Search for the Standard Model Higgs boson
in the $H \rightarrow WW \rightarrow l\nu q\bar{q}$ decay channel
with the CMS detector at the LHC
# Contents

1 Introduction

1.1 The Standard Model ................................................. 1
  1.1.1 The non-interacting theory .................................... 1
  1.1.2 Gauge invariance principle and particle interactions ......... 3
  1.1.3 Spontaneous symmetry breaking and Higgs mechanism .......... 6

1.2 The search for the Higgs boson .................................... 12
  1.2.1 Theoretical constraints ....................................... 12
  1.2.2 Direct constraints from LEP and Tevatron .................... 14

1.3 Higgs boson searches at the LHC .................................. 17
  1.3.1 Higgs production mechanisms .................................. 17
  1.3.2 Higgs decay modes ............................................ 18

2 The CMS detector at the LHC

2.1 The Large Hadron Collider ....................................... 21

2.2 The CMS detector ................................................... 22
  2.2.1 The silicon tracker .............................................. 24
  2.2.2 The electromagnetic calorimeter - ECAL ....................... 25
  2.2.3 The hadronic calorimeter - HCAL ............................. 27
  2.2.4 The muon system ............................................... 28
  2.2.5 The CMS trigger ................................................. 29

2.3 Object reconstruction at CMS .................................... 30
  2.3.1 Electron reconstruction ........................................ 30
  2.3.2 Muon reconstruction ............................................ 31
  2.3.3 Particle Flow reconstruction for jets and missing transverse energy . . 32

2.4 The 2011 proton-proton run ..................................... 36

3 Description of the experimental final state

3.1 Overview of the channel .......................................... 39

3.2 Signal and backgrounds ........................................... 40
  3.2.1 Standard Model Higgs signal .................................. 42
  3.2.2 Backgrounds ................................................... 43

3.3 Datasets ............................................................. 45
  3.3.1 Data samples ................................................... 45
  3.3.2 Monte Carlo samples ............................................ 46
Chapter 1

Introduction

In this first chapter, a brief introduction to the Standard Model is given, together with the description of the electroweak symmetry breaking mechanism introduced by Brout, Englert and Higgs to explain the origin of the masses of the weak bosons and of the fundamental fermions. The theory does not predict the mass of the quantum associated with the Higgs field (the so-called Higgs boson) and therefore its value must be determined experimentally. Constraints on the Higgs mass, arising both from theoretical considerations and experimental searches at LEP and Tevatron, can nevertheless be set and are presented in section 1.2, updated to the summer/winter of 2011.

An overview of the Higgs production and decay mechanisms at the Large Hadron Collider (LHC), as well as the main search strategies, is presented in section 1.3, together with the preliminary results shown during summer 2011.

1.1 The Standard Model

1.1.1 The non-interacting theory

The Standard Model of particle physics is the theory which currently better describes the electromagnetic, weak and strong interactions between fundamental particles [1, 2]. It is a relativistic quantum field theory based on the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry group; it has a perturbative behaviour at high energies and it is renormalizable [3, 4]. The Standard Model includes 12 elementary particles of spin $\frac{1}{2}$ known as fermions, each one with the corresponding antiparticle having opposite quantum numbers and the same couplings. The fundamental fermions are subdivided in two groups: six leptons and six quarks, with each group further divided into three families or generations, which exhibit a similar physics behaviour (see table 1.1).

The three lepton families are the electron ($e$), the muon ($\mu$) and the tau ($\tau$), each one with the associated neutrino ($\nu_e$, $\nu_\mu$, $\nu_\tau$).

The six quarks are labelled: up ($u$), down ($d$), charm ($c$), strange ($s$), top ($t$) and bottom ($b$). Quarks carry colour charge and therefore interact via the strong force. A phenomenon called “colour confinement” results in quarks being bound to one another, forming colour-neutral composite particles named hadrons, containing either a quark and an antiquark $q\bar{q}$ (mesons) or three quarks $qqq''$ (baryons). The quantum field operators associated with fermions are four-components Dirac spinors, which in the following will be denoted as $\psi$. Introducing the Weyl
1. Introduction

| 1st generation | 2nd generation | 3rd generation | \( Q \)/|e| |
|----------------|----------------|----------------|----------|
| Leptons        | ν\(_e\) 511 keV/c\(^2\) | ν\(\mu\) 105.7 MeV/c\(^2\) | ν\(\tau\) 1.777 GeV/c\(^2\) | -1 |
|                | \( \sim 0 \)                      | \( \sim 0 \)                      | \( \sim 0 \)                      | 0   |
| Quarks         | u 1.7 - 3.1 MeV/c\(^2\)          | c 1.29\(\pm 0.05\) GeV/c\(^2\) | t 172.9\(\pm 1.1 - 1.1\) GeV/c\(^2\) | 2/3 |
|                | d 4.1 - 5.7 MeV/c\(^2\)          | s 100\(\pm 30\) - 20 MeV/c\(^2\) | b 4.19\(\pm 0.18 - 0.06\) GeV/c\(^2\) | -1/3 |

Table 1.1: Fundamental spin-\(\frac{1}{2}\) fermions, with their respective mass and charge.

representation of the \(\gamma\) matrices:

\[
\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}
\]

where the \(\sigma^i\) are the 2\(\times\)2 Pauli matrices, the adjoint spinor can be built as \(\bar{\psi} = \psi\dagger\gamma^0\), with the dagged notation \(\dagger\) representing the conjugate transpose. For a fermion of mass \(m\), the free Lagrangian (i.e. without introducing any kind of interaction) is:

\[
L_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi
\]

from which the following equation of motion (Dirac equation) can be derived:

\[
(i\gamma^\mu\partial_\mu - m)\psi = 0
\]

For convenience, the Dirac spinor \(\psi\) is usually separated into the left-handed and the right-handed spinors \(\psi = \psi_L + \psi_R\), obtained applying the projection operators \(P_L\) and \(P_R\):

\[
\psi_L = P_L\psi = \frac{1}{2}(1 - \gamma^5)\psi \quad \psi_R = P_R\psi = \frac{1}{2}(1 + \gamma^5)\psi
\]

The left and the right handedness of the spinors is called “chirality”.

Interactions between fermions are mediated by the exchange of spin-1 particles (bosons) which arise from the requirement of the invariance of the theory under so-called gauge symmetries (see section 1.1.2). The gauge boson mediators of three types of fundamental interactions out of four have been experimentally observed so far\(^1\).

The \(SU(2)_L \times U(1)_Y\) group is associated with electroweak interactions, which are the unified description of electromagnetism and weak interactions. The long-range electromagnetic interaction is mediated by the massless photon, while the short-range weak force carriers are the massive \(W^+, W^-\) and \(Z^0\) bosons. The \(SU(2)_L\) gauge bosons couple only to the left-handed components \(\psi_L\) of the fermion fields, leading to the observed parity-violating character of weak interactions. The \(U(1)_Y\) gauge boson couples to both the left-handed and the right-handed components. The left-handed projections of the fermion fields are grouped into \(SU(2)_L\) doublets:

\[
f_L = \left\{ \left( \begin{array}{c} \nu_e \\ c \end{array} \right)_L, \left( \begin{array}{c} \nu_\mu \\ \mu \end{array} \right)_L, \left( \begin{array}{c} \nu_\tau \\ \tau \end{array} \right)_L, \left( \begin{array}{c} u \\ d \end{array} \right)_L, \left( \begin{array}{c} c \\ s \end{array} \right)_L, \left( \begin{array}{c} t \\ b \end{array} \right)_L \right\}
\]

while the right-handed components are \(SU(2)_L\) singlets:

\[
f_R = \{ e_R, \mu_R, \tau_R, u_R, d_R, c_R, s_R, t_R, b_R \}
\]

\(^1\)The fourth known fundamental interaction, gravity, is not yet accommodated into a unified theory together with electromagnetic, weak and strong forces. Namely, the existence of a spin 2 “graviton” that, in the framework of QFT, should be associated with the gravitational field, is not confirmed.
To each doublet is associated a so-called “weak isospin” charge $T = \frac{1}{2}$: neutrinos and upper quarks possess a third isospin component $T_3 = \frac{1}{2}$, while $e_L$, $\mu_L$, $\tau_L$ and the down-type left quarks have component $T_3 = -\frac{1}{2}$. To the singlets $f_R$ is conventionally assigned a null weak isospin charge ($T = 0$).

The absence of right-handed neutrinos into $f_R$ will be discussed later.

The $SU(3)_C$ colour group is associated with the strong interaction between quarks, which is described by quantum chromodynamics (QCD). Each quark appears in three different “colour” states, thus belonging to a $SU(3)_C$ triplet, while leptons are colourless singlets. The quanta of the strong interaction field, called gluons, have spin 1, zero mass and carry colour charge. They can also interact among each other and this leads to the already mentioned phenomenon of quark confinement.

A free boson having mass $m$ and spin 0 is represented in QFT by a complex scalar field $\phi$, whose dynamics is described by the following Klein-Gordon Lagrangian, from which the Euler-Lagrange equation of motion can be easily extracted:

$$L_{KG} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - m^2 \phi^\dagger \phi \Rightarrow (\Box + m^2) \phi = 0$$ \hspace{1cm} (1.4)

For vector (i.e. spin-1) bosons, the associated representation is a vector field $A_\mu$, whose dynamics is described by:

$$L_{Proca} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} m^2 A_\mu A^\mu \Rightarrow \partial_\mu F^{\mu\nu} + m^2 A^\nu = 0$$ \hspace{1cm} (1.5)

where the antisymmetric Faraday tensor is defined as $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$.

Up to now, only the free non-interacting theory has been introduced. The Standard Model approach to account for interactions between particles is the requirement of local gauge invariance of the Lagrangian, as explained in the next section.

1.1.2 Gauge invariance principle and particle interactions

The requirement of a symmetry in the Lagrangian translates into a conservation of charges, via Noether’s theorem, and allows for the introduction of new fields and interactions.

In QFT it proves very convenient to require that the Lagrangian is invariant under some symmetry transformation groups. For example, in the simplest case of the Abelian symmetry group $U(1)$, the transformation is just a phase multiplication having the form $\psi \rightarrow \psi' = e^{ie\alpha} \psi$, where $e$ is some constant. If also the phase $\alpha$ is constant in time and space, we speak about a “global” phase transformation, whereas if it differs from point to point ($\alpha = \alpha(x)$) we call it a “local” phase transformation.

The Dirac Lagrangian (1.1) that describes a free fermion field is invariant under global phase transformations because, trivially, $\partial_\mu \psi' = e^{ie\alpha} \partial_\mu \psi$ and the $e^{ie\alpha}$ term cancels out the $e^{-ie\alpha}$ which comes from the adjoint spinor $\bar{\psi}$. Physically speaking, this implies that in the system there is a conserved current which, arranging the constants in the appropriate way, can be interpreted as the electromagnetic current $J_\mu^e = -e\bar{\psi} \gamma^\mu \psi$ and a conserved charge, obtained integrating the fourth component of the current on the three-dimensional space.

If the phase is made dependent from $x^\mu$, the Lagrangian picks up an extra term under the transformation and the invariance is lost. One way to recover it, is to apply a “minimal substitution”, replacing every derivative $\partial_\mu$ with the covariant derivative:

$$D_\mu = \partial_\mu - ieA_\mu$$
where the new field $A_\mu$ changes under the transformation as: $A_\mu \rightarrow A_\mu + \partial_\mu \alpha(x)$.

Therefore, the price paid in order to preserve the invariance of the Lagrangian is the introduction of a new vector field $A_\mu$, called “gauge field”, which, of course, needs its own free term $\mathcal{L} = -\frac{i}{4} F_{\mu\nu} F^{\mu\nu}$ into the full Lagrangian. Moreover, expanding the covariant derivative, it is easy to see that this field couples with the fermion field $\psi$ through a term of the form $-J^\mu_\epsilon A_\mu$.

It is important to notice that, in order not to spoil the local gauge invariance, the new field $A_\mu$ is required to be massless (that is, the term in equation (1.5) that depends from $m^2$ must be dropped).

This is exactly the quantum field description of the electromagnetic interaction (QED), with the boson $A_\mu$ identified as the photon.

More generally, let’s consider the transformation $\psi \rightarrow \psi' = U \psi$, where $U$ is a $N \times N$ unitary matrix ($U^\dagger U = 1$). Any unitary matrix can be written in the form $U = e^{iH}$, where $H$ is Hermitian ($H^\dagger = H$). Moreover, the most general Hermitian $N \times N$ matrix can be decomposed in the form $H = \theta 1 + \sum_k \alpha_k \cdot t_k$, where 1 is the unit matrix and $t_k$ are $N^2 - 1$ matrices that can be identified with the generators of the group.

In group-theoretical language, what we have just said is that $U(N) = U(1) \times SU(N)$. Having already discussed the phase transformations of the $U(1)$ group, we shall now focus on the global and local $SU(N)$ transformations.

Every element of this group can be written as:

$$S = S(\bar{\alpha}) = e^{ig\alpha_k t_k} \quad k = 1, \ldots, N^2 - 1$$

where $g$ is a constant that will determine the interaction strength of the field.

As before, it is straightforward to prove the invariance under a global $SU(N)$ transformation of any Lagrangian which depends from some field $\psi$ and from the derivatives $\partial_\mu \psi$. However, requiring the invariance to hold also locally turns out to be a little bit more subtle compared to the $U(1)$ case because, in general, the symmetry group will be non-Abelian. This means that the generators follow the non trivial commutation relations $[t_i, t_j] = i f_{ijk} t_k$, where the $f_{ijk}$ are called “structure constants” of the group.

Adapting the same idea, we might restore the invariance introducing a set of new vector fields $A^k_\mu$ and replacing the usual derivative with the covariant derivative:

$$D_\mu = \partial_\mu - igA_\mu$$

with

$$A_\mu = \sum_{k=1}^{N^2-1} t^k A^k_\mu$$

Substituting the covariant derivative into equation (1.1), the Lagrangian reads:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi + g\bar{\psi}\gamma^\mu t^k A^k_\mu \psi$$

where the last term expresses the coupling between the fermion field and the new vector fields. To $A_\mu$ (or, equivalently, to the $N^2 - 1$ gauge fields $A^k_\mu$) must be assigned a transformation rule such that $D_\mu \psi \rightarrow S(D_\mu \psi)$.

This translates in the requirement that:

$$A_\mu \rightarrow A'_\mu = SA_\mu S^{-1} - \frac{i}{g} (\partial_\mu S) S^{-1} \quad (1.6)$$

and

$$A^k_\mu \rightarrow A'^k_\mu = A^k_\mu + \partial_\mu \alpha^k - 2gf^{ijk}\alpha^j A^i_\mu \quad (1.7)$$
1.1 The Standard Model

Finally, in order to give these gauge fields a free term, a tensor $F_{\mu\nu}$ antisymmetric in its spatial indexes has to be introduced. The definition $-igF^{k\mu\nu} = [D_\mu, D_\nu]$ or, more explicitly

$$F^{\mu\nu}_k = \partial_\mu A^\nu_k - \partial_\nu A^\mu_k + gf^{ijk} A^j_\mu A^k_\nu$$

preserves the local gauge invariance.

To describe the experimental knowledge of the particles and their interactions at the quantum level, two such non-Abelian symmetries, together with an Abelian one, are necessary and sufficient. First of all, the Lagrangian exhibits a local $U(1)$ phase invariance and we call $B_\mu$ the gauge field associated with it.

A second invariance, under a set of non-Abelian transformations that form a $SU(2)$ group, leads to the introduction of three $W^i_\mu$ fields ($i = 1, 2, 3$), one for each of the generators $\tau^i/2$, where $\tau^i$ is just another common notation for the set of $2 \times 2$ complex Hermitian and unitary Pauli matrices. The third invariance, also non-Abelian, under a set of transformations that form an $SU(3)$ group, requires the introduction of eight $G^a_\mu$ fields ($a = 1, \ldots, 8$). The general transformation is then given by:

$$U = \exp \left[ i \left( g' Y(x) \frac{\tau^i}{2} + g Y(x) \tau^i + g s \lambda^a(x) \lambda^a \right) \right]$$

(1.8)

and the covariant derivative, which ensures the invariance of the theory under all the three transformations, takes the form:

$$D_\mu = \partial_\mu - ig g' B_\mu - ig g Y(x) \frac{\tau^i}{2} W^i_\mu - ig s \lambda^a G^a_\mu$$

(1.9)

where the scalar $Y$ and the matrices $^2 \tau^i$ and $\lambda^i$ are the generators for the $U(1)$ hypercharge, $SU(2)$ weak isospin and $SU(3)$ colour charge groups, respectively. The way fermions behave under gauge transformations depends on the charge they carry with respect to each interaction:

- $SU(3)_C$: only quarks have colour charge, and appear as colour triplets under $SU(3)$ transformations. Other leptons transform as colour singlets;

- $SU(2)_L$: recalling the chiral decomposition of $\psi$ into the $\psi_L$ and $\psi_R$ spinors (equation (1.3)), the weak-isospin charge is experimentally found to be different for left and right-handed particles. Left-handed fermions transform as isospin doublets, while right-handed ones are singlets of 0 weak isospin, and therefore do not interact with the gauge bosons of this symmetry group. This chiral nature of the weak isospin transformations has an immediate consequence. Fermion mass terms in the Lagrangian are written as

$$- m \bar{\psi} \psi = -m \bar{\psi} (P_L + P_R) \psi = -m (\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L)$$

(1.10)

which manifestly violate gauge invariance, since $\psi_L$ is a member of an isospin doublet and $\psi_R$ is a singlet. We are therefore forced to conclude that fermion mass terms cannot be included into the theory in this naive way;

- $U(1)_Y$: the $U(1)$ hypercharge induces transformations as singlets and is non-zero for all fermions except for the right-handed neutrinos. As a convention, the corresponding quantum number for left-handed leptons is chosen to be $Y_L = -1$.

$^2$The $\lambda^a$ are the Gell-Mann traceless and Hermitian matrices.
Since right-handed neutrinos do not couple to any of the previously introduced interactions, they can be regarded as “sterile” and are not included into the theory.

Restricting to the electroweak sector alone, the Lagrangian must include terms for the free gauge fields, which look like:

\[-\frac{1}{4} W_i^\mu W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}\]  

(1.11)

Given the SU(2) algebra, we can write:

\[
W_i^{\mu\nu} = \partial_\mu W_i^\nu - \partial_\nu W_i^\mu + g\epsilon_{ijk} W_j^{\mu} W_k^{\nu},
\]

\[
B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu
\]

(1.12)

From the above equation for $W_{\mu\nu}$, self-interaction terms among the gauge bosons are visible, which are due to the non-Abelian character of the SU(2) gauge symmetry.

Recalling equation (1.7), the following relations can be derived, expressing the transformation law for the vector gauge fields:

\[
B_\mu \rightarrow B'_\mu = B_\mu + \partial_\mu \beta(x)
\]

\[
\tilde{W}_\mu \rightarrow \tilde{W}'_\mu = \tilde{W}_\mu + \partial_\mu \tilde{\alpha}(x) - g (\vec{\alpha}(x) \times \tilde{W}_\mu)
\]

(1.13)

Unlike strong interactions, identified with the SU(3)$_C$ symmetry group, the $U(1)_Y$ and SU(2)$_L$ gauge interactions do not directly correspond to the electromagnetic and weak forces respectively. Instead, the observed interactions are a manifestation of the combined SU(2)$_L \times U(1)_Y$ gauge group, where the physical fields $A_\mu$, $Z_\mu$ and $W^\pm_\mu$ for, respectively, the photon, the Z boson and the $W^\pm$ bosons, arise as combinations of the gauge fields according to:

\[
W^\pm_\mu = \frac{1}{\sqrt{2}} \left( W^1_\mu \mp i W^2_\mu \right)
\]

\[
\begin{pmatrix}
A_\mu \\
Z_\mu
\end{pmatrix} = \begin{pmatrix}
\cos \theta_W & \sin \theta_W \\
-\sin \theta_W & \cos \theta_W
\end{pmatrix} \begin{pmatrix}
B_\mu \\
W^3_\mu
\end{pmatrix}
\]

(1.14)

where $\theta_W$ is the weak mixing angle (Weinberg angle), having a measured value$^3$ of $\sin^2 \theta_W = 0.23153 \pm 0.00016$. The SU(2)$_L$ and $U(1)_Y$ groups cannot therefore be considered separately, since the two components of the doublets have different electric charge. The relation between electric charge, hypercharge and weak isospin is given by the Gell-Mann-Nishijima formula:

\[
Q = T_3 + \frac{Y}{2}
\]

Up to this point, not only are the fermions forced to be massless, but also gauge bosons mass terms are not allowed, if the local gauge symmetry has to be preserved. The transformation of the gauge fields (see equation (1.13)) does not allow for an explicit term $\propto \frac{1}{2} W_i^{\mu} W_i^{\mu\nu}$ or $\frac{1}{2} B_{\mu\nu} B^{\mu\nu}$. A possible solution for the conflict between massless particles, as required by the theory, and massive fermions and vector bosons, as observed experimentally, can be provided by the spontaneous breaking of the symmetry.

### 1.1.3 Spontaneous symmetry breaking and Higgs mechanism

If a theory is described by a Lagrangian which possesses a given symmetry, but its physical “ground state” (that is, the state with the lowest energy) does not, the symmetry is said to be

---

$^3$This is actually the value of $\sin^2 \theta_{\text{lept}}^{\text{eff}}$, the effective Weinberg angle obtained from Z-pole measurements in leptonic final states [5].
spontaneously broken.

A canonical example of a spontaneously broken symmetry is that of a ferromagnetic system. Above the Curie temperature $T_C$, the system shows a $SO(3)$ rotational symmetry, with all the dipoles randomly oriented in the three-dimensional space, yielding a null overall magnetization. For $T < T_C$ the configuration of minimum energy is reached when all the dipoles are aligned in some arbitrary direction (spontaneous magnetization) and the rotational symmetry is broken. The system ground state then chooses a particular configuration among the infinite possible, but once a ground state is assumed, it cannot be changed, unless an external input of energy is introduced into the system to re-orient the dipoles in a different direction.

An important consequence of the spontaneous symmetry breaking in QFT, which happens when the original symmetry is continuous, is the appearance of massless and scalar (spin-0) particles. This statement is in fact the result of a theorem, which goes under the names of Nambu and Goldstone [6], and the newly appeared scalar fields are therefore referred to as “Goldstone bosons”.

The number of Goldstone bosons of the broken theory coincides with the number of continuous symmetries which are broken by the choice of a specific ground state.

An important consequence of the spontaneous symmetry breaking in QFT, which happens when the original symmetry is continuous, is the appearance of massless and scalar (spin-0) particles. This statement is in fact the result of a theorem, which goes under the names of Nambu and Goldstone [6], and the newly appeared scalar fields are therefore referred to as “Goldstone bosons”.

The number of Goldstone bosons of the broken theory coincides with the number of continuous symmetries which are broken by the choice of a specific ground state.

In the Standard Model, one needs an external field to break the electroweak gauge symmetry and this role is taken by the Higgs field [7, 8, 9]. In order to generate masses for the three gauge bosons $W^\pm$ and $Z^0$, without generating a photon mass, at least three degrees of freedom are needed. The simplest solution is to add a complex $SU(2)$ doublet of scalar fields having hypercharge $Y = 1$:

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

(1.15)

This doublet has no colour charge and therefore it will not affect the $SU(3)_C$ sector. The Lagrangian for the Higgs field is given by:

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

with

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda \left( \Phi^\dagger \Phi \right)^2$$

(1.16)

which is manifestly invariant under $SU(2)_L \times U(1)_Y$.

The expected form of the potential is sketched in figure 1.1: for $\mu^2 > 0$ the scalar potential has a

---

**Figure 1.1:** Form of the potential $V(\phi)$ depending on the sign of $\mu^2$, positive on the left and negative on the right.
global minimum at \( \langle 0|\Phi|0 \rangle = \Phi_0 = 0 \), which would not break the electroweak gauge symmetry. For \( \mu^2 < 0 \) the potential has a circle of degenerate minima at

\[
\langle 0|\Phi|0 \rangle = \Phi_0 = -\frac{\mu^2}{2\lambda} = \frac{1}{2}v^2
\]

where \( v \equiv \sqrt{-\frac{\mu^2}{\lambda}} \) (1.17)

\( v \) is the vacuum expectation value (VEV) of the field \( \Phi \).
The spontaneous breaking of the \( SU(2) \) symmetry consists in choosing a particular ground state, around which the Higgs field \( \Phi(x) \) is expanded. The particular vacuum chosen is:

\[
\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}
\]

(1.18)

It is clear that neither \( T_i \) nor \( Y \) cancels \( \phi_0 \), in particular:

\[
T_3\Phi_0 = -\frac{1}{2}\Phi_0 \quad \text{and} \quad Y\Phi_0 = \Phi_0
\]

On the contrary, since \( \Phi_0 \) is neutral, the \( U(1)_Q \) symmetry remains unbroken, that is

\[
Q\Phi_0 = \left( T_3 + \frac{Y}{2} \right)\Phi_0 = 0 \quad \Rightarrow \quad \Phi_0 \to \Phi'_0 = e^{i\delta(x)\frac{Q}{2}}\Phi_0 = \Phi_0
\]

(1.19)

Thus, \( SU(2)_L \) and \( U(1)_Y \) are completely broken separately, but the product group \( SU(2)_L \times U(1)_Y \) is not: after the symmetry breaking, it will remain a residual symmetry generated by \( Q \).

This pattern of symmetry breakdown is then described in formula as: \( SU(2)_L \times U(1)_Y \to U(1)_Q \). If we label the fluctuation of the \( \phi_1, \phi_2, \phi_3 \) and \( \phi_4 \) real scalar fields around the minimum as \( \theta_2, \theta_1, H \) and \( -\theta_3 \) we can write

\[
\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \theta_2(x) + i\theta_1(x) \\ v + H(x) - i\theta_3(x) \end{pmatrix}
\]

\[
\approx \frac{1}{\sqrt{2}} \begin{pmatrix} 1 + i\theta_3/v \\ i(\theta_1 - i\theta_2)/v \\ i(\theta_1 + i\theta_2)/v \\ 1 - i\theta_3/v \end{pmatrix} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}
\]

(1.20)

Thanks to the \( SU(2) \) invariance of the Lagrangian, the three fields \( \theta_i(x) \) in equation (1.20) can be gauged away with a transformation \( U = \exp\left(-i\frac{2\theta_1}{v} \frac{\tau_1}{2}\right) \): these are the massless Goldstone bosons, which do not explicitly appear in the final Lagrangian.

This particular gauge fixing in which the Goldstone boson components are set to zero, making the number of scalar degrees of freedom minimal, is often called “unitary gauge”.

By expanding the scalar Higgs field Lagrangian (1.16) around \( \Phi_0 \) using

\[
\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}
\]

(1.21)

\[4\text{In the following expression we only consider small fluctuations around the minimum.} \]
one finds [10]:
\[
\mathcal{L}_{\text{Higgs}} = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{1}{2} 2 v^2 \lambda H^2 - \frac{1}{3!} 6 v \lambda H^3 - \frac{1}{4!} 6 \lambda H^4 \\
+ \frac{1}{2} v^2 g^2 \frac{1}{4} W^-_\mu W^-_\mu + \frac{1}{2} v^2 g^2 \frac{1}{4} W^+ W^+ \\
+ \frac{1}{2} \frac{1}{4} (g W^3 - g' B^\mu) \frac{2}{\sqrt{g^2 + g'^2}} + 0 \cdot \left( \frac{g W^3 + g B^\mu}{\sqrt{g^2 + g'^2}} \right)^2 \\
+ \frac{1}{2} (2vH + H^2) \left[ g^2 W^- W^+ + \frac{1}{2} (g^2 + g'^2) \left( \frac{W^3 - g' B^\mu}{\sqrt{g^2 + g'^2}} \right)^2 \right]
\]
(1.22)

In the first line, originated from the expansion of the potential \( V(\Phi) \), the kinetic term for the Higgs boson, its mass term and the Higgs boson self-interaction terms are visible. We notice that the Higgs mass itself is equal to \( m_H = v \sqrt{2 \lambda} \). While \( v \) can be put in relation to the Fermi constant \( G_F \) and therefore estimated from precise muon lifetime measurements, i.e. \( v = (\sqrt{2} G_F)^{-1/2} \approx 247 \text{ GeV} \), \( \lambda \) is a free parameter of the model, hence the Higgs mass is unknown.

In the second line, coming from the kinetic term \((D_\mu \Phi)^\dagger (D^\mu \Phi)\), the \( W^\pm \) vector bosons can be identified in the linear combination of the gauge bosons \( W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp i W^2) \). The process of spontaneous symmetry breaking allows them to acquire mass.

The third line provides the right mass terms for the observed \( Z^0 \) and \( \gamma \) vector bosons\(^5\). The first linear combination of the gauge fields \( W^3_\mu \) and \( B_\mu \) comes with an appropriate mass term and it is therefore interpreted as the massive \( Z^0 \) boson. The second combination of fields is orthogonal to the first one and it is added “by hand” with an associated null mass.

We can therefore interpret the results as:

\[
\begin{align*}
m_W &= \frac{1}{2} v g & \text{with} & & W^\pm_\mu &= \frac{1}{\sqrt{2}} \left( W^1_\mu \mp i W^2_\mu \right) \\
m_Z &= \frac{1}{2} v \sqrt{g^2 + g'^2} & \text{with} & & Z_\mu &= \frac{g W^3_\mu - g' B^\mu}{\sqrt{g^2 + g'^2}} \\
m_{\gamma} &= 0 & \text{with} & & A^\mu &= \frac{g' W^3_\mu + g B^\mu}{\sqrt{g^2 + g'^2}}
\end{align*}
\]
(1.23)

The gauge bosons have “eaten” the three massless Goldstone bosons, acquiring a mass. The degrees of freedom of the Goldstone bosons are in fact needed: once the gauge bosons become massive, an additional degree of freedom is required in order to allow them to have a longitudinal polarization. We also notice that the unbroken \( U(1)_Q \) symmetry causes the photon to remain massless.

The mixing of \( W^3_\mu \) and \( B_\mu \) yielding the physical force carriers can be interpreted as a rotation of parameter \( \theta_W \), where we have identified

\[
\frac{g}{\sqrt{g^2 + g'^2}} = \cos \theta_W \quad \text{and} \quad \frac{g'}{\sqrt{g^2 + g'^2}} = \sin \theta_W
\]
(1.24)

Therefore, the following relation between the weak bosons masses can be inferred:

\[
m_Z = \frac{m_W}{\cos \theta_W}
\]
(1.25)

---

\(^5\)The numerical factor \( \sqrt{g^2 + g'^2} \) has been introduced in order to normalize the combinations of gauge fields \( g W^3_\mu - g' B^\mu \) and \( g' W^3_\mu + g B^\mu \).
Finally, in the last line of the Lagrangian (1.22), the cubic and quartic couplings of the Higgs boson to the weak gauge bosons can be deduced. We notice in particular that the coupling of one single Higgs boson to a pair of W or Z bosons is proportional to $m_W$ and $m_Z$, respectively:

$$g_{HWW} = g m_W$$
$$g_{HZZ} = g \frac{2}{\cos \theta_W} m_Z$$

From this, the following relation can be derived for the branching ratio $B$ of the Higgs boson into a pair of vector bosons (valid at tree level for Higgs boson masses well above the kinematic threshold for the production of a diboson pair):

$$B(H \rightarrow WW) B(H \rightarrow ZZ) = \left( \frac{g_{HWW}}{g_{HZZ}} \right)^2 = 4 \cos^2 \theta_W \frac{m_W^2}{m_Z^2} \simeq 2.7$$

The full Standard Model Lagrangian (neglecting the colour part) can be written as

$$\mathcal{L}_{SM} = \mathcal{L}_{GWS} + \mathcal{L}_{Higgs}$$

where its electroweak part (representing the Glashow-Weinberg-Salam model of electroweak unification) is given by:

$$\mathcal{L}_{GWS} = -\frac{1}{4} W^i_{\mu\nu} W^{\mu\nu}_i - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}$$
$$+ i \bar{\nu} L \gamma^\mu \partial_\mu \nu_L + i \bar{e} L \gamma^\mu \partial_\mu e_L + i \bar{e} R \gamma^\mu \partial_\mu e_R +$$
$$+ i \bar{f} L \gamma^\mu (\pm ig \tau^i W^i_\mu - ig' Y^i B_\mu) f_L + i \bar{e} R \gamma^\mu (\mp ig' Y^i B_\mu) e_R$$

Re-expressing the interaction part of the above Lagrangian in terms of the physical fields and writing explicitly the covariant derivative, one obtains:

$$\mathcal{L}_{GWS}^{\text{int}} = \mathcal{L}_{CC}^{\text{int}} + \mathcal{L}_{NC}^{\text{int}}$$
$$= \left\{ e J^\mu_\mu A^\mu + \frac{g}{\cos \theta_W} J^2_\mu Z^\mu \right\} + \left\{ \frac{g}{\sqrt{2}} \left( J^+_\mu W^+ - J^-_\mu W^- \right) \right\}$$

for the neutral and charged part respectively. The electromagnetic coupling constant $e$ has been introduced, identifying $e = g \sin \theta_W$. The following currents have also been defined:

$$J^\text{em}_\mu = Q f_{\gamma_\mu} f$$
$$J^Z_\mu = \frac{1}{2} \bar{f}_{\gamma_\mu} \left( c_\nu^{\ell} - c_A^{\gamma_5} \right) f \quad \text{with} \quad c_\nu^{\ell} = T_3 - 2Q \sin^2 \theta_W, \quad c_A^{\gamma_5} = T_3$$

### 1.1.3.1 Fermion masses

An attractive feature of the Standard Model is that the same Higgs doublet which generates W and Z masses is also sufficient to give mass to leptons and quarks. For the lepton sector, for instance, the following Lagrangian can be added (for each lepton generation $\ell$):

$$\mathcal{L}_{\text{Yukawa}}^{\ell} = -G_\ell \left[ (\bar{\ell}_L \Phi) \ell_R + \bar{\ell}_R (\Phi^\dagger \ell_L) \right]$$
where the Higgs doublet has exactly the required $SU(2)_L \times U(1)_Y$ quantum numbers to couple to $\bar{\ell}_L \ell_R$. After the breakdown of the symmetry, inserting equation (1.21) into the (1.32), one obtains:

\[
\mathcal{L}_{\text{Yukawa}}^\ell = -\frac{G_\ell}{\sqrt{2}} \left\{ (\bar{\nu}_\ell, \bar{\ell})_L \left( \begin{array}{c} 0 \\ v + H \end{array} \right) \ell_R + \bar{\ell}_R (0, v + H) \left( \begin{array}{c} \nu_\ell \\ \ell \end{array} \right)_L \right\} 
\]

\[
= -\frac{G_\ell}{\sqrt{2}} \left\{ v(\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L) + (\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L)H \right\} 
\]

\[
= -\frac{G_\ell}{\sqrt{2}} \left\{ v\bar{\ell}\ell + \bar{\ell}\ell H \right\} 
\]

(1.33)

It is now easy to see from equation (1.33) that an appropriate choice of the coupling factor $G_\ell$ (called “Yukawa coupling”) can generate the required lepton mass:

\[
m_\ell = \frac{G_\ell \cdot v}{\sqrt{2}} 
\]

(1.34)

Rephrasing the above statement from another point of view: the coupling of a fermion to the Higgs boson is proportional to its mass $m_\ell$. This property has of course important consequences for the Higgs production and decay.

The Yukawa Lagrangian for the lepton sector can be rewritten as

\[
\mathcal{L}_{\text{Yukawa}}^\ell = -m_\ell \left\{ \bar{\ell} \ell + \frac{1}{v} \bar{\ell}\ell H \right\} 
\]

(1.35)

This technique allows to generate a mass term for leptons and down-like quarks. For up-like quarks and neutrinos, a different Higgs doublet has to be introduced, defined as:

\[
\Phi_c = i\tau_2 \Phi^* = \left( \begin{array}{c} \phi^0 \cdot^* \\ \phi^+ \cdot^* \end{array} \right) 
\]

(1.36)

Not only have fermions acquired mass thanks to the spontaneous symmetry breaking, but a new coupling between the Higgs boson and the fermions has become manifest. As an important consequence, in case the Higgs boson is observed in a future experiment, the amplitude of a Higgs decay process will be proportional to the second power of the mass of the particle the Higgs decays into (see equations (1.26) and (1.34)).
1.2 The search for the Higgs boson

The constraints on the unknown mass of the Higgs boson are derived both from theoretical considerations and from direct searches spanning all the energy range accessible to past and present colliders.

1.2.1 Theoretical constraints

Although the Higgs boson mass, as we have seen, is not predicted by the theory, both lower and upper limits can be set, based on theoretical considerations [11]. A first upper constraint is found by studying the weak boson scattering process $W_L W_L \rightarrow W_L W_L$. In a scenario where no Higgs boson actually exists, the amplitude for such a process would be proportional to the center-of-mass energy, and thus violate unitarity at high energy, namely $\sqrt{s} \simeq 1.2$ TeV. Diagrams involving the exchange of an Higgs boson among the $W_L$s allow for cancellations, so that the scattering amplitude is regularized and finite at all energies, provided that $m_H \lesssim 700$ GeV/c$^2$.

More restricting bounds on the Higgs mass depend on the energy scale $\Lambda$ up to which the SM is assumed valid, i.e. the scale up to which no new interactions and particles are expected. These bounds are derived from the one-loop Renormalization Group Equation for the Higgs quartic coupling $\lambda$, which describes the evolution of this parameter with the energy. First of all, the Higgs potential described in equation (1.16) is affected by radiative corrections from Higgs couplings to vector bosons and fermions (mostly top loops). These corrections may modify the shape of the potential in such a way that an absolute minimum no longer exists and no stable spontaneous symmetry breaking can occur. The requirement of vacuum stability, i.e. $\lambda$ positive and large enough to avoid instabilities, implies a lower bound on $m_H$ (stability bound).

Another limit can be imposed after observing that the coupling constant evolution with energy presents a singularity (Landau pole) for some energy value. Since $\lambda$ explodes in the proximity of the pole, the theory is no more consistent with a perturbative approach. A possible way to make it meaningful for all energy scales is to set the value of $\lambda$ equal to 0, which would result

![Figure 1.2: Upper and lower theoretical limits on the Higgs mass as a function of the energy scale $\Lambda$ up to which the Standard Model is assumed to hold. The shaded area indicates the theoretical uncertainties in the calculation of the bounds. Here the values $m_t = 175$ GeV/c$^2$ and $\alpha_s(M_Z) = 0.118$ have been used.](image-url)
1.2 The search for the Higgs boson

in a trivial, i.e. non interacting, theory. Another possible approach is to consider the SM as an effective theory up to an energy scale $\Lambda$, after which it leaves the perturbative domain. This requirement imposes an upper limit to the Higgs mass, depending on $\Lambda$ itself (triviality bound). The theoretical constraints on $m_H$ as a function of the energy scale $\Lambda$ are shown in figure 1.2. As it can be seen, this plot suggests that if the SM validity extends up to the scale of Grand Unification Theories ($\Lambda_{\text{GUT}} \simeq 10^{16}$ GeV), the Higgs boson mass has to lie roughly in the $150 - 180$ GeV/$c^2$ range. Conversely, for an Higgs particle lighter than 150 or heavier than 180 GeV/$c^2$, new physics is expected to exist at an energy scale below $\Lambda_{\text{GUT}}$.

Finally, since the Higgs boson enters into one-loop radiative corrections in the Standard Model, precise electroweak measurements can bound its mass. However, the dependence on $m_H$ in all one-loop electroweak parameters is only logarithmic and therefore the limits derived from this method are relatively weak. In contrast, the top quark contributes quadratically to many observables. From precision measurements at LEP and SLD of many electroweak observables, and direct measurements of $m_W$ and $m_t$ at LEP-II and Tevatron, the plots of figure 1.3 can be obtained, where contour curves of 68% probability are shown in $(m_W, m_H)$ and $(m_t, m_H)$ planes. In these plots the combination of the electroweak results is updated to the summer of 2011, while the 95% confidence level exclusion bands come from the final results of the direct search at LEP-II [12] (lower limit of 114.4 GeV/$c^2$ on $m_H$) and from the July 2010 Tevatron combination [13] (158-175 GeV/$c^2$ exclusion).

![Figure 1.3](attachment:image.png)

**Figure 1.3:** Contour curves of 68% probability in (a) the $(m_W, m_H)$ plane and (b) the $(m_t, m_H)$ plane, based on various LEP, LEP-II, SLD and Tevatron results. In (a) the direct measurements of $m_W$ and $\Gamma_W$ are excluded while making the contour plot, while $m_t$ is included. The measured value of $m_W$ is represented separately with a shaded horizontal band of $\pm 1$ sigma width. In (b) the direct measurement of $m_t$ is excluded, while $m_W$ and $\Gamma_W$ are included. The measured value of $m_t$ is represented by the horizontal band. The vertical band shows the 95% confidence level exclusion limit on $m_H$.

In order to obtain the most stringent constraint on the mass of the SM Higgs boson, the whole ensemble of electroweak measurements are used as input to perform a global fit. The fit consists in a $\chi^2$ minimization, where the $\chi^2$ is calculated comparing the measured values of 18 different variables and their errors with their predictions calculated within the framework of the Standard Model, as a function of five input parameters ($\Delta \alpha^{(5)}_{\text{had}}(m_Z^2)$, $\alpha_S(m_Z^2)$, $m_Z$, $m_t$, $m_H$). This
analysis procedure tests quantitatively how well the Standard Model is able to describe the complete set of measurements with just one value for each of the five input parameters. Moreover, the fit yields a best value expected for the only unknown parameter $m_H$. The result of the fit is reported in figure 1.4, where the $\Delta\chi^2(m_H) = \chi^2_{\text{min}}(m_H) - \chi^2_{\text{min}}$ curve is shown.

From the result of the fit it is clear that electroweak measurements seem to favour a light Higgs, with the one-sided 95% C.L. upper limit on $m_H$ being:

$$m_H < 161 \text{ GeV}/c^2$$

(1.37)

Of course, since these bound arises from loop corrections, it may be circumvented by some still unknown new physics entering the same loops.

1.2.2 Direct constraints from LEP and Tevatron

One major direct search for the Higgs boson was carried out at the LEP accelerator at CERN [12]. Data from $e^+e^-$ collisions at a center-of-mass energy up to 209 GeV were used by the four experiments installed on the LEP accelerator to look for hints of the Higgs boson. The main production mechanism at an $e^+e^-$ collider was the so-called “Higgs-strahlung”, where an Higgs boson is radiated by a virtual $Z$ boson. The most probable final state consisted in a couple of $b$-jets coming from the Higgs decay and another couple of jets from the $Z$ decay (although also the leptonic decay modes of the $Z$ were considered). To summarize, the following channels were
analyzed (sorted for relative importance):

\[ e^+e^- \rightarrow Z^*/\gamma^* \rightarrow Z(\rightarrow q\bar{q}) \ H(\rightarrow b\bar{b}) \]
\[ e^+e^- \rightarrow Z^*/\gamma^* \rightarrow Z(\rightarrow \nu\bar{\nu}) \ H(\rightarrow b\bar{b}) \]
\[ e^+e^- \rightarrow Z^*/\gamma^* \rightarrow Z(\rightarrow l^+l^-) \ H(\rightarrow b\bar{b}) \quad \text{with} \quad l = e, \mu, \tau \]

(1.38)

No significant excess\(^6\) of events with a Higgs-compatible topology was found, thus allowing to set a lower bound on the Higgs boson mass at 95% confidence level (as shown in figure 1.5):

\[ m_H > 114.4 \text{GeV}/c^2 \quad \text{(LEP-II – 2003)} \]

(1.39)

![Figure 1.5](image)

**Figure 1.5:** The ratio \( CL_s = CL_{s+b}/CL_b \) for the signal plus background hypothesis. Solid line: observation; dashed line: median background expectation. The green and yellow bands correspond to the 68\% and 95\% probability bands. The intersection of the horizontal line for \( CL_s = 0.05 \) with the observed one, defines the 95\% C.L. lower bound on the mass of the SM Higgs boson. Picture taken from [12].

The search for the Standard Model Higgs continued at Tevatron, a \( p\bar{p} \) collider with a center-of-mass energy of \( \sqrt{s} = 1.96 \) TeV which ended its operation in 2011. Also in this case, the main production mechanism consists in Higgs boson production in association with a vector boson (W or Z), whose decay products are used to tag the event. A large variety of final states was considered by the two experiments built at Tevatron, namely CDF and DØ. As it can be seen in figure 1.6, the combination of all analyses updated to the end of 2011 [14] leads to an exclusion of the SM Higgs boson in a mass region around 160 GeV/c\(^2\) and, at very low mass, in a region already covered by LEP results, namely:

\[ m_H \notin [100, 106] \text{GeV}/c^2 \quad \text{(Tevatron – 2011)} \]

\[ m_H \notin [147, 179] \text{GeV}/c^2 \]

(1.40)

\(^6\)At a mass of 115 GeV/c\(^2\), the ALEPH experiment reported an excess of four events in the four-jet final state, compatible with the production of a SM Higgs boson. This excess was nevertheless not statistically significant, especially after combining it with the results of DELPHI, L3 and OPAL.
Figure 1.6: Observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross section, as a function of the Higgs boson mass for the combined CDF and DØ analyses. The limits are expressed as a multiple of the SM prediction for test masses for which both experiments have performed dedicated searches in different channels. The bands indicate the 68% and 95% probability regions where the limits can fluctuate, in the absence of signal. The limits displayed in this figure are obtained with the Bayesian calculation. Picture from [14].

In the same referenced paper [14], the two collaborations observe an excess of data events with respect to the background estimation in the mass range $115 < m_H < 135 \text{ GeV}/c^2$. The $p$-value for a background fluctuation to produce this excess is $\sim 3.5 \times 10^{-3}$, corresponding to a local significance of 2.7 standard deviations. The global significance reduces to approximately 2.2 standard deviations when taking into account the probability of seeing such an excess anywhere in the full mass range (look-elsewhere effect).
1.3 Higgs boson searches at the LHC

In this section, an overview of the main Higgs production mechanisms and decay channels in a proton-proton collider such as the Large Hadron Collider will be presented. The detailed description of the machine and its detectors is postponed to the next chapter.

1.3.1 Higgs production mechanisms

The Large Hadron Collider (LHC) is a particle accelerator situated at CERN, which started colliding protons at a center-of-mass energy of 7 TeV in 2010. The SM Higgs boson production cross section at a pp hadron collider having that energy is shown in figure 1.7 [15].

The dominating Higgs production process over the entire mass range accessible at the LHC is the gluon fusion ($gg \rightarrow H$) [16]. It proceeds with an heavy quark triangle loop, as shown in figure 1.8(a). Because of the already discussed Higgs couplings to fermions, the $t$-quark loop is by far the most important, due to the heavy top mass, with the next-to-leading contribution being at least a factor $O(m_t^2/m_b^2)$ smaller.

The second largest production mechanism of the Higgs boson is the vector boson fusion (VBF, $qq \rightarrow qqH$). In this process, which is about one order of magnitude smaller than the gluon fusion, the Higgs boson is originated from the fusion of two weak bosons radiated off the incoming quarks (see figure 1.8(b) for a graphical representation). The hadronization of the quarks produces two forward jets of high invariant mass, which can be used to tag the event and differentiate it from backgrounds. Another interesting property is the reduced hadronic activity within the tag jets, since they are colour disconnected.

In the Higgs-strahlung ($q\bar{q}' \rightarrow WH, q\bar{q}' \rightarrow ZH$) and $t\bar{t}$ associated production ($gg, q\bar{q} \rightarrow t\bar{t}H$) processes the Higgs is produced in association with a $W/Z$ boson or a pair of top quarks (see figures 1.8(c) and 1.8(d)). In both cases, their decay products can be used to tag the event.
1. Introduction

Figure 1.8: Feynman diagrams at the tree level for the most important production processes of a SM Higgs boson: (a) gluon fusion, (b) vector boson fusion, (c) Higgs-strahlung, (d) $t\bar{t}$ associated production.

1.3.2 Higgs decay modes

Depending on the Higgs mass, a plethora of different decay channels can be exploited to detect the Higgs. The total decay width and its different decay branching ratios depend on the Higgs couplings to the vector bosons and the fermions, as they appear in the Standard Model Lagrangian given in the previous paragraph. Due to the dependence of the couplings on the particle masses, the Higgs tends to decay into the heaviest particles kinematically allowed. This behaviour is visible in figure 1.9(a), which shows the Higgs decay branching ratios including also NLO QCD and EWK corrections. From this plot it can be seen how new channels open up when the Higgs mass is above a di-lepton/quark or vector boson threshold. Light-fermion decay modes contribute only in the low mass region (up to $\sim 150 \text{ GeV}/c^2$), where the branching ratio is dominated by the $H \rightarrow b\bar{b}$ channel. Once the decay into a pair of weak boson is possible, it quickly dominates. A peak in the WW decay mode is visible, when the production of two on-shell $W$ bosons becomes possible, while a $Z$ pair is still not accessible. At high masses ($\gtrsim 350 \text{ GeV}/c^2$) also $t\bar{t}$ pairs can be produced. The Higgs boson does not couple to photons and gluons at tree level, but such couplings can arise via fermion loops and they give a contribution in the low mass region.

The total width, given by the sum over all the possible decay channels, is shown in figure 1.9(b). It quickly increases with the Higgs mass due to the opening of new channels and it becomes almost as large as the Higgs mass itself around $1 \text{ TeV}/c^2$, thus making it problematic to identify an hypothetic Higgs resonance as a particle: such a heavy Higgs would be more suitably accommodated in theories extending the Standard Model.

The decay $H \rightarrow \gamma\gamma$ is the main channel for the discovery of the Higgs boson at masses of about $140 \text{ GeV}/c^2$ or below. The challenge here is the low branching ratio and therefore the small
signal rate. Large backgrounds come from prompt photon pairs produced by quarks and gluons in the initial state, from one or two jets which fake the photon signature and from Drell-Yan production of electron pairs. The signal signature is two high-$E_T$ isolated photons which can be well identified experimentally. In this channel, we expect the Higgs boson appearance as a narrow peak above a large continuous background.

Due to its very clean signature with 4 isolated leptons in the final state, $H \rightarrow ZZ \rightarrow 4l$ is considered the golden mode for the discovery of the Higgs boson. The backgrounds to this channel are ZZ($^*$) $, t\bar{t}$ and Z$b\bar{b}$ productions, which can be suppressed in an efficient way by some requirements on the lepton isolation, transverse momentum and invariant mass, as well as requirements on the event vertex.

Around $m_H \simeq 160 \text{GeV}/c^2$ the Higgs branching ratio into WW is close to one. This makes $H \rightarrow WW \rightarrow l\nu l\nu$ the discovery channel in this mass range. The signature is two charged leptons and missing energy. Since the mass peak can not be reconstructed due to the neutrinos in the final state, the search strategy is based on event counting, for which an accurate knowledge of all the possible backgrounds is needed. The main backgrounds are electroweak WW, $t\bar{t}$ and W+jets productions. They can be reduced imposing some requirements on the lepton momentum and isolation, applying a jet veto and exploiting the opening angle between the two leptons, whose smallness is due to spin correlations.

As shown in figure 1.10, Monte Carlo studies conducted by the CMS experiment in 2008 indicate that, in case a SM Higgs boson exists, an integrated luminosity of 5 fb$^{-1}$, collected at $\sqrt{s} = 8$ TeV, would grant a discovery significance of at least 3$\sigma$ ($5\sigma$ in a wide part of the mass range) for all Higgs mass hypotheses between 114 and 600 GeV/$c^2$. Otherwise, the same amount of data would be enough to exclude the SM Higgs existence with a confidence level of at least 95% everywhere in the above mass range.

Before summer 2011 both CMS and ATLAS collected around 1 fb$^{-1}$ of data, therefore being able to update the world best limits on the SM Higgs mass [17, 18]. Their results confirmed the Tevatron exclusion around $m_H = 160 \text{GeV}/c^2$, extending the ruled-out masses in the intermediate regions ($m_H \notin [300, 440] \text{GeV}/c^2$ for CMS and $m_H \notin [295, 450] \text{GeV}/c^2$ for ATLAS).
As can be seen from figure 1.11, the very low mass region ($m_H \lesssim 140 \text{ GeV/c}^2$) and the very high mass region ($m_H \gtrsim 400 \text{ GeV/c}^2$) emerged as the most interesting mass ranges where to concentrate the efforts.

**Figure 1.10:** The plots show (a) the projected exclusion limit and (b) the expected observation significance for a SM Higgs search $\sqrt{s} = 8 \text{ TeV}$ and $L = 5 \text{ fb}^{-1}$. Contributions of individual channels used in the overall combination are also shown.

**Figure 1.11:** The observed and expected 95% C.L. upper limits on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for the SM Higgs boson hypothesis, as a function of the Higgs boson mass, for (a) the CMS experiment [17] and (b) the ATLAS experiment [18] with $L = 1.0 - 1.2 \text{ fb}^{-1}$. These results were shown at the EPS conference in July 2011.
Chapter 2

The CMS detector at the LHC

In the first section of this chapter a general overview of the Large Hadron Collider and its detectors is given. A particular emphasis is placed on the CMS detector, which is the main subject of all section 2.2.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [19] is a proton-proton and heavy ions collider which lies in a 27 km long circular tunnel, spanning across the Franco-Swiss border near Geneva, Switzerland. It was built by the European Organization for Nuclear Research (CERN) with the aim of investigating the high-energy domain of particle physics through the collisions of two opposite beams of hadrons, with a nominal luminosity $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and an energy in the center of mass of 14 TeV, one order of magnitude higher than the previous generation of particle accelerators. This scale of energy should allow an accurate experimental exploration of the Standard Model Higgs sector and provide further understanding of the properties of the already known particles enabling, for example, deeper studies of top quark physics or CP violation.

However, the general hope is that this experiment will bring some new physics into the current scenario, verifying or refuting the hypothesis of the existence of Super Symmetric particles at the TeV mass scale or the possible existence of additional quark flavours, beyond the three families predicted by the current Standard Model. Furthermore, the heavy ions experiments, carried out with high energy beams of lead (Pb) ions, will ensure the unique possibility of observing a new state of matter, the quark-gluon plasma, that is a new phase predicted by quantum chromodynamics which consists of almost free quarks and gluons at extremely high temperatures and densities.

In order to achieve such an high energy, the charged proton beams are prepared by the existing CERN accelerator complex, which, in increasing order of energy, consists of two linear accelerators, a Proton Synchrotron (PS) and a Super Proton Synchrotron (SPS), which brings their energy up to 450 GeV, before injecting them into the main LHC ring. Here, 8 radio-frequency resonant cavities oscillating at 400 MHz accelerate the bunches to their final energy with “kicks” of 0.5 MeV per turn. The beams are kept on their path through the use of 1232 dipole magnets and focused with additional 392 quadrupole magnets. The Nb-Ti magnets work in a superconducting regime and therefore have to be kept at a temperature of 1.9 K by means of super-fluid Helium. Since collisions occur between particles of the same charge, two separate beam pipes are required, with two opposite magnetic field configurations. At the four beam intersection points
the main experiments have been installed (see fig 2.1). Two of them, CMS (Compact Muon
Solenoid [20]) and ATLAS (A Toroidal Lhc ApparatuS [21]), are general purpose detectors, one
(LHC-b [22]) deals with the study of CP violation and B-physics and the last one, ALICE (A
Large Ion Colliding Experiment [23]), is devoted to the investigation of high energy ions physics.
At the end of March 2010 the first collisions took place between beams of 3.5 TeV each, giving
a total center-of-mass energy of 7 TeV (half of the nominal one). The same energy was kept for
all the 2011 run, while for 2012 the choice was to increase the energy of the beams up to 4 TeV
each (8 TeV in the center of mass) to profit from the largest Higgs production cross section at
this energy.

2.2 The CMS detector

The Compact Muon Solenoid (fig. 2.2) is a general purpose detector located at the interaction
point number 5 along the LHC ring [24]. It has a cylindrical symmetry around the beam pipe,
with an overall length of 22 m and a diameter of 15 m and it is made of a sequence of substructures
dedicated to the measurement of energy and momentum of all the collision products.
Starting from the innermost layer, the main subdetectors that are contained in CMS are: the
tracking system, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL)
and the muon chambers. The choice which has been made was to keep the tracker and both
the calorimeters inside the 13 m long and 6 m wide superconducting solenoid, hence the name
“compact”. The muon system is placed outside, embedded in the iron magnetic yoke that sus-
tains the structure and drives back the 3.8 T magnetic field produced by the solenoid.
Every part of the overall detector has been designed in order to meet the goals of the LHC
physics program: the innermost tracking system allows the reconstruction of tracks momentum
and interaction vertices, while the high granularity of the calorimeters provides a good spatial
resolution and the possibility to exploit cluster shape-based methods to reject fake electrons.
Moreover, the magnetic field produced by the solenoid allows, through the curvature of their
tracks, an excellent measurement of particles momentum: the goal is to retain a 10% resolution
for muon transverse momenta of 1 TeV.
In a typical proton-proton collision, the fractions $x_a$ and $x_b$ of the parent proton momentum
carried by the interacting partons are in general different, and the rest frame of the hard collision
is boosted along the beam line with respect to the laboratory frame. The reconstruction of the
boost of the system requires the full reconstruction of the remnants of the colliding protons,
Figure 2.2: The Compact Muon Solenoid experiment at LHC.
which is in practice not possible, because of the presence of the beam-pipe and the instrumentation placed at small angles with respect to the beam-line. Because of the unknown energy balance along the beam-line, proton collisions are usually studied in a convenient coordinate system which has been established such that the origin is centered at the nominal collision point inside the experiment, the z-direction is parallel to the beam line, the y-direction is vertical and the x-direction is horizontal, pointing toward the center of the ring. The azimuthal angle, $\phi$, is measured around the beam line in the xy plane, starting from the x-axis, while the polar angle $\theta$ is measured from the z-axis. The polar angle is usually expressed in terms of the pseudorapidity $\eta$, defined as:

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

(2.1)

The advantage of this coordinate frame is the Lorentz invariance of transverse quantities and differences in $\eta$, under Lorentz boosts along the beam-line. As a consequence, a solid angle in ($\eta, \phi$) is also invariant under longitudinal boosts.

To reach the nominal high luminosities, up to 2808 bunches per beam, containing about $1.1 \times 10^{10}$ protons each and having a small transverse size of 15 $\mu$m, will be collided every 25 ns\(^2\). This calls for a powerful trigger system which, in CMS, is based on two levels: the first one is an hardware “Level-1 trigger”, installed on every subdetector, while the second one is a software “High Level Trigger”, running on ordinary computer farms.

### 2.2.1 The silicon tracker

The CMS tracker [25] is the innermost subdetector of CMS and surrounds the interaction point, with a total length of 5.4 m and a diameter of 2.4 m. It consists of 1440 silicon pixel and 15148 silicon strip detector modules, divided into five main parts: the Tracker Outer Barrel (TOB), the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker End Caps (TEC) and the pixer detector (see figure 2.3). The entire tracker covers the region of pseudorapidity $|\eta| < 2.5$, with a barrel-endcap transition region at $0.9 < |\eta| < 1.4$.

![Figure 2.3](image.png)

**Figure 2.3:** The different sub-systems of the CMS silicon tracker. Each line represents a detector module.

---

\(^1\)The pseudorapidity $\eta$ is in fact the high-energy limit of the rapidity $y$, defined as $y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right)$. Differences of rapidities are Lorentz-invariant under boosts along the $z$ axis.

\(^2\)Up to 2012, though, pp collisions have been delivered with a minimal bunch time separation of 50 ns.
The choice of silicon as building material was driven by the harsh environment with which the tracker has to cope: at the LHC design luminosity, an average of about 1000 particles coming from more than 20 overlapping proton-proton interactions will be present for each bunch crossing. Therefore, a detector technology featuring high granularity and fast response is required so that trajectories can be identified reliably and attributed to the correct bunch crossing. The intense particle flux will also cause severe radiation damage to the tracking system, so radiation hardness is a fundamental requirement.

The pixel detector surrounds the beam line: is made out of silicon pixels divided into modular units and its aim is to identify the secondary vertices of the interactions and their long-lived products, such as \( \tau \)'s or hadrons containing bottom quarks. The barrel part of the tracker consists of TIB and TOB modules equipped with chips to read out energy depositions.

For the pixel detector, the estimated resolution on the single hit is of 10\( \mu m \) for the \((r, \phi)\) coordinate and 15\( \mu m \) for \(z\) in the barrel, while it is of 15\( \mu m \) and 20\( \mu m \) respectively in the endcaps. For the silicon strip detector, the expected resolution grows to \(\sim 50 \mu m\) in \((r, \phi)\) and \(\sim 500 \mu m\) along the \(z\) coordinate.

The fact that the tracker is a complex solid state device implies that every particle experiences an energy loss inside it, depending from the amount of material traversed. In the case of the CMS silicon tracker, the radiation length of the material crossed by a particle is very dependent from the particle direction. For the central region of the detector the radiation length is about 0.4\( X_0 \), but this number increases rapidly when moving to forward regions, as can be seen in figure 2.4. A maximum of 2\( X_0 \) is found for the barrel-endcap transition region. The material budget constitutes the main source of error in accurate calorimetric measurements of electrons (which emit bremsstrahlung radiation) and photons (which convert into \(e^+e^-\) pairs).

![Tracker Material Budget](image)

**Figure 2.4:** Tracker material budget as a function of the pseudorapidity in units of radiation length.

### 2.2.2 The electromagnetic calorimeter - ECAL

The EM calorimeter [26] plays an important role into the Higgs search because of its high sensitivity into the di-photon decay channel (\(H \rightarrow \gamma\gamma\)) and the capability of reconstructing electrons and positrons coming from the decays of vector bosons produced through the channels \(H \rightarrow WW\) and \(H \rightarrow ZZ\).

ECAL is made out of nearly 83000 lead tungstate (PbWO\(_4\)) crystals arranged into a barrel
structure which covers the pseudorapidity region $|\eta| < 1.479$ and two endcaps on both sides ($1.479 < |\eta| < 3$), as showed in figure 2.5.

![Diagram of CMS electromagnetic calorimeter](image)

**Figure 2.5:** View in $(r, z)$ of a quarter of the CMS electromagnetic calorimeter.

The lead tungstate has been chosen as scintillation material because of its radiation hardness, high density and fast response. The scintillation decay time is comparable with the LHC bunch crossing time: at a nominal rate of 40 MHz there will be an interaction every 25 ns, which is the time required for a crystal to emit about 80% of its scintillation light. However, PbWO$_4$ presents a low light-yield (from 50 to 80 $\gamma$/MeV) which makes it necessary to use intrinsic high-gain photodetectors, capable of operating in an high magnetic field. Avalanche PhotoDiodes (APDs) and Vacuum PhotoTriodes (VPTs) are used, the former glued to the back side of each barrel crystal, the latter to endcap crystals.

The barrel is made of 61200 crystals, arranged in a cylindrical shape with an inner radius of 1.29 m. Every crystal has a truncated pyramidal shape that, except for a small offset, points toward the interaction vertex, with a depth of 23 cm (corresponding to a radiation length of 25.8 $X_0$) and a front surface that covers the range of $1^\circ$ in the $\eta - \phi$ plane ($22 \times 22$ mm$^2$). EB crystals are grouped in arrays of $5 \times 2$ elements, called sub-modules, assembled in a thin glass-fibre alveolar structure. Sub-modules are then arranged in a module, in the number of 40/50 each and, eventually, 4 modules are assembled with metallic cross plates in between to form the biggest unity of the barrel part, the so-called supermodule. On the whole there are therefore 36 supermodules, 18 for each side of the interaction point.

The two endcaps (EE) consist of identically shaped crystals having a larger front face of $29 \times 29$ mm$^2$ and a shorter length of 22 cm, corresponding to $\sim 25 X_0$. They are grouped into carbon-fiber structures of $5 \times 5$ elements, called supercrystals. Each endcap is divided into 2 halves, or Dees, holding 3662 crystals each.

A preshower (ES) is placed in front of EE crystals with the aim of providing position measurement of the electromagnetic shower to high accuracy and discriminating photons produced in an Higgs boson decay from photons produced in a $\pi^0 \rightarrow \gamma\gamma$ decay. Thin lead radiators are used to initiate the shower while silicon strip sensors are placed beyond them to measure the hit position of the shower. The total material depth corresponds to about $3 X_0$, on a fiducial region of $1.653 < |\eta| < 2.6$.

When a photon or electron hits the surface of the scintillator, an electromagnetic shower develops inside the calorimeter and every crystal emits an amount of light directly proportional to the energy released by the shower inside it. To reconstruct the overall energy of the shower (and of the particle who has initiated it) clusters of square matrices of crystals are considered, centered around the most energetic one. The total energy is then estimated as the sum of the energies measured by all the elements of the cluster.
2.2.3 The hadronic calorimeter - HCAL

The purpose of the CMS Hadronic CALorimeter [27] is to measure the energy and direction of hadronic jets, relying on the hadronic and electromagnetic showers initiated by nuclear interactions. It is a sampling calorimeter consisting of three main parts: an Hadron Barrel (HB), an Hadron Endcap (HE) and a Very Forward calorimeter (HF), as sketched in figure 2.6.

![Figure 2.6: Longitudinal view in (r, z) of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters.](image)

The hadron calorimeter barrel is radially restricted between the outer extent of ECAL \((r = 1.77 \, \text{m})\) and the inner extent of the magnet coil \((r = 2.95 \, \text{m})\). This constrains the total amount of material which can be put in to absorb the hadronic shower. Therefore, an outer hadron calorimeter (HO) is placed outside the solenoid, complementing the barrel calorimetry. The endcap part extends until \(|\eta| = 3\) while, beyond it, forward hadron calorimeters are placed at 11.2 m from the interaction point, extending the pseudorapidity coverage down to \(|\eta| = 5.2\).

The barrel and the endcap parts are made of absorbing layers of brass and plastic scintillators, while the forward calorimeter is made of quartz fibers embedded in steel, with photomultipliers as readout detectors. The good ermeticity of this system and the great pseudorapidity region covered allow the collection of most of the transverse energy of the event. This is needful in order to determine with the greatest possible precision the missing transverse energy \(E_{\text{miss}}^{\text{T}}\), that is that part of momentum that lacks in the transverse plane of the beam and that has to be added in order to obtain an overall balance of zero \(p_T\), consistent with the conservation law. When possible sources of mismeasurement due to imperfections in the detector are corrected, a non-zero missing energy represents the most common signature of non-detectable particles, like SM neutrinos or neutral long lived particles included in some topologies of physics beyond the Standard Model.

According to the test-beam results, the expected energy resolution for single pions interacting in the central part of the calorimeter is:

\[
\frac{\sigma_E}{E} = 94\% \sqrt{E} \oplus 4.5\%
\]

where the energy is measured in GeV. An important degradation of the resolution is expected at \(|\eta| = 1.4\), due to the presence of services and cables. The performance of the very forward
The CMS detector at the LHC

2. The CMS detector at the LHC

2.2.4 The muon system

The CMS muon system [28] is placed outside the magnetic coil, sustained by the iron return yoke of the magnet itself. The return field (whose value goes from \( \sim 1.8 \) T in the barrel region to \( \sim 2.5 \) T in the endcaps) allows the measurement of the momentum and charge: this is particularly important for muons with transverse momentum in the TeV range, for which the complementary tracker measurement degrades. This subdetector consists of four layers of muon chambers in the barrel part and four in the endcap region, each one providing track segments reconstruction from few distributed hits. These tracks will be later combined with the informations coming from the tracker to form a complete muon track. A sketch of the muon system is represented in figure 2.7.

![Figure 2.7: View in \((r, z)\) of a quarter of the CMS muon system layout.](image)

In the barrel part, the four layers of muon chambers follow the cylindrical geometry of CMS, arranged in such a way that a muon traverses at least three of them. They are segmented on the \(z\) coordinate by the 5 wheels of the yoke and divided in 12 sectors on the \(\phi\) plane, for a global coverage from \(|\eta| = 0\) to \(|\eta| \approx 1.3\). Each chamber is made of 12 layers of Drift Tubes (DTs) grouped in 3 sub-units called superlayers. The CMS DTs are gaseous detectors consisting of a long aluminium cell filled with gas, with an anode wire in the center that collects ionization charges. Two of the superlayers have anode wires parallel to the beam line, providing a measurement of the \(r\) and \(\phi\) coordinates; the third one is placed perpendicularly between the others and provides the \(z\) measurement. Each station is designed to measure muon positions with a 100 \(\mu\)m resolution in the \((r-\phi)\) plane and 150 \(\mu\)m in the \((r-z)\) plane.

The endcaps are equipped with Cathode Strip Chambers (CSCs) covering the pseudorapidity interval from \(|\eta| = 0.9\) to \(|\eta| = 2.4\). CSCs are multi-wire proportional chambers, shaped in a trapezoidal form which follows the endcap geometry. Inside the chamber each cathode plane is...
segmented intro strips running across wires: when a muon crosses the chamber, the avalanche developed on a wire induces a charge on several strips of the cathode plane. In every CSC, two coordinates per plane are therefore made available by the simultaneous and independent detection of the signal induced by the same particle on the wires and on the strips.

Resistive Plate Chambers (RPCs) are coupled to the previously described detectors both in the barrel and in the endcap. They are gaseous parallel-plate detectors made of bakelite and operating in avalanche mode. Their space resolution is very poor, due to the large width of the strips in which each plane is segmented (from 2.3 to 4.1 cm); however they have an exceptional time resolution ($\sim 3$ ns), comparable to that of scintillators, and therefore they are used mainly for triggering purposes and for an unambiguous identification of the bunch crossing.

2.2.5 The CMS trigger

At the nominal LHC luminosity, the expected event rate is about $10^9$ Hz. Storing all these collision events is definitely impractical and, even if it were possible to do so, the dominating fraction of them is made of soft pp interactions, which are not interesting for the CMS Physics program. Therefore, a trigger system [29] has been developed with the purpose of providing a large rate reduction factor, whilst maintaining an high efficiency on potentially useful events. The total output rate is reduced by about seven order of magnitudes to $\mathcal{O}(100)$ Hz thanks to a two-level system: a Level-1 (L1) Trigger, which consists of custom-designed, largely programmable electronics, and an High Level Trigger (HLT), which is a software system implemented in a farm of about one thousand commercial processors.

The Level-1 Trigger is hardware-based and has been designed to analyze each 25 ns bunch crossing on a coarse-grain scale, within a latency of no more than $3.2 \mu$s. To deal with these short timescales, it employs the calorimetric and muon informations only, since the tracker algorithms are too slow for this purpose. Therefore, the L1 trigger is organized into a calorimeter and a muon trigger, whose information is transferred to the global trigger which takes the accept-reject decision.

The calorimeter trigger is based on trigger towers (arrays of 5 crystals in ECAL which match the granularity of the HCAL towers) grouped in calorimetric regions each one consisting of $4 \times 4$ towers. The calorimeter trigger identifies the best four candidates for each of the following classes: electrons and photons, central jets, forward jets and $\tau$-jets. The information is then passed to the global trigger, together with the measured $E_T^{\text{miss}}$.

The muon trigger is performed separately for each muon detector. The information is then merged and the best four muon candidates are transferred to the global trigger.

At the end, the rate of the selected events is reduced down to $\mathcal{O}(100)$ kHz.

The High Level Trigger completes the reduction of the output rate down to $\mathcal{O}(100)$ Hz. The idea of the HLT software is the regional reconstruction on demand, that is only those objects in the useful regions are reconstructed and the uninteresting events are rejected as soon as possible. To the first level, using only the muon system and the calorimeters, is added a second level, where the information coming from the tracker pixels is made available, and a third one, exploiting the full event information. Flexibility is maximized since there is complete freedom in the selection of the data to access, as well as in the sophistication of the algorithms, usually referred to as “HLT paths”.

2. The CMS detector at the LHC

2.3 Object reconstruction at CMS

In this section, a description of the most important high-level physics objects used in the CMS analyses and the respective reconstruction efficiencies is presented.

2.3.1 Electron reconstruction

For each electron reaching the ECAL surface, an electromagnetic shower develops within the first centimeters of the ECAL crystals. Most of the electron energy is, in general, collected within a small matrix of crystals centered around the one which has been hit, but the situation could be much more complicated: the tracker material budget, as already discussed in subsection 2.2.1, causes electrons to loose part of their energy radiating photons by bremsstrahlung effect. As the electrons are degraded in energy, the effect of the magnetic field is to enhance the bending of their trajectories, which ultimately results in a spread of irradiated photons along the azimuthal $\phi$ coordinate. Therefore, to obtain an accurate measurement of the electron energy in correspondence of the primary vertex and minimize the cluster containment variations, it is essential to collect bremsstrahlung photons.

This is the purpose of the first stage in the electron reconstruction sequence, which goes under the name of superclustering: some algorithms have been developed to group electron crystals and recollect bremsstrahlung photons [30]. The starting point is a “seed”, that is a single crystal with at least 1 GeV of deposited transverse energy. In the barrel, the algorithm looks for $1 \times 3$ or $1 \times 5$ dominoes of crystals in the $\eta - \phi$ plane, each with a total energy of at least 100 MeV. The dominoes are aligned with the seed crystal along $\eta$ and extend up to $\pm 17$ crystals away from the seed one. Different dominoes are then clustered together along $\phi$, where a valley with less than 100 MeV of deposited energy separates different clusters. The so-obtained cluster of clusters goes under the name of supercluster (see figure 2.8).

![Figure 2.8: Illustration of the clustering algorithm in the ECAL barrel. Picture from [30].](image)

In the endcap, clusters are built connecting rows of crystals along $\phi$ containing energies decreasing monotonically when moving away from the seed crystal. Superclusters are then built collecting other clusters along a $\phi$ road in both directions around each clusters.

Once a supercluster is found, the reconstruction proceeds with the track-building stage. Under both $+1$ and $-1$ charge hypotheses, the supercluster energy and position are back-propagated
in the magnetic field to the nominal vertex, to look for compatible hits in the pixel detector. Once a pair of compatible hits is found, an electron pre-track seed is built. Starting from seeds, compatible hits are searched for on the next available silicon layers. In this pattern-recognition problem, the probability of major energy losses due to bremsstrahlung emission has to be taken into account. Therefore, a dedicated algorithm has been developed, where the electron energy loss pdf, well described by the Bethe-Heitler model, is approximated with a sum of Gaussian functions, in which different components model different degrees of hardness of the bremsstrahlung in the layer under consideration. This procedure, known as Gaussian Sum Filter (GSF) [31], is iterated until the last tracker layer, unless no hit is found in two subsequent layers. A minimum of five hits is finally required to create a track.

Figure 2.9: In the left plot the fractional resolution (effective RMS) is plotted as a function of generated energy $E$ as measured with the ECAL supercluster (downward arrows), the electron track (upward arrows) and the combined track-supercluster (circles). In the right plot correlations between ECAL energy and tracker momentum measurements in the $\eta$ range of the barrel are shown [32].

In the final stage, the supercluster and track information are merged. The energy measurement $E_{sc}$ provided by the electromagnetic calorimeter can be combined with the tracker momentum measurement $p_{tk}$ to improve the estimate of the electron momentum at the interaction vertex for low energy particles. The improvement is expected to come both from the opposite behaviour with $E$ or $p$ of the intrinsic calorimetry and tracking resolutions, and from the fact that $p_{tk}$ and $E_{sc}$ are differently affected by the bremsstrahlung radiation (see figure 2.9).

2.3.2 Muon reconstruction

The standard muon reconstruction sequence is performed in three stages: a local reconstruction inside every muon subdetector, a standalone reconstruction in the muon system and a global reconstruction in the whole detector [32].

In the first stage, local pattern recognition algorithms start from single hits and build track stubs separately in each subdetector (CSCs in the endcap and DTs in the barrel): the result is a three-dimensional segment associated with a single muon layer. In the second stage, track parameters are propagated from one to the next layer of the detectors, working from inside out and taking into account material effects (ionization and bremsstrahlung). A suitable $\chi^2$ cut is applied to reject bad hits and the procedure is iterated until the outermost surface of the muon system is reached. Inclusion of RPC measurements helps in the reconstruction of low $p_T$ muons and of those which escaped through the inter-spaces, leaving only one fired DT station. The final track is then extrapolated to the point of closest approach to the beamline. Due to the large amount of material traversed to reach the muon spectrometers, the momentum resolution
as measured in the muon chambers is degraded by multiple scattering. In the last stage (global reconstruction), muon trajectories are extended until the outermost layer of the tracker system (silicon strips + pixel) and a $\eta - \phi$ region of interest is defined. The compatibility between the muon track and the track parameters of the reconstructed silicon trajectories is checked on a $\chi^2$ basis and, if the result is found in agreement, a global fit is performed with all the hits (tracker + muon).

Figure 2.10 shows how the additional information provided by the muon tracking system is precious for the momentum reconstruction of high-energy muons ($p \gtrsim 100 \text{ GeV}/c$), for which the tracker-only momentum measurement degrades. For low and medium-$p_T$ muons, instead, the resolution of the tracking system is dominating.

![Figure 2.10](image.png)

**Figure 2.10:** The muon momentum resolution versus $p$ using the muon system only, the inner tracker only, or both (“full system”). (a) Barrel, $|\eta| < 0.2$; (b) Endcap, $1.8 < |\eta| < 2$.

### 2.3.3 Particle Flow reconstruction for jets and missing transverse energy

The Particle Flow (PF) is a whole-event reconstruction technique whose purpose is the reconstruction and identification of all the particles produced in each pp collision (namely charged hadrons, photons, neutral hadrons, muons and electrons) with an optimized combination of all the sub-detector informations [33]. The resulting list of particles can then be used to construct a variety of higher-level objects and observables such as jets, missing transverse energy ($E_T^{\text{miss}}$), taus, etc.

While no substantial changes are expected for the reconstruction of high-energy electrons and muons, the Particle Flow allows to significantly improve the resolution of jets and $E_T^{\text{miss}}$ with respect to a standard, pure calorimetric jet reconstruction. Since only about the 15% of a jet energy is carried by neutral, long-lived hadrons (neutrons, $\Lambda$ baryons, etc.), for the remaining 85% carried by charged particles the coarse HCAL information can be combined with the more precise tracker momentum measurements, thus allowing for a largely better jet reconstruction.

Photons (e.g. coming from $\pi^0$ decays or from electron bremsstrahlung) are identified as ECAL energy clusters not linked with the extrapolation to ECAL of any charged particle trajectory. Electrons (e.g. coming from photon conversions in the tracker material or from leptonic decays) are identified as a primary charged particle track and potentially many ECAL energy clusters
which correspond to the track extrapolation on ECAL and to possible bremsstrahlung photons emitted along the way through the tracker material. Muons (e.g. from hadron semi-leptonic decays) are identified as a track in the central tracker consistent with another track in the muon system, associated with an energy deficit in the calorimeters. Charged hadrons are identified as charged particle tracks neither identified as electrons, nor as muons. Finally, neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory, or, if they overlap with charged particles in the calorimeters, as ECAL and HCAL energy excesses with respect to the expected charged hadron energy deposit.

The energy of photons is directly obtained from the ECAL measurement, corrected for zero-suppression effects. The energy of electrons is determined from a combination of the track momentum at the main interaction vertex, the corresponding ECAL cluster energy, and the energy sum of all bremsstrahlung photons associated to the track. The energy of muons is obtained from the corresponding track momentum. The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energy, corrected for zero-suppression effects and calibrated for the nonlinear response of the calorimeters. Finally the energy of neutral hadrons is obtained from the corresponding calibrated ECAL and HCAL energy.

Figure 2.11 shows the composition of a typical minimum-bias event in terms of different particle types. In the central part of the detector, where the tracker allows for charge measurements, the largest fraction of an event energy is carried by charged hadrons (∼65%). Only about 2% is carried by electrons, with neutral hadrons and photons almost equally sharing the remaining part. Outside the tracker acceptance, instead, no distinction can be made between charged and neutral particles. Here, the vast majority of the event energy is carried by hadronic candidates, with purely electromagnetic objects contributing a 10% or less.

The PF approach to the event reconstruction also allows for a natural definition of jet objects.
Once final state, well isolated leptons are excluded from the particle list, all the remained objects can be clustered into jets, as further explained in the following paragraph. The advantage of this approach is that jets and leptons are naturally disentangled, since the same energy deposits or tracker hits cannot belong to different physics objects.

**Jets**

A high-energy, coloured quark or gluon emitted in a hard proton-proton collision does not in the end appear in the detector. As it reaches large distances from the rest of the proton, the strong force potential favours the radiation of softer, often collinear gluons and quarks, until it is reached a point where a non-perturbative transition causes the partons to combine into colourless hadrons. The result is a spray of more-or-less collimated particles, referred to as jet, which, due to energy conservation, reflects at some level the energy and the flight direction of the initial parton.

Jets are detectable in modern experiments as a cluster of tracks and energy deposits in a defined region of the detector, merged together following the prescriptions of some sort of algorithm. Given the fact that the cross section for producing an extra gluon in the final or initial state is divergent in the soft ($p_T \to 0$) and collinear limit, jet algorithms must satisfy a few requirements, so that they can be used to provide finite theoretical predictions. The two conditions to be respected are the following:

- **collinear safety**: the outcome of the jet algorithm must not change if a particle of momentum $p$ is substituted by two collinear particles of momentum $p/2$;
- **infrared safety**: the outcome of the jet algorithm must not change if an infinitely soft particle is added (or subtracted) to the list of particles to be clustered.

In CMS, the adopted clustering algorithm, which respects the two criteria above described, is the so-called anti-$k_T$ [34]. This algorithm proceeds via the definition of two distances for each particle $i$ in the list of particles, namely:

\[
\begin{align*}
    d_{ij} &= \min \left( \frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2} \right) \frac{\Delta R_{ij}^2}{R^2} \\
    d_{iB} &= \frac{1}{p_{T_i}^2}
\end{align*}
\]  

(2.2)

$d_{ij}$ can be interpreted as the “distance” between the particle $i$ and a generic other particle $j$ among those still to be clustered, while $d_{iB}$ represents the “distance” between the particle $i$ and the beam line. $\Delta R_{ij}$ is the distance between the two particles in the $\eta - \phi$ plane, while $R$ is the algorithm radius parameter.

The algorithm looks, for each particle $i$, if there is another particle $j$ such that $d_{ij}$ is smaller than $d_{iB}$. If this happens, then particles $i$ and $j$ are merged by adding together their four-momenta, otherwise the $i$ particle is promoted to jet. The whole procedure is iterated until there are no more particles to merge.

It can be easily seen that particles at a distance greater than $R$ from the jet axis are not clustered together with the jet itself, thus leading to the construction of cone-shaped jets. The standard radius parameter adopted in CMS, and then the approximate jet size in the $\eta - \phi$ plane, is 0.5.
2.3 Object reconstruction at CMS

The jet momentum is determined as the vectorial sum of all particle momenta in it. Although important corrections are already applied at particle level during PF reconstruction, a set of further corrections have to be applied on reconstructed jets so that they can be used as high-level physics objects. The jet correction scheme adopted in CMS is factorized into subsequent steps, each of them addressing a different physic aspect.

- **Level 1** (offset) corrections: the purpose of this first step is to remove from the jet the additional energy coming from spurious particles produced in secondary pp interactions within the same bunch crossing or from the underlying events that randomly overlap with the jet area. This correction is determined both in data and in Monte Carlo on an event-by-event basis. First of all, the charged component of a jet within the tracker acceptance can be removed from the jet, calculating the impact parameter of all jet particles: those which are not compatible with the event primary vertex are not considered in the jet clustering algorithm. To further remove the contribution of neutral particles, or to correct jets with $|\eta| > 2.4$, a different technique is used. All Particle Flow candidates are re-clustered implementing a different algorithm ($k_T$ instead of anti-$k_T$) and after adding a large number of very soft “ghost” particles uniformly to the event. The median energy density $\rho_{PU} = E/\Delta\eta/\Delta\phi$ of the many pseudo-jets so produced is taken as the estimate of the pile-up plus underlying event energy density for that event, and is subtracted from real jets, after being multiplied for the jet area (roughly $\pi R^2$) [35];

- **Level 2** (relative) corrections: these corrections are meant to correct for non-uniformities in the different CMS sub-detectors by equalizing the jet response along $\eta$ to the center of the barrel;

- **Level 3** (absolute) corrections: this last correction factor correctly sets the jet absolute energy scale, and is derived from $\gamma$+jet events, where the event energy balance allows to compare the jet energy to the photon energy, which is measured in ECAL with a better precision.

Level 2 and 3 corrections are derived in simulated events, and further checked on real data via a closure test. Potential differences between data and Monte Carlo are accounted for with residual correction factors, applied to jets in real data.

As an example, figure 2.12 reports the jet energy resolution expected from the simulation and measured in data for PF jets reconstructed with an anti-$k_T$ algorithm having an $R$ parameter equal to 0.5, within the tracker acceptance.

**Missing transverse energy**

In general, $E_T^{\text{miss}}$ is defined as the negative of the vector sum of the transverse momenta of all final-state particles in the event. In the hypothesis that all detectable particles are properly reconstructed, $E_T^{\text{miss}}$ coincides with the sum of the four-momenta of all undetectable particles (i.e. neutrinos, or BSM particles such as neutralinos in more exotic scenarios), since the initial pp collision occurs between two particles of negligible transverse momentum ($\lesssim 1 \text{ GeV}$). In practice, this is not true because the detector is not perfectly hermetic and a fraction of the total event energy is unavoidably lost in the beam pipe or only coarsely reconstructed in the

---

3 The median is used since it is only little affected by the presence of few real high-energy jets in the event.
forward calorimeters. As a result, the measured $E_T^{miss}$ is only an approximation of the transverse momentum of the undetected particle.

Since all detector information is included in the PF-based reconstruction, it is simple to define PF $E_T^{miss}$ as the negative of the vector sum, over all PF candidates particles, of their transverse momenta. Jet factorized energy corrections are propagated to the missing transverse energy computation in order to improve its resolution.

### 2.4 The 2011 proton-proton run

During 2011, LHC collided protons from March until the end of October, operating at a center-of-mass energy equal to 7 TeV (3.5 TeV per beam). During stable beams conditions, the total integrated luminosity delivered to the experiments was 6.10 fb$^{-1}$, out of which 5.56 fb$^{-1}$ were recorded by CMS. At the end of 2011, the amount of all the data taking periods and the luminosity sections which were certified as “good” (meaning that all the sub-detectors were working correctly) gives a total integrated luminosity equal to $(5.0 \pm 0.1) \text{fb}^{-1}$, which is the value used for the analysis presented here.

The plot in figure 2.13 shows the integrated luminosity versus time delivered to, and recorded by CMS during 2011.

The recorded and certified data are divided into two main running periods, called Run2011A and Run2011B. The former extends up to the September technical stop, and can be further divided into four sub-periods, while the latter covers the data taking operations from September to the end of the proton run.
The instantaneous luminosity, as can be seen from figure 2.14, ranges from $\sim 5 \cdot 10^{32}$ cm$^{-2}$s$^{-1}$ at the beginning of the 2011 data taking, to $\sim 4 \cdot 10^{33}$ cm$^{-2}$s$^{-1}$, gaining a factor of about ten.

Figure 2.13: Integrated luminosity versus time for the 2011 proton-proton run.

Figure 2.14: Maximum instantaneous luminosity per day versus time for the 2011 proton-proton run.
Chapter 3

Description of the experimental final state

In this chapter, a detailed description of the final state on which this analysis focuses is given. The first section introduces the main advantages of the WW semi-leptonic decay channel, namely its high branching ratio and the possibility of reconstructing a peak in the Higgs invariant mass distribution. In section 3.2 the signals and the main backgrounds are presented, while in section 3.3 the datasets used are reported. In section 3.4 the cuts imposed on the physical observables, both at reconstruction level and at analysis level, are discussed. Section 3.5 explains in detail the procedure for reconstructing the invariant mass of the Higgs boson. The evaluation of the reconstruction, identification and trigger efficiencies is pointed out in section 3.6.

3.1 Overview of the channel

As already mentioned in section 1.3.2 and illustrated in figure 1.9(a), when the Higgs mass is sufficiently high to kinematically allow its decay into a pair of massive W or Z bosons, these decay modes become the dominant ones. If we follow the subsequent decays of the W and Z and consider the subset of the experimentally-allowed final states, we find the results summarized in table 3.1 for a reference mass $m_H = 400\text{ GeV/c}^2$ and graphically represented in figure 3.1 as a function of the Higgs mass.

<table>
<thead>
<tr>
<th>H decay mode</th>
<th>$B$</th>
<th>$B/B(H \rightarrow l\nu q\bar{q})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WW $\rightarrow l\nu q\bar{q}$</td>
<td>$1.7 \cdot 10^{-1}$</td>
<td>1</td>
</tr>
<tr>
<td>WW $\rightarrow l\nu l\nu$</td>
<td>$2.6 \cdot 10^{-2}$</td>
<td>1/6</td>
</tr>
<tr>
<td>ZZ $\rightarrow l^+ l^- q\bar{q}$</td>
<td>$2.5 \cdot 10^{-2}$</td>
<td>1/7</td>
</tr>
<tr>
<td>ZZ $\rightarrow l^+ l^- \nu \bar{\nu}$</td>
<td>$7.2 \cdot 10^{-3}$</td>
<td>1/24</td>
</tr>
<tr>
<td>ZZ $\rightarrow l^+ l^- l^+ l^-$</td>
<td>$1.2 \cdot 10^{-3}$</td>
<td>1/144</td>
</tr>
</tbody>
</table>

**Table 3.1:** Absolute and relative-to-$l\nu q\bar{q}$ branching ratio values for a 400 GeV/c² Higgs boson decaying into a couple of vector bosons. The branching ratios for the diboson decays are taken from [15], while for the W and Z decays only final states with $l = e, \mu$ are considered.

The diboson Higgs decay in a pair of W is kinematically favoured compared to the one in a pair of Z. Excluding the fully-hadronic decay $H \rightarrow WW \rightarrow q\bar{q} q\bar{q}$, which is extremely challenging to address at a hadron collider such as LHC, it is clear that the channel with the largest branching
3. Description of the experimental final state

Figure 3.1: Accessible final states of Higgs boson decays and their respective cross section times branching ratio values (with \( l = e, \mu \)). Picture from [15].

The starting point of the analysis is therefore to select the events with one well identified and isolated lepton (electron or muon) and large missing transverse energy coming from the decay of the leptonic W, together with at least two high-\( p_T \) jets originating from the fragmentation of the q\( \bar{q} \) pair from the hadronic W decay. To increase the signal efficiency, events with up to three reconstructed jets are retained. The main experimental issue is to control the dominating W+jets background sufficiently well, in order to establish if the observed data are more compatible with the signal+background hypothesis or with the background only hypothesis.

### 3.2 Signal and backgrounds

The total pp cross section at the LHC, for an energy in the center of mass equal to 7 TeV, is about 100 mb (see figure 3.2). This implies an interaction rate of about \( 10^8 \) Hz when running at an instantaneous luminosity of \( 1 \cdot 10^{33} \text{cm}^{-2} \text{s}^{-1} \), similar to the one reached by LHC during the 2011 run (see section 2.4). In most of the cases, these events consist in long-distance collisions between the two incoming protons: the momentum transferred is small and particle scattering at large angle is suppressed. The particles produced in the final state of such interactions have large longitudinal momentum but small transverse momentum (\( p_T \simeq 500 \text{MeV}/c \)) and most of the collision energy escapes down the beam pipe, undetected. These soft scattering events represent by far the majority of pp interactions, but are of no interest. Much more interesting are the events where there is a so-called hard scattering between two partons and the resulting products have an high transverse momentum: it is in these kind of events that massive particles are produced and where one hopes to find signs of new physics.

A pictorial representation of an hard scattering process is shown in figure 3.3: the final state
3.2 Signal and backgrounds

Figure 3.2: Standard Model cross sections at the Tevatron and LHC colliders for some reference processes. Note the logarithmic scale on the vertical axis.

Figure 3.3: A graphical representation of a generic hard scattering process with undefined final states.

could be the production of a massive vector boson, a lepton-antilepton pair (Drell-Yan process), heavy quarks hadronizing into jets, and so on. \( H_1 \) and \( H_2 \) are the two incoming hadrons (protons in our case), each one with a momentum \( p_i \). The two partons who interact carry a fraction \( x_i p_i \) of the hadron momentum, where \( x_i \) ranges between 0 and 1.

The parton model formalism allows to write the cross section for a generic hard scattering process as [36, 37]:

\[
\sigma_{H_1H_2}(p_1,p_2) = \sum_{i,j \text{ partons}} \int_0^1 \, dx_1 dx_2 \, f_i^{(H_1)}(x_1, \mu_F^2)f_j^{(H_2)}(x_2, \mu_F^2) \, \hat{\sigma}_{ij}(x_1 p_1, x_2 p_2, \alpha_S(\mu_R^2), \mu_R^2)
\]

(3.1)
The short distance partonic cross section $\hat{\sigma}$ is calculable as a perturbative expansion in the running coupling constant $\alpha_S$, which depends upon the renormalization scale $\mu_R$. The partonic cross section is process-dependent and is a function of the effective center-of-mass energy $\hat{s} = x_a x_b s$. The $f_i$ are probability density functions (PDFs) which describe the probability of finding a parton $i$ carrying a fraction $x_i$ of the hadron longitudinal momentum: they represent the non-perturbative part of equation 3.1 and, due to their universality (i.e. they do not depend upon the particular process considered), are usually extracted from deep inelastic scattering data. $\mu_F$ is a factorization scale, which can be thought of as the scale that separates the long and short-distance physics.

### 3.2.1 Standard Model Higgs signal

What figure 3.2 immediately suggests is that the expected cross section for the production of a Standard Model Higgs boson is orders of magnitude ($\sim 10^4$ to $10^5$) smaller than the electroweak production of $W$ bosons. When this cross section is multiplied by the branching ratio of an Higgs boson decaying into a pair of $W$ bosons and, subsequently, into a lepton-neutrino and a di-jet pair, the effective cross sections for the signal processes relevant for this analysis are obtained, as listed in table 3.2 as a function of the Higgs mass.

All the cross sections and the branching ratios are provided by the “LHC Higgs Cross Section Working Group” [15].

<table>
<thead>
<tr>
<th>$m_H$ [GeV/c$^2$]</th>
<th>$\sigma_{gg} \times B(H \rightarrow l\nu q\bar{q})$ [pb]</th>
<th>$\sigma_{VBF} \times B(H \rightarrow l\nu q\bar{q})$ [pb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>$0.67^{+0.11}_{-0.10}$</td>
<td>$0.087^{+0.003}_{-0.002}$</td>
</tr>
<tr>
<td>300</td>
<td>$0.48^{+0.08}_{-0.07}$</td>
<td>$0.060^{+0.003}_{-0.002}$</td>
</tr>
<tr>
<td>350</td>
<td>$0.45^{+0.09}_{-0.07}$</td>
<td>$0.0416^{+0.002}_{-0.0012}$</td>
</tr>
<tr>
<td>400</td>
<td>$0.34^{+0.05}_{-0.06}$</td>
<td>$0.0272^{+0.0016}_{-0.0008}$</td>
</tr>
<tr>
<td>450</td>
<td>$0.22^{+0.03}_{-0.04}$</td>
<td>$0.0197^{+0.0013}_{-0.0006}$</td>
</tr>
<tr>
<td>500</td>
<td>$0.13^{+0.02}_{-0.02}$</td>
<td>$0.0150^{+0.0011}_{-0.0005}$</td>
</tr>
<tr>
<td>550</td>
<td>$0.084^{+0.016}_{-0.015}$</td>
<td>$0.0117^{+0.0009}_{-0.0004}$</td>
</tr>
<tr>
<td>600</td>
<td>$0.053^{+0.010}_{-0.010}$</td>
<td>$0.0093^{+0.0008}_{-0.0004}$</td>
</tr>
</tbody>
</table>

Table 3.2: Cross section values and their uncertainties for Higgs boson production via the gluon fusion (left) or vector boson fusion (right) mechanism, multiplied by the branching ratio of the $l\nu q\bar{q}$ final state (with $l = e, \mu$), as a function of the Higgs mass. Values taken from [15].

Of all the four production mechanisms described in section 1.3.1, the Higgs-strahlung and the $t\bar{t}$ associated production are not considered because their contribution is negligible (see the plot reported in figure 1.7). The requirement of having at most three reconstructed jets is expected to reduce the fraction of signal events produced via vector boson fusion, which carry two extra jets coming from the hadronization of the forward scattered quarks. This contribution is nevertheless included in the signal samples and, at the end of the selections, is found to be responsible for about 10% of the total signal events, as will be reported in section 4.1.2 for some reference Higgs mass hypotheses.
3.2 Signal and backgrounds

3.2.2 Backgrounds

A certain number of known SM processes can give an experimental signature similar or identical to the signal we are looking for, therefore resulting in a contamination of the data sample. Generally speaking, all the physical processes with a lepton (either “real” or reconstructed as a lepton by the CMS algorithms), jets and missing transverse energy can be considered as a source of background events which can pass the selection requirements of the analysis. In the following, the most important ones are listed:

- **\( W(\rightarrow l\nu) + \text{jets} \):** this is the production of a single W vector boson decaying leptonically, in association with the radiation of quarks or gluons that can mimic the final state signature if reconstructed as the products of an hadronic W decay (see figure 3.4). Because of the huge production cross section, this background is by far the dominant one.

![W+jets Diagram](image)

**Figure 3.4:** W+jets leading order diagrams. Even if, for the sake of clarity, only the leading order Feynman diagrams have been shown in the figure, these background events may exhibit an higher multiplicity of jets.

- **Drell-Yan \( Z/\gamma^*(\rightarrow l^+l^-) + \text{jets} \):** this is the production of single Z/\(\gamma^*\) bosons in association with quarks or gluons, where one lepton is undetected because of acceptance or inefficiency effects, and the hadronic activity mimics the final state signature and the hadronic W decay products. The leading order diagrams are the same of figure 3.4, with a Z/\(\gamma^*\) boson instead of the W.

- **Diboson:** the production of vector boson pairs constitutes an irreducible background for the analysis. Three different production channels are considered:
  1. **WW:** the final state originating from the non-resonant production of W boson pairs is topologically indistinguishable from signal events and only kinematic selections can help in its reduction.
  2. **WZ:** if the Z decays hadronically and the W leptonically, or if the W decays hadronically and one lepton from the Z decay is not identified by the detector, this process also contributes to the backgrounds.
  3. **ZZ:** the production of Z pairs can originate the same final state of the analysis in case one Z decays hadronically, and one lepton from the other Z boson is not identified by the detector. The contamination level from such events is almost irrelevant due to its tiny cross section, when compared to the other leading backgrounds.

Some leading order diagrams which contribute to the diboson production are represented in figure 3.5.
3. Description of the experimental final state

- **$t\bar{t}+\text{jets}$**: top quark pairs are produced at LHC via the gluon fusion process $gg \rightarrow t\bar{t}$ or via quark annihilation $q\bar{q} \rightarrow t\bar{t}$, as illustrated in figure 3.6. The contamination arising from the semi-leptonic decays of the $t\bar{t}$ pairs can be reduced with requirements on the number of jets and vetoing the presence of bottom-like jets (anti $b$-tagging). The contamination coming from fully-leptonic decays is reduced by requiring the presence of only one good lepton in the event. In any case, because of acceptance and inefficiencies, this background still contaminates the signal phase space, being the second leading one after $W+\text{jets}$.

- **Single top**: top quarks can be produced through three separate channels:
  1. $t$-channel: the top quark is produced after an interaction between a generic quark and a bottom quark, with the exchange of a virtual $W$;
  2. $s$-channel: the top quark is produced in association with an anti-bottom, after the annihilation of a pair of quarks in a weak vertex;
  3. $tW$-channel: the top quark is produced in association with a charged vector boson in a weak process, from a gluon-bottom pair in the initial state.

Some leading order diagrams which contribute to the single top production are represented in figure 3.7.

- **QCD**: to the previously listed backgrounds the one coming from QCD multijet production has to be added, that is the sum of all the processes described by interactions like $gg \rightarrow gg$, $gg \rightarrow q\bar{q}$, $gg \rightarrow qg$ and so on. Even if none of these events contains “real” leptons in the final state, from the experimental point of view it could happen to have “fake” leptons, that is physical objects (mainly pions, kaons or even protons) which are wrongly reconstructed by the detector. Despite the fact that the mis-reconstruction probability is quite low, the fake rate could still be sizeable, because of the huge multijet production cross section at a hadron collider. Moreover, since the fraction of fake muons reconstructed from jets is much lower than the probability of having fake electrons, QCD contamination is more important in the $e\nu_e,q\bar{q}$ final state.
The cross sections for the various background processes taken into account, multiplied by the branching ratios, are reported in table 3.3 below.

<table>
<thead>
<tr>
<th>process</th>
<th>$\sigma$ [pb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W(\rightarrow l\nu)+$jets</td>
<td>$31300 \pm 1600$</td>
</tr>
<tr>
<td>$Z(\rightarrow ll)+$jets</td>
<td>$3050 \pm 130$</td>
</tr>
<tr>
<td>WW</td>
<td>$47.0 \pm 1.5$</td>
</tr>
<tr>
<td>WZ</td>
<td>$18.2 \pm 0.7$</td>
</tr>
<tr>
<td>ZZ</td>
<td>$7.10 \pm 0.15$</td>
</tr>
<tr>
<td>$t\bar{t}$+jets</td>
<td>$163 \pm 14$</td>
</tr>
<tr>
<td>$t$ ($t$-channel)</td>
<td>$41.9 \pm 1.8$</td>
</tr>
<tr>
<td>$\bar{t}$ ($t$-channel)</td>
<td>$22.6 \pm 1.0$</td>
</tr>
<tr>
<td>$t$ ($tW$-channel)</td>
<td>$7.9 \pm 0.6$</td>
</tr>
<tr>
<td>$\bar{t}$ ($tW$-channel)</td>
<td>$7.9 \pm 0.6$</td>
</tr>
<tr>
<td>$t$ ($s$-channel)</td>
<td>$3.19 \pm 0.14$</td>
</tr>
<tr>
<td>$\bar{t}$ ($s$-channel)</td>
<td>$1.44 \pm 0.07$</td>
</tr>
<tr>
<td>QCD (e-enriched)</td>
<td>$\sim 6740000$</td>
</tr>
<tr>
<td>QCD ($\mu$-enriched)</td>
<td>$\sim 84700$</td>
</tr>
</tbody>
</table>

Table 3.3: Cross section values for the considered backgrounds, multiplied by the branching ratio when meaningful (in this table $l = e$, $\mu$, $\tau$).

### 3.3 Datasets

#### 3.3.1 Data samples

The analysis here described has been applied to the data recorded by CMS during all 2011. As already discussed briefly in section 2.4, this period is divided into two main subsets (called Run2011A and Run2011B, respectively) and the total analyzed statistics corresponds to an integrated luminosity equal to $(5.0 \pm 0.1)$ fb$^{-1}$. The datasets in which the events are stored are labelled as SingleElectron or SingleMu, depending on the flavour of the lepton present in the final-state.

The analyzed data samples and the corresponding run periods are summarized in table 3.4(a) together with their integrated luminosity.
3.3.2 Monte Carlo samples

The simulated Monte Carlo samples belong to the official CMS production and contain events of signal (Standard Model Higgs boson, for both the relevant production mechanisms: gluon fusion and VBF) and background (electroweak and QCD induced). The list of signal and background samples used in the analysis is reported in table 3.4(b), together with the equivalent luminosity available for the study.

Given the characteristics of the analysis, only the high mass Higgs hypotheses are considered, with $m_H$ ranging from 250 GeV/$c^2$ to 600 GeV/$c^2$.

The POWHEG-BOX generator [38] is used to produce signal Higgs events with NLO accuracy. Although the analysis requires the presence of an electron or a muon in the final state, the samples in which the leptonic W boson goes into a $\tau \nu_\tau$ pair are taken into account, too. These events can have the same final state searched for in this analysis, when the $\tau$ decays leptonically. Anyway, given the small branching ratio of such decays, and the softer $p_T$ spectrum of the resulting electron or muon, they contribute only a small fraction of the total signal yield at the end of the event selection (see section 4.1.2 for a quantitative estimate).

Due to details in the implementation of the calculation, the resulting Higgs $p_T$ spectrum for the gluon fusion production mechanism is harder compared with the most precise calculation available (NNLO with re-summation to NNLL order), obtained with MCFM. Therefore, an event-by-event reweighting is applied to POWHEG-BOX signal events, based on the generated Higgs $p_T$.

For W+jets, Z+jets, and $t\bar{t}$+jets processes the leading-order MadGraph generator is used, which is able to generate vector bosons/$t\bar{t}$ pairs together with up to four partons at the matrix element level, successively matched to the parton shower. Instead, POWHEG-BOX is chosen for single-top events, generated at next-to-leading order. Finally, diboson and QCD events are generated with PYTHIA [39].

QCD samples

Due to the huge production cross section (and to finite computing resources) it is in general not practical to simulate a sufficient number of QCD multijet events surviving all the analysis selections. As can be seen from table 3.4(b), the equivalent luminosity of the QCD samples is in fact much lower than the total integrated luminosity of the data (5 fb$^{-1}$). The QCD background estimation, therefore, does not rely on the Monte Carlo simulation, but it is based on a data-driven approach that will be described in details in section 4.1.1. The QCD samples reported in table 3.4(b) are used just to perform a closure test, which is done to check the consistence of the data-driven method.

W+jets samples

For what concerns the W+jets background (which, as we know from section 3.2.2, is the dominant one because of the large production cross section), the equivalent luminosity being just half of the acquired data could result in large statistical fluctuations in the physical distributions (invariant masses, $p_T$ spectra, pseudorapidity, and so on) built with this sample. For this reason, the inclusive W+jets sample listed in table 3.4(b) has been replaced with high-statistics exclusive samples belonging to the official CMS production. These samples are listed in table 3.5 together with their equivalent integrated luminosity.

The W+3jets exclusive sample was not available for the analysis, due to problems occurred in the submission of jobs on the distributed computing resources of CMS. Therefore, the events
### 3.3 Datasets

(a) Data samples

<table>
<thead>
<tr>
<th>dataset name: /SingleElectron[SingleMu]XXX/AOD</th>
<th>run range</th>
<th>( \mathcal{L} ) [fb(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/Run2011A-May10Reco-v1/</td>
<td>160431 – 163869</td>
<td>0.211 ± 0.005</td>
</tr>
<tr>
<td>/Run2011A-PromptReco-v4/</td>
<td>165088 – 167913</td>
<td>1.0 ± 0.2</td>
</tr>
<tr>
<td>/Run2011A-05Aug2011-v1/</td>
<td>170249 – 172619</td>
<td>0.390 ± 0.009</td>
</tr>
<tr>
<td>/Run2011A-PromptReco-v6/</td>
<td>172620 – 173692</td>
<td>0.71 ± 0.02</td>
</tr>
<tr>
<td>/Run2011B-PromptReco-v1/</td>
<td>175832 – 180252</td>
<td>2.72 ± 0.06</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>160431 – 180252</strong></td>
<td><strong>5.0 ± 0.1</strong></td>
</tr>
</tbody>
</table>

(b) Monte Carlo samples

<table>
<thead>
<tr>
<th>sample name: /XXX/Fall11-PU.S6.START42.V14B-v1/AODSIM</th>
<th>( \mathcal{L}^{eq} ) [fb(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/GluGluToHToWToLnuJ_7TeV-powheg-pythia6/</td>
<td>( \sim 1000 )</td>
</tr>
<tr>
<td>/VBF_HToHToWToLnuJ_7TeV-powheg-pythia6/</td>
<td>( \sim 9000 )</td>
</tr>
<tr>
<td>/WJetsToLnu_TuneZ2_7TeV-madgraph-tauola/</td>
<td>2.5</td>
</tr>
<tr>
<td>/DYJetsToLL_TuneZ2_50_7TeV-madgraph-tauola/</td>
<td>11.3</td>
</tr>
<tr>
<td>/WW_TuneZ2_7TeV_pythia6_tauola/</td>
<td>98.3</td>
</tr>
<tr>
<td>/WZ_TuneZ2_7TeV_pythia6_tauola/</td>
<td>234.0</td>
</tr>
<tr>
<td>/ZZ_TuneZ2_7TeV_pythia6_tauola/</td>
<td>710.0</td>
</tr>
<tr>
<td>/TTJets_TuneZ2_7TeV-madgraph-tauola/</td>
<td>23.5</td>
</tr>
<tr>
<td>/T_TuneZ2_t-channel_7TeV-powheg-tauola/</td>
<td>93.0</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_7TeV-powheg-tauola/</td>
<td>85.9</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_800GeV-powheg-tauola/</td>
<td>81.5</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_800GeV-powheg-tauola/</td>
<td>81.5</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_1700GeV-powheg-tauola/</td>
<td>41.1</td>
</tr>
<tr>
<td>/T_TuneZ2_s-channel_7TeV-powheg-tauola/</td>
<td>103.0</td>
</tr>
<tr>
<td>/T_TuneZ2_s-channel_800GeV-powheg-tauola/</td>
<td>95.8</td>
</tr>
<tr>
<td>/QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/</td>
<td>0.03</td>
</tr>
<tr>
<td>/QCD_Pt-20_BToE_TuneZ2_7TeV-pythia6/</td>
<td>0.02</td>
</tr>
<tr>
<td>/QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/</td>
<td>0.28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sample name: /XXX/Fall11-PU.S6.START42.V14B-v1/AODSIM</th>
<th>( \mathcal{L}^{eq} ) [fb(^{-1})]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/GluGluToHToWToLnuJ_7TeV-powheg-pythia6/</td>
<td>( \sim 1000 )</td>
</tr>
<tr>
<td>/VBF_HToHToWToLnuJ_7TeV-powheg-pythia6/</td>
<td>( \sim 9000 )</td>
</tr>
<tr>
<td>/WJetsToLnu_TuneZ2_7TeV-madgraph-tauola/</td>
<td>2.5</td>
</tr>
<tr>
<td>/DYJetsToLL_TuneZ2_50_7TeV-madgraph-tauola/</td>
<td>11.3</td>
</tr>
<tr>
<td>/WW_TuneZ2_7TeV_pythia6_tauola/</td>
<td>98.3</td>
</tr>
<tr>
<td>/WZ_TuneZ2_7TeV_pythia6_tauola/</td>
<td>234.0</td>
</tr>
<tr>
<td>/ZZ_TuneZ2_7TeV_pythia6_tauola/</td>
<td>710.0</td>
</tr>
<tr>
<td>/TTJets_TuneZ2_7TeV-madgraph-tauola/</td>
<td>23.5</td>
</tr>
<tr>
<td>/T_TuneZ2_t-channel_7TeV-powheg-tauola/</td>
<td>93.0</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_7TeV-powheg-tauola/</td>
<td>85.9</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_800GeV-powheg-tauola/</td>
<td>81.5</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_800GeV-powheg-tauola/</td>
<td>81.5</td>
</tr>
<tr>
<td>/T_TuneZ2_tW-channel_1700GeV-powheg-tauola/</td>
<td>41.1</td>
</tr>
<tr>
<td>/T_TuneZ2_s-channel_7TeV-powheg-tauola/</td>
<td>103.0</td>
</tr>
<tr>
<td>/T_TuneZ2_s-channel_800GeV-powheg-tauola/</td>
<td>95.8</td>
</tr>
<tr>
<td>/QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/</td>
<td>0.03</td>
</tr>
<tr>
<td>/QCD_Pt-20_BToE_TuneZ2_7TeV-pythia6/</td>
<td>0.02</td>
</tr>
<tr>
<td>/QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/</td>
<td>0.28</td>
</tr>
</tbody>
</table>

Table 3.4: (a) Summary of the datasets used and the corresponding run ranges and integrated luminosity for each sub-period. (b) Summary of Monte Carlo samples used in the analysis and their equivalent integrated luminosity. Higgs masses range from 250 to 600\,GeV/c\(^2\). QCD samples are divided in 20-30, 30-80, and 80-170\,GeV/c \( \hat{p}_T \) bins.
3. Description of the experimental final state

belonging to the inclusive sample have been used instead, after having applied a selection on the number of particles per event at generator level, in order to retain only the events with three additional jets.

<table>
<thead>
<tr>
<th>sample name: /XXX/Fall11-PU_S6_START42_V14B-v1/AODSIM</th>
<th>$L_{\text{eq}}$ [fb$^{-1}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>/W1Jet_TuneZ2_7TeV-madgraph-tauola/</td>
<td>13.8</td>
</tr>
<tr>
<td>/W2Jets_TuneZ2_7TeV-madgraph-tauola/</td>
<td>15.4</td>
</tr>
<tr>
<td>/W4Jets_TuneZ2_7TeV-madgraph-tauola/</td>
<td>26.6</td>
</tr>
</tbody>
</table>

Table 3.5: Summary of the Monte Carlo samples for the W+jets background.

Pile-up description

A “pile-up” happens when more than one proton pair of the colliding beams interact and the products overlap with the ones coming from the main interaction vertex, resulting in the presence of many different primary vertices in the event.

If the multiple interactions happen in the same bunch crossing, it is common to refer at them as “in-time” pile-up. Since the probability that this happens is non negligible in the 2011 data-taking conditions, this is taken into account by adding, in the Monte Carlo samples, simulated minimum-bias events on top of the hard scattering ones. When the spurious interactions come from a contiguous bunch crossing, therefore anticipated or delayed of 50 ns, we speak about “out-of-time” pile-up. This is taken into account in the simulation as well.

In data, the expected number of pile-up (PU) interactions in a given bunch crossing $i$ is expressed by the following formula:

$$\langle N_{\text{PU}}^{\text{true}} \rangle_i = \frac{L_i \cdot \sigma_{\text{min. bias}}}{\nu_{\text{orbit}}}$$

(3.2)

where $L_i$ is the instantaneous luminosity of that specific bunch crossing, $\sigma_{\text{min. bias}}$ is the cross section of minimum-bias interactions and $\nu_{\text{orbit}}$ is the LHC orbit frequency (11246 Hz). A value of 68 mb is assumed for $\sigma_{\text{min. bias}}$, as a result of a best-fit analysis comparing the distribution of the number of reconstructed vertices in data and Monte Carlo for $Z \rightarrow \mu^+\mu^-$ events.

The constant increase in instantaneous luminosity which, as mentioned in section 2.4, grew by one order of magnitude during all 2011, caused a corresponding increase in the average number of pile-up interactions per bunch crossing, which in the following will be denoted as $\langle N_{\text{PU}} \rangle$.

In figure 3.8, $\langle N_{\text{PU}}^{\text{true}} \rangle$ is plotted against time: the increase between Run2011A (up to September 2011) and Run2011B is sizeable.

In Monte Carlo, the actual number of PU interactions superimposed to the hard scattering in each event is randomly taken from a Poisson distribution, which has $N_{\text{PU}}^{\text{true}}$ as mean parameter. The distribution of the mean number of pile-up vertices in MC samples is therefore chosen to be similar to the one observed in data. In figure 3.9(b), such a distribution for a W+jets sample is shown.

Despite the pile-up added to the Monte Carlo simulation, it is necessary to reweight the MC on a event-by-event basis, in order to account for the residual differences and match the two distributions exactly.

The effectiveness of the reweighting procedure is checked observing the distribution of a particular variable, that is the number of reconstructed vertices per event $N_{\text{PV}}$, which is strongly
3.3 Datasets

Figure 3.8: The expected number of PU interactions is shown versus time, averaged on a time lapse of two days (black line). $N^{\text{true}}_{\text{PU}}$ is calculated from the instantaneous luminosity of each luminosity section as in equation 3.2. On the same plot, the integrated luminosity versus time: delivered to (solid red line), recorded by (dotted red line) and accepted by CMS (dashed red line).

Figure 3.9: Distribution of $N^{\text{true}}_{\text{PU}}$ in (a) data and (b) Monte Carlo.

correlated with the number of pile-up interactions.
In figure 3.10 the distribution of $N_{\text{PV}}$ is represented without and with the reweighting of the simulated events. As can be seen, the data-Monte Carlo agreement is much better after the reweighting procedure.

3.3.3 Trigger on data

During the first period of data taking, the $e\nu, q\bar{q}$ final state events are triggered by single electron triggers. For higher instantaneous luminosities, events are selected with a trigger based on one electron plus a transverse mass $m_T$, where the transverse mass is calculated using the decay product of the leptonic W at HLT level, i.e. the online electron, and the HLT Particle Flow
3. Description of the experimental final state

(a) Without PU reweighting

(b) With PU reweighting

Figure 3.10: Data/MC comparison of the number of reconstructed primary vertices (a) before and (b) after having applied the reweighting for pile-up.

missing transverse energy, called $H_{\text{miss}}$.

Electron candidates at trigger level fulfill some identification and isolation criteria based on cluster shape variables, energy deposits in the calorimeters and isolation observables$^1$. The additional requirement on $m_T$ is imposed in order to reduce the trigger rate down to an affordable size ($\sim 10$ Hz). The detailed cut values on the electron identification and isolation variables are reported in table 3.6.

The electron $E_T$ threshold at trigger level is 32 GeV for most of the data-taking period, except for the very beginning of Run2011A, when it was possible to trigger electrons down to 27 GeV, and for other sub-periods before the September technical stop.

The so-called VT,MT,TT working point is used for the first part of 2011, while two working points called WP80(WP70) (efficiency on real electrons $\sim 80(70)\%$) are used for the remaining periods. Depending on the data taking period, the PF $m_T$ was requested to be larger than 40(50) GeV/c$^2$.

In table 3.7(a) a summary of the electron trigger path names used and their range of validity is reported.

The $\mu\nu q\bar{q}$ final state is much easier to trigger because of the cleaner experimental signature. Single isolated muon triggers with a $p_T$ threshold of 24 GeV/c can therefore be used throughout the full data taking period (see table 3.7(b)).

To cope with the increasing interaction rate at the end of 2011, it is sufficient to restrict the pseudorapidity range of muon candidates within $|\eta| < 2.1$: this choice is harmless, because it coincides with the acceptance cut of the analysis.

3.3.4 Trigger on Monte Carlo

An accurate reproduction of the online trigger requirements is needed in order for MC samples to faithfully describe the observed data, which can be influenced by inefficiencies and/or inhomogeneities of the CMS trigger system response. However, the trigger evolution during the

$^1$All these variables are explained and discussed in section 3.4.1; here just the working points are given for reference.
3.4 Event selection

The experimental signature of the final state on which this analysis focuses is represented by an isolated lepton (electron or muon) and missing transverse energy, coming from the leptonic decay of one W boson, together with a pair of jets, coming from the hadronic decay of the other. In section 3.4.1 the requirements that the reconstructed final-state objects must satisfy in order to be identified as leptons or jets are presented. Section 3.4.2 deals with the cuts on the physical observables which are imposed event-by-event.

3.4.1 Object reconstruction

Leptons, jets, and the missing transverse energy $E_T^{\text{miss}}$ are reconstructed using the Particle Flow (PF) approach described in section 2.3.3.
Table 3.7: List of trigger paths used to select events for this analysis in the final state with (a) electrons and (b) muons. VT=T stands for CaloIdVT and is made explicit in Table 3.6, as well as the WP80(70) working points.

<table>
<thead>
<tr>
<th>Period</th>
<th>Run Range</th>
<th>Dataset Name</th>
<th>Trigger Path Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td></td>
<td>Run2011A-May10ReReco-v1</td>
<td>HLT.Ele27.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v6</td>
<td>HLT.Ele20.PPMT50.0V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v3</td>
<td>HLT.Ele20.PPMT50.0V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v1</td>
<td>HLT.Ele20.PPMT50.0V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v2</td>
<td>HLT.Ele20.PPMT50.0V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v4</td>
<td>HLT.Ele20.PPMT50.0V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v5</td>
<td>HLT.Ele20.PPMT50.0V</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v7</td>
<td>HLT.Ele20.PPMT50.0V</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td>Run2011A-PromptReco-v6</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v4</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v2</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v1</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v3</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v4</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v5</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Run2011A-PromptReco-v7</td>
<td>HLT.IsoMu24.0PM10.PP</td>
</tr>
</tbody>
</table>

VT=T stands for CaloIdVT and is made explicit in Table 3.6, as well as the WP80(70) working points.
3.4 Event selection

Tight electrons

For the $e\nu e\bar{q}$ final state, the electron candidates have to pass several requirements in order to be identified as “real” electrons:

- first of all, candidates are required to pass a simple cut-based selection, which has roughly an efficiency of 70% on real electrons ($WP70$). This electron ID relies on four variables:
  - $\sigma_{\eta_{sc}}$: the $\eta$ width of the electron candidate supercluster (sc);
  - $|\phi_{sc} - \phi_{tk}|$: the difference between the $\phi$ of the supercluster back-propagated to the vertex and the $\phi$ of the track;
  - $|\eta_{sc} - \eta_{tk}|$: the difference between the $\eta$ of the supercluster back-propagated to the vertex and the $\eta$ of the track;
  - $H/E$: the fraction of the energy deposited in the hadronic calorimeter and the one deposited in the electromagnetic calorimeter by the electron candidate.

The values of the cuts on these four variables are different for the barrel and the endcap region and are summarized in table 3.8. The cuts applied are at least as tight as the ones required at trigger level (see table 3.6);

- the electron transverse momentum is chosen to be at least 3 GeV/$c$ larger than the HLT threshold to avoid turn-on effects due to the fact that the trigger is not yet fully efficient. The electron $p_T$ cut is then 30–35 GeV/$c$, depending on the trigger period;

- the electron pseudorapidity spans over the whole tracker acceptance region, i.e. $|\eta| < 2.5$. There is an exclusion range due to the ECAL barrel-endcap transition region, defined by $1.4442 < |\eta_{sc}| < 1.566$, being $\eta_{sc}$ the pseudorapidity of the ECAL supercluster;

- in order to make sure that the selected electrons come from the primary hard interaction and not from one of the many pile-up vertices, the $z$ coordinate of the primary vertex and the $z$ coordinate of the electron inner track vertex, evaluated with respect to the beamline, are required to lie within a distance of less than 0.2 cm. A cut on the absolute value of the transverse impact parameter calculated with respect to the primary vertex is also applied, requiring it to be smaller than 0.02 cm;

- the selected electron candidates have to be isolated, i.e. low hadronic activity is expected around them, since they come from the decay of a weak boson. A generic isolation variable $I$ can be defined summing together other particles’ energy deposits in a cone built around the electron track and defined in the $\eta - \phi$ plane by $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2} \leq 0.3$.

First of all, a tracker isolation $I_{tk}$ is defined, which is computed summing the $p_T$ of all the tracks coming from the primary vertex and contained inside the cone. The actual cut is then applied on the relative isolation variable, defined as $I_{rel}^{tk} = I_{tk}/p_{ele}^T$, which is required to be smaller than 0.1. This cut is chosen since the isolation variable computed with tracker deposits is only mildly dependent on the pile-up.

The electron is further required to be isolated simultaneously in the tracker and in the electromagnetic and hadronic calorimeters. The fraction of energy deposited in the calorimeters within the isolation cone by particles coming from pile-up interactions can become sizeable in conditions of high pile-up, as it was in the second part of the data taking period. A correction is therefore applied, which consists in subtracting from the isolation cone the average pile-up energy. The combined relative isolation then becomes:

$$I_{comb}^{rel} = \frac{I_{trk} + I_{em} + I_{had} - \rho_{PU} \cdot \pi \Delta R_{eff}^2}{p_{ele}^T} \quad (3.3)$$
In equation 3.3, $\rho_{PU}$ is the average density of energy coming from pile-up particles and is estimated on an event-by-event basis as the median energy density in the $\eta - \phi$ plane of all jets\(^2\). $\pi \Delta R_{\text{eff}}^2$ is the isolation cone effective area, from which the pile-up energy density is subtracted. To determine the cone effective area, the pile-up energy density as well as the combined isolation variable are studied as a function of the number of reconstructed vertices in the event (i.e., as a function of the number of PU interactions), as shown in figure 3.11. Both quantities exhibit a linearly increasing trend with the number of vertices, although with different slopes. The ratio of the slopes of two linear fits to these quantities can be interpreted as $\pi \Delta R_{\text{eff}}^2$. A value of about 0.20 is found in data for the effective cone size, while in Monte Carlo $\Delta R_{\text{eff}} \simeq 0.18$.

Finally, with this correction applied, the electron candidate is required to have $I_{\text{comb}}^{\text{rel}} < 0.05$ in order to be considered isolated;

- in order to reject events where the electron originates from a conversion of a photon into an $e^+e^-$ pair inside the tracker material, the number of missed inner tracker layers of the electron track is required to be exactly zero. This means that there are no missed layers before the first hit of the electron track from the beam line. In addition, any event in which the selected electron is close in space to a partner track compatible with a photon conversion is rejected (i.e. $|\Delta \cot \theta| < 0.02$ and $|\text{dist}| < 0.02$, where these quantities are the distance of the two tracks in the longitudinal and transverse plane, respectively).

<table>
<thead>
<tr>
<th>variable</th>
<th>WP70 cuts</th>
<th>WP95 cuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\text{etmiss}}$</td>
<td>0.010 0.030</td>
<td>0.010 0.030</td>
</tr>
<tr>
<td>$</td>
<td>\Delta \phi_{\text{tk-sc}}</td>
<td>$</td>
</tr>
<tr>
<td>$</td>
<td>\Delta \eta_{\text{tk-sc}}</td>
<td>$</td>
</tr>
<tr>
<td>$H/E$</td>
<td>0.040 0.025</td>
<td>0.150 0.070</td>
</tr>
</tbody>
</table>

**Table 3.8:** Cut values of electron identification variables for WP70 (barrel and endcap) used for the tight electron selection, and WP95 (barrel and endcap) used in the loose electron selection.

### Tight muons

For the $\mu\nu_\mu q\bar{q}$ final state, the muon candidates have to pass several requirements in order to be identified as “real” muons:

- the muon candidate is required to be reconstructed both in the tracker and in the muon chambers;
- the HLT conditions allow to keep the muon transverse momentum requirement as low as 27 GeV/c for the whole period;
- the muon pseudorapidity spans the muon trigger geometrical acceptance region, that is $|\eta| < 2.1$;

\(^2\)Particle Flow jets built with a $k_T$ algorithm with $R$ parameter of 0.6 are used for $\rho_{PU}$ evaluation. A large number of very soft “ghost” particles is added uniformly to the event before jet clustering in order to allow a uniform tessellation of the $\eta - \phi$ plane.
as done for electrons, cuts on the muon track impact parameter along the $z$ coordinate and in the transverse plane of respectively 0.2 cm and 0.02 cm ensure that the muon does not originate from one of the pile-up vertices;

• muon identification helps in rejecting fake muons coming from QCD jets and consists in cutting on some track quality parameters. In particular, at least one pixel hit found for the inner track of the muon and a total number of tracker hits larger than 10 is required. Finally, the global track reduced chi-square must be smaller than 10.

• the selected muon candidates also have to be isolated. As for electron candidates, both a tracker relative isolation and a combined, PU-corrected, relative isolation are required. The value of the effective cone size for muons is $\Delta R_{\text{eff}} \simeq 0.17$ both in data and Monte Carlo. The cuts applied are then $I_{\text{rel}}^{\text{trk}} < 0.05$ and $I_{\text{comb}}^{\text{rel}} < 0.1$.

![Figure 3.11: Bare combined isolation variable $I_{\text{comb}}$ (purple dots), as well as the energy density from PU particles $\rho_{\text{PU}}$ (empty dots) as a function of the number of vertices in the event $N_{\text{vtx}}$ (data on the left, Monte Carlo on the right). The cone effective area is obtained as the ratio of the slopes of the two quantities. Pile-up corrected isolation is represented with green dots and its flatness proves the effectiveness of the empiric correction.](image)
Loose leptons

For the purpose of rejecting events with more than one lepton, we define loosely-identified leptons relaxing some of the requirements previously described. We consider as “loose” all the electrons which have $p_T > 15 \text{ GeV/c}$, $|\eta| < 2.5$, and $I_{\text{rel}}^{\text{trk}} < 0.2$ and which satisfy electron identification cuts according to WP95 (efficiency on real electrons $\sim 95\%$). The cut values for the identification variables used in the analysis can be found in table 3.8. Similarly, all the muons reconstructed both in the tracker and in the muon chambers and having $p_T > 10 \text{ GeV/c}$, $|\eta| < 2.5$ and $I_{\text{rel}}^{\text{trk}} < 0.2$ are defined as loose.

Jets

Once the final-state isolated electrons or muons are excluded from the list of the PF objects, all the remaining particles are clustered together to form jets. The clustering algorithm used is the anti-$k_T$, with $R$ parameter equal to 0.5. Only jets with $p_T > 30 \text{ GeV/c}$ and $|\eta| < 2.4$ (that is, inside the tracker acceptance region) are considered in the analysis. The Level 1 (offset), Level 2 (relative) and Level 3 (absolute) corrections described in section 2.3.3 are applied both in data and in the simulated samples, while residual corrections applied on data account for additional data/MC differences.

A few requirements are applied on reconstructed jets with the goal to reject fake, badly reconstructed and calorimeter noise jets, while retaining more than 98%–99% of real jets. The quality cuts are the following:

- jets must be made of more than one single constituent, and at least one of them must be a charged particle;
- the jet energy fraction carried by charged hadrons must be greater than exactly 0, while neutral hadron must carry at most 99% of the total jet energy;
- the jet energy fraction carried by charged and neutral electromagnetic objects is required to be smaller than 0.99.

In order to account for the fact that some good leptons might be also reconstructed as jets, we finally require the jets to lie outside a cone of radius $\Delta R = 0.5$ around the lepton selected for the analysis.

3.4.2 Analysis cuts

The selections presented in this section are applied event-by-event in order to reduce the background contamination, while enhancing the potential presence of a signal. The event selection is divided in two subsequent steps. The first one is a “preselection” where minimum criteria are required to be satisfied, with the aim of decreasing the number of events selecting only the desired final state. In the second step further selections are applied, in order to enhance the signal contribution over the large background rate. Both steps are described below:
Preselection

- The events are required to have a good primary vertex (PV). The primary vertex is selected as the one with the highest sum of $p_T^2$ of the associated tracks, and it is required to have a number of degrees of freedom $N_{dof} \geq 4$. In addition, the PV must lie in the central detector region of $|z| \leq 24 \text{ cm}$ and $\rho \leq 2 \text{ cm}$ around the nominal interaction point;

- the events are required to have exactly one tight lepton candidate (electron or muon) fulfilling the criteria described in the first part of section 3.4.1. The events containing one or more additional loose leptons of the same flavour or any other lepton of another flavour are rejected;

- in both the electron+jets and the muon+jets final states the events are required to have large missing transverse energy coming from the undetected neutrino, that is $E_{T}^{\text{miss}} > 30 \text{ GeV}$, and to have a leptonic W transverse mass $m_T > 30 \text{ GeV}/c^2$. The transverse mass is defined as

$$m_T = \sqrt{p_{T}^{\text{lepton}} \cdot E_{T}^{\text{miss}} \cdot (1 - \cos \Delta \phi)}$$

where $\Delta \phi$ is the angle between the lepton momentum $p_{T}^{\text{lepton}}$ and the $E_{T}^{\text{miss}}$ in the transverse plane. These cuts are designed to reduce the background contamination from QCD multi-jet production and Drell-Yan, which are free of real missing transverse energy. For the $e\nu_{e}q\bar{q}$ final states, a higher threshold at 50 GeV/$c^2$ is applied for the periods where the same selection appears online in the HLT requirement;

- events are required to have exactly two or three jets passing the cuts described in section 3.4.1. The two jets with highest $p_T$ are considered as W decay product candidates. This criterion, instead of choosing the jet pair whose invariant mass is closest to $m_W$, is used in order to minimize the bias introduced for background events due to jet combinatorics. Since in final states with three jets the $t\bar{t}$ contribution to the total background is significant, a b-jet veto is applied in those events to reduce its contamination. Events are discarded if one or more jets with $p_T > 30 \text{ GeV}/c$ and $|\eta| < 2.4$ are identified to be generated from a bottom quark, using a b-tag discriminator based on the counting of tracks with large impact parameter within the jet.

Final selections

- The most effective selection which helps in reducing the large W+jets contamination profits from the fact that the jet pair in signal events resonates at the W mass, within jet energy resolution. Conversely, in W+jets events the two jets originate from initial or final state gluon or quark radiation: due to the kinematic selections, the di-jet invariant mass has a broad peak around 80 GeV/$c^2$ with large low- and high-energy tails from jet combinatorics. For the final analysis, only a narrow region around $m_W$ is therefore retained as the signal phase space:

signal region: $65 < m_{jj} < 95 \text{ GeV}/c^2$

Events which fail the di-jet mass requirement are nevertheless kept and will be used to evaluate the W+jets background from data;

- in signal events, the Higgs decay products tend to be emitted in the central part of the detector, unlike W+jets events. Centrality cuts can therefore be applied on both the $\eta$ of the lepton and of the di-jet system. In particular, we require $|\eta_{\text{lepton}}| < 1.5$ and $|\eta_{jj}| < 3$;
• the W bosons from Higgs decay can have a large transverse momentum for medium-to-high Higgs masses. Since the W decay products are emitted at small opening angles for large boosts, we require $|\Delta \eta_{jj}| < 1.5$. This cut reduces the contamination from W+jets background without cutting much in the signal phase space and preserving the $l\nu q\bar{q}$ invariant mass shape (unlike the effect of a cut on $\Delta \phi_{jj}$, for example).

The aforementioned cuts have been tuned using the quantity $S/\sqrt{S+B}$ as a figure of merit, where $S$ and $B$ are respectively the number of signal and background events expected from the Monte Carlo simulation at an integrated luminosity of $5.0 \text{ fb}^{-1}$, in four-body mass ($m_{l\nu jj}$) windows centered around the nominal Higgs mass. The edges of these mass windows are listed in table 3.9. A detailed description of the techniques applied to reconstruct the Higgs invariant mass is postponed to the next section.

<table>
<thead>
<tr>
<th>$m_{H}$ [GeV/c$^2$]</th>
<th>250</th>
<th>300</th>
<th>350</th>
<th>400</th>
<th>450</th>
<th>500</th>
<th>550</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{l\nu jj}^{\text{min}}$ [GeV/c$^2$]</td>
<td>225</td>
<td>270</td>
<td>310</td>
<td>350</td>
<td>390</td>
<td>430</td>
<td>470</td>
<td>515</td>
</tr>
<tr>
<td>$m_{l\nu jj}^{\text{max}}$ [GeV/c$^2$]</td>
<td>275</td>
<td>330</td>
<td>390</td>
<td>450</td>
<td>510</td>
<td>575</td>
<td>640</td>
<td>700</td>
</tr>
</tbody>
</table>

Table 3.9: Definition of reconstructed $l\nu jj$ mass windows for the different Higgs mass hypotheses. Values are extracted from the study of the $S/\sqrt{S+B}$ figure of merit as a function of the window size, as shown in figure 3.15 for three representative mass points (low, intermediate and high $m_{H}$).

The result of this cut optimization is presented in figure 3.13, while figure 3.12 shows the expected significance at preselection level and after having applied the cuts here described.

![Figure 3.12](image_url)  
**Figure 3.12:** Higgs signal significance, expressed in terms of $S/\sqrt{S+B}$, for three different analysis steps: just after preselection (in black), after cutting on the di-jet invariant mass (in red), and after centrality and $\Delta \eta_{jj}$ cuts (in blue). The number of events for signal and background are counted in the mass windows of $m_{l\nu jj}$ defined in table 3.9.
3.4 Event selection

Figure 3.13: Left plots: cut optimization study, using $S/\sqrt{S+B}$ as a figure of merit. The relative impact on the significance of different cuts is represented with respect to a reference significance. In the top plot the reference value of the figure of merit is computed just after pre-selection, while in all other plots the cut on $m_{jj}$ is assumed. Right: the reconstructed four-body invariant mass $m_{l\nu jj}$ after applying the different cuts. The four rows of plots, from top to bottom, refer to cuts on $m_{jj}$, $|\eta_{\text{lep}}|$, $|\eta_{jj}|$, and $\Delta\eta_{jj}$ respectively.
3.5 Higgs mass reconstruction

As previously mentioned, despite the presence of an undetected neutrino, it is possible to reconstruct an Higgs mass peak in the $H \rightarrow WW \rightarrow l\nu qq$ decay channel. On the hadronic side the mass reconstruction is straightforward, because the two jets fully reconstruct the W mass; on the leptonic side it is possible to recover the W invariant mass adding together the lepton four-momentum, the missing transverse energy and imposing a further constraint on the neutrino $p_z$. This procedure is described here below.

3.5.1 Neutrino $p_z$ determination

The transverse components of the neutrino momentum correspond to the transverse missing energy (within the $E_T^{\text{miss}}$ reconstruction resolution), since no other undetectable particle is present in the final state. There is only one unknown parameter: the neutrino $p_z$, which nevertheless can be determined a posteriori, imposing that the the lepton and the neutrino momenta add together to form a W boson of mass $m_W = 80.399 \text{ GeV}/c^2$. Solving the resulting second-order equation one finds:

$$(p^z_\nu)_{1,2} = \frac{p^{\text{lepton}}_T \left( \hat{p}^{\text{lepton}}_T \cdot E_T^{\text{miss}} + m_W^2/2 \right)}{\left( p^{\text{lepton}}_T \right)^2} \pm \sqrt{\Delta}$$

where

$$\Delta = \left( \hat{p}^{\text{lepton}}_T \cdot E_T^{\text{miss}} + m_W^2/2 \right)^2 - \left( p^{\text{lepton}}_T \right)^2 \left( E_T^{\text{miss}} \right)^2.$$ (3.5)

Due to the finite detector resolution and the uncertainties in the reconstruction of $E_T^{\text{miss}}$, it can happen that the discriminant $\Delta$ is smaller than 0, in which cases we force it to 0 and a unique solution is obtained for $p^z_\nu$. Depending on the Higgs mass, this happens in about 31–39% of cases. For the remaining events, the discriminant $\Delta$ is larger than 0, and two distinct solutions exist. In this case, the one yielding the smaller value of $|p^z_\nu|$ for each event is chosen, after verifying via matching with the Monte Carlo truth that this choice corresponds to the correct solution in the majority of cases (61–65% depending on the Higgs mass).

3.5.2 Kinematic fit

The width of the Higgs resonance is given by the convolution of two different terms: the irreducible intrinsic decay width $\Gamma_H$ and an additional smearing term caused by detector resolution effects. The Higgs intrinsic width is the dominant term at high mass, while the impact of finite jet and $E_T^{\text{miss}}$ resolution is sizeable for a light Higgs particle.

A valuable piece of information in the Higgs decay chain is provided by the presence of two intermediate narrow resonances, that is the two W vector bosons: the a-priori knowledge of their mass can be used to improve the mass resolution in the four-body reconstruction. The most efficient way to do so is to implement a two-constraint kinematic fit on both the leptonic and the hadronic side of the Higgs decay chain. On an event-by-event basis, the fit uses the four-momenta $p_i$ of the lepton, neutrino and jets as inputs and modifies them to accommodate the external constraints. Each particle four-momentum is changed accordingly to the known resolution, parametrized as a function of its transverse momentum and pseudorapidity, and the...
best value is found through the minimization of a $\chi^2$ variable having the form:

$$\chi^2 = \sum_{i=\text{lepton},\nu,\text{jet}_1,\text{jet}_2} \frac{(p_i^{\text{fit}} - p_i^{\text{meas}})^2}{\sigma_i^2(\text{PT},\eta)} + \frac{(m_{\nu} - m_W)^2}{\Gamma_W^2} + \frac{(m_{\text{jj}} - m_W)^2}{\Gamma_W^2}$$

(3.6)

where the two constraints have a Gaussian form, with widths equal to the W boson decay widths.

Figure 3.14 shows the effect of the kinematic fit on the four-body mass spectrum for a signal Higgs mass hypothesis of 250 and 400 GeV/$c^2$. The distributions obtained before and after the kinematic fit are reported. Each of them is fitted with a Breit-Wigner function convoluted with a gaussian function, which represents the resolution effects. The parameters of the Breit-Wigner are fixed to the mass and width expected for that Higgs hypothesis, while the parameters of the Gaussian are left free in the fit.

In figure 3.15 the usual figure of merit $S/\sqrt{S+B}$ is used to perform a signal significance scan as a function of the counting mass window. For low and intermediate Higgs mass points, the advantages of a kinematic-fit based reconstruction of the $l\nu\text{jj}$ mass are sizeable. For Higgs masses $\geq 450$ GeV/$c^2$, little margin of improvement is expected, because the total peak width is dominated by the irreducible Higgs intrinsic decay width.

![Figure 3.14: Four-body mass distributions obtained before (dashed black line) and after (solid black line) the kinematic fit, for Higgs mass hypotheses of 250 and 400 GeV/$c^2$. The resolution improvement, in terms of Gaussian $\sigma$, is sizeable (about 15 GeV/$c^2$ subtracted in quadrature).](image)

### 3.6 Reconstruction, selection and trigger efficiencies

Given a certain simulated process having a cross section $\sigma$ and supposing an efficiency $\epsilon$ of the reconstruction procedure, the analysis selections or the trigger conditions, the number of expected events $N_{\text{exp}}$ after having collected an integrated luminosity $L$ is given by:

$$N_{\text{exp}} = \epsilon \cdot \sigma \cdot L$$

(3.7)

In this section, the evaluation of these efficiencies is explained.

Lepton reconstruction, selection and trigger efficiencies are computed using a tag-and-probe technique on $Z \rightarrow l^+l^-$ events having $70 < m_{l^+l^-} < 110$ GeV/$c^2$ [40].
3. Description of the experimental final state

The tag-and-probe method profits from the fact that the Z decays leptonically in a pair of same flavour particles (electrons or muons) to measure the efficiency of some selection criteria applied on these particles. One of the two leptons (the “tag”) is required to satisfy a set of very tight selection criteria, which identify that particular type of particle with a good confidence. The second one, called “probe”, is then selected with looser identification criteria. To ensure that also the probe is a particle of the required type, and not a fake lepton for instance, a pairing with the tag is required, such that the invariant mass of the combination is consistent with the mass of the Z. The sample of tag-probe pairs under the Z peak is, for most of practical cases, almost background free.

The efficiency is then measured by computing the ratio between the number of probes passing the criteria under study and the total number of probes, that is:

$$\epsilon = \frac{N_{\text{passing probes}}}{N_{\text{passing probes}} + N_{\text{failing probes}}} \quad (3.8)$$

Although the requirement of a tag-probe pair close to the Z resonance allows to have a highly pure sample, a possible source of systematic uncertainty in the computation of the efficiencies is given by the contamination of the probe ensemble by fake objects, which present a higher probability to fail the selection criteria under study than the real objects. Since the rate of fakes is expected to increment away from the Z peak, the efficiency calculation is repeated with two different requirements on the tag-probe invariant mass, one tighter ($80 < m_{l^+l^-} < 100\text{ GeV}/c^2$) and one looser ($60 < m_{l^+l^-} < 120\text{ GeV}/c^2$). The maximal excursion of the three different efficiency values is taken as the associated systematic uncertainty.

The tag-and-probe procedure can be applied on both data and on a Drell-Yan Monte Carlo sample. From the efficiency measurements in data and Monte Carlo, scale factors $\rho = \epsilon_{\text{data}}/\epsilon_{\text{MC}}$ can be calculated and used as weighting factors in simulated samples to remove any imperfections in the latter ones.

3.6.1 Electron efficiencies

The electron efficiency can be factorized in the product of three distinct components: the reconstruction efficiency $\epsilon_{\text{RECO}}$, the identification and isolation efficiency $\epsilon_{\text{ID}}$ and the offline trigger.
efficiency on the selected events $\epsilon_{\text{HLT}}$. In formula, we have:

$$
\epsilon_{\text{ele}} = \epsilon_{\text{RECO}} \times \epsilon_{\text{ID}} \times \epsilon_{\text{HLT}}
$$

(3.9)

The reconstruction efficiency $\epsilon_{\text{RECO}}$ is the probability of finding a reconstructed track when the electron supercluster is within the ECAL fiducial volume. As a matter of fact, the probability that every electron releases its energy in the electromagnetic calorimeter and that this energy can be collected in a supercluster is equal to one. The probe is therefore selected as an ECAL supercluster having a reconstructed transverse energy greater than 20 GeV. To reduce backgrounds, a tight selection is applied on the tag and the probe is asked to pass additional loose shower shape and isolation requirements, which are known from simulation to be uncorrelated with the supercluster reconstruction efficiency. $\epsilon_{\text{RECO}}$ is then the fraction of probes reconstructed as electron tracks.

Dividing the probes in mutually exclusive bins of $p_T$ (\([30, 35, 40, 45, 50, 60, 75, 200] \text{ GeV/c}\)) and pseudorapidity $\eta$ (\([-1.5, 0, 1.5]\)), the reconstruction efficiency measured from data is always above 96% and, in almost all cases, higher than 99%.

The isolation and identification efficiency $\epsilon_{\text{ID}}$ is given by the fraction of probes surviving the so-called $\text{WP70}$ working point and the isolation cuts defined in section 3.4. For this step, electron tracks matched to a supercluster are considered as probes.

In the left plot of figure 3.16, the ID efficiency as a function of the probe $p_T$ is shown, for probes having a pseudorapidity between 0 and 1.5. As can be seen, the working point $\text{WP70}$ has an efficiency of about 70% on real electrons of low $p_T$ ($\sim 30 \text{ GeV/c}$), while it increases up to about 85% the more the transverse momentum grows.

Finally, the HLT trigger efficiency $\epsilon_{\text{HLT}}$ is given by the probability of well identified electrons to fire the trigger paths defined in table 3.7(a).

In the right plot of figure 3.16, the HLT efficiency as a function of the probe $p_T$ is shown. There is a clear turn-on in the $p_T$ region corresponding to the trigger cut value (32 GeV/c for most of the data), which is due to the combination of Level-1 and HLT resolution effects, while for high $p_T$ values the efficiency reaches a stable plateau around 97%.

![Figure 3.16](image.png)

Figure 3.16: Efficiency measurements on electron data for the (a) identification and isolation and (b) the HLT step, using the tag-and-probe method. The result is shown as a function of the probe electron $p_T$, for probes having $0 < \eta < 1.5$. 
3.6.2 Muon efficiencies

Since the probability of reconstructing a muon is 100% for all the cases of practical interest, the efficiency for muons can be written as the product of an identification and isolation efficiency $\epsilon_{\text{ID}}$ and a trigger efficiency $\epsilon_{\text{HLT}}$:

$$\epsilon_{\text{muon}} = \epsilon_{\text{ID}} \times \epsilon_{\text{HLT}} \quad (3.10)$$

The isolation and identification efficiency is the probability of a reconstructed muon track to survive the good muon requirements listed in section 3.4, while the HLT trigger efficiency is given by the probability of well identified muons to fire the trigger paths summarized in table 3.7(b).

As done for electrons, also for muons the measurement is binned both in $p_T$ and $\eta$ of the probe muon, covering the intervals relevant for this analysis.

In the left plot of figure 3.17, the efficiency of the isolation plus identification measured on data is shown as a function of the probe muon $p_T$, for probes with $\eta$ between 0 and 1.5: the working point chosen for muons has a very high efficiency on real muons, above 90% at all transverse momenta.

The same thing is applied in the right plot of figure 3.17 for the trigger efficiency. Here a negative trend with the probe $p_T$ is exhibited: this effect originates from a different definition of the isolation cut between the HLT selection and the analysis cut. While offline muons are required to survive a relative isolation cut – which is then less and less tight as the muon $p_T$ increases – the requirement at HLT level consists of an absolute cut on the isolation variable, independent from the muon $p_T$.

![Figure 3.17](image)

*Figure 3.17:* Efficiency measurements on muon data for the (a) identification and isolation and (b) the HLT step, using the tag-and-probe method. The result is shown as a function of the probe muon $p_T$, for probes within $0 < \eta < 1.5$.

In table 3.10 and 3.11, the complete efficiency tables for electrons and muons, respectively, are reported for reference.

3.6.3 Electron+PF$m_T$ trigger efficiency

For what concerns the electron+PF$m_T$ trigger, its efficiency depends both on the electron leg of the trigger and on the transverse mass cut, which can be factorized as:

$$\epsilon_{\text{HLT}} = \epsilon_{\text{ele}} \times \epsilon_{PFm_T} \quad (3.11)$$
3.6 Reconstruction, selection and trigger efficiencies

<table>
<thead>
<tr>
<th>$p_T$ (GeV/$c$)</th>
<th>$\eta$</th>
<th>$\epsilon_{\text{ele}}^{\text{RECO}}$</th>
<th>$\rho_{\text{RECO}}$</th>
<th>$\epsilon_{\text{ele}}^{\text{ID}}$</th>
<th>$\rho_{\text{ID}}$</th>
<th>$\epsilon_{\text{ele}}^{\text{HLT}}$</th>
<th>$\rho_{\text{HLT}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[30, 35]</td>
<td>[−1.5, 0]</td>
<td>0.968±0.006</td>
<td>0.994±0.011</td>
<td>0.731±0.007</td>
<td>0.968±0.022</td>
<td>0.511±0.005</td>
<td>0.742±0.014</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.967±0.006</td>
<td>0.995±0.011</td>
<td>0.722±0.008</td>
<td>0.952±0.022</td>
<td>0.513±0.004</td>
<td>0.753±0.013</td>
</tr>
<tr>
<td>[35, 40]</td>
<td>[−1.5, 0]</td>
<td>0.990±0.002</td>
<td>0.998±0.003</td>
<td>0.761±0.003</td>
<td>0.942±0.009</td>
<td>0.866±0.002</td>
<td>0.951±0.004</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.989±0.002</td>
<td>0.996±0.004</td>
<td>0.762±0.003</td>
<td>0.941±0.009</td>
<td>0.865±0.002</td>
<td>0.953±0.004</td>
</tr>
<tr>
<td>[40, 45]</td>
<td>[−1.5, 0]</td>
<td>0.996±0.001</td>
<td>0.999±0.002</td>
<td>0.797±0.003</td>
<td>0.943±0.006</td>
<td>0.895±0.001</td>
<td>0.953±0.003</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.996±0.001</td>
<td>0.998±0.001</td>
<td>0.793±0.002</td>
<td>0.936±0.006</td>
<td>0.896±0.001</td>
<td>0.957±0.003</td>
</tr>
<tr>
<td>[45, 50]</td>
<td>[−1.5, 0]</td>
<td>0.997±0.001</td>
<td>0.998±0.001</td>
<td>0.809±0.003</td>
<td>0.928±0.007</td>
<td>0.914±0.002</td>
<td>0.957±0.003</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.996±0.001</td>
<td>0.998±0.002</td>
<td>0.812±0.003</td>
<td>0.932±0.007</td>
<td>0.917±0.002</td>
<td>0.959±0.004</td>
</tr>
<tr>
<td>[50, 60]</td>
<td>[−1.5, 0]</td>
<td>0.997±0.001</td>
<td>0.998±0.002</td>
<td>0.819±0.003</td>
<td>0.926±0.008</td>
<td>0.930±0.003</td>
<td>0.958±0.005</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.996±0.001</td>
<td>0.998±0.002</td>
<td>0.816±0.003</td>
<td>0.921±0.008</td>
<td>0.933±0.002</td>
<td>0.962±0.004</td>
</tr>
<tr>
<td>[60, 75]</td>
<td>[−1.5, 0]</td>
<td>0.994±0.004</td>
<td>0.997±0.007</td>
<td>0.828±0.007</td>
<td>0.930±0.016</td>
<td>0.949±0.007</td>
<td>0.972±0.011</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.992±0.004</td>
<td>0.994±0.007</td>
<td>0.831±0.007</td>
<td>0.933±0.016</td>
<td>0.944±0.004</td>
<td>0.965±0.008</td>
</tr>
<tr>
<td>[75, 200]</td>
<td>[−1.5, 0]</td>
<td>0.995±0.004</td>
<td>0.996±0.008</td>
<td>0.851±0.009</td>
<td>0.958±0.020</td>
<td>0.968±0.005</td>
<td>0.984±0.010</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.995±0.004</td>
<td>1.000±0.008</td>
<td>0.845±0.014</td>
<td>0.954±0.030</td>
<td>0.973±0.006</td>
<td>0.989±0.012</td>
</tr>
</tbody>
</table>

Table 3.10: Reconstruction, identification and trigger efficiencies for electrons, and corresponding scale factors ($\rho$). The statistic and systematic uncertainties are combined.

<table>
<thead>
<tr>
<th>$p_T$ (GeV/$c$)</th>
<th>$\eta$</th>
<th>$\epsilon_{\text{ele}}^{\text{ID}}$</th>
<th>$\rho_{\text{ID}}$</th>
<th>$\epsilon_{\text{ele}}^{\text{HLT}}$</th>
<th>$\rho_{\text{HLT}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>[30, 35]</td>
<td>[−1.5, 0]</td>
<td>0.928±0.001</td>
<td>0.991±0.004</td>
<td>0.890±0.001</td>
<td>1.000±0.001</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.929±0.001</td>
<td>0.995±0.001</td>
<td>0.888±0.001</td>
<td>0.999±0.001</td>
</tr>
<tr>
<td>[35, 40]</td>
<td>[−1.5, 0]</td>
<td>0.938±0.001</td>
<td>0.992±0.001</td>
<td>0.883±0.001</td>
<td>0.999±0.001</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.937±0.001</td>
<td>0.995±0.001</td>
<td>0.880±0.001</td>
<td>0.997±0.001</td>
</tr>
<tr>
<td>[40, 45]</td>
<td>[−1.5, 0]</td>
<td>0.948±0.001</td>
<td>0.992±0.001</td>
<td>0.881±0.001</td>
<td>1.000±0.001</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.947±0.001</td>
<td>0.994±0.001</td>
<td>0.877±0.001</td>
<td>0.998±0.001</td>
</tr>
<tr>
<td>[45, 50]</td>
<td>[−1.5, 0]</td>
<td>0.955±0.001</td>
<td>0.993±0.001</td>
<td>0.879±0.001</td>
<td>1.002±0.001</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.954±0.001</td>
<td>0.994±0.001</td>
<td>0.875±0.001</td>
<td>1.000±0.001</td>
</tr>
<tr>
<td>[50, 60]</td>
<td>[−1.5, 0]</td>
<td>0.959±0.001</td>
<td>0.993±0.001</td>
<td>0.873±0.001</td>
<td>1.000±0.002</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.956±0.001</td>
<td>0.994±0.001</td>
<td>0.871±0.001</td>
<td>1.000±0.002</td>
</tr>
<tr>
<td>[60, 75]</td>
<td>[−1.5, 0]</td>
<td>0.959±0.001</td>
<td>0.992±0.001</td>
<td>0.860±0.002</td>
<td>0.999±0.003</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.962±0.001</td>
<td>0.997±0.001</td>
<td>0.863±0.002</td>
<td>1.000±0.003</td>
</tr>
<tr>
<td>[75, 200]</td>
<td>[−1.5, 0]</td>
<td>0.964±0.002</td>
<td>0.993±0.002</td>
<td>0.844±0.003</td>
<td>0.994±0.004</td>
</tr>
<tr>
<td></td>
<td>[0, 1.5]</td>
<td>0.959±0.002</td>
<td>0.990±0.002</td>
<td>0.845±0.003</td>
<td>0.997±0.004</td>
</tr>
</tbody>
</table>

Table 3.11: Identification and trigger efficiencies for muons, and corresponding scale factors ($\rho$). The statistic and systematic uncertainties are combined.

$\epsilon_{\text{ele}}^{\text{HLT}}$ is evaluated with the tag-and-probe technique described above, while $\epsilon_{\text{ele}}^{\text{PFmT}}$ is modeled as a function of the offline $m_T$: events surviving the offline selection and firing the electron part of the cross trigger are used as reference events. The efficiency is computed as the ratio (binned in offline $m_T$ bins) of the events that pass the PF$m_T$ online cut and the total number of reference events. In formula, for each $i$-th $m_T$ bin:

$$\epsilon_{\text{ele}}^{\text{PFmT}}|_i = \frac{\#(\text{offline selected} \land \text{ref. HLT} \land \text{PFmT cut})_i}{\#(\text{offline selected} \land \text{ref. HLT})_i}$$

(3.12)
The resulting histogram is fitted with a Erf-like function of $m_T$ which models the turn-on. The ratio of the two curves in data and Monte Carlo is then used as a correction factor for the simulated samples. The systematic uncertainty related to the precision of this measure is accounted for by calculating the 68% confidence belts of the scaling factors from the fit parameter errors.

Figure 3.18 shows the measured turn-on curves in data and Monte Carlo, for a representative data-taking period.

![Turn-on curves](image)

**Figure 3.18:** (a) Turn-on curves for data during Run2011B (period e-vii of table 3.6(a)). (b) Turn-on curves measured in Monte Carlo samples, corresponding to the test of the HLT_Ele32_WP70_PFMT50_v4 trigger path, using HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v7 as reference trigger for the electron part.
Chapter 4

Limit extraction on the Higgs boson cross section

In this chapter, the analysis strategy developed to extract a limit on the Higgs boson cross section, based on the study of the decay channel $H \rightarrow WW \rightarrow l\nu q\bar{q}$, is presented. In section 4.1 the reconstruction, identification and trigger efficiencies evaluated in the last section of the previous chapter are applied to calculate the MC-expected distributions of the main physical observables, comparing them with the measured ones. Sections 4.2 and 4.3 deal, respectively, with the modelization of the expected background and signal shapes, while section 4.4 is about systematic uncertainties. The final limit extraction on the Higgs cross section is presented in section 4.5.

4.1 Data - Monte Carlo comparisons

After having applied the selection criteria described in section 3.4 to the Monte Carlo samples listed in table 3.4(b), it is possible to compare the simulated kinematic variables with the ones measured in data. Before doing so, an event-by-event reweighting procedure is applied on the MC samples, to take into account the different pile-up conditions between the data and the simulation (cfr. section 3.3.2). In addition, another weight is also assigned to Monte Carlo events, which accounts for data/MC differences in the reconstruction, identification and trigger efficiencies, as explained in section 3.6.

In the first part of this section, the data-driven approach for the QCD background estimation is described. Then, we proceed in listing the expected event yields for signal and backgrounds and in showing the comparison plots for the most significant kinematic observables.

4.1.1 Data - driven QCD estimation

In trying to quantify the contamination coming from multijet events where hadrons are wrongly reconstructed as leptons, the Monte Carlo is of little help because of the very small equivalent luminosity (see table 3.4(b)). The standard procedure is to estimate the shape and the contribution of the QCD background directly from data.

First of all, it is necessary to obtain a high-purity QCD sample from data. The isolation variables are the most effective ones to identify a QCD contamination: while the leptons coming
from weak boson decays are truly isolated (i.e. they have a little hadronic activity surrounding them), the ones coming from multijet events are in fact fake leptons, originating from the misreconstruction of other physical objects (mainly pions or kaons), and therefore have a substantial hadronic activity around them. Of course, it is possible to have real leptons also in the QCD background, if they come from the semi-leptonic decays of bottom and charmed mesons contained in jets, but nevertheless what said before about their poor isolation still holds. Therefore, at the preselection step the cut on the relative tracker isolation $I_{\text{rel}}^{\text{tk}} = I_{\text{tk}}/p_T$ is reversed, requiring $I_{\text{rel}}^{\text{tk}}$ to be greater than 0.1 (0.05) for electrons (muons). The working point of the identification is then relaxed to WP95 for electrons (see the cut values for the different ID variables reported in table 3.8) while it is kept unchanged for what concerns muons. Finally, the cut on the missing transverse energy $E_T^{\text{miss}}$ is relaxed down to 15 GeV.

The sample obtained from data after this modification in the selection flow is a good approximation of the QCD one, as verified on simulation with a closure test, and it is used to model the shape of the QCD contribution to the variables of interest.

A template fit is then performed on the distribution of the missing transverse energy found in data, using the shapes obtained from the Monte Carlo as a template for all the backgrounds except QCD, and the shape obtained from the data with the anti-selection previously described as a template for the QCD. This procedure is performed both for final states containing an electron and for final states containing a muon. The integral of the QCD distribution is allowed to float in the fitting procedure, and the result is taken as a measure of the QCD contamination in the signal region, both for electrons and muons; the integral of all the other backgrounds is fixed to the Monte Carlo expectation.

Among all the possible kinematic observables, the missing transverse energy has been chosen because it has a good discriminating power between the QCD events, which have mainly low $E_T^{\text{miss}}$, and the ones coming from W+jets, which exhibit a Jacobian peak around 40 GeV. The result of these template fits is represented in figure 4.1 for both electrons and muons.

![Figure 4.1](https://example.com/figure4.1.png)

**Figure 4.1:** Result of the QCD template fit for (a) the electron and (b) the muon final state. The QCD shape is taken directly from data, as explained above, while its normalization is left free in the fit. Other backgrounds have shape and normalization fixed to the Monte Carlo simulation. Note that the $E_T^{\text{miss}}$ cut has been lowered to 15 GeV.

The number of QCD events obtained from the fit, for the full 2011 dataset and for both the leptonic final states, is reported in table 4.1 below. The first row of the table, labelled by the
\( E_T^{\text{miss}} \) cut at 15 GeV, contains the integral of the QCD distribution obtained after the template fit, together with the associated uncertainty (\( \sim 2\% \) for the electron channel and \( \sim 7\% \) for the muons). The second row contains the number of QCD events expected when the \( E_T^{\text{miss}} \) cut actually performed in the analysis (30 GeV) is restored. Here the uncertainty is computed such that the relative error is the same as the one obtained from the fit (first row).

In the next column, for reference, the number of events actually measured after all the pre-election and final cuts of section 3.4.2, except the one on \( m_{jj} \), is reported (see also the tables in section 4.1.2): using these numbers is easy to see that the fraction of QCD expected events resulting from the fit amounts to \( \sim 4\% \) for electron data and \( \sim 2\% \) for muon data.

\[
\begin{array}{cccc}
\text{QCD} & \text{data} \\
\text{\( E_T^{\text{miss}} \)} > 15 \text{ GeV} & \begin{array}{ll}
electron{\nu}{jj} \text{ final state} & 11773 \pm 278 \\
\mu\nu{\mu}{jj} \text{ final state} & 5988 \pm 409
\end{array} & \begin{array}{ll}
electron{\nu}{jj} \text{ final state} & 90758 \\
\mu\nu{\mu}{jj} \text{ final state} & 148837
\end{array} \\
\text{\( E_T^{\text{miss}} \)} > 30 \text{ GeV} & \begin{array}{ll}
3564 \pm 84 & 3057 \pm 209
\end{array} & \begin{array}{ll}
90758 & 148837
\end{array}
\end{array}
\]

Table 4.1: Expected QCD events and associated uncertainties as obtained from the fit (first row) and when the analysis selection on \( E_T^{\text{miss}} \) is reintroduced (second row). The number of events observed in data after all the cuts except the one on \( m_{jj} \) is also shown for comparison with the QCD yield.

**Closure tests**

A preliminary test consists in checking the similarity between the data-driven \( E_T^{\text{miss}} \) shape and the simulated one. In figure 4.2 the shape of the \( E_T^{\text{miss}} \) distribution for QCD, as obtained from Monte Carlo, is compared with the one obtained from the anti-selection on data, both for electrons and muons. Despite the large uncertainties, due to the poor statistics in Monte Carlo and data, the two distributions are statistically consistent.

![Normalized Event Count vs \( E_T^{\text{miss}} \) for \( \electron{\nu}{jj} \) final state](a)  

![Normalized Event Count vs \( E_T^{\text{miss}} \) for \( \mu\nu{\mu}{jj} \) final state](b)

Figure 4.2: Comparison of the \( E_T^{\text{miss}} \) shapes for MC vs data-driven QCD events, for (a) electrons and (b) muons.

A more sophisticated test is then performed relying only on the simulated samples. First of all, the whole set of Monte Carlo samples (including the Monte Carlo QCD) is added together to mimic the data. The anti-isolation selection described above is then applied on the resulting dataset and a total \( E_T^{\text{miss}} \) shape is obtained: this will be the new – MC based –
template for the QCD background. As done before, the shape and yields of all the non-QCD backgrounds is taken from Monte Carlo. Finally, a $E_T^{\text{miss}}$ (pseudo) “data” distribution is build summing all the MC samples (including QCD). The same fitting procedure illustrated above is then performed, using these three templates (QCD, the sum of all non-QCD backgrounds and “data”).

A perfectly closed method would result in a QCD yield after the template fit which is exactly the same as the one coming from the QCD Monte Carlo itself. In fact this time we know that, by construction, the only difference between the background shape and the “data” shape is given by the simulated QCD, which in the first case is not added to the other samples, while in the second case it is.

The number of QCD events expected from Monte Carlo when all the cuts, except the one on the di-jet invariant mass, are applied is 2918 for the electron final state and 1863 for muons. The closure fit gives a value of 2503 for the $e\nu_ejj$ channel and 1974 for the $\mu\nu_\mu jj$ one. The percentual difference is equal to 14% for electrons and 6% for muons.

**QCD uncertainties**

In the case of electrons, the error on the number of QCD events coming from the fit (table 4.1, second row) is small, and we conservatively estimate the uncertainty to be one half of the expected value (50% relative error). For muons, the uncertainty is taken to be the one obtained from the fit (7% relative error). Both these numbers are more conservative than the percentual difference estimated with the closure test described above.

### 4.1.2 Event yields

The number of signal events expected from Monte Carlo simulation at 5 fb$^{-1}$ is reported in table 4.2 for four representative mass points: 300 GeV/c$^2$, 400 GeV/c$^2$, 500 GeV/c$^2$ and 600 GeV/c$^2$. The event yields are divided in the two production mechanisms here considered (gluon fusion and VBF), and each category is then further sub-divided according to the different leptonic decay modes of the W. Since only one good electron or muon is required in the analysis, the $W \rightarrow \tau \nu_\tau$ contributes only if the $\tau$ subsequently decays leptonically (the branching ratio $\mathcal{B}(\tau \rightarrow l\nu_{\nu})$ is $\sim 35\%$). The events are computed after each one of the selections described in section 3.4.2, except for the powerful cut on the di-jet invariant mass, which is not applied yet.

As can be seen from these numbers, a gluon-fusion produced Higgs constitutes the vast majority of the total signal, roughly the 90%. The VBF contribution, although not explicitly selected in the analysis cutflow, accounts for a 6–12% of the total expected signal. Looking at the decay products, the channels where the W goes directly into an electron or muon are obviously dominant, while the contribution coming from a $\tau$ decay is just a few percent of the total (4–5%, depending from the Higgs mass).

In table 4.3, similarly, all the Monte Carlo background samples are listed (the W+jets contribution has been split into the different exclusive jet categories) as well as the measured data, in the full $m_{jj}$ range. Each entry of the table shows both the electron and the muon channels, in the form: $e\nu_ejj/\mu\nu_\mu jj$. The QCD event yield is not shown as a function of the cuts because its value is extracted, with the data-driven technique described above, from a template fit on the $E_T^{\text{miss}}$ distribution at the $\Delta_\eta jj$ selection step (with just a loosening of the cut on the $E_T^{\text{miss}}$ itself). The sum of all the backgrounds at the last selection step is also shown for a direct comparison with the signal yields of table 4.2 and the measured data.
Table 4.2: Signal yields for some representative Higgs masses, after each selection cut, as obtained from Monte Carlo simulation at 5 fb$^{-1}$. Both the production mechanisms (gluon fusion and VBF) and the decays into electrons, muons and taus are shown. The selections are the ones described in section 3.4.2, except for the cut on the di-jet invariant mass $m_{jj}$, which is not applied yet.

### 4.1.3 Plots

In the following, the distributions of the most interesting kinematic observables are shown. In each plot, the different backgrounds are represented as coloured stacked histograms, while the black dots are the measured data. The signal distribution is shown with a red dashed line, for a representative Higgs mass $m_H = 400 \text{ GeV}/c^2$. To help the visualization of the signal, its cross section has been multiplied by a factor of 100. At the bottom of every plot, the bin-by-bin data/MC ratio is shown as a function of the variable itself, with the associated 68% confidence level error band in grey.

All the event cuts are applied, except for the one on the invariant mass of the two jets, $m_{jj}$. The stack of Monte Carlo samples is normalized to the integrated 2011 luminosity (5.0 fb$^{-1}$).

Lepton-related variables are shown in figure 4.3, $E_T$ related ones in figure 4.4, and jet distributions in figure 4.5. The distributions of the di-jet system are shown in figure 4.6, while the four-body invariant mass spectrum is presented in figure 4.7. For every group of plots, the column on the left shows the $e \nu_{e}jj$ final state, while the one on the right is for the $\mu \nu_{\mu}jj$.

The only large discrepancy is the one observed in the distribution of $\phi$ variable of the the missing transverse energy. The origin of such effect is known, and resides in an imperfect modelization in the simulation of the HCAL detector noise, which is not symmetric along the $\phi$ coordinate. Nevertheless, it has been verified that no other variables are affected by this effect.

All the other observables are modeled in a quite satisfactory way, although the non-perfect description of some of them (such as the four-body invariant mass) makes it clear that a pure Monte Carlo approach is not suitable to evaluate the expected background level in the final state and extract the signal. The method followed relies on the Monte Carlo only for the construction of a scale factor, for which it is required that the MC itself models the di-jet invariant mass in a satisfactory way around the $55–125 \text{ GeV}/c^2$ signal region. From what can be seen in figure 4.6, this is verified to be a reasonable assumption.

A complete description of the analysis method is presented in the next section.
4. Limit extraction on the Higgs boson cross section

Limit extraction on the Higgs boson cross section

<table>
<thead>
<tr>
<th>Cut</th>
<th>VV</th>
<th>W + jets</th>
<th>0 jets</th>
<th>1 jet</th>
<th>2 jets</th>
<th>3 jets</th>
<th>4 jets</th>
<th>( m_{jj} ) cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preselection</td>
<td>3.81e+03/7.52e+03</td>
<td>2.46e+03/4.40e+03</td>
<td>( \frac{3.34e+03}{7.78e+03} )</td>
<td>( \frac{2.38e+04}{4.26e+04} )</td>
<td>( \frac{7.59e+04}{1.41e+05} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centrality cuts</td>
<td>1.69e+03/3.68e+03</td>
<td>955/2.08e+03</td>
<td>( \frac{1.86e+03}{3.18e+03} )</td>
<td>( \frac{1.71e+04}{2.94e+04} )</td>
<td>( \frac{3.46e+04}{6.12e+04} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \Delta \eta_{jj} ) cut</td>
<td>1.51e+03/2.59e+03</td>
<td>955/2.08e+03</td>
<td>( \frac{1.68e+03}{3.07e+03} )</td>
<td>( \frac{1.10e+04}{1.87e+04} )</td>
<td>( \frac{2.27e+04}{3.98e+04} )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: Expected event yields as obtained from the Monte Carlo simulation at 5.0 fb\(^{-1}\). Each entry shows the number of events for both the final states: \( e\nu \) and \( \mu\nu \). The selections are the ones described in section 3.4.2, except for the cut on the di-jet invariant mass \( m_{jj} \), which is not applied yet. In the last column, the number of data events which survive the cuts is shown. The QCD estimation does not rely on MC and is computed by means of a template fit on the data \( E_{miss}^T \) distribution at the \( \Delta \eta_{jj} \) selection step. The selections are the ones described in section 3.4.2, except for the cut on the di-jet invariant mass \( m_{jj} \). The selections are the ones described in section 3.4.2, except for the cut on the di-jet invariant mass \( m_{jj} \).
4.1 Data - Monte Carlo comparisons

Lepton variables

Figure 4.3: Lepton-related variables. From top to bottom: lepton $p_T$, $\eta$, $\phi$. 

---

*4.1 Data - Monte Carlo comparisons*
$E_T^{\text{miss}}$ variables

Figure 4.4: $E_T^{\text{miss}}$-related variables. From top to bottom: $E_T^{\text{miss}}$, $E_T^{\phi}$, and lepton + $E_T^{\text{miss}}$ mT.
Jet variables

Figure 4.5: Jet-related variables. From top to bottom: leading $p_T$, trailing jet $p_T$, and jet $\eta$ distribution.
Di-jet system variables

Figure 4.6: Di-jet system-related variables. From top to bottom: $m_{jj}$, di-jet $p_T$, and $\Delta\phi$ di-jet separation in the transverse plane.
Higgs system variables

Figure 4.7: Higgs system-related variables. From top to bottom: plain $m_{1\nu jj}$, kinematic fitted $m_{1\nu jj}$, and Higgs reconstructed $\eta$. 
4. Limit extraction on the Higgs boson cross section

4.2 Background estimation

After the final selections, discussed in the previous sections, we are left with the two \( m_{\ell\nu\ell jj} \) distributions of figure 4.7, corresponding to the two categories in which the analysis is sub-divided (electrons and muons). These distributions are the same for every Higgs signal hypothesis and contain events in the full \( m_{jj} \) range, since up to now this cut has not been implemented yet.

The idea is to further categorize the events in two exclusive subsets, according to the value of the di-jet invariant mass: one is the “signal” region defined in section 3.4.2, having \( 65 < m_{jj} < 95 \text{ GeV}/c^2 \), and the other one is a “sideband” region, spanning a complementary \( m_{jj} \) range. In the former, the presence of a potential signal will be very much favoured compared to the latter; therefore, we can safely look at the sideband data events without spoiling the “blindness” of the analysis (that is, its insensitivity to the presence of a signal in the optimization phase or in the background extraction).

Once the sideband region has been defined (we will show in the following how this is done), the \( m_{\ell\nu\ell jj} \) background distribution is extrapolated from the \( m_{jj} \) sideband to the \( m_{jj} \) signal region. One way to do so, is to look at the four-body mass spectrum of the full Monte Carlo background, both in the \( m_{jj} \) signal and sideband regions, and create a ratio \( \alpha(m_{WW}) \) between the two \( m_{\ell\nu\ell jj} \) distributions. This ratio predicts how the sideband data should be scaled (as a function of the Higgs invariant mass) to obtain a background prediction in the signal region:

\[
N_{bkg}(m_{WW}) = N_{sb}(m_{WW}) \times \frac{N_{bkg}^{MC}(m_{WW})}{N_{sb}^{MC}(m_{WW})} = N_{sb}(m_{WW}) \times \alpha(m_{WW})
\]

(4.1)

where \( N_{sb}(m_{WW}) \) is the distribution of the events in the data sideband region, \( N_{bkg}(m_{WW}) \) is the extrapolated background in the data signal region and the MC superscript defines the same quantities at Monte Carlo level.

Once we have this background prediction, we can look at the data signal region and test the possible presence of an Higgs boson with statistical methods, for example counting the events in fixed windows around the nominal Higgs mass and comparing them with the background-only hypothesis, or performing a shape analysis on the full mass range under both the background-only and signal+background assumptions.

4.2.1 Defining the sideband region

While the so-called \( m_{jj} \) signal region is fixed, because it has already been defined optimizing the cut working point, there is some arbitrariness in choosing the edges of the sidebands. Ideally, we would like to find a region where the shape of the four-body invariant mass distribution does not differ too much from the one obtained with the \( m_{jj} \) signal events, and where the statistics is not too low. The first requirement allows to reduce \( \alpha \) to a constant scale factor on the whole \( m_{WW} \) range, the second avoids that the statistical error associated with \( \alpha \) becomes overwhelming.

An optimization procedure has been performed on the whole ensemble of Monte Carlo background samples. First of all, a set of cuts delimiting one or two different sideband regions on \( m_{jj} \) is defined, both above and below the \([65,95]\) GeV\( /c^2 \) range, with the only requirement that one of the edges of the sideband coincides with the 65 or 95 GeV\( /c^2 \) limit of the signal region. An \( m_{\ell\nu\ell jj} \) invariant mass distribution is built for each cut, using both electron and muon Monte Carlo events coming from the sidebands, and is then compared with the MC signal one, after having normalized both to unit area. The statistical comparison is performed via a \( \chi^2 \) test between the two histograms. The window that gives the best agreement between the \( m_{\ell\nu\ell jj} \) shapes coming from the signal and the sideband, while maintaining a sufficiently high statistics, defines the
4.2 Background estimation

limits of the $m_{jj}$ sideband.

The output of the optimization is represented in figure 4.8: for each value of the cut (or for combinations of them, which define two different sidebands surrounding the signal region) the result of the $\chi^2$ test is shown. The double interval $[55,65] \cup [95,125] \mathrm{GeV}/c^2$ is found to have the best compatibility between the shapes, while maintaining a high-enough statistics.

![Figure 4.8: Results of the $\chi^2$ test between the MC four-body mass distributions in the $m_{jj}$ signal (fixed) and sideband regions. The intervals on the horizontal and vertical axes define the sideband itself; the values in each bin are the $\chi^2$/ndf returned by the test: the closer they are to 1, the better is the agreement.](image)

To summarize, three regions on $m_{jj}$ have been defined:

- lower sideband region: $55 < m_{jj} < 65 \mathrm{GeV}/c^2$
- signal region: $65 < m_{jj} < 95 \mathrm{GeV}/c^2$
- upper sideband region: $95 < m_{jj} < 125 \mathrm{GeV}/c^2$

and for each one of them we can build the four-body invariant mass distribution, both for electrons and muons.

The advantage of the sideband approach is that most systematic uncertainties cancel out in the ratios. Supposing that our analysis has just one background contribution, the Monte Carlo simulation tells that the number of expected events in the signal region is $N_{\text{MC}}(m_{WW})$, while the number of sideband events is $N_{\text{sb}}^{\text{MC}}$. Here, the notation $N_{\text{MC}}^{\text{MC}}$ can refer to both the actual number of events or to a distribution, function of $m_{WW}$ (in this case $N_{\text{MC}}^{\text{MC}} = N_{\text{MC}}(m_{WW})$). The quantity $N_{\text{MC}}^{\text{MC}}$ will be given, in general, by the product of “plain” Monte Carlo events (cross section of the process times luminosity) and a weight $\epsilon_{\text{MC}}$, which contains reweighting factors and efficiencies. The uncertainty due to a systematic which has not been treated in a correct way, can be absorbed in a multiplicative coefficient $k$, and the weight factor becomes: $k \cdot \epsilon_{\text{MC}}(m_{WW})$. If this mismodeling does not influence $m_{jj}$ (the variable used to define the signal and the sideband regions), the scale factor becomes:

\[
\alpha_k(m_{WW}) = \frac{k \cdot N_{\text{MC}}^{\text{MC}}(m_{WW})}{k \cdot N_{\text{sb}}^{\text{MC}}(m_{WW})} = \frac{N_{\text{MC}}^{\text{MC}}(m_{WW})}{N_{\text{sb}}^{\text{MC}}(m_{WW})} = \alpha(m_{WW})
\] (4.2)
Actually, the procedure remains safe also if $m_{jj}$ is the wrongly described variable, but in such a way that the bad modeling influences the signal and the sideband in the same way (for example, a global scale factor between Monte Carlo an data, constant on $m_{jj}$, which factorizes out and is canceled in the ratio between the two).

A typical source of uncertainty which is wiped off with this approach is the one on the luminosity, which is nothing but an overall multiplicative factor applied in the same way to both the sideband and signal events.

In real life, the analysis has more than one background contamination and the factorization is not perfect, because the ratio in (4.2) now takes the form:

$$\alpha_{k_1,k_2,\ldots}(m_{WW}) = \frac{k_1 \cdot N_{MC}^{bkg_1} + k_2 \cdot N_{MC}^{bkg_2} + k_3 \cdot N_{MC}^{bkg_3} + \cdots}{k_1 \cdot N_{sb_1}^{MC} + k_2 \cdot N_{sb_2}^{MC} + k_3 \cdot N_{sb_3}^{MC} + \cdots} \quad (4.3)$$

However, we can get back to the previous case if all the backgrounds, except one, are sufficiently well described by the simulation, or if all the backgrounds, except one, have a negligible impact on the analysis.

The plots in figure 4.6 show that the data/MC agreement in the $[55,125] \text{ GeV}/c^2$ $m_{jj}$ region is quite satisfactory, making us more confident about the goodness of this approach.

A further test consists in checking the correlation between $m_{jj}$ and the other main kinematic variables: if they are completely uncorrelated, a potential mismodeling of one of them does not influence the di-jet invariant mass distribution.

The correlation plots between $m_{jj}$ and the main leptonic or hadronic observables are represented in figure 4.9, together with the relative correlation coefficients $\rho$'s. The leptonic variables, such as $p_{T,\text{lepton}}$, $\eta_{\text{lepton}}$ or the $E_{T,\text{miss}}$-related ones, together with the four-body invariant mass, are completely uncorrelated with the di-jet mass. Some correlation arises (as one could expect) with the hadronic ones, such as the jet $p_T$. The $\eta$ separation between jets is almost uncorrelated with $m_{jj}$.

Once the signal and the sideband regions are defined, we can compute the number of events which fall in each one of them. This is done in table 4.4 for all the background samples, at an integrated luminosity of $5 \text{ fb}^{-1}$, and for data. The number of Higgs events expected in the $m_{jj}$ signal region, for all the masses considered in the analysis, is shown in table 4.5.

In both tables, the events are divided according to the leptonic final state (electron or muon), while in table 4.5 there is a further differentiation, based on the Higgs production mechanism (gluon fusion or VBF).

The four-body invariant mass distributions for the events belonging to the two $m_{jj}$ categories are shown in figure 4.10. The coloured stacked histograms are the Monte Carlo background samples (here the different W+jets contributions have been merged together) normalized to the integrated 2011 luminosity, while the black dots are the measured data. The column on the left shows the $e^{\nu}e^{\nu}_{jj}$ final state, while the one on the right is for the $\mu^{\nu}_{\mu}jj$ final state. To maintain the blindness of the analysis until the final limit extraction, only data points corresponding to the signal region are represented.

### 4.2.2 Closure test on data

Before computing the ratio on Monte Carlo and scaling the data sideband distribution, a closure test has been performed on real data, keeping the signal region blinded.

The idea is to construct a pseudo-signal distribution using data events outside the $65 < m_{jj} < 95 \text{ GeV}/c^2$ interval, define a new sideband for these events and then test the rescaling approach
4.2 Background estimation

with a factor $\alpha(m_{WW})$ evaluated from Monte Carlo in these two new regions. Since we are not working with the “real” signal region, and therefore no excesses due to the potential existence of Higgs events are expected, the test should be “closed”, meaning that the extrapolated background should be statistically compatible with the invariant mass distribution in the data.

Figure 4.9: Correlation plots between $m_{jj}$ and, from left to right and from top to bottom: $P_T^{\text{lepton}}$, $E_T^{\text{miss}}$, lepton + $E_T^{\text{miss}}$ $m_T$, leading jet $p_T$, trailing jet $p_T$, $\Delta \eta_{jj}$, $m_{Wjj}$. The number $\rho$ in the upper right corner of every plot is the correlation coefficient between the two variables.
4. Limit extraction on the Higgs boson cross section

<table>
<thead>
<tr>
<th>$m_{\jj}$</th>
<th>0 jets</th>
<th>1 jet</th>
<th>$W + \text{jets}$</th>
<th>2 jets</th>
<th>3 jets</th>
<th>4 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>signal</td>
<td>5.58e+02/1.13e+03</td>
<td>6.17e+02/1.15e+03</td>
<td>1.11e+04/1.99e+04</td>
<td>5.81e+03/9.92e+03</td>
<td>2.22e+03/3.85e+03</td>
<td></td>
</tr>
<tr>
<td>sideband</td>
<td>2.61e+02/6.13e+02</td>
<td>4.20e+02/7.86e+02</td>
<td>1.09e+04/1.94e+04</td>
<td>5.98e+03/1.08e+04</td>
<td>2.62e+03/4.52e+03</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m_{\jj}$</th>
<th>$\jj$</th>
<th>$\text{VV}$</th>
<th>$Z + \text{jets}$</th>
<th>$t\bar{t} + \text{jets}$</th>
<th>single top</th>
<th>QCD (data driven)</th>
</tr>
</thead>
<tbody>
<tr>
<td>signal</td>
<td>8.40e+02/1.45e+03</td>
<td>5.21e+02/9.93e+02</td>
<td>1.20e+03/1.89e+03</td>
<td>4.52e+02/7.67e+02</td>
<td>9.39e+02/9.27e+02</td>
<td></td>
</tr>
<tr>
<td>sideband</td>
<td>3.46e+02/6.06e+02</td>
<td>6.03e+02/1.04e+03</td>
<td>1.46e+03/2.37e+03</td>
<td>5.31e+02/9.08e+02</td>
<td>1.00e+03/8.57e+02</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.4:
Number of events in the $m_{\jj}$ signal and sideband regions, for both Monte Carlo (at 5 fb$^{-1}$) and data. Each entry shows the two final states, in the form: $e\nu_{e}/\mu\nu_{\mu}\jj$.

<table>
<thead>
<tr>
<th>$m_{\jj}$</th>
<th>$\sum \text{bkg}$</th>
<th>data</th>
</tr>
</thead>
<tbody>
<tr>
<td>signal</td>
<td>2.42e+04/4.20e+04</td>
<td>2.48e+04/4.11e+04</td>
</tr>
<tr>
<td>sideband</td>
<td>2.41e+04/4.19e+04</td>
<td>2.53e+04/4.15e+04</td>
</tr>
</tbody>
</table>

### Table 4.5:
Number of expected signal Higgs events in the $m_{\jj}$ signal region (the Higgs events that fall in the sidebands are not considered). The events are divided according to the different production mechanisms, and each entry shows the two final states in the form: $e\nu_{e}/\mu\nu_{\mu}\jj$.

The easiest way to perform this consists in further subdividing the sidebands previously defined in two parts: the events belonging to one of them will form the pseudo-signal, the ones belonging to the other the pseudo-sideband. Specifically, the two new regions, which are defined such that they have a comparable statistics, are the following:

- pseudo-sideband region: $m_{\jj} \in [55, 60] \cup [108, 125]$ GeV/$c^2$
4.2 Background estimation

The scale factor $\alpha(m_{WW})$ is evaluated as the binned ratio between the four-body mass distributions in the pseudo-signal and the pseudo-sideband regions, which are obtained summing all the Monte Carlo backgrounds plus the data-driven QCD. This ratio is then used to multiply, bin-by-bin, the data $m_{\ell\nu\ell\ell}$ distribution coming from the pseudo-sideband and the result is assumed to be an estimate, in both yield and shape, of the background in the data pseudo-signal region. To avoid large fluctuations due to the poor statistics in the tails, the Monte Carlo distributions are rebinned in such a way that the relative error of each bin is equal or less than 20%. This means that the scale factor $\alpha(m_{WW})$ is assumed to be a constant for large values of the $m_{WW}$ invariant mass, while at low masses it is allowed to float bin-by-bin. For the $m_{\ell\nu\ell\ell}$ distributions, a 25 GeV/$c^2$ binning on the range [175,800] GeV/$c^2$, identical to the one used for the plots in figure 4.10, is chosen.

The statistical compatibility of the result of the procedure is checked calculating, for each bin, the pull between the observed data and the extrapolated background, where this quantity is
defined as\textsuperscript{1}:

\[
pull = \frac{N_{\text{obs}} - N_{\text{exp}}}{\sigma_{\text{obs}} \oplus \sigma_{\text{exp}}}
\]

The error \(\sigma_{\text{obs}}\) on the number of observed events is poissonian and is simply given by the square root of the number of events \(N_{\text{obs}}\) contained in the bin; the error on the extrapolated number of background events, \(\sigma_{\text{exp}}\), is the combination of the uncertainty associated with the sideband events (which is also poissonian) and the error of the Monte Carlo scale factor \(\alpha(m_{WW})\).

The results of the test are represented in the two plots of figure 4.11 (electrons) and figure 4.12 (muons). The first plot on the left shows the bin-by-bin ratio between the observed signal events and the expected background (blue dots with error bars) and the ratio between the expected background and itself (orange shaded area): an agreement should manifest in the fact that the blue points are compatible with the shaded area. Large fluctuations are however expected in the tails (at high mass values) due to poor statistics, and are indeed observed in both channels.

At low masses (around 200–250 GeV/c\(^2\)) the muons show a nice agreement, while electrons are slightly worse. This is probably due to an imperfect modeling of the QCD background, which is more important at low masses and which has a larger contribution for electrons than for muons.

In the second plot, on the right, the pull values calculated for each bin are put into an histogram and the resulting distribution is fitted with a Gaussian function. If the two distributions are statistically compatible, the pull values are expected to be distributed around zero, with a width equal to 1. A departure of the mean from zero is the symptom of a bias, while, if the width differs from one, it usually means that the errors have not been computed in the correct way. As can be seen, the \(\sigma\) of the Gaussians is in both cases compatible with 1. The fitted mean value is compatible with zero inside 1.75 (electrons) or 1.2 (muons) standard deviations, respectively.

Since in one case there is a slight over-fluctuation above zero, while in the other there is an under-fluctuation, we are confident that this is simply due to statistical effects and that there is no intrinsic bias in the method.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.11}
\caption{Closure test for electrons. Left: ratio between the observed signal and the expected background as a function of the invariant mass. Right: histogram of the pull values.}
\end{figure}

\subsection{4.2.3 Extrapolating the background shape}

The satisfactory results of the closure test make us confident about the reliability of the method: we are therefore ready to apply it to the sideband and signal regions defined in section 4.2.1.

\textsuperscript{1}The symbol \(\oplus\) indicates the sum of squares between the two uncertainties.
4.2 Background estimation

As done for the closure test, the sideband and signal Monte Carlo distributions are rebinned in order to avoid large fluctuations of the scale factor $\alpha(m_{WW})$, driven by the low statistics in the tails. The requirement for two (or more bins) to be merged together is that the relative bin error must be equal or lower than 20%.

After the rebinning, the $m_{\ell\nu\ell jj}$ distributions are divided bin-by-bin and the ratios represented in figure 4.13 are obtained, for both electrons and muons. This quantity, with the associated uncertainty, will be our scale factor $\alpha(m_{WW})$.

The four-body invariant mass shape built with all the data events which fall in the $m_{jj}$ sideband region is then scaled by this factor and the result is assumed to be our best estimate of the background expected in the data signal region, in both shape and yield. The results of the scaling are shown in figure 4.14 for the two leptonic final states with a blue continuous line which has been superimposed to the signal distribution (black dots).

The pull distributions are shown in figure 4.15, together with the Gaussian fit.

The expected background in the $\mu\nu\ell jj$ final state shows a good agreement with what is actually observed in data (see also the ratio plots in the bottom part of figure 4.14) and this manifests
4. Limit extraction on the Higgs boson cross section

4.3 Signal modeling

In the previous section we have described how it is possible to extrapolate a background yield and shape prediction in the $m_{jj}$ signal region, by means of a re-scaling of the sideband invariant mass distribution. This section, instead, is devoted to the modelization of the signal shape. The number of Higgs events which survive all the cuts, including the $65 < m_{jj} < 95 \text{ GeV}/c^2$ one,
has been shown in table 4.5 for all the hypothetical mass points here considered: they span from a few tens to a few hundreds, depending on \( m_H \), and their yield is therefore much lower than the expected background, dominated by W+jets production (see table 4.4). The possibility of reconstructing the parent invariant mass from the four final state objects, as described in section 3.5, allows to configure the analysis as the search for an excess of events clustered around a certain mass value in the \( m_{\ell\nu\ell j} \) distribution. In order to do so, we don’t need just a prediction of the number of expected Higgs events, but also a modelization of their invariant mass shape.

The Higgs \( m_{\ell\nu\ell j} \) shape is taken from the Monte Carlo event distributions, after having regularized them with a fit. Different fits are performed independently for each mass hypothesis and for each systematic effect that involves also a change in the shape. Moreover, the gluon fusion and the vector boson fusion production mechanisms are considered as distinct signal channels, since they are affected by significantly different systematic uncertainties: as a consequence, each one of them has its own independent mass fit. The two different final states (electronic or muonic) are also fitted separately. Globally, we have 32 different shape modelizations, plus the ones coming from the systematics.

Empirically, a double Crystal-Ball function (i.e. a Gaussian core with power law tails on both sides) has been found to adequately parametrize the signal shapes. The function resulting from the fit is then converted into an histogram on the 175–800 GeV/c\(^2\) mass range, binned as the backround one, and the integral of this histogram is scaled in such a way that its value is equal to the event yields reported in table 4.5. Examples of such fits can be found in figure 4.16, for the gluon fusion production mechanism and the Higgs mass hypotheses of 350 GeV/c\(^2\) (\( \mu\nu\mu j j \) final state) and 500 GeV/c\(^2\) (\( e\nu\mu j j \) final state).

![Figure 4.16: Higgs boson invariant mass shape, as obtained from the simulation, with all the proper weighting factors and normalizations taken into account. The spectrum is fitted with a double Crystal-Ball function (red line). A Higgs mass of 350 GeV/c\(^2\) (500 GeV/c\(^2\)) is assumed in the left (right) plot.](image)
4.4 Systematic uncertainties

This section describes the main sources of systematic uncertainties that affect the measurement. The signal yield and shape estimation are subjected to both theoretical (production cross section, choice of PDF, Higgs boson width modeling at high mass) and instrumental (LHC luminosity, pile-up model in the Monte Carlo, lepton energy scale, jet/\text{\text{E}}^\text{miss}_T\text{ energy scale, lepton reconstruction and trigger efficiencies, b-tag}) uncertainties.

The background yield and shape is extracted using a Monte Carlo-based scale factor, and therefore is affected as well by production cross section, scale and pile-up uncertainties.

4.4.1 Signal systematics

All the sources of systematic uncertainties affecting directly the signal yield or the shape determination are detailed here.

Cross section prediction and other theoretical uncertainties

The inclusive cross sections which have been used for the Higgs signal yield determination are calculated by the “Higgs Cross Section Working Group” [15] and have an associated uncertainty that is computed varying the QCD normalization and factorization scales and that includes the uncertainties of the PDF used for the calculation.

The overall effect is of the order of 13-19%, depending on the Higgs boson mass. The total cross section uncertainty is driven by the Higgs production via the gluon fusion mechanism, which is the dominant one. Nevertheless, since the gg and VBF channels are treated independently in the limit setting procedure, to each production channel is assigned its own cross section systematic.

The uncertainty associated to the PDF above described only accounts for a global normalization of signal events. The effect of the choice of different PDF sets – and of their intrinsic uncertainties – on the event selection efficiency is analysis-dependent and has to be specifically studied. The PDF4LHC recommendations [41] have been followed and the systematic uncertainty has been estimated as the envelope of the signal acceptance differences calculated for three sets of PDFs, namely the CTEQ6.6, MSTW08 and NNPDF2.0.

The values obtained, which vary according to the Higgs mass, range from 1.5% to 3.6% for gluon fusion production and from 0.6% to 1.1% for VBF.

Finally, in the Monte Carlo generators used to produce signal events, an approximation is used, which consists in creating the Higgs on-shell and in simulating the decay via an ad-hoc Breit-Wigner distribution. This approximation, valid at a good level for light Higgs particles, may be inadequate for heavy Higgs ($m_H \gtrsim 300\text{ GeV}/c^2$) searches. A conservative systematic uncertainty has been proposed and is expressed by $150\% \cdot m_H^3$, where the Higgs mass is expressed in TeV/c$^2$. This effect amounts to about 4% for a $300\text{ GeV}/c^2$-mass Higgs, and the systematic grows rapidly with mass, contributing a $\sim 30\%$ effect at $600\text{ GeV}/c^2$.

LHC luminosity

The latest measurement of the CMS luminosity uncertainty is at the level to 2.2% [42]. The luminosity has been extracted by means of a pixel cluster counting technique, which is based on the assumption that the number of hit pixel clusters per bunch crossing is a very linear
function of the number of interactions per crossing, and therefore a good measure of the instantaneous luminosity. The uncertainty on the 2.2% value is dominated by the afterglow\(^2\) correction uncertainty and by differences between measurements performed in distinct periods.

### Pile-up

The presence of pile-up events (i.e. not coming from the primary hard interaction) can affect in a substantial way some physical observables: spurious particles may be clustered together in jets or can be included in isolation cones built around leptons, thus modifying the value of their isolation variables.

Corrections such as the Level 1 or the effective-area energy subtraction are applied in the reconstruction phase of jets and leptons, in order to mitigate these effects. Moreover, a simulation of pile-up is added to Monte Carlo events and a reweighting procedure is performed to take into account residual differences between MC and data.

A source of systematic uncertainty is due to both the uncertainty in the number of interactions observed in data and to uncertainties in the Monte Carlo modelization.

From equation 3.2, we see that an uncertainty in the measured number of pile-up interactions in data is due to an uncertainty in the luminosity estimate (the 2.2% already mentioned) or in the minimum bias cross section \(\sigma_{\text{min, bias}}\). The combination in quadrature of these two sources of error gives a total uncertainty of approximately 3.6% on the estimated number of interactions.

To take into account also the uncertainty in the Monte Carlo number of pile-up events, a conservative estimate of 5% is assumed for the value of this uncertainty. A variation of \(\pm 5\%\) is therefore applied to the number of interactions (see figure 4.17(a)) and then propagated to the weights given to the MC signal samples. This has an effect both on the event yields, which is nevertheless found to be very small (\(\lesssim 1\%)\) for all the mass hypotheses, and on the invariant mass shape. In figure 4.17(b) the shape variation is shown for a representative sample having \(m_H = 400\,\text{GeV/}c^2\), summing both electrons and muons and considering only the gluon fusion production mechanism.

Since the modification on the shape is tiny, this shape-effect is neglected and the pile-up systematic – which is nevertheless taken into account – is assumed to influence only the signal yield.

### Jet Energy Scale

The uncertainty on the jet energy scale (JES) is a function of \(\eta\) and \(p_T\) and affects both the signal efficiency, due to a different acceptance for jets with a given \(p_T\) threshold, and the shape of jet-related quantities, such as the di-jet invariant mass \(m_{jj}\) used for defining the signal and sideband regions.

The systematic due to the JES is taken into account by varying coherently the \(p_T\) of all the reconstructed jets by \(\pm 1\sigma\) of the uncertainty at which the energy scale is known from dedicated CMS measurements [43]. The effect on the signal shape is shown in figure 4.18 for two representative Higgs masses, considering only the gluon fusion production mechanism and merging together the electronic and muonic final states. The overall impact on the signal yields varies according to \(m_H\) and its value is included into the ranges 1.6%-4.6% for the gluon fusion pro-

\(^2\)An afterglow happens when the energy originating from a given bunch crossing creates a small response in subsequent bunch crossings.
4. Limit extraction on the Higgs boson cross section

Figure 4.17: (a) Variation of ±5% on the number of pile-up interactions in data. The dashed red line is the result of an up-scaling, the blue dashed line of a down-scaling. (b) Effect of the pile-up systematic on the shape of a gluon fusion Higgs signal having a mass of 400 GeV/c².

(3.8) and 0.5%–3.4% for the VBF production.

Given the visible modification in the four-body invariant mass shape, this source is handled as both a yield and shape systematic in the limit extraction process.

Figure 4.18: Changes into the signal four-body mass shape due to a ±1σ jet energy scale variation for a 300 GeV/c² (left) and a 500 GeV/c² (right) Higgs boson.

The Jet Energy Resolution (JER) is assumed to be well-described by the simulation and no uncertainty associated with it is included.

Lepton energy scale

As it is done for jets, the energy scales of electrons and muons are rescaled by their respective uncertainties, and the the impact on the signal estimations is calculated.

For eνjj final states, a flat energy scale uncertainty of 2.2% has been assumed for electrons. This is a conservative value, compared to the one which is quoted by ECAL as the result of
4.4 Systematic uncertainties

improved corrections and intercalibrations, but is appropriate for the prompt SingleElectron dataset used for the analysis.

For $\mu\nu_{\ell\ell}$ final states, the muon scale uncertainty has been assumed at the level of 1%.

Lepton efficiency

As explained before, lepton reconstruction, selection and trigger efficiencies are computed using a tag and probe technique on $Z \rightarrow \ell^+\ell^-$ events in data and Monte Carlo. From this measurement, data and MC efficiencies are derived and, from their ratio, scale factors $\rho$ are obtained to reweight the simulated samples. The scale factors have an associated uncertainty $\sigma_\rho$, which is the result of the propagation of each single efficiency measured uncertainty. Such $\sigma_\rho$ is propagated through the full analysis to check its impact on the signal shapes and yields.

The global scale factor $\rho_{\text{TOT}} = \rho_{\text{RECO}} \times \rho_{\text{ID}} \times \rho_{\text{HLT}}$ has been conservatively scaled by $\pm 1\sigma_\rho$, which corresponds to considering the efficiencies of the three single steps fully correlated. The Higgs mass shapes are not affected by variations of lepton efficiencies, while the yields change of a quantity which is assumed to be around 0.3% for electrons and 1% for muons.

In the very same way, the systematic uncertainty associated with the transverse mass cut part of the electron trigger has been estimated independently. The result on the signal yields is flat with the Higgs mass and everywhere at the level of 1.4%.

b-tagging

The tagging efficiencies for bottom and light flavour jets are measured in data and in Monte Carlo and the ratios between the two (scale factors) can be computed in bins of jet $p_T$ and $\eta$ and then used to correct the mismatch between data and Monte Carlo. These measurements have an associated uncertainty which is propagated to the analysis.

The effect of this uncertainty on the shape of $m_{\ell\nu_{\ell\ell}}$ is nevertheless found to be negligible, while for the signal yield a conservative value of 1% is quoted, for both production mechanisms.

4.4.2 Background systematics

The main sources of systematic uncertainties that affect the background yield or shape estimation are detailed here.

Cross section prediction

Since the computation of the scale factor $\alpha(m_{WW})$ relies on Monte Carlo, the uncertainty on the production cross sections for all the different backgrounds might have an impact on it.

Using a notation where $B$ represents the number of total background events and the index $i$ runs on all the considered background sources, we have $B = \sum_{i=1}^{\# \text{bkg}} B_i$.

An uncertainty in the cross section for any of these background processes influences the number of expected Monte Carlo events, and one way to accommodate this is to introduce in the previous formula some coefficients $\beta_i$:

$$B = \sum_{i=1}^{\# \text{bkg}} (1 + \beta_i)B_i = (1 + \beta_1)B_1 + \cdots + (1 + \beta_N)B_N$$  \hspace{1cm} (4.5)
These coefficients can be thought as random numbers sampled from a Gaussian distribution centered in 0 and having a $\sigma$ width equal to the relative uncertainty on the $i$-th background cross section, as extracted from table 3.3. For example, the diboson background has a relative uncertainty around 3% and the coefficient $\beta_{VV}$ is therefore some random number with a 68% probability of having a value between -0.03 and 0.03.

The square of the uncertainty on the total number of background events then becomes:

$$
(\delta B)^2 = \sum_{i=1}^{\# \text{bkg}} \left( \frac{\partial B}{\partial B_i} \delta B_i \right)^2 = \sum_{i=1}^{\# \text{bkg}} (1 + \beta_i)^2 (\delta B_i)^2
$$

(4.6)

where $(\delta B_i)^2$ has both a statistical and a systematic component, other than the cross section uncertainty. The $\beta_i$ factor can therefore be seen as a second-order correction on the systematic one.

Since around 85% of the total background comes from W+jets events (see table 4.3) we are, with a good approximation, in the situation of having just one background contribution to the measured events. Under this assumption, when taking the signal-to-sideband ratio of equation (4.2), every constant factor that enters both in the signal and in the sideband cancels out, leaving the scale factor $\alpha(m_{WW})$ completely unaffected. This, of course, perfectly applies to the $\beta$ factor that has been introduced to model the scaling due to cross section uncertainties.

The corrections for the minority backgrounds can be neglected as well, because this effect amounts to just a few percent and therefore has a negligible impact on the overall error on the estimation of these particular backgrounds.

In conclusion, a cross section uncertainty on the Monte Carlo backgrounds is not expected to have any appreciable influence on the ratio $\alpha(m_{WW})$.

**Pile-up**

As done for the signal samples, the pile-up number distribution (figure 4.17(a)) is scaled up or down of a certain amount (5%) and the Monte Carlo scale factor $\alpha(m_{WW})$ is computed again, therefore obtaining two additional estimates of the expected background in the data signal region. In figure 4.19 the difference in shape is shown for both electrons and muons and the three distributions are found to be in a reasonable good agreement. The difference in yield corresponds to less than 1% for both electronic and muonic final states, and these values are of course insensitive to the value of $m_H$.

Shape-effects are therefore neglected and the pile-up systematic for the background is included as a pure multiplicative (scale) factor on the number of expected events.

**Jet Energy Scale**

In a similar way, the effect of an uncertainty on the jet energy scale is studied rescaling the $p_T$ of all the reconstructed Monte Carlo jets and propagating this change through the scale factor $\alpha(m_{WW})$ and the $m_{t\ell jj}$ background prediction. In figure 4.20 the effect on the pure shape is shown for both final states. For what concerns the event yield, the difference is found to correspond to 0.6% for electrons and 1.4% for muons.

The JES uncertainty on background is included in the limit extraction procedure as both a yield and shape systematic.
4.4 Systematic uncertainties

Figure 4.19: Extrapolated background shapes, normalized to unit area, as obtained from the Monte Carlo samples (black continuous line) and from the same samples, but with the pile-up distribution scaled 5% up (red dashed line) or down (blue dashed line). The bottom plot shows the ratio between the three distributions. Left: $e\nu_{e}jj$ final state; right: $\mu\nu_{\mu}jj$.

Figure 4.20: Extrapolated background shapes, normalized to unit area, as obtained from the Monte Carlo samples (black continuous line) and from the same samples, but with the JES scaled up (red dashed line) or down (blue dashed line). The bottom plot shows the ratio between the three distributions. Left: $e\nu_{e}jj$ final state; right: $\mu\nu_{\mu}jj$.

Background normalization

The error on the number of background events extrapolated in the data signal region is given by the combination between the uncertainty in the number of data events in the sidebands (which is purely statistic) and the uncertainty on the Monte Carlo scale factor $\alpha(m_{WW})$. To the latter contribute both the MC statistics and the systematics (for example the one on the background cross sections described above).

The relative error is simply given by the ratio between the integral of the background four-body distribution when the content of every bin is replaced with the content itself plus the bin error, and the number of expected background events. This quantity corresponds to 5% for $e\nu_{e}jj$ final states and 3.3% for $\mu\nu_{\mu}jj$ final states.
In addition to the overall normalization, a shape systematic is also associated to the background, in a similar way to what has been done for the jet energy scale. The difference between the two is that, while for the JES the systematic is assumed to vary the invariant mass shape in a coherent way (i.e. all the bins are fully correlated), in this case the bins are allowed to float independently one from the other, each one according to its own statistical precision. The difference in shape between the background distributions is shown in figure 4.21.

Figure 4.21: Extrapolated background shapes, normalized to unit area, as obtained from the Monte Carlo samples (black continuous line) and from the same samples, but with each bin content scaled up (red dashed line) or down (blue dashed line) according to the bin error. The bottom plot shows the ratio between the three distributions. Left: $e\nu_{\ell}jj$ final state; right: $\mu\nu_{\ell}jj$.

A summary of the signal and background systematics and their values is given in table 4.6.

<table>
<thead>
<tr>
<th>source of uncertainty</th>
<th>gg $e\nu_{\ell}jj$</th>
<th>gg $\mu\nu_{\ell}jj$</th>
<th>VBF $e\nu_{\ell}jj$</th>
<th>VBF $\mu\nu_{\ell}jj$</th>
<th>background $e\nu_{\ell}jj$</th>
<th>background $\mu\nu_{\ell}jj$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross section (QCD scale)</td>
<td>7.3 – 8.7%</td>
<td>0.4 – 1.5%</td>
<td>8.1 \rightarrow 10%</td>
<td>2.9 \rightarrow 4.6%</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>cross section (pdf)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.3 \rightarrow 32%</td>
<td>/</td>
</tr>
<tr>
<td>pdf acceptances</td>
<td>1.5 – 3.6%</td>
<td>0.6 – 1.1%</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Higgs line-shape</td>
<td></td>
<td></td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>LHC luminosity</td>
<td></td>
<td></td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>pile-up</td>
<td></td>
<td></td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>jet/$E_T^{miss}$ energy scale</td>
<td>1.6 – 4.6%</td>
<td>0.5 – 3.4%</td>
<td>0.6% 1.4%</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>lepton energy scale</td>
<td>2.2% 0.8%</td>
<td>2.2% 0.8%</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>lepton efficiencies</td>
<td>0.3% 1%</td>
<td>0.3% 1%</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>PF$m_T$ trigger efficiency</td>
<td></td>
<td></td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>b-tag veto</td>
<td></td>
<td></td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>background normalization</td>
<td></td>
<td></td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

Table 4.6: Sources of systematics considered in the analysis. The range of variation is intended to span the different Higgs mass points. An arrow between two values indicate that the systematic uncertainty has a trend with the Higgs mass.
4.5 Setting a limit on the SM Higgs boson cross section

In section 4.2 and 4.3 we have obtained and estimate of the expected background and signal invariant mass shapes: the former is given by the rescaling of the $m_{lljj}$ distribution in a sideband region built on data, using a Monte Carlo-driven scale factor; the latter is obtained regularizing the MC shape with a suitable fit function.

Using these two ingredients, it is possible to infer a constraint on the existence of an Higgs boson, for a certain mass hypothesis. If no significant excess is observed in data, an upper limit on the Higgs boson cross section can be established, up to a certain degree of belief. For example, it can be said that there is a 95% C.L. that the Higgs boson cross section is not larger than some value $\sigma_{95\% \text{ C.L.}}$. Otherwise said, if the cross section were larger than that, a statistically-significant excess of data would have emerged over the background. If the upper limit $\sigma_{95\% \text{ C.L.}}$ is as large as the production cross section predicted by the Standard Model ($\sigma_{\text{SM}}$) or smaller, it can be said that the existence of a SM Higgs boson of that mass is excluded at 95% C.L.

The procedure adopted jointly by CMS and ATLAS is based on the modified frequentist CL$_s$ approach [44]. The test statistics needed to define the CL$_s$ is built starting from the following likelihood function:

$$L(\text{data}|\mu, \theta) = P(\text{data}|\mu \cdot s(\theta) + b(\theta)) \cdot p(\hat{\theta}|\theta) \quad (4.7)$$

where $s(\theta)$ and $b(\theta)$ are respectively the signal and background expectations (in the form of event yields or shapes), depending on some set of nuisance parameters $\theta$ with observed value $\hat{\theta}$. The signal strength modifier $\mu$ changes the cross section of all signal processes by the same scale, that is $\sigma = \mu \cdot \sigma_{\text{SM}}$. The likelihood then can be read as the probability of observing a certain amount of data when the expected yield is $\mu \cdot s(\theta) + b(\theta)$, times the probability of measuring a value $\hat{\theta}$ for the nuisance parameter $\theta$.

The appropriate test statistics to quantify an absence of signal is then defined as:

$$q_\mu = \begin{cases} 
-2 \log \frac{L(\text{data}|\mu, \hat{\theta}_\mu)}{L(\text{data}|\hat{\mu}, \hat{\theta})} & \hat{\mu} \leq \mu \\
0 & \hat{\mu} > \mu 
\end{cases} \quad (4.8)$$

$\hat{\mu}$ and $\hat{\theta}$ are the values of the parameters which maximize the likelihood, while $\hat{\theta}_\mu$ is the value of $\theta$ that maximize $L$ for a given assumed $\mu$. The number $\hat{\mu}$ is allowed to assume negative values, but its constraints force the limit to be one sided.

The values of the nuisance parameters that best describe experimental data are obtained through a maximum-likelihood fit in the background-only and in the signal+background hypothesis (i.e. setting $\mu = 0$ or $\mu \neq 0$ in $L$) and these quantities are called $\hat{\theta}_0^{\text{obs}}$ and $\hat{\theta}_\mu^{\text{obs}}$, respectively. These values are then used to generate pseudo-data (simulated counting) in the background-only and signal+background scenarios, where a signal strength $\mu$ is assumed for the latter. From these pseudo-samples, it is possible to obtain the pdf’s of the statistics $q_\mu$ under both hypotheses. An example of these distributions is shown in figure 4.22.

The value of the test statistics actually measured in data, called $q_\mu^{\text{obs}}$, defines two quantities ($p$-values) on these pdf’s: one for the signal+background $(p_\mu)$ and one for the background-only hypothesis $(p_0)$, namely:

$$p_\mu \equiv \text{CL}_{s+b} = P(q_\mu \geq q_\mu^{\text{obs}}|\mu \cdot s(\hat{\theta}_\mu^{\text{obs}}) + b(\hat{\theta}_0^{\text{obs}}))$$

$$p_0 \equiv \text{CL}_b = P(q_\mu \geq q_\mu^{\text{obs}}|b(\hat{\theta}_0^{\text{obs}})) \quad (4.9)$$

$^3$For a shape-based analysis, as the one performed here, $P$ is in fact the product of the Poisson probabilities associated with each bin of a certain distribution.
From these, the $\text{CL}_s (\mu)$ is calculated as:

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{\text{CL}_b} \quad (4.10)$$

If, for $\mu = 1$, $\text{CL}_s < \alpha$, the SM Higgs boson is excluded with $(1 - \alpha) \text{CL}_s$. To find the 95% CL$_s$ upper limit on $\mu$, the constraint $\text{CL}_s = 0.05$ is imposed, and the equation is solved for $\mu$.

Starting from the background estimation only, an upper limit can be built under the assumption that the background-only hypothesis is the true one. This upper limit will not have an unique value, but will statistically fluctuate according to a certain distribution. In order to derive the median-expected upper limit, and the associated $\pm 1\sigma$ and $\pm 2\sigma$ bands, a large set of background-only pseudo-data can be generated and, for each of them, the $\text{CL}_s$ and $\mu_{95\% \text{ C.L.}}$ are calculated. The point at which the cumulative distribution of $\mu_{95\% \text{ C.L.}}$ crosses the 0.5 quantile is taken as the median expected limit. The $\pm 1(2)\sigma$ uncertainty bands are extracted from the values of the crossings of the 16%(2.5%) and 84%(97.5%) quantiles.

![Figure 4.22: Test statistic distributions for ensembles of pseudo-data generated for signal+background and background-only hypotheses. The value of $q_\mu$ measured from real data is indicated with an arrow.](image)

### 4.5.1 Results obtained

The statistical procedure described above has been applied to each one of the 8 hypothetical Higgs mass values studied in this analysis, ranging from 250 GeV/$c^2$ to 600 GeV/$c^2$. All the systematic uncertainties presented in section 4.4 are treated as nuisance parameters of the complete likelihood $L$.

The results for each $m_H$ are put together in the global exclusion plots of figure 4.23(a) ($e\nu, jj$ final state) and figure 4.23(b) ($\mu\nu, jj$ final state). In figure 4.24 the statistical combination of the two is also shown.

The blue dashed line represents the expected exclusion limit (that is, the median) as a function of the Higgs mass, together with the associated uncertainty bands (green and yellow areas). The observed limit is shown with a black continuous line.

No excesses are seen in the considered mass region and the observed limit on the Higgs cross section is found to be compatible with the one expected in the background-only hypothesis. The presence of an Higgs boson having a production cross section around two times as large as the one predicted by the Standard Model is expected to be excluded in the 350–400 GeV/$c^2$ region. The limit becomes less stringent especially at high masses, due to the increasingly lower Higgs
4.5 Setting a limit on the SM Higgs boson cross section

Figure 4.23: The Standard Model Higgs exclusion limit, obtained for (a) the electron and (b) the muon final states.

Figure 4.24: The Standard Model Higgs exclusion limit, obtained combining the electron and muon final states.

production cross section, the sizable width of the resonance and the larger systematics. In this most sensitive region of the analysis, the observed exclusion limit is higher, being around $2.5-3\,\sigma_{SM}$. 
4. Limit extraction on the Higgs boson cross section
Chapter 5

Conclusions

In this work, an analysis technique dedicated to the search for the Standard Model Higgs boson in the semi-leptonic WW decay channel has been presented. The analysis is optimized for the gluon fusion Higgs production mechanism, although a slight contamination coming from vector boson fusion events is expected, and takes into account only electronic and muonic final states. The Higgs search has been performed in the mass region above the WW production threshold, with a maximum sensitivity for Higgs mass values around 400 GeV/$c^2$. The description of the experimental final state, the sources of background contamination and their expected contribution, the event selections and the calculation of the efficiencies have been discussed. A possible method for estimating the expected background level in the final state and extract the number of signal events, based on the rescaling of data from a sideband to a signal region, has been presented. The modeling of the expected signal and background contribution has made possible to extract an upper limit on the Standard Model Higgs production cross section, based on statistical methods. The main sources of systematic uncertainties affecting the analysis have also been examined and included in the limit extraction.

All data collected during 2011 have been analyzed in this final state. In the mass region considered, from 250 GeV/$c^2$ to 600 GeV/$c^2$, no excesses are seen and the observed limit on the cross section is compatible with the one expected in the background-only hypothesis, within two standard deviation error bands. The presence of an Higgs boson having a production cross section around two times as large as the one predicted by the Standard Model is expected to be excluded in the 350–400 GeV/$c^2$ region, while the limit becomes less stringent for lower and higher mass values. The observed exclusion limit is higher, being around $2.5-3 \, \sigma_{SM}$ in the most sensitive region of the analysis.

The result of this analysis paves the way for extending the Higgs exclusion up to the TeV scale, increasing the available statistics with 2012 data. Moreover, in principle, this very same technique can be applied to the search for new resonances having a mass in the range considered here and decaying into a pair of $W$, with a semi-leptonic final state. Further studies may involve the investigation of the WW scattering at high energy.
5. Conclusions
References


References


