



Università degli Studi di Torino

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE NATURALI
Corso di Laura Magistrale in Fisica

TESI DI LAUREA MAGISTRALE

**Study of the signal lineshape in the analysis for the search of
 $H \rightarrow ZZ(*) \rightarrow 4\ell$ at CMS**

Candidato:

Pinna Angioni Gian Luca

Matricola 731768

Relatore:

Nicola Amapane

Correlatori:

Mario Pelliccioni

Chiara Mariotti

Contents

Introduction	3
Istèrrida	5
1 Standard Model Higgs Boson	7
1.1 The "unbroken" Standard Model	7
1.2 The Higgs mechanism	9
1.2.1 Fermion Masses and Couplings	12
1.3 Higgs boson mass	14
1.3.1 Theoretical and experimental constraints	14
1.3.2 Recent LHC result	16
1.4 Higgs phenomenology at LHC	17
1.4.1 Higgs production	17
1.4.2 Higgs decay	20
1.5 Higgs search at LHC	21
1.5.1 Low mass region: $110 \leq M_H \leq 135$ GeV	23
1.5.2 Intermediate mass region: $140 \leq M_H \leq 180$ GeV	23
1.5.3 High mass region: $M > 180$ GeV	24
2 The CMS Detector at the LHC	25
2.1 The Large Hadron Collider	25
2.2 High energy collision kinematic variables	28
2.3 The Compact Muon Solenoid experiment	31
2.3.1 The Magnet	31
2.3.2 Silicon Tracking System	32
2.3.3 The Electromagnetic Calorimeter	33
2.3.4 The Hadron Calorimeter	34
2.3.5 The muon system	35

<i>CONTENTS</i>	2
2.3.6 The CMS Trigger System	37
3 Search for the SM Higgs in the $ZZ(*) \rightarrow 4\ell$ decay channel with CMS	39
3.1 The 4ℓ Final State	39
3.2 Data sample	40
3.3 Simulated Samples	41
3.4 Physics Objects	42
3.4.1 Electrons	43
3.4.2 Muons	43
3.4.3 Lepton Impact Parameter and Isolation	44
3.4.4 Photons	45
3.5 Event Selection	47
3.6 Kinematic Discriminant (MELA)	52
3.6.1 Construction of the MELA discriminant	54
3.7 Statistical Analysis	58
3.8 Results	58
4 Monte Carlo reweighting	67
4.1 Pileup	67
4.2 Efficiency corrections	69
4.3 High-mass line shape	72
5 Signal Modelization	81
5.1 Low and intermediate mass range	81
5.2 High mass range	87
5.3 Signal model uncertainties	94
5.3.1 Efficiency scale factors uncertainties	94
5.3.2 Evaluation of the high-mass corrections systematic uncertainties	96
5.4 Overview of the statistical method	97
Conclusions	99

Introduction

The Standard Model of elementary particles is a relativistic quantum field theory that describes the particles observed in nature and their interactions, except for gravitation. The predictions of this theory coincide with observations in experiments with an incredible precision, making it the most stringently tested existing scientific theory. The only missing piece in the spectrum of the theory is the Higgs boson, the quantum of the field believed to be responsible for the spontaneous breaking of the SM gauge symmetry that also represents the key to the origin of the mass of fundamental particles.

The Compact Muon Solenoid (CMS) is one of the four main experiments at the Large hadron Collider (LHC) and the Higgs search is one of the more important parts of its physics program. In July 2012 the CMS and ATLAS experiment announced the discovery of a new boson at a mass around 125 GeV, with properties compatible with the SM Higgs boson. In this thesis, I present my work in the CMS experiment for the search of the standard model Higgs boson decaying into four leptons. In particular, I have focused my work on the study of the signal lineshape.

My thesis is structured as follows.

Chapter 1 introduces the theoretical framework of the Standard Model together with the Higgs boson properties. It further offers an overview of the constraints on the Higgs boson mass before the LHC era and nowadays, on an experimental and on a theoretical point of view.

In Chapter 2 the LHC accelerator is introduced and the experimental setup of the CMS detector and the low-level reconstruction is summarized.

Chapter 3 concerns the analysis of the Higgs-boson identification in the $H \rightarrow ZZ \rightarrow 4\ell$ channel. The 4ℓ final state signal and background processes are described in detail, for a range of mass from 110 to 1000 GeV. The result of the observation of a new particle in this channel with a mass near 125

GeV is presented.

The following two chapters describe more specifically the work I've done.

In order to estimate the significance of an excess or quantify exclusion, a statistical analysis of the observed data is necessary. For this purpose, a model of the signal lineshape, parametric on the hypothetical Higgs boson mass, is necessary.

The signal modelization is extracted from the Monte Carlo samples, which do not describe immediately the experimental data with the required precision for this analysis. Therefore, it is necessary to correct the Monte Carlo from discrepancy and possible systematic errors. I have been working on the correction of the simulated samples through the use of event weight and per-event correction factors. This is subject of chapter 4. The first part of the chapter is focused on the correction of the Pile-up distribution and of the trigger, reconstruction and identification efficiencies. The second part of the chapter concerns the theoretical lineshape description in the high mass region.

Once these corrections were applied, I could proceed with the study of the modelization of the signal shape. I worked on two different parametrizations, one for the low mass region and the other one for the high mass region. In the latter I have been working on the signal model systematic uncertainty estimation. This is described in chapter 5.

Istèrrida

Su Modellu Istandard (MI) de is partitzeddas elementares est una teoria cuàntica de is campos chi descriet is partitzeddas chi podent èssere osservadas in sa naturalesa e is interatziones issoro, mancu sa gravitatzione. Is previsiones de custa teoria cointzident cun is osservatziones fatas in is esperimentos cun una pretzisione incredibile faghendedda sa prus isperimentada de is teoria iscientificas. S'ùnica parte chi mancat in s'ispetru de sa teoria est su bosone de Higgs, su quantu de su campu chi est cunsideradu èssere responsàbile de sa rotura de simmetria de su *gauge* MI e fintzas sa crae de s'origine de sa massa de is partitzeddas fundamentales.

Su Compact Muon Solenoid (Solenòide muònicu cumpatu) (CMS) est unu de is bator esperimentos printzipales in su Large Hadron Collider (LHC) e sa chirca de s'Higgs est una de is partes prus importantes de su programma de fisica suo. In su mese de trìulas de su 2012 is esperimentos CMS e ATLAS ant annuntziadu s'iscoberta de unu bosone nou cun una massa acanta a 125 GeV, cun propiedades compatibiles cun su bosone de Higgs descritu in su Modellu Istandard. In custa tesi, presento su traballu chi apo fatu in s'esperimentu CMS pro sa chirca de su bosone de Higgs decadende in bator leptones. Mescamente, su traballu miu est cuntzentradu in s'istùdiu de sa forma de su segnale.

Sa tesi cosa mia est istrutturada in custa manera. Su capìtulu 1 introduet sa istruttura teòrica de su Modellu Istandard impare cun is propiedades de su bosone de Higgs. Pustis si donat una bisura de is custringhidura a pitzus de sa massa de su bosone Higgs antis de s'era de su LHC e oe, dae su puntu de bisura teòricu e isperimentale. In su capìtulu 2 introduimus s'atzelleradore LHC, bidimus sa organizazioni isperimentale de su iscobiadore CMS e resumimus sa ricostruzione de livellu bàsciu.

Su capìtulu 3 est a pitzus de s'anàlisi de s'identificatzione de su bosone Higgs in su canale $H \rightarrow ZZ \rightarrow 4\ell$. Su segnale de istadu finale de bator e is

protzessos in background ant a èssere descritos in detàlli pro unu tretu de massa dae 110 fintzas a 1000 GeV. Sunt presentados is resurtados de una osservatzione de una partitzedda noa in custu canale cun una massa acanta de 125 GeV. In is duos capìtulos chi s'ghint est descritu mègius su traballu chi apo fatu.

Pro istimare sa significatividade de unu etzessu o cantificare una esclusione est pretzisa un'anàlisi istatìstica de is datos osservados. Pro fàghere custu pretzisamus de unu modellu de sa forma de su segnale, paramètricu cun sa massa ipotètica de su bosone Higgs. Sa modellizatzione de su segnale est istada pigada dae is campiones de Monte Carlo, mancarì non descriciant is datos isperimentales con sa pretzisione chi netzessitamus. Duncas, bisòngiat currègere is discrepàntzias e is possìbiles errores sistemàticos. Deo apo traballadu in sa curretzione de is campiones simulados cun s'impreu de su pesu de s'eventu e cun fatores de curretzione pro cada eventu. Su capìtulu 4 est a pitzus de custu. Sa primu parte de su capìtulu est cuntzentrada in sa curretzione de sa distributzione de su appiramentu (*pileup*) e de su grillette (*trigger*), de s'efitzèntzia de sa ricostrutzione e de s'identificatzione. Sa parte de duos de su capìtulu est a pitzus de sa descritzione teòrica de sa forma de su segnale in sa regione de massa arta.

Aplicadas custas curretziones apo podidu s'ghire cun s'istùdiu de sa modellizatzione de sa forma de su segnale. Apo traballadu duas diferentes parametrizatziones, una pro sa regione de massa bàscia e un'àtera pro sa regione de massa arta. Pustis apo traballadu in s'istima de s'intzertesa sistemàtica de su modellu de segnale. Custu at a èssere descritu in su capìtulu 5.

Chapter 1

Standard Model Higgs Boson

The Standard Model (SM) of particle physics is a unified framework to describe electromagnetic and weak interactions between quarks and leptons together with strong interactions between quarks. It is composed of the Yang-Mills theory based on the electroweak symmetry group $SU(2)_L \times U(1)_Y$ of Glashow, Weinberg, and Salam [1, 2, 3] and strong $SU(3)_C$ group of QCD [4, 5, 6].

In this chapter, a short overview of the SM and the Electroweak Theory is given, focusing upon the ElectroWeak Symmetry Breaking (EWSB) and the Higgs boson properties.

1.1 The "unbroken" Standard Model

This model has only 2 components, the fermions and gauge fields. The spin 1/2 fermions fields are organized in three generations of fermions, left-handed and right-handed quarks and leptons, $f_{L,R} = \frac{1 \mp \gamma_5}{2} f$. Where f is the fermions field. It is crucial for this consideration that the left-handed fermions are in the weak $SU(2)_L$ isodoublets, while the right-handed fermions are weak isosinglets. Moreover, both left- and right-handed quarks are triplets under the $SU(3)_C$ group, while all leptons are color singlets.

The gauge fields mediate the above-mentioned interactions via (spin-1) bosons. In the electroweak sector, the field B_μ corresponds to the $U(1)_Y$ group and the three fields $W_\mu^{1,2,3}$ correspond to the $SU(2)_L$ group. There

is also an octet of gluon fields G_μ^a which correspond to the color $SU(3)_C$ group. Due to the non-Abelian nature of the $SU(2)$ and $SU(3)$ groups, there are triple and quartic self-interactions between their gauge fields. The fermions fields ψ are minimally coupled to the gauge fields through the covariant derivative D_μ which leads to a unique form of interaction between the fermion and gauge fields, $(-g_i \bar{\psi} V_\mu \gamma^\mu \psi)$, with $g_i = g_s, g_2, g_1$ that are, respectively, the coupling constants of the $SU(3)_C, SU(2)_L,$ and $U(1)_Y$ groups.

The SM Lagrangian, without mass terms for fermions and gauge bosons, is then given by

$$\begin{aligned} \mathcal{L}_{\text{SM}} = & -\frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{L}_i i D_\mu \gamma^\mu L_i + \bar{e}_{Ri} i D_\mu \\ & \gamma^\mu e_{Ri} + \bar{Q}_i i D_\mu \gamma^\mu Q_i + \bar{u}_{Ri} i D_\mu \gamma^\mu u_{Ri} + \bar{d}_{Ri} i D_\mu \gamma^\mu d_{Ri} \end{aligned} \quad (1.1)$$

where

$$D_\mu = \partial_\mu - ig_2 \frac{\tau_a}{2} W_\mu^a - ig_1 \frac{Y_H}{2} B_\mu - ig_s \frac{\lambda_a}{2} G_\mu^a \quad (1.2)$$

and where τ_a, λ_a are the generators of $SU(2)$ and $SU(3)$, respectively.

This Lagrangian is invariant under local $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge transformations for fermion and gauge fields.

Here L_i and Q_i denote the left-handed lepton and quark doublets, respectively, while f_R denotes the relevant right-handed singlets. Summation over fermion generations is implied. Equation 1.1 does not contain mass terms for fermions and gauge bosons.

In the case of the electroweak sector, for instance, one has the gauge transformations:

$$L(x) \rightarrow L'(x) = e^{i\alpha_a(x)T^a + i\beta(x)Y} L(x) \quad (1.3)$$

$$R(x) \rightarrow R'(x) = e^{i\beta(x)Y} R(x) \quad (1.4)$$

$$\vec{W}_\mu(x) \rightarrow \vec{W}'_\mu(x) = \frac{1}{g_2} \partial_\mu \vec{\alpha}(x) - \vec{\alpha}(x) \times \vec{W}_\mu(x) \quad (1.5)$$

$$B_\mu(x) \rightarrow B'_\mu(x) = \frac{1}{g_1} \partial_\mu \beta(x) \quad (1.6)$$

The gauge fields and the fermion fields are massless here. It is interesting to note that in the case of strong interactions (while the gluons are indeed massless particles) the mass terms of the form $-m_q \bar{\psi} \psi$ can be generated for the colored quarks in an SU(3) gauge invariant way. This is due to the fact that all (left- and right-handed) quarks belong only to triplets of the SU(3) color group and all transform in the same manner.

The scenario in the electroweak sector is different. One knows experimentally that the weak gauge bosons are massive and the weak interaction is very short ranged. However, as soon as one adds standard mass terms for the gauge bosons, $\frac{1}{2} M_W^2 W_\mu W^\mu$, one immediately violates the local SU(2) \times U(1) gauge invariance.

In addition, if one includes explicitly the mass term $-m_f \bar{\psi}_f \psi_f$ for SM fermions f in the Lagrangian, this become noninvariant under the weak isospin symmetry transformations discussed above, since e_L is a member of the SU(2)_L doublet, while e_R is the SU(2)_L singlet and, consequently, they change under transformation in a different manner.

Therefore, the mass terms for gauge bosons and fermions result in an obvious breakdown of the local SU(2)_L \times U(1)_Y gauge invariance. The ElectroWeak Symmetry Breakdown (EWSB) provides a mechanism to introduce fermion and weak boson masses while keeping the photon massless.

1.2 The Higgs mechanism

The Standard Model (SM) Lagrangian before EWSB has the form:

$$\mathcal{L}_{\text{SM}} = -\frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \bar{L} i D_\mu \gamma^\mu L + \bar{e}_R i D_\mu \gamma^\mu e_R \dots \quad (1.7)$$

For simplicity, the strong interaction part is omitted. In order to generate masses only for the three gauge bosons the Higgs mechanism is considered. We consider a doublet of complex scalar fields Φ

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

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

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

with

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad (1.10)$$

where both the product

$$\Phi^\dagger \Phi = (\phi^{+*} \phi^{0*}) \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{1}{2} \phi_i \phi^i \quad (1.11)$$

and, consequently, the potential $V(\Phi)$ are invariant under the local gauge transformations

$$\Phi(x) \rightarrow \Phi(x)' = e^{i\alpha_i(x)\tau_i/2} \Phi(x). \quad (1.12)$$

For $\mu^2 < 0$, the potential $V(\Phi)$ has a minimum at

$$\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2}$$

and from (1.11) one can conclude there is a manifold of possible solutions of this equation. The usual and convenient choice is $\phi_1 = \phi_2 = \phi_4 = 0$ in Equation 1.8. Therefore, the neutral component (ϕ_3) of the doublet field Φ produce a nonzero vacuum expectation value

$$\langle \Phi \rangle_0 \equiv \langle 0 | \Phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad \text{with} \quad v = \left(-\frac{\mu^2}{\lambda} \right)^{1/2}. \quad (1.13)$$

Now, as previously, using the pattern of the gauge symmetry of (1.12) one can write the field Φ in the exponential form via four fields $\theta_{1,2,3}(x)$ and $h(x)$:

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

Moving to the unitary gauge by means of a gauge transformation of the field in the form

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

one removes three θ_a fields, chooses only one direction, violates three global initial symmetries of the Lagrangian, and leaves only one invariant (1.11). For simplicity, in what follows for the field $\Phi(x)'$ in the unitary gauge (1.15)

the same notation $\Phi(x)$ will be used.

With $\Phi(x)$ from (1.15) one can expand the kinetic term $(D^\mu\Phi)^\dagger(D_\mu\Phi) \equiv |D_\mu\Phi|^2$ of Lagrangian (1.9)

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

$$\begin{aligned} &= \frac{1}{2}(\partial^\mu h)^2 + \frac{g_2^2}{8}(v+h)^2(W_1^\mu + iW_2^\mu)(W_1^\mu - iW_2^\mu) \\ &\quad + \frac{1}{8}(v+h)^2(g_2W_3^\mu - g_1Y_H B^\mu)^2 = \\ &= \frac{g_2^2 v^2}{8}(W_1^\mu + iW_2^\mu)(W_1^\mu - iW_2^\mu) \quad (1.17) \\ &\quad + \frac{v^2}{8}(g_2W_3^\mu - g_1Y_H B^\mu)^2 + \frac{1}{2}(\partial^\mu h)^2 + \dots \end{aligned}$$

The first term in (1.17) is the mass term $M_W^2 W_\mu^+ W^{-\mu}$ for the charged gauge boson field

$$W^\pm = \frac{1}{\sqrt{2}}(W_1^\mu \mp iW_2^\mu) \quad \text{with} \quad M_W = \frac{1}{2}vg_2. \quad (1.18)$$

In particular, the last relation allows one to fix the vacuum expectation value v in terms of the W boson mass M_W and the Fermi constant G_F

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

The second term in (1.17) mixes two neutral components of the gauge fields W_3^μ and B^μ , but after diagonalization (moving to mass eigenstates Z_μ and A_μ) in the form

$$Z_\mu = \frac{g_2 W_3^\mu - g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, \quad A_\mu = \frac{g_2 W_3^\mu + g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}} \quad (1.20)$$

one can interpret it as a mass term $\frac{1}{2}M_Z^2 Z_\mu Z^\mu$ with

$$M_Z = \frac{1}{2}v\sqrt{g_2^2 + g_1^2}. \quad (1.21)$$

Here $Y_H = 1$ was used. It is very important that the neutral field A_μ , being orthogonal to Z_μ , has no mass term at all. No term like $\frac{1}{2}M_A^2 A_\mu A^\mu$ appears. Therefore, with EWSB we move $SU(2)_L \times U(1)_Y$ to a $SU(2)_I \times U(1)_Q$

Since the $U(1)_Q$ symmetry is still unbroken, the photon, which is the associated boson, remains massless. This is due to the fact that the Lagrangian and the vacuum field $\Phi_0 = \langle \Phi \rangle_0$ of the system both and simultaneously remain invariant under a $U(1)$ transformation, which is a direct consequence of the electric charge conservation. The electric charge of the Higgs field Q is connected with the eigenvalue of the weak $SU(2)$ isospin operator $T_3 \equiv \tau_3$ and $U(1)$ hypercharge for the Higgs field Y_H by means of the Gellman-Nishina equation

$$Q = T_3 + \frac{Y_H}{2}. \quad (1.22)$$

Since we have already fixed the charge of the lower $SU(2)$ component of Φ to zero and the isospin for this component $T_3 = -\frac{1}{2}$, we conclude that $Y_H = 1$. Applying relation (1.22) to the upper ($T_3 = \frac{1}{2}$) component of the doublet Φ , one deduces that it is positively charged (this justifies our notation in (1.8)).

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

$$g_{HW} = \frac{1}{2}v g_1^2 = \frac{2}{v}m_w^2 \quad (1.23)$$

$$g_{HZ} = \frac{1}{2}v(g_1^2 + g_2^2) = \frac{2}{v}m_Z^2. \quad (1.24)$$

A relation between the decay ratios of the Higgs boson to a W pair and to a Z pair can be derived from Equations 1.23 :

$$\frac{BR(H \rightarrow W^+W^-)}{BR(H \rightarrow ZZ)} = \left(\frac{g_{HW}}{\frac{1}{2}g_{HZ}} \right)^2 = 4 \left(\frac{m_W^2}{m_Z^2} \right)^2 \simeq 2.4. \quad (1.25)$$

1.2.1 Fermion Masses and Couplings

The Higgs mechanism can also generate the fermion masses, by introducing in the SM Lagrangian an $SU(2)_I \otimes U(1)_Y$ invariant term, called *Yukawa term*, which represents the interaction between the Higgs and the fermion

fields. Since ϕ is an isodoublet, while the fermions are divided in left-handed doublet and right-handed singlet, the Yukawa terms (one for each fermion generation) must have the following expression for leptons:

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

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.

$$\mathcal{L}_\ell = -\frac{G_{H\ell}}{\sqrt{2}} v \bar{\ell} \ell - \frac{G_{H\ell}}{\sqrt{2}} H \bar{\ell} \ell. \quad (1.26)$$

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

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

which becomes in the unitary gauge

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

Therefore, the Yukawa Lagrangian is

$$\mathcal{L}_Y = -G_{H\ell} \bar{L}_L \phi l_R - G_{Hd} \bar{Q}_L \phi d_R - G_{Hu} \bar{Q}_L \phi^c u_R + h.c.. \quad (1.27)$$

From eq 1.26, the mass of a fermion (apart from neutrinos) and its coupling constant to the Higgs boson are found to be

$$m_f = \frac{G_{Hf}}{\sqrt{2}} v \quad (1.28)$$

$$g_{Hf} = \frac{G_{Hf}}{\sqrt{2}} = \frac{m_f}{v}. \quad (1.29)$$

$$(1.30)$$

$$(1.31)$$

Since G_{Hf} is a free parameter of the theory, the mass of the fermions cannot be predicted.

1.3 Higgs boson mass

The Higgs boson mass is the only yet unknown free parameter of the SM, and cannot be predicted by the theory. The mass of the Higgs boson is generated by the Higgs self-interaction and it depends on the parameters v and λ , the coupling of Higgs self-interaction. But while v can be estimated experimentally, λ is characteristic of the field ϕ and it can be determined only by measuring the Higgs mass itself. However theoretical indications exists and experimental constraints exists, from direct and indirect searches at collider experiment.

1.3.1 Theoretical and experimental constraints

An upper limit on the Higgs mass can be given imposing unitarity of the scattering matrix. Considering the elastic scattering of longitudinally polarized Z bosons:

$$Z_L Z_L \rightarrow Z_L Z_L \quad (1.32)$$

in the limit $s \ll m_{2Z}$, the unitarity bound on the corresponding amplitude implies

$$M_H < \sqrt{\frac{16\pi}{3}} v \sim 1 \text{ TeV} \quad (1.33)$$

A more restrictive bound of 800 GeV can be found considering other scattering processes, such as $Z_L W_L \rightarrow Z_L W_L$. This argument is valid only in the limit of perturbative regime.

For what concerns the experimental constraints, they can be divided in two categories: direct ones, deriving from the searches performed at older colliders such as LEP at CERN and the Tevatron at Fermilab, and indirect ones, arising mainly from precision measurements of the electroweak parameters. Direct searches at LEP-II allowed to set a lower limit of 114.4 GeV (95% C.L.) on the Higgs boson mass [8], while from the Tevatron experiment resulted an exclusion of the mass range of 156 to 177 GeV (95% C.L.) [9]

The indirect constraints were determined through other electro-weak observables that are sensitive to the value of the Higgs mass. In such processes the coupling to the Higgs boson modifies the total amplitude through loop corrections and the vacuum polarization of Z and W bosons. It is found that these corrective terms have a logarithmic dependance of m_H [10].

All the precision electroweak measurements performed by LEP experiments and by SLD, CDF and $D\phi$ [21][22] have been combined together and fitted, assuming the SM a correct theory and using the Higgs mass as free parameter. The result of this procedure is summarised in Figure 1.1, where the $\Delta\chi^2$ is the variation of the χ^2 from its minimum value of the fit, and m_H is Higgs mass. The solid curve is the result of the fit, while the blue band represents the theoretical uncertainty due to unknown higher order corrections. The yellow area shows the regions excluded by LEP-II and Tevatron direct measurements. The indirectly measured value of the Higgs boson mass, corresponding to the minimum of the curve in Figure 1.1, is $m_H = 92 + 34 \text{ GeV}$, where the errors represent the experimental uncertainty at 68% C.L. derived from the black line, thus not taking the theoretical uncertainty into account. An upper limit of $161 \text{ GeV}/c^2$ can also be set at 95% C.L., including also the theoretical uncertainty. Including the direct search limit of 114.4 GeV the limit increases to 185 GeV . Since the loop corrections take into account contributions from known physics only, these results are model-dependent.

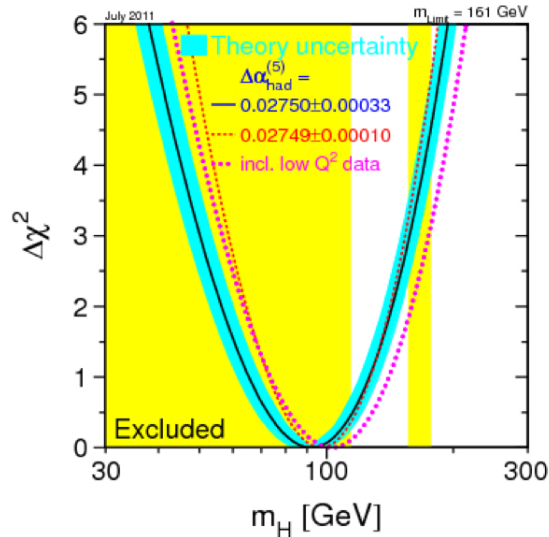


Figure 1.1: $\Delta\chi^2$ of the fit to the electroweak precision measurements performed at LEP, SLC and Tevatron as a function of the Higgs boson mass (July 2011) [11]. The black, solid line represents the result of the fit, and the blue, shaded band is the theoretical error from unknown higher-order corrections. The yellow area represents the regions excluded by LEP-II and Tevatron.

1.3.2 Recent LHC result

Recently, direct searches for the SM Higgs boson have been performed by the ATLAS and the CMS experiments in proton-proton collision at $\sqrt{s} = 7$ TeV and 8 TeV. The CMS experiment, using data samples corresponding to integrated luminosities of up to 5.1 fb^{-1} at 7 TeV and 5.3 fb^{-1} at 8 TeV, has found an excess above the expected background. This excess has been observed at a mass near 125 GeV, with a local significance of 5.0σ (Figure 1.2), a sufficient value to claim the observation of a new particle [28]. The excess is most significant in the two decay modes with the best mass

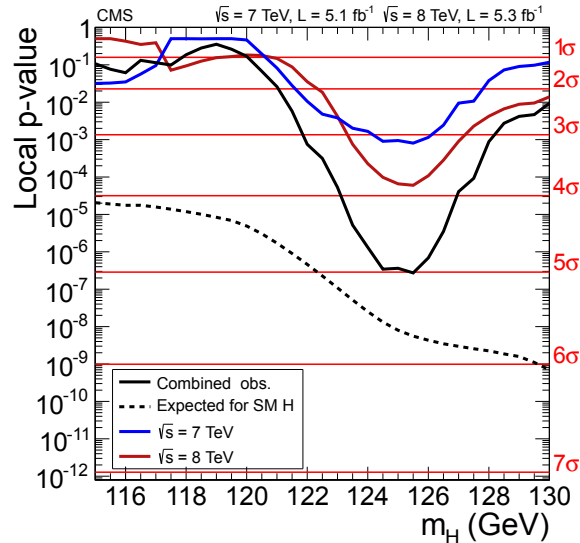


Figure 1.2: The observed local p-value for 7 TeV and 8 TeV data, and their combination as a function of the SM Higgs boson mass [28]. The dashed line shows the expected local p-values for a SM Higgs boson with a mass m_H .

resolution, $\gamma\gamma$ and ZZ , and a fit to these signals gives a mass of 125.3 ± 0.4 (stat.) ± 0.5 (syst.) GeV. The decay to two photons indicates that the new particle is a boson with spin different from one. ATLAS experiment has obtained results consistent with CMS ones[12]. The results presented by the two experiments are consistent, within uncertainties, with expectations for a standard model Higgs boson. A further amount of data resulting in a more rigorous test for this conclusion is needed.

1.4 Higgs phenomenology at LHC

In comparison to lepton collisions where two point-like particles interact, the strong interaction between hadrons is a more complicated process due to their complex internal structure. This structure is described by parton density functions (PDFs), which give the probability to find a quark, or a gluon, with a momentum fraction x of the proton at the momentum scale Q^2 . The energy regions reached by parton collisions at LHC ($\sim TeV$) correspond to a region where the gluon density is dominant with respect to the quarks one. Due to this fact LHC can be thought as a "gluon-gluon collider". Experiments at LHC will search for the Higgs boson within a mass range between 100 GeV to about 1 TeV.

1.4.1 Higgs production

There are four main production mechanisms of the SM Higgs boson at the LHC: gluon-gluon fusion (GGF), vector-boson fusion (VBF), Higgsstrahlung, i.e. the production in association with a vector boson (VH , $V = W, Z$) and associated production with heavy quarks (ttH and bbH). The dominant Higgs production at LHC is the gluon fusion process for all possible Higgs masses. The other processes are interesting because of the special signatures they can provide for identification of the Higgs. In Figure 1.3 [16] the Higgs particle cross section is shown as a function of the mass for different centre of mass energy scenarios. As shown, the total production cross section at 7 TeV is up to one order of magnitude lower than at 14 TeV.

Gluon-Gluon fusion

As stated above, gluon gluon fusion is the dominant production mechanism at hadrons colliders. This is due to the higher luminosity of gluons with respect to quarks. The process diagrams are shown in Figure 1.4. The loop is totally dominated by the top quark because of the strong Higgs coupling (Equation 1.29). 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*

$$K = \frac{\sigma_{HO}}{\sigma_{LO}} \quad (1.34)$$

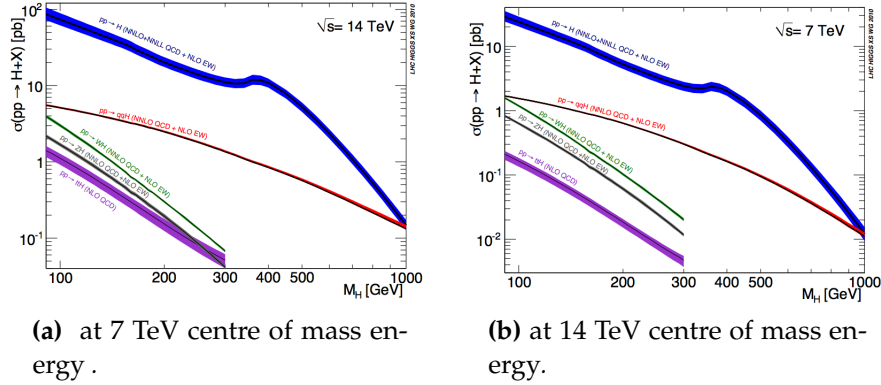


Figure 1.3: Cross section for the different Higgs boson production channels, as a functions of the Higgs boson mass [16].

The K-factor for gluon fusion is ~ 1.7 [13] The results in the computation of the cross section for this process used in the analysis, include next-to-next-to-leading order (NNLO) QCD contributions, complemented with next-to-next-to-leading log (NNLL) resummation, and next-to-leading order (NLO) electroweak corrections. An uncertainty of 15 – 20% on the calculation of this cross section is assumed, principally depending on the parton density functions choice and on the uncalculated higher-order QCD radiative corrections.

Vector Boson Fusion

In the VBF mechanism, the Higgs boson is produced from the fusion of two weak gauge bosons (W^+W^- or ZZ), radiated by the incoming quarks (Figure 1.4d). Its cross section is approximately one order of magnitude lower than gluon fusion for a large range of m_H values. Nevertheless, the peculiar signature of the process (two jets events) makes its identification easier. Also for this process, NNLO QCD and NLO EW calculations are available. The uncertainties are lower than for the gluon fusion mode, of the order of 10%.

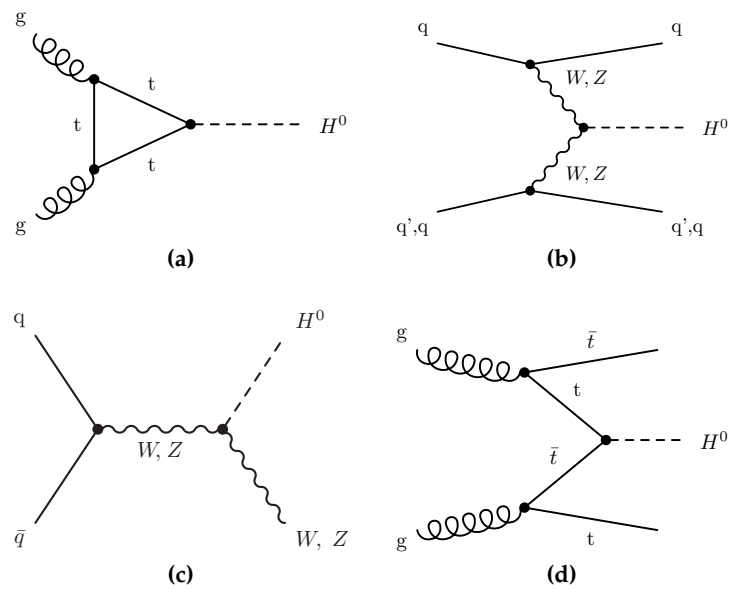


Figure 1.4: Higgs production mechanisms at tree level in proton-proton collisions: (a) gluon-gluon fusion, (b) vector boson fusion, (c) W and Z associated production (or Higgsstrahlung), and (d) $t\bar{t}$ associated production.

Associated Production

The third Higgs production process in order of magnitude is the Higgsstrahlung or W and Z associated production. Even if the cross section is low compared to the total one, this final state can be interesting since they have a characteristic signature that can be used to tag the event. This process is known at the NNLO QCD and NLO EW level. The LO cross section increases about 20 – 25% by the inclusion of the other available contributions.

The last process, illustrated in Figure 1.4d, 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 that of gluon-gluon and vector boson fusion. The presence of the $t\bar{t}$ pair in the final state can provide a good experimental signature. For this cross section, NLO QCD calculations are available.

1.4.2 Higgs decay

The Higgs Boson decays in several ways and the relative branching ratios change dramatically across the possible mass range. The amplitude for the decay of a Higgs boson into a pair of fermions is given by:

$$M = i\bar{u}^{s_1}(p_1)\left(\frac{im_f}{v}\right)v^{s_2}(p_2) \quad (1.35)$$

and the corresponding partial decay width is:

$$\Gamma(H \rightarrow f\bar{f}) = N_c \frac{m_f^2}{v^2} \frac{M_H}{8\pi} \beta_f^3 \quad (1.36)$$

where $\beta_f = (1 - 4m_f^2/M_H^2)^{1/2}$. Fermion decay dominate up to $\sim 150\text{GeV}/c^2$ (low mass region). Above all, the channel $H \rightarrow b\bar{b}$ gives the main contribution, since the b quark is the heaviest fermion available. At high masses, above 350GeV , also $t\bar{t}$ pairs can be produced.

The Higgs boson decay into a vector boson pair, like W^+W^- , ZZ , is mediated at leading order by the S -wave coupling:

$$M_{H \rightarrow AA} = \frac{2M_A^2}{v} \epsilon_{r_2}^\mu \epsilon_{\mu, r_3} \quad (1.37)$$

For $M_H > 2M_V$, the Born width into a pair of on-shell vectors V is given by

$$\Gamma(H \rightarrow VV) = \delta_V \frac{1}{4\pi} \frac{M_V^4}{M_H^2 v^2} \left(1 - \frac{4M_V^2}{M_H^2}\right)^{1/2} \left(3 + \frac{1}{4} \frac{M_H^4}{M_V^4} - \frac{M_H^2}{M_V^2}\right) \quad (1.38)$$

where $\delta_V = 2(1)$ for $V = W(Z)$ accounts for the presence of identical particles in the final state[17]. For large enough M_H values, the decay widths grows like M_H^3 i.e. is very large for $M_H \gg M_W$. Indeed, when the decay channels into vector boson pairs start to be kinematically allowed, they quickly dominate. A peak in the $H \rightarrow W^+W^-$ decay is visible around 160 GeV, where the production of two on-shell W bosons becomes possible and the production of a real ZZ pair is still not allowed. For small M_H , one (two) gauge bosons can be off-shell.

Higgs bosons do not couple to massless particles at tree-level; however couplings can be induced by loops of heavy particles. We have vertex diagrams with fermion (top only) and W boson exchange for $H \rightarrow \gamma\gamma, Z\gamma$ and only top loops for $H \rightarrow gg$.

Considering only top quarks contribution, (the other fermions have negligible Yukawa coupling) the width into gamma-gamma is given by:

$$\Gamma(H \rightarrow \gamma\gamma) = \frac{M_H^3}{9v_2} N_c^2 e_f^4 \frac{\alpha^2}{16\pi^3} \quad (1.39)$$

including the W contribution in the limit $M_H \ll M_W$ the $\Gamma(H \rightarrow \gamma\gamma)$ becomes:

$$\Gamma(H \rightarrow \gamma\gamma) = \frac{M_H^3}{9v_2} \frac{\alpha^2}{16\pi^3} \left| \sum_f N_c e_f^2 - \frac{21}{4} \right|. \quad (1.40)$$

Figure 1.5 shown the branching ratios as a function of the Higgs boson mass.

1.5 Higgs search at LHC

In this section, the different decay channels accessible to experimental searches at LHC depending off the Higgs mass are explored.

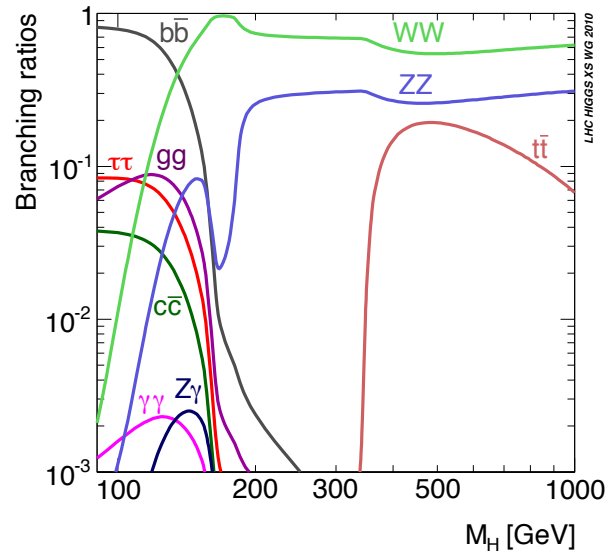


Figure 1.5: The standard model Higgs boson production decay branching fractions [?].

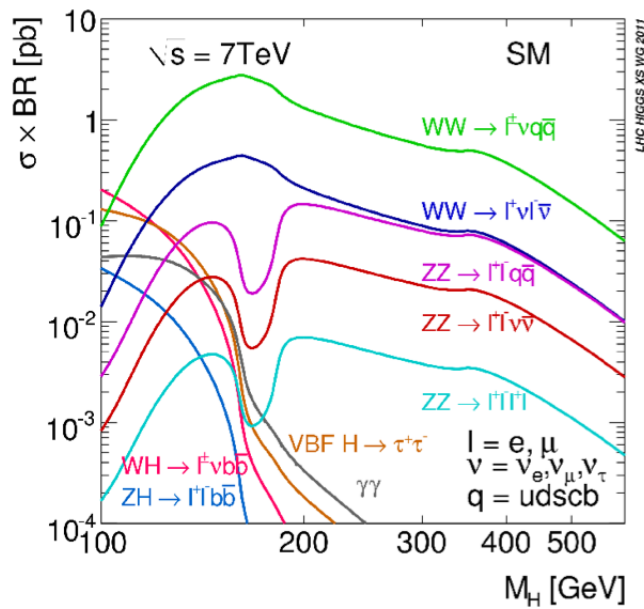


Figure 1.6: Standard Model Higgs boson production cross section times branching ratio as a function of the Higgs boson mass [?].

1.5.1 Low mass region: $110 \leq M_H \leq 135$ GeV

The Higgs boson decays mostly to b quarks ($\text{BR} \approx 50\% - 75\%$), but due to the enormous background from di-jet productions this channel, when produced via gluon-gluon fusion, is experimentally very difficult to search. The scenario improves when produced in association with weak vector bosons (V). Analysis have been performed for the following states: $W(\mu\nu)H$, $W(e\nu)H$, $Z(\mu\mu)H$, $Z(ee)H$ and $Z(\nu\nu)H$. [14]

The second larger BR is $H \rightarrow \tau\tau$ ($\text{BR} \approx 5\% - 7\%$). Also in this case, the Higgs 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 solve this problem this channel is searched in the VBF production; the two forward tagging jets from the incoming quarks which radiate the vector bosons are used to distinguishing the Higgs boson production from SM background processes. The final states $e\tau \rightarrow \text{hadrons}$, $\mu\tau \rightarrow \text{hadrons}$, $e\mu$ are studied.

Despite the small $H \rightarrow \gamma\gamma$ BR, the di-photon final state is important in the lower part of the mass spectrum. This is due to the simple signature of signal. The background is dominated by the irreducible direct di-photon production and the reducible $pp \rightarrow \gamma + \text{jet}$ and $pp \rightarrow \text{jet} + \text{jet}$.

The $H \rightarrow W^+W^- \rightarrow 2\ell 2\nu$ channel can play a role even in this low mass region despite the not favoured branching ratio (Figure 1.6).

As for the $H \rightarrow ZZ$ channels, the only one that plays a role in the low mass region is $H \rightarrow ZZ(*) \rightarrow 4\ell$. In fact despite the $\sigma \times \text{BR}$ (Figure 1.6) is lower than $2\ell 2q$ and $2\ell 2\nu$ ($\ell\ell = e$ or μ) $H \rightarrow ZZ \rightarrow 4\ell$ has clear experimental signature, a high reconstruction efficiency and a very low background. The four mass lepton resolution is about 1% and can contribute to the upper limit for the Higgs boson exclusion as well as to the discovery.

1.5.2 Intermediate mass region: $140 \leq M_H \leq 180$ GeV

When the $2m_W$ mass range is reached, the dominant decay mode becomes $H \rightarrow W^+W^-$ and $H \rightarrow ZZ$. As shown in Figure 1.5, the branching ratio of $H \rightarrow WW$ is higher than $H \rightarrow ZZ$, because of the stronger coupling of the Higgs boson to charged current than to neutral current. Furthermore, this decay mode is particularly important in the mass region between $2m_W$ and $2m_Z$: in this region, the two W produced are both on-shell while in the $H \rightarrow ZZ$ case one of the boson is produced off-shell.

1.5.3 High mass region: $M > 180$ GeV

This region corresponds to values of the Higgs boson mass above the threshold both of $2M_W$ and $2M_Z$, where the Higgs analysis are focused completely on the Higgs decays into a couple of vector bosons. The main channels of interest, as in the intermediate mass region, are those where the two vector bosons decay leptonically. The clear experimental signature of these events compensates for their low branching ratio, which is about one order of magnitude lower than the hadronic ones. Moreover for masses higher than 250 GeV the $H \rightarrow ZZ \rightarrow 2\ell 2\nu$ and $H \rightarrow ZZ \rightarrow 2\ell 2q$ can provide an important contribution to the combined sensitivity to the Higgs boson exclusion.

Chapter 2

The CMS Detector at the LHC

In this chapter, the Large Hadron Collider and the CMS detector are described.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the world major particle accelerator and probably the most complex scientific instrument ever built. The machine was installed in a tunnel 27 km in circumference, on the borderline between France and Switzerland, 50-175 m below ground. The tunnel was built in the 1980s for the previous accelerator, the Large Electron Positron collider (LEP). The LHC produces head-on collisions between two beams of particles, either protons or lead ions. These beams are created in the CERN's accelerator complex and then injected into the LHC (see Figure 2.1).

The collisions occur between particles of the same charge, therefore two separate acceleration cavities with two different magnetic field configurations are needed. The bending power needed to keep the beam circulating is the limiting factor to the achievable centre of mass energy. At the design energy of 14 TeV, a magnetic field of more than 8 T is needed to keep the proton beams in their orbit. Superconducting dipole magnets that are cooled with helium are used to achieve this. Boosts are given by 400 MHz superconducting radio-frequency cavities with a voltage between 8 and 16 MV. The LHC is designed to reach, during its life, a centre-of-mass energy of 14 TeV. The first longer run period in 2010 was performed at only 3.5 TeV energy per beam corresponding to a centre of mass energy of 7 TeV and

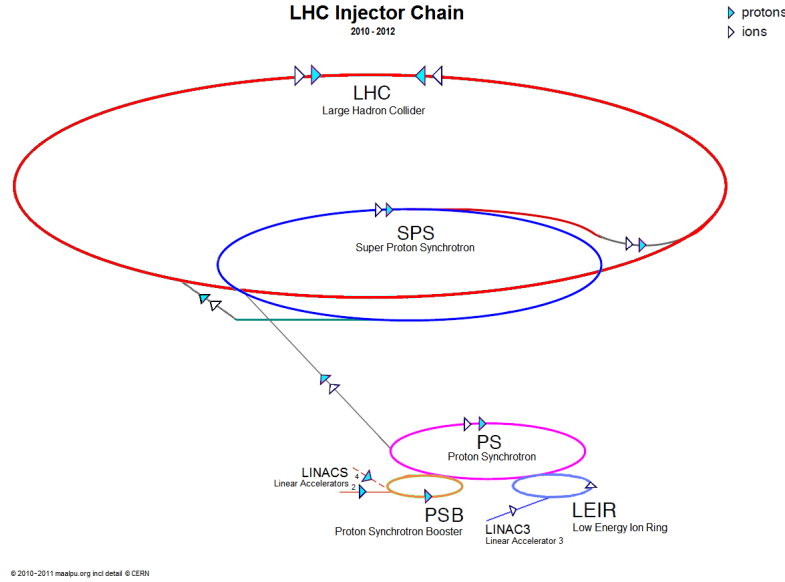


Figure 2.1: LHC chain

after that, the energy has been increased up to 8 TeV for 2012 runs.

The luminosity \mathcal{L} of a collider machine is defined by the ratio of the rate dN/dt of a process and its cross section σ :

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma \quad (2.1)$$

The luminosity can also be written as

$$\mathcal{L} = f \frac{n_b N_1 N_2}{4\pi\sigma_x\sigma_y} \quad (2.2)$$

with f being the revolution frequency, n_b the number proton bunches per beam, N_1 and N_2 number of protons in the bunches, σ_x and σ_y the rms transverse beam sizes in the horizontal (bend) and vertical directions. The instantaneous luminosity obtained in the 2011 and 2012 runs are substantially below the design values. In the next years LHC will progressively increase the luminosity up to design one of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. During 2011 the total integrated luminosity delivered by LHC is 5.73 fb^{-1} of which 5.22 fb^{-1} has been recorded by CMS. In 2012: LHC delivered 21.6 fb^{-1} so far (Figure 2.2). The main parameters of the LHC are summarized in table 2.1.

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

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

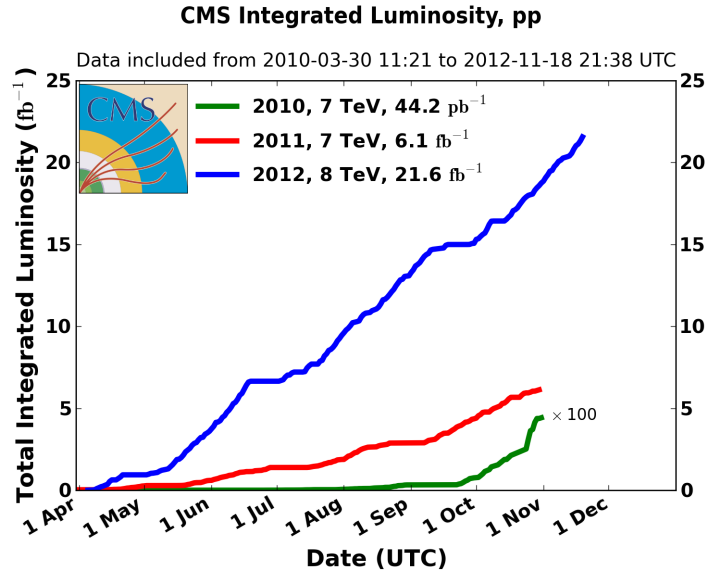


Figure 2.2: Integrated luminosity delivered by the LHC (in red) and recorded by CMS (in blue) in proton-proton collisions for 2011 and 2012.

2.2 High energy collision kinematic variables

To describe the phenomenology at hadron colliders is necessary introduce the common variables and definitions used. The reference frame used in this work is the standard CMS reference frame (Figure 2.3).

The x-axis points toward the centre of the collider, the y-axis points upwards and the z-axis is along the beam direction. The latter direction is taken as longitudinal, instead the $x - y$ plane, orthogonal to the beam line, is called transverse plane. However, most of the time a pseudo-spherical coordinate system is used, which is defined by the radius r and the azimuthal angle ϕ with respect to y-axis. Instead of the polar angle θ with respect to the z-axis the pseudo rapidity η (described in the following) is used.

Rather than the usual momentum vector, the transverse momentum p_t and rapidity y are used. They are defined as:

$$\mathbf{p}_t = p_x + p_y \quad (2.3)$$

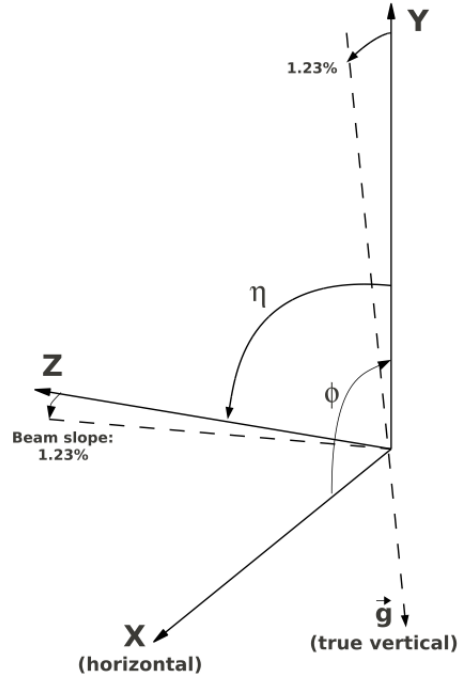


Figure 2.3: coordinate system of CMS.

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \quad (2.4)$$

The former has the property to be invariant under Lorentz transformations along z-axis. The latter, instead, has the property of being additive under Lorentz transformations along the z direction, i.e. it is simply shifted by a constant. For high-energy particles, rapidity can be approximated by pseudorapidity

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

which only depends on the polar angle θ of the particle momentum, i.e. its angle with respect to the z axis.

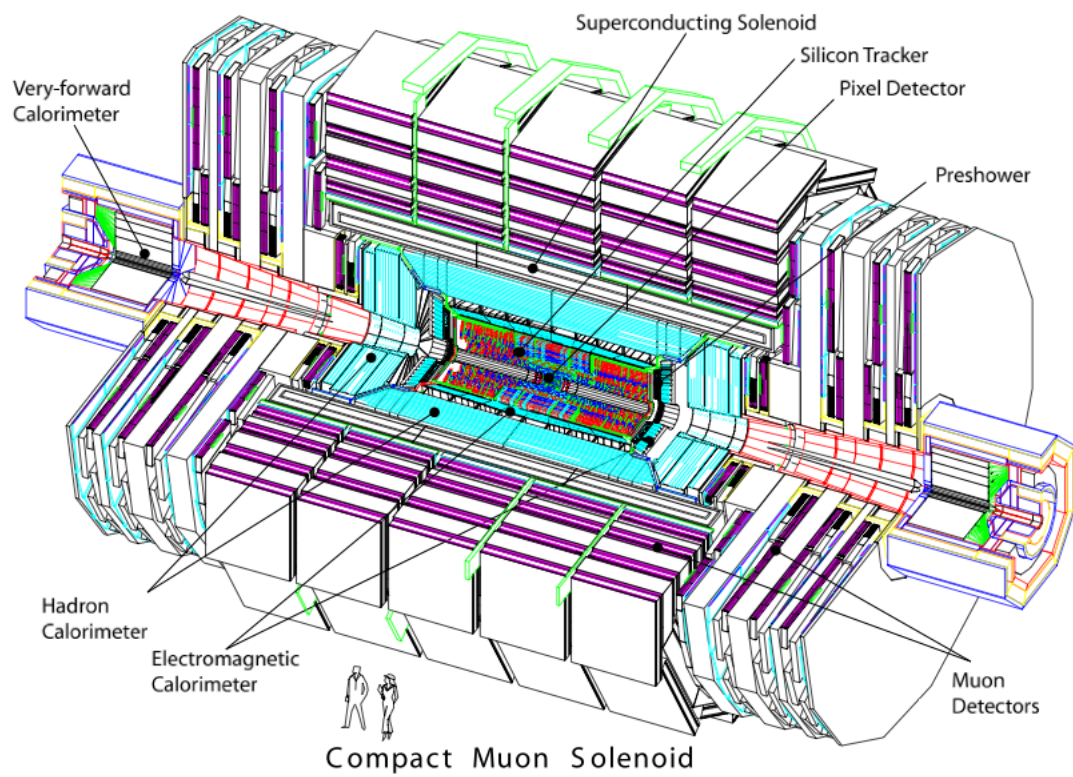


Figure 2.4: Schematic picture of CMS experiment at the LHC.

2.3 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) experiment is one of the four main experiments at the LHC. The main points of its physics program are the study of electroweak symmetry breaking mechanism linked to the discovery of the Higgs boson, the search for physics Beyond the Standard Model (BSM) and the precision measurements of the Standard Model properties. One of the most characteristic features of the CMS detector is the robust, excellent and redundant muon system. This is due to the use of a large, superconducting solenoidal magnet, capable of producing a 3.8 T field. Precise momentum track and unambiguous charge measurements for muons of transverse momenta up to 1 TeV are possible with such a field. The experiment is also characterized with a good electromagnetic calorimeter, a precise and efficient inner tracking system and a hermetic hadronic calorimeter system, capable of delivering good performance in missing transverse energy reconstruction.

The structure of the CMS detector is shown in Figure 2.4. The interaction point or main vertex is located at the centre of CMS. The subsystems are arranged in a cylindrical structure around it. It is divided into a central section, called barrel, made of several layers coaxial to the beam axis, closed at its ends by two discs orthogonal to the beam (the endcaps), that ensure optimal hermeticity.

2.3.1 The Magnet

In order to achieve a compact and high-resolution muon detection system, a large bending power is required. In the CMS experiment this is achieved by a 13 m long superconducting solenoid, with a diameter of 5.9 m. It provides a uniform magnetic field of 3.8 T in the inner region, storing a total magnetic energy of 2.4 GJ. The system is made of $Ni - Tb$ cables operating at a temperature of 4 K, ensured by a sophisticated helium cooling system. At such temperature, the cable becomes superconducting, allowing a 20 kA current to flow without appreciable loss. The whole magnet is then contained in an enormous vacuum cylinder, which isolates it from the external environment. The magnet flux is returned by a saturated iron yoke which permits an independent p_T measurement outside the solenoid.

2.3.2 Silicon Tracking System

The tracker is the first system that a particle coming from the interaction point traverses ($r \leq 10 \text{ cm}$) and is designed to provide a precise and efficient measurement of the trajectories as well as a precise reconstruction of secondary vertices. The structure occupies a cylindrical volume of 2.4 m diameter and 5.4 m length, where the CMS solenoid provides a homogeneous magnetic field of 4 T over the full volume. The principal feature of the detector are the following:

- A reconstruction efficiency for isolated lepton tracks close to 100% in all the pseudorapidity coverage.
- A good lepton momentum resolution (for $|\eta| < 2$, $\sigma(pT)/pT < 4\%$ for single muons in a large range of pT.
- A particle charge identification of $\sim 100\%$ purity for tracks from particles below 1 TeV.
- A good reconstruction of primary and secondary interaction vertices.
- Low rate of fake tracks

The characteristics written above are achieved thanks to low cell occupancy and large hit redundancy. The first is obtained by high granularity detectors, the second is guaranteed by ten layers of silicon detectors providing many measured hits (12 to 14 on average) per track. The Tracking System (see Figure 2.5) consists of two basic parts, the pixel detector and the silicon strip detector. The pixel detector consists of three barrel layers and two endcap disks for each side and covers an area of about 0.92 m^2 . A spatial resolution of about $10 \mu\text{m}$ in the $r - \phi$ plane and $15 \mu\text{m}$ in the z coordinate in the barrel and about $15 \mu\text{m}$ and $20 \mu\text{m}$ in the endcaps is achieved thanks to $100 \times 150 \mu\text{m}^2$ pixels. The silicon strip detector is installed around the pixel detector and consists of 15148 silicon strip modules.

As the pixel detector, the strip detector consists of barrel layers in the central region and discs in the $r - \phi$ plane in the endcaps. Furthermore, the strip detector is divided into inner and outer subdetectors. Both systems are provided with both single-sided and double-sided microstrip modules. The microstrip ensures a spatial resolution of about $40\text{-}60 \mu\text{m}$ in the $r - \phi$ plane and about $500 \mu\text{m}$ along z .

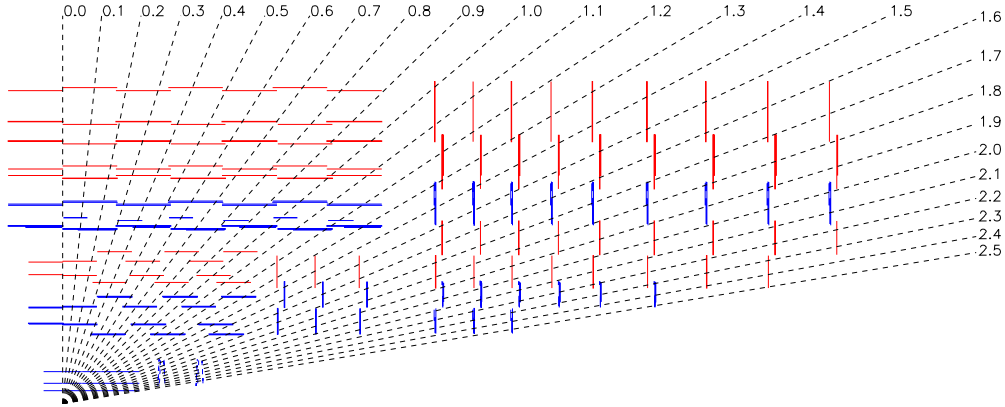


Figure 2.5: Longitudinal view of one quarter of the silicon tracker, including the pixel detector.

2.3.3 The Electromagnetic Calorimeter

The next subdetector enclosing the tracking system is the electromagnetic calorimeter (ECAL). Its principal aim is the identification and measurement of the energy of electrons and photons. The ECAL (Figure 2.6) is made of 74 848 lead tungsten ($PbWO_4$) crystals, which serve as absorber as well as scintillator. They are chosen because of their excellent energy resolution. The advantage of lead tungsten is a high density (8.28 g/cm^3), a short radiation length ($X_0 = 0.89 \text{ cm}$), a small Moliere radius (2.2 cm) and a very short scintillation decay time ($\sim 15 \text{ ns}$). This result in a very fine granularity, a very fast response and, in the same time, a compact structure which permits to place ECAL inside the magnetic coil. The barrel region consists of 61,200 crystals covering the central rapidity region ($|\eta| < 1.48$), with a further 7,324 in each of the endcaps, that extend the coverage up to $|\eta| = 3$. The former covers the pseudorapidity range $|\eta| < 1.479$, the latter the range $1.479 < |\eta| < 3.0$. The crystals are organized in a quasi-projective geometry, so that their axes form a 3° angle with the line that connects them to the nominal interaction point. A single barrel crystal has a front face area of $2.2 \times 2.2 \text{ cm}^2$ (thus matching the Moliere radius) that corresponds to a 0.0174×0.0174 square in the $\eta - \phi$ plane. They are 23 cm long, equivalent to $25.8X_0$. In the endcaps, crystals have $2.47 \times 2.47 \text{ cm}^2$ front face, 22 cm length (corresponding to $24.7X_0$) and are positioned at $z = 3.17 \text{ m}$.

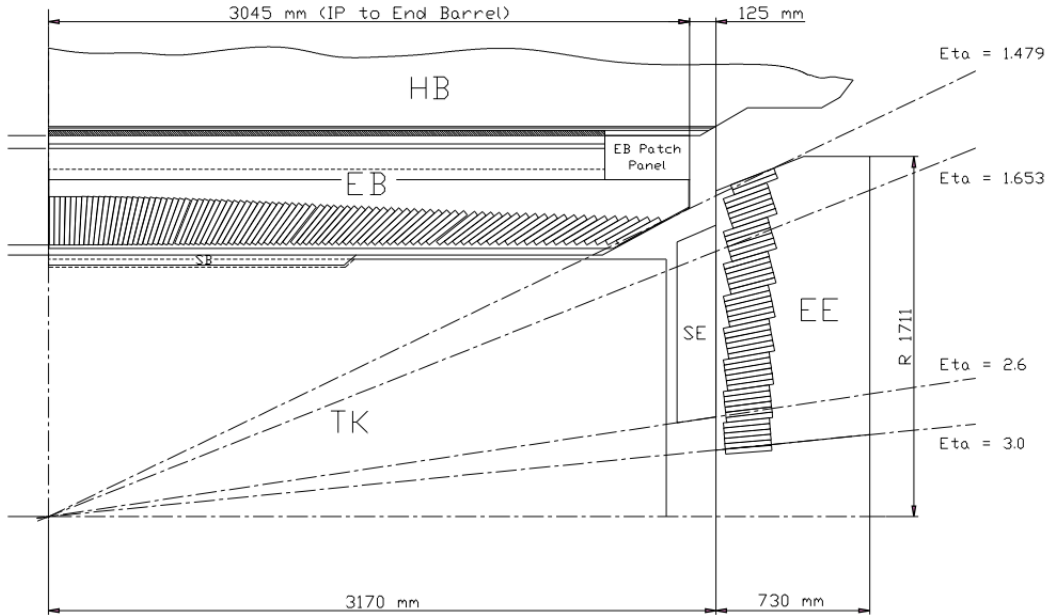


Figure 2.6: Longitudinal view of one quarter of the CMS ECAL.

2.3.4 The Hadron Calorimeter

The hadron calorimeter (HCAL) plays an crucial role in the measurement of hadron jets and neutrinos or exotic particles resulting in apparent missing transverse energy. One of the main design requirements for the HCAL is the high hermeticity. In the forward region, HCAL provides an important information on forward jets, used to reject many backgrounds originating from QCD processes. The hadronic calorimeter system HCAL consists of the barrel (HB), the endcaps (HE), the outer barrel (HO), and the forward calorimeters (HF) (Figure 2.7).

The 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 HE. Both are located inside the solenoid.

The HF is placed around the beam pipe at $|z| = 11m$ and extends the pseudorapidity coverage up to $|\eta| = 5$. Finally, to improve the central shower containment an array of scintillators located outside the magnet, which is referred to as the outer hadronic calorimeter (HO), is used.

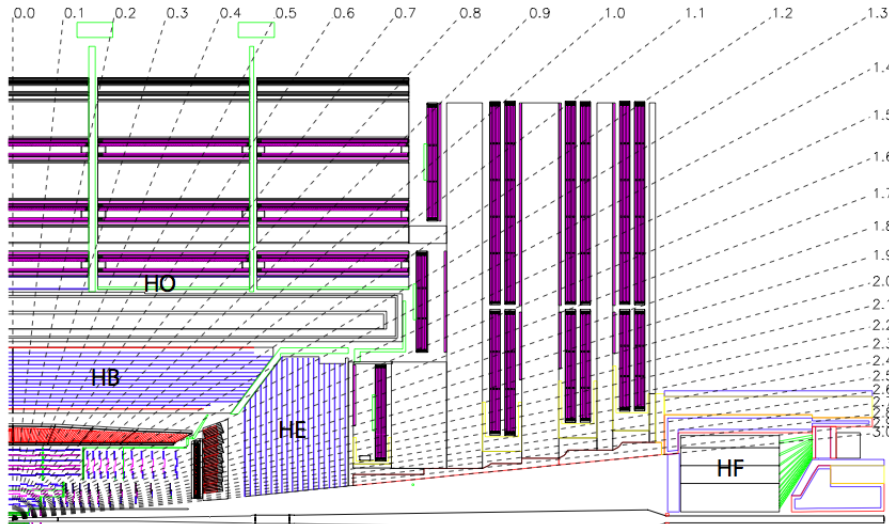


Figure 2.7: Locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters in a longitudinal view of the CMS detector

2.3.5 The muon system

The muon spectrometer [19] is placed outside the magnet and has the purpose of detecting muons, the only charged particles which are not absorbed by the calorimetry. It covers the pseudorapidity region $|\eta| < 2.4$, and provides three functions: muon identification, momentum measurement and triggering. The system is composed of a barrel and two endcaps, which use different technologies. Both regions are organized in four measuring stations, embedded in the iron of the magnet return yoke.

The barrel region ($|\eta| < 1.2$) uses drift tube (DT) stations. The endcaps ($0.9 < |\eta| < 2.4$) instead, are made of cathode strip chamber (CSC) detectors chosen because of the higher magnetic field and a higher particle multiplicity in that region. Both barrel and endcaps are complemented with resistive plate chamber (RCP) detectors.

The DT chamber design is very redundant: each chamber is made up of twelve layers of tubes, organized in three independent subunits, called superlayers (SLs) for a total of 195000 tubes.

CSCs are multi-wire proportional chambers in which the cathode plane has been segmented in to strips running orthogonal to the wires. As written before, the different choice in the endcaps is due to the larger occupancy

expected in these regions, from few Hz/cm^2 to more than $100 \text{ Hz}/\text{cm}^2$ and the intense and non-uniform magnetic field.

The RPCs have a limited spatial resolution, but an excellent time resolution ($\sim 3 \text{ ns}$). This feature makes the RPCs capable of triggering events with muons with high efficiency. They are divided in six barrel and four endcap stations, for a total of 612 chambers.

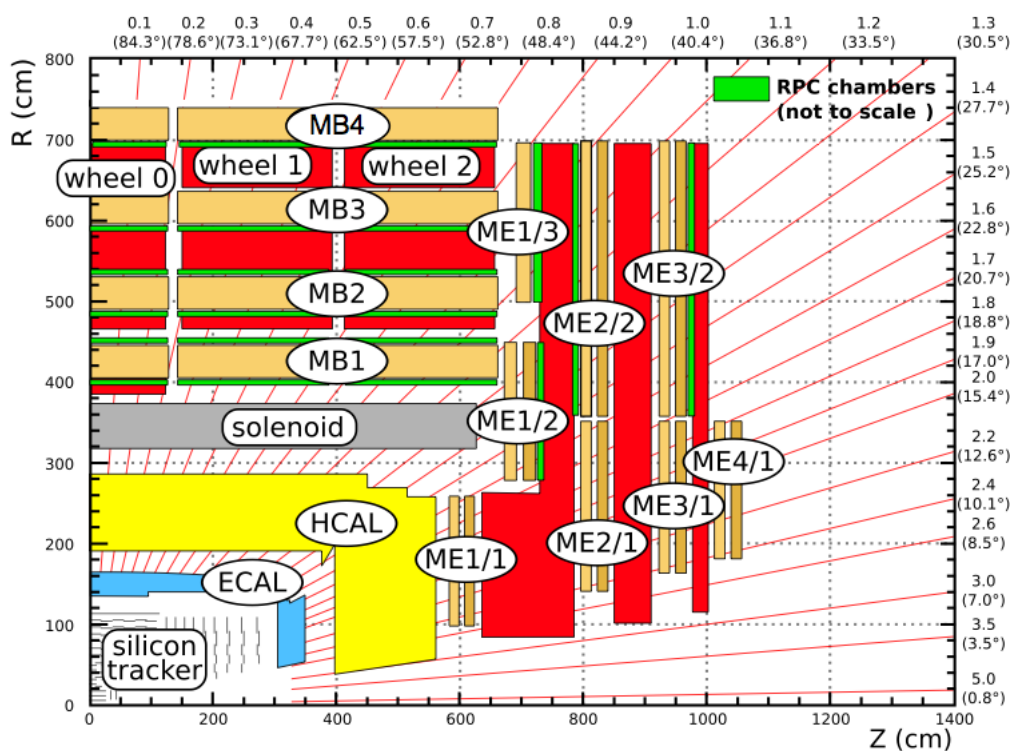


Figure 2.8: An $r - z$ cross-section of a quadrant of the CMS detector with the axis parallel to the beam (z) running horizontally and radius (r) increasing upward. The interaction region is at the lower left corner. The locations of the various muon stations and the steel disks (red areas) are shown. The four drift tube (DT) stations are labelled MB ("muon barrel") and the cathode strip chambers (CSC) are labelled ME ("muon endcap"). Resistive plate chambers (RPC, in green) are in both the barrel and the endcaps of CMS.

2.3.6 The CMS Trigger System

The LHC provides proton-proton and heavy-ion collisions at interaction rates, up to 10^9 at the nominal LHC luminosity. For protons the beam crossing interval is 25 ns, corresponding to a crossing frequency of 40 MHz. Since the typical raw event size is about 1 MB, it is impossible to store and process the large amount of data associated with the resulting high number of events, therefore a drastic rate reduction has to be achieved. Furthermore, most of the interactions are not interesting for the search-oriented physics program CMS intends to pursue. Therefore the aim of the trigger system is that to lower the rate of acquired events to ~ 100 Hz.

This task is performed by the trigger system, which is the start of the physics event selection process. The rate is reduced in two steps, the Level-1 (L1) Trigger and the High-Level Trigger (HLT) [15].

The L1 system is made of a series of custom-designed, largely programmable hardware processors, whereas the HLT is a software system implemented in a computer farm which is organised in several logical levels of increasing complexity, each accessing more data than the precedent one.

The Level-1 Trigger

The Level-1 trigger reduces the rate of selected events down to 50-100 kHz, which is the design output rate limit.

While waiting for the trigger decision the data are stored in pipeline memories, whose depth is technically limited to 128 bunch crossings. Therefore, in the case of a L1 confirmation choice has to be taken in $3.2 \mu\text{s}$; in case of L1 assent, the data are moved to be processed by the HLT. To solve this task, the L1 trigger employs a part of the calorimeter and muon information only, since it has to compute the decision in a time too short to read data from the whole apparatus.

The L1 trigger is organized into a Calorimeter Trigger and a Muon Trigger, whose informations are transferred to final L1 stage, the Global Trigger, which takes the final accept-reject decision.

The Calorimeter Trigger identifies the best four candidates of each of the following classes: electrons and photons, central jets, forward jets and measured calorimetric missing transverse energy. The Muon Trigger is ran separately for each muon detector. The information are then merged

achieving an improved momentum resolution and efficiency compared to the stand-alone systems. The Global Trigger takes the accept-reject decision from both the characteristic of the single objects and of combinations of them. All data from trigger primitives, regional energy sums, muon candidates, and final trigger objects are sent in parallel to the data acquisition (DAQ) system for each accepted event.

High-Level Trigger

The High Level Trigger acquires all events that pass the L1 Trigger filter and has access to the complete read-out data. It can consequently perform complex calculations similar to those made in the offline analysis. The concept of the HLT trigger software is to minimizing the overall CPU usage through the reconstruction of solely those part of each physics object that can be used for selection. The first HLT process selection (Level-2 Trigger), take as input the full Level-1 rate and uses only information from the calorimeter and muon detectors. The second step of the selection (Level-3 Trigger), includes also the reconstruction of full tracks in the tracker. The HLT software is organized in a set of algorithms which are designed to select specific event topologies. Those algorithms reconstruct the information in the CMS detector only partially, taking the accept-reject decision through a regional reconstruction. The HLT system finally reduces the output rate from 100 kHz to about 100 Hz.

Chapter 3

Search for the SM Higgs in the $ZZ(*) \rightarrow 4\ell$ decay channel with CMS

In this chapter a search for the Higgs boson in the $H \rightarrow ZZ^*$ decay channel, with each Z boson decaying to an electron or a muon, is reported. The search covers Higgs boson mass hypotheses in the range $110 < m_H < 1000$ GeV.

3.1 The 4ℓ Final State

In the following, state both signal and background processes are described. In the analysis described in this thesis all Higgs boson production mechanisms described in Section 1.4.1 are considered as part of the signal, but are not discriminated between each other. The three final states examined are 4μ , $4e$ and $2\mu 2e$. The former is the cleanest channel while the latter has the highest BR.

The background can be subdivided in three categories: irreducible, reducible and instrumental. The irreducible background is characterized by a topology similar to the signal one. Therefore, it cannot be removed with the use of kinematic cuts. The main source of this type of background events is given by four leptons event generated from non resonant di-boson production ($ZZ \rightarrow 4\ell$).

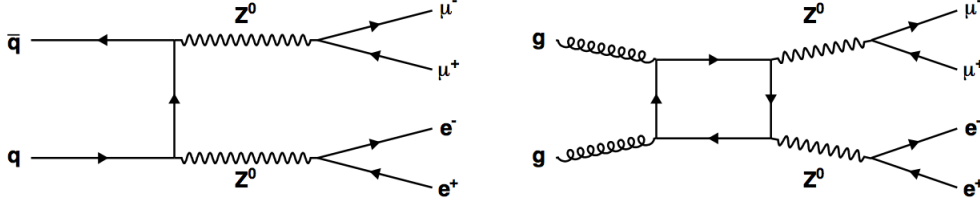


Figure 3.1: Lowest order diagrams for the $qq \rightarrow ZZ^*/Z\gamma^*$ process (left) and for the $gg \rightarrow ZZ^*/Z\gamma^*$ process (right).

The lowest order production mechanism is given by $qq \rightarrow ZZ^*/Z\gamma^*$ (Figure 3.1 left). The gluon-induced ZZ background (Figure 3.1, right), contributes to the NNLO corrections to the process. Above $2m_Z$, it amounts for a non-negligible contribution.

The reducible backgrounds for the $H \rightarrow ZZ \rightarrow 4\ell$ 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 $Zb\bar{b}$ (and $Zc\bar{c}$) associated production, with $Z \rightarrow \ell^+\ell^-$, and the production of top quark pairs in the decay mode $t\bar{t} \rightarrow WbW\bar{b} \rightarrow \ell^+\ell^-\nu\bar{b}b\bar{b}$. These backgrounds are called reducibles as the experimental signature of leptons from b decay can be separated from that of leptons Z decays.

The category of instrumental backgrounds is used to indicate background events with final state leptons from mis-identification of other particles, i.e. mis-reconstructed jets (usually originating from light-flavour quarks) and electrons from γ conversions.

3.2 Data sample

The data sample used in this analysis was recorded by the CMS experiment during 2011 and 2012. The CMS standard selection of runs and luminosity sections is applied which requires high quality data together with a good functioning of the different sub-detectors. Thus, similar detector operation conditions are imposed for the validation of the data to be used for the analysis of the $4e$, 4μ and $2e$, 2μ channels.

CMS recorded in 2011 5.05 fb^{-1} of $\sqrt{s} = 7 \text{ TeV}$ data delivered by LHC.

All of the 2011 statistics is used. For 2012, CMS recorded 21.6 fb^{-1} of $\sqrt{s} = 8$ TeV data delivered by LHC so far. For this analysis, 12.01 fb^{-1} are used from this sample. The absolute pp luminosity is known with a precision of 2.2% [20] in 2011 and 4.4% in 2012 [23].

The data samples are selected by a trigger that combines several collections of High Level Triggers (HLT). The requirement for the 2011 data set relies on the presence of a pair of same sign leptons. The 2012 data, in addition considers also a cross-trigger requiring an electron and a muon in order to recover few percents of inefficiency in the $2e2\mu$ channel at low Higgs boson masses. Moreover a transverse momenta for the first and second lepton of 17 and 8 GeV respectively is required. The trigger efficiency for a Higgs boson signal with $m_H > 120$ GeV within the acceptance of this analysis is greater than 99% (96%, 98%) in the 4μ ($4e$, $2e2\mu$) channels.

3.3 Simulated Samples

SM Higgs boson signal samples, as well as SM background samples, for a large variety of electroweak and QCD-induced processes, been generated using detailed Monte Carlo (MC) simulations. All datasets make use of the full simulation of the CMS detector, and were subjected to full reconstruction and event selection. The signal and background samples have been used for the optimization of the event selection strategy prior to the analysis of the experimental data. Furthermore they are used in this analysis for the comparisons 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. All the signal and background processes cross sections are re-weighted to NLO. The simulated samples are expected to differ from the real data for several reasons. The topic will be discussed in detail in Chapter 4.

Signal sample

The Higgs boson samples used in this analysis are generated with POWHEG which incorporates NLO gluon fusion ($gg \rightarrow H$) and weak-boson fusion $q\bar{q} \rightarrow q\bar{q}H$. Additional samples with WH , ZH and $t\bar{t}H$ associated production are produced with PYTHIA. Generator level events are re-weighted

according to the total cross section $\sigma(pp \rightarrow H)$ which comprises the gluon fusion contribution up to NNLO and NNLL and the weak-boson fusion contribution at NNLO [16]. The total cross section is then scaled by the BR ($H \rightarrow 4\ell$) computed with PROPHECY4F. This calculation includes NLO QCD and electroweak corrections and all interference effects at NLO. In comparison to the $2\mu 2e$ channel, the $4e$ and 4μ channels have an increase of the cross section due to the presence of identical leptons in the final states and the resulting interference. This effect is relevant only in the case of an off-shell Z boson. For the 2011(2012) production, a total of 34(49) Monte Carlo samples were produced in the range 115 to 1000 GeV.

The samples used in the current analysis originate from gluon fusion production mechanism only and have been rescaled to the total cross section including all other processes (weak-boson fusion, WH , ZH and $t\bar{t}H$ associated production).

Background samples

The background samples describing $q\bar{q} \rightarrow ZZ^* \rightarrow 4\ell$ is produced with POWHEG, and includes the complete NLO simulation [24]. The $gg \rightarrow ZZ^* \rightarrow 4\ell$ background is generated with the dedicated tool GGZZ. Both samples are interfaced to PYTHIA [26] for showering, hadronization and decays.

$Z + jets \rightarrow 2\ell + jets$ samples were generated with MADGRAPH at NNLO [25]. Finally the $t\bar{t} \rightarrow 2\ell 2\nu 2b$ sample is generated with POWHEG and include NNLO corrections.

3.4 Physics Objects

For the reconstruction of the SM Higgs boson in the decay chain $H \rightarrow ZZ^* \rightarrow 4\ell$ high-performance lepton reconstruction, identification and isolation are required, together with an excellent lepton energy-momentum measurements. In addition it is important to measure the photons originating from the final state radiation, particularly in the final state with muons. This improves the four lepton invariant mass resolution and the selection efficiency.

Leptons

Because of the low cross section of the signal compared to the total production cross-section observable in the 4ℓ channels, a very high lepton reconstruction efficiency is required. The identification of isolated leptons emerging from the event primary vertex allows for a drastic reduction of QCD-induced sources of misidentified ("fake") leptons. The precision energy-momentum measurement translates into a precision Higgs boson mass measurement $m_{4\ell}$. In the Higgs mass region $m_H < 2m_Z$, a pair of leptons will come from an off-shell Z^* boson.

3.4.1 Electrons

The electron reconstruction combines both ECAL and tracker information. A cluster of energy deposited in ECAL is associated with hits in the outer tracker layers. The combination is used to find electron candidates. Trajectories in the tracker volume are reconstructed using a dedicated algorithm (Gaussian Sum Filter) that takes into account the typical electron energy loss via brehmsstrahlung. Electron candidates are preselected using loose cuts on track-cluster matching observables in order to preserve the highest possible efficiency while removing a portion of the QCD background. For physics analysis, electron candidates are required to have a transverse momentum $p_T^e > 7$ GeV and a reconstructed $|\eta_e| < 2.5$.

Electron candidates are selected using a particular electron identification tool. This tool adopts multivariate techniques which makes use of three main categories of variables: information from purely the calorimetric measurements, from the tracker, and the combined tracker-calorimeter information.

3.4.2 Muons

The reconstruction of muon tracks is done, initially, independently for inner tracker (tracker track) and for the muon system (standalone muon tracks). Afterwards, the two reconstructed objects are combined in two different ways: the Global Muon reconstruction (outside-in) and the Tracker Muon reconstruction (inside-out). The former compares the parameters of two tracks propagated in a common surface in order to find a matching.

Once found the match the *Global-muon track* is fitted. This technique gives appropriate results in the typical p_T ranges used in this analysis.

In the Tracker Muon reconstruction all tracker tracks with $p_T > 0.5$ GeV and the total momentum $p > 2.5$ GeV are considered as possible muon candidates. Taking into account the magnetic field, the average expected energy losses and multiple scattering in the detector material, the algorithm extrapolates a stand alone muon track, for every tracker track. If at least one muon segment (i.e., a short track stub) matches the extrapolated track, the corresponding tracker track is qualified as Tracker Muon. Tracker Muon reconstruction is more efficient than the Global Muon reconstruction at low momenta ($p > 5$ GeV). Candidates found both by the Global Muon and the Tracker Muon approaches that share the same tracker track are merged into a single candidate. The combination of different algorithms provides a robust and efficient muon reconstruction. The muon identification used in this analysis is the *Particle Flow Muon selection*. This method is based on various muon identification variables and combines the information from all subdetectors to identify and reconstruct individually muons produced in the collision.

3.4.3 Lepton Impact Parameter and Isolation

For the process studied in this analysis the electron or muon pairs from Z decays should originate from the primary vertex. To account for this, the significance of the impact parameter on the event vertex is used:

$$|SIP_{3D}| = \frac{|IP|}{\sigma_{IP}} \quad (3.1)$$

where IP is the lepton impact parameter (the distance of closest approach of the track to the primary vertex) in three dimensions, and σ_{IP} is its error. A lepton is selected if $|SIP_{3D}| < 4$.

To ensure that a lepton is an isolated particle in the event, a cut is required. If a lepton comes from a jet decay chain, the presence of additional particles in a solid angle close to the lepton momentum can be observed. Defining a ΔR cone as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ the isolation is computed by the scalar sum of the transverse momentum of the particle flow candidates

reconstructed in a ΔR cone of 0.4, defined as:

$$Iso = \frac{\sum^{chargedhadrons} p_t + \sum^{neutralhadron} p_t + \sum^{photon} p_t}{p_t^{lepton}} \quad (3.2)$$

Because of the presence of multiple interactions within the same bunch crossing, the isolation can be spoiled by the presence of extra tracks coming from the additional interaction vertices. 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 take this effect into consideration. The degradation of isolation performances due to pile-up can be partly mitigated associating the charged particle flow candidates to the primary vertices. For the neutral component (neutral hadrons and photons), the association can not be trivially done, because no tracking information is available. For this reason, a special treatment is needed for neutral particles. Considering pile-up correction it is possible to redefine the neutral isolation sum as:

$$\sum_{neutral}^{correct} p_t = \max \left(\sum_{neutral}^{uncorrect} p_t - \rho \cdot A_{eff}, 0 GeV \right) \quad (3.3)$$

where ρ is the jets median energy density for event and the effective area (A_{eff}) of a given component is defined as the ratio between the slope of the average isolation iso and ρ as a function of number of vertices.

3.4.4 Photons

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

In the electrons case, their measured energies automatically includes the energy of a large fraction of the emitted photons in the associated electromagnetic super-cluster, as the photon is most often collinear with one of the leptons.

On the contrary, muon measured momenta do not include the emitted photons, since the calorimeter information is not used in the muon reconstruction algorithms. 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, in the 4μ and in the $2e2\mu$ final states. It is also expected to reduce the efficiency of the lepton isolation cut when the emitted photon is within the lepton isolation cone.

In order to guarantee an excellent four lepton invariant mass resolution and a large selection efficiency is necessary to recover the FSR photons with high efficiency and purity. Moreover, it is important to remove the energy of the recovered photons from the lepton isolation cones and to measure the mass of the Higgs boson candidate using the momenta of the leptons together with the recovered photons.

In addition to being collinear with the leptons, final state radiation also tends to favour low energy photon emission.

Photons are reconstructed and identified using the particle-flow reconstruction with a specific clustering algorithm. The determination of the photons energy and direction is validated in the data through $\pi_0 \rightarrow \gamma\gamma$ decays and it is shown to be accurate, reliable, and in agreement with the simulation predictions.

The photon isolation is determined considering charged hadrons, photons and neutral hadrons identified by the particle-flow reconstruction in a cone of size $\Delta R = 0.3$ around the photon direction.

The absolute photon isolation is defined as the sum of the transverse momenta of all these iso-deposits. To discriminate against photons that are produced in pileup interactions, an additional isolation deposit is defined that corresponds to the charged particle sum from the vertices other than the primary vertex. Finally, the pileup-corrected relative isolation is obtained by dividing the absolute isolation by the photon transverse momentum, p_T^γ . and is given by:

$$I_\gamma^{tot} = \frac{I_{ch} + I_\gamma + I_{neut} + I_{PU}}{p_T} \quad (3.4)$$

In case an FSR photon candidate is selected in the event, the isolation sum may have to be modified.

Photons are accepted as possible FSR candidates if:

- $p_T^\gamma > 2 \text{ GeV}$
- the photons have $|\eta| < 2.4$

- the minimum ΔR distance with any of the Z leptons is smaller than $\Delta R < 0.5$.
- the $p_T^\gamma > 4$ GeV and PF isolation smaller than 1.0 if the distance of the photon to the closest lepton is between 0.07 and 0.5

Once photons have been accepted with the above criteria the best photon is selected as following:

- If there is at least one photon with $p_T > 4$ GeV the one with the highest transverse momentum is associated to the Z boson
- If there is no photon with $p_T > 4$ GeV the closest photon to any of the leptons is associated to the Z

The performance of the FSR selection algorithm has been measured using MC simulation samples and the rate was verified by means of single-Z data events. 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 $H \rightarrow 4\mu(2e2\mu, 4e)$ events in this analysis.

3.5 Event Selection

The goal of the analysis in 4ℓ channel ($4\mu, 4e, 2\mu 2e$) is to enhance the signal to background ratio. The event selection is optimized to reject the maximum amount of reducible and instrumental background contributions while maintaining the highest possible efficiency for the signal.

To achieve this, the highest possible lepton reconstruction, identification and isolation efficiencies are required, consistently with a quasi-negligible reducible and instrumental background, in the acceptance range used for this analysis ($p_T > 7$ GeV and $|\eta_e| < 2.5$ for electrons and $p_T^\mu > 5$ GeV and $|\eta^\mu| < 2.4$ for muons).

The first common selection is applied directly to primary data sets (PD) for data reduction. It requires:

- at least one good primary vertex (PV) satisfying the following criteria: high number of degree of freedom ($NPV > 4$), collisions restricted along the z -axis ($z_{PV} < 24$ cm) and a radius of the $PV < 2$ 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 and $p_{T,2} > 10$ GeV both for electron and muon objects, where 1 and 2 refer to the lepton with largest transverse momentum, or to the second highest;
- Invariant mass $m_{2\ell} > 40$ GeV for two same flavour leptons;

For the following selection criteria two kind of leptons are defined:

- loose leptons are electrons within the geometrical acceptance of $|\eta_e| < 2.5$, with $p_T^e > 7$ GeV and having 0 or 1 expected missing inner hits, or muons (global or tracker) satisfying $|\eta_\mu| < 2.4$, $p_T^\mu > 5$ GeV. All leptons 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;
- good leptons are loose leptons on which additional criteria are imposed. Electrons should pass the electron identification criteria as described in section 3.4.1 and muons should meet the Particle Flow Muons requirements in 3.4.2. Both electrons and muons should also have $\Delta R(e, \mu) < 0.05$, $PF_{Iso} < 0.4$ and $|SIP_{3D}| < 4$ (see section 3.4.3).

The loose leptons are used in the estimation of reducible background, the good leptons are used for the rest of the selection, which is performed through a sequence of cuts of steps with ever-increasing tightness. The sequence of cuts is the following:

1. First Z: a pair of good lepton candidates of opposite charge and matching flavour (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 also satisfy $40 < m_{Z_1} < 120$ GeV.
2. Four or more leptons: at least another good lepton pair with opposite charge and same flavour.

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 composed of the highest p_T leptons is chosen. The selected pair should satisfy $12 < m_{Z_1} < 120$ GeV. At this stage, it is required that any two leptons of the pairs selected have $p_{T,i} > 20$ GeV and $p_{T,j} > 10$ GeV.
4. QCD suppression: the reconstructed mass of opposite-sign and same-flavour lepton pair must satisfy $m_{\ell\ell} > 4$ GeV.
5. $m_{4\ell}, Z$ and Z^* kinematics: with $m_{4\ell} > m_{4\ell}^{\min}, m_{Z_1}^{\min} < m_{Z_1} < 120$ GeV and $m_{Z_2}^{\min} < m_{Z_2} < 120$ GeV, where $m_{Z_1}^{\min}$ and $m_{Z_2}^{\min}$ are defined below.

The first step ensures that the leptons in the preselected events lie on the high efficiency plateau for the trigger. After this step the control sample for the $Z + jet, Zb\bar{b}/c\bar{c}$ and $t\bar{t}$ backgrounds are obtained. The second step allows for control of the three-lepton event rates which include WZ diboson production events. The first four steps are designed to reduce the contribution of the instrumental backgrounds from QCD multi-jets and Z +jets, preserving, at same time, the maximal signal efficiency and the phase space for the evaluation of background systematics. Finally, three sets of kinematic cuts are introduced to allow studies of the s-channel contribution and to maximize the sensitivity in different ranges of Higgs boson mass hypothesis:

- $Z \rightarrow 4\ell$ phase space analysis defined by requiring $m_{Z_2}^{\min} \equiv 4$ GeV, $m_{Z_1}^{\min} \equiv 40\text{GeV}/c^2, m_{4\ell}^{\min} \equiv 70$ GeV.
- Baseline Higgs phase space analysis defined by requiring $m_{Z_2}^{\min} \equiv 60$ GeV, $m_{Z_1}^{\min} \equiv 60\text{GeV}/c^2$ and $m_{4\ell} \equiv 110$ GeV. This provides a best sensitivity for $m_H < 130$ GeV.
- High-mass phase space analysis is defined by requiring $m_{Z_2}^{\min} \equiv 12$ GeV and $m_{Z_1}^{\min} \equiv 40\text{GeV}/c^2$.

Given the very small cross section \times branching ratio for signal, it is necessary to enlarge the phase space using the baseline selection, in order to increase the signal acceptance in spite of a larger background. Above ~ 130 GeV it is convenient to reduce the phase space in order to better suppress

the background. For Higgs boson masses above $\sim 2m_Z$, further restricting the phase space of the pair of Z bosons can be made without significant loss of acceptance for signal, with the benefit of a modest reduction of the ZZ^* background. The signal detection efficiencies from MC for a 4ℓ system within the geometrical acceptance as a function of Higgs boson mass hypothesis are shown in Figure 3.2 for the baseline and high mass selection. The total efficiency is evaluated to be rising from about 45% (72%, 54%) at $m_H = 190$ GeV to about 59% (82%, 71%) at $m_H = 400$ GeV for the $4e$ ($4\mu, 2e2\mu$) channel.

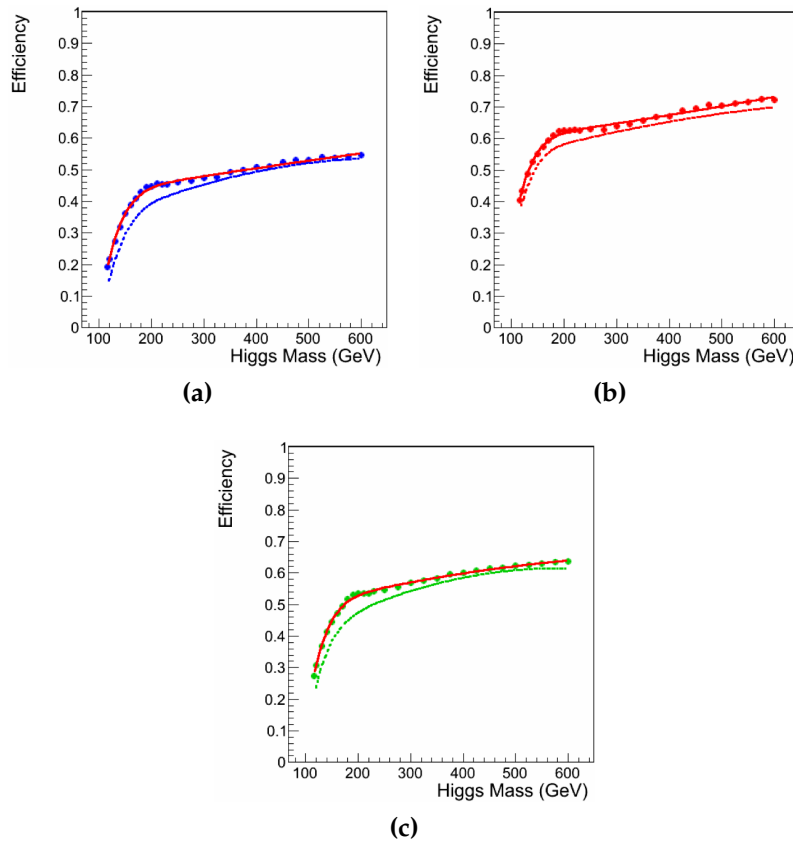


Figure 3.2: Signal detection efficiencies from MC for a 4ℓ system within the geometrical acceptance in the (a) $4e$, (b) 4μ and (c) $2e2\mu$ channels as a function of Higgs boson mass hypothesis. Dashed line are efficiencies as measured for the 2011 PRL analysis.

In Figures 3.3 and 3.4 are shown the events yields as a function of the selection steps for the baseline selection in the $4e$, 4μ .

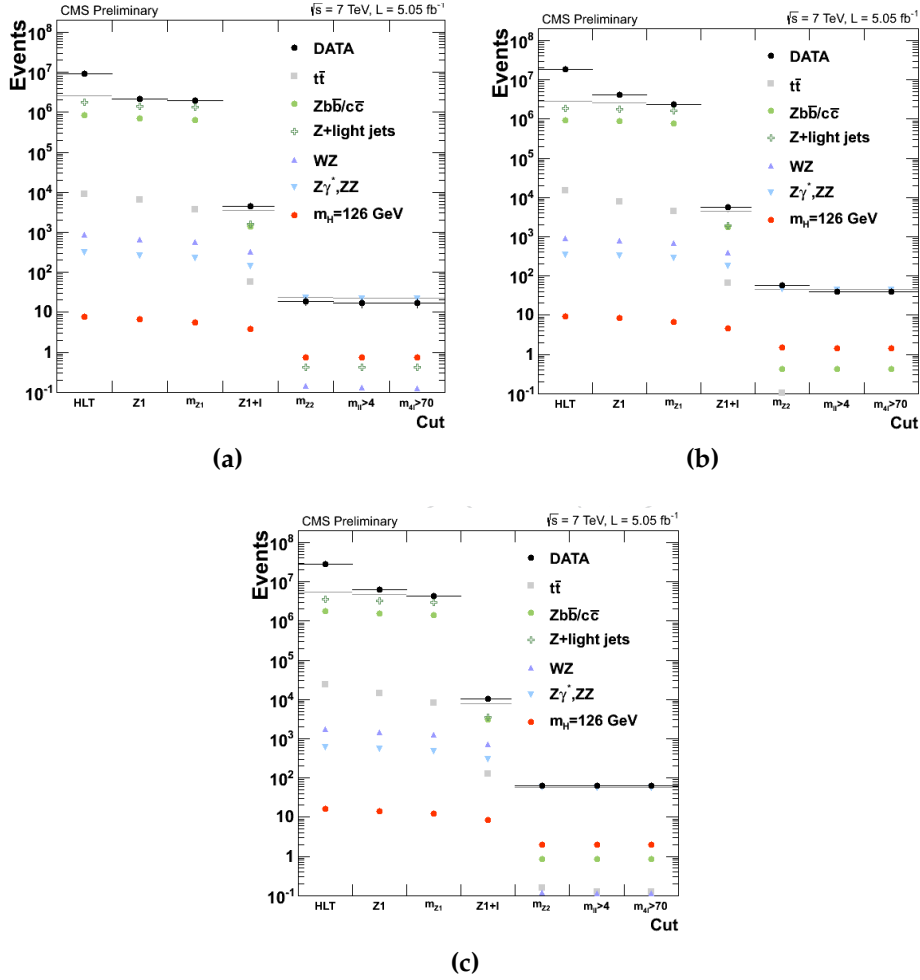


Figure 3.3: Event yields in the (a) $4e$, (b) 4μ and (c) $2e2\mu$ channels as a function of the event selection steps. The MC yields are not corrected for background expectation. The samples correspond to an integrated luminosity of $\mathcal{L} = 5.1 \text{ fb}^{-1}$ of 7 TeV data.

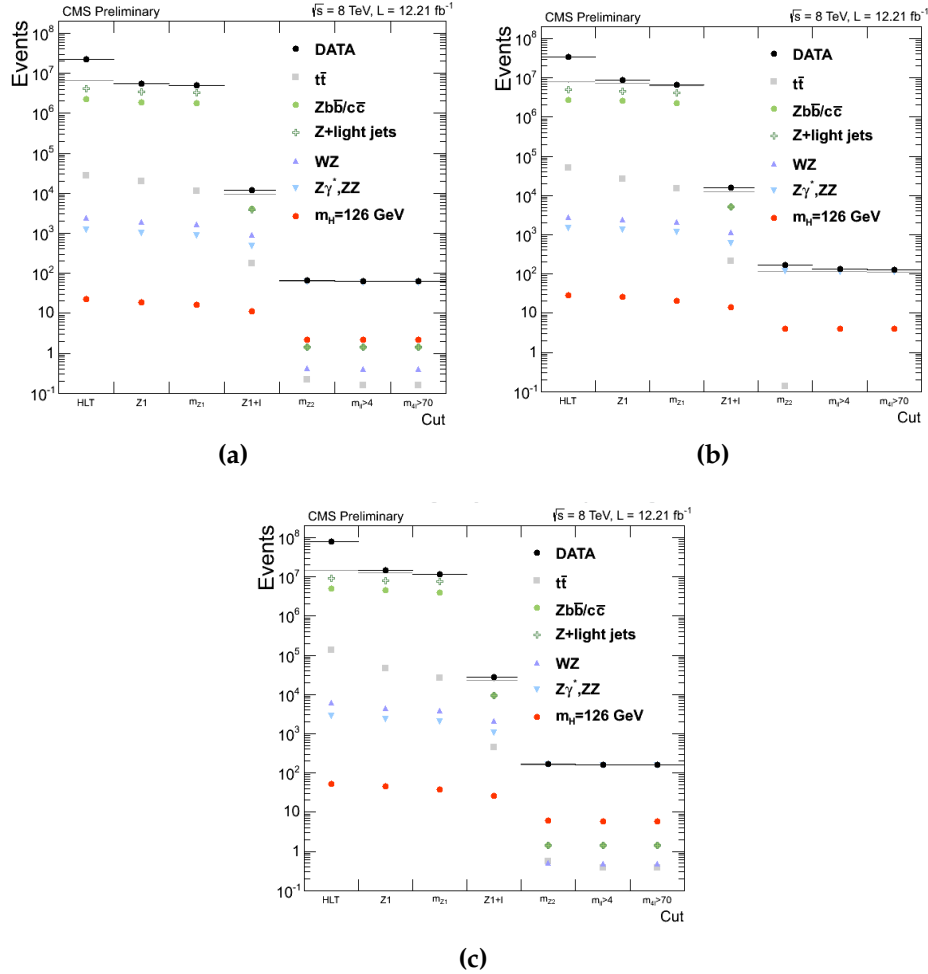


Figure 3.4: Event yields in the (a) $4e$, (b) 4μ and (c) $2e2\mu$ channels as a function of the event selection steps. The MC yields are not corrected for background expectation. The samples correspond to an integrated luminosity of $\mathcal{L} = 12.2 \text{ fb}^{-1}$ of 8 TeV data.

3.6 Kinematic Discriminant (MELA)

It is natural to assume that an experimental analysis will gain greater sensitivity when more information from the observed events is used in the analysis. A complete set of angular observables was introduced in [22] in

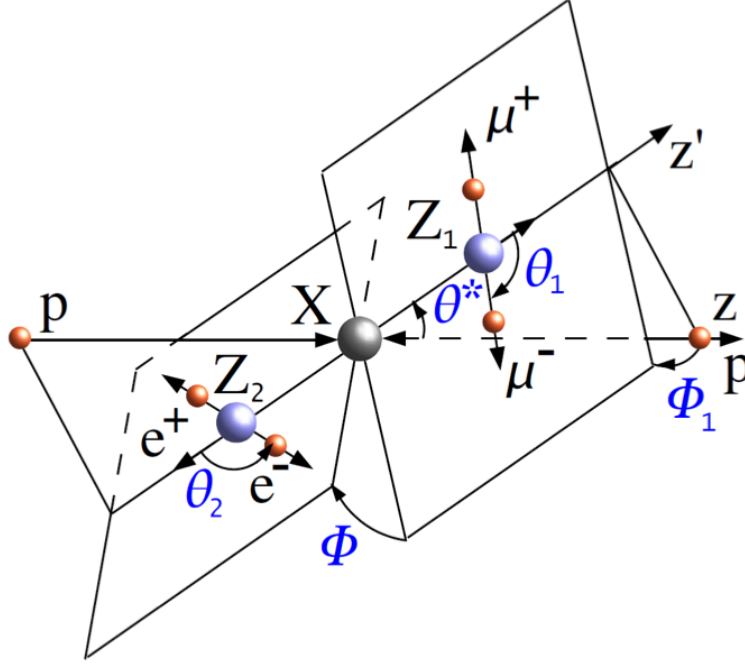


Figure 3.5: Illustration of a particle X production and decay $ab \rightarrow X \rightarrow ZZ \rightarrow 4\ell$ with the two production angles θ^* and Φ_1 shown in the X rest frame and three decay angles θ_1 , θ_2 , and Φ shown in the P_i rest frames [27].

order to help in background rejection through the construction of a kinematic discriminant. In this approach, the signal-to-background probability to be signal or background is created using an analytical or empirical multi-dimension likelihood for an event. In the following, the analytical MELA (Matrix Element Likelihood Approach) method is introduced.

In Figure 3.5 the angular distribution in the production and decay chain $ab \rightarrow X \rightarrow ZZ \rightarrow 4\ell$ 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 θ^* and Φ_1 . The full kinematics of an X resonance $ab \rightarrow X \rightarrow Z_1 Z_2 \rightarrow 4\ell$ can be described with 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 $\Omega = \theta^*, \Phi_1, \theta_1, \theta_2, \Phi$ (Figure 3.5);

- longitudinal boost of the resonance, expressed as rapidity Y ;
- transverse momentum of the resonance p_T and its azimuthal angle;
- one arbitrary azimuthal angle Φ^* reflecting the overall orientation of the system.

The MELA method uses a kinematic discriminant KD from the seven masses and the angular observables

$$KD = F(m_1, m_2, \theta^*, \Phi_1, \theta_1, \theta_2, \Phi) \quad (3.5)$$

and performs a 2D shape fit with the two observables ($m_{4\ell}$, KD).

In Figure 3.6 the separation power between signal and background for each observable is shown. The advantage of the KD discriminant is the combination of all input observables in the most optimal way in a single observable. The observables depending on QCD kinematics (Y , p_t and Φ^*) are removed from the KD creation. Only observables connected with well-understood electroweak processes in the Higgs or ZZ continuum production are kept. The same seven observables are also the key input to measure the new boson properties, such as spin and CP quantum numbers, should a new boson be discovered.

3.6.1 Construction of the MELA discriminant

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

$$\mathcal{P}_{sig}(m_1, m_2, \vec{\Omega} | m_{4\ell}) \quad (3.6)$$

$$\mathcal{P}_{bkg}(m_1, m_2, \vec{\Omega} | m_{4\ell}) \quad (3.7)$$

The probabilities are normalized with respect to the seven observables. $m_{4\ell}$ is instead treated as a conditional parameter. Then the KD discriminant is constructed as follows:

$$KD = \frac{\mathcal{P}_{sig}}{\mathcal{P}_{sig} + \mathcal{P}_{bkg}} = \left[1 + \frac{\mathcal{P}_{sig}(m_1, m_2, \vec{\Omega})}{\mathcal{P}_{bkg}(m_1, m_2, \vec{\Omega})} \right]^{-1} \quad (3.8)$$

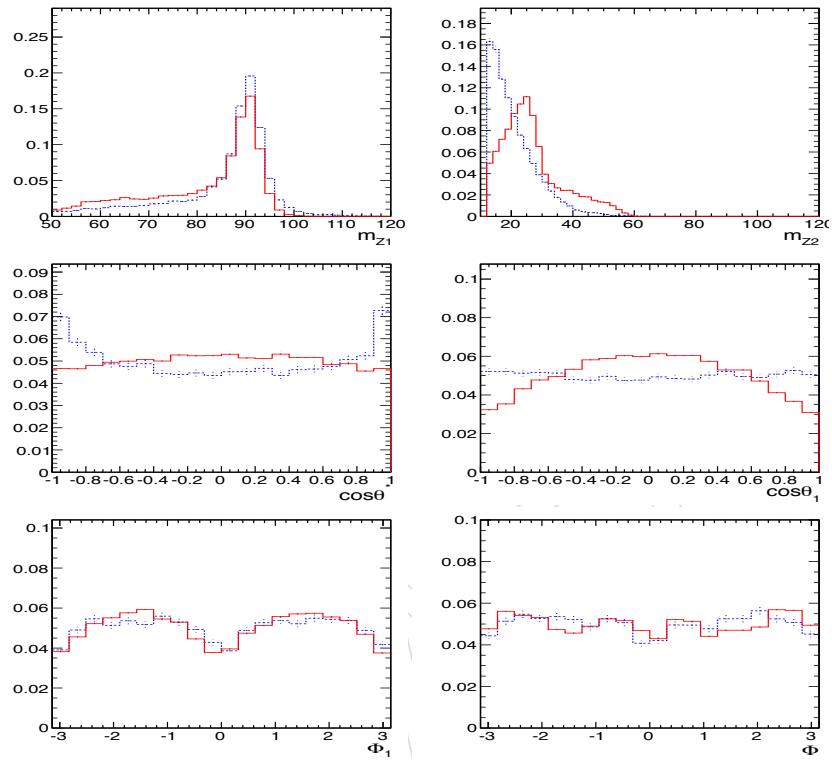


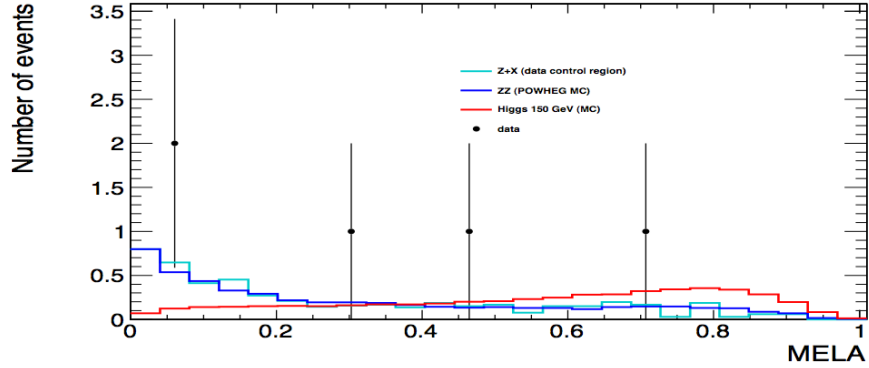
Figure 3.6: Distribution of Higgs signal events with $m_H = 120$ GeV (solid red) and background ZZ events (dashed blue) in the range $100 < m_{4\ell} < 135$ GeV. Top-left: m_1 ; top-right: m_2 ; middle-left: $\cos\theta^*$; middle-right: $\cos\theta_{1,2}$ (both angles have the same distribution); bottom-left: Φ_1 ; bottom-right: ϕ .

This discriminant is continuously distributed between 0 and 1, with signal being closer to 1 and background closer to 0. The signal probability is parametrized as a function of m_{4l} instead of m_H . This choice allows continuous selection of the data-sample independently from the m_H hypothesis. Both signal and background probabilities are normalized for any given value of m_{4l} , which removes unnecessary correlations of KD with m_{4l} and allows, in addition, a more robust fit implementation. The parametrization doesn't include detector effects, and therefore does not require any training. This choice is made possible by the identical detector acceptance effects for signal and background that would cancel in the ratio in Equation 3.8.

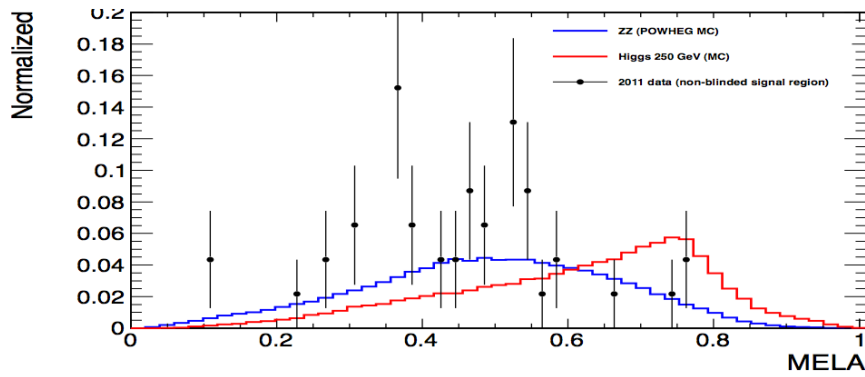
One can note that there is no separation power in the m_1 and m_2 observables above $2m_Z$ threshold. For this reason the choice has been made to make KD a function of only five angles above the threshold.

In order to simplify the background analytical parametrization below threshold, this is substituted with a correlated template distribution (the analytical parametrization is used for validation).

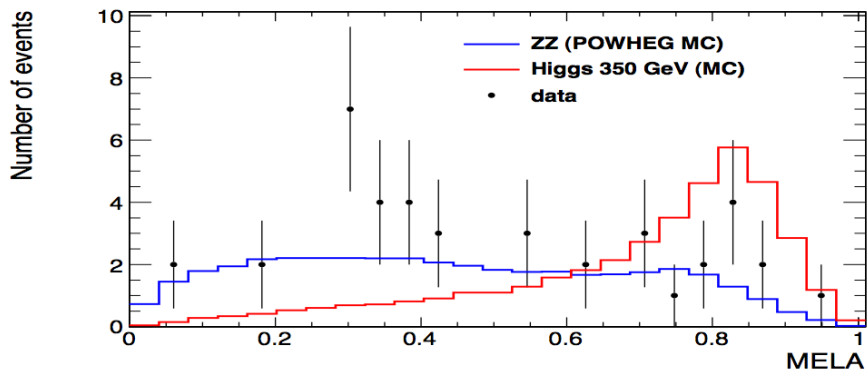
In Figure 3.7 the resulting MELA KD distribution both for signal and background in three different mass ranges with three different Higgs mass hypothesis are shown. As can be seen, there is significant separation between signal and background.



(a)



(b)



(c)

Figure 3.7: The KD distributions for signal and background in three mass ranges: $140 < m_{4\ell} < 160$ GeV (a), $200 < m_{4\ell} < 300$ GeV (b), and $250 < m_{4\ell} < 450$ GeV (c). The signal (red solid histogram) is shown for $H = 150, 250,$ and 350 GeV, respectively. The ZZ continuum background is shown as blue solid histogram. The top plot also shows the Z + X background estimated from data control regions.

3.7 Statistical Analysis

The $(m_{4\ell}, KD)$ unbinned distributions of the selected events are split into six categories based on the three final states (4μ , $4e$, $2\mu 2e$) and two running periods (7 and 8 TeV). These events are examined for 183 hypothetical Higgs boson masses m_H in a range between 110 GeV and 1000 GeV, where the mass steps are optimized to account for the expected for the intrinsic width and experimental resolution [30]. For each mass hypothesis, we perform a simultaneous likelihood fit of the six two-dimensional $(m_{4\ell}, KD)$ distributions using the statistical approaches discussed in [30]. The modified frequentist construction *CLs* [30], [31] as the primary method for reporting limits is adopted.

The probability distribution of $\mathcal{P}^{1D}(m_{4\ell})$ for the background is parameterized 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 differently for two mass regions and will be treated in detail in Chapter 5. The correlated two-dimensional $(m_{4\ell}, KD)$ distribution is described by the two-dimensional probability distribution $\mathcal{P}(m_{4\ell}, KD)$ for signal and background as follows:

$$\mathcal{P}_{sig}(m_{4\ell}, KD) = P_{sig}^{1D}(m_{4\ell}) \times \mathcal{T}_{sig}(m_{4\ell}, KD) \quad (3.9)$$

$$\mathcal{P}_{bkg}(m_{4\ell}, KD) = P_{bkg}^{1D}(m_{4\ell}) \times \mathcal{T}_{bkg}(m_{4\ell}, KD) \quad (3.10)$$

where $P_{sig}^{1D}(m_{4\ell})$ and $P_{bkg}^{1D}(m_{4\ell})$ are the analytical 1D distributions used in 1D invariant mass fits. The 2D distributions $\mathcal{T}_{sig}(m_{4\ell}, KD)$ and $\mathcal{T}_{bkg}(m_{4\ell}, KD)$ are 2D histograms. When the normalized upper limit is lower than unity, it means that the measured cross section maximum value, in that specific range of m_H , is lower than the SM prediction. Consequently, the considered mass region is excluded within the SM. On the other hand, if the normalized upper limit is larger than unity, no exclusion within the SM scenario can be made.

3.8 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 3.8, including the mass below 100 GeV (not used for analysis). It is compared to expectations from the SM background processes.

As expected the $Z \rightarrow 4\ell$ decay gives a clean resonant peak in the four-lepton invariant mass distribution around $m_{4\ell} = m_Z$. At higher mass, the observed distribution is dominated by the irreducible ZZ background, in agreement with expectations. The $Z + X$ reducible background distribution shows that the reducible and instrumental backgrounds contributions are very small. The ZZ and signal normalization and shapes are taken from MC samples. The observed distribution is found compatible with the expectation from SM continuum production of ZZ^* pairs. The number of candidates observed for 7 TeV and 8 TeV are reported in Tables 3.1 and 3.2 together with the estimated background in the signal region and the expected number of signal events for several Higgs hypotheses.

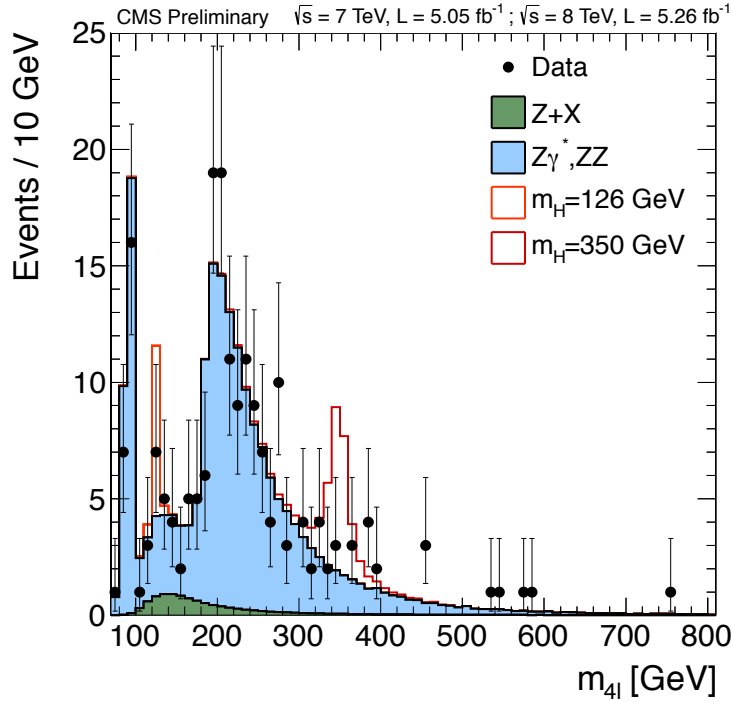


Figure 3.8: Distribution of the four-lepton reconstructed mass in the full mass range for the sum of the $4e$, 4μ and $2e2\mu$ channels. The distributions are presented as stacked histograms. The distribution contains all the selected data events collected with the 7 and 8 TeV statistics [23]

Figures 3.9 and 3.10 show the results for the Upper Limit (UL) and p-value¹ for the entire mass range studied. The SM Higgs boson is excluded by the search in the four-lepton channels at 95% CL in the range 131 – 650 GeV. An excess is observed near the 125 GeV region (Figures 3.11, 3.12), where the minimum of local p-value is reached ($m_{4\ell} = 125.9$ GeV). The signal strength μ , relative to the expectation for the SM Higgs boson, is measured to be $\mu = 0.80^{+0.35}_{-0.28}$ at 126 GeV. The minimum of the local p-value corresponds to a local significance of 4.5σ , 3.1σ in the 1D fit without the MELA KD. The average expected significance for a SM Higgs boson at this mass is 5.0σ and 4.3σ for the 2D and 1D fits, respectively. Using simulation

¹The local p-values represent the significance of local excesses relative to the background expectation

Channel	$4e$	4μ	$2e2\mu$
ZZ background	15.06 ± 1.71	22.58 ± 2.23	35.66 ± 3.64
Z + X	$1.88^{+2.45}_{-1.13}$	$0.99^{+1.49}_{-0.6}$	$2.97^{+2.97}_{-1.49}$
All background expected	$16.94^{+2.99}_{-2.05}$	$23.58^{2.68}_{-2.31}$	$38.63^{+4.7}_{-3.93}$
$m_H = 125$ GeV	0.65 ± 0.1	1.2 ± 0.13	1.54 ± 0.18
$m_H = 126$ GeV	0.72 ± 0.11	1.3 ± 0.14	1.71 ± 0.2
$m_H = 200$ GeV	4.12 ± 0.48	5.86 ± 0.59	9.84 ± 1.02
$m_H = 350$ GeV	2.34 ± 0.28	3.21 ± 0.34	5.59 ± 0.61
$m_H = 500$ GeV	0.81 ± 0.1	1.08 ± 0.12	1.89 ± 0.21
observed	14	20	43

Table 3.1: Number of candidates observed at 7 TeV, compared to the mean expected background and signal rates for each final state. The results are given integrated in the mass range from 110 to 160 GeV.

Channel	$4e$	4μ	$2\mu 2e$
ZZ background	40.2 ± 4.7	60.14 ± 6.66	98.35 ± 10.71
Z + X	$4.46^{+7.58}_{-3.12}$	$2.08^{+2.29}_{-1.04}$	$7.04^{+11.26}_{-4.22}$
All background expected	$44.66^{+8.92}_{-5.64}$	$62.22^{+7.04}_{-6.74}$	$105.39^{+15.54}_{-11.51}$
$m_H = 125$ GeV	1.8 ± 0.25	3.47 ± 0.41	4.52 ± 0.54
$m_H = 165$ GeV	1.99 ± 0.28	3.79 ± 0.44	4.99 ± 0.6
$m_H = 200$ GeV	12.68 ± 1.52	17.82 ± 2.02	30.4 ± 3.38
$m_H = 350$ GeV	7.66 ± 0.94	10.27 ± 1.19	18.2 ± 2.07
$m_H = 500$ GeV	2.8 ± 0.35	3.68 ± 0.43	6.51 ± 0.75
Observed	45	75	119

Table 3.2: Number of candidates observed at 8 TeV, compared to the mean expected background and signal rates for each final state. The results are given integrated in the mass range from 110 to 160 GeV.

it was found that the MELA KD distribution for signal at a mass around $m_H = 126$ GeV is similar for a scalar, pseudo-scalar, or a spin-two resonance with the minimal couplings [22].

Finally, Figure 3.13 shows the $m_{4\ell}$ reconstructed distribution in the mass range 100-160 GeV together with the signal hypothesis for $m_H = 126$ GeV. A good agreement is found.

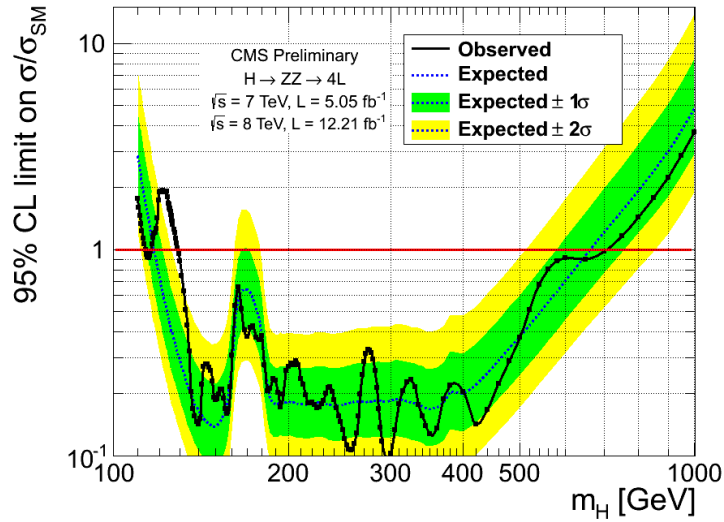


Figure 3.9: Observed and expected 95% CL upper limit on the ratio of the production cross section to the SM expectation with the $2D$ fit. 2011 and 2012 data-samples are used. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively, for the full mass range.

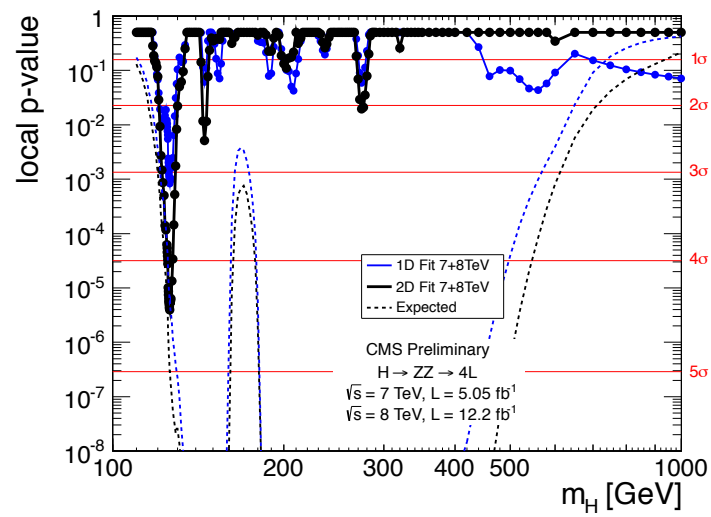


Figure 3.10: Significance of the local fluctuations with respect to the standard model expectation as a function of the Higgs boson mass for an integrated luminosity of 5.1 fb^{-1} at 7 TeV and 12.21 fb^{-1} at 8 TeV with a 1D (blue) and 2D (black) analysis in the full mass range (110-1000 GeV). With respect to the background only SM expectation Dashed line shows the mean expected significance.

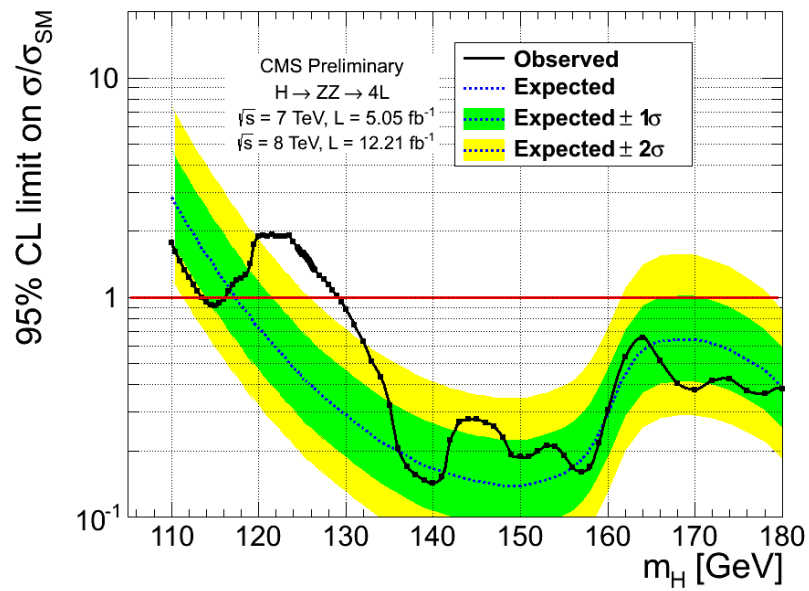


Figure 3.11: Observed and expected 95% CL upper limit on the ratio of the production cross section to the SM expectation with the 2D fit. 2011 and 2012 data-samples are used. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively in the mass range (110-180 GeV).

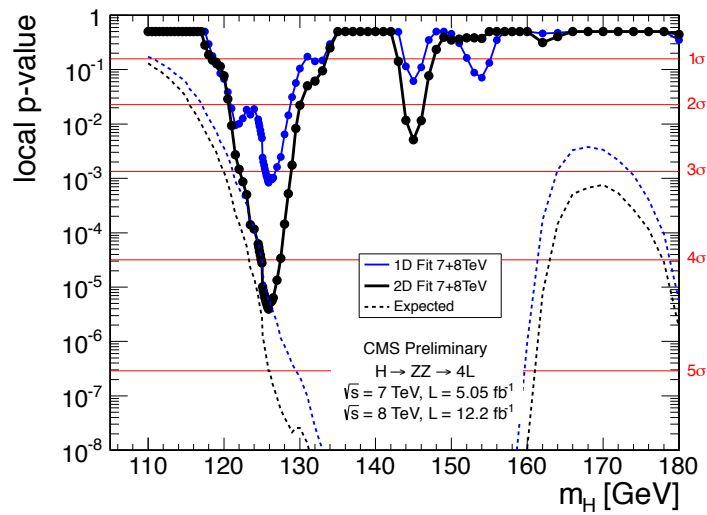


Figure 3.12: Significance of the local fluctuations with respect to the standard model expectation as a function of the Higgs boson mass for an integrated luminosity of 5.1 fb^{-1} at 7 TeV and 12.2 fb^{-1} at 8 TeV with a 1D (blue) and 2D (black) analysis in the mass range (110-180 GeV). With respect to the background only SM expectation Dashed line shows the mean expected significance.

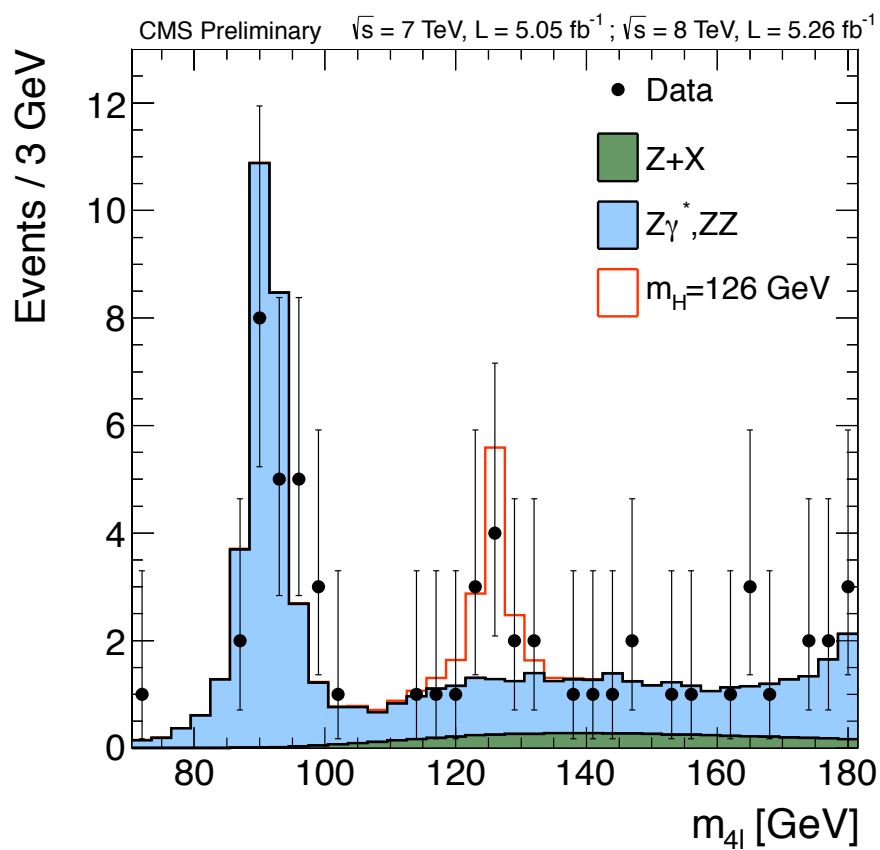


Figure 3.13: Distribution of the four-lepton reconstructed mass in the mass range (110-180 GeV) for the sum of the $4e$, 4μ and $2e2\mu$ channels. The distributions are presented as stacked histograms. The measurements are presented for the sum of the data collected at 7 TeV and 8 TeV, together with the signal hypothesis for $m_H = 126 \text{ GeV}$. [23]

Chapter 4

Monte Carlo reweighting

The Higgs boson signal samples, as well as samples for background processes obtained using Monte Carlo simulation, do not describe out-of-the-box the experimental data with the accuracy required for this analysis. In order to be able to model the signal line shape, and to correctly describe selection efficiencies, I took care of applying proper corrections to simulated samples, as well as estimating the relative systematic uncertainties, as described in the present chapter. The first part is focused on simulated pileup distribution which needs a correction to correctly reproduce the distribution found in the data. The second part concerns corrections applied to reproduce the observed signal efficiencies, while in the last part the theoretical corrections to the line shape in the high mass region is presented.

4.1 Pileup

In high-luminosity colliders, there is a non-negligible probability that one single bunch crossing may produce several separate events. Pileup occurs when the readout of a particle detector includes information from more than one event. This specifically refers to cases where there are other (background) collisions somewhere within a timing window around the signal collision.

In order to let the generated Monte Carlo samples reproduce the luminosity conditions of different data-taking period, the simulated number of pileup events is drawn from a distribution extended enough to cover any possible data distribution.

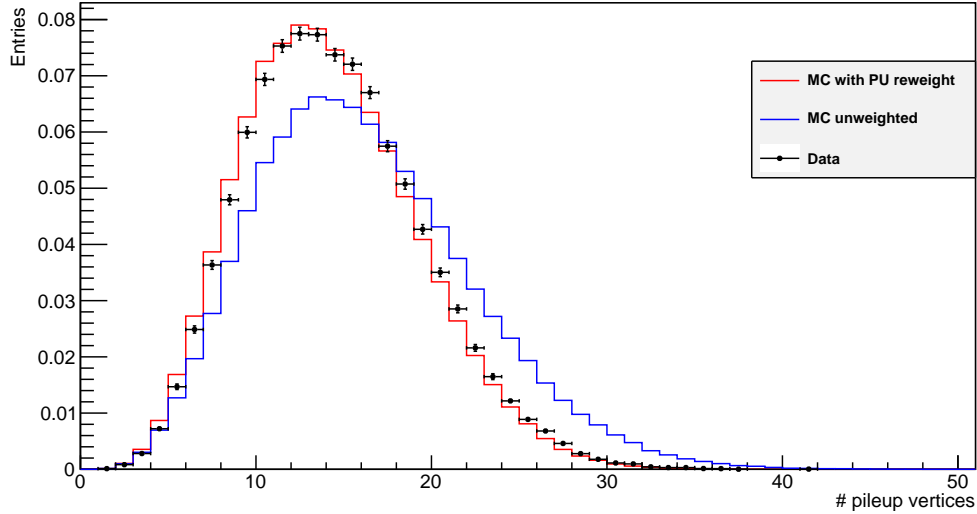


Figure 4.1: Distribution of the number of reconstructed primary vertices before the selection criteria at 8 TeV. The results are shown for data (black), Monte Carlo with the pileup correction (red) and Monte Carlo without the correction (blue).

To be able to compare the Monte Carlo to the data, simulated events have to be re-weighted according to the number of pileup vertices simulated in the event. The weight is the result of the ratio between the data and Monte Carlo distributions. Figure 4.1 shows the distribution of the number of vertices for the Data and Monte Carlo with and without the correction.

Signal efficiency and the *Tag and Probe* method

The efficiencies for reconstruction, identification and trigger for electrons and muons are measured by the tag-and-probe technique directly from the data through the selection of events of inclusive single Z production ($Z \rightarrow \ell\ell$) [28]. This method is based on the tight selection on one lepton, that represents the *Tag*, and a more relaxed selection on the other one, the *probe*. The efficiency is defined as the ratio of the number of probes passing the selection criterion studied, divided the total number of probes. It is important to perform such a measurement with leptons selected with the same criteria used in the analysis. Through appropriate definitions for

probes, the total efficiency per lepton can be factorized in a series of terms, that can be measured independently:

$$\epsilon = \epsilon_{RECO|trackorclustering} \times \epsilon_{ID|RECO} \times \epsilon_{ISO|ID} \times \epsilon_{SIP|ISO} \quad (4.1)$$

where each term represents the probe efficiency to pass a given selection or reconstruction step, given that it passes the previous one. All the efficiencies are measured in several p_T and $|\eta|$ bins, to account to different detector performance in different acceptance and momentum regions.

The measured identification+isolation+|SIP3D| efficiencies for single electrons as a function of the probe p_T , together with MC efficiencies, are shown in 1.19fig:electronefficiency for 2012 data. In Figure 4.3 the reconstruction and identification efficiency for muons for 8 TeV data is shown.

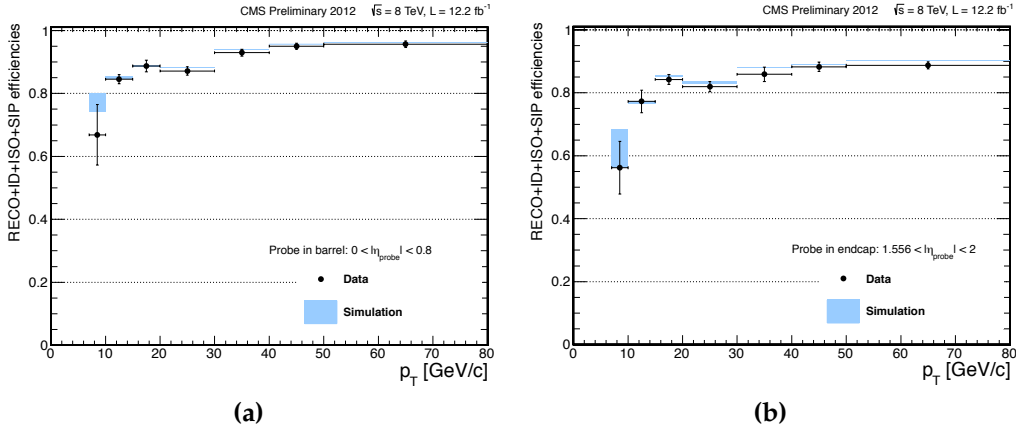


Figure 4.2: Electron identification+isolation+|SIP_{3D}| efficiencies computed with the tag-and- probe method as a function of the probe p_T , in the barrel (a) and endcaps (b). Results are for 8 TeV data [23].

4.2 Efficiency corrections

In order to correct the Monte Carlo, the Tag and Probe method was applied to both data and simulation: in this way it is possible to calculate a correction factor as the ratio between the two efficiencies for each η, p_t bin. The

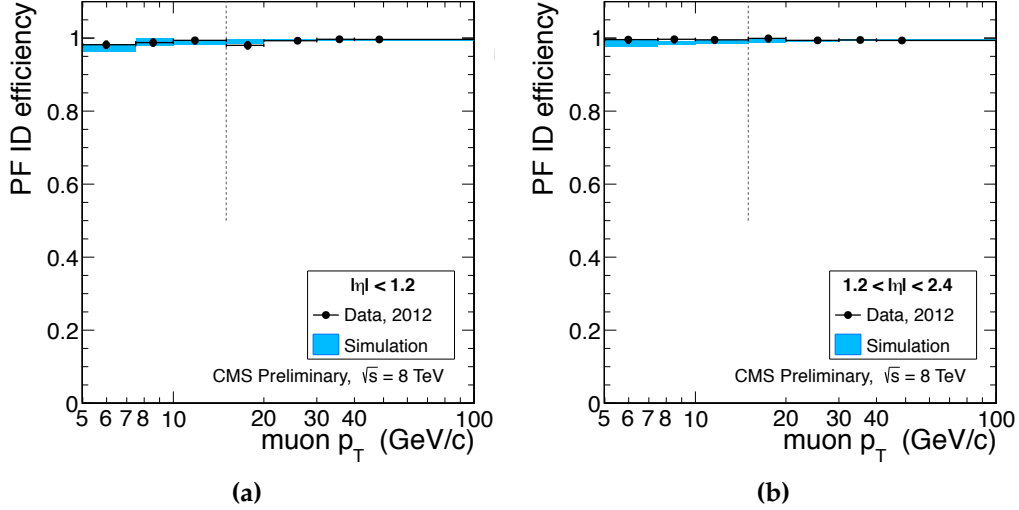
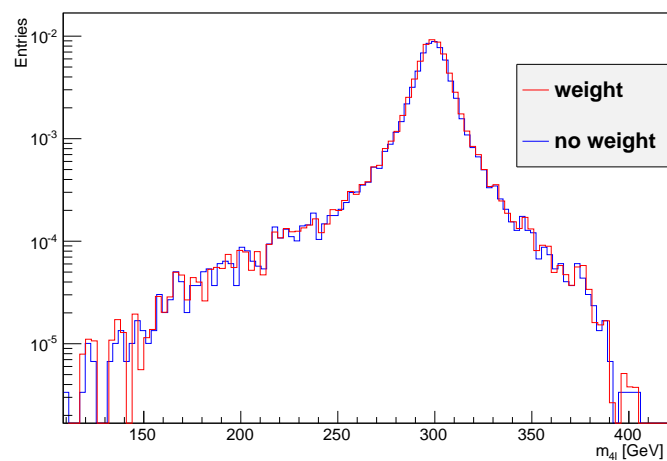


Figure 4.3: Muon reconstruction and identification efficiency for Particle Flow muons, measured with the tag-and-probe method as function of muon p_T , in the barrel (a) and endcaps (b). Results are for 8 TeV data [23].

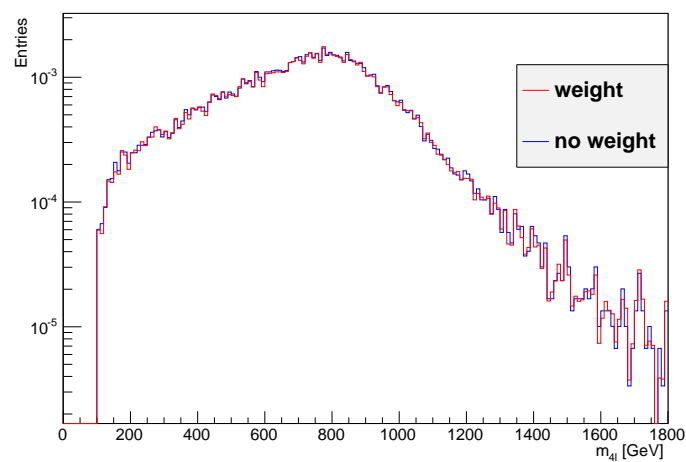
overall lepton weight can be written as follows:

$$w_\ell^{tot} = w_\ell^{iso} \times w_\ell^{ID} \times w_\ell^{SIP} \quad (4.2)$$

Finally, the overall correction factor for a four-lepton event is the product $w_{4\ell}^{tot} = w_{\ell_1}^{tot} \times w_{\ell_2}^{tot} \times w_{\ell_3}^{tot} \times w_{\ell_4}^{tot}$ which is used as a weight for each event. A small effect is expected on the shape distribution. One also expects a variation in the total normalization of the MC. In Figure 4.4 the mass distribution for three different mass hypothesis is shown, with and without the weight. Although the shape variation is small, an average increase of the final yields of about 2%, 3% and 5% for the $4\mu, 2\mu 2e$ and $4e$ channels respectively is found. Once this weights are applied, one must take into account the related systematic uncertainty. The topic will be discussed in detail in Section 5.3.1.



(a)



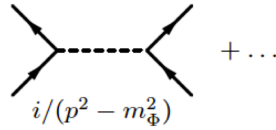
(b)

Figure 4.4: Mass distribution for Monte Carlo events with and without weight for the $m_H = 300$ (a) and $m_H = 99$ (b) GeV Higgs mass hypotheses at 8 TeV.

4.3 High-mass line shape

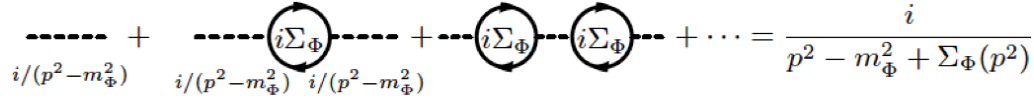
The simulation for a heavy Higgs boson is made assuming an on-shell (stable) Higgs boson production. This means that the Higgs boson production cross section is sampled with a Breit-Wigner distribution (with a running-width approximation) in Monte Carlo simulations. The standard solution for a propagator, in a generic resonance process ($f_1 f_2 \rightarrow \Phi \rightarrow f_3 f_4$) can be computed as follow:

- First order computation



$$i/(p^2 - m_\Phi^2) + \dots$$

- Inclusion of the perturbative corrections



$$i/(p^2 - m_\Phi^2) + \frac{i\Sigma_\Phi}{i/(p^2 - m_\Phi^2)} + \frac{i\Sigma_\Phi i\Sigma_\Phi}{i/(p^2 - m_\Phi^2)} + \dots = \frac{i}{p^2 - m_\Phi^2 + \Sigma_\Phi(p^2)}$$

This approximation can be considered correct in presence of signal only. In the Higgs production and decay, there are other processes at LHC with the same "incoming" and "outgoing" particles that should be taken into account. Processes of this kind will cause quantum interference with signal. On the other hand, the on-shell approximation can be considered acceptable if $p^2 \sim M_\Phi$. This means that for resonances where Γ_Φ is small and consequently the phase space is narrowly distributed around the pole, this approximation can be considered acceptable. Considering this simplified approach, the propagator in momentum space is given by:

$$\tilde{G}(p) = \frac{i}{p^2 - M^2 + iM\Gamma}. \quad (4.3)$$

The accuracy of such propagator is of the order $\sim \Gamma/m$. In order to evaluate this approach to the Higgs search, it is useful to examine the behaviour of Higgs total width in function of the Higgs Mass (Figure 4.5). The width

below the $2m_W$ threshold is of the order of the MeV while above the threshold it rapidly increases by many orders of magnitude. We conclude that this approximation holds for low mass hypotheses only. For example for $m_H \lesssim 350$ GeV the accuracy it is $\Gamma/m_H \lesssim 5\%$ and for $m_H \gtrsim 450$ GeV it is $\Gamma/m_H \gtrsim 10\%$. In other words, for masses above 400 GeV a correction has to be made.

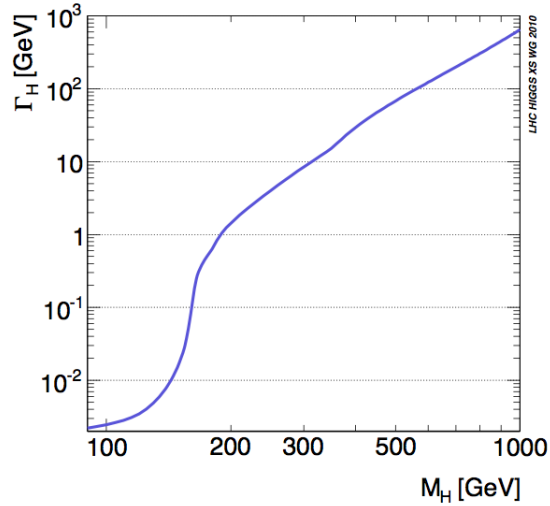


Figure 4.5: total decay width of the Higgs boson as a function of its mass[16].

There are several different theoretical inputs to the off-shellness of the Higgs boson. A possible solution known as Complex Pole Scheme (CPS) is introduced in [32]. This is to-date the most correct approach to describe the Higgs lineshape at high mass. This lead to the introduction of the H complex pole, a gauge-invariant property of the S-matrix. In Figure 4.6 the mass spectrum for $m_H = 800$ GeV is shown, both using the standard Breit-Wigner approach and the CPS scheme.

Moreover, interference effects are expected to be strong between the Higgs boson and SM in ZZ production backgrounds at high invariant masses due to the large intrinsic width of the Higgs boson in that region. At LHC interference between Higgs boson and continuum contributions in $gg \rightarrow ZZ$ is expected. These effects are not taken into account in the original Monte Carlo samples. There are serious problems in including in Monte Carlo generation the signal interference with background in gluon-gluon fusion. The whole cross-section of this process can be written as

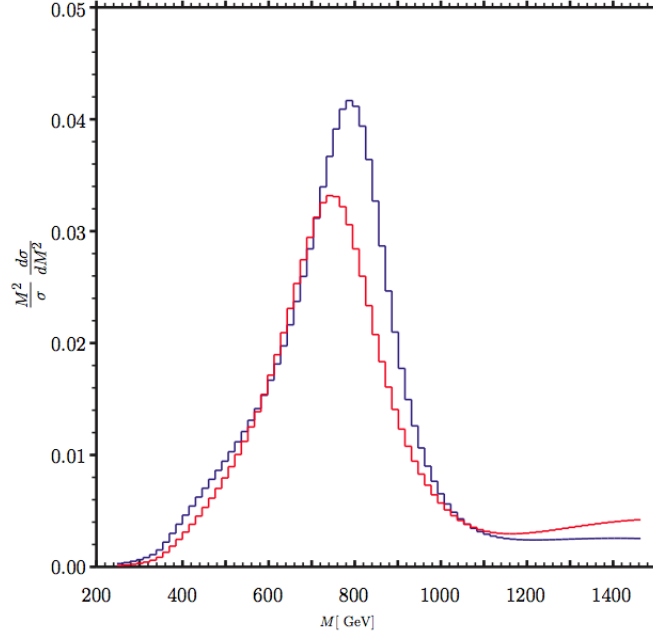


Figure 4.6: The normalized invariant mass distribution in the CPS scheme (blue) and the standard scheme (red) at 800 GeV [32].

follows:

$$\sigma_{gg \rightarrow ZZ} = \sigma_{gg \rightarrow ZZ}(S) + \sigma_{gg \rightarrow ZZ}(I) + \sigma_{gg \rightarrow ZZ}(B), \quad (4.4)$$

where S , B and I stand for signal ($gg \rightarrow H \rightarrow ZZ$), background ($gg \rightarrow ZZ$) and interference.

In Section 3.3 it has been shown that the production cross-section $\sigma_{gg \rightarrow H}$ is known at NNLO and $\Gamma_{H \rightarrow ZZ}$ is known at NLO with leading NNLO effects in the limit of large Higgs boson mass. However the background (continuum $gg \rightarrow ZZ$) and the interference are only known at LO. Let introduce a factor K as:

$$K = \frac{\sigma_{sig}^{NNLO}}{\sigma_{sig}^{LO}} \quad (4.5)$$

This factor shows how much the signal cross section increases at NNLO with respect to the LO. In order to compare the interference with signal at NNLO it is necessary to make some hypothesis about how much the interference can increase from LO to NNLO.

Consider any distribution D , i.e.

$$D = \frac{d\sigma}{dM_{ZZ}} \quad \text{or} \quad D = \frac{d\sigma}{dp_T^Z} \quad \text{etc.} \quad (4.6)$$

where M_{ZZ} is the invariant mass of the ZZ -pair and p_T^Z is the transverse momentum. Two possible options for D_{NNLO} are:

- additive

$$D_{NNLO} = D_{NNLO}(S) + D_{LO}(I) + D_{LO}(B) \quad (4.7)$$

- multiplicative

$$D_{NNLO} = K_D (D_{LO}(S) + D_{LO}(I)) + D_{LO}(B) \quad (4.8)$$

where $K_D = \frac{D^{NNLO}}{D^{LO}}(S)$

Furthermore, it is possible to have the hypothesis:

- intermediate

$$D_{NNLO} = K_D D_{LO}(S) + (K_D^{gg})^{1/2} D_{LO}(I) + D_{LO}(B) \quad (4.9)$$

where K_D^{gg} includes only the gluon gluon contribution and can be written as:

$$K_D^{gg} = \frac{D^{NNLO}(gg \rightarrow H(gg) \rightarrow ZZ(g))}{D^{LO}(gg \rightarrow H \rightarrow ZZ)}. \quad (4.10)$$

In Figure 4.7 two examples of m_{ZZ} distribution are shown, both of them with the correction and with the three different hypotheses described above. Between the three different options the intermediate one is chosen and the other two are used to define a theoretical uncertainty. While the effect of the EW corrections and of the interference on the Higgs lineshape may be different in gluon-gluon production and Vector Boson Fusion (VBF), in the present inclusive analysis we use the $gg \rightarrow H$ lineshape to describe both cases since the gluon-gluon contribution dominates in of whole Higgs mass spectrum up to 1 TeV.

In order to match the Higgs lineshape predicted with CPS and including interference effects, Monte Carlo samples have been reweighted with correction factors depending on the generated mass.

Figure 4.8 shows the corrected distributions for several Higgs mass values. The bands represent the theoretical uncertainty estimated by the two other hypothesis.

The reconstructed m_{4l} distributions after reweighting are shown in Figure 4.9. Figure 4.10 shows the reconstructed invariant mass is shown, for $m_H = 400$ (900) GeV with high mass correction and together with the alternative shapes, describing the lineshape uncertainty. The theoretical uncertainty is used as a systematic uncertainty, as discussed in the next chapter.

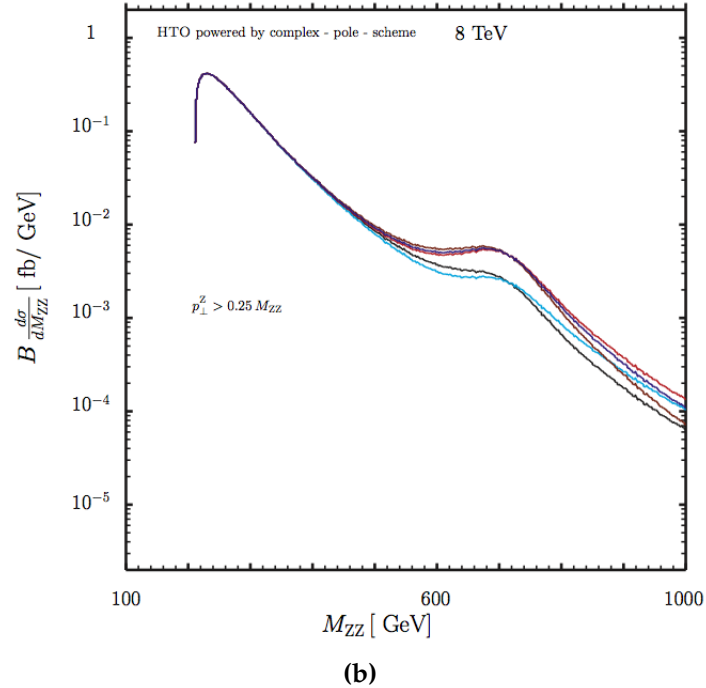
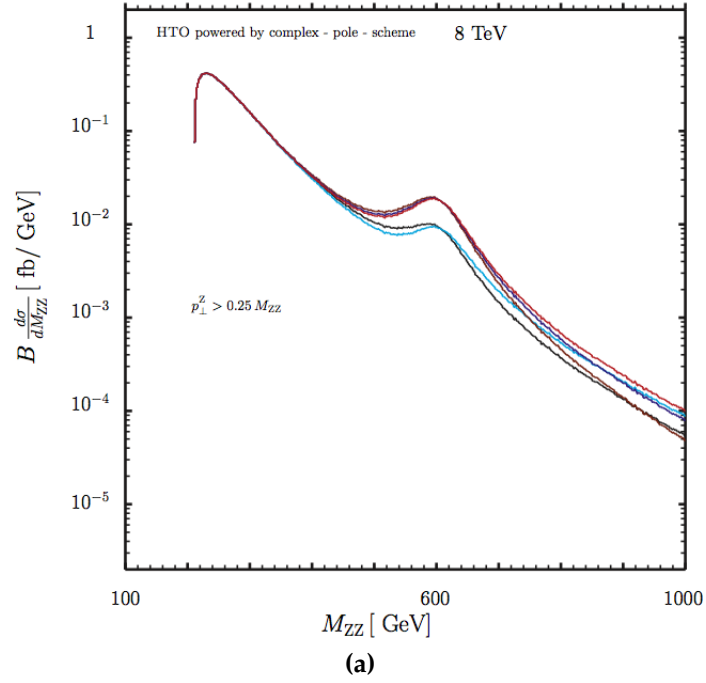


Figure 4.7: The ZZ invariant mass distribution in the CPS scheme for the Higgs hypothesis $m_H = 600$ GeV (a) and $m_H = 800$ GeV (b). The black line is the full LO $gg \rightarrow ZZ$ result, the brown line gives the multiplicative option, the red line is the additive option while the blue line is the intermediate option. The cyan line gives signal plus background (LO) neglecting interference [33].

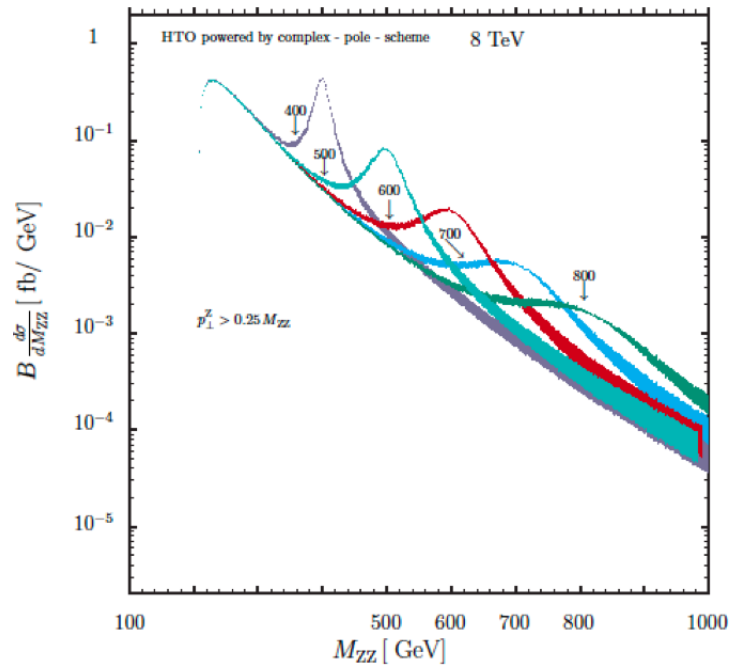
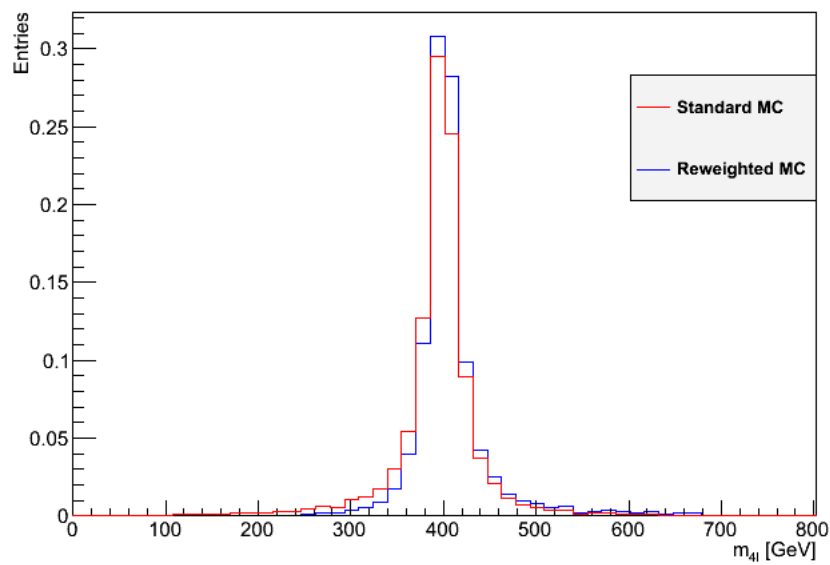
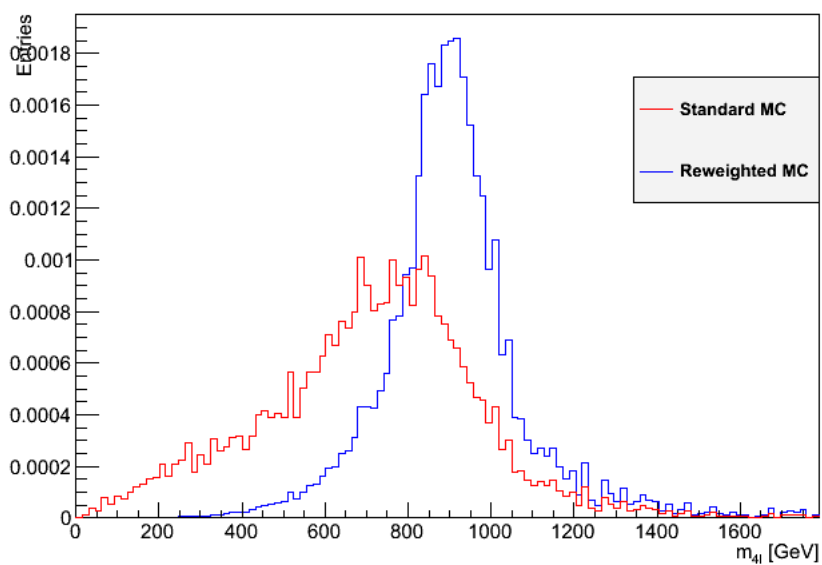


Figure 4.8: Effective NNLO ZZ invariant-mass distribution for $m_H = 400, 500, 600, 700, 800$ GeV including theoretical uncertainty. $B = 4.36 \cdot 10^{-3}$ is the BR for both Z bosons to decay into e or μ [33].

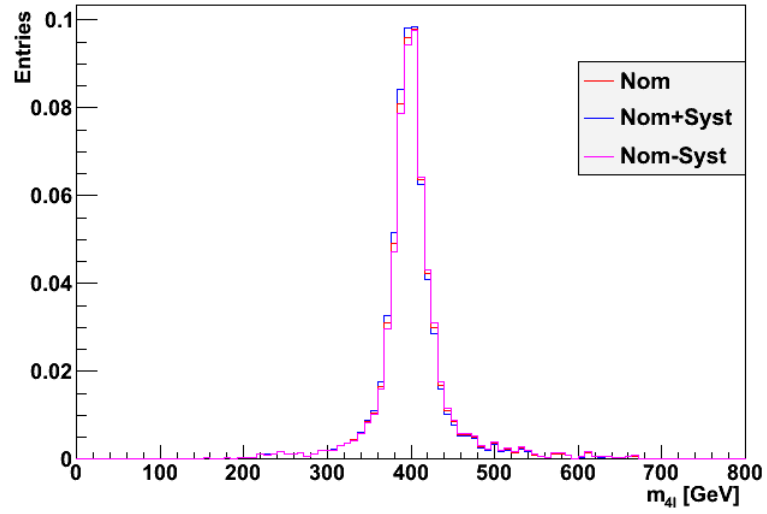


(a)

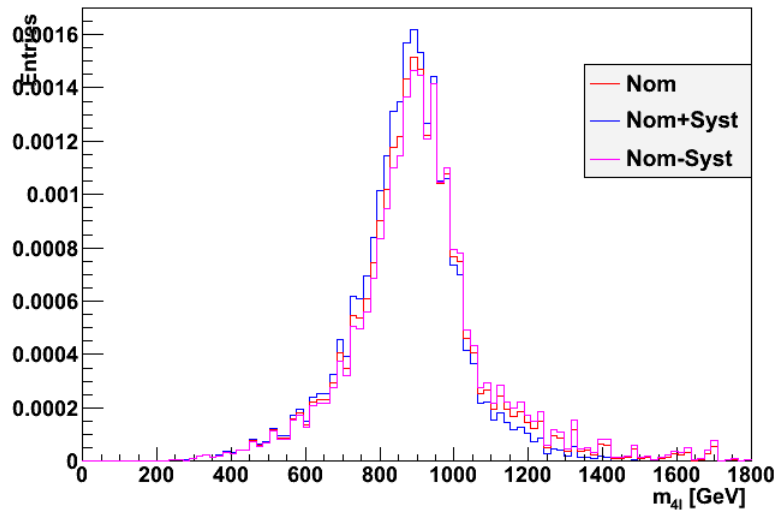


(b)

Figure 4.9: Monte Carlo mass distribution with and without the CPS + interference reweight for 400 GeV (a), 900 GeV (b) at 8 TeV in 4μ channel.



(a)



(b)

Figure 4.10: Monte Carlo mass distribution with the high mass correction at 8 TeV in 4μ channel. The red line is the nominal value, the blue line is nominal value plus the uncertainty and violet line is the nominal value minus the uncertainty, for $m_H = 400$ GeV (a) and $m_H = 900$ GeV (b)

Chapter 5

Signal Modelization

In Section 3.3 it has been shown that, in order to study a large range of mass, samples between 120 and 1000 GeV have been created, 34 for 7 TeV and 49 for 8 TeV. However, for the computation of statistical analysis, a larger number of mass points are needed. A modelization of the signal shape allows a parametrization of the signal mass distributions which, consequently, permits computing the limit for any mass value within the studied range. In order to correctly describe the mass distribution, it is necessary to take into account both the intrinsic line shape and the effect of the detector response. This is achieved using a theoretical function convoluted with a resolution function. The modelling of the signal is performed differently for the two ranges of mass: $m_H = 115 \text{ GeV} - m_H = 400 \text{ GeV}$ (low and intermediate mass range) and $m_H = 400 \text{ GeV} - m_H = 1000 \text{ GeV}$ (high mass range).

5.1 Low and intermediate mass range

For the low mass range the signal $f(m_{4\ell}, m_H)$ Probability Distribution Function (PDF) is modelled as the convolution of the relativistic Breit-Wigner $pdf_1(m_{H^*}, m_H)$ and Double Crystal Ball function. The relativistic Breit-Wigner has the form:

$$f(m_{4\ell}, m_H) = \frac{\Gamma_{gg}(m_{4\ell}) \cdot \Gamma_{ZZ}(m_{4\ell}) \cdot m_{4\ell}}{(m_{4\ell}^2 - m_H^2)^2 + m_{4\ell}^2 \cdot \Gamma^2(m_H)} \quad (5.1)$$

where $\Gamma(m_{4\ell})$ is the total Higgs boson width and $\Gamma_{gg}(m_{4\ell})$ and $\Gamma_{ZZ}(m_{4\ell})$ are the Higgs boson partial width of $H \rightarrow gg$ and $H \rightarrow ZZ$ respectively. Note that m_H is the only free parameter of the function.

The Double Crystal Ball (DCB) has six parameters and is composed of a Gaussian core and two exponential tails. The Gaussian part describes the four-lepton mass resolution through the parameter σ_m and the systematic mass shift with the parameter Δm_{H^*} . The left- and right-hand tail are used to characterize the radiative losses and the not-Gaussian contributions. Its analytical form is:

$$DCB(\xi) = N \cdot \begin{cases} A_L \cdot (B_L - |\xi|)^{-n_L}, & \text{for } \xi < \alpha_L \\ \exp\left(-\frac{\xi^2}{2}\right), & \text{for } \alpha_L \leq \xi \leq \alpha_R \\ A_R \cdot (B_R + |\xi|)^{-n_R}, & \text{for } \xi > \alpha_R \end{cases} \quad (5.2)$$

where $\xi = (m_{4\ell} - m_{H^*} - \Delta m_{H^*})/\sigma_m$ and $n_L, \alpha_L, n_R, \alpha_R$ are the parameters of the tails. Parameters A_L, B_L, A_R and B_R are not independent; they are defined by requiring the continuity of the function itself and its first derivatives. N is the normalizing constant. Figures 5.1 and 5.2 show the fits for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events simulated with $\sqrt{s} = 7$ TeV and 8 TeV.

Once fitted to the reconstructed mass, the shape parameters for all available MC mass points are determined. Parameters, therefore, are plotted as function of the invariant mass and their behaviour fitted with a polynomial function. In this way, a corresponding signal PDF for any value of the Higgs mass hypothesis can be obtained. The interpolation of the parameters is shown in Figures 5.3 and 5.4.

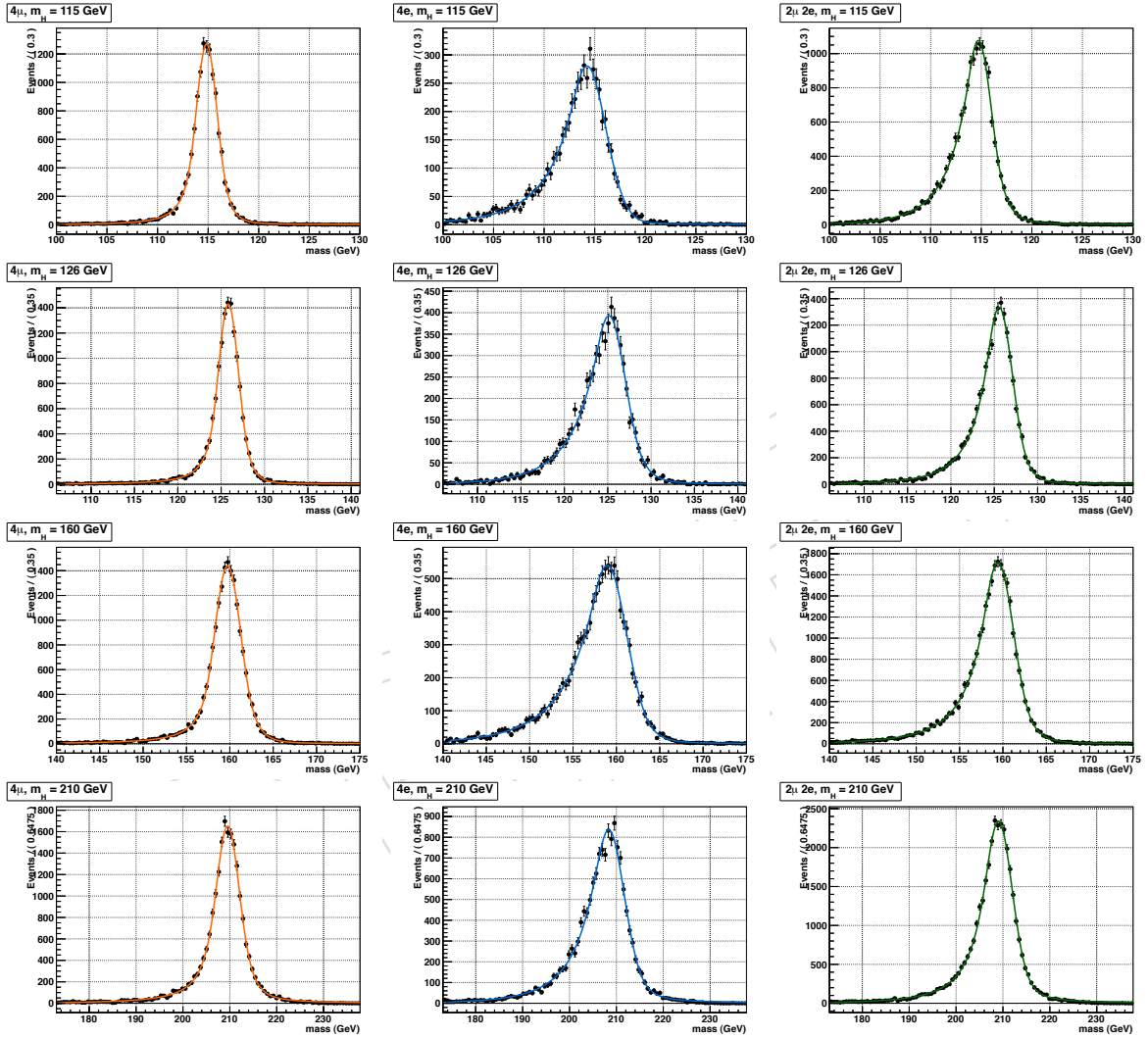


Figure 5.1: Probability density functions $f(m_{4\ell}, m_H)$ for the Higgs boson mass in the low/intermediate mass region at the reconstruction level after the full lepton and event selections are applied. The distributions obtained from 7 TeV MC samples are fitted with the model described in the text for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events.

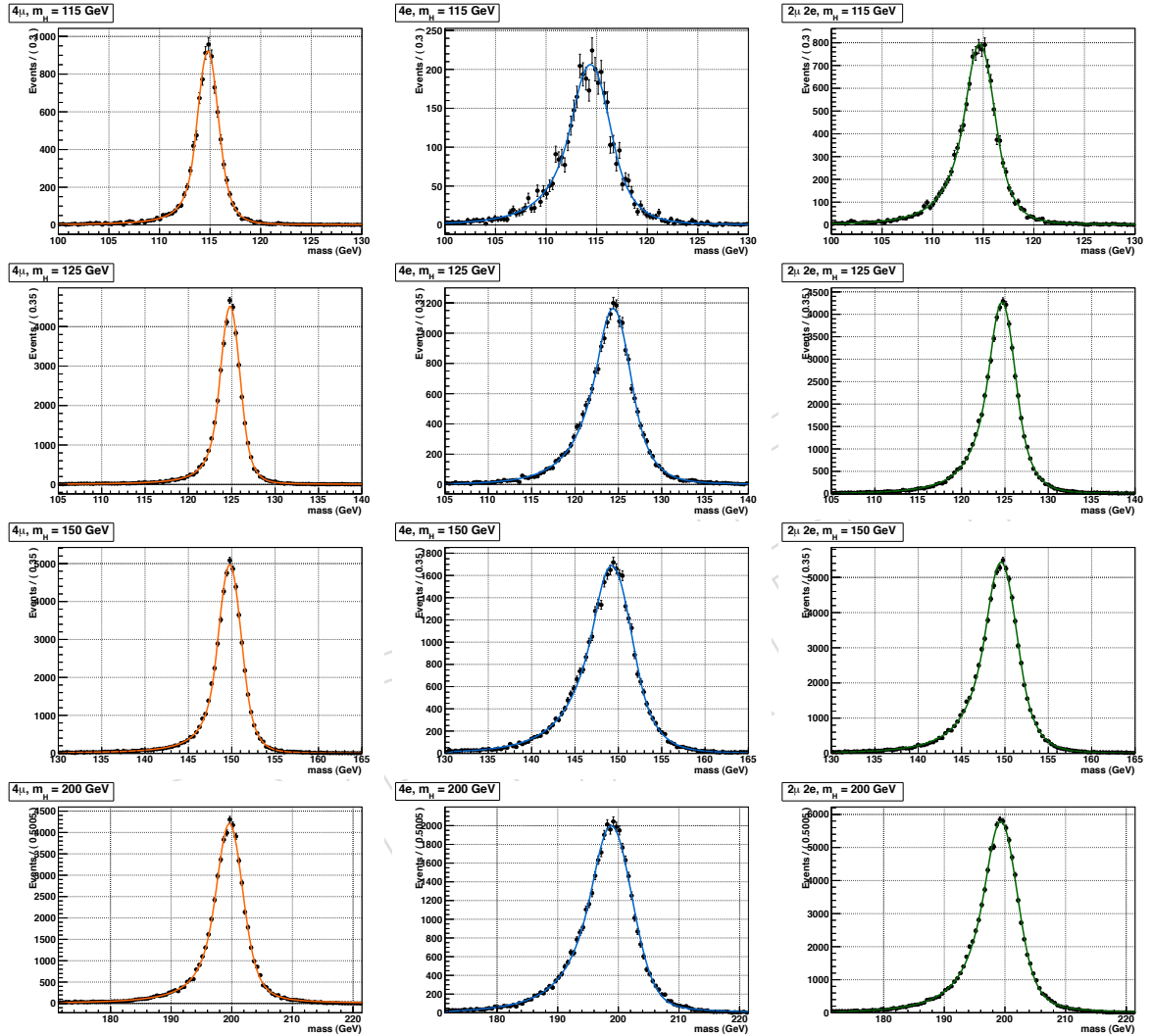


Figure 5.2: Probability density functions $f(m_{4\ell}, m_H)$ for the Higgs boson mass in the low/intermediate region at the reconstruction level after the full lepton and event selections are applied. The distributions obtained from 8 TeV MC samples are fitted with the model described in the text for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events.

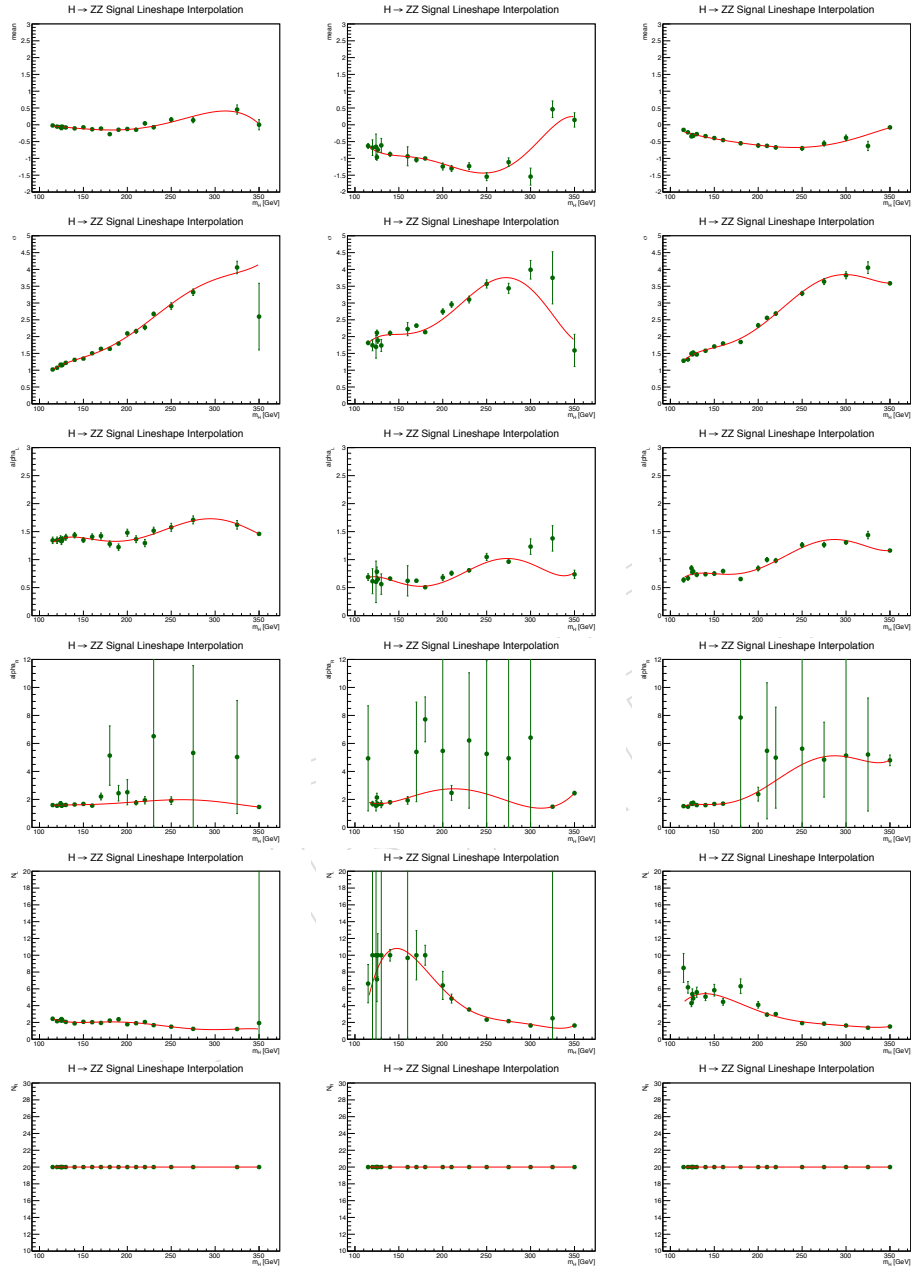


Figure 5.3: Linear and constant fits of the 4 parameters describing the signal $f(m_{4\ell}, m_H)$ pdf as a function of m_H in low/intermediate mass range. The pdf is modelled as a double crystal-ball function convoluted with the relativistic Breit-Wigner function described in the text. From the first row to the last one, σ , α_1 , α_2 , n_1 and n_2 parameters of the Crystal-Ball, are shown respectively for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events simulated for $\sqrt{s} = 7$ TeV.

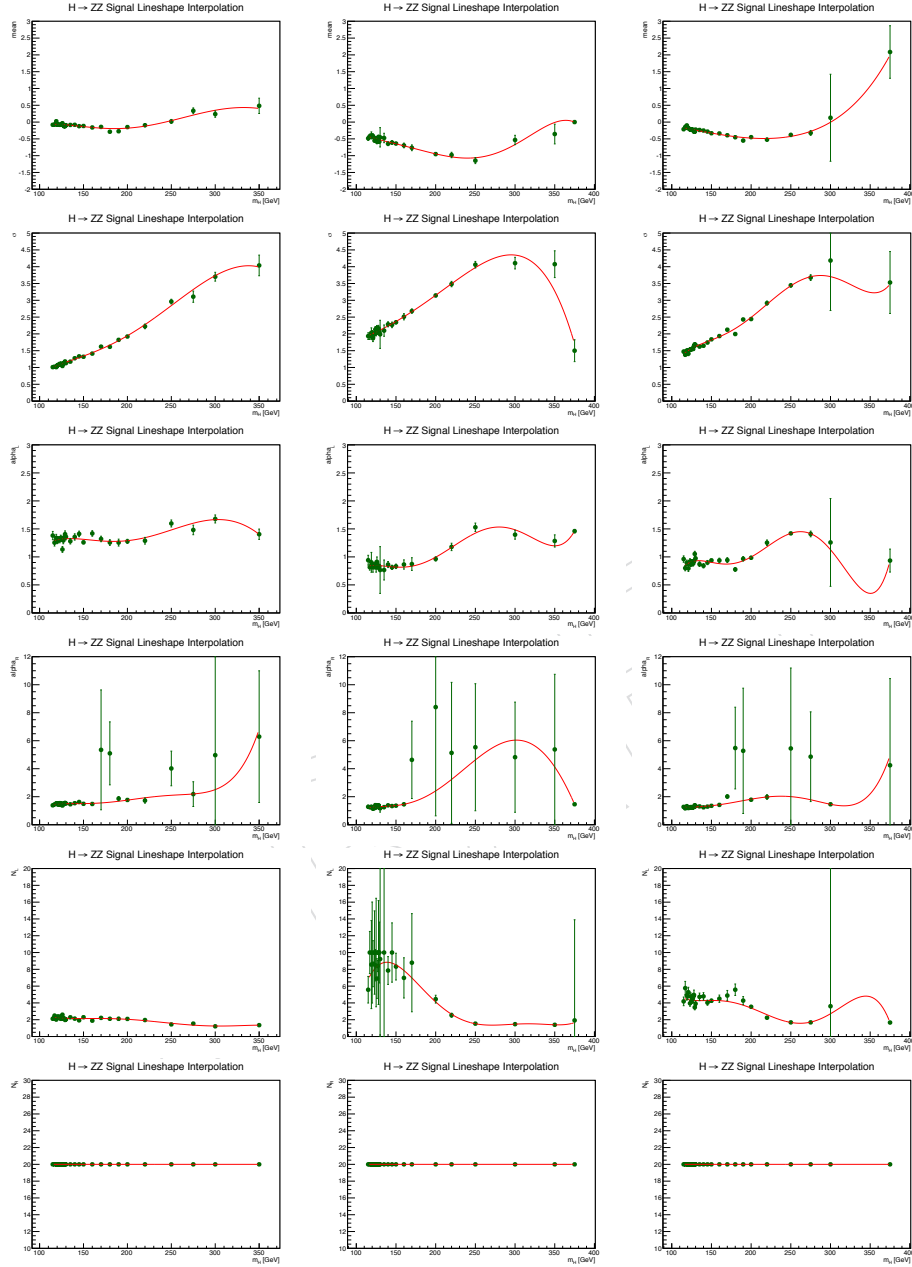


Figure 5.4: Linear and constant fits of the 4 parameters describing the signal $f(m_{A\ell}, m_H)$ pdf as a function of m_H in low/intermediate mass range. The pdf is modelled as a double crystal-ball function convoluted with the relativistic Breit-Wigner function described in the text. From the first row to the last one, σ , α_1 , α_2 , n_1 and n_2 parameters of the Crystal-Ball, are shown respectively for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events simulated for $\sqrt{s} = 8$ TeV.

The good fits obtained for all data samples assures us full appropriateness of the PDF used. On the other hand the parameters interpolation is difficult because the parameters of the Double Crystal Ball functions are highly correlated with each-others, and for several mass values a good fit can be achieved by multiple combinations of parameter values. This results in large uncertainty of the individual parameters. Either way, it was verified that the signal models obtained with the parameters given by the interpolation reproduce correctly the original distributions.

5.2 High mass range

In the high mass range the Higgs width is larger than the mass resolution. For this reason the sensitivity of the fit to the mass resolution is very low and the shape parametrization will be controlled by the theoretical function. The signal $f(m_{4\ell}, m_H)$ pdf, as in the low mass region, is modelled as the convolution function of the relativistic Breit-Wigner $pdf_1(m_H^*, m_H)$ and a Double Crystal Ball. However, instead of using the typical form, a modified version of the Breit-Wigner is used:

$$f(m_{4\ell}, m_H) = \frac{m_{4\ell}}{(m_{4\ell}^2 - m_H^2)^2 + m_{4\ell}^2 \cdot \Gamma^2}. \quad (5.3)$$

This function has the peculiarity that the Γ parameter is initialized at the theoretical value but is then left to float in the fit.

In this way, the additional degree of freedom with respect to the low mass case provided by letting the Gamma float in the fit procedure helps to obtain a good fit quality. Moreover, at high mass the lineshape corrections to the theoretical width described in Section 1.4.2 have an associated systematic uncertainty, that can be associated to the Γ parameter. The topic will be treated in Section 5.3.2.

In order to get a smoothly (and monotonically) varying function of the Γ of the Breit Wigner and σ_{DCB} , a constrained likelihood fit on the signal Monte Carlo events is performed. This constriction is achieved assuming that the physical width of the Higgs for $m_H > 400$ GeV is larger than the experimental resolution which is regulated by σ_{DCB} . Figures 5.5 and 5.6 shows the fits for some of the high Higgs mass points. The parametrization of the signal mass distributions is obtained as in the previous case. The

interpolation of the parameters are shown in Figures 5.7 and 5.8 for 7 TeV and 5.9 and 5.8 for 8TeV.

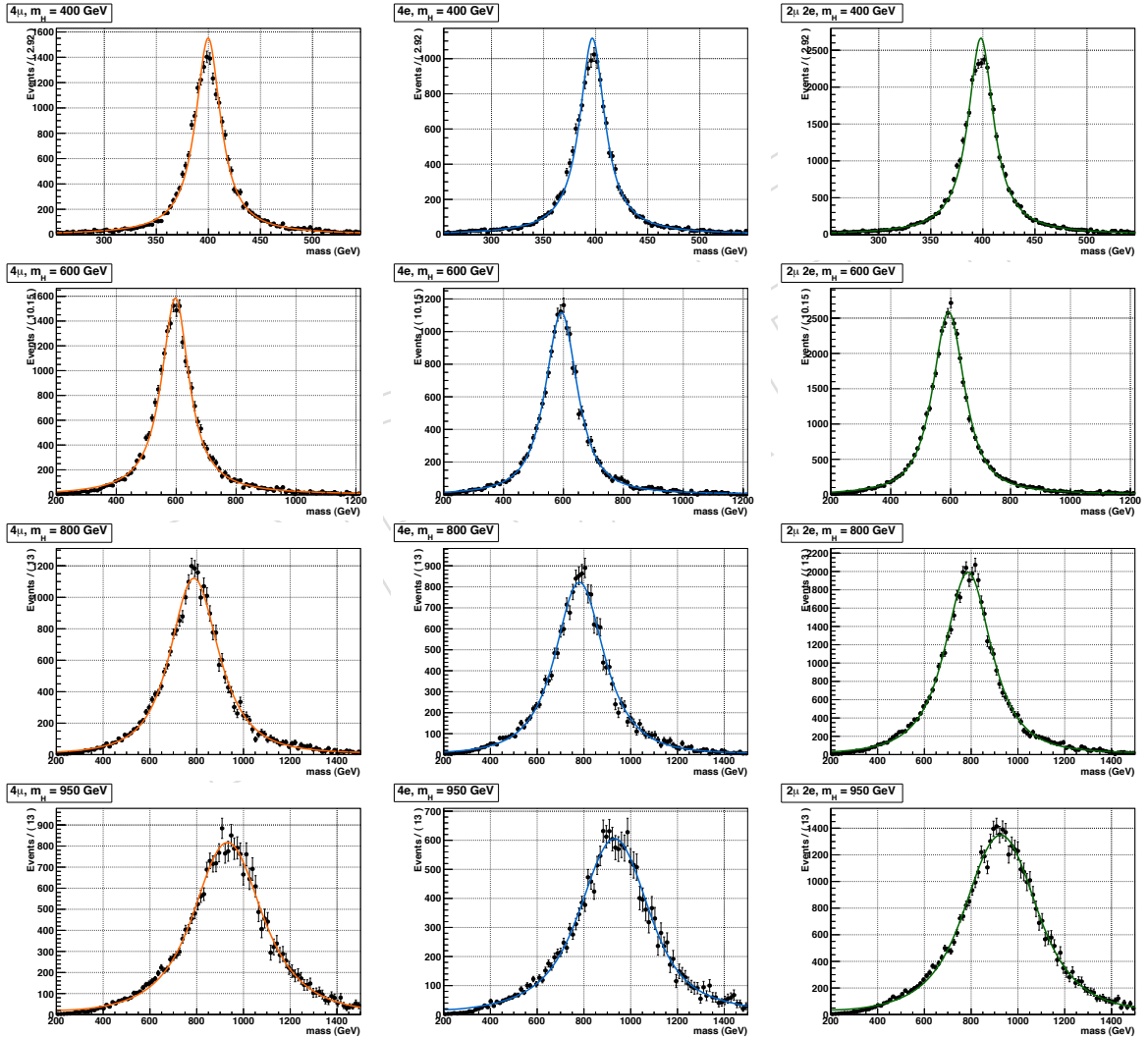


Figure 5.5: Probability density functions $f(m_{4\ell}, m_H)$ for the Higgs boson mass in the high mass region at the reconstruction level after the full lepton and event selections are applied. The distributions obtained from 7 TeV MC samples are fitted with the model described in the text for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events.

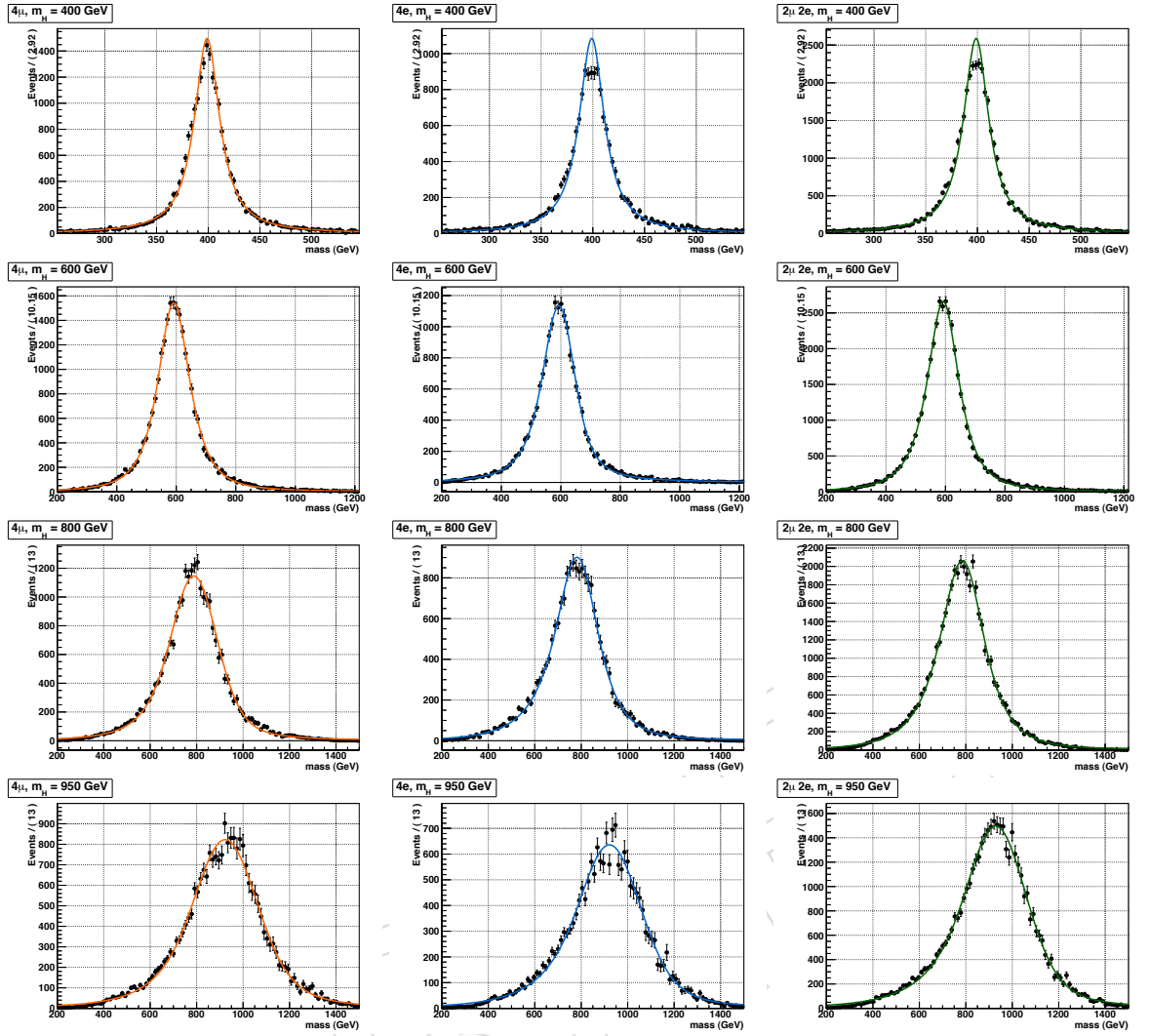


Figure 5.6: Probability density functions $f(m_{\Delta\ell}, m_H)$ for the Higgs boson mass in the high mass region at the reconstruction level after the full lepton and event selections are applied. The distributions obtained from 8 TeV MC samples are fitted with the model described in the text for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events.

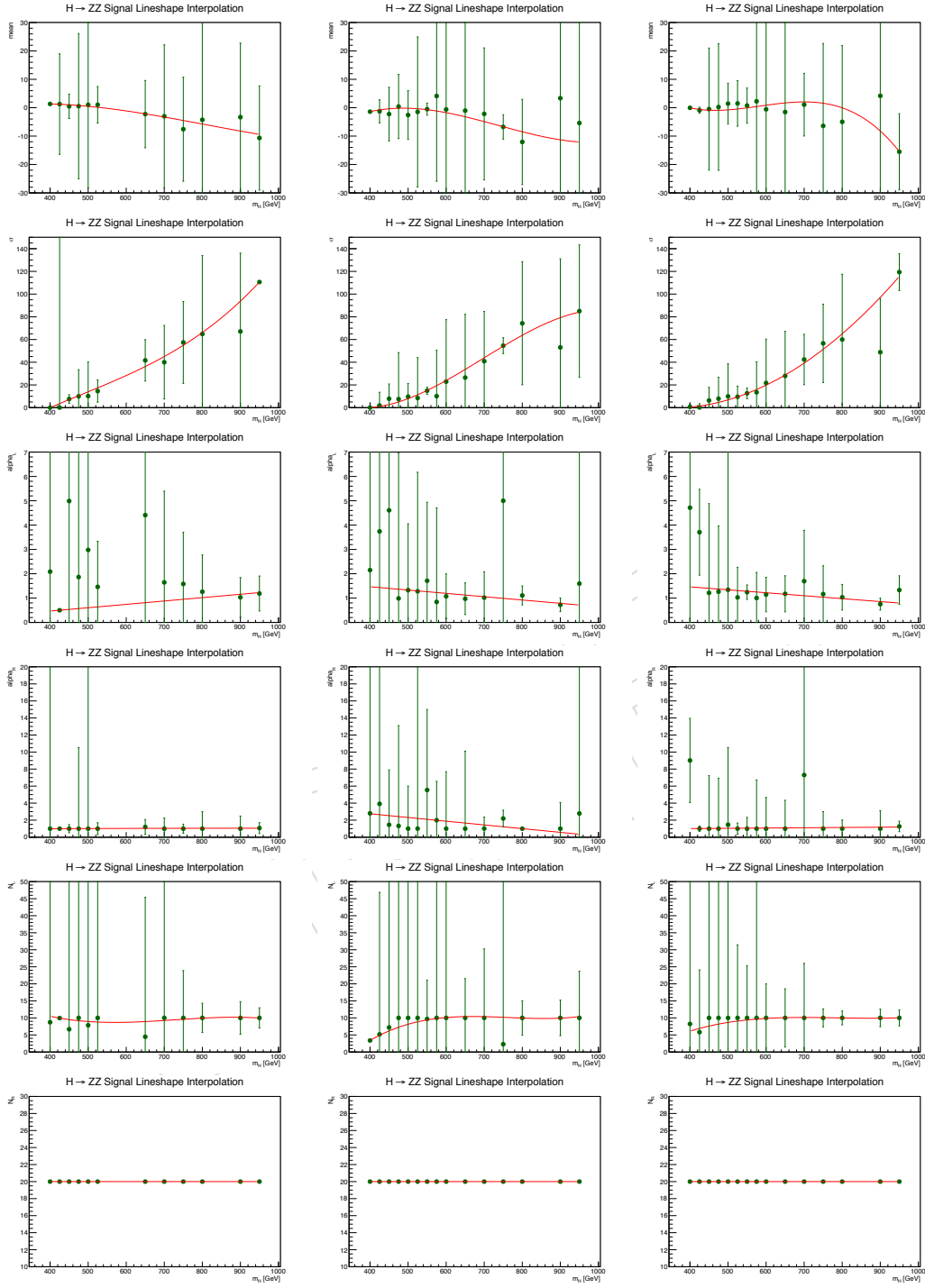
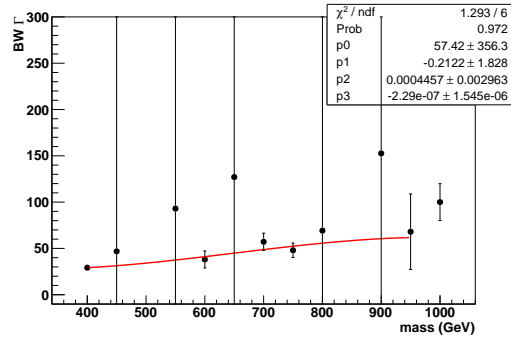
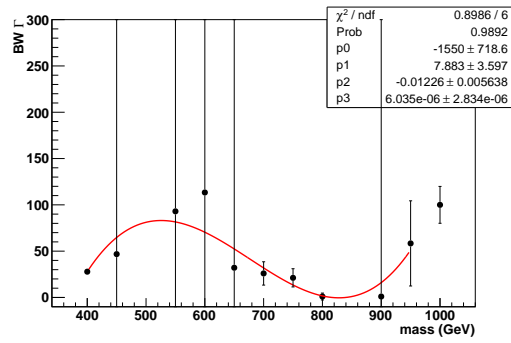


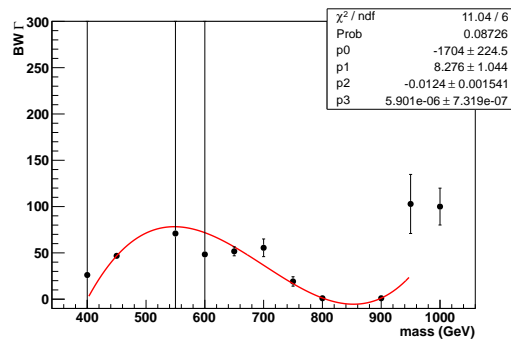
Figure 5.7: Linear and constant fits of the 4 parameters describing the signal $f(m_{4\ell}, m_H)$ pdf as a function of m_H high mass range. The pdf is modelled as a double crystal-ball function convoluted with the relativistic Breit-Wigner function described in the text. From the first row to the last one, the crystal-ball's mean, σ , α_1 , α_2 , n_1 and $4n_2$ parameters are shown respectively for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events simulated for $\sqrt{s} = 7$ TeV.



(a)



(b)



(c)

Figure 5.8: Linear fit of the Γ parameter describing the signal $f(m_{A\ell}, m_H)$ pdf as a function of m_H high mass range in the (a) $4\mu, 4e$ and (c) $2/\mu 2e$ for 7 TeV. The pdf is modelled as a double crystal-ball function convoluted with the relativistic Breit-Wigner function described in the text.

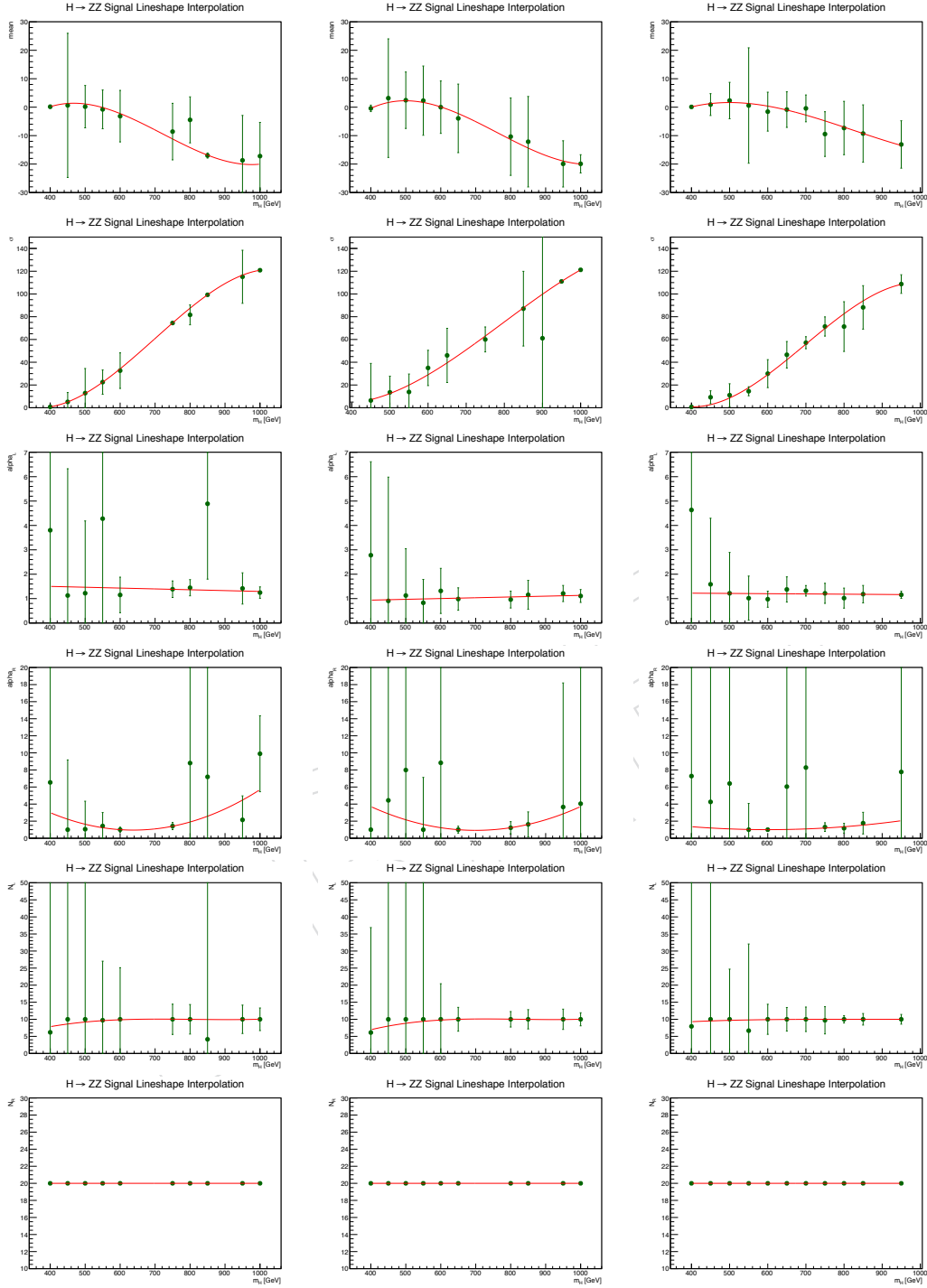
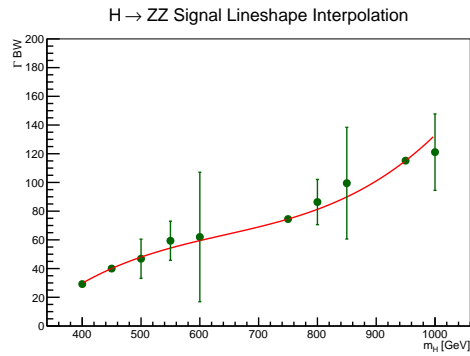
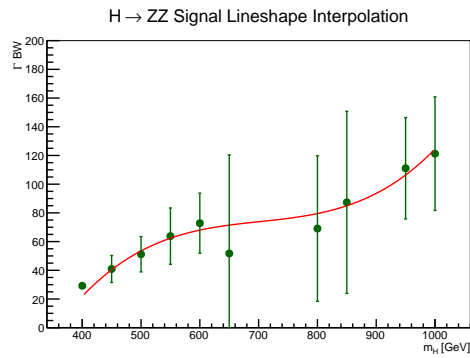


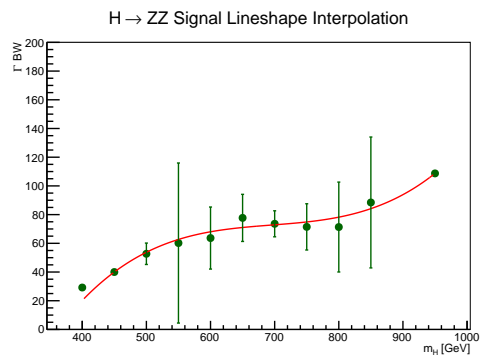
Figure 5.9: Linear and constant fits of the 4 parameters describing the signal $f(m_{4\ell}, m_H)$ pdf as a function of m_H in high mass range. The pdf is modelled as a double crystal-ball function convoluted with the relativistic Breit-Wigner function described in the text. From the first row to the last one, the crystal-ball's mean, σ , α_1 , α_2 , n_1 and $4n_2$ parameters are shown respectively for 4μ (left), $4e$ (center) and $2\mu 2e$ (right) events simulated for $\sqrt{s} = 8$ TeV.



(a)



(b)



(c)

Figure 5.10: Linear fit of the Γ parameter describing the signal $f(m_{4\ell}, m_H)$ pdf as a function of m_H high mass range in the (a) $4\mu, 4e$ and (c) $2/\mu\nu 2e$ for 8 TeV. The pdf is modelled as a double crystal-ball function convoluted with the relativistic Breit-Wigner function described in the text.

As in the low/intermediate mass 5.1, the parameter interpolation is made difficult by the large correlations between parameters. As explained above, the sensitivity of the fit to the mass resolution is very low and the shape parametrization should be regulated essentially by the Γ parameter of the Breit-Wigner. Moreover, in addition to the parameters correlation described for low/intermediate mass range, the correlation between Γ and σ of the Double Crystal Ball is ~ 1 . This further problem in the fit procedure increases the probability to obtain good fits for several parameter values combinations raising the magnitude of the errors and the fluctuations of the fit parameters. Also in this case the goodness of signal models obtained using the parameters given by the interpolation has been checked.

5.3 Signal model uncertainties

Systematic uncertainties on signal model can be factorized into (a) those affecting the overall event yield and (b) uncertainties affecting the shape of signal event distributions.

The uncertainties affecting the overall event yields are as follows:

- Theoretical total cross section uncertainties; Systematic errors on the signal total cross section for each production mechanism and for all Higgs boson masses are fully defined [16].
- Theoretical uncertainties on the $H \rightarrow ZZ$ branching fraction; The uncertainty on $BR(H \rightarrow 4\ell)$ is taken to be 2% [16] and assumed to be m_H -independent.
- Uncertainties on the efficiency scale factors discussed in 4.2;

The only systematic uncertainties on the line shape taken into account is the theoretical uncertainty on the reweighting at high mass.

5.3.1 Efficiency scale factors uncertainties

In Section 4.2 the method used to evaluate the the efficiency scale factors it has been shown. The related systematic uncertainty is calculated as follows:

- For every bin of η, p_t the weight value is floated in accordance with the corresponding error assuming a Gaussian Probability Density

Function (PDF). A new set of data/MC efficiency corrections is thus produced.

- The total normalization is recalculated with the new set of weights.
- The procedure is iterated 1000 times.
- The distribution of the normalization values of the 1000 iterations are plotted.
- Finally the RMS of the normalization distribution is considered as systematic uncertainty.

The resulting systematics are reported in the Figures 5.11 and 5.12 for the 7 and 8 TeV analysis respectively.

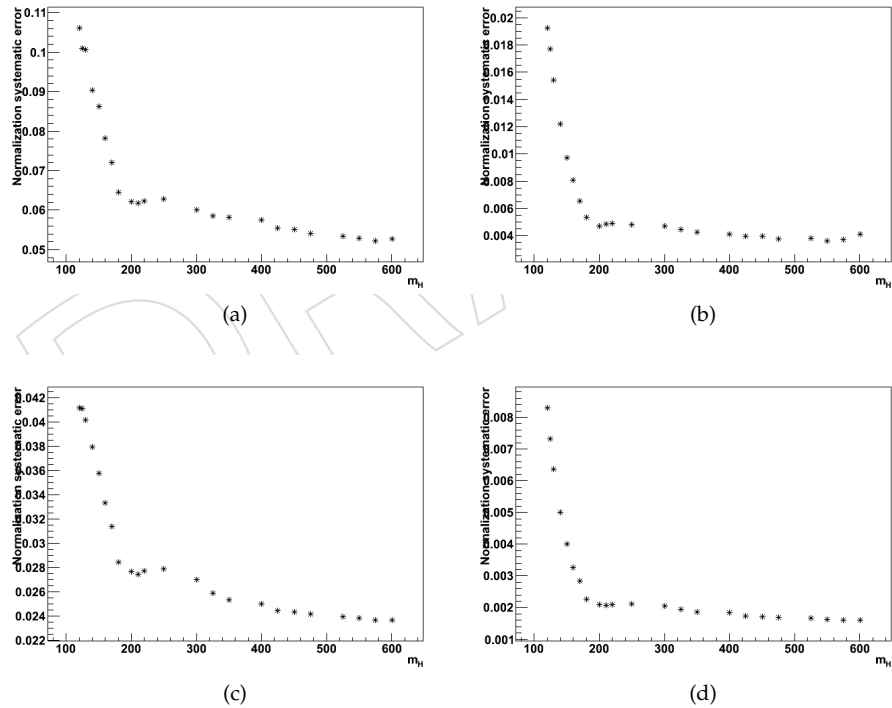


Figure 5.11: Uncertainties related to data/MC efficiency corrections for reconstruction, identification, isolation and $|SIP|$ selections, as a function of m_H , for (a) 4e channel, (b) 4 μ channel (c) 2e2 μ channel (electron only uncertainties), (d) 2e2 μ channel (muon only uncertainties). Results are for 7 TeV data.

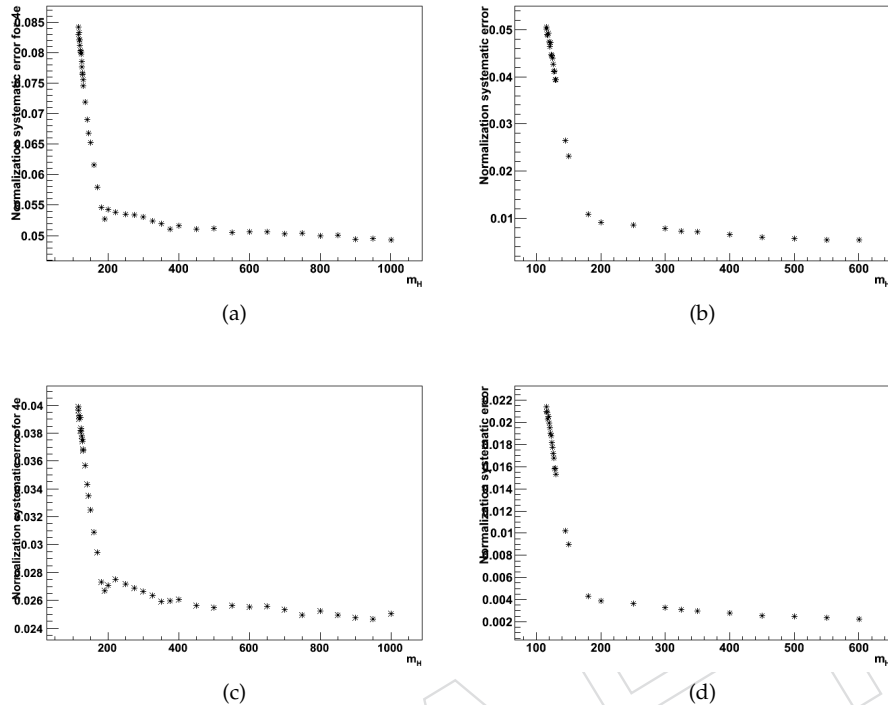


Figure 5.12: Uncertainties related to data/MC efficiency corrections for reconstruction, identification, isolation and $|SIP|$ selections, as a function of m_H , for (a) $4e$ channel, (b) 4μ channel (c) $2e2\mu$ channel (electron only uncertainties), (d) $2e2\mu$ channel (muon only uncertainties). Results are for 8 TeV data.

5.3.2 Evaluation of the high-mass corrections systematic uncertainties

Figure 4.10 shows the size of the uncertainties on the shape given by the high mass corrections described in Section 4.3. In order to propagate this systematic effect on the UL and p-value calculations, we refit the signal shape function with the two alternative hypotheses. In this fit, all the parameters describing the signal PDF are fixed to the values obtained from the nominal fit, except for the Γ of the Breit-Wigner function. In this way it is possible to propagate the uncertainty on the high-mass corrected shapes to the parameter representing the width of the theoretical PDF. Once the fits on the two alternative hypotheses are performed, the difference is calculated between the nominal value of Γ and the value determined by the

alternative fits, and the largest variation between the two is considered as a systematic uncertainty. This procedure is performed for all the mass points. It has been found a systematic effect on Γ between 3% and 5% among the whole spectrum. A systematic error of 5% is chosen for any Higgs mass hypothesis.

5.4 Overview of the statistical method

In section 3.7 an overview of the entire procedure for computing upper-limit and p-values has been shown. The aim of the modelization given above, as refers to this procedure, is to allow the construction of a test statistic for any mass value within the studied range. Indeed, for the test statistic construction the signal and background expectations in the form of event yields is needed. Without the parameterization, defining a test statistic would be possible only for the limited number of MC samples produced for this analysis.

The statistical approach used in the study for the search Higgs boson is structured as follows.

Firstly a signal strength modifier $\mu = \sigma/\sigma_{SM}$ is defined, where σ is the "true" cross section and σ_{SM} is the predicted SM cross section. Then the test statistic q_μ , as the ratio of likelihoods, is defined by:

$$q_\mu = -2\ln \frac{\mathcal{L}(data|\mu s + b)}{\mathcal{L}(data|\hat{\mu} s + b)} \quad (5.4)$$

where s and b are the yields for signal and background respectively and $\hat{\mu}$ is the value that maximizes the likelihood $\mathcal{L}(data|\hat{\mu} s + b)$.

$\mathcal{L}(data|rate)$ is simply a product of Poisson probabilities for number of either observed or simulated events in each sub- channel, given the expected signal and background rates that are evaluated through the modelization of the signal and the background parameterization.

In order to take into account the systematics, each independent source of systematic uncertainty is assigned a nuisance parameter θ_i . The expected Higgs boson and background yields become functions of these nuisance parameters, allowing modifications to the test statistic itself and/or the way pseudo-data are generated. A detailed description can be found elsewhere [30].

Limits

Having defined the test statistic, one constructs PDFs of test statistic q under the signal+background and the sole background hypothesis by means of "tossing" toy pseudo-observations according to the very same Poisson probabilities. Using these PDFs, one can then evaluate the probability $P(q_\mu \geq q_\mu^{data} | \mu s + b)$ for the observed value q^{data} to be as or less compatible with the background+signal hypothesis and the probability $P(q_\mu \geq q_\mu^{data} | b)$ for the background hypothesis only. Then the quantity CL_s is calculated as the ratio of these two probabilities:

$$CL_s = \frac{CL_{s+b}}{CL_b}. \quad (5.5)$$

In the study of the search for the Higgs boson CL_s is required to be less than or equal to 0.05, in order to declare the 95% C.L. exclusion.

P-value

In the case of observing an excess of events, its characterization begins with evaluating the p-value of the upward fluctuation of the background-only hypothesis. This can be done by "tossing" background-only pseudo-data and building up the corresponding PDF for the test statistic

$$q_0 = -2 \ln \frac{\mathcal{L}(data|b)}{\mathcal{L}(data|\hat{\mu}s + b)}. \quad (5.6)$$

The p-value, i.e. the probability of getting an observation as or less compatible as seen in data for the background-only hypothesis, is then defined as:

$$p_0 = P(q_0 \geq q_0^{obs} | b). \quad (5.7)$$

Conclusions

In this thesis I presented my work in the CMS experiment for the search of the Standard Model Higgs boson produced in pp collisions and decaying in the final state $ZZ^* \rightarrow 4\ell$ with $\ell = e, \mu$. Since the signature of the four final state leptons is very clear, this channel was expected to be one of the main decay channels for the Higgs boson discovery. In July, 4th, the discovery of a Higgs boson-like particle has been announced by the CMS and ATLAS experiments around a mass of 125 GeV, where the $H \rightarrow ZZ \rightarrow 4\ell$ analysis has made a substantial contribution.

In this thesis, I first discuss a set of corrections I had to apply to Monte Carlo samples. A first correction is necessary to match the distribution of number of pileup events different in Monte Carlo and data.

A second correction is necessary for a discrepancy in the lepton reconstruction, trigger and identification efficiencies, as well as lepton isolation and significance of the impact parameter efficiencies. This correction leads to an increase in the expected yields of about 2%, 3% and 5% for 4μ , $2\mu 2e$ and $4e$ channels respectively as well as a small change in the signal line shape. Then, by propagating the uncertainties on these efficiency corrections to the event weights, I estimated the uncertainties related to efficiency corrections.

I also took care of applying an additional reweighting to correct the line-shape at high-mass. This is necessary since simulated samples have been made using a narrow-width approximation describing the Higgs lineshape with a Breit-Wigner distribution, which breaks down at high Higgs masses (typically above 400 GeV). Moreover at high Higgs masses the interference between the Higgs signal and the $gg \rightarrow ZZ$ background becomes very large. Recent theoretical studies provide a recipe to correct Monte Carlo samples and to estimate the uncertainty due to missing higher perturbative order on the interference.

Once these corrections were applied, I could proceed with a study on the modelization of the signal line shape. Since a limited numbers of Monte Carlo samples were produced, the aim of my work was to develop a signal model parametrization to allow the statistical analysis of the $H \rightarrow 4\ell$ data for any mass value within the studied range.

I modelled the signal shape using the convolution of a relativistic Breit-Wigner as theoretical distribution and a resolution function in order to take into account the detector response. Two different models have been used for low/intermediate mass range and for high mass range. In the latter one, it was necessary to modify the the Breit-Wigner function to allow a correct modelling of the theoretical line shape with the corrections described above. This choice allowed me to evaluate the related systematic uncertainty on the Higgs Γ that has been found to be of the order $\sim 5\%$.

Bibliography

- [1] S. L. Glashow, "Partial Symmetries of Weak Interactions", *Nucl. Phys.*, vol. 22, pp. 579-588, 1961.
- [2] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264-1266, 1967.
- [3] A. Salam, "Elementary Particle Physics," N. Svartholm ed., Almquist and Wiksell, Stockholm, 1968.
- [4] Gell-Mann M. *Phys. Lett.* V.8. P.214-215, 1964.
- [5] Fritzsche H., Gell-Mann M., and Leutwyler H. *Phys. Lett.* V.B47. P.365-368, 1973.
- [6] Gross D. J. and Wilczek F. *Phys. Rev. Lett.* V.30. P.1343-1346, 1973.
- [7] S. Dawson, "Introduction to electroweak symmetry breaking," 1998. hep-ph/9901280.
- [8] The LEP Working Group for Higgs boson searches, "Search for the standard model Higgs boson at LEP ", *Phys. Lett.*, vol. B565, pp. 61-75, 2003.
- [9] CDF and $D\otimes$ Collaborations, "Combined CDF and $D\otimes$ Upper Limits on Standard Model Higgs-Boson Production with up to 8.6 fb⁻¹ of Data", 2011. *arXiv:1107.5518v2*.
- [10] The LEP Collaborations, the LEP Electroweak Working Group, "A Combination of preliminary electroweak measurements and constraints on the standard model", 2006. *arXiv:0612034 [hep-ex]*.

- [11] The LEP Electroweak Working Group.
http://lepewwg.web.cern.ch/LEPEWWG/.
- [12] ATLAS collaboration, "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC" CERN-PH-EP-2012-218, 2012 *arXiv:1207.7214*
- [13] Wu, Xin, Clark, Allan, Campanelli, Mario, "Effective K-factors: a method to include higher order QCD corrections in parton shower Monte Carlos" 2006-01-01 336-337 *http://dx.doi.org/10.1007/978-3-540-32841-4_63*
- [14] CMS Collaboration "Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks" CMS PAS HIG-12-044 *http://cms-physics.web.cern.ch/cms-physics/public/HIG-12-044-pas.pdf*
- [15] INSTITUTE OF PHYSICS PUBLISHING AND SISSA , THE CERN LARGE HADRON COLLIDER: ACCELERATOR AND EXPERIMENTS, "The CMS experiment at the CERN LHC".
- [16] 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]*.
- [17] A. Djouadi, "The anatomy of electro-weak symmetry breaking", *Phys.Rept.*, 459:1-241, 2008.
- [18] CMS Collaboration, Search for the standard model Higgs boson decaying to a W pair in the fully leptonic final state in pp collisions at $\sqrt{s} = 7$ TeV, CMS-HIG-11-024; CERN-PH-EP-2012-018 *arXiv:1202.1489v1*.
- [19] CMS Collaboration. The Muon Project: Technical Design Report. CERN/LHCC 97-32, (1997).
- [20] CMS Collaboration, "CMS Luminosity" *https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults2011*
- [21] CMS Collaboration Collaboration, "Search for a Higgs boson in the decay channel $H \rightarrow ZZ(*) \rightarrow q\bar{q}\ell^-\ell^+$ in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* 1204 (2012) 036, *arXiv:1202.1416*.

- [22] Y. Gao, A. V. Gritsan, Z. Guo et al., "Spin determination of single-produced resonances at hadron colliders", *Phys.Rev. D*81 (2010) 075022, *doi:10.1103/PhysRevD.81.075022*, *arXiv:1001.3396*.
- [23] The CMS Collaboration, "Search for the standard model Higgs boson in the decay channel $H \rightarrow ZZ \rightarrow 4\ell$ in pp collisions", CMS NOTE AN-12-367.
- [24] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* 11 (2007) 070, *doi:10.1088/1126-6708/2007/11/070*, *arXiv:0709.2092*.
- [25] J. Alwall et al., "MadGraph/MadEvent v4: The New Web Generation", *JHEP* 09 (2007) 028, *doi:10.1088/1126-6708/2007/09/028*, *arXiv:0706.2334*.
- [26] T.Sjöstrand, S.Mrenna, and P.Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP*05(2006)026, *doi:10.1088/1126-6708/2006/05/026*, *arXiv:hep-ph/0603175*.
- [27] Y. Gao, A. V. Gritsan, Z. Guo et al., "Spin determination of single-produced resonances at hadron colliders", *Phys.Rev. D*81 (2010) 075022, *doi:10.1103/PhysRevD.81.075022*, *arXiv:1001.3396*.
- [28] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* 716 (2012) 30-61, *doi:10.1016/j.physletb.2012.08.021*, *arXiv:1207.7235*.
- [29] CMS Pileup Reweighting, July 26th, 2011, <http://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities>
- [30] ATLAS and CMS Collaborations, LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011", ATL-PHYS-PUB/CMS NOTE 2011-11, 2011/005, (2011).
- [31] A. Read, "Modified frequentist analysis of search results (the CLs method)", Technical Report CERN-OPEN-2000-005, CERN, (2000).
- [32] S. Gorla, G. Passarino and D. Rosco, "The Higgs Boson Lineshape " *arXiv:1112.5517 [hep-ph]*.

- [33] G.Passarino,"Higgs Interference Effects in gg-ZZ and their Uncertainty," *arXiv:1206.3824[hep-ph]*.

Acknowledgements

Mille grazie a:

Mario, Nicola e Chiara, per avermi dato l'opportunità di partecipare all'esperimento CMS in questo splendido periodo e perché è davvero bello lavorare in gruppo con voi; in particolare **Mario** per avermi insegnato tanto e **Chiara** per essere sempre stata così solare;

Martina e Linda, per tutto quello che abbiamo condiviso a Ginevra e perché, con l'aiuto di Dilip, siamo diventati un'ottima squadra;

Dilip, perché ci ha fatto imparare lo spirito di squadra e ci ha fatto capire che certe domande non hanno un risposta; **Ignazio e Rita (I miei genitori)**, per avere sempre creduto in me anche negli anni più bui della mia carriera scolastica;

Anna, perché nonostante tutto è riuscita a sopportarmi ed amarmi tutto questo tempo. Questa tesi è dedicata a te;

Violetta, per le mille ore di studio passate assieme e per essere stata una parte fondamentale della mia carriera scolastica;

Alessio, per aver intrapreso assieme questa avventura a Torino;

Irene, Charlotte e Fabio, per aver condiviso la casa con un tipo strano...;

Milo, per essere diventato in poco tempo un vero amico;

Stefano, per avermi accolto subito al mio arrivo;

Villo, Ilaria, Nonno, Elisa, Rocco e tutti gli amici Torinesi perché gli ho trascurati per colpa di questa tesi ma mi vogliono bene comunque;

Keko, Fubu, Dario, Robo, Piras, Prune, Marra, Ire, Geko ecc.. perché sono gli amici di una vita con cui sono cresciuto, maturato ecc..; in particolare **Marra** per avermi aiutato con l'introduzione in limba;

Giammino perché, anche se lontano, mi è stato molto vicino ed è stato un esempio;

Rita per essere stata una seconda mamma nonché amica;

Il bosone di Higgs per avermi aspettato prima di uscire allo scoperto;

