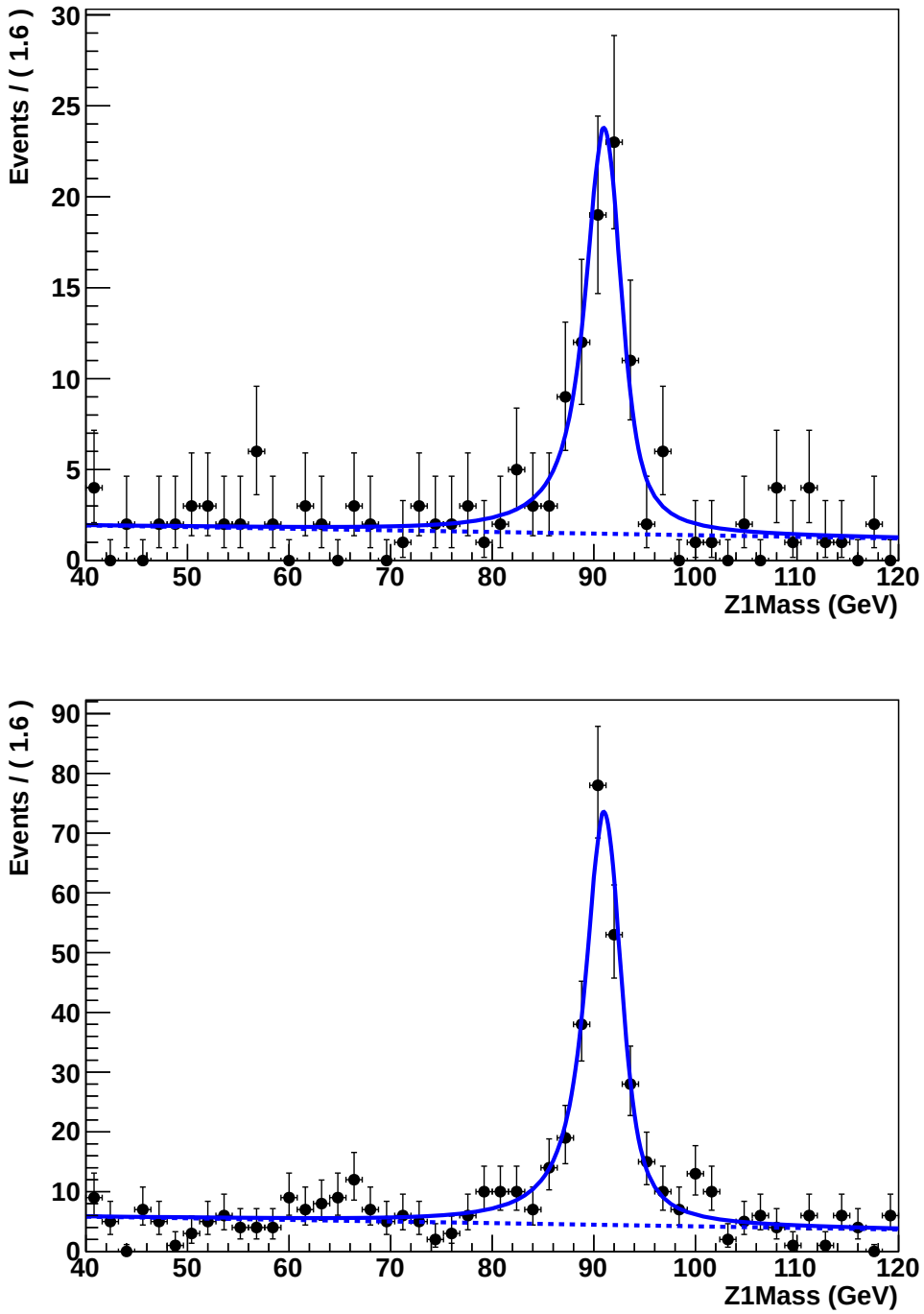


Figure 4.11: *Fit of m_{Z_1} distribution on data events for $Z_1 \to \mu^+\mu^-$ at 7 (top) and 8 (bottom) TeV: points represent data, the continuous line represents the best fit result and the dashed line is the $t\bar{t}$ contribution.*

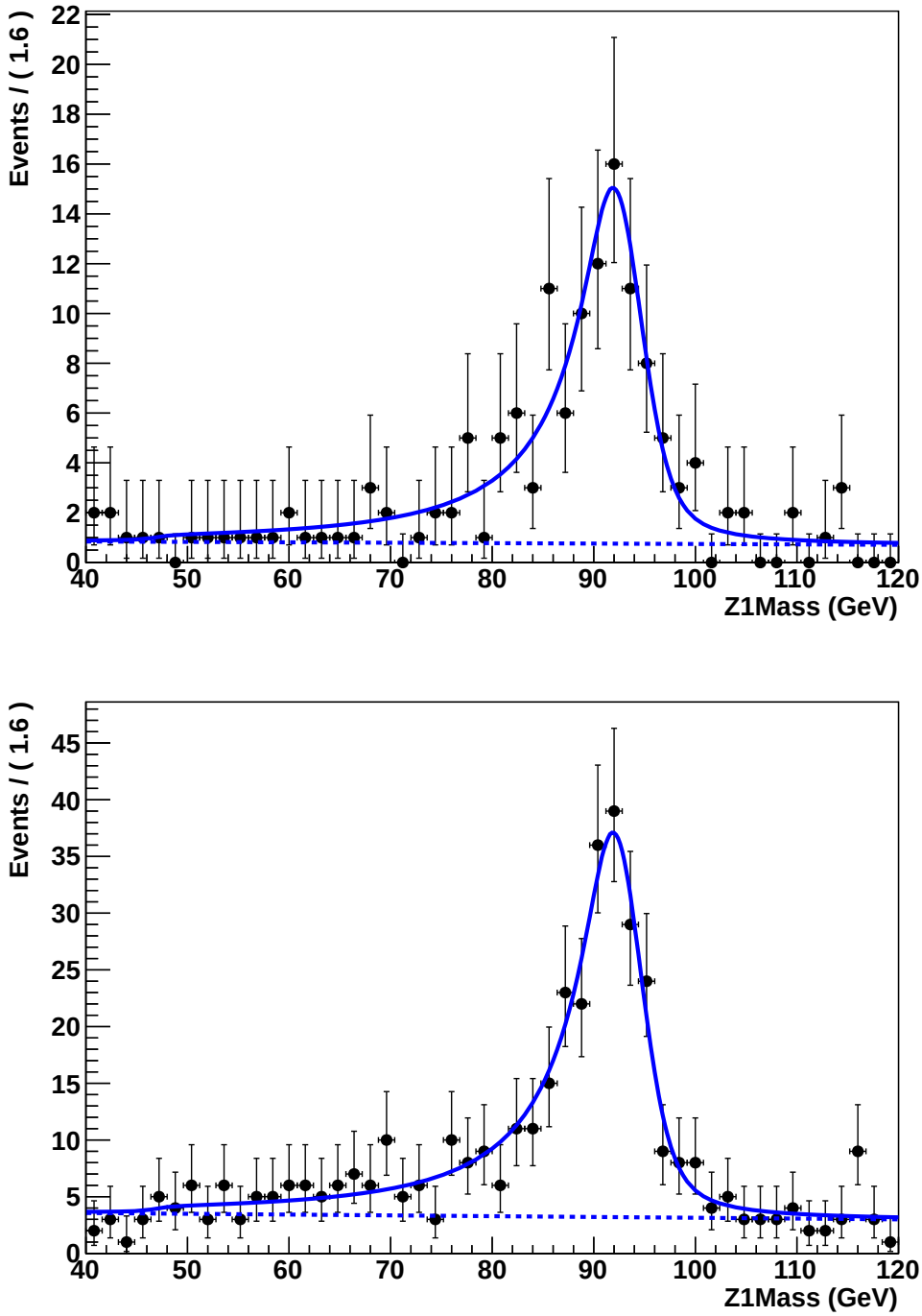


Figure 4.12: *Fit of m_{Z_1} distribution on data events for $Z_1 \to e^+e^-$ at 7 (top) and 8 (bottom) TeV: points represent data, the continuous line represents the best fit result and the dashed line is the $t\bar{t}$ contribution.*

Table 4.5: *Fit parameters obtained from the 7 and 8 TeV data, for $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$.*

Parameter	$\mu^+\mu^-$	e^+e^-
$N_{Zb\bar{b}}(7\ TeV)$	83 ± 11	106 ± 13
$N_{t\bar{t}}(7\ TeV)$	79 ± 11	39 ± 10
$N_{Zb\bar{b}}(8\ TeV)$	258 ± 20	252 ± 23
$N_{t\bar{t}}(8\ TeV)$	237 ± 20	164 ± 21
$\bar{m}$	0.23 ± 0.21	1.20 ± 0.51
σ	1.02 ± 0.29	2.21 ± 0.55
α	1.37 ± 0.37	0.71 ± 0.18
a_0	-0.23 ± 0.11	-0.09 ± 0.15

Table 4.6: *R values calculated for $Zb\bar{b}$ background, for μ and e, 7 and 8 TeV separately.*

$Zb\bar{b}$	$\mu^+\mu^-$	e^+e^-
$7\ TeV$	0.91 ± 0.24	1.28 ± 0.34
$8\ TeV$	1.12 ± 0.19	1.20 ± 0.24

Table 4.7: *R values calculated for $t\bar{t}$ background, for μ and e, 7 and 8 TeV separately.*

$t\bar{t}$	$\mu^+\mu^-$	e^+e^-
$7\ TeV$	1.36 ± 0.45	0.91 ± 0.44
$8\ TeV$	1.53 ± 0.31	1.53 ± 0.47

Values for R are reported in Tables 4.6 and 4.7. As shown, the R values are compatible with unity, within the error. In Figures 4.13 and 4.14 data and MC samples are compared, for $Z_1 \to e^+e^-$ and $Z_1 \to \mu^+\mu^-$ respectively. The green, blue and gray (stacked) histograms represent the MC simulations of $t\bar{t}$, $Zb\bar{b}$ and $Z + light\ jets$ backgrounds, points are data with statistical errors and the red line is the best fit result on data.

4.4 Cross Section Measurements

For the study of backgrounds of a search, it is important to know how much a particular background process is dominant with respect to the others, in order to estimate its contribution. In particular, the $Z + jets \to 2\ell + jets$ and $t\bar{t} \to 2\ell 2\nu 2b$ simulation samples used in the $H \to 4\ell$ analysis are generated at LO level. Since NLO and NNLO corrections may be large, it is useful to obtain measurements that do not depend on the simulation

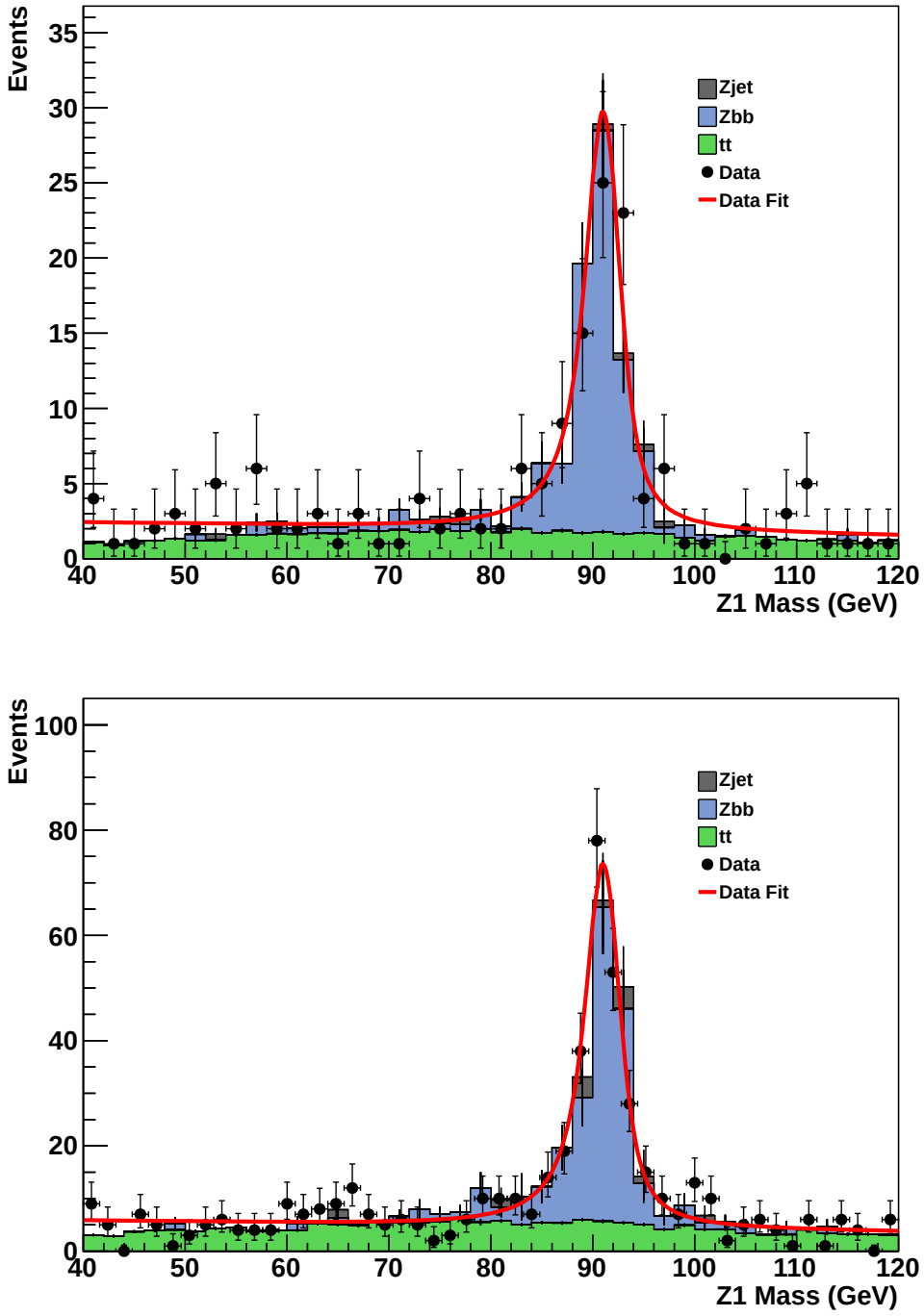


Figure 4.13: *Comparison between data and MC for the reconstructed mass m_{Z_1}, for $Z_1 \to \mu^+\mu^-$ at 7 (top) and 8 (bottom) TeV: points represent data, the continuous line represents the best fit result on data and histograms represent the $t\bar{t}$, $Zb\bar{b}$ and Z + light jets background contributions, obtained from MC.*

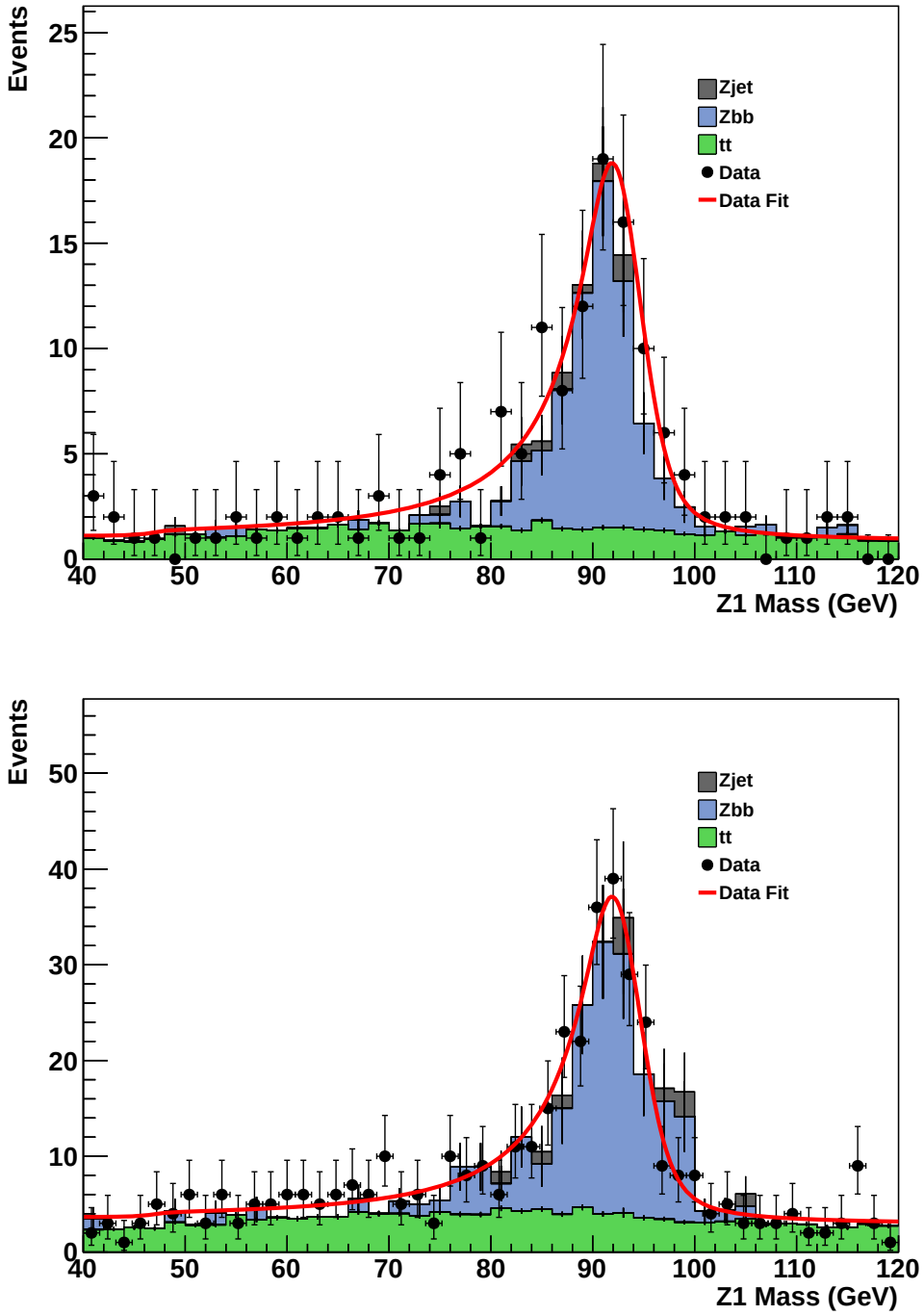


Figure 4.14: *Comparison between data and MC for the reconstructed mass m_{Z_1}, for $Z_1 \to e^+e^-$ at 7 (top) and 8 (bottom) TeV: points represent data, the continuous line represents the best fit result on data and histograms represent the $t\bar{t}$, $Zb\bar{b}$ and Z + light jets background contributions, obtained from MC.*

to verify theoretical predictions. In this Section the measurement of $Zb\bar{b} \to 4\ell$ and $t\bar{t} \to 4\ell$ cross sections is presented using the results previously obtained.
The cross section is given by the formula

$$\sigma = \frac{p \times N}{A \times \epsilon \times \mathcal{L}} \tag{4.4}$$

where

- N is the number of signal events, selected in data;
- p is the purity of the sample;
- A is the fiducial and kinematic acceptance;
- ϵ is the selection efficiency for events inside the acceptance;
- $\mathcal{L}$ is the integrated luminosity.

In this analysis, the number of signal events of $Zb\bar{b} \to 4\ell$ or $t\bar{t} \to 4\ell$ (with $\ell = e$ or μ) is extracted from the invariant mass fit described in Section 4.3.1.

The acceptance A, the efficiency ϵ and the purity p are calculated from MC simulations.
The acceptance takes into account the fraction of final state leptons that cross an active detection region of the experiment. It is defined as

$$A = \frac{N^{gen}(p_T, \eta)}{N^{gen}(all)}, \tag{4.5}$$

in which $N^{gen}(all)$ is the total number of generated events (from MC simulations) of $Zb\bar{b}$ or $t\bar{t}$ decaying in four leptons ($\ell = e, \mu$) and $N^{gen}(p_T, \eta)$ is the number of generated events, passing our kinematic and fiducial cuts. These cuts consists in requiring that all four electrons/muons have transverse momentum $p_T > 7/5\ GeV/c$ and pseudorapidity $|\eta| < 2.5/2.4$.
The signal efficiency of the selection cuts is

$$\epsilon = \frac{N^{gen}_{sel}}{N^{gen}(p_T, \eta)}, \tag{4.6}$$

where N^{gen}_{sel} is the number of selected events passing kinematic and fiducial cuts at generation level.
The number of signal events N determined from the m_{Z_1} invariant mass fit may be contaminated by several contributions. Background events may reproduce the signal m_{Z_1} distribution. For example, for the $Zb\bar{b} \to 4\ell$ final state, one can have a $Z \to 2\ell$ decay together with two fake leptons. To take

these contributions into account, the purity of our selection is determined. It is defined as

$$p = \frac{N_{sel}^{gen}}{N_{sel}^{reco}}, \tag{4.7}$$

where N_{sel}^{reco} is the total number of selected events in the sample (thus also containing events that do not fall into to our signal definition).

The contribution of these three quantities to the cross section, according to Equation 4.4, is thus simply

$$\frac{p}{A \times \epsilon} = \frac{N^{gen}(all)}{N_{sel}^{reco}}, \tag{4.8}$$

but, in order to show how each of these quantities affects the measurement, they are reported separately.
Since A, ϵ and p are derived from Monte Carlo simulations, their values are affected by theoretical uncertainties, linked to PDF and renormalization, as discussed in the following.

The luminosity $\mathcal{L}$ is different for 7 and 8 TeV and it is

$$\mathcal{L}_{7TeV} = 5.05 \pm 0.11 \ fb^{-1}$$

$$\mathcal{L}_{8TeV} = 11.34 \pm 0.50 \ fb^{-1}.$$

4.4.1 $Zb\bar{b} \to 4\ell$ Cross Section

As far as the $Zb\bar{b} \to 4\ell$ sample is concerned, the signal at generation level is defined as generated events with four leptons (e or μ) in the final state, two of which coming directly from a Z boson (no $Z \to \tau^+\tau^- \to \ell^+\ell^-$ events are considered) and with at least two leptons originating from bottom and charm hadrons. Electrons and muons that come from a kaon or a pion are not counted. As written above, electrons/muons must have $p_T > 7/5 \ GeV/c$ and $|\eta| < 2.5/2.4$ and the corresponding acceptance must be determined. The total amount of generated events $N^{gen}(all)$ and of generated events passing acceptance criteria $N^{gen}(p_T, \eta)$ is extracted from the MC samples. N_{sel}^{gen} is instead determined by checking how many reconstructed events selected by the analysis cuts N_{sel}^{reco} come from the $Zb\bar{b} \to 4\ell$ generated subsample. As described in Section 4.2, the mass of the loose lepton pair m_{Z_2} must be smaller than 60 GeV/c^2 and the significance of the impact parameter SIP_{3D} larger than 4. These numbers (reported in Table 4.8 for 7 and 8 TeV) allow calculating the acceptance, the efficiency and the purity of samples, that are

$$A^{7TeV} = (0.295 \pm 0.007)\%$$

$$\epsilon^{7TeV} = (14.23 \pm 0.80)\%$$
$$p^{7TeV} = (74.6 \pm 2.3)\%$$

for the 7 TeV sample and

$$A^{8TeV} = (0.265 \pm 0.007)\%$$
$$\epsilon^{8TeV} = (12.66 \pm 0.82)\%$$
$$p^{8TeV} = (74.8 \pm 2.6)\%$$

for the 8 TeV sample.
Acceptance, efficiency and purity errors are calculated considering the number of events that pass the selection as a binomial variable. The efficiency error is

$$\delta\epsilon = \sqrt{\frac{\epsilon(1-\epsilon)}{N}}. \tag{4.9}$$

An analogous formula is used for the acceptance and for the purity. N is the total of events on which the efficiency, the acceptance and the purity are calculated ($N^{gen}(all)$ for A, $N^{gen}(p_T, \eta)$ for ϵ and N^{reco}_{sel} for p).
The acceptance is affected by a theoretical systematic uncertainty due to the choice of PDF, the QCD energy scale and α_s. This systematic uncertainty has not been computed in the present work and will have to be determined with dedicated studies.
The acceptance is so small mainly because of the cut on the transverse momentum. Considering a general process $X \to x_1 + x_2$, one obtains that

$$M_X^2 = m_1^2 + m_2^2 + 2E_1E_2 - 2p_1p_2cos\theta,$$

where m_1 and m_2 are the masses of decay particles and θ the polar angle between them. If $m_{1,2} << M_X$, $m_{1,2} \simeq 0$ and the previous expression becomes

$$M_X^2 = 2p_1p_2(1 - cos\theta).$$

Assuming that the parallel component to the beam line of the produced particles momenta is equal to zero, one has

$$M_X^2 = 2p_{T1}p_{T2}(1 - cos\phi_{12}),$$

where p_{Ti} are the transverse momenta of the two particles and ϕ_{12} is the angle between them, in the transverse plane. If X is produced at rest, this angle is of 180 degree and so

$$M_X^2 = 4p_{T1}^2 \Rightarrow M_X = 2p_{T1}.$$

A heavy decaying object thus produces particles with a p_T of about $M_X/2$. Leptons coming from a Z boson have, for example, a $p_T \sim 45\ GeV/c$

Table 4.8: *Total number of generated events $N^{gen}(all)$, generated events passing kinematic and fiducial cuts $N^{gen}(p_T, \eta)$, selected events passing acceptance cuts at generation level N^{gen}_{sel} and total number of selected events N^{reco}_{sel} for $Zb\bar{b} \to 4\ell$ signal, for 7 and 8 TeV.*

$Zb\bar{b} \to 4\ell$	7 *TeV*	8 *TeV*
$N^{gen}(all)$	651466	619688
$N^{gen}(p_T, \eta)$	1919	1643
$N^{gen}_{sel}(p_T, \eta)$	273	208
$N^{reco}_{sel}(Zb\bar{b})$	366	278
$N^{reco}_{sel}(Z + light\ jets)$	20	21

and pass the cut. The transverse momentum of leptons originated by the decay of b quarks is instead lower, because B hadrons have a mass of about 5/10 GeV/c^2. It is therefore more probable that these leptons don't pass the p_T cut.

The number N of signal events contained in Equation 4.4 is determined from data and it is reported in Table 4.5, for the $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$ cases and for 7 and 8 *TeV* separately (see Section 4.3.1). The $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$ are merged and a single result is reported. It is obtained, for 7 and 8 *TeV*,

$$N^{7TeV}_{Zb\bar{b}} = 189 \pm 17$$

$$N^{8TeV}_{Zb\bar{b}} = 510 \pm 30.$$

Since, as seen, a $Z + light\ jets$ event fraction is contained in the total amounts, one must subtract the fake contribution. It can be done from MC simulation, adding a new factor on Equation 4.4: it is the "$Zb\bar{b}$ purity" defined as

$$p_{Zb\bar{b}} = \frac{N^{reco}_{sel}(Zb\bar{b})}{N^{reco}_{sel}(Zb\bar{b}) + N^{reco}_{sel}(Z + light\ jets)}, \tag{4.10}$$

i.e. how much $Zb\bar{b}$ events are selected over the sum of $Zb\bar{b}$ plus $Z + light\ jets$. The Equation 4.4 thus becomes

$$\sigma_{Zb\bar{b}\to 4\ell} = \frac{p \times p_{Zb\bar{b}} \times N}{A \times \epsilon \times \mathcal{L}}. \tag{4.11}$$

The $Zb\bar{b}$ purity values are (see Table 4.8)

$$p^{7TeV}_{Zb\bar{b}} = (94.8 \pm 1.1)\%$$

$$p^{8TeV}_{Zb\bar{b}} = (93.0 \pm 1.5)\%,$$

where errors are determined according to Equation 4.9.
Putting together all these information and considering the luminosity values ($\mathcal{L} = 5.051\ fb^{-1}$ at 7 TeV and $\mathcal{L} = 11.341\ fb^{-1}$ at 8 TeV), it is thus possible to calculate the $Zb\bar{b} \to 4\ell$ cross section for 7 and 8 TeV, obtaining

$$\sigma^{7TeV}_{Zb\bar{b}\to 4\ell} = 63.1 \pm 5.7 \pm 4.2 \pm 1.4\ pb$$

$$\sigma^{8TeV}_{Zb\bar{b}\to 4\ell} = 93.2 \pm 5.5 \pm 8.0 \pm 4.1\ pb.$$

The first uncertainty that is reported is the statistical one and it is calculated according to

$$\delta\sigma_{stat} = \frac{p\ p_{Zb\bar{b}}}{A\epsilon\mathcal{L}}\delta N, \tag{4.12}$$

taking into account only the uncertainty on the number of events estimated from the fit to the data. The second term represents systematic errors and it contains the reported uncertainties of the acceptance, the efficiency and the purities (p and p_{Zbb}). It is calculated as

$$\delta\sigma_{syst} = \frac{N}{A\epsilon\mathcal{L}}\sqrt{\left(pp_{Zb\bar{b}}\frac{\delta\epsilon}{\epsilon}\right)^2 + \left(pp_{Zb\bar{b}}\frac{\delta A}{A}\right)^2 + \left(p_{Zb\bar{b}}\frac{\delta p}{p}\right)^2 + \left(p\frac{\delta p_{Zb\bar{b}}}{p_{Zb\bar{b}}}\right)^2} \tag{4.13}$$

The third term is the uncertainty due to the luminosity, i.e.

$$\delta\sigma_{lumi} = \frac{p\ p_{Zb\bar{b}}\ N}{A\epsilon\mathcal{L}}\frac{\delta\mathcal{L}}{\mathcal{L}}. \tag{4.14}$$

The theoretical error related to the acceptance determination is not yet calculated.
Up to now no theoretical calculation exists at NLO level for $Zb\bar{b} \to 4\ell$ cross section, thus it is not possible to make a comparison between the measurement and a theoretical prediction. Theoretical calculations are presently being developed.

4.4.2 $t\bar{t} \to 4\ell$ Cross Section

As far as the $t\bar{t} \to 4\ell$ sample is concerned, the signal at generation level is defined as generated events with four leptons (e or μ) in the final state, at least two of which coming directly from a W boson (no $W \to \tau\ \nu_\tau \to \ell\ \nu_\ell\ \nu_\tau$ events are considered) and with at least two leptons deriving from bottom and charm hadrons. Electrons and muons that come from a kaon or a pion are not counted. As in the $Zb\bar{b} \to 4\ell$ case, electrons/muons must have $p_T > 7/5\ GeV/c$ and $|\eta| < 2.5/2.4$ and the relative acceptance must be determined. The total amount of generated events $N^{gen}(all)$ and of generated events passing acceptance criteria $N^{gen}(p_T, \eta)$ is extracted from the MC production. N^{gen}_{sel} is instead determined by checking how many

Table 4.9: *Total number of generated events $N^{gen}(all)$, generated events passing kinematic and fiducial cuts $N^{gen}(p_T, \eta)$, selected events passing acceptance cuts at generation level N^{gen}_{sel} and total number of selected events N^{reco}_{sel} for $t\bar{t} \to 4\ell$ signal, for 7 and 8 TeV.*

$t\bar{t} \to 4\ell$	7 *TeV*	8 *TeV*
$N^{gen}(all)$	1063474	1434006
$N^{gen}(p_T, \eta)$	212779	276932
$N^{gen}_{sel}(p_T, \eta)$	9897	-
N^{reco}_{sel}	17108	-

reconstructed events selected by the analysis cuts N^{reco}_{sel} come from the $t\bar{t} \to 4\ell$ generated subsample. As previously described, the mass of the loose lepton pair m_{Z_2} must be smaller than 60 GeV/c^2 and the significance of the impact parameter SIP_{3D} larger than 4. The obtained numbers (reported in Table 4.9 for 7 and 8 *TeV*) allow calculating

$$A^{7TeV} = (20.00 \pm 0.04)\%$$

$$\epsilon^{7TeV} = (4.651 \pm 0.046)\%$$

$$p^{7TeV} = (57.85 \pm 0.38)\%$$

for the 7 *TeV* sample and

$$A^{8TeV} = (19.31 \pm 0.3)\%$$

for the 8 *TeV* sample.
In this last case it was not possible, for technical reasons, to determine the efficiency and the purity with the available MC samples. It is thus assumed the 7 *TeV* values even for the 8 *TeV* sample too.
Efficiency and purity errors are calculated using Equation 4.9.

The number N of signal events contained in Equation 4.4 is determined from data and it is reported in Table 4.5, for $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$ cases and for 7 and 8 *TeV* separately (see Section 4.3.1). The $Z_1 \to \mu^+\mu^-$ and $Z_1 \to e^+e^-$ are merged and a single result is reported. It is obtained, for 7 and 8 *TeV*,

$$N^{7TeV}_{t\bar{t}} = 118 \pm 15$$

$$N^{8TeV}_{t\bar{t}} = 401 \pm 29.$$

Putting together all these information and considering the luminosity values ($\mathcal{L} = 5.051\ fb^{-1}$ at 7 *TeV* and $\mathcal{L} = 11.341\ fb^{-1}$ at 8 *TeV*), it is thus possible to calculate the $t\bar{t} \to 4\ell$ cross section for 7 and 8 *TeV*, obtaining

$$\sigma^{7TeV}_{t\bar{t}\to 4\ell} = 1.452 \pm 0.184 \pm 0.017 \pm 0.032\ pb$$

$$\sigma^{8TeV}_{t\bar{t}\to 4\ell} = 2.277 \pm 0.215 \pm 0.035 \pm 0.100 \ pb.$$

The first uncertainty that is reported is the statistical one, determined taking into account only the uncertainty of the number of events estimated from data (see Equation 4.12). The second term represents systematic uncertainties and it contains the reported uncertainties of the acceptance, the efficiency and the purity (Equation 4.13). The third term is the uncertainty due to the luminosity obtained from Equation 4.14.
An additional systematic uncertainty is present due to the assumption that the efficiency and the purity at 8 TeV are the same of the ones determined for 7 TeV. This uncertainty has not yet been evaluated.
The theoretical uncertainty related to the acceptance determination is not yet calculated.
As far as theoretical predictions are concerned, the theoretical cross section for $t\bar{t}$ process is $\sim 164/252$ pb at $7/8$ TeV [19]. The measured cross sections are compatible with these predictions within error, once the branching ratio of W bosons to leptons (μ or e) and the branching ratio of b quarks to leptons (both $b \to \ell$ and $b \to c \to \ell$ cases) are taken into account:

$$\sigma^{7TeV\ Th}_{t\bar{t}\to 4\ell} \sim 1.25 \ pb$$

$$\sigma^{8TeV\ Th}_{t\bar{t}\to 4\ell} \sim 1.92 \ pb.$$

Chapter 5

New Method for Instrumental and Reducible Backgrounds Estimation

In this chapter a new data-driven procedure to separately measure instrumental and reducible backgrounds is presented. These backgrounds remain after the first steps of the "$H \to ZZ \to 4\ell$" selection. The method used in the $H \to 4\ell$ CMS analysis estimates them in a inclusive way (see Section 3.5.2). As soon as the available integrated luminosity increases, it becomes useful to separate instrumental from reducible background in order to better evaluate their contributions and reduce systematic errors.

In the following, a control sample is defined in order to study the contributions of instrumental and reducible backgrounds. A fit is performed on the SIP_{3D} distributions, that is the observable that better separates $Z + light\ jets$ from $Zb\bar{b}$ and $t\bar{t}$. The number of expected $Z + light\ jets$ and $Zb\bar{b} + t\bar{t}$ events is obtained in the control sample and then it is extrapolated to the Higgs signal region.

5.1 Definition of the Control Sample

As explained in Section 3.5.2, the instrumental background indicates background events with final state leptons from misidentification of other particles, such as QCD multi-jets and $Z/W + light\ jets$ processes where lepton candidates mainly come from jets faking leptons. $Z + light\ jets$ (for simplicity just called $Z + jets$ from now on) events thus present in the final state non-isolated leptons. Moreover this background is independent from charge and flavor of misidentified particles.

The control sample used to evaluate instrumental and reducible back-

grounds from experimental data is obtained as set of the events that satisfy the "*first Z*" step of preselection (see Section 3.4), that have the reconstructed four-lepton invariant mass $m_{4\ell} > 100\ GeV/c^2$ and the invariant mass of the additional pair of reconstructed leptons (called *loose leptons*) of $12 < m_{Z_2} < 60\ GeV/c^2$.
The data control sample is divided in three different subsets, according to the charge and the flavor of the *loose lepton* pair:

- same-sign sample (SS): it is defined by requiring *loose leptons* of same sign ($\mu^{\pm}\mu^{\pm}$, $e^{\pm}e^{\pm}$, $\mu^{\pm}e^{\pm}$ and $e^{\pm}\mu^{\pm}$);
- opposite-sign-different-flavor sample (OS-DF): it is defined by requiring *loose leptons* of opposite sign and different flavor ($\mu^{\pm}e^{\mp}$ and $e^{\pm}\mu^{\mp}$);
- opposite-sign-same-flavor sample (OS-SF): it is defined by requiring *loose leptons* of opposite sign and same flavor ($\mu^{\pm}\mu^{\mp}$ and $e^{\pm}e^{\mp}$).

The division SS/OS is due to the fact that the number of $Zb\bar{b}$ events, unlike the $Z + jets$ events, depends on the sign of leptons. B mesons can in fact decay semileptonically in different ways: $b \to \ell^- c$ with a branching ratio $\mathcal{B} \sim 10.7\%$, that includes $b \to \ell^- c \to \ell^-\ell^-$ (with $\mathcal{B} \sim 2\%$), and $b \to \bar{c} \to \ell^+$ with $\mathcal{B} \sim 8\%$. The $Zb\bar{b} + t\bar{t}$ distribution has thus a different normalization in these cases. Moreover the opposite-sign sample is split according to the lepton flavor, in order to take into account that the OS-SF case can contain a contribution from ZZ and possibly from Higgs boson decaying to $\ell^+\ell^-\ell'^+\ell'^-$, while the OS-DF sample is formed only by instrumental and reducible backgrounds.

5.2 Fitting SIP_{3D} Distribution

In order to distinguish $Z + jets$ events from reducible backgrounds ($Zb\bar{b}$ and $t\bar{t}$), the distribution of the largest SIP_{3D} value between the *loose leptons* is studied (see Sections 3.3.1 and 4.2). The impact parameter significance is the observable that better distinguishes the two different contributions. As one can see from Figure 5.1, $Z + jets$ MC events are mainly concentrated at low SIP_{3D} values, while the $Zb\bar{b}$ and $t\bar{t}$ distributions present a longer tail at large values. From Figure 5.1 it can be noted that there is a significant discrepancy between data and MC distributions, in particular at low SIP_{3D} values. Since, as seen in Section 4.3.2, the MC samples of $Zb\bar{b}$ and $t\bar{t}$ events describe well the data, it is assumed that this discrepancy is due mainly to the $Z + jets$ contribution.
To separate instrumental from reducible backgrounds, the SIP_{3D} distributions of the SS, OS-DF and OS-SF samples are fitted, using two different

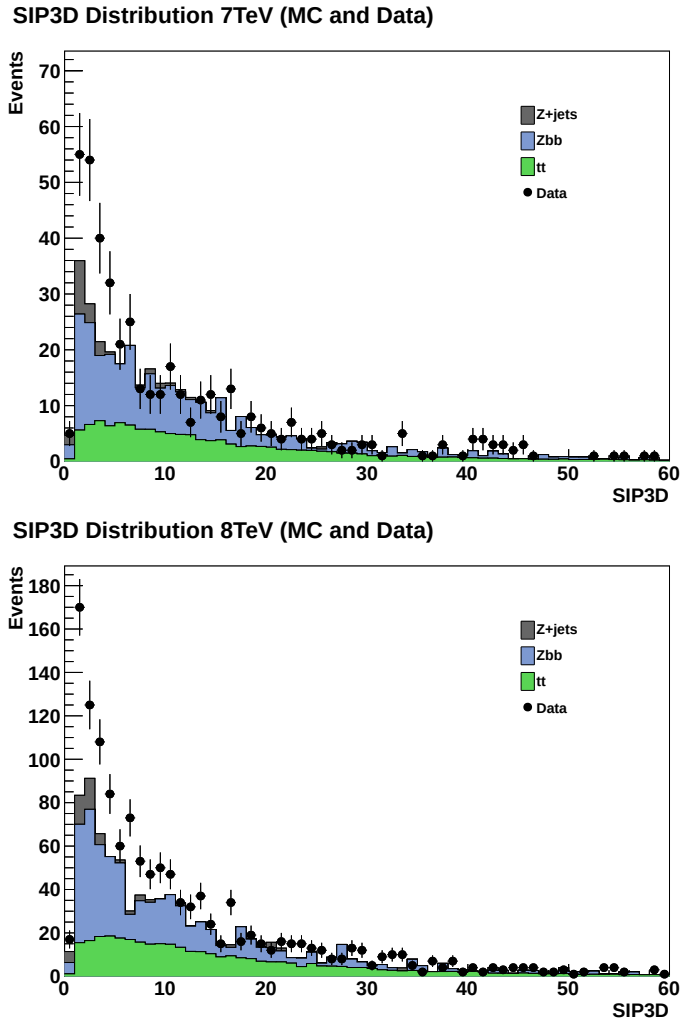


Figure 5.1: *SIP_{3D} distribution of $Z + jets$, $Zb\bar{b}$ and $t\bar{t}$ MC samples and data, for a luminosity $\mathcal{L} = 5.05\ fb^{-1}$ at 7 TeV (top) and $\mathcal{L} = 11.34\ fb^{-1}$ at 8 TeV (bottom).*

Landau functions for $Z + jets$ and $Zb\bar{b} + t\bar{t}$. Because of the low number of events and the amount of parameters to determine, a simultaneous fit of these samples is performed. The shape parameters of the probability distribution functions (mean and sigma of each Landau) are calculated over the whole statistics, while the normalization of the two contributions is obtained separately for the SS, OS-DF and OS-SF subsets. It is assumed therefore that the shapes of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ are the same independently from sign and flavor of the *loose leptons*, while the relative normalization is not. Since $Z + jets$ and $Zb\bar{b} + t\bar{t}$ cross sections depend on the center-of-mass energy, the simultaneous fit is performed for 7 and 8 *TeV* data samples separately. Due to the low number of events, fits are

unbinned.
The results of simultaneous fits are shown in Figures 5.2 and 5.3, for 7 and 8 *TeV* respectively. The number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events for the SS, OS-DF and OS-SF samples and distribution parameters obtained from the fit are reported in Tables 5.1 and 5.2, where only statistical errors are shown. A discrepancy between data and MC simulation is present in all three subsets. This indicates that such discrepancy comes from a poor modelling of the $Z + jets$ MC sample.
In both $Z + jets$ and $Zb\bar{b} + t\bar{t}$ cases, at least two leptons are generated inside jets and they are thus not isolated. In order to verify this characteristic, the same simultaneous fit described above is performed, after a cut to select high values of the isolation variable R_{Iso} defined in Sections 3.3.1 and 4.2. SS, OS-DF and OS-SF subsets are therefore re-defined, requiring that the least isolated lepton of each event has $R_{Iso} > 0.4$, and fitted. The result is shown in Figures 5.4 and 5.5. The number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events for the three different samples and distribution parameters obtained from the fits are reported in Tables 5.1 and 5.2, where only statistical errors are shown.

Table 5.1: *Fit parameters obtained from the 7 TeV sample, with and without the R_{Iso} cut: $N_{Zb\bar{b}+t\bar{t}}$ and N_{Z+jets} are the normalization constants obtained for SS, OS-DF and OS-SF subsets; $m_{1,2}$ and $\sigma_{1,2}$ are the mean and the sigma of the Landau function that fits the $Zb\bar{b} + t\bar{t}$/$Z + jets$ distribution.*

Parameter	$R_{Iso}(All)$	$R_{Iso} > 0.4$
$N_{Zb\bar{b}+t\bar{t}}(SS)$	97 ± 19	92 ± 17
$N_{Z+jets}(SS)$	69 ± 19	64 ± 16
$\chi^2_R(SS)$	2.23	2.14
$N_{Zb\bar{b}+t\bar{t}}(OS-DF)$	54 ± 17	51 ± 14
$N_{Z+jets}(OS-DF)$	60 ± 17	56 ± 15
$\chi^2_R(OS-DF)$	1.37	1.27
$N_{Zb\bar{b}+t\bar{t}}(OS-SF)$	71 ± 17	70 ± 14
$N_{Z+jets}(OS-SF)$	71 ± 17	50 ± 14
$\chi^2_R(OS-SF)$	1.77	1.86

Parameter	$R_{Iso}(All)$	$R_{Iso} > 0.4$
$\bar{m}_1$	8.9 ± 2.1	9.2 ± 1.7
σ_1	3.68 ± 0.46	3.60 ± 0.43
$\bar{m}_2$	1.69 ± 0.22	1.79 ± 0.23
σ_2	0.63 ± 0.14	0.66 ± 0.14

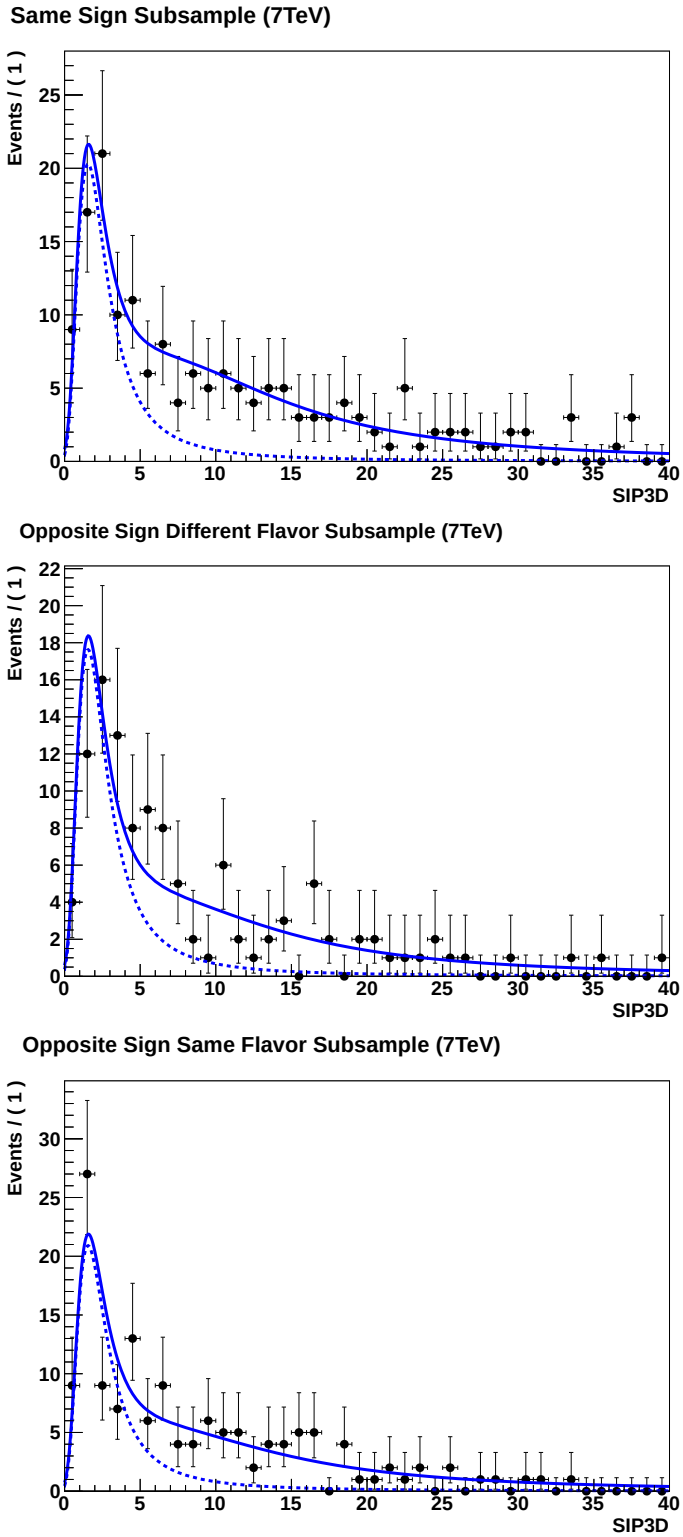


Figure 5.2: *SIP_{3D} distributions of Z + jets and $Zb\bar{b} + t\bar{t}$ for SS (top), OS-DF (center) and OS-SF (bottom) samples at 7 TeV: points represent data, the continuous line represents the best fit result and the dashed line is the Z + jets contribution.*

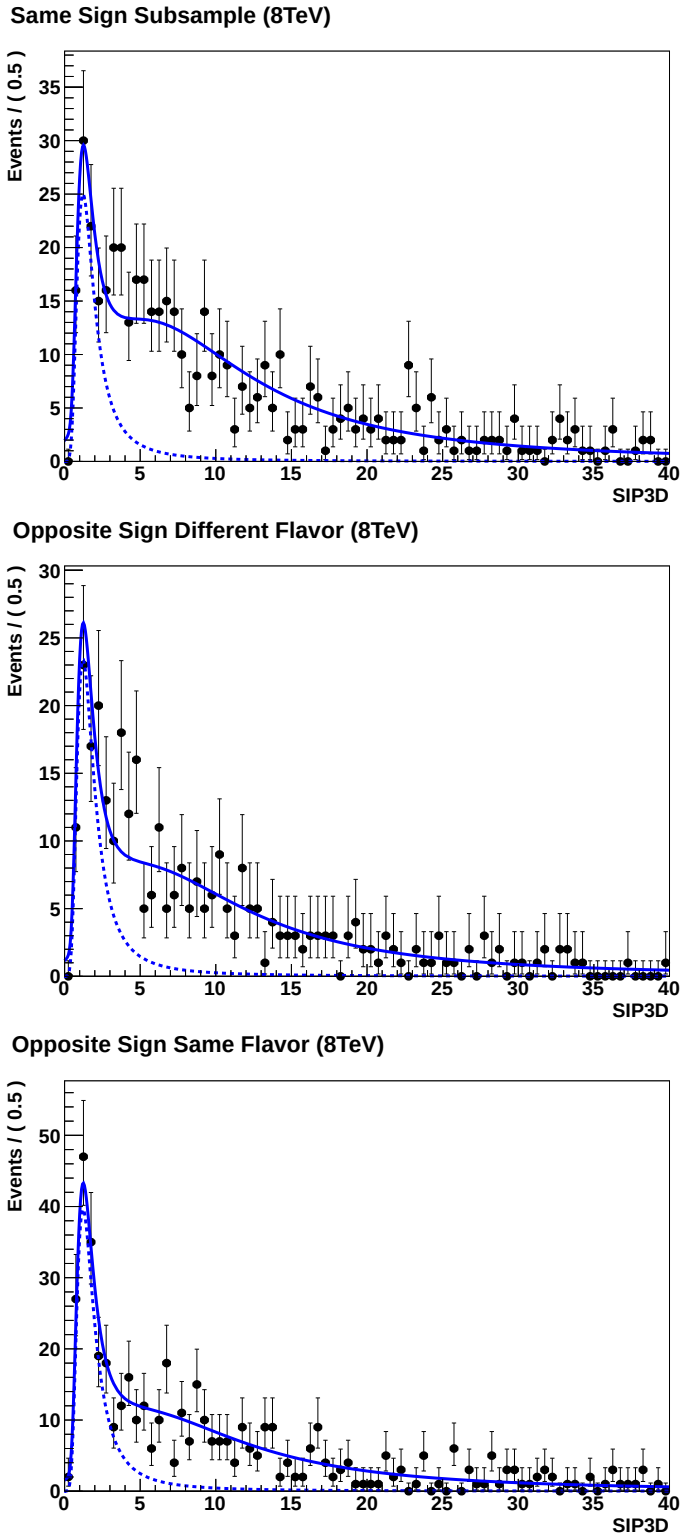


Figure 5.3: *SIP_{3D} distributions of Z + jets and $Zb\bar{b} + t\bar{t}$ for SS (top), OS-DF (center) and OS-SF (bottom) samples at 8 TeV: points represent data, the continuous line represents the best fit result and the dashed line is the Z + jets contribution.*

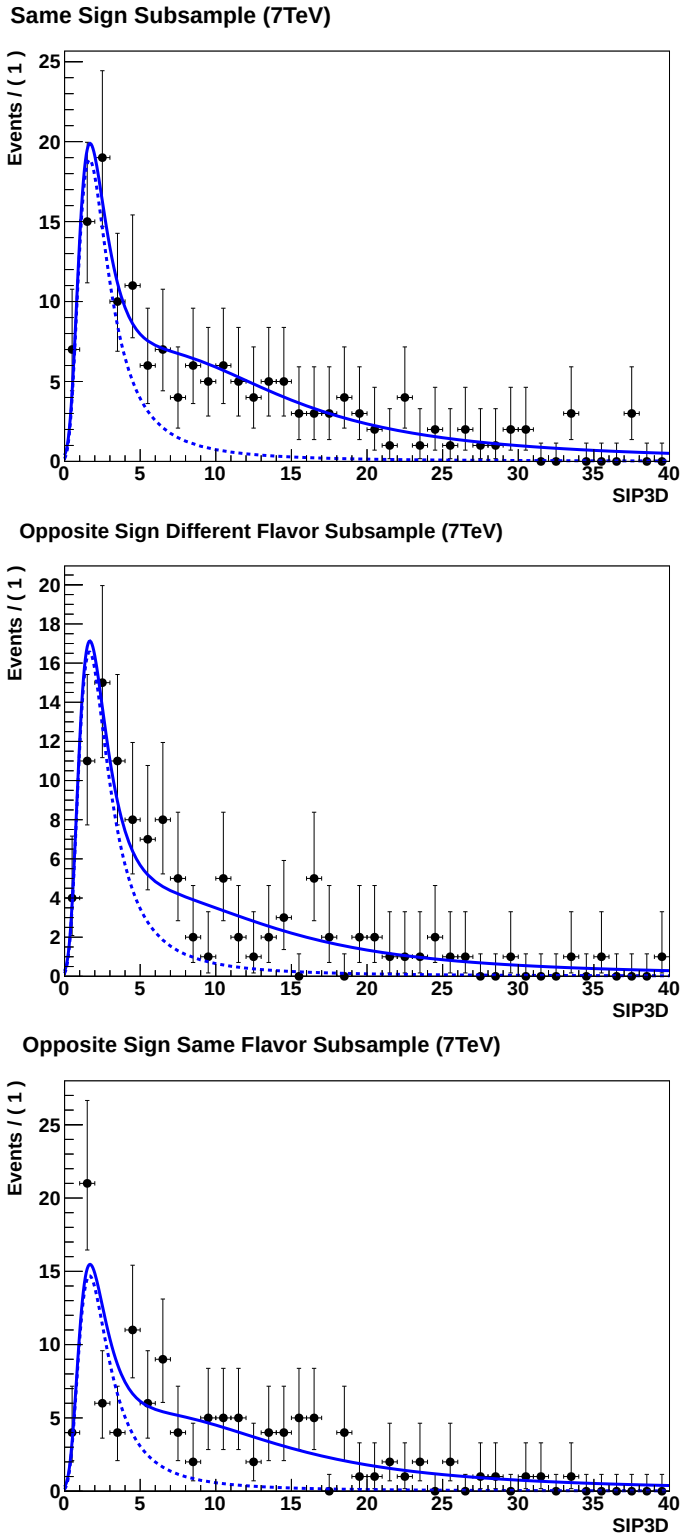


Figure 5.4: *SIP_{3D} distributions of Z + jets and $Zb\bar{b} + t\bar{t}$ for SS (top), OS-DF (center) and OS-SF (bottom) samples at 7 TeV applying $R_{Iso} > 0.4$: points represent data, the continuous line represents the best fit result and the dashed line is the Z + jets contribution.*

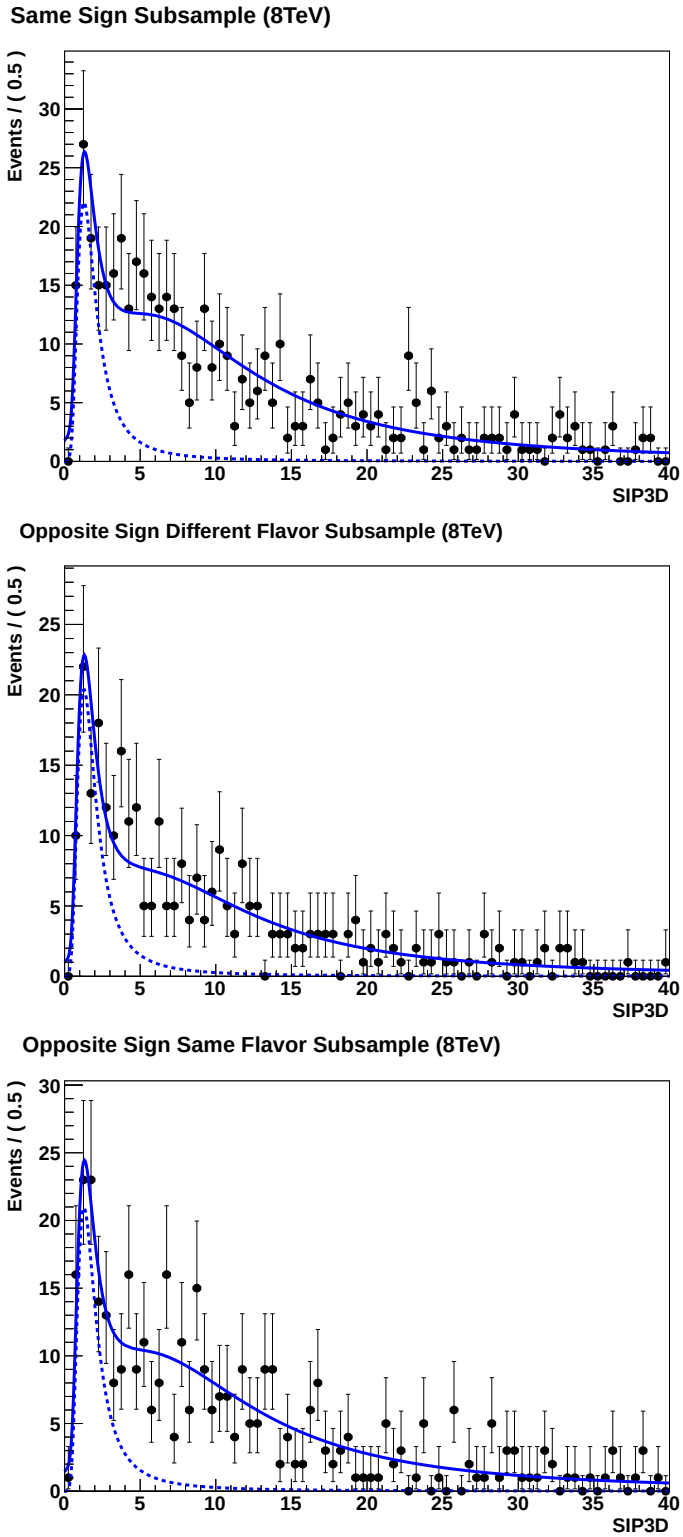


Figure 5.5: *SIP_{3D} distributions of Z + jets and $Zb\bar{b} + t\bar{t}$ for SS (top), OS-DF (center) and OS-SF (bottom) samples at 8 TeV applying $R_{Iso} > 0.4$: points represent data, the continuous line represents the best fit result and the dashed line is the Z + jets contribution.*

Table 5.2: *Fit parameters obtained from the 8 TeV sample, with and without the R_{Iso} cut: $N_{Zb\bar{b}+t\bar{t}}$ and N_{Z+jets} are the normalization constants obtained for SS, OS-DF and OS-SF subsets; $m_{1,2}$ and $\sigma_{1,2}$ are the mean and the sigma of the Landau function that fits the $Zb\bar{b} + t\bar{t}/Z + jets$ distribution.*

Parameter	$R_{Iso}(All)$	$R_{Iso} > 0.4$
$N_{Zb\bar{b}+t\bar{t}}(SS)$	375 ± 25	358 ± 26
$N_{Z+jets}(SS)$	98 ± 18	93 ± 21
$\chi^2_R(SS)$	4.55	4.35
$N_{Zb\bar{b}+t\bar{t}}(OS-DF)$	223 ± 19	204 ± 20
$N_{Z+jets}(OS-DF)$	92 ± 16	86 ± 17
$\chi^2_R(OS-DF)$	2.76	2.54
$N_{Zb\bar{b}+t\bar{t}}(OS-SF)$	301 ± 22	291 ± 23
$N_{Z+jets}(OS-SF)$	156 ± 19	88 ± 18
$\chi^2_R(OS-SF)$	3.49	3.51

Parameter	$R_{Iso}(All)$	$R_{Iso} > 0.4$
$\bar{m}_1$	6.86 ± 0.47	7.02 ± 0.57
σ_1	3.17 ± 0.19	3.23 ± 0.20
$\bar{m}_2$	1.30 ± 0.19	1.37 ± 0.10
σ_2	0.36 ± 0.19	0.38 ± 0.06

Comparing the two sets of fit parameters, it can be seen that the distributions obtained not applying and applying the R_{Iso} cut are compatible within the statistical error. It means that, as expected, $Z + jets$ and $Zb\bar{b} + t\bar{t}$ contributions are characterized by non-isolated leptons with large R_{Iso} values. The only appreciable difference is on the number of $N_{Z+jets}(OS - SF)$, that is considerably reduced after cutting on the isolation variable. The R_{Iso} cut in fact completely eliminates the contributions from ZZ and possibly from Higgs decays, which generate isolated leptons that could populate the lower R_{Iso} region.

5.3 Determination of the $Z + jets$ and $Zb\bar{b} + t\bar{t}$ Contributions in the $X + e$ and $X + \mu$ Samples

Because of the different resolution, electrons and muons have a different SIP_{3D} distribution. For the same impact parameter value, the significance for electrons is in fact smaller, due to the worse resolution at the denominator. This means that the SIP_{3D} distribution for electrons is shifted to lower values with respect to the muon one. In order to take into account this difference, it is thus necessary to separate the case in which the lepton with the worse SIP_{3D} is an electron from the case in which it is a muon. These two samples are defined by the same cuts of SS, OS-DF and OS-SF subsets (*first* Z step, $m_{4\ell} > 100\ GeV/c^2$ and $12 < m_{Z_2} < 60\ GeV/c^2$) and they are called $X + e$ and $X + \mu$ samples. Again, samples with and without the R_{Iso} cut are considered.
In order to determine the number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events of 7 and 8 TeV data samples, the $X + e$ and $X + \mu$ subsets are separately fitted using a Landau function for $Z + jets$ and one for $Zb\bar{b} + t\bar{t}$. The shape parameters of the two PDFs are fixed from the previous simultaneous fit, both for sample without and with R_{Iso} cut. The normalization values are not fixed and they are determined by the fit procedure. Due to the low number of events, unbinned fits are performed.
The results of fits are shown in Figures 5.6, 5.7, 5.8 and 5.9 for 7 and 8 TeV. The values both in the full R_{Iso} range and with $R_{Iso} > 0.4$ are reported. The number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ of $X + e$ and $X + \mu$ samples and distribution parameters obtained are reported in Tables 5.3 and 5.4, where only statistical errors are shown.

Table 5.3: *Number of $Zb\bar{b} + t\bar{t}$ and Z + jets events for the $X + e$ and $X + \mu$ samples at 7 TeV, for the whole range of R_{Iso} ($R_{Iso}(all)$) and for $R_{Iso} > 0.4$.*

$X + e$ (7 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	39 ± 8	41 ± 7
N_{Z+jets}	108 ± 11	92 ± 10

$X + \mu$ (7 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	198 ± 17	173 ± 15
N_{Z+jets}	77 ± 13	77 ± 12

Table 5.4: *Number of $Zb\bar{b} + t\bar{t}$ and Z + jets events for the $X + e$ and $X + \mu$ samples at 8 TeV, for the whole range of R_{Iso} ($R_{Iso}(all)$) and for $R_{Iso} > 0.4$.*

$X + e$ (8 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	198 ± 17	186 ± 16
N_{Z+jets}	144 ± 15	100 ± 13

$X + \mu$ (8 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	702 ± 29	678 ± 28
N_{Z+jets}	201 ± 18	156 ± 17

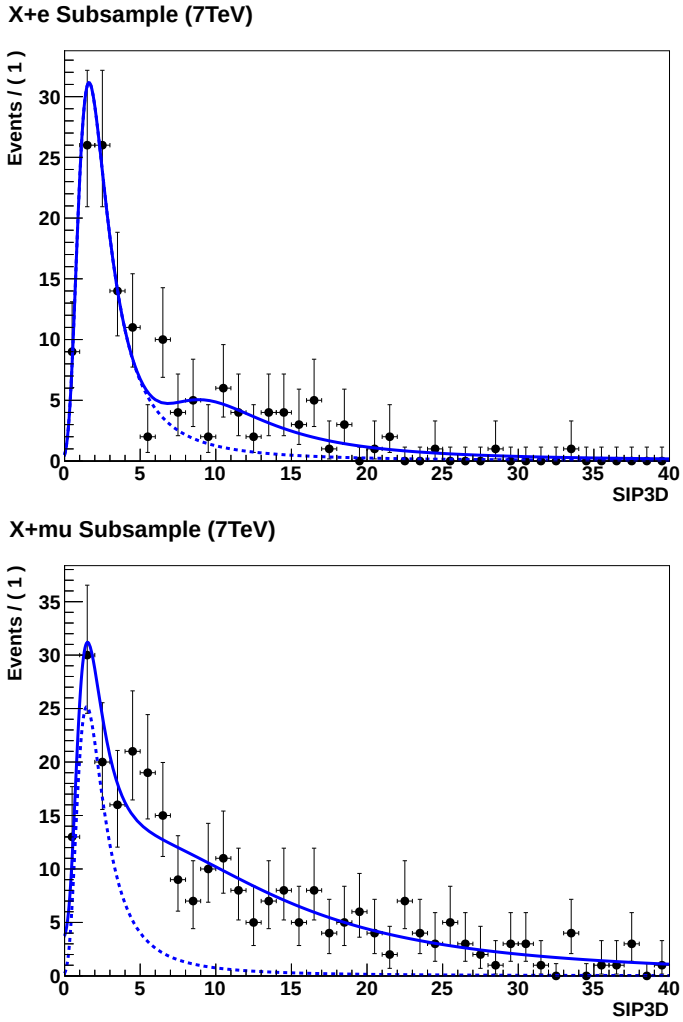


Figure 5.6: *SIP_{3D} distributions of $X + e$ (top) and $X + \mu$ (bottom) samples for 7 TeV: points represent data, the continuous line represents the best fit result and the dashed line is the $Z + jets$ contribution.*

5.4 Extraction of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ Contributions to the Signal Region

With the fit procedure described in the previous section, the number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events (for electrons or muons with the worse SIP_{3D} value) is found in the defined control sample. In order to extrapolate these contributions in the signal region, where backgrounds must be evaluated, it is necessary to consider the signal selection cuts (see Section 3.4) that consist in:

1. a pair of good lepton candidates of opposite charge and matching

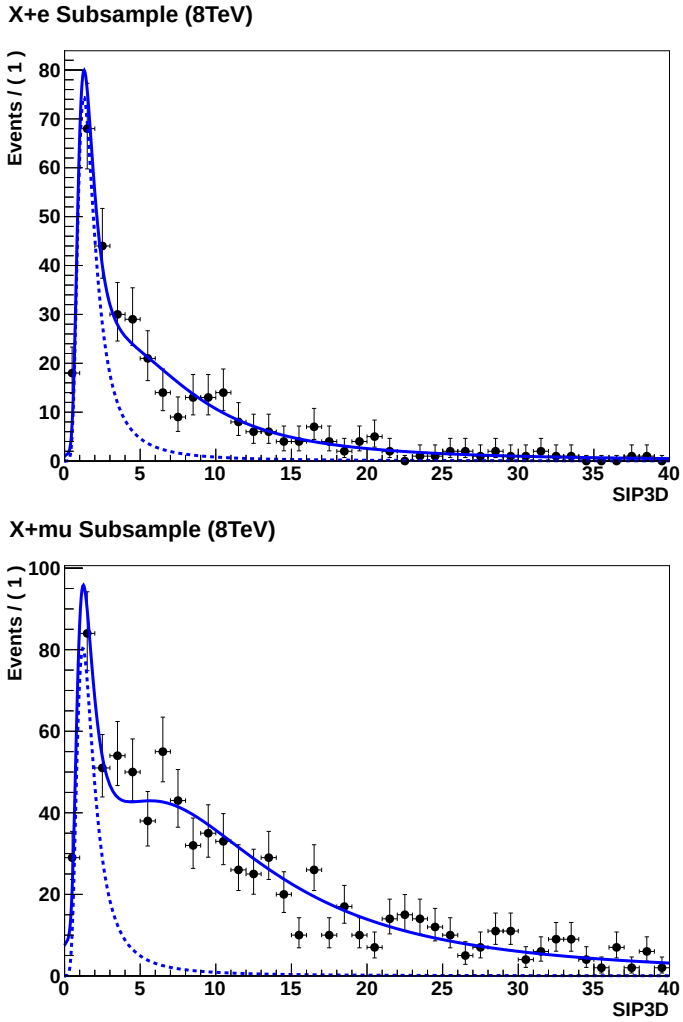


Figure 5.7: *SIP_{3D} distributions of $X + e$ (top) and $X + \mu$ (bottom) samples for 8 TeV: points represent data, the continuous line represents the best fit result and the dashed line is the Z + jets contribution.*

flavor (e^+e^-, $\mu^+\mu^-$) with reconstructed mass $m_{1,2}$ closest to the nominal Z boson mass (denoted Z_1), that satisfies $40 < m_{Z_1} < 120\ GeV/c^2$ (*first* Z step);

2. another good lepton pair with same flavor and opposite charge (denoted Z_2) and with $12 < m_{Z_2} < 120\ GeV/c^2$;
3. the reconstructed four-lepton invariant mass of $m_{4\ell} > 100\ GeV/c^2$;
4. $SIP_{3D} > 4$ for each lepton;
5. $R_{Iso} < 0.4$ for each lepton;

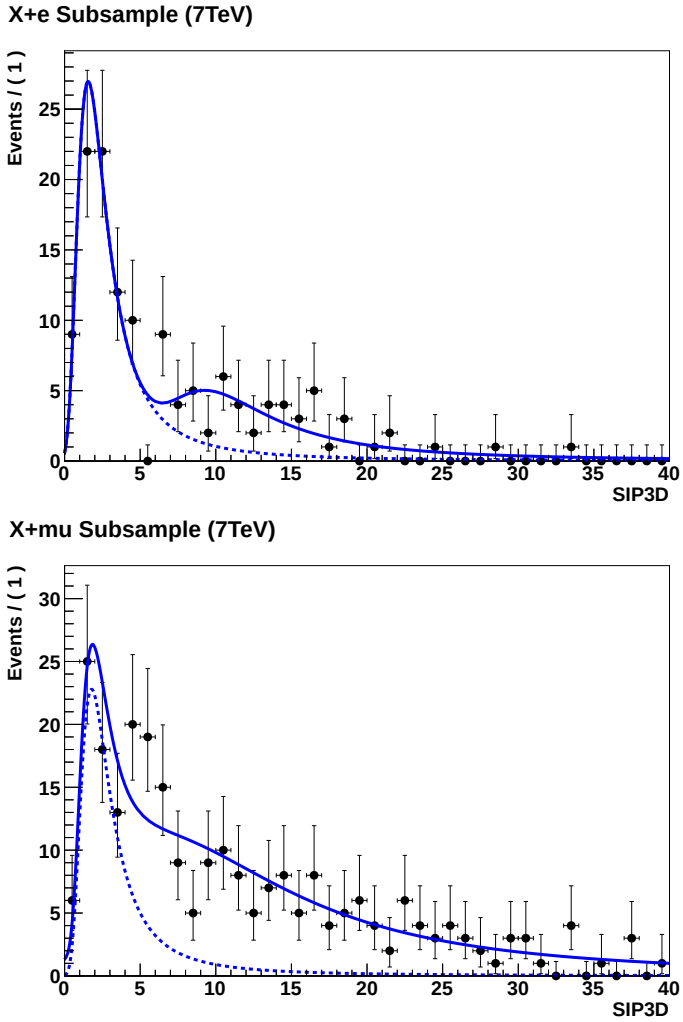


Figure 5.8: *SIP_{3D} distributions of $X + e$ (top) and $X + \mu$ (bottom) samples for 7 TeV with $R_{Iso} > 0.4$: points represent data, the continuous line represents the best fit result and the dashed line is the $Z +$ jets contribution.*

In the $X + e$ and $X + \mu$ samples the first and the third criterion are required. To reproduce the signal region, it is necessary also to apply the other selection criteria. As far as the SIP_{3D} cut is concerned, it is sufficient to integrate over the SIP_{3D} range (from 0 to 4) the PDFs obtained from the fit and multiply the result for the total number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events. This procedure is applied for $X + e$ and $X + \mu$ samples both in the whole R_{Iso} range ($R_{Iso}(all)$) and with the $R_{Iso} > 0.4$ condition, for $Z + jets$ and $Zb\bar{b} + t\bar{t}$ contributions and for 7 and 8 TeV separately.
The results are shown in Tables 5.5 and 5.6. In order to obtain the number of events with the $R_{Iso} < 0.4$ constraint, the result obtained for the $R_{Iso} >$

Table 5.5: *Number of $Zb\bar{b} + t\bar{t}$ and Z + jets events for the $X + e$ and $X + \mu$ samples at 7 TeV, integrating over the SIP_{3D} range (from 0 to 4) and considering the whole range of R_{Iso} ($R_{Iso}(all)$), $R_{Iso} > 0.4$ and $R_{Iso} < 0.4$.*

$X + e$ (7 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$	$R_{Iso} < 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	$(3.4 \pm 0.7)10^{-3}$	$(1.7 \pm 0.3)10^{-3}$	$(1.7 \pm 0.7)10^{-3}$
N_{Z+jets}	77 ± 8	66 ± 7	10 ± 11

$X + \mu$ (7 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$	$R_{Iso} < 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	27 ± 2	16 ± 1	11 ± 3
N_{Z+jets}	58 ± 9	54 ± 8	4 ± 13

Table 5.6: *Number of $Zb\bar{b} + t\bar{t}$ and Z + jets events for the $X + e$ and $X + \mu$ samples at 8 TeV, integrating over the SIP_{3D} range (from 0 to 4) and considering the whole range of R_{Iso} ($R_{Iso}(all)$), $R_{Iso} > 0.4$ and $R_{Iso} < 0.4$.*

$X + e$ (8 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$	$R_{Iso} < 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	37 ± 3	33 ± 3	3 ± 4
N_{Z+jets}	123 ± 13	83 ± 11	39 ± 17

$X + \mu$ (8 TeV)	$R_{Iso}(all)$	$R_{Iso} > 0.4$	$R_{Iso} < 0.4$
$N_{Zb\bar{b}+t\bar{t}}$	88 ± 4	84 ± 4	4 ± 5
N_{Z+jets}	172 ± 16	133 ± 14	39 ± 21

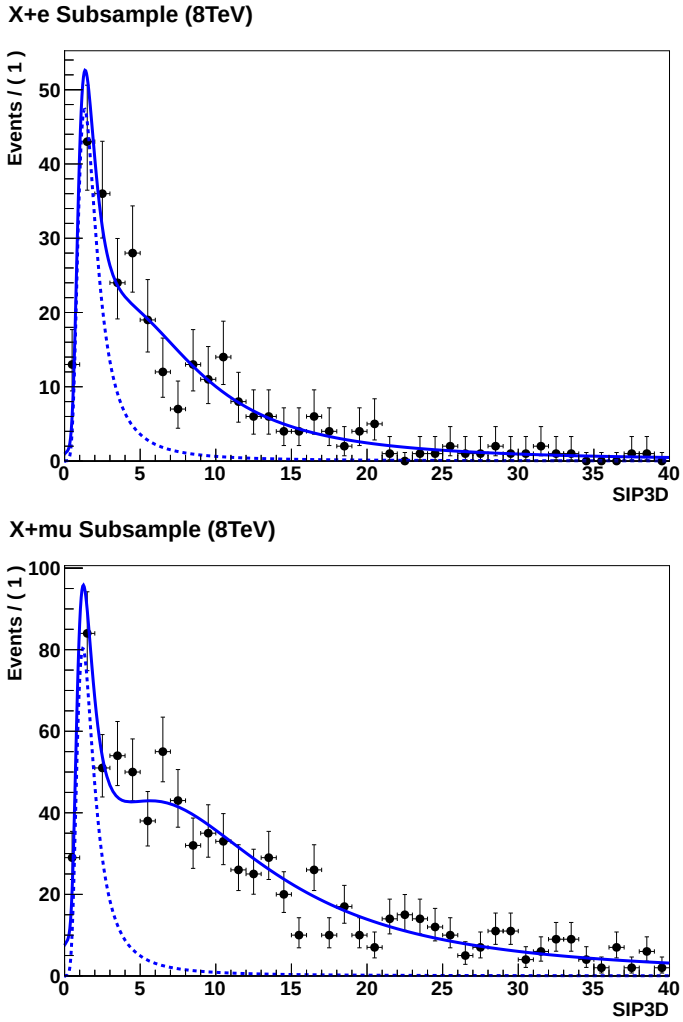


Figure 5.9: *SIP_{3D} distributions of $X + e$ (top) and $X + \mu$ (bottom) samples for 8 TeV with $R_{Iso} > 0.4$: points represent data, the continuous line represents the best fit result and the dashed line is the Z + jets contribution.*

0.4 sample is subtracted from the $R_{Iso}(all)$ contribution.
As far as m_{Z_2} observable is concerned, in the $X + e$ and $X + \mu$ control samples it is required that $12 < m_{Z_2} < 60\ GeV/c^2$, that is a smaller range with respect to the corresponding Higgs selection cut. In order to take into account the events that are lost because of this more restrictive cut, it is thus necessary multiply the number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events for a correction factor f_{Z_2}. The f_{Z_2} factor is obtained from MC simulation and it is defined as

$$f_{Z_2} = \frac{N(m_{Z_2} < 120 | R_{Iso}, SIP_{3D})}{N(m_{Z_2} < 60 | R_{Iso}, SIP_{3D})},$$

where the numerator (denominator) is the number of $Z + jets$ or $Zb\bar{b} + t\bar{t}$ events with $R_{Iso} < 0.4$, $SIP_{3D} < 4$ and $12 < m_{Z_2} < 120\ GeV/c^2$ ($12 < m_{Z_2} < 60\ GeV/c^2$). Because of the low MC samples number of events after R_{Iso} and SIP_{3D} cuts, the correction factor is calculated as

$$f_{Z_2} \sim \frac{N(m_{Z_2} < 120)}{N(m_{Z_2} < 60)},$$

without applying cuts. It is thus assumed that the m_{Z_2} distributions ratio does not depend on R_{Iso} and SIP_{3D} cuts, obtaining

$$f_{Z_2}^{Zb\bar{b}+t\bar{t}}(7TeV) = 1.099 \pm 0.052$$

$$f_{Z_2}^{Z+jets}(7TeV) = 1.22 \pm 0.23$$

$$f_{Z_2}^{Zb\bar{b}+t\bar{t}}(8TeV) = 1.077 \pm 0.043$$

$$f_{Z_2}^{Z+jets}(8TeV) = 1.28 \pm 0.25.$$

Since f_{Z_2} is the inverse of an efficiency ($f_{Z_2} = 1/\epsilon$), its error is

$$\delta f_{Z_2} = \frac{\delta\epsilon}{\epsilon^2},$$

where $\delta\epsilon$ is defined by Equation 4.9.
Finally, it is observed that the $X + \ell$ sample (with $\ell = e, \mu$) contains several contributions that are not in the signal region. The *loose lepton pair* in fact can be formed by $\ell^{\mp}\ell^{\pm}$, $\ell^{\pm}\ell^{\pm}$, $\ell'^{\pm}\ell^{\pm}$ and $\ell'^{\mp}\ell^{\pm}$, while in the signal region only the first combination contributes. For this reason, the obtained results are multiplied for an additional correction factor $f_{\ell^+\ell^-} = 0.25$.
The number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events in the signal region N_{sig} is thus determined by

$$N_{sig} = f_{\ell^+\ell^-} f_{Z_2} N_{CR}, \tag{5.1}$$

where N_{CR} are the values obtained with $R_{Iso} < 0.4$ reported in Tables 5.5 and 5.6.
Putting together all these information, the number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events expected in the signal region is obtained, separating the electron and the muon case, for 7 and 8 TeV. Results are reported in Table 5.7. Errors are calculated with

$$\delta N_{sig} = f_{\ell^+\ell^-} \sqrt{(f_{Z_2}\delta N_{CR})^2 + (\delta f_{Z_2} N_{CR})^2}. \tag{5.2}$$

Some of the contributions are compatible with zero, given their uncertainties.

Summing the $Z + jets$ and $Zb\bar{b} + t\bar{t}$ contributions, it is found that

Table 5.7: *Number of $Zb\bar{b} + t\bar{t}$ and $Z + jets$ events extrapolated in the signal region at 7 and 8 TeV.*

$X + e$	7 TeV	8 TeV
$N_{Zb\bar{b}+t\bar{t}}$	$(4.7 \pm 2.0)10^{-4}$	1 ± 1
N_{Z+jets}	3 ± 3	13 ± 6

$X + \mu$	7 TeV	8 TeV
$N_{Zb\bar{b}+t\bar{t}}$	3 ± 1	1 ± 1
N_{Z+jets}	1 ± 4	12 ± 7

$$N_{Z+X}^{7TeV} = 7 \pm 5$$

$$N_{Z+X}^{8TeV} = 27 \pm 9.$$

Comparing the 7 *TeV* values with those obtained with the method used in the $H \to 4\ell$ analysis (see Table 3.1), it can be observed that the results are compatible.

5.5 Evaluation of Systematic Uncertainties

As described in the previous section, the number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events extracted in the signal region is obtained from Equation 5.1. Therefore the main contributions to the systematic uncertainties come from the f_{Z_2} and N_{CR} factors, while $f_{\ell^+\ell^-}$ is taken without any error.
The f_{Z_2} systematic uncertainty is due mainly to the assumption that the m_{Z_2} distributions ratio does not depend on R_{Iso} and SIP_{3D} cuts. The contributions of other scale factors are canceled by the ratio. In order to estimate this systematic uncertainty, the f_{Z_2} factors is calculated with inverted cuts, that are $R_{Iso} > 0.4$ and $SIP_{3D} > 4$. These values are indicated with $f_{Z_2}^s$, where s stays for "systematic", while the "nominal" values, i.e. values reported in the previous section, are indicated with $f_{Z_2}^n$. The systematic error is calculated according to

$$\delta f_{Z_2}^{syst} = f_{Z_2}^n - f_{Z_2}^s. \tag{5.3}$$

Finally, for the f_{Z_2} factors it is obtained

$$f_{Z_2}^{Zb\bar{b}+t\bar{t}}(7TeV) = 1.099 \pm 0.052 \pm 0.025$$

$$f_{Z_2}^{Z+jets}(7TeV) = 1.22 \pm 0.23 \pm 0.04$$

$$f_{Z_2}^{Zb\bar{b}+t\bar{t}}(8TeV) = 1.077 \pm 0.043 \pm 0.019$$

$$f_{Z_2}^{Z+jets}(8TeV) = 1.28 \pm 0.25 \pm 0.05,$$

Table 5.8: *Number of $Zb\bar{b}+t\bar{t}$ and $Z+jets$ events for the $X+e$ and $X+\mu$ samples at 7 TeV, integrating over the SIP_{3D} range (from 0 to 4) and considering the whole range of R_{Iso} ($R_{Iso}(all)$), $R_{Iso}>0.4$ and $R_{Iso}<0.4$ (fit parameters not fixed).*

$X+e$ (7 TeV)	$R_{Iso}(all)$	$R_{Iso}>0.4$	$R_{Iso}<0.4$
$N_{Zb\bar{b}+t\bar{t}}$	$(3.39\pm0.96)10^{-3}$	$(1.67\pm0.37)10^{-3}$	$(1.72\pm1.03)10^{-3}$
N_{Z+jets}	77 ± 10	66 ± 8	10 ± 13

$X+\mu$ (7 TeV)	$R_{Iso}(all)$	$R_{Iso}>0.4$	$R_{Iso}<0.4$
$N_{Zb\bar{b}+t\bar{t}}$	27 ± 5	16 ± 4	11 ± 6
N_{Z+jets}	58 ± 29	54 ± 27	4 ± 39

Table 5.9: *Number of $Zb\bar{b}+t\bar{t}$ and $Z+jets$ events for the $X+e$ and $X+\mu$ samples at 8 TeV, integrating over the SIP_{3D} range (from 0 to 4) and considering the whole range of R_{Iso} ($R_{Iso}(all)$), $R_{Iso}>0.4$ and $R_{Iso}<0.4$ (fit parameters not fixed).*

$X+e$ (8 TeV)	$R_{Iso}(all)$	$R_{Iso}>0.4$	$R_{Iso}<0.4$
$N_{Zb\bar{b}+t\bar{t}}$	37 ± 6	33 ± 6	3 ± 7
N_{Z+jets}	123 ± 27	83 ± 27	39 ± 38

$X+\mu$ (8 TeV)	$R_{Iso}(all)$	$R_{Iso}>0.4$	$R_{Iso}<0.4$
$N_{Zb\bar{b}+t\bar{t}}$	88 ± 5	84 ± 5	4 ± 7
N_{Z+jets}	172 ± 27	133 ± 26	39 ± 38

where the last contribution is the systematic uncertainty.
The systematic uncertainty of the fit method used to extract N_{CR} is determined by comparing the obtained value with the one obtained with the same procedure, but without keeping fixed the parameters of the fit on the $X+e$ and $X+\mu$ samples. Again, the integration over the SIP_{3D} range (from 0 to 4) of the PDFs obtained from the fit is calculated and the result is multiplied for the total number of $Z+jets$ and $Zb\bar{b}+t\bar{t}$ events. This procedure is applied for $X+e$ and $X+\mu$ samples both in the whole R_{Iso} range ($R_{Iso}(all)$) and with the $R_{Iso}>0.4$ condition, for $Z+jets$ and $Zb\bar{b}+t\bar{t}$ contributions and for 7 and 8 TeV separately. The results are shown in Tables 5.8 and 5.9. It can be seen that the normalization constants do not change with respect to the case with fixed parameters, but errors are significantly larger. In order to obtain the number of events with the $R_{Iso}<0.4$ cut and its uncertainty, the result obtained for the $R_{Iso}>0.4$ sample is subtracted from the $R_{Iso}(all)$ contribution and the errors are propagated.

Table 5.10: *Number of $Zb\bar{b}+t\bar{t}$ and $Z+jets$ events extrapolated in the signal region at 7 and 8 TeV, considering systematic errors.*

$X+e$	7 *TeV*	8 *TeV*
$N_{Zb\bar{b}+t\bar{t}}$	$(4.7 \pm 2.0 \pm 2.8)10^{-4}$	$1 \pm 1 \pm 2$
N_{Z+jets}	$3 \pm 3 \pm 4$	$13 \pm 6 \pm 13$

$X+\mu$	7 *TeV*	8 *TeV*
$N_{Zb\bar{b}+t\bar{t}}$	$3 \pm 1 \pm 2$	$1 \pm 1 \pm 2$
N_{Z+jets}	$1 \pm 4 \pm 12$	$12 \pm 7 \pm 12$

5.6 Final Results

In order to determine the systematic uncertainty on the number of events in the signal region, the Equation 5.2 is used. In this case δf_{Z_2} and δN_{CR} are the systematic uncertainties obtained in the previous section, while f_{Z_2} and N_{CR} are the "nominal" values. Results are reported in Table 5.10. Summing the $Z+jets$ and $Zb\bar{b}+t\bar{t}$ contribution, the final results are

$$N_{Z+X}^{7TeV} = 7 \pm 5 \pm 13$$

$$N_{Z+X}^{8TeV} = 27 \pm 9 \pm 18,$$

where the first term is the statistical uncertainty and the second one the systematic uncertainty.

Conclusions

The work presented in this thesis was carried on in the context of the search for the Higgs boson in the decay channel $H \to ZZ^{(\star)} \to 4\ell$, using proton-proton collision data recorded by the CMS detector at the LHC. This channel was expected to be one of the main decay channels for the Higgs boson discovery, since the signature of the four final state leptons is very clear. 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 and the contribution of the $H \to ZZ \to 4\ell$ analysis was significant.

In this thesis, I presented a study of the reducible and instrumental backgrounds, that the search for a Higgs signal in the $H \to ZZ^{(\star)} \to 4\ell$ channel has to cope with.
I first focused my attention on the $Zb\bar{b}$ and $t\bar{t}$ backgrounds, defining the control region used to evaluate these contributions. After applying the control region selection cuts, I fitted the invariant mass distribution of the on-mass shell Z boson m_{Z_1}, in which the Z mass peak of $Zb\bar{b}$ events and the non-resonant $t\bar{t}$ background are easily disentangled. I verified that Monte Carlo and data results are compatible within the errors.
Using the fit results, I measured for the first time the $Zb\bar{b} \to 4\ell$ and $t\bar{t} \to 4\ell$ cross sections (with $\ell = e, \mu$), obtaining

$$\sigma^{7TeV}_{Zb\bar{b}\to 4\ell} = 63.1 \pm 5.7 \pm 4.2 \pm 1.4 \ pb$$

$$\sigma^{8TeV}_{Zb\bar{b}\to 4\ell} = 93.2 \pm 5.5 \pm 8.0 \pm 4.1 \ pb.$$

for the $Zb\bar{b} \to 4\ell$ process and

$$\sigma^{7TeV}_{t\bar{t}\to 4\ell} = 1.452 \pm 0.184 \pm 0.017 \pm 0.032 \ pb$$

$$\sigma^{8TeV}_{t\bar{t}\to 4\ell} = 2.277 \pm 0.215 \pm 0.035 \pm 0.100 \ pb.$$

for the $t\bar{t} \to 4\ell$ process. In this last case there is a good agreement with theoretical predictions. Statistical and systematical uncertainties are reported, while theoretical uncertainties have not been yet calculated.

In Chapter 5 I presented a new data-driven procedure to separately measure instrumental and reducible backgrounds that remain after the first steps of the $H \to ZZ \to 4\ell$ selection. I defined a control sample to study the two different contributions and I fitted the distributions of the lepton impact parameter significance SIP_{3D} , the observable that better distinguishes the $Z + jets$ and $Zb\bar{b} + t\bar{t}$ backgrounds. I finally extrapolated the number of $Z + jets$ and $Zb\bar{b} + t\bar{t}$ events from the control sample to the signal region of the $H \to 4\ell$ analysis.
Summing the $Z + jets$ and $Zb\bar{b} + t\bar{t}$ contribution, I obtained that

$$N_{Z+X}^{7TeV} = 7 \pm 5 \pm 13$$

$$N_{Z+X}^{8TeV} = 27 \pm 9 \pm 18,$$

where the first term is the statistic uncertainty and the second one the systematic uncertainty. Comparing the 7 TeV values with those obtained with the method presently used in the $H \to 4\ell$ analysis in CMS (see Table 3.1), I found that the results are compatible.

Bibliography

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

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

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

[4] 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), arXiv:1101.0593 [hep-ph].

[5] 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), arXiv:1201.3084 [hep-ph].

[6] LHC Higgs Cross Section Working Group. http//twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[7] *LHC Design Report*, CERN 2004-003, 2004.

[8] R. Adolphi et al., *The CMS experiment at the CERN LHC*, JINST, vol. 3, p. S08004, 2008.

[9] 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.

[10] The CMS Collaboration, *Search for the standard model Higgs boson in the decay channel $H \to ZZ \to 4\ell$ in pp collisions*, CMS NOTE AN-12-141.

[11] The CMS Collaboration, *Evidence for a new state 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*, CMS PAS HIG-12-016.

[12] The Geant4 Collaboration, *GEANT Detector Description and Sim- ulation Tool*, http://geant4.web.cern.ch/geant4/.

[13] S. Frixione, P. Nason, and C. Oleari,*Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP, vol. 11, p. 070, 2007.

[14] T. Binoth, N. Kauer, and P. Mertsch, *Gluon-induced QCD corrections to* $pp \to ZZ \to \ell\ell\ell'\ell'$, 2008.

[15] M. Cacciari, S. Frixione, M. Mangano et al., *Updated predictions for the total production crosssections of top and of heavier quark pairs at the Tevatron and at the LHC*, `arXiv:0804.2800v3`.

[16] G. Cowan, *Statistical Data Analysis*, Clarendon Press, Oxford, 1998.

[17] J. Beringer et al. (Particle Data Group), *The Review of Particle Physics*, Phys. Rev. D86, 010001 (2012).

[18] *RooFit Users Manual*, v2.91, Root, CERN.

[19] Hathor, Comp.Phys.Commun. 182(2011) 1034.

Acknowledgements

Semplicemente grazie a...

Chiara, perché ha sempre il sorriso e l'ottimismo è contagioso;
Nicola e Mario, per il fondamentale aiuto e i numerosi insegnamenti;
Nicolò per le sue spiegazioni illuminanti;
Marty e Gianlu, perché con loro tutto ha avuto inizio;
Giulia, Vio ed Ema, perché vivere nella Dilip House senza di loro non sarebbe stata la stessa cosa;
Max, Luca ed Alberto, perché non abitavano nella Dilip House ma facevano comunque parte della famiglia;
Stefano, perché è stato il primo a volere una copia;
Dilip, perché non tutte le domande hanno una risposta e, se anche ce l'avessero, a volte sarebbe meglio non conoscerla;
i compagni dell'open space, perché un "hi ladies" di prima mattina fa sempre piacere;
Miky, perché le giornate non sarebbero iniziate con il piede giusto senza un suo saluto;
Claudia, perché quando ci sentivamo mi capiva senza che dicessi nulla;
Ale, perché il 4 luglio è il suo compleanno;
Monia, perché riusciva ad essere abbracciosa pur essendo fisicamente lontana;
Elisa, per aver condiviso questo mio ultimo meraviglioso anno in via Berthollet;
gli amici valdo-torinesi, perché non ce la facevano più con tutte quelle feste di addio e non vedevano l'ora che partissi;
la sezione percussioni (anche chi è da un po' che non ne fa parte), perché l'antimojito è il mio cocktail preferito;
di nuovo (l'avevo promesso) Marty e Giulia, perché con pazienza mi hanno insegnato che non è mai troppo tardi;
le nonne, perché nonostante quello che dicono sono ancora in gran forma;
Mamma, Papà e Chiara, per tutto (ed è davvero tantissimo).

Infine un grazie particolare al bosone di Higgs (??), che ha gentilmente deciso di mostrarsi in pubblico appena due settimane dopo il mio arrivo

al CERN, permettondomi così di vivere un'esperienza indimenticabile. Come tutte le persone appena elencate sopra, del resto. Grazie davvero.