# Università degli Studi di Milano Bicocca Facoltà di Scienze Matematiche Fisiche e Naturali

Scuola di Dottorato di Scienze Corso di Dottorato di Ricerca in Fisica ed Astronomia



# Search for a Higgs Boson in the $H \to W^+ W^- \to \ell \nu \ell \nu$ channel at CMS

Coordinatore della Scuola di Dottorato: prof. Giuseppe Chirico Relatore:

prof. Stefano Ragazzi

Tesi di dottorato di Andrea Massironi matricola nº 063516

Ciclo XXV Anno Accademico 2012/2013

Le grandi opere si compiono meno colla forza che colla perseveranza (Cesare Cantù)

# CONTENTS

1	The	The Standard Model 1							
	1.1	Standar	rd Model	1					
	1.2	Gauge	invariance	4					
	1.3	Sponta	neously broken symmetry	7					
		1.3.1	Limits on the Higgs mass	12					
		1.3.2	Higgs at LHC	15					
2	LHC	and CI	MS	19					
	2.1	LHC .		19					
		2.1.1	LHC operation	20					
		2.1.2	LHC properties	20					
		2.1.3	LHC phenomenology	22					
		2.1.4	LHC kinematics	24					
	2.2	CMS .		24					
		2.2.1	The Tracker	27					
		2.2.2	The electromagnetic calorimeter: ECAL	28					
		2.2.3	The hadronic calorimeter: HCAL	30					
		2.2.4	The muon system	31					
		2.2.5	The trigger system	32					
3	Phys	sics obie	ect reconstruction	33					
-	3.1	Muons		33					
	3.2	Electro	ns	34					
	0.2	3.2.1	Supercluster (SC) reconstruction	35					
		3.2.2	Electron track	36					
		3.2.3	GSF Electron	36					
		324	Energy corrections	37					
	33	Jets	merely concourses a second s	37					
	3.4	Missing	y transverse energy MET	30					
	3.5	Particle	e Flow reconstruction	40					

4	ECA	ECAL Alignment 43							
	4.1	Alignment procedure							
		4.1.1 Samples used for the analysis							
		4.1.2 Event selection $\ldots \ldots 46$							
	4.2	Monte Carlo alignment and expected precision							
	4.3	Alignment performances							
	4.4	Conclusions $\ldots \ldots \ldots$							
_	<b>-</b> .								
5	Ihe	Di-Jet studies 55							
	5.1 5.0	Theoretical motivations							
	5.2 5.0	Trigger and event selection							
	5.3	Analysis							
	5.4 	Systematic uncertainties							
	6.6	Results							
6	The	HWW Analysis 65							
Ŭ	6 1	Main backgrounds 68							
	6.2	Analysis strategy 73							
	0.2	6.21  Trigger							
		6.2.2 Primary Vertex Reconstruction 75							
		6.2.3 Muon Selection $76$							
		6.2.4 Electron Selection 77							
		6.2.5 Missing Energy 78							
		6.2.6 Z. Veto 80							
		6.2.7 Let Counting $81$							
		6.2.8 Top Tagging 81							
		6.2.9 Other Preselection Requirements 82							
	63	Higgs Signal Extraction Strategy 82							
	6.4	Background Estimation 87							
	0.1	6.41 Jet Induced Backgrounds: W + jets and OCD 89							
		6.4.2 Top Background 91							
		6.4.3 Drell-Van Background $Z/\gamma^* \rightarrow \ell^+ \ell^-$ 93							
		$6.4.4  \text{Drell-Van} \to \tau\tau \qquad $							
		6.4.5 Other Backgrounds 98							
	6.5	Efficiency Measurements 98							
	6.6	Systematics 98							
	6.7	Limit extraction and discovery significance							
	0.1								
7	The	VBF Analysis 105							
	7.1	VBF selections							
	7.2	Background estimation							
		7.2.1 The top background							
		7.2.2 The $Z/\gamma^* \to \ell^+ \ell^-$ background $\ldots \ldots \ldots$							
	7.3	Systematics							
	7.4	Results							
-		····							
8	The	VH Analysis 123							
	8.1	VH selections							
	8.2	Background estimation							
		8.2.1 The top background $\ldots$ 128							

		8.2.2 The $Z/\gamma^* \to \ell^+ \ell^-$ background	129			
		8.2.3 The WWV background	132			
	8.3	Systematics	132			
	8.4	Results	133			
9	The	Final Results	139			
	9.1	VBF and VH Standard Model Higgs boson $H \to W^+W^-$ combined limit	139			
	9.2	Standard Model Higgs boson $H \to W^+W^-$ combined limit	140			
	9.3	Standard Model Higgs boson combined limit	143			
	9.4	Fermiophobic Higgs	146			
10	The	Conclusions	149			
List of tables vi						
List of figures x						
11	11 Acknowledgements x					

CHAPTER 1

# THE STANDARD MODEL

Apri la mente a quel ch'io ti paleso e fermalvi entro; ché non fa scïenza, sanza lo ritenere, avere inteso.

Dante Alighieri, Paradiso

Elementary particles and their interactions are today best described by the Standard Model, a field theory which combines special relativity and quantum mechanics. In this chapter the basic concepts of this model are introduced. The electroweak symmetry breaking and the Higgs mechanism are then described as the way foreseen to explicate the origin of particles masses. The limits on the Higgs mass, set on both theoretical and experimental grounds, are discussed. An overview of the Higgs production and decay mechanisms at the LHC is given.

## 1.1 Standard model

The theory that summarizes the current experimental knowledge of elementary particles and their interactions is called Standard Model (SM)[1] [2]. It is a relativistic quantum field theory based on the group of symmetries  $SU(3) \bigotimes SU(2) \bigotimes U(1)$ . These three local symmetry groups dictate the three interactions between the particles in the SM. The theory is perturbative at sufficiently high energies and renormalizable. According to the SM, the constituents of matter are spin-1/2 particles called fermions. The fundamental fermions observed up to today in experiments are subdivided in leptons and quarks. These two groups of particles come in three families or generations, that behave almost identically under interactions. The three known lepton families are the electron (e), the muon  $(\mu)$  and the tau  $(\tau)$ . Each of them comes with its associated neutrino,  $\nu_e$ ,  $\nu_{\mu}$  and  $\nu_{\tau}$ . The six quark flavours are labelled as up u, down d, charm c, strange s, top t and bottom b. All fermions are summarized in Table 1.1.

Up to now, quarks have only been observed into bound states of  $q\bar{q}$  pairs, called mesons, or  $qqq/\bar{q}\bar{q}\bar{q}$  aggregates, called baryons.

Each of these fermions is in addition accompanied by an anti-particle with opposite quantum numbers and exactly the same coupling of its counterpart.

	1	$s^{st}$ gen.		2 <sup>nd</sup> gen.		$3^{rd}$ gen.	Q
loptong	$\nu_e$	$\sim 0$	$ u_{\mu} $	$\sim 0$	$\nu_{ au}$	$\sim 0$	0
reprons	e	$511  \mathrm{keV}/c^2$	$\mu$	$105.7{\rm MeV}/c^2$	$\tau$	$1.777{ m GeV}/c^2$	-1
quarka	u	$\sim 2{\rm MeV}/c^2$	c	$\sim 1.3{\rm GeV}/c^2$	t	$\sim 173.5{\rm GeV}/c^2$	2/3
quarks	d	$\sim 5{\rm MeV}/c^2$	s	$\sim 95  {\rm MeV}/c^2$	b	$4.2{ m GeV}/c^2$	-1/3

Table 1.1: Spin-1/2 matter constituents, divided into their generations and their electric charge [3].

The quantum field operators associated to fermions are 4-components Dirac spinors  $\psi$ . For a free fermion of mass m the associated Lagrangian is

$$\mathcal{L}_{Dirac} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi, \qquad (1.1)$$

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

$$(i\gamma^{\mu}\partial_{\mu} - m)\psi = 0. \qquad (1.2)$$

The adjoint spinor  $\bar{\psi} = \psi^{\dagger} \gamma^{0}$  has been introduced. For further convenience, the Weyl spinor representation<sup>1</sup> has been introduced, which allows to write

$$\psi = \psi_L + \psi_R = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \chi_R \end{pmatrix}.$$
(1.3)

The two-components objects  $\psi_L$  and  $\psi_R$  are referred to as left-handed and right-handed Weyl spinors respectively, and are obtained from the spinor  $\psi$  through the projection operators

$$\psi_L = P_L \psi = \frac{1}{2} (1 - \gamma^5) \psi \qquad \psi_R = P_R \psi = \frac{1}{2} (1 + \gamma^5) \psi.$$
 (1.4)

The left or right-handedness of spinors is called chirality.

Interactions between fermions happen through the exchange of spin-1 particles, called bosons, which arise from invariances of the theory under the so-called gauge symmetries. Three types of fundamental interactions between fermions have been observed.

The  $SU(2)_L \times U(1)_Y$  groups are associated to the electroweak interaction, which is 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-violation characteristic of weak interactions. The  $U(1)_Y$  gauge boson couples to both the left-handed and the right-handed components.

<sup>&</sup>lt;sup>1</sup>Here the adopted Weyl representation of the  $\gamma$  matrices corresponds to  $\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $\gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix}$ and  $\gamma^5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ 

The left-handed projections of the fermion fields form  $SU(2)_L$  doublets

$$\mathbf{f}_{L} = \begin{pmatrix} \nu_{e} \\ e \end{pmatrix}_{L}, \begin{pmatrix} \nu_{\mu} \\ \mu \end{pmatrix}_{L}, \begin{pmatrix} \nu_{\tau} \\ \tau \end{pmatrix}_{L}, \begin{pmatrix} u \\ d \end{pmatrix}_{L}, \begin{pmatrix} c \\ s \end{pmatrix}_{L}, \begin{pmatrix} t \\ b \end{pmatrix}_{L}, \quad (1.5)$$

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

$$f_R = e_R, \ \mu_R, \ \tau_R, \ u_R, \ d_R, \ c_R, \ s_R, \ t_R, \ b_R.$$
(1.6)

To each doublet it is associated a so-called weak isospin charge  $T = \frac{1}{2}$ : neutrinos and "up" quarks possess a third isospin component  $T_3 = \frac{1}{2}$ , whereas leptons and "down" quarks exhibit a  $T_3 = -\frac{1}{2}$ .

The  $SU(3)_C$  symmetry group is related to the strong interaction between quarks, which is governed by QCD. Each quark appears in three different colour states, thus belonging to a  $SU(3)_C$  triplet, while leptons are colourless singlets. The gluons, which are the quanta of the strong interaction field, acting between colour-charged quarks, have zero mass and carry colour-charge.

The fourth known fundamental interaction, gravity, is much weaker than the other three interactions and cannot be accommodated yet into an unified theory together with electromagnetic, weak and strong forces.

A free bosonic force carrier of mass m and spin 0 is represented in Quantum Field Theory by a complex scalar field  $\phi$ , whose dynamics is described by the Klein-Gordon Lagrangian

$$\mathcal{L}_{KG} = \left(\partial_{\mu}\phi\right)^{\dagger} \left(\partial^{\mu}\phi\right) - m^{2}\phi^{\dagger}\phi \,. \tag{1.7}$$

By deriving the Eulero-Lagrange law of motion from the above Lagrangian, the following equation is obtained:

$$\left(\Box + m^2\right)\phi = 0. \tag{1.8}$$

For vector (i.e. spin 1) bosons, the associated operator is a vector field  $A_{\mu}$  in the four dimensional Minkowski time-space, whose dynamics is described by the Proca Lagrangian:

$$\mathcal{L}_{Proca} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} m^2 A_{\mu} A^{\mu} , \qquad (1.9)$$

where the kinetic term has been introduced through the antisymmetric tensor  $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ . The equation of motion is

$$\left(\Box + m^2\right) A_{\mu} = 0. \tag{1.10}$$

Up to now, only the free non-interacting theory has been introduced. The SM approach to account for interactions between particles is the requirement of local gauge invariance of the Lagrangian.

### 1.2 Gauge invariance

Since Maxwell's unification of electric and magnetic interaction, the concept of gauge invariance has played a strategic role in the understanding and the description of the fundamental forces of Nature. The requirement of a symmetry in the Lagrangian of a theory accounts for conservation of charges, via the Noether's theorem, and allows for the introduction of new fields and interactions in the theory.

In QFT it proves very convenient to require the Lagrangian invariance under gauge transformations, i.e. under an internal phase transformations of the form  $\psi \to e^{i\alpha}\psi$ . If the phase  $\alpha$  is constant in time and space, it's called "global gauge", whereas if the phase differs from point to point,  $\alpha = \alpha(x)$ , it's "local gauge" transformations.

Considering for instance a local gauge transformation of the form

$$\psi \to \psi' = U\psi = e^{i\alpha(x)}\psi \tag{1.11}$$

into some internal group G of generators  $T^a$ , with a = 1, ..., n. The group algebra is defined by the structure constants  $f^{abc} = [T^a, T^b]$ . The phase can be expressed as  $\alpha(x) = \varepsilon^a(x)T^a$ , with  $\varepsilon^a(x)$  being the rotation parameter. The quantum-mechanical observables, which depend only on  $|\psi|^2$  are invariant under Eq. 1.11, whereas a Lagrangian such as the one in Eq. 1.1 in general is not. This is due to the extra term  $\partial_{\mu}\alpha(x)$  in the derivative transformation:

$$\partial_{\mu}\psi \to \partial_{\mu}\psi' = e^{i\alpha(x)}\partial_{\mu}\psi + i\psi\partial_{\mu}\alpha(x). \qquad (1.12)$$

A possible way to make the theory manifestly invariant under the gauged symmetry is to introduce a set of new vector fields  $A^a_{\mu}$  and replace the usual time-space derivative  $\partial_{\mu}$  with the so-called covariant derivative

$$\mathcal{D}_{\mu} = \partial_{\mu} - igT^a A^a_{\mu}, \qquad (1.13)$$

with an arbitrary parameter g that will determine the interaction strength of the field. Substituting the covariant derivative into the Lagrangian of Eq. 1.1 yields

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\mathcal{D}_{\mu}\psi - m\bar{\psi}\psi$$
  
$$= i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi - ig\bar{\psi}\gamma^{\mu}T^{a}A^{a}_{\mu}\psi, \qquad (1.14)$$

where the last term expresses the coupling between the fermion field and the new vector fields. The substitution of the standard derivative with the covariant one allows for the Lagrangian invariance under the U transformation, provided that the new introduced  $A^a_{\mu}$  vector fields transform under U so as to exactly compensate the extra term in Eq. 1.14. By demanding that

$$\mathcal{D}'_{\mu}\psi' = U(\mathcal{D}_{\mu}\psi) \tag{1.15}$$

the following transformation laws for the vector fields  $A^a_{\mu}$  can be derived:

$$A_{\mu} \to A'_{\mu} = U A_{\mu} U^{\dagger} - \frac{i}{g} (\partial_{\mu} U) U^{\dagger} . \qquad (1.16)$$

Finally, in order to give a kinetic term to these gauge fields, a tensor  $F_{\mu\nu}$  has to be introduced, which must be antisymmetric in its two spatial indexes. It is natural to define  $-igF^a_{\mu\nu}T^a = [\mathcal{D}_{\mu}, \mathcal{D}_{\nu}]$ , or more explicitly

$$F^{a}_{\mu\nu} = \partial_{\mu}A^{a}_{\nu} - \partial_{\nu}A^{a}_{\mu} + gf^{abc}A^{b}_{\mu}A^{c}_{\nu}, \qquad (1.17)$$

which preserves the local gauge invariance.

For example, if  $U = e^{i\alpha(x)}$  represents a U(1) phase abelian transformation, the covariant derivative is given by  $\mathcal{D}_{\mu} = \partial_{\mu} - igA_{\mu}$  and the gauge field transformation law is simply given by  $A'_{\mu} = A_{\mu} + \frac{1}{g}\partial_{\mu}\alpha(x)$ . The kinetic tensor has the form  $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ . No mass term of the form  $\frac{1}{2}m_{A}^{2}A^{\mu}A_{\mu}$  can be added without spoiling the symmetry, because of the  $A_{\mu}$  transformation law: the gauge boson is intrinsically massless. This is nothing but the quantum field description of the electromagnetic interaction (QED), with the boson  $A_{\mu}$  identified as the photon.

Theories with a local non-abelian phase invariance are also possible and go under the name of Yang-Mills theories. To describe the experimental knowledge of the particles and their interactions at the quantum level, two such symmetries, together with an abelian symmetry, are necessary and sufficient. First of all, the Lagrangian exhibits a local U(1) phase invariance. The gauge field associated to it is called  $B_{\mu}$ . 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  $\tau^i/2$ . 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

$$U = \exp\left\{i\left(\beta(x)\frac{Y}{2} + \alpha^{i}(x)\frac{\tau^{i}}{2} + \gamma^{a}(x)\frac{\lambda^{a}}{2}\right)\right\},\tag{1.18}$$

and the covariant derivative which ensures all the three invariances of the theory takes the form

$$\mathcal{D}_{\mu} = \partial_{\mu} - ig' \frac{Y}{2} B_{\mu} - ig \frac{\tau^{i}}{2} W^{i}_{\mu} - ig_{s} \frac{\lambda^{a}}{2} G^{a}_{\mu} , \qquad (1.19)$$

where the scalar Y and the matrices<sup>2</sup>  $\tau^i$  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 interactions:

- $SU(3)_C$ : 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 into Weyl spinors (Eq. 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 gauge bosons. This chiral nature of the

 $<sup>^{2}\</sup>tau^{i}$  are the set of 2×2 complex Hermitian and unitary matrices called Pauli matrices. The  $\lambda^{a}$  are the Gell-Mann traceless and Hermitian matrices.

weak isospin transformations has an immediate consequence: fermion mass terms in the Lagrangian are written as

$$-m\bar{\psi}\psi = -m\bar{\psi}\left[\frac{1}{2}(1-\gamma^5) + \frac{1}{2}(1+\gamma^5)\right]\psi = -m\left(\bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L\right)\,,\qquad(1.20)$$

which manifestly violates gauge invariance, since  $\psi_L$  is a member of an isospin doublet and  $\psi_R$  is a singlet, then, at this point, fermion mass terms must be excluded from the theory;

•  $U(1)_Y$ : 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 is chosen Y = -1 for left-handed leptons. Since right-handed neutrinos do not couple to any of the introduced interactions, they are sterile and do not form a part of the theory.

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

$$-\frac{1}{4}W^{i}_{\mu\nu}W^{\mu\nu}_{i} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}$$
(1.21)

and given the SU(2) algebra

$$W^{i}_{\mu\nu} = \partial_{\mu}W^{i}_{\nu} - \partial_{\nu}W^{i}_{\mu} - ig\epsilon_{ijk}W^{j}_{\mu}W^{k}_{\nu},$$
  

$$B_{\mu\nu} = \partial_{\mu}B_{\nu} - \partial_{\nu}B_{\mu}.$$
(1.22)

From the above equation for  $W_{\mu\nu}$ , self-interaction terms among the gauge bosons are visible, due to the non-Abelian character of SU(2) gauge symmetry. Recalling Eq. 1.16, the following relations can be derived, expressing the transformation law for the vector gauge fields:

$$B_{\mu} \to B'_{\mu} = B_{\mu} + \frac{1}{g'} \partial_{\mu} \beta(x)$$
  
$$\vec{W}_{\mu} \to \vec{W}'_{\mu} = \vec{W}_{\mu} + \frac{1}{g} \partial_{\mu} \vec{\alpha}(x) - \vec{\alpha}(x) \times \vec{W}_{\mu}$$
(1.23)

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

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}} \left( W_{\mu}^{1} \mp i W_{\mu}^{2} \right) ,$$
  
$$\begin{pmatrix} A_{\mu} \\ Z_{\mu} \end{pmatrix} = \begin{pmatrix} \cos \vartheta_{W} & \sin \vartheta_{W} \\ -\sin \vartheta_{W} & \cos \vartheta_{W} \end{pmatrix} \begin{pmatrix} B_{\mu} \\ W_{\mu}^{3} \end{pmatrix} , \qquad (1.24)$$

where  $\vartheta_W$  is the weak mixing angle (Weinberg angle), measured  $\sin^2 \vartheta_W = 0.22295$  [3].  $SU(2)_L$ and  $U(1)_Y$  cannot therefore be considered separately, since the two components of doublets have different electric charge. The relation between electric charge, hypercharge and weak isospin is given by the Gell-Mann-Nishijima formula:

$$Q = T_3 + \frac{Y}{2} \,. \tag{1.25}$$

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 (cfr. 1.23) does not allow for an explicit term such as  $\frac{1}{2}W^i_{\mu}W^{\mu}_i$  or  $\frac{1}{2}B_{\mu}B^{\mu}$ . A possible solution of 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.

### 1.3 Spontaneously broken symmetry

If a theory is described by a Lagrangian which possesses a given symmetry, but its physical vacuum state does not, the symmetry is said to be spontaneously broken. A canonical example of a spontaneous 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 hidden.

It is an important consequence of the spontaneous symmetry breaking in QFT the appearance of massless and spinless particles, when the original symmetry is continuous. This affirmation is in fact the result of a theorem, which goes under the names of Nambu and Goldstone [4], and the new appeared scalar fields are referred to as Goldstone bosons. The number of Goldstone bosons of the broken theory coincides to 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: it is called the Higgs field. 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 realization is to add a complex SU(2) doublet of scalar fields of hypercharge Y = 1:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}.$$
(1.26)

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

$$\mathcal{L} = (\mathcal{D}^{\mu}\phi)^{\dagger} (\mathcal{D}_{\mu}\phi) - V(\phi,\phi^{\dagger}) \quad \text{with} \quad V(\phi,\phi^{\dagger}) = \mu^{2}\phi^{\dagger}\phi + \lambda \left(\phi^{\dagger}\phi\right)^{2}.$$
(1.27)

The expected form of the potential is sketched in Fig. 1.1: for  $\mu^2 > 0$  the scalar potential has a global minimum at  $\phi = 0$ , which would not break the electroweak gauge symmetry. For  $\mu^2 < 0$  the potential has a circle of degenerate minima at

$$\phi^{\dagger}\phi = \frac{1}{2}\left(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2\right) = -\frac{\mu^2}{2\lambda} \equiv \frac{1}{2}v^2, \qquad (1.28)$$

with v the vacuum expectation value of the field  $\phi$  (vev = 246 GeV).



Figure 1.1: Form of the potential  $V(\phi, \phi^{\dagger})$  depending on the sign of  $\mu^2$ , positive on the left and negative on the right.

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

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0\\ v \end{pmatrix} . \tag{1.29}$$

The operators  $T_i$  and Y don't cancel  $\phi_0$ , in particular

$$T_3\phi_0 = -\frac{1}{2}\phi_0$$
 and  $Y\phi_0 = \phi_0$ 

but

$$Q\phi_0 = \left(T_3 + \frac{Y}{2}\right)\phi_0 = 0.$$
 (1.30)

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 symmetry breaking, there remains a residual symmetry generated by Q. This pattern of symmetry breakdown is then described by the following:

$$SU(2)_L \times U(1)_Y \to U(1)_Q. \tag{1.31}$$

If the fluctuation of the  $\phi_1$ ,  $\phi_2$ ,  $\phi_3$  and  $\phi_4$  real scalar fields around the minimum are labelled as  $\vartheta_2$ ,  $\vartheta_1$ , H and  $-\vartheta_3$ , the  $\phi(x)$  is expanded as:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \vartheta_2(x) + i\vartheta_1(x) \\ v + H(x) - i\vartheta_3(x) \end{pmatrix} \\
\approx \frac{1}{\sqrt{2}} \begin{pmatrix} 1 + i\vartheta_3/v & i(\vartheta_1 - i\vartheta_2)/v \\ i(\vartheta_1 + i\vartheta_2)/v & 1 - i\vartheta_3/v \end{pmatrix} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \\
\approx \frac{1}{\sqrt{2}} e^{i\frac{2\vartheta^i(x)}{v}\frac{\tau^i}{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}.$$
(1.32)

Thanks to the SU(2) invariance of the Lagrangian, the three fields  $\vartheta_i(x)$  in Eq. 1.32 can be gauged away with a transformation  $U = e^{-i\frac{2\vartheta^i(x)}{v}\frac{\tau^i}{2}}$ : these are the massless Goldstone bosons, which do not explicitly appear in the final Lagrangian. By expanding the scalar Higgs field Lagrangian in Eq. 1.27 around  $\phi_0$  as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0\\ v + H(x) \end{pmatrix}$$
(1.33)

one finds [5]:

$$\mathcal{L}_{Higgs} = \left\{ \frac{1}{2} \partial_{\mu} H \partial^{\mu} H - \frac{1}{2} 2v^{2} \lambda H^{2} \right\} + \left\{ -\frac{1}{3!} 6v \lambda H^{3} - \frac{1}{4!} 6\lambda H^{4} \right\} \\
+ \left\{ \frac{1}{2} \frac{v^{2} g^{2}}{4} W_{\mu}^{-\dagger} W^{-\mu} + \frac{1}{2} \frac{v^{2} g^{2}}{4} W_{\mu}^{+\dagger} W^{+\mu} \right\} \\
+ \left\{ \frac{1}{2} \frac{v^{2} (g^{2} + g'^{2})}{4} \left( \frac{g W_{\mu}^{3} - g' B^{\mu}}{\sqrt{g^{2} + g'^{2}}} \right)^{2} + 0 \left( \frac{g' W_{\mu}^{3} + g B^{\mu}}{\sqrt{g^{2} + g'^{2}}} \right)^{2} \right\} \\
+ \left\{ \frac{1}{4} \left( 2v H + H^{2} \right) \left[ g^{2} W_{\mu}^{-} W^{+\mu} + \frac{1}{2} (g^{2} + g'^{2}) \left( \frac{W_{\mu}^{3} - g' B_{\mu}}{\sqrt{g^{2} + g'^{2}}} \right)^{2} \right] \right\}.$$
(1.34)

In the first line, originated from the expansion of the potential  $V(\phi, \phi^{\dagger})$ , the kinetic term for the Higgs boson, its mass term and the Higgs boson self-interaction terms are visible. The Higgs mass itself is equal to  $m_H = v\sqrt{2\lambda}$  and it is not predicted by the theory, being  $\lambda$  a free parameter. In the second line, coming from the kinetic term  $(\mathcal{D}_{\mu}\phi)^{\dagger}(\mathcal{D}^{\mu}\phi)$ , the  $W^{\pm}$  vector bosons can be identified in the linear combination of the gauge bosons  $W^{\pm} = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$ . The process of spontaneous symmetry breaking allows them to acquire mass: from the expected form of the mass term the W mass is found to be

$$m_W = \frac{1}{2}vg$$
 with  $W^{\pm}_{\mu} = \frac{1}{\sqrt{2}} \left( W^1_{\mu} \mp i W^2_{\mu} \right)$ . (1.35)

The third line provides the right mass terms for the observed  $Z^0$  and  $\gamma$  vector bosons<sup>3</sup>. The first linear combination of the gauge fields  $W^3_{\mu}$  and  $B_{\mu}$  comes with an appropriate mass term and is

<sup>&</sup>lt;sup>3</sup>The numerical factor  $\sqrt{g^2 + {g'}^2}$  has been introduced in order to normalize the combinations of gauge fields  $gW^3_{\mu} - g'B_{\mu}$  and  $g'W^3_{\mu} + gB_{\mu}$ .

therefore interpreted as the massive Z boson. The second combination of fields is orthogonal to the first one and has null mass. The results can be interpreted as

$$m_Z = \frac{1}{2}v\sqrt{g^2 + {g'}^2}$$
 with  $Z_\mu = \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + {g'}^2}}$  (1.36)

$$m_A = 0$$
 with  $A_\mu = \frac{g' W_\mu^3 + g B_\mu}{\sqrt{g^2 + {g'}^2}}$ . (1.37)

The mixing of  $W^3_{\mu}$  and  $B_{\mu}$  yielding the physical force carriers can be interpreted as a rotation of parameter  $\vartheta_W$ , where

$$\frac{g}{\sqrt{g^2 + {g'}^2}} = \cos \vartheta_W \qquad \text{and} \qquad \frac{g'}{\sqrt{g^2 + {g'}^2}} = \sin \vartheta_W \tag{1.38}$$

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

$$m_Z = \frac{m_W}{\cos\vartheta_W} \tag{1.39}$$

Finally, in the last line of the Lagrangian of Eq. 1.34, the cubic and quartic couplings of the Higgs boson to the weak gauge bosons can be deduced. The coupling of one single Higgs boson to a pair of W or Z bosons is proportional to  $m_W$  and  $m_Z$  respectively:

$$g_{HWW} = gm_W$$
  

$$g_{HZZ} = \frac{g}{2\cos\vartheta_W}m_Z.$$
(1.40)

From this, the following relation for the branching ratios of the Higgs boson into a pair of vector bosons (valid at tree level) can be derived:

$$\frac{BR(H \to W^+W^-)}{BR(H \to ZZ)} = \left(\frac{g_{HWW}}{g_{H_ZZ}}\right)^2 = 4\cos^2\vartheta_W \frac{m_W^2}{m_Z^2} \simeq 2.7.$$
(1.41)

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

$$\mathcal{L}_{SM} = \mathcal{L}_{GWS} + \mathcal{L}_{Higgs} \,, \tag{1.42}$$

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

$$\mathcal{L}_{GWS} = -\frac{1}{4} W^{i}_{\mu\nu} W^{\mu\nu}_{i} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + i \bar{\nu}_{L} \gamma^{\mu} \partial_{\mu} \nu_{L} + i \bar{e}_{L} \gamma^{\mu} \partial_{\mu} e_{L} + i \bar{e}_{R} \gamma^{\mu} \partial_{\mu} e_{R} + + -i \bar{\mathbf{f}}_{L} \gamma^{\mu} \left( -i g \frac{\tau^{i}}{2} W^{i} - i g' \frac{Y}{2} B_{\mu} \right) \mathbf{f}_{L} - i \bar{e}_{R} \gamma^{\mu} \left( -g' \frac{Y}{2} B_{\mu} \right) e_{R}.$$
(1.43)

Re-expressing the above Lagrangian in terms of the physics fields and writing explicitly the covariant derivative, one obtains

$$\mathcal{L}_{GWS} = \mathcal{L}_{CC} + \mathcal{L}_{NC}$$
$$= \left\{ eJ_{\mu}^{em}A^{\mu} + \frac{g}{\cos\vartheta_W}J_{\mu}^Z Z^{\mu} \right\} + \left\{ \frac{g}{\sqrt{2}} \left( J_{\mu}^+ W^{+\mu} + J_{\mu}^- W^{-\mu} \right) \right\}$$
(1.44)

for the neutral and charged part respectively. The electromagnetic coupling constant e has been introduced, thus identifying

$$e = g \sin \vartheta_W \,. \tag{1.45}$$

The following currents have also been defined:

$$J^{em}_{\mu} = Q\bar{f}\gamma_{\mu}f, \qquad (1.46)$$

$$J_{\mu}^{Z} = \frac{1}{2} \bar{f} \gamma_{\mu} \left( c_{V} - c_{A} \gamma_{5} \right) f \qquad c_{V} = T_{3} - 2Q \sin^{2} \vartheta_{W}, \ c_{A} = T_{3} , \qquad (1.47)$$

$$J_{\mu}^{+} = \frac{1}{2} \bar{\nu} \gamma_{\mu} (1 - \gamma_{5}) e. \qquad (1.48)$$

### Fermion masses

An attractive feature of the Standard Model is that the same Higgs doublet which generates  $W^{\pm}$ and  $Z^0$  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 l):

$$\mathcal{L}_{lY} = -G_l \left[ (\bar{l}_L \phi) l_R + \bar{l}_R (\phi^{\dagger} l_L) \right] , \qquad (1.49)$$

where the Higgs doublet has exactly the required  $SU(2)_L \times U(1)_Y$  quantum numbers to couple to  $\bar{l}_l l_R$ . After the breakdown of the symmetry, inserting Eq. 1.33 into Eq. 1.49, one obtains

$$\mathcal{L}_{lY} = -\frac{G_l}{\sqrt{2}} \left[ v(\bar{l}_L l_R + \bar{l}_R l_L) + (\bar{l}_L l_R + \bar{l}_R l_L) H \right] = -\frac{G_l}{\sqrt{2}} \left[ v\bar{l}l + \bar{l}lH \right].$$
(1.50)

In Eq. 1.50,  $G_l$  (called Yukawa coupling) was chosen so as to generate the required lepton mass term, that is

$$m_l = \frac{G_l v}{\sqrt{2}} \tag{1.51}$$

and the Yukawa Lagrangian for the lepton sector can be rewritten as

$$\mathcal{L}_{lY} = -m_l \left[ \bar{l}l + \frac{1}{v} \bar{l}lH \right] \,. \tag{1.52}$$

Not only have leptons acquired mass thanks to the spontaneous symmetry breaking, but a new coupling between the Higgs boson and the leptons has become manifest. As an important consequence the amplitude of a Higgs decay process is proportional to the second power of the mass of the particle the Higgs decays into.

### 1.3.1 Limits on the Higgs mass

Although the Higgs boson mass is not predicted by the theory, both lower and upper limits can be set on theoretical backgrounds [6]. A first upper constraint is found considering 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 ( $\sqrt{s} \simeq 1.2 \text{ TeV}$ ). Diagrams involving the exchange of a Higgs boson among the  $W_L$  allow for a cancellation, which results in a unitary scattering matrix at all energies, provided that  $m_H \lesssim 700 \text{ GeV}/c^2$ .

More restricting bounds on the Higgs mass depend on the energy scale  $\Lambda$  up to which the SM is 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 the constant with energy.

First of all, the Higgs potential described in Eq. 1.27 is affected by radiative corrections which involve the mass of fermions and bosons and depend on the renormalization scale. These radiative corrections may modify the shape of the potential in a way such that an absolute minimum no longer exists and no stable spontaneous symmetry breaking occurs. The requirement of vacuum stability, the  $\lambda$  coefficient positive and large enough to avoid instability up to a certain scale  $\Lambda$ , implies a lower bound on  $m_H$  (stability bound).

Another limit can be imposed since the coupling constant evolution with energy presents a singularity (Landau pole) for some energy value. In order to preserve the perturbativity of the theory, the SM can be considered as an effective theory up to an energy scale  $\Lambda$ : 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 Fig. 1.2.



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 valid. The shaded area indicates the theoretical uncertainties in the calculation of the bounds.

Experimental bounds on the Higgs boson mass are provided by measurements at different experiments. The most sensitive direct search, up to the LHC startup, has been carried out at the LEP accelerator at CERN. No evidence for a signal was observed in data from  $e^+e^-$  collisions

up to center of mass energies of 209 GeV at LEP-II (Fig. 1.3). An experimental lower bound is set to  $m_H > 114.4 \text{ GeV}/c^2$  at the 95% confidence level [7].



Figure 1.3: Observed and expected behaviour of the test statistics  $-2 \ln Q$  as a function of the Higgs mass. Q is the ratio between the signal plus background likelihood and the background only likelihood. The result is the combination of the data collected by the four LEP experiments. Green and yellow shaded bands represent the 68% and 95% probability C.L.

The search for the Standard Model Higgs particle has been performed at Tevatron, both with the CDF and DØ detectors. Fig. 1.4 (left) shows the most recent update for the joint CDF/DØ sensitivity curves, together with the Standard Model prediction [8]. The Tevatron experiments have excluded the presence of a SM Higgs boson between 147 and 179 GeV/c<sup>2</sup> with a confidence level of 95%. A mild excess ( $3\sigma$ , see Fig. 1.4) is found at low masses between 120 and 135 GeV/c<sup>2</sup>, that is interpreted as evidence for the presence of a new particle consistent with the standard model Higgs boson, which is produced in association with a weak vector boson and decays to a bottom-antibottom quark pair [9].

An indirect measurement of  $m_H$  within the Standard Model framework is possible using the precision measurements of the fundamental parameters, e.g.,  $m_Z$ ,  $m_W$  etc., since the Higgs boson mass enters logarithmically the loop corrections [10]. Such measurements have been performed by several experiments and a global fit to these electroweak observables with the Higgs boson mass as a free parameter sets limits on  $m_H$  [11]. Fig. 1.5 shows the  $\chi^2$  curve, derived from high- $Q^2$  electroweak precision measurements, performed at LEP and by SLD, CDF, and D $\emptyset$ , as a function of the Standard Model Higgs boson mass. The preferred value for its mass, corresponding to the minimum of the curve, is at 87 GeV/ $c^2$ , with an experimental uncertainty of  $-27/ + 36 \text{ GeV}/c^2$  (at 68% C.L., derived from  $\Delta \chi^2 = 1$ ). An upper limit at 160 GeV/ $c^2$  is fixed at 95% C.L., increasing to 190 GeV/ $c^2$  when including the LEP-II direct search limit.



Figure 1.4: The observed and expected (median, for the background-only hypothesis) 95% C.L. upper limits on the ratios to the SM cross-section, as functions of the Higgs boson test mass, for the combination of all CDF and DØ analyses (a). Low mass region zoom, with the excess interpreted as a Higgs boson decaying into two b quarks (b).



Figure 1.5: Global fit of all precision electroweak measurements. The yellow shaded area corresponds to the LEP-II direct exclusion [11]. The fit depends logarithmically on the Higgs mass.

#### 1.3.2 Higgs boson production, decay and detection at LHC

The Higgs boson production cross-sections at a pp hadron collider are shown in Fig. 1.6 for a center of mass energy  $\sqrt{s} = 7$  TeV and  $\sqrt{s} = 8$  TeV, that are respectively the energies at which LHC operated in 2011 and 2012.



Figure 1.6: Higgs production cross-sections for the various processes at  $\sqrt{s} = 7$  TeV and  $\sqrt{s} = 8$  TeV as a function of the Higgs mass [12, 13].

As shown in Fig. 1.7, the leading-order diagrams for the interesting production processes are:

- The gluon-gluon fusion  $(gg \rightarrow H)$ , the dominating Higgs production process over the entire mass range accessible at the LHC [14]. It proceeds with a heavy quark triangle loop. Because of the Higgs couplings to fermions, the *t*-quark loop will be dominating.
- The vector boson fusion (VBF,  $qq \rightarrow qqH$ ). In this process, which is about one order of magnitude below the gluon-gluon fusion, the Higgs boson is originated from the fusion of two weak bosons radiated off the incoming quarks (see Chapter 7).
- 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 t quarks (see Chapter 8).

The main decay channels of the Higgs are summarized in Fig.1.8:

- $H \to \gamma \gamma$ , the main channel for the discovery of the Higgs boson at masses below 140 GeV. 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 energetic isolated photons which can be well identified experimentally and the Higgs boson can be detected as a narrow peak above a large background. Since the Higgs width is well below 1 GeV at low masses, the energy resolution of the electromagnetic calorimeter is crucial. The fake photon signals due to  $\pi^0 \to \gamma \gamma$  decays can be rejected by photon isolation and using shower shape variables in the electromagnetic calorimeter.
- $H \to ZZ^* \to 4l$ , that due to its very clean signature with 4 isolated leptons in the final



Figure 1.7: Feynman diagrams for the most important LO production processes for SM Higgs boson.

state, is considered the golden mode for the discovery of the Higgs boson. The backgrounds to this channel are ZZ, tt and  $Z \rightarrow bb$  productions, which can be suppressed in an efficient way by some requirements on the leptons isolation, transverse momentum and invariant mass and by requirements on the event vertex. The Higgs discovery in the ZZ channel is possible with an integrated luminosity of 10 fb<sup>-1</sup> in the whole range of masses between 120 and 500 GeV, apart for the small region around mH ~ 165 GeV.

•  $H \to WW^* \to 2l_2\nu$ , the discovery channel in the mass region  $2m_W < m_H < 2m_Z$  where the Higgs branching ratio into W W is close to one. 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 accurate knowledge of all the possible backgrounds is needed. The main backgrounds are WW,  $t\bar{t}$  and W + jets productions. They can be reduced by requirements on the leptons momentum and isolation, by a jet veto and exploiting the small opening angle between the two leptons which is due to spin correlations. Due to the very large branching ratio this channel extended its feasibility both in the high mass region and in the low mass region, covering, even with a low mass resolution (30 GeV/c<sup>2</sup>), the whole Higgs mass spectrum, 100-600 GeV.



Figure 1.8: Higgs branching ratios as a function of the Higgs mass [12, 13].

# CHAPTER 2\_

# LHC AND CMS

Il lettore che lo sa potrà saltare alcune pagine per riprendere il filo della storia: e per me lo consiglio di far così: giacchè le parole che mi sento sulla punta della penna sono tali da annojarlo.

Alessandro Manzoni, Fermo e Lucia

In this chapter the motivation and some of the features of the Large Hadron Collider (LHC) are exposed (Sec. 2.1). Moreover, a brief discussion about the physics programme of the Compact Muon Solenoid (CMS) detector together with its description according to each of the subdetectors (Sec. 2.2) is presented.

# 2.1 The Large Hadron Collider LHC

The Large Hadron Collider (LHC) is a two-ring, superconducting accelerator and collider installed in the 27 km long Large Electron-Positron (LEP) [15]. The prime motivation of the LHC accelerator is the investigation of the electroweak symmetry breaking for which the Higgs mechanism is presumed to be responsible [16].

The search for a possible Higgs boson is not the only motivation of the LHC. The new accelerator allows the study of the consistency of the Standard Model (SM) at the scale  $\Lambda \simeq 1$  TeV. In addition, precision studies of QCD, electroweak and flavour physics are possible. These measurements might also open a window onto new physics. The LHC will state the last word about the existence of Super Symmetric particles or high-mass intermediate vector bosons like the Z' at the TeV mass scale. In addition the manifestation of extra dimensions could lie just beyond the electroweak energy scale. Furthermore, the understanding of heavy ion collisions physics will experience a giant leap: LHC also endows the ions with an energy of  $\sqrt{s} = 5.5$  TeV to be compared with the 200 GeV attainable by RHIC [17]. Given the machine design centre of mass energy and luminosity (7+7 TeV and  $\mathscr{L} = 10^{34}$  cm<sup>-2</sup>s<sup>-1</sup> for *p-p* collisions), the LHC represents an unprecedented challenge from the point of view of technologies and human resources involved.



Figure 2.1: The LHC collider ring and the various caverns dedicated to the four experiments that will take data: CMS, ATLAS, ALICE, LHCb. At the bottom of the picture is indicated the SPS synchrotron which will inject the beams into the LHC machine

### 2.1.1 LHC operation

The LHC delivered the first proton-proton collisions on the 23<sup>rd</sup> November 2009 at a center of mass energy of 0.9 TeV (450+450 GeV). A new energy record was reached 7 days after, with a collisions with beam at 1.18 TeV each, delivering a total integrated luminosity of few  $\mu b^{-1}$ . During the 2011 run LHC worked with a center of mass energy of 7 TeV and a peak luminosity (increasing during the data taking period) of  $\mathscr{L} = 4 \times 10^{33} \text{ cm}^{-2} \text{s}^{-1}$  in *p*-*p* collisions, while the 2012 run LHC had a center of mass energy of 8 TeV and a luminosity  $\mathscr{L} = 6 \times 10^{33}$ , almost flat during the whole year (3/5 of design luminosity), as shown in Fig. 2.2.

The time evolution of the total integrated luminosity, during stable beams for pp collisions, is shown in Fig. 2.3 for 2011 at  $s = \sqrt{7}$  TeV (left) and for 2012 at  $s = \sqrt{8}$  TeV (right). The integrated luminosity delivered by the LHC is plotted in red and compared to the one recorded by CMS in blue. The integrated luminosity available for physics results is about 95% of the recorded one, given the performances of all detectors in CMS. The LHC timeline foresees operation for proton physics until the end of 2012, with a shutdown from March 2013 that will allow further upgrades to enable collisions at higher energies (13 TeV).

### 2.1.2 LHC properties

The high beam intensities implied by a luminosity of  $\mathscr{L} = 10^{34} \text{ cm}^{-2} \text{s}^{-1}$  exclude the use of anti-proton beams and one common vacuum and magnet system for both circulating beams (as



Figure 2.2: Peak delivered luminosity per day in 2011 and 2012 [18].



Figure 2.3: Total integrated luminosity versus time in 2011 and 2012 [18].

it is done in the TEVATRON) and imply the use of two proton beams. To collide two beams of equally charged particles requires opposite magnet dipole fields in both beams. The LHC is therefore designed as a proton-proton collider with separate magnet fields and vacuum chambers in the main arcs and with common sections only at the insertion regions where the experimental detectors are located. The two beams share an approximately 140 m long common beam pipe along the interaction regions.

The nominal number of bunches for each proton beam is 2808 with a nominal bunch spacing of 25 ns, leading to a nominal bunch crossing rate is 40 MHz. The actual separation during the 2011 and 2012 run is 50 ns. There is not enough room for two separate rings of magnets in the LEP tunnel. Therefore the LHC uses twin bore magnets which consist of two sets of coils and beam channels within the same mechanical structure and cryostat.

The peak beam energy in a storage ring depends on the integrated dipole field along the storage ring circumference. Aiming at peak beam energies of up to 7 TeV inside the existing LEP tunnel implies a peak dipole field of 8.33 T and the use of superconducting magnet technology, namely 9300 liquid Helium cooled superconducting magnets made of a Niobium-Titanium compound and with a running temperature of 1.9 K.

The bunches of protons are prepared and accelerated to 26 GeV in the Proton Synchrotron (PS). Then they are injected in the Super Proton Synchrotron (SPS) where they are brought to the energy of 450 GeV for the final injection into the LHC (see Fig. 2.1).

The interaction points are four, one for each of the big experiments: CMS (Compact Muon Solenoid), ATLAS (A Toroidal LHC ApparatuS), ALICE (A Large Ion Colliding Experiment) and LHCb (the Large Hadron Collider beauty experiment).

CMS and ATLAS are two general purpose experiments, with complementary features and detector choices. CMS will be described in detail in the next section. The LHCb collaboration will aim to perform precision measurements on CP violation and rare decays in order to reveal possible indications for new physics [19]. ALICE is dedicated to heavy ions physics and the goal of the experiment is the investigation of the behaviour of the strongly interacting hadronic matter resulting from high energy Pb nuclei collisions. In those extreme energy densities the formation of a new phase of matter, the quark gluon plasma, is expected [20].

### 2.1.3 Phenomenology of proton-proton collisions

Even if the LHC is a proton collider, the actual particles that interact are the partons (quarks and gluons). The kind of events at a hadron collider can be divided into two main categories: long range and short range collisions. The former occur when there is a small momentum transfer between the incoming partons (soft collisions) and a suppression of particles scattering at large angle. The particles produced in the final state of such interactions have large longitudinal momentum, but small transverse momentum relative to the beam line and most of the collision energy escapes down the beam pipe. The latter are characterized by head-on collisions between two partons of the incoming protons, with a large momentum transfer (hard scattering). In these conditions, final state particles can be produced at large angles with respect to the beam line with creation of massive particles. However, these are rare events compared to the soft interactions. The total proton-proton cross section at 7 TeV is approximately 110 mb, while, for example, the production of a W boson through the annihilation of a quark-antiquark pair has a cross-section of about 90 pb. A comparison of the cross sections of the typical processes at the LHC is shown in Fig. 2.4



Figure 2.4: Cross sections at LHC for different center of mass energies [21].

### 2.1.4 Hadron collider kinematics

A convenient set of kinematic variables for particles produced in hadronic collisions is the transverse momentum  $p_{\rm T}$ , the rapidity Y and the azymuthal angle  $\phi$ , defined as

$$p_T = \sqrt{p_x^2 + p_y^2}, \quad Y = \frac{1}{2} ln \frac{E + p_z}{E - p_z}, \quad p_x = p_T cos\phi, \quad p_y = p_T sin\phi.$$
 (2.1)

with the collision axis is the z axis, in the CM frame of the collision, a particle with energy E and three momentum  $\vec{p} = (p_x, p_y, p_z)$ . These variables have simple transformation properties under longitudinal boosts (i.e. boosts along the beam line direction),  $p_T$  and  $\phi$  being invariant, and

$$Y \to Y + \frac{1}{2} ln \frac{1+\beta}{1-\beta},\tag{2.2}$$

where  $\beta$  is the boost velocity along the z direction. Then the variational variable  $\Delta Y = Y_1 - Y_2$ , is invariant under z-boosts. Experimentally, the rapidity is substituted by the pseudo-rapidity variable, defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right),\tag{2.3}$$

where  $\theta$  is the polar angle between  $\vec{p}$  and the z-axis. The pseudo-rapidity is equal to the rapidity for mass-less particles, then the difference in  $\eta$ , namely  $\Delta \eta$ , is, in first approximation, invariant under longitudinal boosts.

### 2.2 Compact Muon Solenoid CMS

The Compact Muon Solenoid (CMS) is a general purpose detector that is installed at the interaction point number 5 along the LHC tunnel [22]. It is 22 m long, its diameter is 15 m and its weight is about 12500 t. The main requests that CMS needs to satisfy to meet the goals of the LHC Physics program can be summarised as follows [16]:

- Good muon identification and momentum resolution, 1% di-muon mass resolution at 100  $\text{GeV/c}^2$ , and the ability to determine unambiguously the charge of muons with a momentum up to 1 TeV/c.
- Good charged particle momentum resolution and reconstruction efficiency in the tracker. Efficient triggering and offline tagging of  $\tau$ 's and *b*-jets and high vertex reconstruction efficiency.
- Good electromagnetic energy resolution, di-photon and di-electron mass resolution, hermeticity and efficient photon and lepton isolation at high luminosities.
- Good  $E_T^{miss}$  and dijet mass resolution requiring hadron calorimeters with large hermetic coverage and fine lateral segmentation.

The magnetic field in CMS of 3.8 T, enough to bend charged particles to achieve the goal momentum precision, is performed by means of a superconducting solenoid (see Fig. 2.6), 13 m long and with a inner diameter of 5.9 m. The tracker and the calorimetry are placed in the solenoid, while the return magnetic field is large enough to saturate the 1.5 m of iron of the

holding structure that contains 4 layers of muon detectors for the outer muon tracking. The coordinate system adopted by CMS has the origin centred at the nominal collision point. The y-axis points vertically upwards while the x-axis points radially inward toward the LHC ring while the z-axis points along the beam direction toward the Jura mountains from LHC Point 5 (see Fig. 2.5). The azimuthal angle  $\phi$  is measured from the x-axis in the x-y plane. The polar



Figure 2.5: The CMS coordinate system.

angle  $\theta$  is measured from the z-axis.

CMS is composed of four principal subdetectors, from the inside to the outside: the tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the muon chambers. Each of the subdetectors follows the barrel-endcap scheme. In the following part of the chapter a description of each subdetector is given.





### 2.2.1 The Tracker

The CMS tracker[23] is composed of silicon devices. It is 5.4 m long and it has a 2.4 m external diameter. Its volume is 24.4 m<sup>3</sup> and its running temperature is  $-10^{\circ}$  C. A layout of the detector is shown in Fig. 2.7. Surrounding the beam line, the Silicon Pixel Detector (SPD), organised in



Figure 2.7: The CMS tracker detector: an overview. This device is composed of Silicon Pixel Detectors and Silicon Strip Detectors. The Silicon Strip Tracker is composed of three sub-detectors: Tracker Inner Barrel (TIB), Tracker Inner Disk (TID), Tracker Outer Barrel (TOB) and Tracker End Cap (TEC).

three layers, is meant to reconstruct the secondary vertexes of the interactions. It is made of silicon pixels with  $150 \times 150 \ \mu \text{m}^2$  surface and  $250 \ \mu \text{m}$  thickness, divided into modular units of  $6.4 \times 1.6 \text{ cm}^2$ . It covers the region  $|\eta| < 2.6$ . A total of 1440 pixels modules are mounted in three barrel layers at radii between 4.4 cm and 10.2 cm and two endcap disks on each side of the barrel.

A second sub-detector, the Silicon Strip Detector (SSD), surrounds the SPD. It is composed of silicon microstrips arranged in 10 cylindrical layers and 9 endcap disks per endcap. It covers the region  $|\eta| < 2.5$ . A total of 15148 silicon strip modules are arranged in 10 barrel detection layers extending outward to a radius of 1.1 m and 9 disks on each side of the barrel. The active silicon area is about 200 m<sup>2</sup> active, then making the CMS tracker the largest silicon tracker ever built.

In Tab. 2.1 the space resolutions of the various sub-detectors are reported.

detector	resolution $(\mathbf{r}, \phi)$	resolution $z$
SPD barrel	$15 \ \mu { m m}$	$11-17 \ \mu \mathrm{m}$
SPD endcap	$15~\mu{ m m}$	$90~\mu{ m m}$
SSD	$15~\mu{ m m}$	$1 \mathrm{mm}$

Table 2.1: Tracker sub-detectors space resolution, in the (r, $\phi$ ) plane and along the z direction.

One of the major constraints in the design of a tracking system is to reduce as much as possible the amount of material distribution in front of the subsequent calorimeters. For the CMS tracker, the

material budget, shown in Fig. 2.8, constitutes the main source of error in accurate calorimetric measurements of electrons and photons (which convert into  $e^+e^-$  pairs).



Figure 2.8: CMS Tracker budget material in units of radiation length  $X_0$  as a function of  $\eta$ . The maximum is reached in the region of transition between the barrel and the endcaps.

#### 2.2.2 The electromagnetic calorimeter: ECAL

The electromagnetic calorimeter (ECAL) plays an essential role in the study of the Physics of electroweak symmetry breaking, particularly through the exploration of the Higgs sector [24]. The search for the Higgs at the LHC strongly relies on information from ECAL: by measuring the two-photon decay mode for  $m_H \leq 150$  GeV, and by measuring the electrons and positrons from the decay of Ws and Zs originating from the  $H \rightarrow ZZ$  and  $H \rightarrow WW$  decay chain for 140 GeV  $\leq m_H \leq 700$  GeV.

ECAL is a hermetic, homogeneous calorimeter composed of 61200 lead tungstate (PbWO<sub>4</sub>) crystals mounted in the central barrel part, closed by two endcaps including 7324 crystals each (see Fig. 2.9). The endcaps cover the pseudorapidity range  $1.479 < |\eta| < 3$  and consist of identically shaped crystals, grouped into carbon-fibre structures of  $5 \times 5$  elements, called supercrystals. The barrel part of the ECAL covers the pseudorapidity range  $|\eta| < 1.479$ . Crystals for each half-barrel are grouped in 18 supermodules each subtending  $20^{\circ}$  in  $\phi$ . Each supermodule comprises four modules of 500 crystals in the first module and 400 crystals in the remaining three. For simplicity of construction and assembly, crystals have been grouped in arrays of  $2 \times 5$  crystals which are contained in a very thin wall (200  $\mu$ m) alveolar structure and form a submodule. All five alveoli in  $\eta$  contain crystals of the same type; therefore, the number of crystal types is reduced to 17. The front face of the crystals is at a radius of 1.29 m and each crystal has a square cross-section of about  $22 \times 22$  mm<sup>2</sup> and a length of 230 mm corresponding to 25.8 radiation lengths (X<sub>0</sub>). The crystals are truncated pyramid-shaped and mounted in a geometry which is off-pointing with respect to the mean position of the primary interaction vertex, with a 3° tilt both in  $\phi$ and  $\eta$  (see Fig. 2.10). The crystal cross-section corresponds to  $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$  (1°). The crystal volume in the barrel amounts to  $8.14 \text{ m}^3$  (67.4 t). The barrel granularity is 360-fold


Figure 2.9: CMS ECAL geometry schema. The ECAL barrel (EB) is made of 36 SuperModules: 18 in EB+ (z > 0) and 18 in EB- (z < 0) as depicted in (a). The ECAL endcap (EE) is divided in 4 Dees: 2 Dees in EE+ (z > 0) and 2 Dees in EE+ (z > 0) as depicted in (b).

in  $\phi$  and  $(2 \times 85)$ -fold in  $\eta$ . This granularity leads to a natural discrete labelling of the ECAL crystals: the crystals in a single supermodule can be associated with a couple of discrete values  $(\eta, \phi)$  with  $1 \le \eta_d \le 85$  and  $1 \le \phi_d \le 20$ .

CMS has chosen the lead tungstate scintillating crystals for its ECAL since they are dense (8.2 g/cm<sup>3</sup>), have short radiation lenght ( $X_0 = 0.89$  cm) and Molière radius (21.9 mm), are fast (80% of the light is emitted within 25 ns) and radiation hard (see Table 2.2).

material	$\tau$ (ns)	$ ho~({ m g/cm^3})$	light yield $(\gamma/{ m MeV})$
PbWO <sub>4</sub>	25	8.2	50 - 80
BGO $(20^\circ)$	300	7.13	8200
CsI(Tl)	800	4.51	60000

Table 2.2: Three scintillating materials parameters [25]. Belle detector at KEK and Babar at PEP-II present a CsI(Tl) electromagnetic calorimeter while the L3 calorimeter at LEP was made of BGO crystals.

The relatively low light yield of PbWO<sub>4</sub> (50  $\gamma$ /MeV) requires the use of photodetectors with intrinsic gain that can operate in a high magnetic field. Silicon avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps. The sensitivity of both the crystals and the APD response to temperature changes requires a stabilising system (the goal is to keep variations below  $0.1^{\circ}$ C).

Because of the harsh radiation environment (15 rad/h), the crystals behaviour during the LHC runs are affected by the radiation damage. According to the radiation damage models supported by direct measurements [24], while the scintillation mechanism of PbWO<sub>4</sub> stays unaffected, irradiation modifies the PbWO<sub>4</sub> crystal transparency via creation of colour centers that absorb and



Figure 2.10: The disposal of CMS ECAL crystals. The crystal tilt in a transverse view (a) and construction of the crystal  $\varphi$  tilt (b). To produce a non-pointing geometry in  $\eta$ , crystal longitudinal axes are all inclined by 3° with respect to the line joining the crystal front face centre to the interaction point.

scatter the light. At the ECAL working temperature (18° C) the damage anneals and the balance between damage and annealing results in a dose-rate dependent equilibrium of the optical transmission. In the varying conditions of LHC running the result is a cyclic transparency behaviour between LHC collision runs and machine refills. The evolution of the crystal transparency is measured using laser pulses injected into the crystals via optical fibers. Two laser wavelengths are used for the basic source. One, blue, at  $\lambda = 440$  nm, is very close to the scintillation emission peak and is used to follow the changes in transparency due to radiation; the other, near infrared, at  $\lambda = 796$  nm, far from the emission peak, and very little affected by changes in transparency, can be used to verify the stability of other elements in the system. The crystal response to laser light is normalized by the laser pulse magnitude measured using silicon PN photodiodes.

To provide continuous monitoring during LHC runs, laser pulses are sent to ECAL detector elements in the LHC beam gaps of  $3.17\mu s^1$ . Only about 1% of the available beam gaps are used for the ECAL monitoring data taking, with a corresponding data rate of about 100 Hz. The scan of the entire ECAL takes about 30 minutes.

#### 2.2.3 The hadronic calorimeter: HCAL

The CMS Hadronic CALorimeter (HCAL) is meant to measure the energy and the direction of the hadronic jets. Furthermore, it allows evaluation of the transverse missing energy, a typical signature of events with neutral long lived particles escaping detection. To accomplish these goals, the detector must have a high granularity and a good hermeticity, covering the largest possible solid angle.

<sup>&</sup>lt;sup>1</sup>The abort gap in LHC is the area without any bunches in the bunch train that fits the time required for building up the nominal field of the LHC dump kicker

In the region  $|\eta| < 3$  the calorimeter is designed to follow the typical barrel-endcaps structure. It is composed of 3 to 8 cm thick absorbing layers of brass (30% Zn and 70% Cu) lying among active layers of plastic scintillators. The readout is performed by wavelength shifters. The hadron calorimeter barrel is radially restricted between the outer extent of the electromagnetic calorimeter (r = 1.77 m) and the inner extent of the magnet coil (r = 2.95 m). This constrains the total amount of material which can be put in to absorb the hadronic shower. Therefore, an outer hadron calorimeter is placed outside the solenoid complementing the barrel calorimetry. The forward regions  $3 < |\eta| < 5$  are covered by two cylindrical shaped frontal hadronic calorimeters (HF), placed at  $\pm 11$  m from the nominal interaction point along the z region. HF is made of quartz fibres embedded in bulky steel and the readout is performed by photomultipliers. The choice of this technique is related to its very high radiation resistance, needed to survive in the very forward direction. In Fig. 2.11 a lateral view of HCAL is shown.



Figure 2.11: 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.

To cover the very high pseudorapidity region  $(5.2 < |\eta| < 6.6)$  CASTOR[26] detector (Centauro And STrange Object Research) has been placed at 14.37 meters away from the interaction point. CASTOR is made of tungsten absorbers and quartz plates. Cherenkov light is generated inside the quartz plates as they are traversed by the fast charged particles of the shower developed in tungsten.

#### 2.2.4 The muon system

A very accurate muon detection system is placed outside the magnetic coil (Fig. 2.12). Three types of gaseous detectors are integrated in the iron return yoke of the magnet: Drift Tubes (DT), Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC). The choice of the detector technologies has been driven by the very large surface to be covered and by the different radiation environments. DT chambers are used in the barrel region ( $|\eta| < 1.2$ ), where the neutron induced background is small and the muon rate as well as the residual magnetic field is low. In the two endcaps ( $0.9 < |\eta| < 2.4$ ), where the muon rate as well as the neutron induced background rate



is high and the magnetic field is also high, CSC are deployed.

Figure 2.12: The CMS muon stations integrated in the iron return yoke. The trajectory of a typical muon is displayed in the lateral view.

#### 2.2.5 The trigger system

Since the production rate  $(40 \cdot 10^6 \text{ evt/s})$  is high compared to the affordable acquisition rate  $(\sim 100 \text{ evt/s}^2)$ , a powerful trigger system has been implemented, based on two levels. The first one is hardware implemented on each subdetector (Level-1 trigger), the second is the High Level Trigger (HLT) and runs on a dedicated farm of commercial PCs[27].

The Level-1 triggers involve the calorimetry and muon systems, as well as some correlation of information between these systems. The Level-1 decision is based on the presence of trigger primitive objects such as photons, electrons, muons, and jets above set transverse energy thresholds. It also employs global sum of  $E_{\rm T}$  and  $E_{\rm T}^{\rm miss}$ . Reduced-granularity and reduced-resolution data are used to form trigger objects. The maximum rate provided by Level-1 trigger is 100 kHz. The HLT reduces the output rate down to few hundreds 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. This leads to the development of three virtual trigger levels: at the first level only the full information of the muon system and of the calorimeters is used, in the second level the data from the tracker pixels are added and in the third the full event information is available.

In addition to triggers where all events that satifies the requirements are saved (unprescaled triggers), a set of utility triggers, whose rate would be too high due to band saturation, have been developed, where a good event is saved only once every N times (prescaled triggers, with a prescale parameter N). The latter ones are used for detector studies and to study kinematical regions of object reconstruction, such as low  $p_{\rm T}$  leptons.

 $<sup>^{2}</sup>$ In 2012 the affordable acquisition rate is increased to 600 Hz

# CHAPTER 3\_

# PHYSICS OBJECT RECONSTRUCTION

"Si parva licet componere magnis"

Virgilio, Georgiche, IV, 176

The raw data from CMS, skimmed by the trigger system, are analyzed to identify stable particles travelling in the detector (physics objects), to be used in the description of the final states of interest. In this chapter a summary of the techniques used for the events reconstruction is presented. In particular, in Sec.3.1 and in Sec.3.2 the algorithms for muons and electrons are exposed. Sec.3.3 describes jet reconstruction and in Sec.3.4 missing transverse energy definitions are shown. Sec.3.5 describes the general idea of Particle Flow methodology and its application to leptons, jets and missing transverse energy.

## 3.1 Muons

The reconstruction of muons in CMS exploits both tracking and calorimetry information [28, 29]. The high-level muon physics objects are reconstructed in several ways, with the final collection being comprised of three different muon types, Stand-alone, Global and Tracker muons. The ability to reconstruct muons over a wide range of energies and in the whole geometric acceptance of the detector is crucial to the proper recognition of physics signatures at the LHC. The CMS detector is designed to meet these requirements by using several different types of sub-detectors with complementary capabilities. While each sub-detector is able to measure part of a muon's properties, a global muon is the result of combining information from the sub-detectors in order to obtain the best description of the muon.

The reconstruction in the muon spectrometer starts with the determination of hit positions in the DT, CSC and RPC subsystems. Hits within each DT and CSC chamber are then geometrically matched to form segments. The segments are collected and matched to each other to generate seeds that are used as a starting point for the actual track fit of DT, CSC and RPC hits. The result is a reconstructed trajectory in the muon spectrometer, called stand-alone muon. Tracker muons are muon objects reconstructed with an algorithm that starts from a silicon tracker track and looks for compatible segments in the muon chambers. Stand-alone muon tracks are then

matched with tracker tracks to generate global muon tracks, featuring the full CMS resolution. A unique collection of muon objects is assembled from the standalone, global and tracker muon collections.

Figure 3.1 shows the measurements of muon transverse momentum resolution versus  $\eta$  in data and in simulation.



Figure 3.1: Relative transverse momentum resolution  $\sigma(p_T)/p_T$  in data and simulation measured by applying two different methods (MuScleFit, using Z lineshape convoluted with gaussian function, and SIDRA, using the full simulation of Z decay) to muons produced in the decays of Z bosons. The thin line shows the result of MuScleFit on data, with the grey band representing the overall (statistical and systematic)  $1\sigma$  uncertainty of the measurement. The circles are the result of MuScleFit on simulation. The downward-pointing and upward-pointing triangles are the results from SIDRA obtained on data and simulation, respectively; the resolution in simulation was evaluated by comparing the reconstructed and true  $p_T$ . The uncertainties for SIDRA are statistical only and are smaller than the marker size [30].

## 3.2 Electrons

Electrons interact in CMS with the tracker system and the electromagnetic calorimeter. Therefore, they need two main pieces of information in order to be reconstructed and identified: an ECAL cluster and a corresponding fitted charged particle track.

For a single electron reaching the ECAL surface, an electromagnetic shower starts within the first centimeters of the ECAL crystals and most of the electron energy is collected within a small

#### **3.2. ELECTRONS**

matrix of crystals around the hit one. The interaction with the tracker (tracker material budget) causes electrons to loose part of their energy radiating photons by bremsstrahlung effect before reaching the calorimeter. As the electrons loose energy, the effect of the magnetic field is to enhance the bending of their trajectories, which ultimately results in a spread of the irradiated photons along the  $\phi$  coordinate. Therefore, to obtain an accurate measurement of the electron energy, it is essential to account for bremsstrahlung photons. This is the purpose of the first stage in the electron reconstruction sequence, which goes under the name of superclustering, that results in a measurement of the electrons energy. Once energy is measured, the reconstruction proceeds with the track-building stage and the geometrical matching between tracker and ECAL information.

#### 3.2.1 Supercluster (SC) reconstruction

Two algorithms, known as Hybrid and Island algorithms, are used to group crystals interested by electrons energy deposit and recollect bremsstrahlung photons [31]. Both algorithms start from single crystals seeds with at least 1GeV of measured transverse energy.

The first one, used in the barrel, then looks for  $1 \times 3$  or  $1 \times 5$  dominoes of crystals in the  $\eta - \phi$  plane around the seed, 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 along  $\phi$ . Different dominoes are then grouped together along  $\phi$ , thus obtaining a set of clusters that goes under the name of supercluster. Figure 3.2 shows the schemes of the Hybrid clustering algorithm.

The Island algorithm in the endcap, instead, builds clusters by connecting crystals in rows along  $\phi$  containing energies decreasing monotonically when moving away from the seed crystal. Superclusters are built by collecting other Island clusters along a  $\phi$  road in both directions around each Island clusters. Each crystal of transverse energy above the 0.18 GeV threshold seeds a cluster, provided it represents a local maximum in energy when compared to its four neighbours by side. A 5×5 matrix of crystals is built around the seed, including only those not already belonging to another cluster.



Figure 3.2: Schemes of the Hybrid clustering algorithm used in the ECAL barrel region[31].

#### 3.2.2 Electron track

Once a supercluster is found, the reconstruction proceeds with the track-building stage. Under both +1 and -1 charge hypotheses, the supercluster position is back-propagated in the magnetic field to the nominal vertex, to look for compatible hits in the pixel detector. Once track seeds (pairs or triplets of hits) in the inner tracker layers are found, electron tracks are built: trajectories are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian Sum Filter (GSF). This procedure approximates the electron energy loss probability density function, well described by the Bethe-Heitler model [32], with a sum of Gaussian functions, in which different components model different degrees of hardness of the bremsstrahlung in the layer under consideration [33].

#### 3.2.3 GSF Electron

In the final stage, the supercluster and track information are merged. The energy measurement  $E_{\rm sc}$  provided by the electromagnetic calorimeter is combined with the tracker momentum measurement  $p_{\rm tk}$  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 energy of the intrinsic calorimetry and tracking resolutions, and from the fact that  $p_{\rm tk}$  and  $E_{\rm sc}$  are differently affected by the bremsstrahlung radiation. For high energy electrons (E> 15 GeV) the resolution is dominated by the ECAL performance, while for low energy electrons the tracker momentum resolution is the most important, as shown in Fig. 3.3.



Figure 3.3: The fractional resolution (effective RMS) is plotted as a function of generated energy E as measured with the ECAL supercluster (downward triangles), the electron track (upward triangles) and the combination of the two (circles) [34].

#### 3.2.4 Energy corrections

The raw energy as measured in the ECAL is spoilt by detector effects and the ability to recover all the electron energy (shower containment leakage). The measured energy of the electromagnetic objects results underestimated, which is mainly due to the interaction of electromagnetic particles with the tracker material, producing bremsstrahlung and photon conversions. A set of multiplicative corrections to the raw energy are measured to recover this loss. In addition to a better energy scale, the corrections allow to improve the resolution which is crucial for precision analyses involving photons or electrons. The correct superclusters energy is computed as described in Eq. 3.1:

$$E_{e,\gamma} = F_{e,\gamma}(\eta, E_T) \cdot \sum_{xtal \in cluster} G \cdot c_{xtal} \cdot A_{xtal}$$
(3.1)

where  $A_{xtal}$  is the signal amplitude in ADC counts in the single crystals that compose the supercluster, G is a global scale calibration term and  $c_{xtal}$  is the crystal by crystal inter-calibration coefficient, that uniforms the response of each crystal. The correction  $F_{e,\gamma}(\eta, E_T)$  takes into account energy containment losses and the response of the clustering algorithm to bremsstrahlung losses.

Once the supercluster is built, its position is reconstructed in the ECAL, by means of a weighted sum of the energy depositions:

$$\vec{X} = \sum_{i} w_i \vec{x}_i / \sum_{i} w_i .$$
(3.2)

In the previous formula the sum runs over the crystals of the supercluster, while  $\vec{x}_i$  is the position of the crystal *i*. The value  $w_i$  is a weight of the crystal [31], based on the logarithm of the fraction of the cluster energy contained in the crystal, calculated with the formula:

$$w_i = w_0 + \log\left(\frac{E_i}{\sum_j E_j}\right) \tag{3.3}$$

where the weight is constrained to be positive, or is otherwise set to zero and  $w_0$  is a parameter that controls the smallest energy fraction a crystal can have to be considered in the sum, whose optimized value is 4.2 [31].

The crystals in the CMS ECAL are quasi-projective (see Sec.2.2.2) and do not exactly point to the nominal interaction vertex. So the lateral position  $(\eta, \phi)$  of the crystal axis depends on depth. A depth  $t_{max}$  is defined as the longitudinal centre of gravity of the shower, and its optimal mean value varies logarithmically with the shower energy:  $A \cdot (B + log(E))$ , where the parameters are different for electrons and photons, as photons penetrate deeper in the crystals before showering.

#### 3.3 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 and collinear gluons and quarks, until a point where a non-perturbative transition causes the partons combine into colourless hadrons. The result is a spray of collimated particles, referred to as a jet, which, due to energy conservation, reflects the energy and the flight direction of the initial parton.

Jets are detected as a cluster of tracks and energy deposits in a defined region of the detector. Due to the intrinsic compositeness of such objects, a jet cannot be defined until an algorithmic procedure to recombine different particles is chosen. Given the high probability of a collinear or soft gluon to be emitted by a parton, jet algorithms must satisfy basic 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 within the shower is substituted by two collinear particles of momentum p/2;
- infrared safety: the outcome of the jet clustering 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- $k_T[35]$ . This algorithm proceeds via the definition of two distances for each particle *i* in the list of particles, namely

$$d_{ij} = \min\left(\frac{1}{p_{\mathrm{T}i}^2}, \frac{1}{p_{\mathrm{T}j}^2}\right) \frac{\Delta R_{ij}^2}{R^2}$$

$$d_{iB} = \frac{1}{p_{\mathrm{T}i}^2}$$
(3.4)

In the above equation,  $d_{ij}$  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  $\eta \times \phi$  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 recombined by adding together their four-momenta, otherwise the *i* particle is promoted to jet. The whole procedure is iterated and the algorithm stops when only jets are left.

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

The jet momentum is determined as the vectorial sum of all particle momenta in it. A set of corrections have to be applied on reconstructed jets to reflect the energy of the starting parton. The jet correction scheme adopted in CMS is factorized into subsequent steps, each of them addressing a different aspect of the jet reconstruction.

- 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 proton-proton interactions within the same bunch crossing or from the underlying event that randomly overlaps with the jet area. This correction is determined both in data and in Monte Carlo on a event-by-event basis.
- Level 2 (relative) corrections: these corrections are meant to equalize 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  $\gamma$ +jets events, where the event energy balance allows to compare the jet energy to the photon, precisely measured in ECAL.

Level 2 and 3 corrections are derived in simulated events, and further checked on real data. Potential differences between data and Monte Carlo are accounted for with residual correction factors for jets in real data. Further details about the performances of jet reconstruction at CMS with the first 36 pb<sup>-1</sup> of data collected in 2010 can be found at [36]. As an example, Figure 3.4 reports the jet energy resolution expected from the simulation and measured in data for jets reconstructed with an anti- $k_T$  algorithm of R parameter 0.5 within the tracker acceptance.



Figure 3.4: Data measurements, compared to the Monte Carlo truth resolution before (red dashed line) and after correction for the measured discrepancy between data and simulation (red solid line) compared to data for PF jets in different  $\eta$  ranges [36].

### 3.4 Missing transverse energy, MET

Missing transverse energy,  $\vec{E_{T}}^{miss}$ , is the only physics quantity, definable at hadron colliders, used as signature of invisible particles like neutrinos.  $\vec{E_{T}}^{miss}$  is defined as the negative vector sum of the transverse momenta of all final-state particles in the event. In the hypothesis that all detectable particles are properly reconstructed,  $\vec{E_{T}}^{miss}$  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 ( $\leq 1 \text{GeV}$ ). In practice, this is not possible: since a fraction of the total event energy is unavoidably lost in the beam pipe or only coarsely reconstructed in the forward

calorimeters, additional contributors besides undetectable particles affect  $\vec{E_T}^{\text{miss}}$ . Depending on the basic objects used in the sum, three different  $E_{\text{T}}^{\text{miss}}$  types can be considered [37]:

$$\vec{E}_T^{calo} = -\sum_{i,caloTower} \vec{E}_T^i - \sum_{j,muon} \vec{p}_T^j + \sum_{j,muon} \vec{E}_T^j$$
(3.5)

since muons are minimum ionizing particles, their contribution is estimated from tracker momentum measurement.

$$\vec{E}_T^{calo+trk} = -\sum_{i,caloTower} \vec{E}_T^i - \sum_{j,muon} \vec{p}_T^j + \sum_{j,muon} \vec{E}_T^j + \sum_{k,track} \vec{p}_T^k - \sum_{k,track} \vec{E}_T^k$$
(3.6)

$$\vec{E}_T^{PF} = -\sum_{i.PFcand} \vec{E}_T^i. \tag{3.7}$$

See Sec. 3.5 for further information about Particle Flow.

### 3.5 Particle Flow reconstruction

The Particle Flow (PF) is a whole-event reconstruction technique whose purpose is the reconstruction and identification of each single particle produced in each proton-proton collision with an optimized combination of all sub-detectors information [38, 39]. In this process, the identification of the particle type (photon, electron, muon, charged hadron, neutral hadron) plays a crucial role in the determination of the particle direction and energy.

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

In particular the PF reconstruction follows the following scheme:

- Photons (e.g. coming from  $\pi^0$  decays or from electron bremsstrahlung) are identified as ECAL energy clusters not linked to the extrapolation of any charged particle trajectory to the ECAL.
- Electrons (e.g. coming from photon conversions in the tracker material or from hadron leptonic decays) are identified as a primary charged particle track and potentially many

ECAL energy clusters, corresponding to this track extrapolation to the ECAL and to possible bremsstrahlung photons emitted along the way through the tracker volume.

- Muons (e.g. from hadrons leptonic decays) are identified as a track in the central tracker consistent with either a track or several hits 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.
- Neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory or 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. 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 attached 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.

The list of particles resulting from the operation of the PF algorithm on a whole event represents the best description of the event at the particle level, according to the information provided by the CMS detector and the intrinsic energy and position resolutions of the different sub-detectors. Figure 3.5 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 that remains can be clustered into jets, as explained in the Sec. 3.3. In this approach, jets and leptons are naturally disentangled, since the same energy deposits or tracker hits cannot have contributed to the reconstruction of two distinct objects.



Figure 3.5: Reconstructed jet energy fractions as a function of pseudorapidity in data and in Monte Carlo. From bottom to top in the central region: charged hadrons (from pileup and from jets), photons, neutral hadrons, electrons and muons. In the forward regions: hadronic deposits, electromagnetic deposits[40].

# CHAPTER 4\_

## ECAL ALIGNMENT

There's no sense in being precise when you don't even know what you're talking about

John von Neumann

As described in Sec.2.2.2, the high granularity of the electromagnetic calorimeter allows to reconstruct not only the energy of a particle (electrons and photons in primis) but also its impact position on the detector surface. To translate the local position measurement in ECAL to a global one in the laboratory rest frame, the elements of the detector need to be aligned with respect to CMS. Having an accurate position measurement of particles impacting on the calorimeter is very important for the photons, since it is the only way to measure their trajectory and therefore is critical to reconstruct invariant masses, for example for the search of the Higgs boson in the  $H \rightarrow \gamma \gamma$  channel.

Similar accurancy is required to correctly match the energy deposits in ECAL to the hits in the tracker detector both for the trigger and the offline electron reconstruction. In fact, to identify and remove fake electrons, typical electron identification asks for a difference of at most  $4 \cdot 10^{-3}$  units in pseudorapidity and 20 mrad in  $\varphi$  between the position extrapolated from the tracker and the one reconstructed by ECAL. Studies based on simulations [31] demonstrate that a spatial resolution of about  $10^{-3}$  units in  $\eta$  and 1.6 mrad in  $\varphi$  can be reached on 35 GeV electrons using ECAL only.

After the ECAL installation in the CMS detector, the position of each supermodule has been determined, by means of photogrammetry measurements, with an accuracy of about 1 mm [41]. However, after the switching on of the magnet, it is necessary to perform a further alignment, with a dedicated *in situ* data analysis, that reduces this error below the requirements for the electron identification and photon pairs invariant mass reconstruction.

In Sec. 4.1 the ECAL alignment procedure, by means of track measurement of isolated electrons produced by the W decay, is described. Sec. 4.2 is dedicated to the expected precision based on a Monte Carlo study and the goal of ECAL alignment. The alignment performances on data are reported in Sec. 4.3, while Sec. 4.4 summarizes the results obtained.

# 4.1 Alignment procedure

The alignment procedure compares for each electron the position provided by the ECAL detector (supercluster position, SC) to the point of closest approach of the tracker track to the ECAL position.

The distribution of the energy which is deposited by an electron in a supercluster depends on the impact position of the electron. Therefore, as described in Section 3.2, it is possible to measure the position with ECAL, defined as the point of maximum shower activity in the crystals matrix, as shown in Fig. 4.1 (a), by means of a weighted average of the positions of the crystals involved in the energy deposit [33]:

$$\vec{X} = \frac{\sum_{xtal \in SC} w_{xtal} \vec{x}_{xtal}}{\sum_{xtal \in SC} w_{xtal}} \quad \text{with} \quad w_i = w_0 + \log\left(\frac{E_i}{\sum_{j \in SC} E_j}\right). \tag{4.1}$$

A position resolution of about 1 mm is obtained using ECAL information as shown in [42] during test beam studies.

The tracker based position is the point of closest approach to the supercluster position, extrapolating from the innermost track position and direction [33], as shown in Fig. 4.1.



Figure 4.1: The tracker based position is the point of closest approach to the supercluster position, extrapolating from the innermost track position and directions. In (a) the position reconstructed by ECAL is indicated with a cross [24], while in (b) the position extrapolated from the tracker is the red spot. The vector that connects the ECAL position to the tracker position is used to define  $\Delta \varphi$  and  $\Delta \eta$  of Eq. 4.3.

The distance along the  $\eta$  and  $\varphi$  directions of the two points are used to build the following  $\chi^2$  function:

$$\chi^2 = \chi_+^2 + \chi_-^2 \tag{4.2}$$

The functions  $\chi^2_{\pm}$  depend on the charge of the lepton (electron or positron) and are calculated according to Eq. 4.3:

$$\chi_{\pm}^{2} = \sum_{lepton} \frac{(\Delta \varphi - \left\langle \Delta \varphi_{\pm}^{MC} \right\rangle)^{2}}{\varepsilon_{\varphi}^{2}} + \frac{(\Delta \eta - \left\langle \Delta \eta^{MC} \right\rangle)^{2}}{\varepsilon_{\eta}^{2}}$$
(4.3)

where  $\varepsilon$  is the error associated to the SC position determination and depends on electron energy [42]. The  $\Delta \varphi$  and  $\Delta \eta$  measured from data are shifted to Monte Carlo expectation values  $\langle \Delta \varphi^{MC} \rangle$  and  $\langle \Delta \eta^{MC} \rangle$ , that are not expected to be zero because of the interplay between the electron trajectory bending in the magnetic field and the tilt of the crystals (see Sec. 2.2.2). The  $\Delta \varphi^{MC}$  used is different for electrons and positrons because of their opposite  $\varphi$  bending in the magnetic field (see Sec. 4.2).

For each ECAL element (Supermodules in the barrel, Dees in the endcap), see Fig. 2.9, the  $\chi^2$  function is calculated and is minimized with respect to the three-dimensional traslations and the three Euler angles, describing a possible tilt of the ECAL modules with respect to their nominal positions, defined in the CMS reference system.

#### 4.1.1 Samples used for the analysis

The first ECAL alignment was performed in 2010 showing a big improvement in ECAL position reconstruction. During 2011 and 2012 data taking the ECAL alignment corrections were found to be small thanks to better tracker alignment precision and the increased statistics available: while in 2010 only translational degrees of freedom were considered for each ECAL subregion, starting from 2011 also rotations have been considered.

Electrons used for the study are selected by means of single electron triggers and the so called W/Z skim has been used: with respect to standard analysis in CMS, low level information, such as ECAL crystals response and tracker hits, is needed for ECAL alignment, since a full reconstruction of electrons is required in order to take into account new tracker alignment and to test the performances of the ECAL alignment itself. The skim is also based on isolation and identification requirements for electrons and on kinematic cuts <sup>1</sup>, namely:

- $p_{\rm T} > 20 \text{ GeV}$
- Tracker Isolation /  $p_{\rm T} < 0.1$
- HCAL Isolation /  $p_{\rm T} < 0.1$
- ECAL Isolation /  $p_{\rm T} < 0.1$
- HCAL / ECAL < 0.1
- $\sigma_{\eta\eta} < 0.014 \ (0.035 \text{ in EE})$
- $E_{\rm T}^{\rm miss}$  > 12 GeV (for W $\rightarrow$ e $\nu$ ) or  $M_{ee}$  > 40 GeV/c<sup>2</sup> (for Z $\rightarrow$ ee)

These selections have an efficiency on W $\rightarrow e\nu$  events of 60% and a multijet background rejection of 99.5%.

<sup>&</sup>lt;sup>1</sup> In 2011 the momentum thresholds for electrons were increased to cope with the increased instantaneous luminosity.

#### 4.1.2 Event selection

The main sources of background to this analysis come from the presence of non isolated electrons, produced in the decay of intermediate states in QCD events, or fake electrons due to misidentification of jets. To keep under control background rate the standard electron identification and isolation has been applied [33]. Electrons from the W decay are selected by means of a minimal threshold on the electron transverse energy and a missing transverse energy requirement. The selections applied [43] are:

- golden electron selections [33]:
  - only one ECAL cluster
  - low bremsstrahlung radiation,  $f_{brem} = (p_{in} p_{out})/p_{in} < 0.5$ , where  $p_{in}$  and  $p_{out}$  are respectively the momentum measured from the first layers of the tracker and one from the last layers.
  - the ratio between the energy measured by ECAL and momentum measured by the tracker (E/p) must be more than 90%
- $E_T^{SC} > 20$  (30) GeV in 2010 (2011)
- $E_{\rm T}^{\rm miss} > 20 \text{ GeV} (30 \text{ GeV in EE})$
- $m_T > 30$  (50) GeV in 2010 (2011)
- electron isolation (WP80 [33])

Fig. 4.2 shows the  $\eta$  spectrum of the selected electrons. As it can be seen, the analysis selects mainly the products of Ws decay. The golden electron selection ensures that the supercluster associated to the electron is composed of only one basic cluster, therefore removing superclustering and bremsstrahlung effects from position reconstruction. During the 2011/2012 run, the electron identification selections have been tightened in order to follow the evolution of the trigger thresholds and to have a cleaner sample of W $\rightarrow e\nu$  events.

## 4.2 Monte Carlo alignment and expected precision

The  $\Delta \varphi^{MC}$  and  $\Delta \eta^{MC}$  distributions versus  $\eta$  are not flat in the simulation due to the interplay between the electron trajectory bending in the magnetic field and the tilt of the crystals in each Supermodule. Fig. 4.3 shows the distribution of  $\Delta \varphi^{MC}$  for electrons and positrons: the behaviour in the two emispheres is different both for positive and for negative particles, and is not centered at zero. Fig. 4.4 shows the distribution of  $\Delta \eta^{MC}$ , where electrons and positron contributions are added since they have the same behaviour as far as  $\Delta \eta$  is concerned. Even in this case a difference in the  $\Delta \eta$  distribution for negative and positive values of  $\eta$  is visible.

In order to evaluate the error on the alignment performed with isolated electrons, the procedure has been applied to a simulated  $W \rightarrow e\nu$  sample, with the same statistics available in data, where each electron has been artificially mis-aligned with deviations comparable to the ones measured in data. The difference between the obtained alignment offsets and the artificial mis-alignment is used to estimate the precision of the procedure. Fig. 4.5 shows the distribution of the difference for barrel (left) and endcaps (right) respectively: the first one shows a spread of about 0.5 mm, the second of about 0.7 mm. Considering the mean distance between the interaction point and



Figure 4.2: The pseudorapidity distribution of electrons mainly due to  $W \rightarrow e\nu$  process. Data are superimposed to the Monte Carlo samples, which are summed and normalized to the data.



Figure 4.3: The  $\Delta \varphi^{MC}$  versus  $\eta$  distribution in Monte Carlo sample for electrons (left) and positions (right). The different behaviour in the two emispheres is apparent for both positive and negative particles.



Figure 4.4: The  $\Delta \eta^{MC}$  versus  $\eta$  distribution in Monte Carlo sample.

the supercluster (1.3 m), these figures correspond to about  $10^{-3}$  units in both  $\Delta\varphi$  and  $\Delta\eta$ . Figure 4.6 shows the distribution of the difference for barrel in x, y and z separately. The precision on the z axis is better than the one on the x-y plane: this is due to the fact that the z axis is constrained by the  $\Delta\eta$  measurement, that is less biased by electron bremsstrahlung.



Figure 4.5: Difference between the artificial mis-alignment introduced in the simulation, and the obtained one after the re-alignment of the sample, for the ECAL barrel (left) and endcaps (right) respectively.

## 4.3 Alignment performances

After measuring the position of ECAL by means of the alignment technique, the events are reconstructed with the new conditions and a closure test is performed by comparing the  $\Delta\varphi$ and  $\Delta\eta$  distributions after alignment and reconstruction to the MC ones, as shown in Fig. 4.7. In Fig. 4.8, 4.9, 4.10 and 4.11 the same distributions are shown in more detail, considering separately barrel and endcap contributions, electrons and positrons. The distribution with the previous ECAL alignment coefficients are also shown: without the new ECAL alignment, the



Figure 4.6: Difference between the artificial mis-alignment introduced in the simulation, and the obtained one after the re-alignment of the sample, for the ECAL Barrel, in  $\Delta x$  (left),  $\Delta y$  (middle) and  $\Delta z$  (right).



Figure 4.7: The  $\Delta \varphi$  and  $\Delta \eta$  distributions for collision data electrons with the full 2011 sample. Data are superimposed to the Monte Carlo prediction, which is normalized to the data. Data points shown are after ECAL alignment procedure is performed.



position reconstructed by ECAL would be wrong, thus preventing a good electron identification and a good photon position reconstruction. The main improvement is found to be in the endcap.

Figure 4.8: The  $\Delta \varphi$  distributions for collision data electrons in the barrel with the full 2011 sample. Data are superimposed to the Monte Carlo prediction, which is normalized to the data. Data points shown are after ECAL alignment procedure is performed [44].

### 4.4 Conclusions

The ECAL detector has been aligned *in situ* exploting the comparison of isolated electrons position measured in the ECAL and the extrapolated position of the associated trajectory measured



Figure 4.9: The  $\Delta \varphi$  distributions for collision data electrons in the endcap with the full 2011 sample. Data are superimposed to the Monte Carlo prediction, which is normalized to the data. Data points shown are after ECAL alignment procedure is performed [44].



Figure 4.10: The  $\Delta \eta$  distributions for collision data electrons with the full 2011 sample. Data are superimposed to the Monte Carlo prediction, which is normalized to the data. Data points shown are after ECAL alignment procedure is performed [44].



Figure 4.11: The  $\Delta \eta$  distributions for collision data electrons with the full 2011 sample. Data are superimposed to the Monte Carlo prediction, which is normalized to the data. Data points shown are after ECAL alignment procedure is performed [44].

53

in the tracker. The first alignment was performed in 2010 with few pb<sup>-1</sup>, then updated with the data collected during 2011 and 2012. The goal precision for the electrons identification and di-photon resonances reconstruction has been met: a precision of  $2 \cdot 10^{-3}$  rad in  $\Delta \varphi$  and  $2 \cdot 10^{-3}$  units in  $\Delta \eta$  has been obtained.

# CHAPTER 5

### THE DI-JET STUDIES

Qui e nuce nuculeum esse volt, frangit nucem.

Tito Maccio Plauto

Using the early data collected by the CMS detector at LHC during 2010, with proton-proton collisions at a center of mass energy of 7 TeV, it was possible to study the production of forward jets in conjunction with a central jet. The CMS calorimeter coverage allows this measurement to be done over a pseudo-rapidity range never investigated before, extending to  $|\eta| < 5.2$ . Such final state can give informations on multi-parton interactions and multi-jet production, in particular, such measurements can allow the study of different types of parton radiation dynamics. In addition, understanding the dynamics of forward jet production is essential for the control of the backgrounds in searches of the Higgs boson produced via the vector-boson fusion mechanism. In this Chapter the measurement of the differential cross sections  $d^2\sigma/dp_T^f d\eta^f$  and  $d^2\sigma/dp_T^c d\eta^c$  for the simultaneous production of at least one forward jet (f) and at least one central jet (c) is presented.

The luminosity analyzed corresponds to the first  $3.14 \text{ pb}^{-1}$  of proton-proton collision data [45, 46], when the trigger for low  $p_{\rm T}$  di-jet was not prescaled and the pile-up was very low (the average number of interactions per bunch crossing was 0.3). In addition to the intrinsic value of the measurement, the di-jet study is one of the first measurements involving jets performed by CMS.

In Sec. 5.1 the motivations of this measurement are described in detail. Sec. 5.2 describes the triggers and the event selection, Sec. 5.3 is dedicated to the details of the measurement while in Sec. 5.4 the systematics of this analysis are presented. The results are reported in Sec. 5.5.

### 5.1 Theoretical motivations

The study of the production of forward jets in conjunction with a central jet gives information on multi-parton interaction and multi-jet production. In particular such measurements can allow the study of different types of parton radiation dynamics as implemented in the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [47, 48, 49, 50] or the Balitski-Fadin-Kuraev-Lipatov (BFKL) [51, 52, 53], or the Ciafaloni-Catani-Fiorani-Marchesini (CCFM) evolution equations [54, 55, 56, 57]. In addition, understanding the dynamics of forward jet production is essential for the control of the backgrounds in searches for the Higgs boson produced via the vector-boson fusion mechanism. In general, the understanding of these QCD processes is also an important ingredient for the measurement of the vector-boson scattering cross section, which is fundamental to understand the Electroweak Symmetry Breaking mechanism [58].

The inclusive cross sections for hard production processes are calculated in terms of the parton model formula:

$$\sigma_{H_1H_2}(p_1, p_2) = \sum_{ij} \int dx_1 dx_2 \ f_i^{H_1}(x_1, \mu_f) f_j^{H_2}(x_2, \mu_f) \ \hat{\sigma}_{ij}(x_1p_1, x_2p_2, \mu_f, \mu_R)$$
(5.1)

where the main ingredients are:

- $\sigma_{H_1H_2}(p_1, p_2)$  represents the hard process cross section (e.g.  $t\bar{t}$ , W, Z, ...);
- i and j are generic parton indices (quarks, anti-quarks and gluons);
- $f_i^H(x, \mu_f)$  are called parton distribution functions (PDF); they represent the probability to find a parton *i* with a fraction *x* of the total four-momentum of the incoming proton;
- $\hat{\sigma}_{ij}(x_1p_1, x_2p_2, \mu_f, \mu_R) = \hat{\sigma}_{ij}^0(\hat{p}_1, \hat{p}_2) + \alpha_s(\mu_R^2)\hat{\sigma}_{ij}^1(\hat{p}_1, \hat{p}_2, \mu_f^2) + O[\alpha_s(\mu_R^2)^2]$  is the short distance cross section calculated in the perturbation theory framework, at a given QCD perturbative order.

Eq. 5.1 depends on two arbitrary scales: the renormalization scale  $\mu_R$ , which scans the evolution of the strong coupling constant, and the factorization scale  $\mu_f$ , that enters in the theory in order to regularize collinear divergences from initial state emission.

In the case of the production of a forward jet, it is possible to introduce a relation between jet rapidity and x of one of the two incoming partons. The higher the jet rapidity the lower is x, according to [59]:

$$x^{min} = \frac{x_T e^{-\eta}}{2 - x_T e^{\eta}}, \quad \text{with} \quad x_T = 2p_T / \sqrt{s}$$
 (5.2)

With the rapidity ranges that can be probed at LHC (up to  $\eta \simeq 5$ ), the transverse energy of the jets ( $\simeq 50 \text{GeV}$ ) and the center of mass energy ( $\sqrt{s} = 7 \text{ TeV}$ ) it is possible to probe x down to  $10^{-5}$ , thus allowing to study the PDF in a different phase space with respect to previous experiments, as reported in Fig. 5.1. Thus the measure of the differential cross section of forward jets is also a test of the PDF used for LHC.

The results obtained are then compared with the predictions from different Monte Carlo event generators, PYTHIA 6.422 with D6T and Z2 tune [60, 61], PYTHIA 8.135 with Tune 1 [62], HERWIG 6.510.3 [63] (with underlying events modelled with Jimmy [64]), HERWIG++ [65], CASCADE 2.2.04 [66, 67], HEJ [68, 69] and POWHEG [70] matched with PYTHIA and with HERWIG parton showers. The PYTHIA and HERWIG Monte Carlo event generators are based on the DGLAP parton evolution equation. The CASCADE event generator, based on the CCFM approach, accounts not only for parton evolution in virtuality ( $Q^2$ ), but also in Bjorken x. The HEJ event generator provides, at parton level, an all-order description of the dominant radiative corrections for hard, wide-angle emissions.



7 TeV LHC parton kinematics

Figure 5.1: Phase space in x and the energy squared of the process  $Q^2$  that can be looked at at LHC and in the previous experiments, such as HERA [21].

### 5.2 Trigger and event selection

For this study, two sets of data at  $\sqrt{s} = 7$  TeV were analysed: data collected with a minimum bias trigger at instantaneous luminosity of  $\approx 10^{29}$  cm<sup>-2</sup>s<sup>-1</sup>, corresponding to 400  $\mu$ b<sup>-1</sup> of integrated luminosity, were used to measure the trigger efficiencies; data collected at instantaneous luminosity of  $\approx 10^{30}$  cm<sup>-2</sup>s<sup>-1</sup>, selected by a di-jet trigger with a raw calorimeter energy threshold of  $(E_{T,1} + E_{T,2})/2 = 15$  GeV integrated within  $|\eta| < 5.2$ , corresponding to 3.14 pb<sup>-1</sup> of integrated luminosity, were used for the cross section measurement. These low thresholds in the event selection were possible only in the very beginning of 2010 run and are strictly related to the instantaneous luminosity.

The selected events were required to have a good primary vertex consistent with the measured transverse position of the beam: to be reconstructed from at least 5 tracks and to lie within 24 cm in the longitudinal direction with respect to the nominal interaction point. This selection is highly efficient ( $\approx 100\%$ ) for this analysis and rejects non-collision background. The events were required to contain at least ten tracks, of which at least 25% should satisfy a high purity requirement [71].

Particles produced in the event have been clustered into jets with the anti-kT algorithm [35] with a jet size of R = 0.5. The algorithm was applied to calorimetric energy deposits, as done in the inclusive forward jet measurement [72].

The forward and central regions were defined respectively as  $3.2 < |\eta| < 4.7$  and  $|\eta| < 2.8$ . An event was accepted if there was at least one reconstructed jet [73] with axis within each one

of the pseudorapidity ranges with transverse momentum  $p_T > 35$  GeV. If more than one jet is present in the central or forward region, the one with the highest  $p_T$  is considered. The upper limit of the forward region takes into account the clustering jet size of 0.5, since the hadronic forward calorimeter (HF) extends up to  $|\eta| = 5.2$ .

The efficiency of the High-Level-Trigger (HLT) (DijetAve15, requiring two jets with sum  $(E_{T,1} + E_{T,2})/2 > 15$  GeV) is shown in Fig. 5.2, as a function of the forward and central jet  $p_T$  separately. The efficiency is determined as the ratio of HLT-triggered events over events passing the minimum bias trigger: the efficiency for the forward (central) jet is calculated requiring that the jet in the central (forward) region has  $p_T>35$  GeV, and it is almost 100%.



Figure 5.2: The High-Level trigger efficiency after the analysis selections, evaluated from a data sample of minimum bias events (red squares) and from the simulation (blue line). The efficiency is plotted separately as a function of the corrected  $p_T$  for the forward jets (left) and central jets (right).

### 5.3 Analysis

The distribution that has been measured is the differential cross section for the simultaneous production of at least one forward jet (f) and at least one central jet (c):

$$\frac{d^4\sigma}{dp_T^c dp_T^f d\eta^c d\eta^f} \tag{5.3}$$

integrated over  $\eta$ , and shown in the forward (f) and central (c) regions separately:

$$\left(\begin{array}{c}
\frac{d^{2}\sigma}{dp_{T}^{f}\Delta\eta^{f}} = \frac{1}{\Delta\eta^{f}} \cdot \frac{d^{4}\sigma}{dp_{T}^{c} dp_{T}^{f} d\eta^{c} d\eta^{f}} \bigg|_{p_{T}^{c} > 35 \text{GeV} \land |\eta^{c}| < 2.8} \\
\frac{d^{2}\sigma}{dp_{T}^{c}\Delta\eta^{c}} = \frac{1}{\Delta\eta^{c}} \cdot \frac{d^{4}\sigma}{dp_{T}^{c} dp_{T}^{f} d\eta^{c} d\eta^{f}} \bigg|_{p_{T}^{f} > 35 \text{GeV} \land 3.2 < |\eta^{f}| < 4.7}
\end{array}$$
(5.4)

The binning for the  $p_T$  spectra is chosen taking into account the jet energy resolution, to minimize

the effect of bin migration. For each bin, the average cross-section value, normalized to the bin width and to the width of the corresponding  $\eta$  region, is plotted at the true centre of the  $p_T$  distribution in the bin [74].

Fig. 5.3 shows the reconstructed  $p_T$  spectrum of forward (left) and central (right) jets after the selections, for data and simulation. The simulated events are normalized to the integrated luminosity. The simulated events went through the whole CMS detector simulation, thus allowing a comparison at detector level of the kinematics distributions.



Figure 5.3: The reconstructed  $p_T$  spectrum of forward (left) and central (right) jets, for the analyzed data (black dots) and the various MC models considered.

To obtain the cross section for the hadronic final state (i.e. fully devonvoluted from detector interaction and reconstruction effects) and being able to compare the results with different MC generators, a bin-by-bin correction is calculated on the simulated samples, which have been reweighted at hadron level to match the measured data distributions. This correction is evaluated by comparing the observables after the full simulation of the detector with those after hadronization. The ratio of the two is used as a multiplicative factor to the reconstructed  $p_T$  spectra:

$$f_{\text{HAD}}(v) = f_{\text{DET}}(v) \times \frac{f_{\text{HAD}}^{MC}(v)}{f_{\text{DET}}^{MC}(v)} , \qquad (5.5)$$

where v is the observable, and  $f_{\text{DET}}(v)$  and  $f_{\text{HAD}}(v)$  are the values determined after full simulation and after hadronization respectively.

The bin-by-bin ratios between the spectra for the hadronic final state and the ones after the full simulation are shown in Fig. 5.4. The black line is the average correction factor of different MC that went through the detector simulation process. The band is the envelope of all the correction factors using different MC and is used to evaluate the systematic uncertainty related to the choice of the model for the correction factor calculation. By using the bin-by-bin correction factors, Eq. 5.5 can be applied to determine the cross sections for the hadronic final state from the corresponding data distributions (Fig. 5.3).

Different methods to unfold the distribution at detector level to the distribution at hadron level have been tested, such as Singular Value Decomposition (SVD, [75]), bidimensional matrix



Figure 5.4: The bin-by-bin unsmearing correction as a function of  $p_T$  for forward (left) and central (right) jets. The black line is the average value of the factors computed with different MC samples. The band is the envelope of all the correction factors.

inversion unfolding and bayesian unfolding [76]. The bin-by-bin unfolding procedure has been adopted, given its simplicity and low systematic error. Alternative unfolding procedures have been used as cross-checks of the result.

### 5.4 Systematic uncertainties

In all  $p_T$  bins of the measured cross section, the statistical uncertainty (of the order of 1-2% in the low  $p_T$  bin and 5-10% in the highest one) is subdominant with respect to the systematic one, which amounts to  $\approx 30\%$ , dominated by the uncertainty on the jet enegy scale (JES).

The following effects have been taken into account as sources of systematic uncertainty:

- The jet energy scale calibration uncertainties have been evaluated as a function of the jet  $p_T$  and  $\eta$  with typical values between ~2.5% and ~3.5%. The uncertainty on the cross sections is estimated by coherently varying the energy of all the jets according to these values [77]. The JES uncertainty propagated to the steeply falling jet spectrum results in a systematic uncertainty of the order of  $\pm 25\%$  in the final cross section for the full  $p_T$  range.
- The uncertainty due to the jet energy resolution has been studied for the hadronic final state by applying artificial smearings of 9% and 11% to jets  $p_T$  reconstructed from the hadronic final state. This is the expected range of variation due to the uncertainty on the jet energy resolution. The effect is less that 5% over the whole  $p_T$  range, both in the forward and central region.
- The uncertainty due to the possible presence of pile-up is studied in a data-driven way: the  $p_T$  distributions have been produced with the requirement of one single primary vertex in the event and without this constraint. The forward and central  $p_T$  spectra calculated in these two conditions differ by less than 5%.
- The knowledge of the integrated proton-proton luminosity at the CMS interaction point

(dominated by the LHC beam currents during the Van der Meer scans) results in an overall 4% normalization uncertainty [78].

- The uncertainty on the bin-by-bin correction is estimated as the difference between correction factors calculated with different Monte Carlo samples (see Fig. 5.4), in each of the  $p_T$  bins.
- The uncertainty on the HLT efficiency is evaluated from the turn-on curves in the central and forward regions: since the plateau at 100% is reached for all the phase space of the analysis, the contribution to the systematic uncertainty is considered negligible.

Systematic uncertainties are shown as a function of the jet  $p_T$  in Fig. 5.5, for forward (left) and central (right) jets. The grey area shows the overall uncertainty, while the colors correspond to each single contribution.



Figure 5.5: Total systematic uncertainties as a function of the jet  $p_T$  for the forward (left) and central (right) jets. The grey area shows the overall uncertainty, while the colors correspond to each single contribution, independently of the others.

### 5.5 Results

The fully corrected cross section for simultaneous production of at least one central and at least one forward jet is presented in Fig. 5.6 as a function of the forward and central jet  $p_T$ . The uncertainty bands take into account both the statistical and the systematic uncertainties, summed in quadrature. Different MC generators have been compared to data. Table 5.1 tabulates the final  $p_T$ -differential forward and central jet cross sections.

To better evaluate the compatibility of the Monte Carlo predictions with the measured cross section, the ratio of various Monte Carlo simulations over the data are plotted on top of the band corresponding to the total uncertainty in Fig. 5.7. The ratios are shown as a function of the jet  $p_T$  for the hadronic final state. The HERWIG MC event generator, which uses angular ordering for the showering, describes the data best. The other MC event generators, with different tunes, do not describe the data well over the full  $p_T$  range.

The result of this work is published [79] and it is currently used to constrain theoretical models



Figure 5.6: The cross sections for the hadronic final state, as a function of  $p_T$ , for the forward jets on the left plots, and the central jets on the right plots compared to different sets of MCs.

to describe di-jet production [80, 81, 82]. Furthermore, this analysis can be considered as a first benchmark for the Higgs analysis described in the following chapters, since it involves di-jet production in the very same phase space (but with also leptons in the event).

$\Delta p_T \; [\text{GeV}]$	35 - 45	45 - 57	57 - 72
$\langle p_T \rangle   [\text{GeV}]$	39.3	50.0	62.9
$\frac{d^2\sigma}{dp_T^f\Delta\eta^f} \left[\frac{\rm pb}{\rm GeV}\right]$	$21.08 \pm 0.21 \pm {}^{+5.4}_{-0.45} \times 10^3$	$9.17 \pm 0.12 \pm {}^{+2.2}_{-1.79} \times 10^3$	$2.94 \pm \ 0.06 \ \pm \ ^{+0.7}_{-0.57} \ \times 10^3$
$\langle p_T \rangle   [\text{GeV}]$	39.6	50.2	63.2
$\frac{d^2\sigma}{dp_T^c\Delta\eta^c} \left[\frac{\rm pb}{\rm GeV}\right]$	$10.27 \pm 0.11 \pm {}^{+2.64}_{-2.15} \times 10^3$	$5.17 \pm 0.07 \pm {}^{+1.23}_{-0.99} \times 10^3$	$1.89 \pm 0.03 \pm {}^{+0.43}_{-0.35} \times 10^3$
$\Delta p_T \; [\text{GeV}]$	72 - 90	90 - 120	120 - 150
$\langle p_T \rangle$ [GeV]	78.7	100.0	130.3
$\frac{d^2\sigma}{dp_T^f\Delta\eta^f}  \left[\frac{\rm pb}{\rm GeV}\right]$	$695\pm\ 26\ \pm\ ^{+171}_{-142}$	111.9 $\pm$ 7.9 $\pm$ $^{+26}_{-24}$	$9.7\pm2.3\pm^{+2.9}_{-2.9}$
$\langle p_T \rangle   [\text{GeV}]$	79.0	101.1	132.7
$\frac{d^2\sigma}{dp_{x}^c\Delta n^c} \left[\frac{\mathrm{pb}}{\mathrm{GeV}}\right]$	$586 \pm 17 \pm ^{+133}_{-118}$	$133.9\pm6.2\pm^{+33.0}_{-24.8}$	$28.0\pm2.9\pm^{+6.6}_{-5.2}$

Table 5.1: Measured  $p_T$ -differential forward and central jet cross sections. The first (second) error is the statistical (systematic) uncertainty.



Figure 5.7: Ratio of the deconvoluted  $p_T$ -differential jet cross section from the various Monte Carlo over data. The yellow band corresponds to the total uncertainty for the forward region on the left plots, and the central one on the right plots.
# снартег б

# THE HWW ANALYSIS

When I start off to find somebody, I find them. That's why they pay me.

The bad, The good, the bad and the ugly

One of the open questions in the standard model (SM) of particle physics [83, 84, 85] is the origin of the masses of fundamental particles. Within the SM, vector boson masses arise from the spontaneous breaking of the electroweak symmetry by the Higgs field [86, 87, 88, 89, 90, 91]. The discovery or the exclusion of the SM Higgs boson is one of the central goals of the CERN Large Hadron Collider (LHC) physics program.

The search for the Higgs boson in the  $H \to W^+W^- \to 2\ell 2\nu$  final state, where  $\ell$  is an isolated charged lepton, electron or muon, and  $\nu$  a neutrino, is one of the main searches in a broad Higgs mass spectrum, given the high  $H \to W^+W^-$  branching ratio and the clean signature, as described in Sec. 1.3.2.

The search discussed in the following chapters is performed over the mass range 110–600 GeV, and the data sample used corresponds to  $4.9 \text{ fb}^{-1}$  of integrated luminosity collected in 2011 at a center of mass energy of 7 TeV and  $12.1 \text{ fb}^{-1}$  of integrated luminosity collected in 2012 at a center of mass energy of 8 TeV. Given the two different energies, the analysis strategy changed from 2011 to 2012. In the following, the final analysis with the full statistics available in October 2012 is described. I refer to [92] and [93] for details about analysis at the major milestones.

The  $H \rightarrow W^+W^-$  search is sensitive to different production mechanisms. In particular, at CMS, the analysis is divided into exclusive channels, defined by the number of jets with transverse momentum greater than 30 GeV reconstructed in the event. The so called zero-jet-bin analysis (with no jets with  $p_T > 30$  GeV) is sensitive to the gluon fusion production mechanism (ggH), depicted in Fig. 6.1. In the presence of one jet in the event (one-jet-bin) the main Higgs production mechanism is still gluon fusion, with an initial state radiation of a gluon, detected as a jet in the CMS sensitive volume. If two jets are reconstructed (two-jet-bin), it is possible to develop an analysis that is sensitive to different Higgs production mechanisms, such as vector boson fusion VBF (qqH), and associated production (VH).

In a VBF process, the Higgs is produced through the WWH and ZZH couplings, which give name to the mechanism. The two vector bosons are radiated by quarks, coming from the protons, that have enough transverse energy to deviate from their initial direction (beam line) and are detected



Figure 6.1: Feynman diagram for Higgs production via gluon fusion via a top-quark loop.

as two jets, called *tag jets*, usually in the forward pseudorapidity region. Even if the inclusive cross section for a production of a Higgs via VBF is roughly one tenth compared to the gluon initiated process, tight selections on the *tag jets*, such as a big separation in the  $\eta$  direction and a large invariant mass, can lead to a significant reduction of the backgrounds, as well as a low contamination from the ggH process, thus allowing for a pure VBF search.

Analyses sensitive to different production mechanisms allow to measure the couplings of the Higgs boson to different Standard Model particles and then to test the SM, or see deviations and hints of new physics.

Fig. 6.2 shows the Feynman diagram for the  $H \rightarrow W^+W^-$  production via vector boson fusion.



Figure 6.2: Feynman diagram for Higgs production via vector boson fusion.

A final state with two jets can be obtained also with the production of a Higgs boson in association with a vector boson (W or Z) as depicted in Fig. 6.3. Even if the cross section of an associated production of a Higgs boson is one order of magnitude smaller with respect to VBF, exploiting the high branching ratio of W and Z boson into quarks ( $\approx 70\%$ ) and looking for events with two jets with an invariant mass around the W/Z one, it is possible to reduce the background contamination and to probe the Higgs-Strahlung process.

In Fig. 6.4 and Fig. 6.5 the actual number of events produced in  $1fb^{-1}$  through different mechanisms, taking into account also the branching ratios of W<sup>+</sup>W<sup>-</sup>  $\rightarrow 2\ell 2\nu$ , according to the Standard Model predictions is shown, for a center of mass energy of 7 and 8 TeV.

In addition, the VBF and VH searches are also a benchmark to test beyond Standard Model



Figure 6.3: Feynman diagram for Higgs production via Higgs-Strahlung process.



Figure 6.4: Expected number of events for different Higgs production mechanism in the  $W^+W^- \rightarrow 2\ell 2\nu$  decay channel, for 7 (dashed line) and 8 TeV of center of mass energy.



Figure 6.5: Expected number of events for different Higgs production mechanism in the  $W^+W^- \rightarrow 2\ell 2\nu$  decay channel, for 7 (dashed line) and 8 TeV of center of mass energy, in the low Higgs mass range.

theories, such as fermiophobic Higgs scenarios.

In the following sections the main features of the  $H \rightarrow W^+W^- \rightarrow 2\ell 2\nu$  analysis are reported: the list of the main backgrounds (Sec. 6.1), the analysis strategies (Sec. 6.2), the main data driven background estimates (Sec. 6.4), and the systematics (Sec. 6.6), common to all jet bin categorizations. The limit extraction and discovery significance procedure is described in Sec. 6.7.

A detailed description of the VBF analysis (Sec. 7) and the VH analysis (Sec. 8) is presented as well as the final combination of all channels and the Higgs search results (Sec. 9).

# 6.1 Main backgrounds

All the processes with two isolated leptons and missing energy in the final state have to be considered as possible sources of background, as well as processes with jets reconstructed as leptons by the detector. The main background sources are due to non-resonant diboson production  $(W^+W^-, WZ, ZZ, W\gamma, Z\gamma)$ , Drell-Yan production (DY), top production  $(t\bar{t} \text{ and } tW), W + \text{jets}$  production, and QCD multijet processes in which two jets are misidentified as isolated leptons. These processes are summarized in Table 6.1 with their respective cross sections, multiplied by the branching ratio when meaningful, at 7 and 8 TeV.

Several Monte Carlo event generators are used to simulate the signal and background processes. The POWHEG program [94] provides event samples for the  $H \rightarrow W^+W^-$  signal (VBF and gluon fusion) and the tt and tW processes. The  $qq \rightarrow W^+W^-$ , Drell-Yan and W + jets processes are generated using the MADGRAPH 5.1.3 [95] event generator, the  $gg \rightarrow W^+W^-$  process using GG2WW [96], and the remaining processes using PYTHIA 6.424 [60]. For leading-order generators, the default set of parton distribution functions (PDF) used to produce these samples is

process	7 TeV $\sigma$ (pb)	8 TeV $\sigma$ (pb)
$\overline{\qquad WW \rightarrow \ell \nu \ell \nu}$	4.94	5.99
WZ	18.2	22.4
ZZ	7.67	9.03
$W\gamma  ightarrow \ell  u$	429	553
$Z\gamma  ightarrow \ell\ell$	96.6	132.6
$Z/\gamma^* \to \ell^+ \ell^- \ (m_{ll} > 50)$	3048	3533
${ m t} {ar t} { m +jets}$	163	225
${ m tW}$	15.7	22.4
$t \; ( ext{t-channel})$	64.57	85.53
$t \ (s-channel)$	4.63	5.65
W + jets	31314	37509

Table 6.1: Cross-section values for the backgrounds, multiplied by the branching ratio when meaningful.

CTEQ6L [97], while CT10 [98] is used for next-to-leading order (NLO) generators. Cross section calculations [12] at next-to-next-to-leading order (NNLO) are used for the  $H \rightarrow W^+W^-$  process, while NLO calculations are used for background cross sections. For all processes, the detector response is simulated using a detailed description of the CMS detector, based on the GEANT4 package [99].

### The $\mathrm{t}\bar{\mathrm{t}}$ and $\mathrm{t}$ production

The  $t\bar{t}$  pairs are produced at LHC via the gluon fusion process  $gg \to t\bar{t}$  or via QCD quark annihilation  $q\bar{q} \to t\bar{t}$ , as shown in Fig. 6.6, at tree level. Since the top decays into a W boson and a b-quark with a branching ratio of almost 100%, the  $t\bar{t}$  final state well reproduces signal topology since it consists of two opposite sign leptons, missing energy and two jets. Thus, the top production is one of the main backgrounds in the two-jet-bin analysis.

Single top production proceeds through three separate sub-processes at LHC [100], as shown in Fig. 6.7:

- t-channel: the dominant process involves the exchange of a space-like W boson.
- s-channel: involves the production of a time-like W boson, which then decays into a top and a bottom quark.
- tW-channel: top quark in association with a W boson is produced through a weak interaction between a gluon and a b quark from the proton sea.

In order to reduce top contamination a dedicated b-jet veto is applied, as described in Sec.6.2. The main top contamination comes from tW, given the two leptons that are produced in the hard scattering: one coming from the leptonic decay of the W, and the other from the decay of the top into Wb, and the consequent decay of the W into  $\ell\nu$ .



Figure 6.6: Tree level diagrams for  $t\bar{t}$  production via gluon fusion $gg \rightarrow t\bar{t}$  (a,b) and via QCD quark annihilation  $q\bar{q} \rightarrow t\bar{t}$  (c).



Figure 6.7: Feynman diagrams for single t production.

# The $Z/\gamma^* \rightarrow \ell^+ \ell^-$ + jets production

The production of a single Z boson in association with jets is described by the tree-level diagrams in Fig. 6.8. The Z boson is produced together with a quark or a gluon, which then hadronizes in one hard jet. Other jets are produced via parton splitting, and are mainly soft. The production of a Z boson via vector boson fusion mimics the VBF Higgs boson signature, but, given its low cross sections ( $\simeq 0.1$  pb), it is not one of the main backgrounds in the analysis [101, 102]. To suppress the  $Z/\gamma^* \rightarrow \ell^+\ell^-$  contribution, a tight cut on the missing transverse energy is applied: in a pure  $Z/\gamma^* \rightarrow \ell^+\ell^-$  event, missing energy comes mainly from mis-reconstruction of energies in the detector, thus leading to a fake lack of momentum balancing. In addition a veto of events with two electrons or two muons with an invariant mass close to Z one is applied.



Figure 6.8: Feynman diagrams for Z production. While the first two diagrams have a high cross section, but the jet kinematics are different with repsect to the one used in the analysis, the VBF Z boson production mimics the VBF Higgs boson production but it has a lower cross section ( $\simeq 0.1$  pb).

#### The W + jets production and QCD multijet processes

The production of a single W boson in association with jets is described by the tree-level diagrams in Fig. 6.9. Since a jet can be reconstructed as an electron, a W + jets event has a signature similar to the signal one. The lepton identification has been developed in order to minimize the jet mis-identification probability while keeping a good efficiency on real leptons.

An optimized lepton identification reduces also QCD multijet processes, where two jets are reconstructed as leptons. In addition, jets may contain real electrons and muons from leptonic b quark decays, but those leptons are vetoed by the isolation requirements.



Figure 6.9: Feynman diagrams for W production.

# The WW/WZ/ZZ production

The main diagrams at tree level for the production of two vector boson are shown in Fig. 6.10 and 6.11 (t and s, Triple Gauge Coupling, channels are depicted).

The WW sample has the same signature of signal (two leptons and two neutrinos in final state). With a leptonic decay of a Z associated with the decay of a Z into two neutrinos, the ZZ sample has a final state similar to signal one. Leptonic and hadronic decay of WZ can lead to two leptons of opposite charge and missing energy, in case of a neutrino from W, together with jets.



Figure 6.10: Feynman diagrams for WW production.



Figure 6.11: Feynman diagrams for ZZ and WZ production.

## The $W\gamma$ and $Z\gamma$ production

The associated production of a vector boson and a prompt photon is also a background of this analysis, when the conversion of a photon produces a pair of electrons. The main diagrams at tree level for these productions are shown in Fig. 6.12.

Veto on electrons coming from conversions based on tracks information are used to reject  $W\gamma$  and  $Z\gamma$  events, thus making these backgrounds negligible after all selections are applied.



Figure 6.12: Feynman diagrams for  $W\gamma$  on the left and  $Z\gamma$  production on the right.

# 6.2 Analysis strategy

As stated in the introduction, the search strategy for  $H \to W^+W^-$  is based on the final state in which both W bosons decay leptonically, resulting in a signature with two isolated, oppositely charged, high  $p_T$  leptons (electrons or muons) and large missing transverse momentum,  $E_T^{\text{miss}}$ , due to the undetected neutrinos. To improve the signal sensitivity, the events are separated according to the jet multiplicity into three mutually exclusive categories, which are characterized by different signal yields and signal-to-background ratios. Furthermore, the search strategy splits signal candidates into three final states denoted by  $e^+e^-$ ,  $\mu^+\mu^-$ , and  $e^\pm\mu^\mp$ , and the analysis is developed independently in the  $(e^+e^-)+(\mu^+\mu^-)$ , called same flavour final state (SF), and  $e^\pm\mu^\mp$ , called different flavour final state (DF). The bulk of the signal arises through direct W decays to electrons or muons of opposite charge, where the small contribution proceeding through an intermediate  $\tau$  lepton is implicitly included.

## 6.2.1 Trigger

A suite of signal and control triggers appropriate for this analysis were designed: double lepton and single lepton unprescaled triggers (see Chapter 2.2.5). The dilepton triggers have a high efficiency to collect Higgs boson events and are sufficiently loose to collect control events to estimate fake lepton backgrounds and selection efficiencies with adequate precision. The single lepton triggers, given the tight lepton identification requirements, are used to recover events were one lepton passes tight identification and kinematic thresholds, while the second is on the turn-on curve of the dilepton trigger. The list of triggers are summarized in Table 6.2.

Final state	trigger paths	
SingleElectron, e	HLT_Ele27_WP80	
SingleMu, $\mu$	HLT_IsoMu24_eta2p1	
DoubleElectron, ee	HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_	
	$\_Ele8\_CaloIdT\_CaloIsoVL\_TrkIdVL\_TrkIsoVL$	
<b>DoubleMu</b> , $\mu\mu$	HLT_Mu17_Mu8	
	HLT_Mu17_TkMu8	
MuEG, $e\mu$	HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL	
	HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL	

Table 6.2: Analysis triggers. The identification and isolation requirements are described in<br/>Table 6.3.

The main dielectron triggers require two HLT electron candidates with loose shower shape and calorimeter isolation requirements on both legs and a match to a Level-1 seed for the leading leg. Since the offline selections are  $E_T > 20, 10$  GeV for the leading and trailing electron respectively,  $E_T > 17, 8$  GeV is required at the HLT level. Having a total trigger rate compatible with the band-width is challenging in the dielectron channel, due to large fake electron background rates. Additional requirements must be added to the track-to-cluster matching and track isolation to control the total trigger rate (see Sec 4 for details about tracker-ECAL matching).

The identification and isolation requirements are described in Table 6.3. Because the electron HLT uses simplified algorithms compared to the offline selections, the variables used online and

name	criterion
CaloId_L	H/E < 0.15(0.10)
	$\sigma_{\eta\eta} < 0.014 \; (0.035)$
CaloId_T	H/E < 0.15(0.10)
	$\sigma_{\eta\eta} < 0.011 \; (0.031)$
CaloId_VT	H/E < 0.05(0.05)
	$\sigma_{\eta\eta} < 0.011 \ (0.031)$
TrkId_VL	$ \Delta\eta  < 0.01 \ (0.01)$
	$\Delta \phi < 0.15 \ (0.10)$
TrkId_T	$ \Delta\eta  < 0.008 \ (0.008)$
	$\Delta \phi < 0.07 \; (0.05)$
CaloIso_VL	$ECalIso/E_{T} < 0.2 (0.2)$
	$HCalIso/E_{T} < 0.2 (0.2)$
CaloIso_T	$ECalIso/E_T < 0.15 (0.075)$
	$\rm HCalIso/E_T < 0.15~(0.075)$
CaloIso_VT	$ECalIso/E_T < 0.05 (0.05)$
	$HCalIso/E_T < 0.05 (0.05)$
$TrkIso_VL$	$TrkIso/E_T < 0.2 (0.2)$
$TrkIso_T$	$TrkIso/E_T < 0.15 (0.075)$
$TrkIso_VT$	$TrkIso/E_T < 0.05 (0.05)$
WDOO	H/E < 0.10(0.05)
WP80	$\sigma_{\eta\eta} < 0.01 \ (0.03)$
	$ \Delta \eta  < 0.007 \ (0.007)$
	$ \Delta \phi  < 0.06 \ (0.03)$
	$\left \frac{1}{E} - \frac{1}{p}\right  < 0.05 \ (0.05)$
	$ECalIso/E_{T} < 0.15 (0.10)$
	$HCalIso/E_T < 0.10 (0.10)$
	$TrkIso/E_T < 0.05 (0.05)$

offline do not always correspond exactly. Nevertheless, the efficiencies of the offline requirements with respect to the online trigger selections are above 99%.

Table 6.3: Summary of requirements applied to electrons in the triggers used for this analysis. The selection requirements are given for electrons in the barrel (endcap). L=Loose, VL=Very loose, T=Tight, VT=Very Tight.

The main dimuon triggers require two HLT muon candidates with transverse momentum larger than 17/8 GeV/c and a match to a Level-1 seed is required for both legs. These are described in Table 6.2.

In the electron-muon channel two complementary triggers, which require both muon and electron HLT candidates, are used and summarised in Table 6.2.

Additional triggers are used to collect control or calibration events not covered by the main analysis triggers. Because the main dielectron analysis triggers put requirements on both legs, events collected with them cannot be used to measure efficiencies without introducing biases. Thus, to measure the electron and muon selection and trigger efficiencies specialised tag and probe triggers have been designed to maximise the number of  $Z/\gamma^* \rightarrow \ell^+ \ell^-$  events for both low and high  $p_{\rm T}$  leptons, while keeping the total trigger rate at a reasonable level, as summarized in

lepton flavor	trigger paths	
muon efficiency	HLT_IsoMu24_eta2p1	
	HLT_IsoMu30_eta2p1	
	HLT_Mu40_eta2p1	
	HLT_Mu50_eta2p1	
electron efficiency	HLT_Ele27_WP80	

Table 6.4. The tag and probe method is described in Sec. 6.5.

Table 6.4: Trigger paths used in data for studying trigger efficiencies.

Another set of specialised triggers is used to record events enriched in fake electrons and muons for the measurement of jet induced backgrounds. This is done using the fake rate method, which is described in detail in Sec. 6.4.1. Three triggers for the electron and two for muon fake rate measurements have been introduced, as described in Table 6.5. For electrons, since these triggers are prescaled, the first three impose different  $p_{\rm T}$  thresholds to collect a sufficient sample over a large  $p_{\rm T}$  range. For muons, two different  $p_{\rm T}$  thresholds are used to collect a sufficient sample over a large  $p_{\rm T}$  range since these triggers are prescaled.

lepton flavor	trigger paths
electron fakes	HLT_Ele8_CaloIdT_TrkIdVL
	HLT_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL
	HLT_Ele17_CaloIdL_CaloIsoVL_TrkIdVL_TrkIsoVL
muon fakes	HLT_Mu8
	HLT_Mu17

Table 6.5: Trigger paths used in data for studying fake rates.

The sum of the rates of the analysis triggers is about 20 Hz.

Given these triggers, the events are divided into different Primary Datasets, namely Single-Electron, SingleMu, DoubleElectron, DoubleMu and MuEG (Muon-ElectronGamma). Only the subset of runs and luminosity blocks which have passed all the quality tests of the Physics Validation Team are considered.

No trigger requirement is made on the simulated events, but scale factors to take into account differences between data and MC are measured as a function of kinematic variables of the leptons  $(p_{\rm T} \text{ and } \eta)$ .

# 6.2.2 Primary Vertex Reconstruction

For a given event, it is necessary to reconstruct the vertex associated to the hard scattering and to remove the ones coming from multiple interactions. Vertices are reconstructed using the Deterministic Annealing (DA) clustering of tracks [103]. Reconstructed vertices are required to have a z position within 24 cm of the nominal detector center and a radial position within 2 cm of the beamspot.0.7 From the set of vertices in the event passing these selection cuts, the vertex with the largest summed squared- $p_{\rm T}$  of the associated tracks is chosen as the event primary vertex. Reconstructed leptons will be required to have small impact parameters with respect to this vertex.



Figure 6.13: The distribution of the mean number of interactions per bunch crossing in 2011 (a) and 2012 (b).

The simulated samples are reweighted to represent the distribution of the number of protonproton interactions per bunch crossing (pile-up) as measured in the data. The average number of pile-up events per beam crossing in data is about 9 in 2011 and 21 in 2012, as shown in Fig. 6.13. The good agreement in the distribution of the number of reconstructed vertices, as shown in Fig. 6.14, guarantees the correct simulation of multiple interactions per bunch crossing.

#### 6.2.3 Muon Selection

The muons are required to be reconstructed both in the tracker and in the muon chambers  $^{1}$ . In addition, to reject muons from jets, a high track quality is required: more than 10 hits in the inner tracker and at least one pixel hit.

Cuts on the muon track impact parameters ensure that the muon does not originate from a pile-up vertex: in the transverse plane  $|d_0| < 0.02 \ (0.01)$  cm for muons with  $p_{\rm T}$  greater (smaller) than 20 GeV/c, and along the z coordinate  $|d_z| < 0.1$  cm, calculated with respect to the primary vertex.

Furthermore, the muons are required to have a pseudorapidity  $|\eta|$  smaller than 2.4 and to have a relative  $p_{\rm T}$  resolution better than 10%.

In order to reduce the contamination from the non-isolated muons originating from jets, an isolation algorithm is applied, built on the energy deposits measured in five concentric rings around the muon direction of size 0.1 in the  $\eta \times \phi$  plane. These values are combined by means of a MVA and corrected for the the average density of energy from pile-up particles.

A muon will be considered to be isolated when its MVA isolation value is greater than a given threshold, optimized separately for different  $p_{\rm T}$  ranges ( $p_{\rm T} \ge 20 \,{\rm GeV}$ ) and barrel (endcap).

<sup>&</sup>lt;sup>1</sup> GlobalMuon, see Sec. 3, with  $\chi^2/\text{ndof} < 10$  on the global fit, must have at least one good muon hit, and at least two matches to muon segments in different muon stations; or *TrackerMuon*, provided it satisfies the "Tracker Muon Last Station Tight" selection requiring at least two muon segments matched at  $3\sigma$  in local X and Y coordinates, with one being in the outermost muon station



Figure 6.14: The distribution of the number of reconstructed vertices in 2011 (a) and 2012 (b). A good agreement between data and MC is observed, thus assuring the correct description of pile-up events. The average number of vertexes changes between 2011 and 2012 due to the increase of instantaneous luminosity, moving from 7 to 15 events as shown in the picture.

A linear cut on the isolation variable has been used in the 2011 analysis, defined as the scalar sum of the  $p_{\rm T}$  of the particle flow candidates satisfying the following requirements:

- $\Delta R < 0.3$  to the muon in the  $\eta \times \phi$  plane,
- distance along z coordinate measured at the primary vertex between the charged PF candidate and the muon less than 0.1 cm,
- $p_{\rm T}>10$  GeV, if the PF candidate is classified as a neutral hadron or a photon.

#### 6.2.4 Electron Selection

In 2011 a cut based approach to select electrons was used, as described in detail in [104, 105]. In 2012 a more sophisticated electron identification is used: a multivariate variable, that exploits a large set of inputs, is trained against jets and non isolated electrons, that would pass electron selections if not properly rejected. The variables used in the MVA-based electron identification are:

- kinematics :  $p_{\rm T}, \eta$
- shower shape :  $\sigma_{i\eta i\eta}$  ,  $\sigma_{i\phi i\phi}$  ,  $\Delta\phi_{SC}$  ,  $\Delta\eta_{SC}$  ,  $E_{3\times3}$  / $E_{5\times5}$ ,  $E1\times5$  / $E5\times5$
- track fit quality  $(\chi^2)$
- number of tracker layers

- cluster-track matching (geometry) :  $\Delta \phi_{SC-Tk}$  and  $\Delta \eta_{SC-Tk}$
- cluster-track matching (energy-momentum) : E/p
- fraction of energy carried away by bremstraulung :  $f_{brem} = (p_{in} p_{out})/p_{in}$ , where  $p_{in}$  is the momentum reconstructed with the first layers of the tracker, while  $p_{out}$  is the one reconstructed with the last layers of the tracker,
- ratio of hadronic energy to electromagnetic energy : HCAL/ECAL
- impact parameter: transverse and 3D impact parameters with respect to the primary vertex
- Preshower contribution:  $E_{ES} / E_{SC}$

Isolation requirements are then imposed by computing the particle flow isolation, defined as the scalar sum of the  $p_{\rm T}$  of the particle flow candidates satisfying the following requirements:

- $\Delta R < 0.4$  to the electron in the  $\eta \times \phi$  plane,
- PF electrons and muons are vetoed,
- for gamma PF candidates, require that they are outside the footprint veto regoin of  $\Delta R < 0.08$ ,
- for charged hadron PF candidates, require that they are outside the footprint veto region of  $\Delta R < 0.015$  ,
- for charged hadron PF candiates, require that they are associated with the primary vertex,

Neutral components are corrected by subtracting pileup contribution which is calculated by  $\rho \times A_{eff}$ , where  $\rho$  (kt6PFJets) is the event-by-event energy density and  $A_{eff}$  is the effective area. The isolation variable is defined as:

$$\frac{Iso_{PF}}{p_{T}} = (Iso_{charged\ hadron} + Iso_{gamma} + Iso_{neutral\ hadron} - \rho \times A_{eff})\frac{1}{p_{T}}$$
(6.1)

where  $Iso_{charged\ hadron}$ ,  $Iso_{gamma}$ , and  $Iso_{neutral\ hadron}$  are the scalar sum of the  $p_{\rm T}$  of charged hadron, gamma and neutral hadron PF candidates, respectivlely, in the isolation cone of 0.4 around the electron. The value  $\frac{Iso_{PF}}{p_{\rm T}}$  is required to be less than 0.15.

In order to reject events where an electron originates from a conversion of a photon into a  $e^+e^-$  pair in the tracker material, the number of missed inner tracker layers of the electron track is required to be exactly zero. In addition any event in which the selected electron is close in space to a track, and the pair electron-track is compatible with a photon conversion, is rejected:  $|\Delta \cot \theta| < 0.02$  and dist < 0.02, being these quantities the distance of the two tracks in the longitudinal and transverse plane respectively [106].

Finally to reduce fake electrons from non-prompt sources, the transverse and longitudinal impact parameters with respect to the primary vertex are required to be less than 0.02 and 0.1 cm respectively.

### 6.2.5 Missing Energy

The missing transverse energy is used to reject background events where there is no natural source of missing energy, like in Drell-Yan and QCD events, while in a  $H \to WW \to \ell \nu \ell \nu$  event large missing energy is expected due to neutrinos.

However there are events that may mimic a  $H \to WW \to \ell \nu \ell \nu$  event, but whose kinematics are different, such as  $Z/\gamma^* \to \tau^+ \tau^-$ . In the  $Z/\gamma^* \to \tau^+ \tau^-$  process, given the large difference in the masses of  $\tau$  and Z, the taus are produced with large boost and their decay products, including neutrinos, are aligned with the leptons, as depicted in Fig. 6.15.



Figure 6.15:  $Z \rightarrow \tau \tau$  decay.

Therefore a transverse component of the missing energy with respect to the leptons direction is a measure of missing energy in the event, not originating from  $\tau$  decay. To reject such background events with a small opening angle between  $E_{\rm T}^{\rm miss}$  and one of the leptons, the projected  $E_{\rm T}^{\rm miss}$  for event selection is used, defined as:

$$proj - E_{\rm T}^{\rm miss} = \begin{cases} E_{\rm T}^{\rm miss} & \text{if } \Delta \phi_{min} > \frac{\pi}{2}, \\ E_{\rm T}^{\rm miss} \sin(\Delta \phi_{min}) & \text{if } \Delta \phi_{min} < \frac{\pi}{2} \end{cases}$$
(6.2)

with 
$$\Delta \phi_{min} = min(\Delta \phi(\ell_1, E_{\mathrm{T}}^{\mathrm{miss}}), \Delta \phi(\ell_2, E_{\mathrm{T}}^{\mathrm{miss}}))$$
 (6.3)

where  $\Delta \phi(\ell_i, E_T^{\text{miss}})$  is the angle between  $E_T^{\text{miss}}$  and lepton *i* in the transverse plane, as shown in Fig. 6.16.

Furthermore, in the presence of high multiple-interactions (pile-up), the instrumental  $E_{\rm T}^{\rm miss}$  tail in  $Z/\gamma^* \rightarrow \ell^+ \ell^-$  events increases significantly, with  $\ell = e/\mu$ . To improve the signal over background performance of  $E_{\rm T}^{\rm miss}$  selections in the presence of pile-up, the tracker  $E_{\rm T}^{\rm miss}$  is used, reconstructed using only charged particles originating from the primary vertex. The trk-MET is defined as

trk-MET 
$$\equiv -\overrightarrow{p_T}(l_1) - \overrightarrow{p_T}(l_2) - \sum_i \overrightarrow{p_T}(i),$$
 (6.4)

where  $\overrightarrow{p_T}(l_1)$  and  $\overrightarrow{p_T}(l_2)$  are the transverse momentum vectors of the two leptons passing the lepton selections described in Sec. 6.2.3 and Sec. 6.4.1, and  $\overrightarrow{p_T}(i)$  represent the transverse momentum vectors of the charged PFC and takes satisfying the following requirements:



Figure 6.16: Met projection procedure. In figure (a) the projection procedure for a background event (such as  $Z/\gamma^* \to \tau^+\tau^-$ ). The minimum between the two  $E_{\rm T}^{\rm miss}$ lepton angles is considered (the "electron" in this picture) and the met is projected in the transverse plane. This procendre reduces  $E_{\rm T}^{\rm miss}$  for  $Z/\gamma^* \to \tau^+\tau^$ events, as depicted in Fig. 6.15. In figure (b) the projection procedure for signal events, where  $E_{\rm T}^{\rm miss}$  and leptons are expected to be in opposite directions, then the projection leaves the  $E_{\rm T}^{\rm miss}$  as it is, because  $\Delta\phi_{min} > \frac{\pi}{2}$ .

- the track matched to PF Candidate has  $\Delta z < 0.1$  cm with respect to the signal primary vertex;
- the track has  $\Delta R > 0.1$  with respect to both leptons, to avoid double-counting of the leptons.

Compared to the projected PFMet, the projected trk-MET has a larger tail in  $Z/\gamma^* \to \ell^+ \ell^$ background events. However these two  $E_{\rm T}^{\rm miss}$  values are weakly-correlated in  $Z/\gamma^* \to \ell^+ \ell^$ backgrounds with no geninue  $E_{\rm T}^{\rm miss}$ , and strongly correlated for the signal processes with geninue  $E_{\rm T}^{\rm miss}$ . Therefore the signal over background ratio is improved by selecting the events based on the mininum of these two projected  $E_{\rm T}^{\rm miss}$  values:

$$min - proj E_{\rm T}^{\rm miss} \equiv min({\rm proj}_{\rm trk-MET}, {\rm proj}_{\rm PFMET}).$$
 (6.5)

Events are selected if a value of  $min - projE_{T}^{miss}$  greater than 20 GeV is observed.

### 6.2.6 Z Veto

To further reduce the Drell–Yan background in the  $e^+e^-$  and  $\mu^+\mu^-$  final states, events with a dilepton invariant mass within 15 GeV of the Z are vetoed. Events with a dilepton invariant mass below 12 GeV/c<sup>2</sup> are rejected to suppress contributions from low mass resonances, such as  $J/\psi$  (3 GeV), Y(1S) (9.5 GeV), Y(2S) (10.0 GeV), Y(3S) (10.4 GeV), as shown in Fig 6.17.



Figure 6.17: Low mass resonances removed by requiring the invariant mass of the dilepton system to be greater than 12 GeV.

## 6.2.7 Jet Counting

Jets are reconstructed using calorimeter and tracker information using a particle flow algorithm [107]. The anti-k<sub>T</sub> clustering algorithm [35] with R = 0.5 is used (see Sec. 3.3). To exclude electrons and muons from the sample, jets are required to be separated from the selected leptons in  $\Delta R$  by at least  $\Delta R^{\text{jet-lepton}} > 0.3$ .

In this analysis high  $p_T$  jets are used to define the analysis jet bin categories (0/1/2) and low  $p_T$  jets to do the top events veto:

- counted jet: a reconstructed jets with  $p_{\rm T} > 30$  GeV within  $|\eta| < 4.7$ ;
- low  $p_T$  jet: a reconstructed jets with 10  $< p_T < 30$  GeV within  $|\eta| < 4.7$

### 6.2.8 Top Tagging

Top backgrounds pose a significant challenge, since the production cross-section is substantially higher than the signal cross-section. To reduce it, two top tagging methods were introduced, relying on the fact that top quarks decay to Wb with almost certainty.

The first method vetoes events containing soft muons from the *b*-quark decays. The requirements used to select soft muons are:

- $p_{\rm T} > 3 \,\,{\rm GeV};$
- reconstructed as a TrackerMuon;
- meet TMLastStationAngTight muon id requirements;
- number of valid inner tracker hits > 10;
- transverse impact parameter with respect to the Primary Vertex,  $|d_0| < 0.2$  cm;
- longitudinal impact parameter with respect to the Primary Vertex  $|d_z| < 0.1$  cm;

• non-isolated (Iso<sub>Total</sub>/ $p_{\rm T} > 0.1$ ) if  $p_{\rm T} > 20$  GeV.

The second method uses standard *b*-jet tag, that looks at the tracks composing a jet. All the tracks associated to a jet are taken into account and the impact parameter (IP), defined as the distance between the track and the vertex at the point of closest approach, as shown in Fig. 6.18, is measured and used to estimate the significance of that track, defined as IP /  $\sigma_{IP}$ , where  $\sigma_{IP}$  is the precision of the measurement of IP. Then, the track counting algorithm identifies a jet as originating from a b quark if it contains at least N tracks each with a significance of the impact parameter exceeding a threshold. The discriminator associated with N = 2 is called track counting high efficiency (TCHE)[108, 109].



Figure 6.18: Illustration of the sign of the impact parameter of a track: the sign is positive (negative) if the angle  $\theta$  between the impact parameter direction and the jet axis is smaller (larger) than 90°.

Events containing low  $p_{\rm T}$  jets tagged with the TCHE algorithm with a discriminator value greater than 2.1 are vetoed. In addition, in the VBF (VH) analysis the two high  $p_{\rm T}$  selected jets are required to have the TCHE value smaller than 2.1 (1.6).

#### 6.2.9 Other Preselection Requirements

To reduce the background from diboson processes, events containing an additional lepton meeting the previously described selection requirements with  $p_{\rm T} > 10 \text{ GeV}/c$  are vetoed. This removes about 60% of the WZ component but only 10% on the ZZ component, which is dominated by ZZ  $\rightarrow 2l2\nu$  decays after the full event selection and surviving the Z veto. The efficiency for  $WW \rightarrow 2l2\nu$  events is 99.9%. Finally, the angle in the transverse plane between the dilepton system and the di-jet system must be smaller than 165 degrees in the  $ee/\mu\mu$  final states. This requirement rejects  $Z/\gamma^* \rightarrow \ell^+\ell^-$  events, where the Z boson recoils against two jets.

# 6.3 Higgs Signal Extraction Strategy

To enhance the sensitivity to the Higgs boson signal a set of selections has been optimized. In order to see an excess of events, and then the signal, a precise knowledge of background contamination is needed. In the next chapters a description of the main background data-driven procedure is given. In the analysis, a common set of selections are defined historically as WW  $level^2$ , where a sanity check of the main variables distribution is performed. The WW level is used in the zero-jet-bin analysis to measure the W<sup>+</sup>W<sup>-</sup> cross section in CMS [110, 111]. and to perform measure of electron to muon efficiencies used in the estimation of the various backgrounds.

The WW level is defined by the following  $cuts^3$ :

- Lepton preselection:
  - at least two opposite-sign leptons are reconstructed in the event; with a  $|\eta| < 2.5$  for electrons and  $|\eta| < 2.4$  for muons;
  - $-p_{\rm T} > 20$  GeV for the leading lepton. For the trailing lepton, the transverse momentum is required to be larger than 10 GeV.
- Lepton selection: both leptons have to pass the identification and isolation requirements.
- Extra lepton veto: the event is required to have two and only two opposite-sign leptons passing the lepton selection (no extra lepton with  $p_{\rm T} > 10$  GeV).
- $E_{\rm T}^{\rm miss}$  preselection: PF  $E_{\rm T}^{\rm miss} > 20$  GeV for  $e\mu$  and  $\mu e$  events and PF  $E_{\rm T}^{\rm miss} > 45$  GeV for ee and  $\mu\mu$  events.
- Low mass resonances rejection:  $m_{\ell\ell} > 12$  GeV.
- Z-peak veto:  $|m_{\ell\ell} m_Z| > 15$  GeV for ee and  $\mu\mu$  events.
- Projected  $E_{\rm T}^{\rm miss}$  selection: the  $min proj E_{\rm T}^{\rm miss}$  variable described in Sec. 6.2.5 is required to be larger than 20 GeV.
- Soft muon veto: the event is required to not have soft muons as defined in Sec. 6.2.8.
- Anti b-tagging: the event is required to not have any soft jet passing the b-tagging selection described in Sec. 6.2.8.
- Kinematical cut: the transverse momentum of the di-lepton system is required to be greater than 45 GeV ( $p_T^{\ell\ell} > 45$  GeV).
- Jet requirement: the event is required to have at least two jets with  $p_{\rm T} > 30$  GeV and  $|\eta| < 4.7$ . The jets are also required to fail a tight b-tagging selection, TCHE less than 2.1 (1.6) for VBF (VH) analysis.

As shown in Fig. 6.19 and Fig. 6.20 the kinematic distributions of jets and leptons in the event at WW level are well reproduced, thus assuring the goodness of MC simulation. These variables are used in the analyses to discriminate signal from background. A test of the good description of the variables by MC samples is needed, since, while for some backgrounds it is possible to estimate their contributions after selections using data-driven estimations, see Sec. 6.4, for others and for the signal the analyses rely on MC predictions. In Fig. 6.19 and Fig. 6.20 the MC predictions from different processes are stacked. The MC shaded area represents the error due to MC statistics and normalization systematic uncertainties. Shape uncertainties, such as the one due to lepton energy scale, are not shown. The MC components are scaled to data-driven

<sup>&</sup>lt;sup>2</sup>The WW level is defined as common as possible between different jet categorisations. In the zero-jet-bin analysis it is highly dominated by the W<sup>+</sup>W<sup>-</sup> contribution, while in the two-jet-bin one, despite the name, the main backgrounds are Top and W<sup>+</sup>W<sup>-</sup>.

<sup>&</sup>lt;sup>3</sup>For sake of simplicity the WW level defined in the 2012 analysis is reported. In 2011 a slightly different definition is used, but with a similar signal and background efficiency. The details are exposed in [104, 105, 112, 113].

estimations, when available, and the error on the scale factor is propagated to the plot. Even if signal is not visible in these plots, due to relaxed selections, it is added to the stacked MC plots. The black dots are the number of data events in each bin and their error is statistical. The ratio of data over MC predictions is shown: the grey band is the error on the simulation while the statistical error is shown on the data points.

Fig. 6.19 shows jets kinematic variables: transverse momentum of the leading and the trailing jet, the invariant mass of the di-jet system and the pseudorapidity difference between the two jets.

Fig. 6.20 shows leptons kinematic variables: the invariant mass and the transverse momentum of the di-lepton system, the azimuthal opening between the two leptons. In all the distributions the last bin is the overflow one.



Figure 6.19: Jets distributions at WW level. Good agreement between data and simulation is observed.



(c)  $\Delta \phi_{\ell\ell}$ 

Figure 6.20: Lepton distributions at WW level. Good agreement between data and simulation is observed.

# 6.4 Background Estimation

A combination of data-driven methods and detailed Monte Carlo simulation are used to estimate background contributions to the signal region. From data it is possible to estimate the following backgrounds: W + jets,  $Z/\gamma^* \rightarrow \ell^+ \ell^-$ ,  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  and top. The remaining processes are taken from simulation.

Background composition and yields depend on the final state and on the Higgs boson mass hypothesis under study. In the **zero-jet-bin** final state, the non-resonant  $W^+W^-$  background dominates, while W+jets background contribution becomes sizable in the low Higgs mass cases. In the 1-jet and 2-jet final states, the largest contribution comes from top decays, while the non-resonant  $W^+W^-$  background contribution is the second largest one.

All the background estimations can be categorized into two main approaches: AB and ABCD methods. The AB method defines two regions:

- A; the signal region, that is defined by the nominal selections applied in the analysis;
- B; a background dominated region (signal free), where the most important selections used to remove the specific background are reversed.

The ratio for a given background between the contamination in region A and B,  $R^{A/B} = N^A/N^B$  is taken from MC while a normalization of the background under study is extracted from data, measuring the number of events in data  $N^B$ . The number of background event in region A is then estimated by means of Eq. 6.6

$$N^{A} = R_{MC}^{A/B} \ N_{DATA}^{B} = \frac{N_{MC}^{A}}{N_{MC}^{B}} \ N_{DATA}^{B}.$$
(6.6)

A sketch of the AB method is depicted in Fig. 6.21.



Figure 6.21: The AB method for background estimation. The ratio between background events in region A and B is taken from MC  $(R^{A/B} = N^A/N^B)$  while the expected number of events in region A is extracted from data measuring the number of background events in region B and scaling by  $R^{A/B}$ .

The main hypothesis of this method is that the ratio  $R^{A/B}$  is well modelled in the MC. In addition, the region  $N_{DATA}^B$  must be dominated by the selected background under investigation: other components have to be subtracted and the error due to their subtraction is propagated to the final estimation of  $N^A$ .

The second method, ABCD, defines four regions:

- A; the signal region, that is defined by the nominal selections applied;
- *B*; a background dominated region (signal free), where the most important selections used to remove the specific background are reversed;
- C; a background dominated region (signal free), with the same selections of region A except one cut;
- D; a background dominated region (signal free), with the same selections of region B except one cut, the same of region C.

Under the hypothesis of the independence between the selection that defines the regions AB with respect to CD, to be tested on a MC sample, the number of background event in region A can be estimated by means of Eq. 6.7.

$$N^A = \frac{N^C_{DATA}}{N^D_{DATA}} N^B_{DATA} . aga{6.7}$$



Figure 6.22: The *ABCD* method for background estimation. The ratio between background events in region *A* and *B* is taken from data measuring the number of background events in two "parallel" regions (*C* and *D*). The expected number of events in region *A* is extracted from data measuring the number of background events in region *B* and scaling by  $\frac{N^{C}}{N^{D}}$ . The assumption of this method (to be checked on MC or on a different phase space), is that the ratio  $\frac{N^{C}}{N^{D}}$  for the background is the same of  $\frac{N^{A}}{N^{B}}$ .

The hypothesis of this method is that the ratio  $\frac{N^C}{N^D}$  is the same as  $\frac{N^A}{N^B}$ : the main issue is the estimation of the degree of belief of this assumption and then the measurement of the systematics related to that. In addition, the region B, C and D must be dominated by the selected background

under investigation: other components are subtracted and the error due to their subtraction is propagated to the final estimation of  $N^A$ . In Fig. 6.22 a schematic view of the *ABCD* method is shown.

In the following the estimation of the these backgrounds are described:

- jet induced backgrounds (W + jets and QCD), [ABCD]
- top, [ABCD]
- $Z/\gamma^* \to \ell^+ \ell^-$ , [AB and ABCD]
- $Z/\gamma^* \to \tau^+ \tau^-$  (ad hoc approach).

### 6.4.1 Jet Induced Backgrounds: W + jets and QCD

Jet induced fake leptons are an important source of background for many physics channels. In this analysis the main sources of fake leptons are W + jets and QCD events, where at least one of the jets or one of its constituent is misidentified as an isolated lepton. The dominant background is W + jets because there is already one prompt, well isolated, lepton from the W boson decay in the event. Fake non-prompt leptons arise from the leptonic decay of heavy quarks, misidentified hadrons or electrons from photon conversions.

A data-driven approach is pursued to estimate this background. A set of loosely selected leptonlike objects, referred to as the "fakeable object" or "denominator" from here on, is defined in a sample of events dominated by di-jet production. The efficiency for these denominator objects to pass the full lepton selection critera is measured. This background efficiency, typically referred to as the "fake rate" ( $\epsilon_{\text{fake}}$ ), is parameterized as a function of the  $p_{\text{T}}$  and  $\eta$  of the denominator object in order to capture any dependence on kinematic and geometric quantities. These fake rates are, then, used as weights to extrapolate the background yield from a sample of denominator objects to the sample of fully selected leptons.

#### Fakeable object

The higher instantaneous luminosity delivered by the LHC leads to tighter selection requirements in the high level trigger for electrons, thus limiting the choice of possible denominator object definitions. The denominator object definition used for electrons is:

- $\sigma_{i\eta i\eta} < 0.01/0.03$  (barrel/endcap)
- $|\Delta \phi_{in}| < 0.15/0.10$
- $|\Delta \eta_{in}| < 0.007/0.009$
- HCAL/ECAL < 0.12/0.10
- full conversion rejection (see Sec.)
- $|d_0| < 0.02 \text{ cm}$

• 
$$\frac{\sum_{\text{trk}} E_{\text{T}}}{p_{\text{T}}^{\text{ele}}} < 0.2$$

•  $\frac{\sum_{\text{ECAL}} E_{\text{T}}}{p_{\text{T}}^{\text{ele}}} < 0.2$ 

•  $\frac{\sum_{\mathrm{HCAL}} E_{\mathrm{T}}}{p_{\mathrm{T}}^{\mathrm{ele}}} < 0.2$ 

For muons, the selection requirements differ from the tight selection of Sec. 6.2.3 only in less stringent cuts on  $d_0$  and MVA based isolation:

- $|d_0| < 0.2 \text{ cm}$
- MVAoutput > -0.6

#### Fake rate measurement

The lepton fake rates are measured in a sample dominated by QCD di-jets events, which still may contain real leptons from W or Z leptonic decays. The muons from W decays are removed by requiring the event to have PF  $E_{\rm T}^{\rm miss} < 20$  GeV. The W transverse mass has to be lower than 20 GeV as well. The muons from resonances are removed with the  $m_{\mu\mu} > 20$  GeV and the  $m_{\mu\mu} \notin [76, 106]$  GeV constraints. For electrons the W transverse mass cut is not applied, and the Z-peak veto is enlarged to  $m_{\rm ee} \notin [60, 120]$  GeV. Finally, both muon and electron candidates are required to be well separated from the leading jet of the event,  $\Delta\phi(\ell, j) > 1$ .

From these selected event samples, the fake rate ( $\epsilon_{\text{fake}}$ ) is measured by counting the number of denominator objects which pass the full lepton selection, in bins of  $p_T$  and  $\eta$ , as defined in Eq. 6.8

$$\epsilon_{\text{fake}}\left(p_{\text{T}},\eta\right) = \frac{N_{pass}}{N_{pass+fail}}\tag{6.8}$$

The  $\eta$  ranges considered are [0,1], (1,1.479], (1.479,2] and (2,2.5], while the  $p_{\rm T}$  ranges are (10,15], (15,20], (20,25], (25,30] and (30,35].

#### Application of Fake rates

Once the fake rates are measured, parameterized in the kinematic quantities of interest, they are used as weights in order to extrapolate the yield of the sample of loose leptons to the sample of fully selected leptons. This is done by selecting events passing the full event selection, with the exception that one of the two lepton candidates is required to pass the denominator selection cuts but it fails the full lepton selection ones. This lepton is from here on denoted the "failing leg". The other lepton is required to pass the full selection. The data sample selected in this way is denoted the "tight + fail" sample. Each of the events passing this selection is given a weight computed from the fake rate in the particular  $p_T$  and  $\eta$  bin of the failing leg, as follows:

$$w_i = \frac{\epsilon_{\text{fake}}(p_{\text{Ti}}, \eta_i)}{1 - \epsilon_{\text{fake}}(p_{\text{Ti}}, \eta_i)}$$
(6.9)

where *i* is an index denoting the failing leg, and  $p_{\text{Ti}}$  and  $\eta_i$  are its transverse momentum and pseudorapidity. Summing the weights  $w_i$  over all such events in the tight + fail sample yields the total jet induced background prediction.

This tight + fail extrapolation prediction double counts the QCD component of the background, where both leptons are jet induced fakes. This is essentially a combinatorial artifact, due to

the fact that in the tight + fail selection, one is unable to uniquely distinguish which lepton is required to be the tight one and which lepton is required to be the failing one. This double fake background is typically very small and accounts for roughly a few percent of the total jet induced background. In order to estimate the amount of double counting, the fake rate extrapolation is performed on both lepton legs, selecting events which pass all event selection criteria, except that both leptons are required to pass the denominator selection, but fail the full lepton selection. This event sample is denoted as the "fail + fail" sample. Events in the fail + fail sample are then given weights as follows:

$$w_{i,j} = \frac{\epsilon_{\text{fake}}(p_{\text{Ti}}, \eta_i)}{1 - \epsilon_{\text{fake}}(p_{\text{Ti}}, \eta_i)} \times \frac{\epsilon_{\text{fake}}(p_{\text{Tj}}, \eta_j)}{1 - \epsilon_{\text{fake}}(p_{\text{Tj}}, \eta_j)}$$
(6.10)

where *i* and *j* denote the two failing leg, and  $p_{\text{Ti/j}}$  and  $\eta_{i/j}$  are the transverse momentum and pseudorapidity of the first and second leg. Summing the weights  $w_{i,j}$  over all such events in the fail + fail sample yields the total QCD double fake background. This prediction is then subtracted from the tight + loose prediction in order to account for the double counting.

A more robust approach is also used as a cross check, that takes into account contamination from prompt leptons in the fake rate application, as described in [114].

#### **Closure Test and Systematic Uncertainties**

The fake rate method for estimating fake lepton backgrounds crucially relies on the assumption that fake rates can be transferred from jets in QCD events to jets in W+jets events. The degree to which this assumption is incorrect must be reflected in the systematic uncertainties of the fake lepton background prediction. In order to test the validity of the assumption and to extract quantitative measure of the systematic uncertainties, a closure test is performed on the W+jets Monte Carlo simulation sample by comparing the background yield predicted by the Monte Carlo simulation with the yield predicted using the fake rate procedure applied on it. To be consistent, the QCD Monte Carlo simulation is used to measure the fake rates, that are then applied to the tight + fail sample selected in the W+jets Monte Carlo sample. The degree of disagreement yields a quantitative measure of the systematic uncertainty of the method, that is found to be about 36%. This value is used as a normalization systematic error for the jet induced background.

Further closure test on the fake lepton background estimate is performed using data events with two same sign leptons. This control sample is highly enriched in W+Jet background and can serve as an additional cross-check of the systematic uncertainties estimated above from Monte Carlo simulation. The result of this cross-check is perfectly consistent with the uncertainties estimated in the previous section, demonstrating that the extrapolation systematics estimated from the Monte Carlo simulation is applicable to data.

#### 6.4.2 Top Background

The general strategy for determining the residual top events in the signal region relies on the measure of the top tagging efficiency in an orthogonal region of phase space in data. Then, using this efficiency, the contamination in the signal region is extrapolated from a control region defined inverting the main top rejection cut, namely the b-tag cut (TCHE). The number of expected top events in the signal region is therefore:

$$N_{bveto} = N_{btag} \cdot \frac{1 - \varepsilon_{\rm top}}{\varepsilon_{\rm top}} \tag{6.11}$$

where  $N_{btag}$  is the number of events in the control region and  $\epsilon_{top}$  is the b-tagging efficiency as measured in data. The Eq. 6.11 is applied in  $\eta$  bins of the most central jet.

The  $\eta$ -dependent b-tag efficiency is measured for the most central jet (cj) in a top-enriched control region, at the WW level (without btag veto), where the most forward jet (fj) has been b-vetoed and with the additional request of different lepton flavors in the final state to suppress the Drell-Yan contamination. The efficiency as a function of  $|\eta^{cj}|$  is calculated as the ratio between the distribution of the central jets that survive the b-tagging cut and the distribution of the central jets without that requirement, as shown in Equation 6.12.

$$\epsilon(|\eta^{cj}|) = \frac{N_{btag}^{cj,\text{control}}}{N^{cj,\text{control}}}$$
(6.12)

Fig. 6.23 shows the  $\eta$  distribution for the most central jet in the b-tagged region at WW level on the left, and the inclusive  $\eta$  distribution on the right. The efficiency is the ratio between top events on the left and top events on the right.



Figure 6.23: The  $\eta$  distribution for the most central jet in the b-tagged region at WW level on the left, and the inclusive (without b-tag selections)  $\eta$  distribution on the right. The efficiency is the ratio between top events on the left and top events on the right.

Following Equation 6.11, the number of top events in the b-veto region are given by:

$$N_{bveto}^{\text{data driven}} = \int d\eta \ N_{bveto}^{\text{data driven}}\left(\eta\right) = \int d\eta \ \left(N_{btag}^{DATA}\left(\eta\right) \frac{1-\varepsilon\left(\eta\right)}{\varepsilon\left(\eta\right)}\right) \tag{6.13}$$

The contribution from other backgrounds is subtracted to the number of events in the btagged region,  $N_{btag}^{DATA}$ , when applying Equation 6.11. Given the high purity in btagged region (about 90%) the effect of this subtraction is small (about 10%).

The method relies on the fact that the  $\eta$  efficiencies are the same in the phase space where they have been measured and after all the analysis selections are applied.

The systematic error on top estimation is due to three effects:

- the measurement of the b-tag efficiency (15%),
- the variation of the b-tag efficiency at WW level and after all selections are applied (10%),
- the subtraction of non-top contribution in the btagged region (10%).

The numbers reported are just representative of the order of magnitude of the systematic effect. The exact numbers will be reported in Sec. 7 for the VBF case and in Sec. 8 for the VH one for each Higgs mass hypothesis analysis.

While the error coming from the measurement of the b-tag efficiency decreases with more statistics available, the other factors will remain the same, unless a phase space closer to the signal one is used to measure  $\varepsilon$ .

# 6.4.3 Drell-Yan Background $Z/\gamma^* \rightarrow \ell^+ \ell^-$

The expected contributions from  $Z/\gamma^* \to \ell^+ \ell^-$  events outside the Z-mass region in data ("out" region) can be estimated by counting the number of events in the Z mass region in data ("in" region), subtracting from it the non-Z contributions, and scaling it by a ratio  $R_{out/in}$  defined as the fraction of events outside and inside the Z-mass region in the simulation. The non-Z contributions in the Z-mass region (WW and  $t\bar{t}$  in primis) in data is estimated from the number of events in the  $e^{\pm}\mu^{\mp}$  final state  $(N_{in}^{e\mu})$ , applying a correction factor that accounts for the differences in the detection efficiency between electrons and muons,  $k^{ee/\mu\mu}$ , as shown in Eq. 6.14<sup>4</sup>:

$$N_{\ell\ell}^{\text{background}} = \frac{1}{2} k^{\ell\ell} \cdot N_{\text{e}\mu}$$
(6.14)

where

$$k^{ee} = \sqrt{\frac{N_{ee}}{N_{\mu\mu}}} k^{\mu\mu} = \sqrt{\frac{N_{\mu\mu}}{N_{ee}}}$$
(6.15)

In Eq. 6.15,  $N_{\rm ee}$  and  $N_{\mu\mu}$  are measured at WW level.

Fig. 6.24 shows the distribution of the invariant mass of the di-lepton system  $m_{\ell\ell}$  at WW level. It is clearly visible the Z peak, that defines the "in" region, and the "out" region, where the signal lies.

The ratio  $R_{out/in}$  can be obtained both from simulation (AB method, see Sec. 6.4) and data (ABCD method, see Sec. 6.4). In the simulation it is defined as the ratio  $N_{out}^{MC}/N_{in}^{MC}$ . While the first method was used at the beginning of data-taking, when MC statistics was big enough to have a reasonable small error, the second is currently used, thanks to the increased integrated luminosity and the smaller systematic error coming from the measurement of  $R_{out/in}$  from data. To ratio  $R_{out/in}$  from data is calculated separately for the di-electron and di-muon cases according

<sup>&</sup>lt;sup>4</sup> WW and  $t\bar{t}$  are assumed to have an equally probable production of the flavour pairs:  $(e, e), (\mu, \mu), (e, \mu), (\mu, e)$ .



Figure 6.24: The invariant mass of the di-lepton system  $m_{\ell\ell}$  at WW level. The "in" region (under the Z peak) and the "out" regions (where the signal is expected) are shown.

to Eq. 6.16 in a control region (CR), defined by applying all the analysis selections but the missing energy cut, thus obtaining an enriched  $Z/\gamma^* \to \ell^+ \ell^-$  phase space:

$$R_{\mu\mu} = \frac{N_{\mu\mu}^{\text{out}} - \frac{1}{2} k^{\mu\mu,\text{out}} \cdot N_{e\mu}^{\text{out}} - N_{\mu\mu,ZV}^{\text{out},MC}}{N_{\mu\mu}^{\text{in}} - \frac{1}{2} k^{\mu\mu,\text{in}} \cdot N_{e\mu}^{\text{in}} - N_{\mu\mu,ZV}^{\text{in},MC}}$$

$$R_{\text{ee}} = \frac{N_{\text{ee}}^{\text{out}} - \frac{1}{2} k^{\text{ee,out}} \cdot N_{e\mu}^{\text{out}} - N_{\text{ee},ZV}^{\text{out},MC}}{N_{\text{ee}}^{\text{in}} - \frac{1}{2} k^{\text{ee,in}} \cdot N_{e\mu}^{\text{in}} - N_{\text{ee},ZV}^{\text{out},MC}}$$
(6.16)

where:

- *in* stands for "under the Z peak", and *out* is "outside the Z peak"
- $N_{ZV}^{MC}$  represents the expected peaking ZZ and ZW contributions, estimated from simulation

Also in the measurement from data of the ratio  $R_{out/in}$  the WW/ $t\bar{t}$  contributions are subtracted using the number of events in the  $e^{\pm}\mu^{\mp}$  final state, while the ZZ and ZW contributions are subtracted using MC expectations.

Fig. 6.25 shows the distribution of the  $PF-E_T^{miss}$  in ee/ $\mu\mu$  events. The low  $PF-E_T^{miss}$  region is an enriched  $Z/\gamma^* \rightarrow \ell^+\ell^-$  phase space (CR), where the *R* factor is measured from data.

The amount of Drell-Yan background in the signal region ("out") is estimated by means of  $R_{\ell\ell}$ 



Figure 6.25: The PF $-E_{\rm T}^{\rm miss}$  distribution in ee/ $\mu\mu$  events. The low PF $-E_{\rm T}^{\rm miss}$  region is an enriched  $Z/\gamma^* \rightarrow \ell^+\ell^-$  phase space (CR), where the *R* factor is measured from data.

as described in Eq. 7.3:

$$N_{\ell\ell,DY}^{\text{out,data}} = \left(N_{\ell\ell}^{\text{in,data}} - \frac{k^{\ell\ell,\text{in}}}{2} \cdot N_{e\mu}^{\text{in,data}} - N_{ZV}^{\text{in,MC}}\right) \cdot R_{\ell\ell}$$
(6.17)

In addition to the statistical uncertainty on the number of DY events extrapolated from data, a systematic uncertainty is added, due to the hypothesis that the ratio between the number of DY events under Z peak and outside Z peak is the same in the control region (CR) and in the signal region (SR).

To test this assumption, the control region has been splitted into two subregions, and the variation of  $R_{\ell\ell}$  measured in the two sub-regions is taken as an estimation of the systematic error due to the hypothesis  $R_{\ell\ell}^{CR} = R_{\ell\ell}^{SR}$ . The systematic error will be reported in Sec. 7 for the VBF case and in Sec. 8 for the VH one for each Higgs mass hypothesis, and the order of magnitude due to this assumption is 20%.

### 6.4.4 Drell-Yan $\rightarrow \tau \tau$

The low  $E_{\rm T}^{\rm miss}$  threshold applied in the  $e\mu$  final state allows for a significant contribution of events from  $Z/\gamma^* \to \tau^+\tau^-$ , that is infact estimated from data. This is accomplished by using  $Z \to \mu^+\mu^-$  events, replacing in each event muons with  $\tau$ s with the same kinematics of the muon, and simulating the  $\tau \to l\nu_\tau \bar{\nu_e}$  decay. After replacing muons from  $Z \to \mu^+\mu^-$  decays with simulated  $\tau$  decays, the set of pseudo  $Z \to \tau^+\tau^-$  events undergoes the reconstruction step. In this way, the  $Z/\gamma^* \to \tau^+\tau^-$  sample used in the analysis is completely data driven, and it reflects correctly the distribution of jet, leptons and  $E_{\rm T}^{\rm miss}$  as observed in data. The procedure of removing the muon signature and place the  $\tau$  one, usually called *embedding*, is depicted in Fig. 6.26. The online and offline cuts to select  $Z \to \mu^+ \mu^-$  events are loose enough not to introduce bias in the  $Z/\gamma^* \to \tau^+ \tau^-$  results.



Figure 6.26: Embedding procedure for  $Z/\gamma^* \to \tau^+\tau^-$ . Starting from a  $Z \to \mu^+\mu^-$  event, muons from  $Z \to \mu^+\mu^-$  decays are replaced with simulated  $\tau$  decays, and the set of pseudo  $Z \to \tau^+\tau^-$  events undergoes the reconstruction step.

To test the correctness of the embedding procedure, the kinematic distributions of leptons and jets from the  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  data-driven sample have been compared to the ones obtained from MC as shown in Fig. 6.27: a good agreement is found.

In addition, a highly pure  $Z/\gamma^* \to \tau^+ \tau^-$  phase space is used to normalize the whole  $Z/\gamma^* \to \tau^+ \tau^-$  embedded sample: if just one jet is required in the event and looking at the low  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  region, as defined in Eq. 6.18, most of events are expected to come from  $Z/\gamma^* \to \tau^+ \tau^-$ , as shown in 6.28.

$$m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}} = \sqrt{2 \cdot p_T^{\ell\ell} \cdot E_{\rm T}^{\rm miss}} \left(1 - \cos\Delta\phi_{\ell\ell, E_{\rm T}^{\rm miss}}\right) \tag{6.18}$$

In this phase space a comparison of kinematic variables distributions between data and  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  embedded sample is performed in order to test the  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  sample that will be used in the analysis: Fig. 6.29 shows good agreement, then confirming the good description obtained with the embedding procedure.



Figure 6.27: Kinematic distributions for leptons and jet in the embedded  $Z/\gamma^* \rightarrow \tau^+ \tau^$ sample and in the MC  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  at WW level, for two-jet-bin analysis: a good agreement is found. The two set of histograms are normalized to the same area.



Figure 6.28: Comparison data/MC in the one-jet-bin phase space. Low  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  is dominated by  $Z/\gamma^* \to \tau^+\tau^-$  sample. This region ( $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$ <40 GeV) is used to normalize the embedded  $Z/\gamma^* \to \tau^+\tau^-$  sample to data and to check the good agreement between shapes in data and the embedded  $Z/\gamma^* \to \tau^+\tau^-$  sample, see Fig. 6.29.



Figure 6.29: Comparison between data and  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  embedded sample in a higly pure  $Z/\gamma^* \rightarrow \tau^+ \tau^-$  phase space (one-jet-bin,  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}} < 40 \text{ GeV}$ ). Good data/MC agreement is found.

### 6.4.5 Other Backgrounds

There are four processes which need to be estimated from Monte Carlo simulation, after applying the proper data corrections for lepton, trigger and jet veto efficiencies: WW, WZ, ZZ and  $W\gamma$ .

For the WW background, that is the most important among these, to estimate the degree of belief on the MC expectation, two different MC generators are used, namely MC@NLO [115] and Madgraph [95]. The difference between these two generators is taken as systematic error on the estimation of WW.

# 6.5 Efficiency Measurements

The tag and probe method on  $Z/\gamma^* \to \ell^+ \ell^-$  events is used to provide an unbiased, high-purity, lepton sample with which to measure both online and offline lepton selection efficiencies. The method relies on tagging of  $Z/\gamma^* \to \ell^+ \ell^-$  events using one good reconstructed letpon (the tag) and an additional lepton that passes loose selections (the probe). The invariant mass of the pair is required to be close to  $M_Z$ . The efficiency is measured counting the number of events where the probe passes the tight selections. The measurement is performed in bins of  $\eta$  and  $p_T$ of the probe lepton to guarantee a good description of geometrical acceptancies of the detector. In this way, both trigger and lepton identification efficiencies are measured from data and a scale factor is applied to MC in order to correct for small effects not taken into account by the simulation of the interaction with the detector.

# 6.6 Systematics

Because of the impossibility to reconstruct an invariant mass peak, the analysis is to a large extent a counting experiment. Therefore it is important to understand the signal efficiency and the background predictions. The uncertainty due to some assumptions in signal and background predictions, such as a good modelling of lepton energy scales, the choice of a specific MC generator, the knowledge of the delivered luminosity must be taken into account.

The following experimental systematic uncertainties have been considered:

- Luminosity. Based on the CMS online luminosity monitoring the uncertainty is currently 2.2% in 2011 [78] and 4.5% in 2012 [116].
- **Trigger efficiency.** Trigger data-driven measurement has its own error. The uncertainty on the trigger efficiency is propagated through the whole analysis workflow and the error on the electrons and muons yields is less than 1%.
- Lepton reconstruction and identification efficiencies. The uncertainty on the lepton reconstruction and identification efficiency is at the order of 2%.
- Muon momentum and electron energy scale. The electron energy was varied by 2% in the barrel and 5% in the endcap. The systematic uncertainty is about 2% per electron. For the muons the uncertainty is much smaller, and a conservative 0.5% is considered.
- $E_{\rm T}^{\rm miss}$  modeling: a data-driven method to estimate the  $Z/\gamma^* \to \ell^+ \ell^-$  background, which is affected by the  $E_{\rm T}^{\rm miss}$  resolution, is used. Events with neutrinos giving real  $E_{\rm T}^{\rm miss}$  in the final state also have a small uncertainty. This uncertainty is estimated on the event selection efficiency by varying the  $E_{\rm T}^{\rm miss}$  in signal events by an additional 10%. An uncertainty on the event selection efficiency of around 2% is found.
- Jet energy scale (JES). It affects both the jet multiplicity and jet kinematics This error is estimated applying variations of the official jet uncertainties on the JES (which depend on  $\eta$  and  $p_{\rm T}$  of the jet) and compute the variation of the selection efficiency. It turns out to be less than 7%.
- B-mistag modelling. The uncertainties on the selection of not-b jets (TCHE cut) is taken into account looking at the efficiency of the b-veto for events in a DY enriched phase space. The ratio between the efficiency measured in data and in Monte Carlo is considered as an estimation of the scale factor related to the b-mistag modelling, to be applied to all samples that are not data-driven (such as the signal). The scale factor is found to be 0.97 with an uncertainty of 2%. Fig. 6.30 shows the b-tag distribution in the enriched Z/γ\* → ℓ<sup>+</sup>ℓ<sup>-</sup> phase space.
- **Pile-up.** The simulation has been reweighted according to the data instantaneous luminosity. An uncertainty of 8% in the knowledge of the number of interactions was propagated to the pile-up re-weighting procedure. The obtained variation in the expected events is of about 2%.

The following theoretical systematic uncertainties have been considered:

- Higgs boson production cross-section. The uncertainties on the inclusive cross-section for the Higgs sample have been taken from the LHC Higgs Cross Section working group report [12] and are about 20% for gluon fusion contribution, about 2% for VBF one and about 5% for associated Higgs production.
- **PDFs uncertainties.** They have been estimated according to the recipe provided by the LHC Higgs Cross Section working group [12]. Different sets of Parton Density Functions (PDFs) have been tested which change the acceptance of the measurement. The effect on the selection efficiency is 1%, 2% and 1% for the signal, respectively Higgs strahlung, gluon



Figure 6.30: The b-tag distribution in a  $Z/\gamma^* \rightarrow \ell^+ \ell^-$  enriched phase space. The number of b-tagged events and the number of b-vetoed events in compared in data and MC, obtaining a data over MC scale factor of 0.97, with an uncertainty of 2%.

fusion and VBF.

- QCD scale uncertainties. They have been estimated according to the recipe provided by the LHC Higgs Cross Section working group [12], varying the normalization and factorization scales in the production of MC events. The effect on the selection efficiency is between 10% and 30%, depending on the MC considered.
- UEPS. The uncertainty on the underlying event (UE) and parton shower (PS) models has been estimated by comparing the signal efficiencies with different parton showers and different tunes of the underlying event generation. The effect is of the order of 30%.

The limited statistics of some MC samples is then considered as an additional uncertainty:

• Monte Carlo statistics. It contributes as an uncertainty of about 15% to the signal efficiencies and about 20% for most of backgrounds.

The detailed values of the systematics in the two analyses (VBF and VH) will be reported in the dedicated chapters (Sec. 7 and Sec. 8).
### 6.7 Limit extraction and discovery significance

Upper limits on the product of the Higgs boson production cross section and the  $H \to W^+W^$ branching fraction,  $\sigma_{Higgs} \times BR(H \to W^+W^-)$ , with respect to the SM expectation, i.e.  $\sigma_{95\%}/\sigma_{SM}$ , are derived with two different statistical methods. The first method is based on Bayesian inference [117] and the second one, known as CLs , is the modified frequentist approach [118, 119, 120].

The Bayesian and the classical frequentist, with a number of modifications, are two statistical approaches commonly used in high energy Physics for characterising the absence of a signal. Both methods allow to quantify the level of incompatibility of data with a signal hypothesis, which is expressed as a confidence level (CL). The probabilistic interpretation of CL as the chance of being right or wrong when stating the non-existence of a signal is not straightforward. In addition, in an analysis targeting a specific signal production mechanism and a particular decay mode, one can also set approximately model-independent limits on signal cross section times branching ratio ( $\sigma \times BR$ ) or somewhat better defined limits on cross section times branching ratio times experimental acceptance ( $\sigma \times BR \times A$ ). The latter are less useful for testing various theories unless a model of the experimental acceptance A is also provided.

In a combination of multiple analyses sensitive to different signal production mechanisms and different decay modes, presenting results in a form of limits on  $\sigma \times BR$  or  $\sigma \times BR \times A$  is impossible. The customary alternative for SM Higgs searches is to set limits on a common signal strength modifier  $\mu$  that is taken to change the cross sections of all production mechanisms by exactly the same scale. Decay branching ratios are assumed to be those given by the Standard Model. The Standard Model Higgs is said to be excluded at, 95%CL, when the 95% CL limit on  $\mu$  drops to one, i.e.  $\mu_{95\%CL} = 1$ .

The limit extraction procedure is described in detail in [121]. In the following the basic steps are summarized.

• Construct likelihood function  $\mathscr{L}(data|\mu, \vartheta)$ 

$$\mathscr{L}(data|\mu,\vartheta) = prob\left(data|\mu \cdot s(\vartheta) + b(\vartheta)\right) \cdot p(\vartheta|\vartheta) \tag{6.19}$$

Here "data" represents either the actual experimental observation or pseudo-data used to construct sampling distributions to be discussed further below. The parameter  $\mu$  is the signal strength modifier and  $\vartheta$  represents the full set of nuisance parameters (systematics). The probability distribution  $p(\tilde{\vartheta}|\vartheta)$  is the prior distribution of the systematics and  $\tilde{\vartheta}$  is the default value of the systematics.

 $prob (data | \mu \cdot s + b)$  stands either for a product of Poisson probabilities to observe  $n_i$  events in bins:

$$\prod_{i} \frac{(\mu \cdot s_{i} + b_{i})^{n_{i}}}{n_{i}!} e^{-\mu \cdot s_{i} - b_{i}}$$
(6.20)

or for an unbinned likelihood over k events in the data sample:

$$k^{-1} \prod_{i} \left( \mu \cdot Sf_s(x_i) + Bf_b(x_i) \right) e^{-\mu \cdot S - B}$$
(6.21)

In the latter equation,  $f_s(x)$  and  $f_b(x)$  are pdfs of signal and background of some observ-

able(s) x, while S and B are total event rates expected for signal and backgrounds.

• To compare the compatibility of the data with the background-only and signal+background hypotheses, where the signal is allowed to be scaled by some factor  $\mu$ , the test statistic  $\tilde{q}_{\mu}$  is constructed based on the profile likelihood ratio:

$$\tilde{q}_{\mu} = -2ln \frac{\mathscr{L}(data|\mu, \hat{\vartheta}_{\mu})}{\mathscr{L}(data|\hat{\mu}, \hat{\vartheta})}, \text{ with a constraint} 0 \le \hat{\mu} \le \mu$$
(6.22)

where  $\hat{\vartheta}_{\mu}$  refers to the conditional maximum likelihood estimators of  $\vartheta$ , given the signal strength parameter  $\mu$  and "data" that, as before, may refer to the actual experimental observation or pseudo-data (toys). The pair of parameter estimators  $\hat{\mu}$  and  $\hat{\vartheta}$  correspond to the global maximum of the likelihood.

The lower constraint  $0 \leq \hat{\mu}$  is dictated by physics (signal rate is positive), while the upper constraint  $\hat{\mu} \leq \mu$  is imposed by hand in order to guarantee a one-sided (not detached from zero) confidence interval. Physics-wise, this means that upward fluctuations of the data such that  $\hat{\mu} > \mu$  are not considered as evidence against the signal hypothesis.

- Find the observed value of the test statistic  $\tilde{q}_{\mu}^{obs}$  for the given signal strength modifier  $\mu$  under test.
- Find values of the nuisance parameters  $\hat{\vartheta}_{0}^{obs}$  and  $\hat{\vartheta}_{\mu}^{obs}$  best describing the experimentally observed data (i.e. maximising the likelihood as given in Eq. 6.19), for the background-only and signal+background hypotheses, respectively.
- Generate toy Monte Carlo pseudo-data to construct pdfs  $f(\tilde{q}_{\mu}|\mu, \hat{\vartheta}_{\mu}^{obs})$  and  $f(\tilde{q}_{\mu}|0, \hat{\vartheta}_{0}^{obs})$  assuming a signal with strength  $\mu$  in the signal+background hypothesis and for the background-only hypothesis ( $\mu = 0$ ). An example of the pdfs of  $\tilde{q}_{\mu}$  is shown in Fig. 9.7.

For the purpose of generating a pseudo-dataset, the nuisance parameters are fixed to the values  $\hat{\vartheta}^{obs}_{\mu}$  and  $\hat{\vartheta}^{obs}_{0}$  obtained by fitting the observed data, but are allowed to float in fits needed to evaluate the test statistic. This choice, in which the nuisance parameters are fixed to their maximum likelihood estimates, has good coverage properties [122].



Figure 6.31: Distribution of the test statistic  $\tilde{q}_{\mu}$  with pseudo-data generated for signal plus background and background-only hypotheses.

• Having constructed  $f(\tilde{q}_{\mu}|\mu, \hat{\vartheta}_{\mu}^{obs})$  and  $f(\tilde{q}_{\mu}|0, \hat{\vartheta}_{0}^{obs})$  distributions, two p-values are defined to be associated with the actual observation for the signal+background and background-only hypotheses,  $p_{\mu}$  and  $p_{b}$ :

$$p_{\mu} = P(\tilde{q}_{\mu} \ge \tilde{q}_{\mu}^{obs} | \text{signal} + \text{background}) = \int_{\tilde{q}_{\mu}^{obs}}^{\infty} f(\tilde{q}_{\mu} | \mu, \hat{\vartheta}_{\mu}^{obs}) d\tilde{q}_{\mu}$$
(6.23)

$$1 - p_b = P(\tilde{q}_\mu \ge \tilde{q}_\mu^{obs} | \text{background} - \text{only}) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu | 0, \hat{\vartheta}_0^{obs}) d\tilde{q}_\mu$$
(6.24)

and calculate CLs  $(\mu)$  as a ratio of these two probabilities

$$CLs(\mu) = \frac{p_{\mu}}{1 - p_b} \tag{6.25}$$

- If, for  $\mu = 1$ , CLs  $\leq \alpha$ , the SM Higgs boson is excluded with  $(1 \alpha)$  CLs confidence level. It is known that the CLs method gives conservative limits, i.e. the actual confidence level is higher than  $(1 - \alpha)$ .
- To quote the 95% Confidence Level upper limit on  $\mu$ , to be further denoted as  $\mu_{95\% CL}$ ,  $\mu$  is varied until CLs = 0.05 is reached.

On the other hand, the Bayesian method is based on interpreting the likelihood (Eq. 6.19) as a probability distribution function with a flat prior for the signal strength and a set of pdfs for nuisance parameters, which are often approximated with the log-normal distribution. Integrating over the nuisance parameters the upper limit for the signal strength is calculated. The results obtained using the two methods may differ but in most cases they are very close.

The software package RooStats [123] has been used to perform the computation of the limits.

From 2010 until the beginning of 2012 the main results were reported by means of exclusion limits. With the same statistical approach it was possible, since 2012, to report not only exclusion limits but also discovery significance. The attention has been focused on "number of standard deviations" with respect to the background-only hypothesis (number of  $\sigma$ s): the probability for a background-only hypothesis fluctuation to have a result more signal-like than the one actually observed is measured (p-value, 1-p<sub>b</sub> in Eq. 6.24) and expressed in n<sup>o</sup> $\sigma$  using the conversion Table 6.6.

number of $\sigma$	probability
1	0.159
2	$2.28 \cdot 10^{-2}$
3	$1.30 \cdot 10^{-3}$
4	$3.0 \cdot 10^{-5}$
5	$5 \cdot 10^{-7}$

Table 6.6: Number of  $\sigma$  and probability in the gaussian parent distribution hypothesis. This conversion table is used to express p-value in terms of number of  $\sigma$  for discovery. A conventional number of  $5\sigma$  is commonly accepted to state discovery with great confidence level.

In the following chapters both the exclusion limit and the discovery plots (when meaningful) will be reported.

## CHAPTER 7

### THE VBF ANALYSIS

Sed omnia praeclara tam difficilia, quam rara sunt.

Baruch Spinoza, De potentia intellectus seu de libertate humana

The VBF analysis is a key ingredient for the exclusion or discovery of a Higgs boson at LHC. It probes a specific production mode of the Higgs and, at the same time, it is a crucial benchmark for the search of new Physics looking at the scattering of two W bosons. The SM cross section measurement of the longitudinal scattering of W bosons is not calculable for high center of mass energies [124], therefore measuring the WW cross section is a key ingredient in absence of a low mass Higgs and for the search of new Physics as deviation with respect to SM prediction.

In Fig. 7.1 the Feynman diagram of the VBF Higgs production mechanism and its decay is shown, while in Fig. 7.2 the expected number of events in  $1 \text{fb}^{-1}$  are reported.



Figure 7.1: Feynman diagram for Higgs production via Vector Boson Fusion and decaying into  $W^+W^- \rightarrow \ell^+ \nu \ell^- \bar{\nu}$ , with  $\ell = e/\mu$ .

In the following the selections applied in the analysis are summarized in Sec. 7.1, the specific aspects of background estimation are revisited in Sec. 7.2. The systematics are summarized in Sec. 7.3 and the results are reported in Sec. 7.4.



Figure 7.2: The expected number of events in the VBF H  $\rightarrow$  WW  $\rightarrow \ell \nu \ell \nu$  production and decay mode in 1fb<sup>-1</sup> at 7 (top) and 8 (bottom) TeV of center of mass energy. A zoom in the low mass region is shown. The error bands correspond to the combination of the uncertainties related to QCD scale, pdf uncertainty and  $\alpha_{QCD}$  [12, 13].

### 7.1 VBF selections

On top of WW level (see Sec. 6.3), summarized in Table 7.1, specific selections to enhance VBF Higgs events with respect to backgrounds have been optimized. A set of selections is common to different Higgs mass searches, while a list of Higgs mass dependent cuts have been applied.

Variable	selection
$p_{\rm T} 1^{st}$ lepton	> 20  GeV
$p_{\rm T} 2^{nd}$ lepton	$> 20 { m GeV}$
jet $p_{\mathrm{T}}$	$> 30  { m GeV}$
opposite lepton charge	$q^1 \cdot q^2 = -1$
$3^{rd}$ lepton veto	no extra leptons with $p_{\rm T} > 10 \text{ GeV}$
$\mathrm{PF}E_{\mathrm{T}}^{\mathrm{miss}}$	$ >20~{ m GeV}~{ m in}~{ m e}\mu/\mu{ m e}~{ m and}>45~{ m GeV}~{ m in}~{ m ee}/\mu\mu$
$min - proj E_{\rm T}^{\rm miss}$	$> 20 \mathrm{GeV}$
$m_{\ell\ell}$	$> 12  { m GeV}  { m and}  otin m_{ m Z} \pm 15 { m GeV}  { m in}  { m ee}/\mu\mu$
$\operatorname{Proj-}E_{\mathrm{T}}^{\mathrm{miss}}$	$> 20 { m GeV}$
Soft muon veto	
jet TCHE	< 2.1
$p_T^{\ell\ell}$	> 45  GeV
$\Delta \phi_{\ell\ell,jj}$	$< 165^{\circ}$

Table 7.1: WW level selections for VBF analysis (see Sec. 6.3).

Because of the impossibility to reconstruct an invariant mass peak, in the  $H \to WW \to \ell \nu \ell \nu$ analysis, a mass-like variable has been developed: the Higgs transverse mass  $m_T^{\ell \ell E_T^{miss}}$ , as defined in Eq. 7.1, computed with the two leptons and the missing energy, is required to be in a window optimized with respect to the Higgs mass.

$$m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}} = \sqrt{2 \cdot p_T^{\ell\ell} \cdot E_{\rm T}^{\rm miss} \left(1 - \cos\Delta\phi_{\ell\ell, E_{\rm T}^{\rm miss}}\right)}$$
(7.1)

The cuts applied are:

- Lepton selections: Higgs boson mass dependent selections summarized in Table 7.2. In particular, the variables used to discriminate signal from background are:
  - lepton  $p_{\rm T}$
  - azimuthal angle between the two leptons  $(\Delta \phi_{\ell\ell})$
  - invariant mass of di-lepton system  $(m_{\ell\ell})$ .
- $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  window selection: 30 GeV  $< m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}} < \max m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$ . Fig. 7.3 shows the  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  distribution for different Higgs mass hypotheses. For high mass Higgs hypotheses ( $m_{\rm H}>155\,{\rm GeV}$ ) the maximum value of  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  is the Higgs mass itself, that is  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}} < m_{\rm H}$ . For low Higgs mass hypotheses tighter selections are applied, as summarized in Table 7.2.
- Lepton centrality; leptons are required to be within the  $\eta$  acceptance region defined by the two tag jets.

- Jet selections:
  - $-\Delta\eta_{ij} > 3.5$
  - invariant mass of di-jet system,  $m_{jj} > 500$  GeV
  - central jet veto (CJV); no jets with  $p_{\rm T} > 30$  GeV between the two tag jets.



Figure 7.3: The  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  distribution at WW level for different Higgs mass hypotheses. All the distributions are normalized to unity.

Because of the scalar nature of the Higgs boson and the vectorial one of the W, to conserve the angular momentum the spins of the W bosons produced in  $H \rightarrow WW$  decay have to be anticorrelated. The z axis is the decay direction of the WW system in the Higgs rest frame and the longitudinal  $(L, S_z = \pm 1)$  and transverse  $(T, S_z = 0)$  polarizations with respect to such axis are considered. In the Higgs rest frame, only the decays  $H \to W_T W_T$  and  $H \to W_L W_L$ are possible and the decay  $H \to W_T W_L$  is forbidden. The W polarizations are not directly observable, instead the final state charged leptons are observed. The decay rate of  $W_T^+ \to \ell^+ \nu$  is proportional to  $(1 + \cos \theta)^2$ , where  $\theta$  is the angle between the lepton direction and the  $W_T^+$  spin, therefore the right-handed lepton is emitted in most of cases in the same direction as the  $W_T^+$ spin. Similarly, the left handed electron is emitted in the opposite direction with respect to the  $W_T^-$  spin since its decay follows a  $(1 - \cos \theta)^2$  distribution. Being the two Ws anti-correlated, the electrons are mainly emitted in the same direction. Similar considerations also apply to the case of longitudinally polarized Ws [125]. A schematic cartoon is given in Fig. 7.4. In the case of the WW background, the initial state is unpolarized, therefore the combinations  $W_T W_T$ ,  $W_L W_L$  and  $W_T W_L$  are all allowed and the directions of the two electrons are not correlated. A small opening angle between the two leptons  $\Delta \phi_{\ell\ell}$  is therefore a good discriminating variable to separate the signal from the background, as summarized in Table 7.2.

Fig.7.5 shows the di-jet pseudorapidity separation at WW level: in a VBF Higgs event, the tag jets are expected to have a big  $\eta$  difference. The di-jet system invariant mass at WW level is



Figure 7.4: Scheme of the spin correlations which characterize the H  $\rightarrow$  WW  $\rightarrow \ell \nu \ell \nu$  decay. The two leptons are mainly emitted in the same direction then a  $\Delta \phi_{\ell \ell}$  selection can discriminate H  $\rightarrow$  W<sup>+</sup>W<sup>-</sup> from WW background.

Higgs mass [GeV]	$p_{\mathrm{T}}^{lep} \ 1^{st} \ [\mathrm{GeV}]$	$p_{\rm T}^{lep} 2^{nd}  [{\rm GeV}]$	$\Delta \phi_{\ell\ell}$	$m_{\ell\ell}[\text{GeV}]$	$m_{\mathrm{T}}^{\ell\ell E_{\mathrm{T}}^{\mathrm{miss}}}[\mathrm{GeV}]$
110	20	10	$< 115^{\circ}$	< 40	$[ \ 30 \ , \ 110 \ ]$
115	20	10	$< 115^{\circ}$	< 40	$[\ 30\ ,\ 110\ ]$
120	20	10	$< 115^{\circ}$	< 40	$[\ 30\ ,\ 120\ ]$
125	23	10	$< 110^{\circ}$	< 43	$[\ 30\ ,\ 123\ ]$
130	25	10	$< 90^{\circ}$	< 45	$[\ 30 \ , \ 125 \ ]$
135	25	12	$< 90^{\circ}$	< 45	$[\ 30\ ,\ 128\ ]$
140	25	15	$< 90^{\circ}$	< 45	$[\ 30\ ,\ 130\ ]$
145	25	15	$< 90^{\circ}$	< 48	$[\ 30\ ,\ 140\ ]$
150	27	25	$< 90^{\circ}$	< 50	$[\ 30\ ,\ 150\ ]$
155	27	25	$< 90^{\circ}$	< 50	$[\ 30 \ , \ 155 \ ]$
160	30	25	$< 60^{\circ}$	< 50	$[\ 30\ ,\ 160\ ]$
170	34	25	$< 60^{\circ}$	< 50	$[\ 30 \ , \ 170 \ ]$
180	36	25	$< 70^{\circ}$	< 60	$[\ 30\ ,\ 180\ ]$
190	38	25	$< 90^{\circ}$	< 80	$[\ 30\ ,\ 190\ ]$
200	40	25	$< 100^{\circ}$	< 90	$[\ 30\ ,\ 200\ ]$
250	55	25	$< 140^{\circ}$	< 150	$[\ 30\ ,\ 250\ ]$
300	70	25	$< 175^{\circ}$	< 200	$[\ 30\ ,\ 300\ ]$
350	80	25	$< 175^{\circ}$	< 250	$[\ 30\ ,\ 350\ ]$
400	90	25	$< 175^{\circ}$	< 300	$[\ 30\ ,\ 400\ ]$
450	110	25	$< 175^{\circ}$	< 350	$[\ 30\ ,\ 450\ ]$
500	120	25	$< 175^{\circ}$	< 400	$[\ 30\ ,\ 500\ ]$
550	130	25	$< 175^{\circ}$	< 450	$[\ 30\ ,\ 550\ ]$
600	140	25	$< 175^{\circ}$	< 500	$[\ 30\ ,\ 600\ ]$

Table 7.2: Lepton selections for different Higgs mass searches.



shown in Fig.7.6: it is evident that the VBF contribution is enhanced in the high mass tail.

Figure 7.5: The  $\Delta \eta_{jj}$  distribution at WW level. The VBF Higgs signature is characterized by two jets with large pseudorapidity separation. A cut on this variable at 3.5 reduces the background and the gluon fusion contamination. On the left the distribution normalized to the luminosity with data and stacked MC distributions. On the right all MC distributions are normalized to unity, thus showning the effect of the selection.

One of the main differences between VBF Higgs production and most of the backgrounds (top in primis) is represented by a low hadronic activity between the two hardest jets, due to the exchange of colorless vector bosons in the electroweak t-channel. This property is taken into account by the central jet veto (CJV) selection, requiring no hard jets ( $p_{\rm T} > 30$  GeV) within the leading ones.

The tight selections on the invariant mass of the di-jet system and the high pseudorapidity difference between the tag jets select a highly pure VBF sample: gluon fusion Higgs contamination is reduced to less than 20%. The cut optimization was driven by having the best signal over background ratio, that is the best discovery/exclusion sensitivity, and by assuring a reasonable number of signal events after the selections are applied. A smooth variation of kinematic thresholds are applied to distinguish different Higgs mass hypotheses.

### 7.2 Background estimation

The main background in the VBF analysis is Top. The complete list of backgrounds with a data-driven estimation is:

- Top (Sec. 6.4.2)
- $Z/\gamma^* \rightarrow \ell^+ \ell^-$ (Sec. 6.4.3)
- $Z/\gamma^* \to \tau^+ \tau^-$ (Sec. 6.4.4)



Figure 7.6: The  $m_{jj}$  distribution at WW level. The VBF Higgs signature is characterized by high di-jet invariant mass. A cut on this variable at 500 GeV reduces the background and the gluon fusion contamination. On the left the distribution normalized to the luminosity with data and stacked MC distributions. On the right all MC distributions are normalized to unity, thus showning the effect of the selection.

• W+jets(Sec. 6.4.1)

The details of the methods have been described in the related sections. In the following the results of the data-driven estimation for top and  $Z/\gamma^* \to \ell^+ \ell^-$  and the differences with respect to what was described in Section 6.4 are reported. The estimated number of  $Z/\gamma^* \to \tau^+ \tau^-$  and W+jets events are listed in the summary tables.

### 7.2.1 The top background

With respect to what was described in Sec. 6.4.2, the main problem related to top estimation in the VBF phase space is that, after applying tight selections of the jets, such as  $m_{jj} > 500$  GeV and  $\Delta \eta_{jj} > 3.5$ , in about the 10% of the events even the most central of the two jets is outside the tracker acceptance ( $|\eta| < 2.5$ ). In this case both jets pass the b-veto requirement, since no tracks are reconstructed and the jet is based on calorimeter information only. The data-driven top estimation must take into account for this fact and estimate also the number of events outside the tracker acceptance.

The b-tag efficiency is measured at WW level as described in Sec. 6.4.2. In Table 7.3 the efficiencies measured on data are shown, as well as the efficiencies on a top MC sample at WW level (control region, CR) and after all selections are applied (signal region, SR). The hypothesis of the data-driven estimation is that the b-tag efficiency does not change after selections are applied: as shown in Table 7.3, within the Monte Carlo statistics this assumption is confirmed. The data-driven estimation relies also on the fact that the  $\eta$  distribution of the most central jet in top MC events is well reproduced, as shown in Fig. 7.7. A global scale factor for top events is

$ \eta^{cj} $ bin	$\varepsilon^{\mathrm{DATA}}( \eta^{cj} )$	$\varepsilon^{\mathrm{MC,CR}}( \eta^{cj} )$	$\varepsilon^{\mathrm{MC,SR}}( \eta^{cj} )$
$0 <  \eta  < 0.5$	$0.672 \pm 0.008$	$0.704 \pm 0.004$	$0.88\pm0.50$
$0.5 <  \eta  < 1$	$0.664 \pm 0.011$	$0.700\pm0.006$	$0.70\pm0.28$
$1 <  \eta  < 1.5$	$0.606 \pm 0.017$	$0.633 \pm 0.008$	$0.48\pm0.15$
$1.5 <  \eta  < 2.5$	$0.474 \pm 0.026$	$0.562\pm0.012$	$0.47\pm0.10$

Table 7.3: The b-tagging efficiency measured in a control region, according to Equation 6.12, in bins of  $|\eta|$  of the most central jet.

measured from data by means of the Eq. 7.2 and applied to MC prediction.

$$N_{bveto}^{\text{data driven},|\eta|<2.5} = \int_{|\eta|<2.5} d\eta \ N_{bveto}^{\text{data driven}}(\eta)$$
$$= \int_{|\eta|<2.5} d\eta \ \left[ \left( N_{btag}^{DATA} - N_{btag}^{\text{other backgrounds}} \right)(\eta) \frac{1-\varepsilon(\eta)}{\varepsilon(\eta)} \right]$$
(7.2)

Non-top background contamination in the b-tagged region,  $N_{btag}^{other backgrounds}$ , are subtracted using simulation or data-driven estimation, when available.

In order to have a reasonable number of events in the b-tagged region to estrapolate in the b-vetoed region, the top estimation has been performed in all final state together, that is SF and DF. The division between SF and DF events is measured in data at WW level and estrapolated at Higgs level, that is after all selections are applied, using MC simulation.

In Table 7.4 the detailed composition of the top estimation is reported: the extrapolated number of top events, the number of b-tagged events, the error coming from the measurement of b-tag efficiency, the error coming form the variation of the efficiency at WW level and at Higgs level and the error coming from the subtraction of non-top contributions in the btagged region.

$m_H$	top estimation	$N^{b-tagged}$	$\varepsilon$ error	$\varepsilon \text{ MC error}$	MC Subtraction error
110	$2.4 \pm 1.3$	3	7.2 %	4.3 %	0.6~%
125	$2.2\pm1.2$	3	7.6~%	8.2~%	0.6 %
140	$1.4 \pm 1.0$	2	$5.1 \ \%$	6.1~%	0.6 %
160	$6.0\pm3.3$	4	6.1~%	34~%	0.5~%
200	$6.2\pm2.6$	5	$7.2 \ \%$	$17 \ \%$	1.2 %
400	$23.1\pm9.7$	15	6.6~%	34~%	1.1 %
600	$4.6\pm3.0$	4	5.8~%	41 %	4.2 %

Table 7.4: For each Higgs mass hypothesis the extrapolated number of top events (top estimation), the number of b-tagged events  $(N^{b-tagged})$ , the error coming from the measurement of b-tag efficiency ( $\varepsilon$  error), the error coming form the variation of the efficiency in control region and in signal region ( $\varepsilon$  MC error) and the error coming from the uncertainty related to the subtraction of other backgrounds in the btagged region (MC Subtraction error) are reported.



Figure 7.7: The  $\eta$  distribution for the most central jet at Higgs level, without applying bveto. The top data-driven estimation relies also on the fact that the  $\eta$  distribution of the most central jet in top MC events is well reproduced. A Good agreement is found.

## 7.2.2 The $Z/\gamma^* \rightarrow \ell^+ \ell^-$ background

The  $Z/\gamma^* \to \ell^+ \ell^-$  data-driven estimation is described in Sec. 6.4.3. In order to enrich the "in" region (under the Z peak) the tight  $p_{\rm T}$  selections for the leptons are not applied. Then the "in" and "out" regions are defined as:

- *in* ; under Z peak ( $m_{\ell\ell} \in m_Z \pm 15 \text{GeV}$ ) and without lepton  $p_T$  tight cuts (requiring  $p_T^{lep,max} > 20 \text{ GeV}$  and  $p_T^{lep,min} > 10 \text{ GeV}$ )
- *out*; all cuts defined in the analysis are applied.

To summarize, the amount of Drell-Yan background at Higgs level, is estimated by means of  $R_{\ell\ell}$ , measured inverting the PF- $E_{\rm T}^{\rm miss}$  selection (30 GeV< PF- $E_{\rm T}^{\rm miss}$ < 45 GeV), as described in Eq.7.3:

$$N_{\ell\ell,DY}^{\text{out,data}} = \left( N_{\ell\ell}^{\text{in,data}} - \frac{k^{\ell\ell,\text{in}}}{2} \cdot N_{e\mu}^{\text{in,data}} - N_{ZV}^{\text{in,MC}} \right) \cdot R_{\ell\ell}$$
(7.3)

In addition to the statistical uncertainty on the number of DY events extrapolated from data, a systematic uncertainty is added, due to the hypothesis that the ratio between the number of DY events in the "in" zone and in the "out" zone is independent on the PF- $E_{\rm T}^{\rm miss}$  cut, since it has been measured requiring PF- $E_{\rm T}^{\rm miss} < 45$  GeV and  $R_{\ell\ell}$  is applied after selecting events with PF- $E_{\rm T}^{\rm miss} > 45$  GeV.

To test this assumption, for each mass point, the PF- $E_{\rm T}^{\rm miss} < 45$  GeV region has been splitted into two subregions, namely  $\alpha$  and  $\beta$ , defined in Eq. 7.4.

$$\alpha \text{ zone}: \quad 30 < \text{PF} - E_{\text{T}}^{\text{miss}} < 35 \text{GeV}$$

$$\beta \text{ zone}: \quad 35 < \text{PF} - E_{\text{T}}^{\text{miss}} < 45 \text{GeV}$$
(7.4)

The difference in the ratio  $R_{\ell\ell}$  between  $\alpha$  and  $\beta$  regions is considered as systematic uncertainty. The  $R_{\ell\ell}$  values measured for different Higgs mass working points are reported in Table 7.5. Fig. 7.8 shows the  $m_{\ell\ell}$  distribution at Higgs level. The "in" and "out" regions are delimited by red lines. The big statistical uncertainty is clearly visible, just counting the number of events in the "in" region.

The estimated contamination of Drell-Yan contribution at the Higgs level is reported in Table 7.6.

$m_{\rm H} \; [{\rm GeV}]$	$R_{\mu\mu} \pm \text{stat} \pm \text{syst}$	$R_{\rm ee} \pm { m stat} \pm { m syst}$
110	$0.153 \pm 0.048 \pm 0.053$	$0.148 \pm 0.068 \pm 0.110$
125	$0.143 \pm 0.043 \pm 0.074$	$0.211 \pm 0.078 \pm 0.039$
140	$0.128 \pm 0.040 \pm 0.005$	$0.211 \pm 0.078 \pm 0.040$
160	$0.044 \pm 0.022 \pm 0.046$	$0.147 \pm 0.062 \pm 0.036$
200	$0.055 \pm 0.025 \pm 0.068$	$0.147 \pm 0.062 \pm 0.036$
400	$0.033 \pm 0.019 \pm 0.024$	$0.074 \pm 0.043 \pm 0.076$
600	$0.022 \pm 0.016 \pm 0.020$	$0.049 \pm 0.035 \pm 0.022$

Table 7.5: The  $R_{\ell\ell}$  values and their errors, both the systematic error, coming from the comparison of the subregions  $\alpha$  and  $\beta$ , and the statistical one, coming from the estimation of  $R_{\ell\ell}$ , are reported for different Higgs mass working points.

$m_{\rm H} \; [{\rm GeV}]$	$N_{\mu\mu}^{ m out,data}$	$N_{\mu\mu}^{\rm in,data}$	$R_{\mu\mu}$	$N_{ m ee}^{ m out,data}$	$N_{\rm ee}^{\rm in,data}$	$R_{ m ee}$
110	$2.5\pm1.3$	17	$0.153 \pm 0.072$	$1.25 \pm 1.20$	9	$0.148 \pm 0.128$
125	$3.7 \pm 2.4$	27	$0.143 \pm 0.086$	$2.4 \pm 1.2$	12	$0.211 \pm 0.088$
140	$3.5 \pm 1.3$	28	$0.128 \pm 0.04$	$3.1 \pm 1.5$	15	$0.211 \pm 0.088$
160	$1.5\pm1.7$	37	$0.044 \pm 0.051$	$2.9 \pm 1.5$	22	$0.147 \pm 0.072$
200	$2.1\pm2.8$	43	$0.055 \pm 0.073$	$3.5 \pm 1.9$	27	$0.147 \pm 0.072$
400	$1.4\pm1.4$	48	$0.033 \pm 0.031$	$1.9 \pm 2.3$	29	$0.074 \pm 0.087$
600	$0.96\pm0.69$	48	$0.022 \pm 0.016$	$1.3 \pm 1.1$	29	$0.049 \pm 0.041$

Table 7.6: Estimation of the Drell-Yan background at the Higgs selection level, for various Higgs masses. The results are reported separately for the *ee* and  $\mu\mu$  final state. The ratio  $R^{\ell\ell}$ , its error and the number of events under the Z peak (statistical error) are reported.



Figure 7.8: The invariant mass of the di-lepton system  $m_{\ell\ell}$  at Higgs level. The "in" region (under the Z peak, without tight  $p_{\rm T}$  selections) and the "out" regions (where the signal is expected) are shown.

## 7.3 Systematics

The signal efficiency is estimated using simulations. All Higgs production mechanisms are considered: the gluon fusion process, the associated production of the Higgs boson with a W or Z boson, and the VBF process, even if after tight VBF selections the last production mechanism is the dominant. Experimental effects, theoretical predictions, and the choice of Monte Carlo event generators are considered as sources of uncertainty and their impact on the signal efficiency is assessed. The experimental uncertainties on lepton efficiency, momentum scale and resolution,  $E_{\rm T}^{\rm miss}$  modeling, jet energy scale and resolution, and pile-up simulation are applied to the reconstructed objects in simulated events by smearing and scaling the relevant observables and propagating the effects to the kinematic variables used in the analysis.

The systematic uncertainties due to theoretical ambiguities are separated into two components, which are assumed to be independent. The first component is the uncertainty on the fraction of events categorized into the different jet categories and the effect of jet bin migration. The second component is the uncertainty on the lepton acceptance and the selection efficiency of all other requirements. The effect of variations in parton distribution functions and the value of  $\alpha_s$ , and the effect of higher-order corrections, are considered for both components. The uncertainty in the parton shower model and the underlying event are also considered by comparing different generators and it is about 30%. The uncertainties related to the diboson cross sections are calculated using the MCFM program [126].

The overall signal efficiency uncertainty is estimated to be about 20% and is dominated by the theoretical uncertainty due to missing higher-order corrections and PDF uncertainties. The total uncertainty on the background estimations in the  $H \rightarrow W^+W^-$  signal region is about 15%, which

is dominated by the statistical uncertainty on the observed number of events in the backgroundcontrol regions.

All systematic uncertainties taken into account in this analysis for MC based samples are summarized in Table 7.7.

Source	$H \to W^+W^-$	$\begin{array}{c} qq \rightarrow \\ W^+W^- \end{array}$	$gg \rightarrow W^+W^-$	non-Z resonant VV
Luminosity	4.5	4.5	4.5	4.5
Trigger efficiencies	1.5	1.5	1.5	1.5
Muon efficiency	1.5	1.5	1.5	1.5
Electron id efficiency	2.5	2.5	2.5	2.5
Momentum scale	1.5	1.5	1.5	1.5
$E_{\rm T}^{\rm miss}$ resolution	2.0	2.0	2.0	2.0
Jet counting	7-20	5.4	5.4	5.4
Higgs cross section	5 - 15			_
Monte Carlo statistics	5	5	5	4

Table 7.7: Summary of all systematic uncertainties (relative, in % units) for MC based samples.

## 7.4 Results

The standard VBF analysis is based on a simple cut-and-count analysis, that is counting the number of events that pass the selections from data and from MC and see if the data are more compatible with the background-only hypothesis or the signal plus background hypothesis.

The expected number of signal and background events from the data-driven methods (when available) after all selections for the same flavour (SF) and different flavour (DF) final states are shown respectively in Table 7.8 and in Table 7.9.

ta		~	_	1	1	~	_	0			~		~	_	_	0	_	~		~	~	01
da	1~								<u>ь</u> ,	<u>ь</u> ,								~				
signal	$0.26\pm0.03$	$0.50\pm0.05$	$0.96\pm0.10$	$1.47\pm0.16$	$2.33\pm0.23$	$3.32\pm0.35$	$3.68\pm0.38$	$5.34\pm0.56$	$4.77\pm0.49$	$7.42\pm0.72$	$7.82\pm0.78$	$7.50\pm0.72$	$6.76\pm0.65$	$4.92 \pm 0.47$	$2.79\pm0.28$	$3.26\pm0.32$	$3.24\pm0.24$	$2.81\pm0.21$	$1.99\pm0.15$	$1.64\pm0.12$	$1.33\pm0.09$	$1.16\pm0.08$
sum	$5.14\pm2.36$	$5.14\pm2.36$	$7.31\pm3.58$	$8.12\pm4.11$	$8.61\pm3.44$	$8.80\pm2.96$	$8.04\pm2.85$	$10.41\pm3.18$	$7.14\pm6.76$	$7.40\pm5.58$	$7.18\pm4.49$	$9.03\pm5.07$	$10.47\pm8.59$	$10.10\pm7.87$	$9.43\pm7.91$	$9.96\pm4.36$	$12.88\pm4.74$	$14.01\pm5.45$	$10.66\pm3.93$	$6.87\pm2.64$	$6.50\pm2.55$	$4.76\pm2.26$
MM	$0.36\pm0.15$	$0.36\pm0.15$	$0.52\pm0.19$	$0.60\pm0.20$	$0.60 \pm 0.20$	$0.68\pm0.23$	$0.48\pm0.18$	$0.90\pm0.28$	$0.56\pm0.21$	$0.51\pm0.20$	$0.70\pm0.24$	$1.05\pm0.29$	$1.29\pm0.31$	$1.48\pm0.34$	$1.42\pm0.32$	$1.98\pm0.39$	$2.09\pm0.41$	$1.61\pm0.34$	$1.14\pm0.29$	$0.94 \pm 0.25$	$0.78 \pm 0.22$	$0.68\pm0.21$
ggWW	$0.01\pm0.01$	$0.01\pm0.01$	$0.01\pm0.01$	$0.01 \pm 0.01$	$0.01\pm0.01$	$0.01\pm0.01$	$0.03\pm0.02$	$0.06\pm0.03$	$0.03\pm0.02$	$0.06\pm0.03$	$0.05\pm0.03$	$0.12\pm0.05$	$0.15\pm0.05$	$0.15\pm0.05$	$0.15\pm0.06$	$0.12\pm0.05$	$0.08\pm0.04$	$0.08\pm0.05$	$0.05\pm0.04$	$0.05\pm0.04$	$0.04\pm0.04$	$0.00 \pm 0.00$
VV	$0.00\pm0.00$	$0.00\pm0.00$	$0.01\pm0.01$	$0.01\pm0.01$	$0.02\pm0.01$	$0.02\pm0.01$	$0.02\pm0.01$	$0.02\pm0.01$	$0.02\pm0.01$	$0.03\pm0.01$	$0.04\pm0.02$	$0.05\pm0.02$	$0.07\pm0.02$	$0.08\pm0.02$	$0.10\pm0.03$	$0.12\pm0.03$	$0.12\pm0.03$	$0.09\pm0.02$	$0.07\pm0.02$	$0.07\pm0.02$	$0.04\pm0.02$	$0.04\pm0.02$
WJet	$0.05\pm0.02$	$0.05\pm0.02$	$0.46\pm0.21$	$0.41\pm0.18$	$0.41\pm0.18$	$0.44\pm0.20$	$0.43\pm0.20$	$0.40\pm0.18$	$0.00\pm0.00$	$0.00 \pm 0.00$	$0.00\pm0.00$	$0.00 \pm 0.00$	$0.40 \pm 0.19$	$0.38\pm0.17$	$0.35\pm0.16$	$0.27\pm0.12$	$0.25\pm0.12$	$0.26\pm0.12$	$0.07\pm0.04$	$0.07\pm0.04$	$0.07\pm0.04$	$0.08\pm0.05$
Top	$0.98\pm0.58$	$0.98\pm0.58$	$0.96\pm0.63$	$0.92\pm0.55$	$0.93\pm0.55$	$1.01\pm0.61$	$0.55\pm0.40$	$1.93\pm0.92$	$1.74\pm1.13$	$2.42 \pm 1.49$	$2.21 \pm 1.42$	$3.14\pm1.70$	$2.33 \pm 1.14$	$2.33 \pm 1.14$	$2.18\pm1.14$	$5.18\pm2.10$	$7.55\pm3.13$	$8.62\pm3.77$	$6.61\pm2.97$	$3.50\pm1.76$	$3.33 \pm 1.62$	$1.72\pm1.12$
$\mathrm{DY}\!\to\ell\ell$	$3.74\pm2.28$	$3.74\pm2.28$	$5.35\pm3.51$	$6.17\pm4.06$	$6.64 \pm 3.38$	$6.64 \pm 2.88$	$6.53 \pm 2.81$	$7.10\pm3.02$	$4.79\pm6.66$	$4.38\pm5.37$	$4.18\pm4.25$	$4.67\pm4.76$	$6.23\pm8.51$	$5.68\pm7.78$	$5.23\pm7.82$	$2.29\pm3.80$	$2.79\pm3.54$	$3.35\pm3.92$	$2.72\pm2.56$	$2.24 \pm 1.95$	$2.24 \pm 1.95$	$2.24 \pm 1.95$
Mass	110	115	120	125	130	135	140	145	150	160	170	180	190	200	250	300	350	400	450	500	550	600

Table 7.8: Background contributions and data yields for  $12.1 \text{ fb}^{-1}$  of integrated luminosity in the two-jet-bin for the same flavor final states. The data-driven corrections are applied.

### 7.4. RESULTS

نے																						
date	2	2	2	2	4	с,	2	2	с С	4	4	5	9	~	10	15	14	2	3	2	3	ۍ ا
signai	$0.53\pm0.06$	$0.83\pm0.08$	$1.65\pm0.15$	$2.79\pm0.24$	$4.35\pm0.44$	$5.15\pm0.51$	$5.79\pm0.58$	$7.55\pm0.61$	$8.23\pm0.80$	$11.76\pm1.14$	$12.01\pm1.21$	$11.55\pm1.10$	$10.52\pm1.06$	$9.28\pm0.66$	$6.63\pm0.62$	$5.78\pm0.54$	$4.93\pm0.35$	$3.84\pm0.28$	$2.73\pm0.20$	$2.16\pm0.15$	$1.71\pm0.12$	$1.37\pm0.09$
IIINS	$2.51\pm0.88$	$2.51\pm0.88$	$2.37\pm0.77$	$2.84\pm0.86$	$3.17\pm0.89$	$3.17\pm0.89$	$2.45\pm0.68$	$4.82\pm1.29$	$4.25\pm1.78$	$5.01\pm2.20$	$5.76\pm2.67$	$8.22\pm3.32$	$6.82 \pm 1.92$	$7.22\pm1.93$	$9.58\pm2.31$	$18.49\pm4.63$	$19.75\pm6.14$	$18.66\pm 6.35$	$13.22\pm4.57$	$8.12\pm2.50$	$7.02\pm2.28$	$4.91\pm1.91$
VV VV	$0.71\pm0.26$	$0.71\pm0.26$	$0.87\pm0.28$	$1.05\pm0.31$	$1.32\pm0.35$	$1.32\pm0.35$	$1.22\pm0.33$	$1.70\pm0.40$	$1.12\pm0.31$	$1.13\pm0.30$	$1.14\pm0.29$	$1.37\pm0.33$	$2.09\pm0.41$	$2.46\pm0.45$	$4.03\pm0.60$	$4.49\pm0.64$	$3.98\pm0.61$	$3.65\pm0.57$	$2.72\pm0.50$	$2.76\pm0.50$	$2.01\pm0.42$	$1.63\pm0.37$
gg W W	$0.01\pm0.01$	$0.01 \pm 0.01$	$0.01\pm0.01$	$0.01 \pm 0.01$	$0.05\pm0.04$	$0.10\pm0.05$	$0.10\pm0.05$	$0.16\pm0.07$	$0.14\pm0.06$	$0.10\pm0.05$	$0.10\pm0.05$	$0.15\pm0.06$	$0.19\pm0.07$	$0.22\pm0.07$	$0.22\pm0.07$	$0.24\pm0.07$	$0.16\pm 0.06$	$0.17\pm0.06$	$0.12\pm0.05$	$0.10\pm0.05$	$0.08\pm0.04$	$0.08\pm0.04$
۷ ۷ ۱	$0.05\pm0.02$	$0.05\pm0.02$	$0.06\pm0.02$	$0.08\pm0.02$	$0.09\pm0.02$	$0.09\pm0.02$	$0.09\pm0.02$	$0.09\pm0.02$	$0.04\pm0.01$	$0.04\pm0.02$	$0.04\pm0.02$	$0.06\pm0.02$	$0.06\pm0.02$	$0.08\pm0.02$	$0.10\pm0.03$	$0.11\pm0.03$	$0.09\pm0.02$	$0.10\pm0.03$	$0.05\pm0.02$	$0.04\pm0.02$	$0.03\pm0.01$	$0.02\pm0.01$
VV JET	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.00\pm0.00$	$0.15\pm0.08$	$0.25\pm0.12$	$0.23\pm0.11$	$0.21\pm0.10$	$0.62\pm0.26$	$0.80\pm0.33$	$0.16\pm0.07$	$0.09\pm0.04$	$0.00\pm0.00$	$0.14\pm0.09$	$0.13\pm0.07$	$0.26\pm0.14$
Top	$1.42\pm0.83$	$1.42\pm0.83$	$1.11\pm0.70$	$1.32\pm0.78$	$1.36\pm0.80$	$1.35\pm0.80$	$0.80\pm0.57$	$2.63 \pm 1.22$	$2.69\pm1.74$	$3.57\pm2.18$	$4.16\pm2.65$	$6.18\pm3.30$	$3.83 \pm 1.86$	$3.83 \pm 1.86$	$4.31\pm2.21$	$12.64 \pm 4.57$	$15.19\pm6.11$	$14.53\pm6.32$	$10.21 \pm 4.54$	$4.96\pm2.44$	$4.71\pm2.24$	$2.86 \pm 1.86$
$DY \rightarrow \ell\ell$	$0.32\pm0.15$	$0.32\pm0.15$	$0.32\pm0.15$	$0.38\pm0.17$	$0.35\pm0.16$	$0.31\pm0.15$	$0.24\pm0.13$	$0.24\pm0.13$	$0.26\pm0.14$	$0.17\pm0.11$	$0.17\pm0.11$	$0.21\pm0.12$	$0.42\pm0.19$	$0.42\pm0.19$	$0.30\pm0.15$	$0.21\pm0.12$	$0.17\pm0.11$	$0.12\pm0.09$	$0.12\pm0.09$	$0.12\pm0.09$	$0.06\pm0.06$	$0.06\pm0.06$
Mass	110	115	120	125	130	135	140	145	150	160	170	180	190	200	250	300	350	400	450	500	550	600

пnа 5 NPTT Ĩ -jetnwo Ξ Shree littegraueu Б 01 Background contributions and data yields for 12.1 states. The data-driven corrections are applied. . n. table

Upper limits for VBF Higgs production are derived on the ratio of the product of the Higgs boson production cross section and the  $H \rightarrow W^+W^-$  branching fraction,  $\sigma_H \times BR(H \rightarrow W^+W^-)$ , and the SM Higgs expectation,  $\sigma/\sigma_{SM}$ .

In Fig. 7.9 the exclusion limit with 7 TeV data are shown[127]. With the whole data collected during the 7 TeV run, the VBF channel is sensitive to a Higgs boson with a mass between 150 and 190 GeV.



Figure 7.9: Expected and observed 95% CL upper limits on the cross section times branching fraction,  $\sigma_{\rm H} \times {\rm BR}({\rm H} \to {\rm W}^+{\rm W}^-)$ , relative to the SM Higgs expectation, using the 7 TeV data only. Results are obtained using the CL<sub>s</sub> approach.

In Fig. 7.10 the exclusion limit with 8 TeV data are splitted into same flavour final state and different flavour one. As expected, the different flavor final state is more powerful with respect to the same flavour one, since there is no  $Z/\gamma^* \rightarrow \ell^+ \ell^-$  contamination. The results are summarized in Table 7.10.



Figure 7.10: Expected and observed 95% CL upper limits on the cross section times branching fraction,  $\sigma_{\rm H} \times {\rm BR}({\rm H} \to {\rm W}^+{\rm W}^-)$ , relative to the SM Higgs expectation, using the 8 TeV data only in the same flavour final state (a) and in the different flavour final state (b), VBF analysis. Results are obtained using the CL<sub>s</sub> approach.

The 8 TeV analysis excludes the presence of a Higgs boson with mass in the range 135–200 GeV at 95% CL. The observed (expected) upper limits are about 1.56 (1.63) times the SM expectation

			8 TeV	7 analysis		
TT: M	Observed	Expected	6007	05.07	Observed	Expected
Higgs Mass	$\operatorname{Limit}$	$\operatorname{Limit}$	08% range	95% range	significance	significance
110	8.23	8.09	[5.84, 11.24]	[4.39, 14.94]	0.00	0.54
115	5.18	4.98	[3.60,  6.92]	[2.70, 9.20]	0.00	0.71
120	2.56	2.59	[1.87,  3.59]	[1.40,  4.77]	0.00	1.06
125	1.56	1.63	[1.18, 2.27]	[0.89,  3.01]	0.00	1.47
130	1.44	1.14	[0.83,  1.59]	[0.62, 2.11]	0.62	1.97
135	0.87	0.90	[0.65, 1.25]	[0.49,  1.67]	0.00	2.27
140	0.70	0.74	[0.54, 1.03]	[0.40,  1.37]	0.00	2.65
145	0.48	0.68	[0.49, 0.95]	[0.37, 1.26]	0.00	2.60
150	0.51	0.63	[0.45, 0.87]	[0.34, 1.16]	0.00	2.40
155	0.44	0.51	[0.37,  0.71]	[0.28,  0.94]	0.00	2.90
160	0.39	0.47	[0.34,  0.65]	[0.26,  0.87]	0.00	2.71
170	0.42	0.49	[0.35,  0.68]	[0.26,  0.90]	0.00	2.56
180	0.40	0.55	[0.40,0.77]	[0.30,  1.02]	0.00	2.31
190	0.51	0.59	[0.43,  0.82]	[0.32,  1.09]	0.00	2.75
200	0.83	0.76	[0.55,  1.05]	[0.41,  1.39]	0.35	2.40
250	1.34	1.25	[0.90,1.74]	[0.68, 2.31]	0.23	1.57
300	1.60	1.71	[1.23, 2.38]	[0.93,  3.16]	0.00	1.14
350	1.54	1.98	[1.43, 2.75]	[1.07,  3.65]	0.00	0.88
400	1.20	2.10	[1.52, 2.92]	[1.14, 3.88]	0.00	0.71
450	1.09	2.27	[1.64, 3.16]	[1.23, 4.20]	0.00	0.68
500	1.24	2.35	[1.70, 3.27]	[1.28, 4.34]	0.00	0.82
550	1.74	2.82	[2.04,  3.92]	[1.53, 5.20]	0.00	0.76
600	2.14	3.05	[2.20, 4.24]	[1.66,  5.64]	0.00	0.78

Table 7.10: Expected and observed 95% CL upper limits on the cross section times branching fraction,  $\sigma_{\rm H} \times {\rm BR}({\rm H} \to {\rm W}^+ {\rm W}^-)$ , relative to the SM Higgs expectation, using the 8 TeV data only. Results are obtained using the CL<sub>s</sub> approach. The 68% and 95% ranges are also given. The observed and expected discovery significance is reported.

for  $m_{\rm H} = 125$  GeV. Combining the 7 TeV and the 8 TeV analysis the excluded limit is extended up to 250 GeV as shown in Fig. 7.11.

No excess is observed for low Higgs boson masses, but this channel is still limited by the statistics, and has not reached yet the sensitivity to the SM Higgs boson at 125 GeV.



Figure 7.11: Expected and observed 95% CL upper limits on the cross section times branching fraction,  $\sigma_{\rm H} \times {\rm BR}({\rm H} \to {\rm W}^+{\rm W}^-)$ , relative to the SM Higgs expectation, using the 8 TeV data only (a,b) and the combined 7 TeV and 8 TeV data (c,d), VBF analysis. Results are obtained using the CL<sub>s</sub> approach.

# CHAPTER 8

## THE VH ANALYSIS

Est modus in rebus.

Orazio, Satire

The VH analysis probes the third most probable production mechanism of the Standard Model Higgs boson at LHC. The final state considered is characterized by the decay of the Higgs boson into two W bosons, that subsequently decay into leptons, while the vector boson that radiates the Higgs decays into quarks. In Fig. 8.1 the Feynman diagram of the Higgs production mechanism and its decay are looked for in this analysis is shown. The final state has the same objects of the VBF analysis, but with two jets with an invariant mass next to W/Z ones.



Figure 8.1: Feynman diagram for Higgs production via Higgs-Strahlung process and decaying into  $W^+W^- \rightarrow \ell^+ \nu \ell^- \bar{\nu}$  with the vector boson (W/Z), that radiates the Higgs, decaying into two quarks.

The search is performed at the center of mass energy of 7 TeV, in the Higgs mass hypothesis range of 120-190  $\text{GeV/c}^2$ . The data sample corresponds to an integrated luminosity of 4.9 fb<sup>-1</sup> collected during the 2011 run [112, 113] and the expected number of events in 1fb<sup>-1</sup> are shown in Fig. 8.2.



Figure 8.2: The expected number of events in the VH, H  $\rightarrow$  WW  $\rightarrow \ell \nu \ell \nu$  and V $\rightarrow jj$ production and decay mode in 1fb<sup>-1</sup> at 7 (a) and 8 (b) TeV of center of mass energy. The low production cross section (see Fig. 6.4) is mitigated by the high branching ratio of the vector boson (W/Z) into jets (67.6% for W and 69.9% for Z [3]).

In the following the selections applied in the analysis are summarized in Sec. 8.1, while the specific aspects of background estimation are revisited in Sec. 8.2. The systematics are summarized in Sec. 8.3 and the results are reported in Sec. 8.4.

### 8.1 VH selections

In order to enhance VH Higgs events with respect to backgrounds a set of specific selections have been optimized. On top of WW level (see Sec. 6.3), summarized in Table 7.1, an additional set of selections, partly Higgs boson mass dependent, have been applied, as summarized in the following:

- The events must have only two jets ( $p_T > 30$  GeV) and they must be central ( $|\eta| < 2.5$ ).
- The TCHE b-tag of the two jets must be less than 1.6: a tighter selection with respect to the VBF analysis is required in order to reduce the top background.
- The invariant mass of the jets  $m_{jj}$  is required to be in the range [60 110] GeV to select the jets from the hadronic decay of the initial state vector boson (W/Z). Figure 8.3 shows the distribution of the  $m_{jj}$  at WW level. With this selection there is no intersection with the VBF analysis, where  $m_{jj}$  is required to be > 500 GeV.
- The  $\eta$  distance between the two jets is required to be  $\Delta \eta_{jj} < 2.1$ . This selection reflects the boost of the W/Z boson involved in the Higgs production. Figure 8.4 shows the distribution of the  $\Delta \eta_{jj}$  at WW level. In addition, with this selection there is no intersection with the VBF analysis, where  $\Delta \eta_{jj}$  is required to be > 3.5.
- $\Delta \phi_{\ell\ell,jj}$  selection: for same flavor events, the azimuthal angle  $\phi$  between the di-jet system and the di-lepton system is required to be less than 165 degrees in order to reduce the



Figure 8.3: The invariant mass of the jets  $m_{jj}$  at WW level normalized to luminosity (left) and to 1 (right), for shape comparison. Backgrounds are normalized with data driven techniques and signals are plotted one on top of the other. The error band on the expected rates reflects the Monte Carlo statistics and the systematic error coming from data driven estimation. In the analysis  $m_{jj}$  is required to be in the range [60 - 110]GeV.



Figure 8.4: The pseudorapidity separation of the jets  $\Delta \eta_{jj}$  at WW level normalized to luminosity (left) and to 1 (right), for shape comparison. Backgrounds are normalized with data driven techniques and signals are plotted one on top of the other. The error band on the expected rates reflects the Monte Carlo statistics and the systematic error coming from data driven estimation. In the analysis  $\Delta \eta_{jj}$  is required to be less than 2.1.

Drell-Yan contribution. In addition  $\Delta \phi_{\ell\ell,jj}$  is required to be larger than 75 degrees.

• The invariant mass of the leptons  $m_{\ell\ell}$ : to reject top contamination  $m_{\ell\ell}$  is required to be less than 70 GeV. Figure 8.5 shows the distribution of the  $m_{\ell\ell}$  at WW level. The search for VH production is only a low Higgs mass analysis, then no mass dependent selections are applied for  $m_{\ell\ell}$ .



Figure 8.5: The invariant mass of the di-lepton system  $m_{ll}$  at WW level normalized to luminosity (left) and to 1 (right), for shape comparison. Backgrounds are normalized with data driven techniques and signals are plotted one on top of the other. The error band on the expected rates reflects the Monte Carlo statistics and the systematic error coming from data driven estimation. In the analysis  $m_{ll}$  is required to be less than 70GeV.

- The  $\Delta R_{\ell\ell}$  distance in the  $\eta \times \phi$  plane ( $\Delta R_{\ell\ell} = \sqrt{\Delta \phi_{\ell\ell}^2 + \Delta \eta_{\ell\ell}^2}$ ) between the two leptons is required to be  $\Delta R_{\ell\ell} < 1.3$ . Figure 8.6 shows the distribution of  $\Delta R_{\ell\ell}$  at WW level.
- The transverse mass  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$ , defined in Eq. 7.1, is required to be in the window 50 GeV  $< m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}} < m_{Higgs}$ . Fig. 8.7 shows the  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  distribution for different Higgs mass hypotheses.
- $min projE_{\rm T}^{\rm miss}$  preselection:  $min projE_{\rm T}^{\rm miss} > 20$ GeV for e $\mu$  and  $\mu$ e events and  $min projE_{\rm T}^{\rm miss} > 37 + N_{vtx}/2$ GeV for ee and  $\mu\mu$  events. The cut on  $min projE_{\rm T}^{\rm miss}$  instead of PF- $E_{\rm T}^{\rm miss}$  is the optimized selection for all the analyses during 2011 run at 7 TeV. The cut depends on the number of vertexes,  $N_{vtx}$ , in order to have a constant efficiency with respect to the number of vertexes. This effect is due to the fact that the  $E_{\rm T}^{\rm miss}$  is expected to increase with higher instantaneous luminosity, that is linearly correlated with the number of vertexes.

The application of ortogonal selections with respect to VBF analysis allows for a combination of the two analyses since they are probing disjoint phase spaces.



Figure 8.6: The  $\Delta R_{\ell\ell}$  distance in the  $\eta \times \phi$  plane between the two leptons at WW level normalized to luminosity (left) and to 1 (right), for shape comparison. Backgrounds are normalized with data driven techniques and signals are plotted one on top of the other. The error band on the expected rates reflects the Monte Carlo statistics and the systematic error coming from data driven estimation. In the analysis  $\Delta R_{\ell\ell}$  is required to be less than 1.3.



Figure 8.7: The  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  distribution at WW level for different Higgs mass hypotheses. All the distributions are normalized to unity.

### 8.2 Background estimation

The main background in the analysis is Top. The list backgrounds with a data-driven estimation is:

- Top (Sec. 6.4.2)
- $Z/\gamma^* \rightarrow \ell^+ \ell^-$  (Sec. 6.4.3)
- $Z/\gamma^* \to \tau^+ \tau^-$  (Sec. 6.4.4)
- W+jets (Sec. 6.4.1)

The details of the methods are described in the respective Sections reported. In the following the results of the data-driven estimation for top and  $Z/\gamma^* \to \ell^+ \ell^-$  and the differences with respect to what is described before are reported. The estimated number of  $Z/\gamma^* \to \tau^+ \tau^-$  and W+jets events are listed in the summary tables.

### 8.2.1 The top background

With respect to the VBF case, in the VH analysis both jets are in the acceptance region of the tracker, then both the jets have a reliable TCHE b-tag value. No extrapolation in the high  $\eta$  region is needed, and, in addition, a cross check of the top estimation is possible, comparing the results using the most central jet and the most forward one.

The b-tag efficiency is measured at WW level as described in Sec. 6.4.2. In Table 8.1 the efficiencies measured on data are shown, as well as the efficiencies on a top MC sample at WW level (control region, CR) and after all selections are applied (signal region, SR). The values are different with respect to the VBF analysis at 7 TeV since the TCHE threshold considered is different. The hypothesis of the data-driven estimation is that the b-tag efficiency does not change after selections are applied: as shown in Table 8.1, within the Monte Carlo statistics this assumption is confirmed.

$ \eta^{cj} $ bin	$\varepsilon^{\text{DATA}}( \eta^{cj} )$	$\varepsilon^{\mathrm{MC,CR}}( \eta^{cj} )$	$\varepsilon^{\mathrm{MC,SR}}( \eta^{cj} )$
$0 <  \eta  < 0.5$	$0.77\pm0.03$	$0.79 \pm 0.01$	$0.79\pm0.05$
$0.5 <  \eta  < 1$	$0.77\pm0.03$	$0.78\pm0.01$	$0.81\pm0.08$
$1 <  \eta  < 1.5$	$0.65\pm0.05$	$0.75\pm0.01$	$0.78\pm0.08$
$1.5 <  \eta  < 2.5$	$0.66\pm0.09$	$0.64 \pm 0.02$	$0.68\pm0.10$

Table 8.1: The b-tagging efficiency measured in a control region, according to Equation 6.12, in bins of  $|\eta|$  of the most central jet.

The top background data-driven estimation is performed by means of the Eq. 8.1 using the most central jet.

$$N_{bveto}^{\text{data driven},|\eta|<2.5} = \int d\eta \ N_{bveto}^{\text{data driven}}(\eta)$$
$$= \int d\eta \ \left[ \left( N_{btag}^{DATA} - N_{btag}^{\text{other backgrounds}} \right)(\eta) \frac{1-\varepsilon(\eta)}{\varepsilon(\eta)} \right]$$
(8.1)

As a cross check, the same analysis is performed using the most forward jet, measuring the b-tag efficiency, and extrapolating the top events reversing the b-veto cut. Similar results are obtained, as summarized in 8.2, thus assuring the robustness of the technique.

$m_H$	top (central jet)	top (forward jet)
118	$5.4 \pm 2.0$	$5.7 \pm 2.4$
120	$5.4\pm2.0$	$5.8 \pm 2.4$
124	$7.0 \pm 2.3$	$7.4\pm2.8$
128	$8.2\pm2.5$	$7.8\pm3.0$
130	$8.2\pm2.5$	$8.2\pm3.0$
135	$8.1\pm2.5$	$8.9\pm3.3$
140	$8.8\pm2.5$	$11.0\pm3.5$
150	$10.4 \pm 2.8$	$14.1\pm4.3$
160	$11.1\pm2.9$	$15.4\pm4.3$
170	$11.7\pm3.0$	$15.2\pm4.3$
180	$11.6\pm3.0$	$15.3\pm4.3$
190	$12.2 \pm 3.1$	$16.0 \pm 4.5$

Table 8.2: Estimate of t and  $t\bar{t}$  contamination in the signal region, according to Equation 8.1 for the various working points of the analysis, using central jet and forward jet to define b-tagged and b-vetoed region. The various mass points are highly correlated: the top estimation using central jets or forward jets may be sistematically lower (bigger), but this is not a symptom of bad behaviour of the method.

In Table 8.3 the various steps of the top estimation are reported: the extrapolated number of top events, the number of b-tagged events, the error coming from the measurement of b-tag efficiency, the error coming form the variation of the efficiency at WW level and at Higgs level and the error coming from the subtraction of non-top contribution in the btagged region.

#### 8.2.2 The $Z/\gamma^* \rightarrow \ell^+ \ell^-$ background

The  $Z/\gamma^* \to \ell^+ \ell^-$  data-driven estimation is described in Sec. 6.4.3. In order to enrich the "in" region the  $\Delta R_{\ell\ell}$  selections for the leptons are not applied, since low  $\Delta R_{\ell\ell}$  means low  $m_{\ell\ell}$  then it would remove most of the events under the Z-peak. In addition, in order to have statistical lever, the number of events in the "in" region must be bigger than the number of events in the "out" region. The "in" and "out" regions are then defined as:

- in; under Z peak ( $m_{\ell\ell} \in m_Z \pm 15 \text{ GeV}$ ) and without  $\Delta R_{\ell\ell}$  selection
- out; all cuts defined in the analysis applied

		ath taggod		110	
$m_H$	top estimation	N <sup>o-laggea</sup>	$\varepsilon$ error	$\varepsilon$ MC error	MC Subtraction error
118	$5.4 \pm 2.0$	20	18 %	$11 \ \%$	15 %
120	$5.4 \pm 2.0$	20	18 %	$11 \ \%$	15 %
124	$7.0\pm2.3$	26	18 %	$11 \ \%$	12 %
128	$8.2\pm2.5$	29	18 %	8.3~%	10 %
130	$8.2\pm2.5$	29	18 %	7.4~%	10 %
135	$8.1\pm2.5$	29	18 %	$8.5 \ \%$	10 %
140	$8.8\pm2.5$	30	$17 \ \%$	3.2~%	$9.7 \ \%$
150	$10 \pm 3$	34	16 %	2.1~%	8.6 %
160	$11 \pm 3$	35	16 %	4 %	8.1 %
170	$12 \pm 3$	37	17 %	0.37~%	7.8 %
180	$12 \pm 3$	36	16 %	2.1~%	7.9 %
190	$12 \pm 3$	38	17 %	0.48~%	7.5 %

Table 8.3: For each Higgs mass hypothesis the extrapolated number of top events, the number of b-tagged events, the error coming from the measurement of b-tag efficiency, the error coming form the variation of the efficiency in control region and in signal region, and the error coming from the uncertainty related to the subtraction of other backgrounds in the btagged region are reported.

To summarize, the amount of Drell-Yan background at Higgs level is estimated by means of  $R_{\ell\ell}$ , measured inverting the  $min-projE_{\rm T}^{\rm miss}$  selection (20 GeV  $< min-projE_{\rm T}^{\rm miss} < 37 + N^{vtx}/2$ GeV), as described in Eq. 8.2:

$$N_{\ell\ell,DY}^{\text{out,data}} = \left(N_{\ell\ell}^{\text{in,data}} - \frac{k^{\ell\ell,\text{in}}}{2} \cdot N_{e\mu}^{\text{in,data}} - N_{ZV}^{\text{in,MC}}\right) \cdot R_{\ell\ell}$$
(8.2)

In addition to the statistical uncertainty on the number of DY events extrapolated from data, a systematic uncertainty is added, due to the hypothesis that the ratio between the number of DY events under Z peak and outside Z peak is the same in the low  $E_{\rm T}^{\rm miss}$  region (CR) and in the high  $E_{\rm T}^{\rm miss}$  one (SR). To test this assumption, for each mass point, the low  $E_{\rm T}^{\rm miss}$  region has been splitted into two subregions, namely  $\alpha$  and  $\beta$ , defined in Eq. 8.3.

$$\alpha \text{ zone : } 20 < \min - proj E_{\mathrm{T}}^{\mathrm{miss}} < 20 + \frac{1}{2} \cdot (17 + N_{vtx}/2) \text{ GeV}$$

$$\beta \text{ zone : } 20 + \frac{1}{2} \cdot (17 + N_{vtx}/2) < \min - proj E_{\mathrm{T}}^{\mathrm{miss}} < (37 + N_{vtx}/2) \text{ GeV}$$

$$(8.3)$$

The difference in the ratio  $R_{\ell\ell}$  between  $\alpha$  and  $\beta$  regions is considered as systematic uncertainty. The  $R_{\ell\ell}$  values measured for different Higgs mass working points are reported in Table 8.4. Eventually, the estimated contamination of Drell-Yan contribution at the Higgs level is reported in Table 8.5.

$m_{\rm H} \ [\text{GeV}]$	$R_{\mu\mu} \pm \text{stat} \pm \text{syst}$	$R_{\rm ee} \pm { m stat} \pm { m syst}$
118	$0.116 \pm 0.022 \pm 0.056$	$0.077 \pm 0.022 \pm 0.005$
120	$0.115 \pm 0.022 \pm 0.053$	$0.076 \pm 0.021 \pm 0.006$
124	$0.114 \pm 0.022 \pm 0.049$	$0.076 \pm 0.021 \pm 0.008$
128	$0.114 \pm 0.022 \pm 0.047$	$0.075 \pm 0.021 \pm 0.011$
130	$0.114 \pm 0.022 \pm 0.047$	$0.074 \pm 0.021 \pm 0.014$
135	$0.113 \pm 0.022 \pm 0.044$	$0.073 \pm 0.021 \pm 0.015$
140	$0.113 \pm 0.022 \pm 0.042$	$0.072 \pm 0.020 \pm 0.017$
150	$0.112 \pm 0.022 \pm 0.040$	$0.072 \pm 0.020 \pm 0.020$
160	$0.112 \pm 0.021 \pm 0.038$	$0.071 \pm 0.020 \pm 0.021$
170	$0.111 \pm 0.021 \pm 0.034$	$0.071 \pm 0.020 \pm 0.021$
180	$0.115 \pm 0.023 \pm 0.033$	$0.069 \pm 0.020 \pm 0.017$
190	$0.115 \pm 0.023 \pm 0.033$	$0.069 \pm 0.020 \pm 0.017$

Table 8.4: The  $R_{\ell\ell}$  values and their errors, both the systematic error, coming from the comparison of the subregions  $\alpha$  and  $\beta$ , and the statistical one, coming from the estimation of  $R_{\ell\ell}$  in CR, are reported for different Higgs mass working points.

$m_{\rm H} \ [{\rm GeV}]$	$N_{\mu\mu}^{ m out,data}$	$N_{\mu\mu}^{\rm in,data}$	$R_{\mu\mu}$	$N_{ m ee}^{ m out,data}$	$N_{ m ee}^{ m in,data}$	$R_{ m ee}$
118	$0.46\pm0.31$	5	$0.116 \pm 0.060$	$0.72 \pm 0.31$	10	$0.077 \pm 0.022$
120	$0.46\pm0.31$	5	$0.115 \pm 0.058$	$0.79 \pm 0.33$	11	$0.076 \pm 0.022$
124	$0.79\pm0.46$	8	$0.114 \pm 0.054$	$0.78 \pm 0.33$	11	$0.076 \pm 0.023$
128	$1.00\pm0.55$	10	$0.114 \pm 0.051$	$0.77 \pm 0.34$	11	$0.075 \pm 0.024$
130	$1.34\pm0.71$	13	$0.114 \pm 0.051$	$0.83 \pm 0.37$	12	$0.075 \pm 0.025$
135	$1.56\pm0.79$	15	$0.113 \pm 0.049$	$0.89\pm0.39$	13	$0.073 \pm 0.025$
140	$1.47 \pm 0.72$	15	$0.113 \pm 0.047$	$0.85\pm0.39$	13	$0.072 \pm 0.027$
150	$1.79\pm0.83$	18	$0.112 \pm 0.045$	$0.90 \pm 0.43$	14	$0.072 \pm 0.028$
160	$1.88\pm0.85$	19	$0.112 \pm 0.044$	$0.97 \pm 0.47$	15	$0.071 \pm 0.029$
170	$1.96\pm0.83$	20	$0.111 \pm 0.040$	$1.03 \pm 0.50$	16	$0.071 \pm 0.029$
180	$2.02\pm0.84$	20	$0.115 \pm 0.040$	$1.07 \pm 0.48$	17	$0.069 \pm 0.026$
190	$2.13\pm0.88$	21	$0.115 \pm 0.040$	$1.13 \pm 0.51$	18	$0.069 \pm 0.026$

Table 8.5: Estimation of the Drell-Yan background at the Higgs selection level, for various Higgs masses. The results are reported separately for the *ee* and  $\mu\mu$  final state. The ratio  $R^{\ell\ell}$ , its error and the number of events under the Z peak (statistical error) are reported.

### 8.2.3 The WWV background

Standard Model triple gauge boson production WWV has been considered as a background. The number of WWW for the different working points of the analysis in 4.9 fb<sup>-1</sup> is shown in Table 8.6. As expected these contributions can be considered negligible for the analysis, since the cross section of WWW decaying into a pair of leptons and neutrinos and two jets (WWW  $\rightarrow \ell\nu\ell\nu jj$ ) is about 0.0039 pb (calculated with Madgraph5 [95]). Also the WWZ contribution that has a similar final state, W  $\rightarrow \ell\nu$  and Z  $\rightarrow jj$ , has a cross section of 0.0022 pb and is therefore negligible.

Mass	WWW events
120	$0.00197 \pm 0.00079$
130	$0.00288\pm0.00095$
140	$0.0038\pm0.0011$
150	$0.0041 \pm 0.0011$
160	$0.0044\pm0.0012$
170	$0.0048\pm0.0012$
180	$0.0048\pm0.0012$
190	$0.0050\pm0.0013$

Table 8.6: Expected contamination of WWW for the different working points of the analysis with 4.9 fb<sup>-1</sup>. As expected their contribution can be considered negligible.

## 8.3 Systematics

The following uncertainties have been considered for simulated backgrounds and signal samples:

- Luminosity. Based on the CMS online luminosity monitoring the uncertainty is currently 2.2% [78].
- Trigger efficiency. This uncertainty for both electrons and muons is less than 1%.
- Lepton reconstruction and identification efficiencies. The uncertainty on the lepton reconstruction and identification efficiency is at the order of 2%.
- Muon momentum and electron energy scale. The electron energy is varied by 2% in the barrel and 5% in the endcap. The systematic uncertainty is about 2% per electron. For the muons the uncertainty is much smaller, and a conservative 0.5% is considered.
- Jet energy scale. It affects both the jet multiplicity the jet kinematics. This error is estimated applying variations of the official jet uncertainties on the JES (which depend on  $\eta$  and  $p_{\rm T}$  of the jet) and computing the variation of the selection efficiency. It turns out to be less than 7%.
- **b-mistag modelling.** The uncertainties on the selection of not-b jets (TCHE cut) is taken into account looking at efficiency of b-vetoing for events in a DY enriched phase space. The ratio between the efficiency measured in data and in Monte Carlo is a scale factor related to the b-mistag modelling, to be applied to all Monte Carlo samples. The scale factor is

found to be 0.97 with an uncertainty of 2%.

- Higgs cross-section. The uncertainties on the inclusive cross-section for the Higgs sample have been taken from the LHC Higgs Cross Section working group [12] and are about 20% for the gluon fusion contribution, about 2% for the VBF one and about 5% for the associated Higgs production.
- **PDFs uncertainties**. They have been estimated according to the recipe provided by the LHC Higgs Cross Section working group [12]. Different sets of Parton Density Functions (PDFs) have been tested which change the acceptance of the measurement. The effect on the selection efficiency is 1%, 2% and 1% for the signal, respectively associated production, gluon fusion and VBF.
- QCD scale uncertainties. They have been estimated according to the recipe provided by the LHC Higgs Cross Section working group [12], varying the normalization and factorization scales in the production of MC events. The effect on the selection efficiency is between 10% and 65%, depending on the MC considered.
- **Pile-up.** The simulation has been reweighted according to the data instantaneous luminosity. An uncertainty of 8% in the knowledge of the number of interactions was propagated to the pile-up re-weighting procedure. The obtained variation in the expected events is of about 2%.
- Monte Carlo statistics. It contributes as an uncertainty of about 15% to the signal efficiencies and about 20% for most of backgrounds.

systematics name	VH	ggH	WW	WZ/ZZ
lepton efficiency	3%	3%	3%	3%
electron scale	2%	5%	8%	1%
${ m muon \ scale}$	2%	1%	1%	1%
$_{ m JES}$	3%	2%	6%	4%
B-mistag modelling	2%	2%	2%	2%
luminosity	2.2%	2.2%	2.2%	2.2%
Pile up	3%	4%	2%	6%
$\operatorname{pdf}$	2%	10%	4%	4%
QCD scale	7%	65%	18%	3%
Monte Carlo statistics	11%	11%	9%	12%

The systematics are summarized in the Table 8.7.

Table 8.7: Systematics for signal and main Monte Carlo based samples.

### 8.4 Results

The full analysis is performed on  $4.9 \text{ fb}^{-1}$  of integrated luminosity. The expected number of signal and background events from the data-driven methods after all selections are shown in Table 8.8 for different Higgs mass hypotheses. The expected number of signal and background events from the data-driven methods after all selections for the same flavour and different flavour final states are shown respectively in Table 8.9 and in Table 8.10. The limit computation has been performed both dividing the same flavour and different flavour final state, and merging

them. Similar results are obtained. In the following the exclusion plot with the same and different flavour final states treated as separate channels and then combined is shown. The  $m_T$  distribution is shown in Fig. 8.8 at Higgs level without the  $m_T^{\ell\ell E_T^{\text{miss}}}$  cut itself, while Fig. 8.9 shows the invariant mass of the two jets pair, where the resonance from the hadronic W and Z decay for the signal VH process can be seen.

		$m_H$	data	to	ot bkg	tot s	sig		VH		qqH	ggH	
		118	6	10.	$6 \pm 2.2$	$0.38 \pm$	0.12	0.18	$\pm 0.03$	0.01	$5 \pm 0.002$	$0.18 \pm 0.12$	
		120	6	10.	$8 \pm 2.2$	$0.51 \pm$	0.17	0.24	$\pm 0.04$	0.02	$3 \pm 0.004$	$0.25 \pm 0.16$	
		124	6	12.	$9 \pm 2.5$	$0.68 \pm$	0.24	0.30	$\pm 0.05$	0.02	$7 \pm 0.004$	$0.35 \pm 0.23$	
		128	8	14.	$6 \pm 2.7$	$0.98~\pm$	0.37	0.39	$\pm 0.06$	0.04	$9 \pm 0.009$	$0.55 \pm 0.37$	
		130	8	15.	$2 \pm 2.8$	$1.3 \pm$	0.5	0.54	$\pm 0.08$	0.05	$1 \pm 0.009$	$0.66 \pm 0.44$	
		135	12	15.	$6 \pm 2.9$	$1.5 \pm$	0.6	0.56	$\pm 0.09$	0.04	$4 \pm 0.008$	$0.86 \pm 0.57$	
		140	12	16.	$4 \pm 3.2$	$2.2 \pm$	0.8	0.93	$\pm 0.13$	0.09	$6 \pm 0.017$	$1.20 \pm 0.79$	
		150	16	18.	$7 \pm 3.8$	$3.2 \pm$	1.3	1.20	$\pm 0.15$	0.1	$3 \pm 0.03$	$1.9 \pm 1.2$	
		160	16	20.	$2 \pm 4.1$	4.1 ±	1.7	1.3	$\pm$ 0.2	0.2	$0 \pm 0.03$	$2.6 \pm 1.7$	
		170	18	21.	$4 \pm 4.2$	$4.6 \pm$	2.0	1.4	$\pm 0.2$	0.1	$8 \pm 0.03$	$3.0 \pm 2.0$	
		180	18	21.	$3 \pm 4.3$	$3.6 \pm$	1.5	1.1	$\pm 0.2$	0.1	$5 \pm 0.03$	$2.3 \pm 1.5$	
		190	18	22.	$2 \pm 4.5$	$2.3 \pm$	1.0	0.68	$\pm$ 0.09	0.1	$0 \pm 0.02$	$1.5 \pm 1.0$	
$m_H$	W.	Jets	$V\gamma$		ggV	VW	W	W	WZ/Z	ZZ	top	DY	DY $\rightarrow \tau \tau$
118	1.2 :	$\pm 0.7$	$0.17~\pm$	0.18	0.098 ±	E 0.050	2.0 ±	0.7	$0.20~\pm$	0.04	$5.4 \pm 2.0$	$1.2 \pm 0.44$	$0.3 \pm 0.1$
120	1.2 =	$\pm 0.7$	$0.17~\pm$	0.18	0.098 ∃	⊨ 0.045	2.1 ±	0.8	$0.21~\pm$	0.04	$5.4 \pm 2.0$	$1.2 \pm 0.44$	$0.3 \pm 0.1$
124	1.2 :	$\pm 0.7$	$0.17~\pm$	0.18	0.11 ±	$\pm 0.05$	2.3 ±	- 0.9	$0.22~\pm$	0.04	$7.0\pm2.3$	$1.6 \pm 0.51$	$0.3 \pm 0.1$
128	1.2 :	± 0.7	$0.17~\pm$	0.18	0.12 ±	E 0.05	2.6 ±	1.1	$0.23~\pm$	0.04	$8.2 \pm 2.5$	$1.8 \pm 0.54$	$0.3 \pm 0.1$
130	1.2 :	$\pm 0.7$	$0.17~\pm$	0.18	0.12 ±	$\pm 0.05$	2.8 ±	: 1.1	$0.24~\pm$	0.04	$8.2 \pm 2.5$	$2.2\pm0.62$	$0.3 \pm 0.1$
135	1.1 :	± 0.7	$0.17~\pm$	0.18	0.13 ±	$\pm 0.05$	3.1 ±	= 1.3	$0.25~\pm$	0.04	$8.1 \pm 2.5$	$2.4 \pm 0.66$	$0.3 \pm 0.1$
140	1.1 :	$\pm 0.7$	$0.17~\pm$	0.18	0.13 ±	E 0.06	3.3 ±	= 1.5	$0.28~\pm$	0.04	$8.8\pm2.6$	$2.3 \pm 0.61$	$0.3 \pm 0.1$
150	0.9 =	$\pm 0.7$	$0.17~\pm$	0.18	0.15 ±	E 0.06	3.7 ±	1.8	$0.30~\pm$	0.05	$10.4 \pm 3.2$	$2.7\pm0.67$	$0.3 \pm 0.1$
160	1.0 :	$\pm 0.7$	$0.17~\pm$	0.18	0.16 ∃	⊨ 0.06	4.3 ±	= 2.0	$0.34~\pm$	0.05	$11.1 \pm 3.5$	$2.8\pm0.69$	$0.3 \pm 0.1$
170	0.9 =	$\pm 0.7$	$0.17~\pm$	0.18	0.18 🗄	⊨ 0.07	4.8 ±	2.3	$0.34~\pm$	0.05	$11.7 \pm 3.4$	$3 \pm 0.71$	$0.3 \pm 0.1$
180	0.9 =	$\pm 0.7$	$0.17~\pm$	0.18	0.17 ∃	⊨ 0.07	4.9 ±	2.4	0.33 $\pm$	0.05	$11.6 \pm 3.4$	$3.1 \pm 0.75$	$0.2 \pm 0.1$
190	0.9 =	$\pm 0.7$	$0.17 \pm$	0.18	0.18 ±	E 0.07	$5.0 \pm$	= 2.5	$0.35~\pm$	0.06	$12.2 \pm 3.6$	$3.3 \pm 0.78$	$0.2 \pm 0.1$

Table 8.8: Expected number of signal and background events from the data-driven methods for an integrated luminosity of 4.9 fb<sup>-1</sup> for all analysis working points. Uncertainties used in datacards are reported, where the errors from different sources are summed in quadrature.

Since no significant excess is observed with respect to the background prediction, 95% confidence level (CL) upper limits are calculated for the Higgs boson cross section with respect the expected one in Standard Model using the modified frequentist construction  $CL_s$ . The expected and observed upper limits are shown in Fig. 8.10, and in Table 8.11. The observed (expected) upper limits at the 95% confidence level are about 9.4 (12.9) times larger than the SM expectation for  $m_{\rm H} = 125 \text{ GeV/c}^2$ . As expected, since the sensitivity for the Standard Model Higgs boson is of the order of 10 times the nominal cross section for this channel, good agreement between the background expectations and data is observed.

	$m_H$	data	t	ot bkg	to	t sig		VH		qqH		ggH		
	118	2	5.	$7 \pm 1.8$	0.084	$\pm 0.035$	0.0	$46 \pm 0.023$	0.0	$0097 \pm 0.000$	36	0.036 $\pm$	0.026	
	120	2	5.	$9 \pm 1.8$	0.099	$\pm 0.044$	0.0	$41 \pm 0.018$	0.	$0038 \pm 0.001$	5	$0.054~\pm$	0.040	
	124	2	7.	$5 \pm 2.2$	0.17	$\pm 0.07$	0.0	$93 \pm 0.037$	(	$0.003 \pm 0.001$	.	$0.079~\pm$	0.061	
	128	3	8.	$5 \pm 2.2$	0.24	$\pm 0.10$	0.0	$87 \pm 0.021$	(	$0.013 \pm 0.005$		$0.14~\pm$	0.10	
	130	3	9.	$0 \pm 2.3$	0.41	$\pm 0.17$	0.	$17 \pm 0.04$	(	$0.010 \pm 0.004$	.	$0.23~\pm$	0.16	
	135	5	9.	$4 \pm 2.4$	0.40	$\pm 0.19$	0.	$12 \pm 0.04$	(	$0.014 \pm 0.006$		$0.27~\pm$	0.19	
	140	5	9.	$9 \pm 2.6$	0.74	$\pm 0.31$	0.	$26~\pm~0.05$	(	$0.025 \pm 0.007$	·	$0.45~\pm$	0.30	
	150	5	11	$.7 \pm 2.9$	1.1	$\pm 0.5$	0.	$39 \pm 0.07$	(	$0.034 \pm 0.010$		$0.72~\pm$	0.48	
	160	5	12	$.3 \pm 3.1$	1.5	$\pm$ 0.6	0.	$52 \pm 0.13$	(	$0.068 \pm 0.015$		$0.88~\pm$	0.59	
	170	6	12	$.9 \pm 3.0$	1.7	$\pm 0.8$	0.	$51 \pm 0.09$	(	$0.074 \pm 0.014$	.	1.10 $\pm$	0.76	
	180	6	12	$.9 \pm 3.1$	1.4	$\pm$ 0.6	0.	$45 \pm 0.07$	(	$0.052 \pm 0.011$		$0.94~\pm$	0.63	
	190	6	13	$.4 \pm 3.3$	0.94	$\pm 0.41$	0.	$28 \pm 0.05$	(	$0.043 \pm 0.013$		$0.62~\pm$	0.41	
$m_H$	WJe	ets	$V\gamma$	ggW	W	WW		WZ/ZZ		top	DY	$f \rightarrow \mu \mu$	$DY \rightarrow ee$	
118	$0.05~\pm$	0.11	-	0.035 $\pm$	0.041	$0.59 \pm 0$	.31	$0.041 \pm 0.0$	013	$3.9 \pm 1.7$	0.46	$5 \pm 0.31$	$0.72 \pm 0.31$	1
120	0.05 $\pm$	0.11	-	0.035 $\pm$	0.027	$0.60 \pm 0$	.33	$0.043 \pm 0.0$	)14	$3.9 \pm 1.7$	0.46	$5 \pm 0.31$	$0.79 \pm 0.33$	3
124	$0.05~\pm$	0.11	-	$0.035~\pm$	0.025	$0.69 \pm 0$	.40	$0.049 \pm 0.0$	)13	$5.2 \pm 2.0$	0.79	$0 \pm 0.46$	$0.78 \pm 0.33$	3
128	$0.02~\pm$	0.11	-	0.035 $\pm$	0.025	$0.84 \pm 0$	.46	$0.054 \pm 0.0$	014	$5.8 \pm 2.0$	1.00	$0 \pm 0.55$	$0.77 \pm 0.34$	1
130	$0.02~\pm$	0.11	-	0.035 $\pm$	0.025	$0.89 \pm 0$	.46	$0.056 \pm 0.0$	015	$5.8 \pm 2.1$	1.34	$1 \pm 0.71$	$0.83 \pm 0.37$	7
135	$0.08~\pm$	0.14	-	0.047 $\pm$	0.035	$0.97 \pm 0$	.47	$0.062 \pm 0.0$	017	$5.8 \pm 2.2$	1.56	$5\pm0.79$	$0.89 \pm 0.39$	)
140	$0.08~\pm$	0.14	-	0.047 $\pm$	0.031	$1.0 \pm 0$	.5	$0.074 \pm 0.0$	017	$6.4 \pm 2.4$	1.47	$2 \pm 0.72$	$0.85 \pm 0.39$	)
150	$0.08~\pm$	0.14	-	0.047 $\pm$	0.032	$1.2 \pm 0$	.6	$0.087 \pm 0.0$	019	$7.7 \pm 2.7$	1.79	$0 \pm 0.84$	$0.90 \pm 0.43$	3
160	$0.08~\pm$	0.14	-	$0.056~\pm$	0.038	$1.4 \pm 0$	.8	$0.11 \pm 0.0$	$\mathbf{D2}$	$ 7.8 \pm 2.8 $	1.88	$8 \pm 0.85$	$0.97 \pm 0.47$	7
170	$0.09~\pm$	0.14	-	$0.065~\pm$	0.042	$1.6 \pm 0$	.8	$0.11 \pm 0.0$	$\mathbf{D2}$	$8.0 \pm 2.7$	1.96	$5\pm0.83$	$1.03 \pm 0.50$	)
180	$0.09~\pm$	0.14	-	$0.065~\pm$	0.042	$1.7 \pm 0$	.9	$0.11 \pm 0.0$	)2	$7.8 \pm 2.8$	2.02	$2 \pm 0.84$	$1.07 \pm 0.48$	3
190	$0.09~\pm$	0.14	-	$0.065~\pm$	0.042	$1.8 \pm 1$	.0	$0.12 \pm 0.0$	)3	$8.1 \pm 2.9$	2.13	$3 \pm 0.88$	$1.13 \pm 0.51$	1

Table 8.9: Expected number of signal and background events from the data-driven methods for an integrated luminosity of 4.9 fb<sup>-1</sup> for all analysis working points in the same flavour final state (ee, $\mu\mu$ ). Uncertainties used in datacards are reported, where the errors from different sources are summed in quadrature.

	$m_H$	data	tot bkg	tot sig	VH	qqH	ggH	
	118	4	$4.7 \pm 1.4$	$0.29 \pm 0.10$	$0.13 \pm 0.02$	$0.014 \pm 0.002$	$2 0.14 \pm 0$	.10
	120	4	$4.9 \pm 1.5$	$0.41 \pm 0.13$	$0.20\pm0.03$	$0.019 \pm 0.003$	$0.19 \pm 0$	.13
	124	4	$5.4 \pm 1.6$	$0.50\pm0.18$	$0.21\pm0.04$	$0.024 \pm 0.004$	$1 0.27 \pm 0$	.18
	128	5	$6.1 \pm 1.8$	$0.75 \pm 0.28$	$0.30\pm0.05$	$0.035 \pm 0.006$	$0.41 \pm 0$	.28
	130	5	$6.2 \pm 1.9$	$0.84\pm0.30$	$0.37\pm0.06$	$0.041 \pm 0.007$	$7 0.43 \pm 0$	.29
	135	7	$6.4 \pm 1.8$	$1.10 \pm 0.41$	$0.44 \pm 0.08$	$0.029 \pm 0.003$	$5 \mid 0.60 \pm 0$	.40
	140	7	$6.7 \pm 1.8$	$1.5 \pm 0.5$	$0.66 \pm 0.11$	$0.072 \pm 0.013$	$0.75 \pm 0$	.51
	150	11	$7.0 \pm 1.9$	$2.1 \pm 0.8$	$0.80\pm0.12$	$0.097 \pm 0.018$	$3 1.2 \pm 0$	.8
	160	11	$8.0 \pm 2.0$	$2.6 \pm 1.1$	$0.79\pm0.11$	$0.130 \pm 0.020$	$1.7 \pm 1$	.1
	170	12	$8.6 \pm 2.2$	$2.9 \pm 1.3$	$0.92\pm0.13$	$0.110 \pm 0.018$	$3   1.9 \pm 1$	.3
	180	12	$8.4 \pm 2.2$	$2.1 \pm 0.9$	$0.64 \pm 0.10$	$0.100 \pm 0.023$	$3   1.4 \pm 0$	.9
	190	12	$8.8 \pm 2.3$	$1.3 \pm 0.6$	$0.40\pm0.06$	$0.062 \pm 0.013$	$0.86 \pm 0$	.57
$m_H$	WJ	ets	$V\gamma$	ggWW	WW	WZ/ZZ	top	$DY \rightarrow \tau \tau$
118	$1.1 \pm$	0.7	$0.17 \pm 0.18$	$0.063 \pm 0.036$	$1.4 \pm 0.6$	$0.16 \pm 0.03$	$1.5 \pm 1.1$	$0.3 \pm 0.1$
120	$1.1 \pm$	0.7	$0.17 \pm 0.18$	$0.063 \pm 0.032$	$1.5 \pm 0.6$	$0.17\pm0.03$	$1.5 \pm 1.1$	$0.3\pm0.1$
124	$1.1 \pm$	0.7	$0.17 \pm 0.18$	$0.073 \pm 0.033$	$1.6 \pm 0.6$	$0.17\pm0.03$	$1.9 \pm 1.3$	$0.3  \pm  0.1$
128	$1.2 \pm$	0.7	$0.17 \pm 0.18$	$0.081 \pm 0.039$	$1.8 \pm 0.7$	$0.18\pm0.03$	$2.4 \pm 1.5$	$0.3  \pm  0.1$
130	$1.2 \pm$	0.7	$0.17 \pm 0.18$	$0.081 \pm 0.038$	$1.9 \pm 0.8$	$0.18\pm0.03$	$2.4 \pm 1.5$	$0.3  \pm  0.1$
135	$1.2 \pm$	0.7	$0.17 \pm 0.18$	$0.081 \pm 0.039$	$2.1 \pm 0.9$	$0.19\pm0.03$	$2.3 \pm 1.3$	$0.3  \pm  0.1$
140	$1.2 \pm$	0.7	$0.17 \pm 0.18$	$0.087 \pm 0.045$	$2.3 \pm 1.1$	$0.20 \pm 0.04$	$2.5 \pm 1.2$	$0.3  \pm  0.1$
150	$1.0 \pm$	0.7	$0.17 \pm 0.18$	$0.10\pm0.05$	$2.6 \pm 1.2$	$0.22 \pm 0.04$	$2.6 \pm 1.2$	$0.3  \pm  0.1$
160	$1.1 \pm$	0.7	$0.17\pm0.18$	$0.10\pm0.05$	$2.8 \pm 1.3$	$0.23 \pm 0.04$	$3.2 \pm 1.4$	$0.3~\pm~0.1$
170	$0.98~\pm$	0.70	$0.17\pm0.18$	$0.12 \pm 0.05$	$3.2 \pm 1.5$	$0.24 \pm 0.04$	$3.6 \pm 1.4$	$0.3\pm0.1$
180	$0.98~\pm$	: 0.70	$0.17\pm0.18$	$0.11 \pm 0.05$	$3.2 \pm 1.6$	$0.22 \pm 0.04$	$3.6 \pm 1.4$	$0.2~\pm~0.1$
190	$0.98~\pm$	0.70	$0.17 \pm 0.18$	$0.11 \pm 0.05$	$3.2 \pm 1.6$	$0.22 \pm 0.04$	$3.9 \pm 1.5$	$0.2 \pm 0.1$

Table 8.10: Expected number of signal and background events from the data-driven methods for an integrated luminosity of 4.9 fb<sup>-1</sup> for all analysis working points in the different flavour final state  $(e\mu)$ . Uncertainties used in datacards are reported, where the errors from different sources are summed in quadrature.



Figure 8.8: The transverse mass  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$ , as defined in Eq. 6.18, at Higgs level without the  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  cut itself. In the analysis different selections are applied according to Higgs mass hypothesis, as listed in Sectin 8.1. Backgrounds are normalized with data driven techniques and signals are plotted one on top of the other. The error band on the expected rates reflects the Monte Carlo statistics and the systematic error coming from data driven estimation.

Mass	Observed	Expected	68% probability band	95% probability band
118	17.9	21.15	[14.82, 32.60]	[8.81,48.72]
120	13.0	15.50	$[\ 10.86 \ , 23.74 \ ]$	$\left[ \begin{array}{c} 6.21 \end{array}, 35.24 \end{array}  ight]$
124	9.4	12.89	$[ \ 9.13 \ , \ 19.89 \ ]$	$\left[ \; 4.51 \; , \; 29.29 \;  ight]$
128	7.5	9.49	$\left[ \; 6.51 \; , \; 14.82 \;  ight]$	$\left[ \hspace{0.1 cm} 3.90 \hspace{0.1 cm} , \hspace{0.1 cm} 19.23 \hspace{0.1 cm}  ight]$
130	5.8	8.03	$\left[ \; 5.45 \; , \; 12.38 \; \right]$	$\left[ \hspace{0.1 cm} 3.43 \hspace{0.1 cm} , \hspace{0.1 cm} 15.18 \hspace{0.1 cm}  ight]$
135	7.1	7.10	$\left[ \; 4.93 \; , \; 10.91 \;  ight]$	$\left[ \ 3.10 \ , \ 14.68 \  ight]$
140	4.3	4.75	$\left[ {\ 3.29} \ ,  7.28 \  ight]$	$\left[ \; 2.05 \; , \; 10.70 \;  ight]$
150	4.5	3.71	$\left[ {\ 2.65} \ ,  5.62 \  ight]$	$[ \ 1.79 \ , 8.40 \ ]$
160	3.3	3.13	$\left[ {\ 2.17} \ , \ 4.83 \ \right]$	$[ \ 1.32 \ , \ 7.30 \ ]$
170	3.1	2.91	$\left[ {\ 2.02} \ ,  4.46 \  ight]$	$[ \ 1.34 \ , \ 6.81 \ ]$
180	4.0	3.81	$\left[ {\ 2.64} \ ,  5.95 \  ight]$	$[ \ 1.64 \ , 8.93 \ ]$
190	6.1	6.17	$\left[ \begin{array}{c} 4.31 \end{array}, 9.61  ight]$	$\left[ \hspace{0.1 cm} 3.17 \hspace{0.1 cm} , \hspace{0.1 cm} 14.55 \hspace{0.1 cm}  ight]$

Table 8.11: Expected and observed upper limits with 4.9  $fb^{-1}$  in the SM Higgs scenario.


Figure 8.9: The invariant mass of the two jets  $m_{jj}$  at Higgs level without the  $m_{jj}$  cut ( [60-110] GeV/c<sup>2</sup>). Backgrounds are normalized with data driven techniques and signals are plotted one on top of the other. The error band on the expected rates reflects the Monte Carlo statistics and the systematic error coming from data driven estimation.



Figure 8.10: Upper limits at 95% CL for 4.9  $\rm fb^{-1}for$  SM Higgs boson. Logarithmic and linear plots are shown.

# CHAPTER 9

### THE FINAL RESULTS

When you have excluded the impossible, whatever remains, however improbable, must be the truth

Sherlock Holmes

Since the two analyses, VBF and VH, are probing ortogonal phase spaces, (high and low  $m_{jj}$  regions, high and low  $\Delta \eta_{jj}$  regions) the measurements from the two can be combined in order to have a clear and unique picture for the discovery or the esclusion of the Standard Model Higgs boson. In addition the VBF and VH analysis are part of the H  $\rightarrow$  W<sup>+</sup>W<sup>-</sup> combination, together with the zero-jet-bin and one-jet-bin categorization (see Chapter 6), and part of the effort combining all the Higgs searches in CMS to discover the new boson, namely H  $\rightarrow \gamma\gamma$ , H  $\rightarrow$  ZZ, H  $\rightarrow \tau\tau$  and H  $\rightarrow bb$ . The VBF and VH H  $\rightarrow$  W<sup>+</sup>W<sup>-</sup> results have been also interpreted assuming different coupling of the Higgs boson with vector bosons and fermions, such as the Fermiophobic model, where the coupling to fermions is set to zero.

The combination of the VH and VBF analyses are summarized in Sec. 9.1, while the  $H \rightarrow W^+W^-$  combination is described in Sec. 9.2. The Sec. 9.3 briefly summarizes the CMS Higgs combined limit and in Sec. 9.4 the Fermiophobic results, in which VBF and VH analysis are one of the main players, are reported.

# 9.1 VBF and VH Standard Model Higgs boson ${\rm H} \rightarrow {\rm W}^+ {\rm W}^-$ combined limit

As described in Section 6.7 the strength modifier  $\mu$  is taken coherently for all production mechanisms (gluon fusion, VBF and associated production). Therefore, it is possible to combine the results from VBF analysis and VH one and derive upper limits for the Higgs boson as the ratio of the product of the Higgs boson cross section and the  $H \rightarrow W^+W^-$  branching fraction,  $\sigma_H \times BR(H \rightarrow W^+W^-)$ , and the SM Higgs expectation,  $\sigma/\sigma_{SM}$ . The results are also summarized in Fig. 9.1. The combined analysis excludes the presence of a Higgs boson with mass in the range 135–250 GeVat 95% CL. The observed (expected) upper limits are about 1.5 (1.5) times the SM



Figure 9.1: Expected and observed 95% CL upper limits on the cross section times branching fraction,  $\sigma_{\rm H} \times {\rm BR}({\rm H} \to {\rm W}^+ {\rm W}^-)$ , relative to the SM Higgs expectation using the combined 7 TeV and 8 TeV data and both the VBF and VH analysis. Results are obtained using the CL<sub>s</sub> approach.

expectation for  $m_{\rm H} = 125$  GeV. No significant excess is observed for low Higgs boson masses, but the combination of these two channels is still limited by the statistics, and has not reached yet the sensitivity to the SM Higgs boson at 125 GeV.

Comparing the exclusion plot for VBF only Fig. 7.11 and the combined one Fig. 9.1, the main contribution to the combined limit comes from the VBF analysis. However it is important to perform also the VH analysis since it probes a different production mode, expected in the Standard Model, that needs to be checked to see if it is consistent with the predictions, or if there are deviations, and then hints for new physics.

### 9.2 Standard Model Higgs boson $H \rightarrow W^+W^-$ combined limit

The  $H \rightarrow W^+W^-$  total combination is the sum of the performances of the zero-jet-bin, one-jet-bin and two-jet-bin categories (see Chapter 6). In this combination the VH analysis is currently excluded due to its low sensitivity with the current available statistics. It will be combined, together with other  $H \rightarrow W^+W^-$  analyses (such as, WH analysis, where the Higgs radiating W decays into an electron or a muon) at the end of 2012 data-taking.

The zero-jet-bin and one-jet-bin analyses are divided into same flavour and different flavour sub-categories as well, because of different contamination of Drell-Yan. While the same flavour is a cut based analysis, as for VBF, for the different flavour analysis both a cut based approach and a shape based one is performed. In addition, there are two separated shape based analyses, both with similar performances: one that relies on a multidimensional discriminant (BDT) trained for each Higgs mass point, and one based on a bidimensional shape analysis in the  $m_{\ell\ell}$  and  $m_{\rm T}^{\ell\ell E_{\rm T}^{\rm miss}}$  plane [128].

The separeted contribution of the channels in the  $H \rightarrow W^+W^-$  combined analysis are reported in Fig. 9.2. The VBF different flavour analysis is the fourth in terms of exclusion potential.

The combination of the results from the 8+7 TeV analysis excludes a Higgs boson in the mass



Figure 9.2: Comparison of the expected limits for different analyses at 8 TeV for low Higgs mass hypotheses. The VBF analysis is the fourth in terms of exclusion potential.

range 128–600 GeV at 95% CL. The expected exclusion range for the background only hypothesis is 118–565 GeV. An excess of events is observed for an hypothetical low Higgs boson masses, which makes the observed limits weaker than the expected ones. Due to the poor mass resolution of this channel the excess extends over a large mass range. The results are also summarized in Fig.9.3. The shaded area reflects the expected exclusion limit in the presence of a Higgs boson with mass 125 GeV. The expected and observed significance for a SM Higgs with a mass of 125 GeV are respectively 4.1 and 3.1 standard deviations, as shown in Fig. 9.4.



Figure 9.3: Expected and observed 95% CL upper limits on the cross section times branching fraction,  $\sigma_{\rm H} \times {\rm BR}({\rm H} \to {\rm W}^+{\rm W}^-)$ , relative to the SM Higgs expectation using the combined 7 TeV and 8 TeV data and all the H  $\to {\rm W}^+{\rm W}^-$  analyses. Results are obtained using the CL<sub>s</sub> approach. The expected limits in the presence of the Higgs with  $m_{\rm H} = 125$  GeV and its associated uncertainty are also shown in red shaded area.



Figure 9.4: The observed and expected under the  $m_{\rm H} = 125$  GeV Higgs signal hypothesis significances for the combined 7+8 TeV analysis.

#### 9.3 Standard Model Higgs boson combined limit

The relevant decay modes of the SM Higgs boson depend strongly on its mass  $m_{\rm H}$ . The decay channels considered in CMS are:  ${\rm H} \to \gamma \gamma$ ,  ${\rm H} \to \tau \tau$ , followed by leptonic and hadronic decays of  $\tau$ -leptons,  ${\rm H} \to bb$ ,  ${\rm H} \to {\rm W}^+{\rm W}^-$ , followed by  $W^+W^- \to \ell^+\nu\ell^-\bar{\nu}$  and  $W^+W^- \to \ell\nu jj$  decays, and  ${\rm H} \to {\rm ZZ}$ , followed by ZZ decays to  $4\ell$ ,  $2\ell 2\nu$ ,  $2\ell 2q$ , and  $2\ell 2\tau$ . The WW and ZZ decay modes are used over the entire explored mass range. The  $\gamma\gamma$ ,  $\tau\tau$ , and bb decay modes are used only for  $m_{\rm H} < 150$  GeV since their expected sensitivities are not significant compared to WW and ZZ for higher Higgs boson masses.

For a given hypothesis for the Higgs boson mass, the sensitivity of the search depends on the production cross section of the Higgs boson, its decay branching fraction into the chosen final state, the signal selection efficiency, the mass resolution, and the level of standard model backgrounds. For low values of the Higgs boson mass, the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ \rightarrow 4\ell$  channels play a special role due to the excellent mass resolution for the reconstructed di-photon and four-lepton final states. The  $H \rightarrow WW \rightarrow \ell\nu\ell\nu$  channel provides high sensitivity but has relatively poor mass resolution due to the presence of neutrinos in the final state. In the high mass range, the search sensitivity is dominated by the WW and ZZ modes.

Table 9.1 shows modes used in the searches for a light Higgs boson ( $110 < m_H < 145 \text{GeV}$ ). The search modes are grouped in this table by the production and decay modes specifically targeted by the corresponding analyses. The naming convention reflects the signature targeted. None of these signatures is 100% pure.

	untagged	VBF-tag	VH-tag	<i>tt</i> H-tag
$H \rightarrow \gamma \gamma$	$\checkmark$	$\checkmark$	$\checkmark$	
$H \rightarrow bb$			$\checkmark$	$\checkmark$
$H \to \tau \tau$	$\checkmark$	$\checkmark$	$\checkmark$	
$H \to WW$	$\checkmark$	$\checkmark$	$\checkmark$	
$H \rightarrow ZZ$	$\checkmark$			

Table 9.1: Summary of production mechanisms and decay channels explicitly targeted in the searches for a low mass Higgs boson ( $m_H < 145 \text{GeV}$ ). Untagged searches include gluon-gluon fusion  $gg \rightarrow H$  plus any phase space not covered by searches with explicit tags for enriching datasets with events from VBF, VH, and ttH production. V stands for W or Z. All analyses targeting a particular production mechanism are never 100% pure and have a mixture of other production mechanisms.

As an illustration of the search sensitivity of the different channels, Fig. 9.5 shows the median expected 95% CL upper limit on the ratio of the signal cross section,  $\sigma$ , and the predicted SM Higgs boson cross section,  $\sigma_{\rm SM}$ , as a function of the SM Higgs boson mass hypothesis. The  $H \rightarrow WW \rightarrow \ell \nu \ell \nu$  analysis is one of the most important in mostly the whole Higgs boson mass range. Fig. 9.6 shows the expected sensitivities for the observation of the SM Higgs boson in terms of *p*-values and significances.

To quantify the inconsistency of the observed excesses with the background-only hypothesis, in Fig. 9.7 (left) a scan of the local *p*-value  $p_0$  for 7 TeV, 8 TeV, and the overall combination is shown. Both 7 TeV and 8 TeV datasets exhibit excesses for a Higgs boson mass around 125 GeV. In the overall combination, this results in a 6.9  $\sigma$  excess at  $m_{\rm H} = 125$  GeV. Fig. 9.7 (right) gives *p*-values for sub-combinations by decay channel. Table 9.2 summarises the median expected



Figure 9.5: The median expected 95% CL upper limits on the cross section ratio  $\sigma/\sigma_{\rm SM}$ in the absence of a Higgs boson as a function of the SM Higgs boson mass in the range 110–1000GeV (a) and 110–145GeV (b), for the five Higgs boson decay channels. Here  $\sigma_{\rm SM}$  denotes the cross section predicted for the SM Higgs boson. A channel showing values below unity (dotted red line) would be expected to be able to exclude a Higgs boson of that mass at 95% CL. The jagged structure in the limits for some channels results from the different event selection criteria employed in those channels for different Higgs boson mass sub-ranges.



Figure 9.6: The expected sensitivities for the observation of the SM Higgs boson in terms of p-values and significances is shown as a function of the SM Higgs boson mass in the range 110–145GeV, for the five Higgs boson decay channels.



Figure 9.7: (Left) The observed local *p*-value  $p_0$  for 7-TeV, 8-TeV data, and their combination as a function of the Higgs boson mass. (Right) The observed local *p*-value  $p_0$  for five sub-combinations by decay mode and the overall combination as a function of the Higgs boson mass. The dashed lines show the expected local *p*-values  $p_0(m_{\rm H})$ , should a SM Higgs boson with a mass  $m_{\rm H}$  exist.

and observed local significances for a SM Higgs boson mass hypothesis of  $125.8 \,\mathrm{GeV^1}$  for the individual decay modes and their various combinations.

Decay mode or Combination	Expected $(\sigma)$	Observed $(\sigma)$
ZZ	5.0	4.4
$\gamma\gamma$	2.8	4.0
WW	4.3	3.0
bb	2.2	1.8
au au	2.1	1.8
$\gamma\gamma + ZZ$	5.7	5.8
$\gamma\gamma+ZZ+WW+ au au+bb$	7.8	6.9

Table 9.2: The significance of the median expected and observed event excesses in individual decay modes and their various combinations for a SM Higgs boson mass hypothesis of 125.8 GeV.

<sup>&</sup>lt;sup>1</sup>The Higgs mass with largest significance

### 9.4 Fermiophobic Higgs

Several extensions of the Standard Model (SM) with an enlarged scalar sector have been proposed. It is possible that the electroweak symmetry breaking mechanism responsible for generating the masses of the gauge bosons is independent of the mechanism that generates the fermion masses. In this case there should exist a Higgs boson that, at tree level, couples only to W and Z bosons. In the literature this model is usually referred to as the fermiophobic (FP) Higgs [129, 130]. Its decay to W and Z bosons is SM-like, while the decay to photons is via W loops.

In the framework of the FP model, the branching ratios are different with respect to the SM: Fig. 9.8 shows the BR for a SM Higgs and for a FP one. The di-photon decay and the WW decay are enhanced in the low Higgs mass region. Fig. 9.9 shows the ratio of the BR of  $H \rightarrow W^+W^-$ 



Figure 9.8: Higgs branching ratios for Standard Model (a) and Fermiophobic model. The  $H \rightarrow W^+W^-$  is enchanced in the low mass region.

between FP and Standard Model Higgs. The FP model Higgs has the SM-like production cross section for vector boson fusion (VBF) and associated production with a W or Z (VH), while gluon fusion and top quark fusion are absent or have a negligibly small cross section. Previous searches at LEP [131] and at the Tevatron [132] ruled out a FP Higgs boson lighter than 119 GeV at 95% confidence level (CL).

The results of the VBF [127] and VH [113] analyses can thus be interpreted in the Fermiophobic scenario. The 95% CL upper limit for a fermiophobic Higgs boson is shown in Fig. 9.10 for the VH analysis while Fig. 9.11 shows the 95% CL upper limit for a fermiophobic Higgs boson obtained with the VBF analysis.

The VBF analysis has been combined, together with VBF  $H \rightarrow \gamma\gamma$  analysis in order to get a unique exclusion interval: while the  $H \rightarrow \gamma\gamma$  analysis is the most performing in the low mass range (<130 GeV) the VBF  $H \rightarrow W^+W^-$  analysis is the main player in the intermediate mass range (130 GeV <  $m_{\rm H}$  < 200 GeV) [133]. Fig. 9.12 shows the exclusion range from the combination: the presence of a Higgs boson with Fermiophobic couplings with mass in the range 110-192 GeV is excluded at 95% CL. The excess at 125 GeV is due to the SM Higgs boson.



Figure 9.9: Ratio between the Higgs branching ratios of  $H \rightarrow W^+W^-$ ,  $H \rightarrow ZZ$  and  $H \rightarrow \gamma\gamma$  for Fermiophobic Higgs model and Standard Model one.



Figure 9.10: Upper limits at 95% CL for 4.9 fb<sup>-1</sup>in the fermiophobic Higgs scenario using the results of the VH analysis. Logarithmic and linear plots are shown [113].



Figure 9.11: Upper limits at 95% CL for 4.9 fb<sup>-1</sup> in the fermiophobic Higgs scenario using the results of the VBF  $H \rightarrow W^+W^-$  analysis [127]. A low Higgs boson mass is shown in (a), while the whole spectrum is depicted in (b).



Figure 9.12: Upper limits at 95% CL for 4.9 fb<sup>-1</sup>in the fermiophobic Higgs scenario combining the VBF  $H \rightarrow W^+W^-$  results and the  $H \rightarrow \gamma\gamma$  ones from CMS [133].

## CHAPTER 10\_

### THE CONCLUSIONS

(...) il sugo di tutta la storia. La quale, se non v'è dispiaciuta affatto, vogliatene bene a chi l'ha scritta, (...) Ma se in vece fossimo riusciti ad annoiarvi, credete che non s'è fatto apposta.

Alessandro Manzoni, Promessi Sposi

During my PhD work I mainly focused on the search for a Higgs boson in the  $H \to WW \to \ell \nu \ell \nu$  channel. In particular I concentrated on the vector boson fusion (VBF) and vector associated (VH) production of the Higgs.

Given the leptonic decays of WW, my commissioning work was focused on the electromagnetic calorimeter (ECAL) performances: I developed and performed the alignment of the ECAL achieving the goal needed for a good electron identification. In addition the ECAL alignment is requested by the  $H \rightarrow \gamma \gamma$  analysis in order to reconstruct the Higgs mass peak from the di-photon kinematics, whose direction relies completely on ECAL position reconstruction.

Both the VBF and the VH analyses relies not only on the goodness of leptons, but also on the capability of CMS to reconstruct jets. I performed one of the first measurements of CMS, with data collected during 2010 run, looking at the simultaneous production of one central jet and one forward jet. The results of this analysis, in addition of being one of the first measurements involving jets, have been used as inputs by theoriticians to improve the description of Monte Carlo simulations needed for LHC particular phase space.

I performed the VBF H  $\rightarrow$  WW  $\rightarrow \ell \nu \ell \nu$  analysis from Monte Carlo studies and optimizations, through the first checks with few inverse femtobarns collected, till the very last result, using the full luminosity recorded up to October 2012 and reported in the HCP conference. The analysis excludes a Higgs boson with 95% confidence level with a mass between 135 and 200 GeV. The sensitivity at 125 GeV is not high enough in order to see a Higgs boson produced by the fusion of vector bosons and decaying in to two leptons and neutrinos pairs. However with the full statistics before the long shutdown (2013-2014), it will be possible to have the first glance of this production mechanism.

I proposed, optimized and performed the associated production analysis of a Higgs boson decaying into  $W^+W^- \rightarrow \ell^+\nu\ell^-\bar{\nu}$ , with the vector boson that radiates the Higgs decaying into two jets. The analysis covers an orthogonal phase space with repsect to VBF one, but, even if less sensitive, it is an important ingredient in the Higgs search because it probes a different production mechanism.

Both the analyses, VBF and VH, have been combined with the numerous channels in CMS, thus leading to the exclusion of the Higgs boson in almost the whole mass range, and to the discovery of a Higgs boson at 125 GeV.

All Higgs searches will be updated with the full statistics collected in 2011 and 2012, thus leading to a definitive and clear picture of the Higgs boson and its characteristics, such as its mass, its couplings with SM particles, its spin and parity.

### BIBLIOGRAPHY

- [1] Francis Halzen and Alan Martin. Quarks & Leptons: An introductory course in modern particle physics. John Wiley & Sons, New York, USA, 1984.
- [2] M. E. Peskin and D. V. Schroeder. An Introduction to quantum field theory. Westview Press, 1995.
- [3] C. Amsler et al. Particle Data Group. http://pdg.lbl.gov/, 2012.
- [4] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken symmetries. Phys. Rev., 127(3):965-970, Aug 1962.
- [5] Ho-Kim Quang and Pham Xuan-Yem. Elementary Particles and Their Interactions.
- [6] G. Altarelli and G. Isidori. Lower limit on the higgs mass in the standard model: An update. *Physics Letters B*, 337(1-2):141 – 144, 1994.
- [7] Search for the standard model higgs boson at lep. Physics Letters B, 565:61 75, 2003.
- [8] The TEVNPH Working Group. Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10.0<sup>-1</sup> of Data. 2012.
- [9] D0 Collaboration CDF Collaboration. Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron. 2012.
- [10] H. Veltman. The process  $f\bar{f} \rightarrow wlwl$ . Phys. Rev. D, 43(7):2236–2258, Apr 1991.
- [11] The LEP Eletroweak Working Page. http://lepewwg.web.cern.ch/LEPEWWG/.
- [12] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011.
- [13] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, CERN, Geneva, 2012.
- [14] H. M. Georgi, S. L. Glashow, M. E. Machacek, and D. V. Nanopoulos. Higgs bosons from two-gluon annihilation in proton-proton collisions. *Phys. Rev. Lett.*, 40(11):692–694, Mar

1978.

- [15] LEP design report. CERN, Geneva, June 1984, CERN-LEP/84-01, June 1996, CERN-AC/96-01 (LEP2). Copies shelved as reports in LEP, PS and SPS libraries.
- [16] G. L. Bayatian et al. CMS physics: Technical design report. CERN-LHCC-2006-001.
- [17] RHIC website. http://www.bnl.gov/rhic/.
- [18] https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.
- [19] LHCb webpage. http://lhcb.web.cern.ch/lhcb/.
- [20] ALICE webpage. http://aliceinfo.cern.ch/.
- [21] http://www.hep.phy.cam.ac.uk/~wjs/plots/plots.html.
- [22] *Technical proposal.* LHC Tech. Proposal. Cover title : CMS, the Compact Muon Solenoid : technical proposal.
- [23] CMS Collaboration. The cms tracker system project: Technical design report. Technical Design Report CMS, CMS-TDR-005, 1997.
- [24] CMS Collaboration. The cms electromagnetic calorimeter project: Technical design report. Technical Design Report CMS, CMS-TDR-004, 1997.
- [25] S.Derenzo. Ieee short course. 2002.
- [26] CASTOR webpage. http://cmsdoc.cern.ch/cms/castor/.
- [27] CMS Collaboration. Cms tridas project : Technical design report; 1, the trigger systems. Technical Design Report CMS, CMS-TDR-006-, 2000.
- [28] CMS Collaboration. The cms muon project : Technical design report. Technical Design Report CMS, CMS-TDR-003, 1997.
- [29] Muon reconstruction in the cms detector. CMS AN 2008/097, 2008.
- [30] The CMS collaboration. Performance of CMS muon reconstruction in pp collision events at  $\sqrt{s} = 7$  TeV. Journal of Instrumentation, 7:2P, October 2012.
- [31] Meschi, Emilio and Monteiro, T and Seez, Christopher and Vikas, Pratibha. Electron reconstruction in the cms electromagnetic calorimeter. Technical report, CERN, Geneva, Jun 2001.
- [32] H. Bethe and W. Heitler. On the Stopping of Fast Particles and on the Creation of Positive Electrons. Royal Society of London Proceedings Series A, 146:83–112, August 1934.
- [33] W. Adam, et al. Electron reconstruction in cms. CMS AN, CMS-AN-2009-164.
- [34] CMS Collaboration. CMS Physics Technical Design Report Volume I: Detector Performance and Software. Technical Design Report CMS. CERN, Geneva, 2006.
- [35] Cacciari, Matteo and Salam, Gavin P. and Soyez, Gregory. The Anti- $k_T$  jet clustering algorithm. *JHEP*, 0804:063, 2008.
- [36] Jet energy resolution in cms at sqrt(s)=7 tev. CMS-PAS-JME-10-014, 2011.
- [37] Missing et in 0.9 and 2.36 tev pp collisions. 2010.
- [38] CMS Collaboration. Commissioning of the particle-flow event reconstruction with the first

lhc collisions recorded in the cms detector. CMS Physics Analysis Summaries, PFT-10-001, 2010.

- [39] CMS Collaboration. Commissioning of the particle-flow reconstruction in minimum-bias and jet events from pp collisions at 7tev. CMS Physics Analysis Summaries, PFT-10-002, 2010.
- [40] https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsJME2012JEC.
- [41] https://twiki.cern.ch/twiki/bin/view/cms/ecalalignment.
- [42] P. Meridiani, C. Rovelli. Position resolution at the 2006 ecal testbeam. CMS DN, CMS DN-2007/011.
- [43] CMS collaboration. Measurements of inclusive w and z cross sections in pp collisions at  $\sqrt{s} = 7$  tev. Dec 2010.
- [44] CMS Collaboration. Ecal performance and electromagnetic object reconstruction in cms. EGM-11-001, 2012.
- [45] A. Massironi. Forward jets and forward-central jets at cms, dis2011 proceedings. American Institute of Physics.
- [46] M. T. Lucchini. Simultaneous production cross section of a central and a forward jet in proton-proton collisions at  $\sqrt{s} = 7$  tev. Master thesis.
- [47] V.N. Gribov and L.N. Lipatov. Sov. J Nucl. Phys., 15, 1972.
- [48] L.N. Lipatov. Sov. J Nucl. Phys., 20:94, 1975.
- [49] Yu.L. Dokshitzer. Sov. Phys., JETP 46:641, 1977.
- [50] G. Altarelli and G.Parisi. Nucl. Phys., B 126:298, 1977.
- [51] E.A. Kuraev, L.N. Lipatov and V.S. Fadin. Sov. Phys., JETP 45:199, 1977.
- [52] Ya. Ya. Balitskii and L.N. Lipatov. Sov. J. Nucl. Phys., 28:822, 1978.
- [53] C. Ewerz et al. J. Phys. G: Nucl. Part. Phys., 26:696, 2000.
- [54] Marcello Ciafaloni. Coherence Effects in Initial Jets at Small  $q^2/s$ . Nucl. Phys., B296:49, 1988.
- [55] S. Catani, F. Fiorani, and G. Marchesini. QCD Coherence in Initial State Radiation. Phys. Lett., B234:339, 1990.
- [56] S. Catani, F. Fiorani, and G. Marchesini. Small x Behavior of Initial State Radiation in Perturbative QCD. Nucl. Phys., B336:18, 1990.
- [57] Giuseppe Marchesini. QCD coherence in the structure function and associated distributions at small x. Nucl. Phys., B445:49–80, 1995.
- [58] A. Ballestrero and E. Maina. WW Scattering. arXiv:0902.0293, 2009.
- [59] Salim Cerci and David d'Enterria. Low-x QCD studies with forward jets in proton-proton collisions at  $s^{**}(1/2) = 14$ -TeV. AIP Conf. Proc., 1105:28–32, 2009.
- [60] T. Sjøstrand, S. Mrenna and P. Skands. Pythia 6.4 physics and manual. JHEP, 05:026, 2006.

- [61] R. Field. Early lhc underlying event data finding and surprises. arXiv:1010.3558v1, 2010.
- [62] T. Sjøstrand and P. Skands. Comput. Phys. Comm., 178:852, 2008.
- [63] G. Marchesini et al. Herwig: A monte carlo event generator for simulating hadron emission reactions with interfering gluons. version 5.1 - april 1991. Comput. Phys. Commun., 67:465– 508, 1992.
- [64] J.M. Butterworth et al. Multiparton interaction in photoproduction at hera. Zeit. fur Phys., C72:637-646, 1996.
- [65] M. Bahr et al. HERWIG++ Physics and Manual. Eur. Phys. J., C58:639-707, 2008.
- [66] H. Jung and G. P. Salam. Hadronic final state predictions from CCFM: The hadron- level Monte Carlo generator CASCADE. Eur. Phys. J., C19:351-360, 2001.
- [67] H. Jung, S. Baranov, M. Deak, A. Grebenyuk, F. Hautmann, M. Hentschinski, A. Knutsson, M. Krämer, K. Kutak, A. Lipatov, and N. Zotov. The CCFM Monte Carlo generator CASCADE 2.2.0. The European Physical Journal C - Particles and Fields, 70:1237–1249, 2010.
- [68] J. R. Andersen, J. M. Smillie. Constructing All-Order Corrections to Multi-Jet Rates. JHEP, 1001:039, 2010.
- [69] J. R. Andersen, J. M. Smillie. Multiple Jets at the LHC with High Energy Jets. arXiv:1101.5394, 2010.
- [70] S. Alioli et al. Jet pair production in POWHEG. 2010.
- [71] CMS Collaboration. Tracking and vertexing results from first collisions. CMS Physics Analysis Summary, TRK-10-001, 2010.
- [72] CMS Collaboration. Inclusive forward jet production cross sections in proton $\widetilde{A}$  R proton collisions at  $\sqrt{s} = 7$  tev. CMS Physics Analysis Summary, FWD-10-003, 2010.
- [73] CMS Collaboration. Calorimeter jet quality criteria for the first cms collision data. CMS Physics Analysis Summary, JME-09-008, 2010.
- [74] G.D. Lafferty and T.R. Wyatt. Where to stick your data points: The treatment of measurements within wide bins. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 355(2aÅŞ3):541 – 547, 1995.
- [75] Andreas Hocker and Vakhtang Kartvelishvili. SVD approach to data unfolding. Nucl.Instrum.Meth., A372:469–481, 1996.
- [76] G. D'Agostini. Improved iterative Bayesian unfolding. ArXiv e-prints, October 2010.
- [77] CMS Collaboration. Determination of the jet energy scale in cms with pp collisions at  $\sqrt{s} = 7$  tev. CMS Physics Analysis Summary, JME-10-010, 2010.
- [78] CMS Collaboration. Luminosity estimation for 2010 data update. CMS Physics Analysis Summary, EWK-11-001, 2011.
- [79] Measurement of the inclusive production cross sections for forward jets and for dijet events with one forward and one central jet in pp collisions at  $\sqrt{s} = 7$  tev. oai:cds.cern.ch:1421821. J. High Energy Phys., 06(arXiv:1202.0704. CMS-FWD-11-002. CERN-PH-EP-2011-179):036. 37 p, Feb 2012. Comments: Submitted to the Journal of

High Energy Physics.

- [80] Simone Alioli, Jeppe R. Andersen, Carlo Oleari, Emanuele Re, and Jennifer M. Smillie. Probing higher-order corrections in dijet production at the LHC. *Phys. Rev.*, D85:114034, 2012.
- [81] Michal Deak, Francesco Hautmann, Hannes Jung, and Krzysztof Kutak. Matrix Elements with Vetoes in the CASCADE Monte Carlo Event Generator. 2012.
- [82] Jennifer M. Smillie. The High Energy Jets Framework. 2012.
- [83] S. L. Glashow. Partial symmetries of weak interactions. Nucl. Phys., 22:579–588, 1961.
- [84] S. Weinberg. A model of leptons. Phys. Rev. Lett., 19:1264–1266, 1967.
- [85] Abdus Salam. Weak and electromagnetic interactions. In Nils Svartholm, editor, *Elementary particle physics: relativistic groups and analyticity*, page 367. Almquvist & Wiskell, 1968. Proceedings of the eighth Nobel symposium.
- [86] F. Englert and R. Brout. Broken symmetries and the masses of gauge bosons. Phys. Rev. Lett., 13:321–323, 1964.
- [87] P. W. Higgs. Broken symmetries, massless particles and gauge fields. Phys. Rev. Lett., 12:132-133, 1964.
- [88] P. W. Higgs. Broken symmetry and the mass of gauge vector mesons. Phys. Rev. Lett., 13:508, 1964.
- [89] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, 1964.
- [90] P. W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156-1163, 1966.
- [91] T. W. B. Kibble. Symmetry breaking in non-Abelian gauge theories. Phys. Rev., 155:1554– 1561, 1967.
- [92] CMS. Search for the standard model higgs boson decaying to in the fully leptonic final state in pp collisions at. *Physics Letters B*, 710(1):91 113, 2012.
- [93] CMS. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. oai:cds.cern.ch:1471031. *Phys. Lett. B*, 716(arXiv:1207.7214. CERN-PH-EP-2012-218):1-29. 39 p, Aug 2012. Comments: 24 pages plus author list (38 pages total), 12 figures, 7 tables, revised author list.
- [94] S. Frixione, P. Nason and C. Oleari. Matching nlo qcd computations with parton shower simulations: the powheg method. arXiv:0709.2092v1, 2007.
- [95] J. Alwall et al. Madgraph. JHEP, 0709:028, 2007.
- [96] T. Binoth, M. Ciccolini, N. Kauer and M. Kramer. Gluon-induced w-boson pair production at the lhc. JHEP0612:046, 2006.
- [97] Hung-Liang Lai et al. Uncertainty induced by QCD coupling in the CTEQ-TEA global analysis of parton distributions. 2010.
- [98] Huang-Liang Lai, Marco Guzzi, Joey Huston, Zhao Zi, Pavel Nadolsky, Jon Pumplin, and C.-P. Yuan. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

- [99] S. Agostinelli et al. GEANT4: A simulation toolkit. Nucl. Instrum. Meth., A506:250, 2003.
- [100] M. Mohammadi Najafabadi. Single top production at LHC. 2006.
- [101] Pietro Govoni and Chiara Mariotti. Z Production via Vector Boson Fusion at LHC. 2010.
- [102] Measurement of the electroweak production cross section of the z boson with two forwardbackward jets in pp collisions at 7 tev. 2012.
- [103] . Description and performance of the cms track and primary vertex reconstruction. CMS AN, CMS-AN-2011-172.
- [104] . Search for higgs boson decays to two w bosons in the fully leptonic final state with full 2011 pp dataset at  $\sqrt{s} = 7$  tev. CMS AN, CMS-AN-2012-432.
- [105] CMS Collaboration. Search for the higgs boson in the fully leptonic  $w^+w^-$  final state. CMS-PAS-HIG-2011-024, 2011.
- [106] Study of photon conversion rejection at CMS. CMS AN-2009/159, 2010.
- [107] Jet Performance in pp Collisions at  $\sqrt{s} = 7$  TeV. CMS PAS, JME-10-003, 2010.
- [108] CMS collaboration. Commissioning of b-jet identification with pp collisions at sqrt(s) = 7TeV. BTV-10-001, 2010.
- [109] Evaluation of udsg mistags for b-tagging using negative tags.
- [110] J. Brochero et al., Ww cross section measurement in the fully leptonic final state with 4.63 fb-1 at 7 tev. CMS AN, CMS AN-2012/036.
- [111] J. Brochero et al., Measurement of the production cross section of pp->ww at 8 tev. CMS AN, CMS AN-2012/205.
- [112] A.Massironi et al., Study of associated higgs (vh) production with h > ww > lnln; and hadronic v decay with 4.9 fb-1 at 7 tev. CMS AN, CMS AN-2012/086.
- [113] CMS Collaboration. Vh with  $h \rightarrow ww \rightarrow \ell \nu \ell \nu$  and  $v \rightarrow jj$ . 2012.
- [114] Computing the contamination from fakes in leptonic final states. AN-10-261, 2010.
- [115] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. JHEP, 0206:029, 2002.
- [116] Cms luminosity based on pixel cluster counting summer 2012 update. 2012.
- [117] A. O'Hagan. Kendall's Advanced Theory of Statistics, volume Bayesian Inference, volume 2B.
- [118] Thomas Junk. Confidence level computation for combining searches with small statistics. Nucl. Instrum. Meth., A434:435-443, 1999.
- [119] CERN. oai:cds.cern.ch:411537. 1st Workshop on Confidence Limits, Geneva, 2000. CERN.
- [120] A L Read. Modified frequentist analysis of search results (the  $cl_s$  method). oai:cds.cern.ch:451614. (CERN-OPEN-2000-205), 2000.
- [121] Procedure for the lhc higgs boson search combination in summer 2011. Technical Report ATL-PHYS-PUB-2011-011, CERN, Geneva, Aug 2011.
- [122] K. Cranmer. Statistical Challenges for Searches for New Physics at the LHC. In L. Lyons

and M. Karagöz Ünel, editors, Statistical Problems in Particle Physics, Astrophysics and Cosmology, page 112, 2006.

- [123] L. Moneta, K. Cranmer, G. Schott, and W. Verkerke. The RooStats project. In Proceedings of the 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research. February 22-27, 2010, Jaipur, India., 2010.
- [124] Adam Falkowski, Stefan Pokorski, and J.P. Roberts. Modelling strong interactions and longitudinally polarized vector boson scattering. *Journal of High Energy Physics*, 2007(12):063, 2007.
- [125] Michael Dittmar and Herbert K. Dreiner.  $h0 \rightarrow W+W- \rightarrow l+l'-\nu/l \bar{\nu}/l'$  as the dominant SM Higgs search mode at the LHC for M(h0) = 155-GeV to 180- GeV. 1996.
- [126] John M. Campbell and R.K. Ellis. MCFM for the Tevatron and the LHC. Nucl. Phys. Proc. Suppl., 205-206:10–15, 2010.
- [127] CMS Collaboration. https://twiki.cern.ch/twiki/bin/view/cmspublic/hig11024twiki. 2011.
- [128] CMS Collaboration. Evidence for a particle decaying to  $w^+w^-$  in the fully leptonic final state in a standard model higgs boson search in pp collisions at  $\sqrt{s} = 8$  tev. CMS-PAS-HIG-2012-042, 2012.
- [129] A.G. Akeroyd. Higgs triplets at lep2. Phys. Lett. B, 353:519, 1995.
- [130] A.G. Akeroyd. Fermiophobic higgs bosons at the tevatron. Phys. Lett. B, 368:89, 1996.
- [131] ALEPH, DELPHI, L3, OPAL Collaborations, and The LEP Working Group for Higgs Boson Searches. Search for higgs bosons decaying into photons: Preliminary combined results using LEP data collected at energies up to 209 gev. 2001.
- [132] CDF, D0. Combined cdf and d0 upper limits on fermiophobic higgs boson production with up to 8.2 fb-1 of ppbar data. 2011.
- [133] Serguei Chatrchyan et al. Search for a fermiophobic Higgs boson in pp collisions at  $\sqrt{s} = 7$  TeV. *JHEP*, 1209:111, 2012.

### \_\_\_\_\_LIST OF FIGURES

1.1	The potential $V(\phi, \phi^{\dagger})$
1.2	Theoretical limits on the Higgs mass 12
1.3	LEP exclusion plot
1.4	Tevatron Higgs analysis
1.5	Global fit of all precision electroweak measurements
1.6	Higgs production cross-sections
1.7	Higgs production processes
1.8	Higgs Branching ratios
21	The LHC ring 20
2.1	Instantaneous luminosity by LHC 91
2.2	Integrated luminosity by LHC 21
$\frac{2.0}{2.4}$	Cross sections at LHC 23
2.4	CMS coordinate system
2.0	the CMS detector: 3D view 26
$\frac{2.0}{2.7}$	CMS trackor
2.1	CMS Tracker budget material
2.0	ECAL geometry geheme 20
2.3 2.10	ECAL geometry schema
2.10 9.11	ECAL Crystals
2.11	IICAL
2.12	$\mathbf{M} \mathbf{U} \mathbf{O} \mathbf{I} \mathbf{d} \mathbf{e} \mathbf{t} \mathbf{e} \mathbf{c} \mathbf{t} \mathbf{O} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} \mathbf{S} S$
3.1	Muon momentum resolution
3.2	Hybrid clustering
3.3	ECAL resolution
3.4	Jets resolution
$\frac{1}{3.5}$	Particle Flow jet composition
	5 1
4.1	ECAL and tracker association
4.2	The $\eta$ distribution
4.3	$\Delta \varphi^{MC}$ versus $\eta$ in Monte Carlo sample
4.4	$\Delta \eta^{MC}$ versus $\eta$ in Monte Carlo sample
4.5	Position difference
46	Position difference in EB 49

4.7 $\Delta \varphi$ and $\Delta \eta$ with 2011 data	49 50 51 52 52
5.1LHC and HERA x phase space5.2HLT for di-jet5.3 $p_{\rm T}$ distribution at detector level5.4Bin-by-bin unfolding procedure5.5Systematics di-jet distributions5.6Hadron level $p_{\rm T}$ distributions5.7Comparison with different MC	$57 \\ 58 \\ 59 \\ 60 \\ 61 \\ 62 \\ 64$
<ul> <li>6.1 Gluon Fusion Higgs production</li> <li>6.2 Vector Boson Fusion Higgs production</li> <li>6.3 Associated production of the Higgs</li> <li>6.4 Different Higgs production mechanisms and expected events</li> <li>6.5 Different Higgs production mechanisms and expected events (low mass)</li> </ul>	66 66 67 67 68
$6.6$ $t\bar{t}$ background $$ $6.7$ $t$ background $$ $6.8$ Z background $$ $6.9$ W background $$ $6.10$ WW kashmenn d	70 70 71 71 71
6.10WW background6.11ZZ and WZ background6.12 $W\gamma$ and $Z\gamma$ background6.13Number of PU distribution6.14Number of reconstructed vertexes	72 72 72 76 77
$6.15 \ Z \rightarrow \tau \tau \ decay$ $6.16 \ E_T^{miss}$ projection procedure $6.17 \ Low \ mass \ resonances$ $6.18 \ Track \ counting \ for \ b-jets$ $6.19 \ Jets \ distributions \ at \ WW \ level$	79 80 81 82 85
6.20 Lepton distributions at WW level	86 87 88 92 94
6.25 PF- $E_{\rm T}^{\rm miss}$ distribution at WW level $\gamma^* \rightarrow \ell^+ \ell^- R^{out/in}$ estimation 6.26 Embedding procedure for $Z/\gamma^* \rightarrow \tau^+ \tau^-$	94 95 96 97 97 98
6.30 The b-tag distribution in a $Z/\gamma^* \to \ell^+ \ell^-$ enriched phase space	100 102 105 106
7.3 $m_{\mathrm{T}}^{\ell\ell E_{\mathrm{T}}^{\mathrm{miss}}}$ at WW level	108

7.4	Higgs $\rightarrow$ WW decay
7.5	$\Delta \eta_{jj}$ at WW level
7.6	$m_{jj}$ at WW level
7.7	Tth $\eta$ distribution for central jet at Higgs level $\ldots \ldots \ldots$
7.8	$Z/\gamma^* \to \ell^+ \ell^- m_{\ell\ell}$ distribution at Higgs level
7.9	VBF analysis, SM Higgs upper limit 7 TeV
7.10	VBF analysis, SM Higgs upper limit 8 TeV
7.11	VBF analysis, SM Higgs upper limit
8.1	VH, $H \to WW \to \ell \nu \ell \nu$ and $V \to jj$
8.2	VH, $H \to WW \to \ell \nu \ell \nu$ and $V \to jj$ , expected events
8.3	VH analysis, invariant mass of the jets $m_{ij}$ at WW level
8.4	VH analysis, $\Delta \eta_{ij}$ at WW level
8.5	VH analysis, invariant mass of the leptons $m_{\ell\ell}$ at WW level
8.6	VH analysis, $\Delta R_{\ell\ell}$ at WW level
8.7	VH analysis, $m_{\rm T}^{\ell \ell E_{\rm T}^{\rm miss}}$ at WW level
8.8	VH analysis, $m_{\rm T}^{\ell \ell E_{\rm T}^{\rm miss}}$ at Higgs level
8.9	VH analysis, $m_{jj}$ at Higgs level
8.10	VH analysis, SM Higgs upper limit
9.1	VBF and VH analysis, SM Higgs upper limit
9.2	$H \rightarrow W^+W^-$ contributions in SM Higgs expected upper limit
9.3	$H \rightarrow W^+W^-$ combined analyses, SM Higgs upper limit
9.4	$H \rightarrow W^+W^-$ significance
9.5	Higgs combination exclusion limit for different channels
9.6	Higgs combination significance for different channels
9.7	Higgs combination discovery <i>p</i> -value
9.8	Higgs branching ratios
9.9	Higgs branching ratios FP/BR
9.10	VH analysis, FP Higgs upper limit
9.11	VBF analysis, FP Higgs upper limit
9.12	FP Higgs upper limit

### LIST OF TABLES

1.1	Fermions
2.1 2.2	CMS tracker sub-detectors space resolution
5.1	$p_{\rm T}$ central and forward jet
$\begin{array}{c} 6.1 \\ 6.2 \end{array}$	Background cross sections    69      Analysis triggers    73
$\begin{array}{c} 6.3 \\ 6.4 \end{array}$	Electron definition       74         Trigger paths used in data for studying trigger efficiencies.       74
$\begin{array}{c} 6.5 \\ 6.6 \end{array}$	Trigger paths used in data for studying fake rates. $\ldots$ $\ldots$ $\ldots$ $\ldots$ $$ 75 Number of $\sigma$ and probability $\ldots$ $\ldots$ $\ldots$ 105
7.1	WW level selections
7.2 7.3	Higgs mass dependent lepton selections       109         B-tag efficiecy for central jet       112
$7.4 \\ 7.5$	VBF top estimation       112 $B_{\ell\ell}$ for VBF       114
7.6	$Z/\gamma^* \to \ell^+ \ell^-$ for VBF
7.7	$VBF \ ee/\mu\mu \qquad \dots \qquad$
$7.9 \\ 7.10$	VBF $e\mu$ 118VBF limit120
8.1	B-tag efficiency for central jet
8.2 8.3	VH top estimation: central and forward jet
8.4	$R_{\ell\ell} \text{ for VH} \qquad 13$
8.5 8.6	$Z/\gamma^* \rightarrow \ell^+ \ell^-$ estimation for VH
8.7 8.8	Systematics    133      VH vields    134
8.9	VH yields in $ee/\mu\mu$
0.10	VII yleids ill $e\mu$

8.11	VH exclusion limit	136
9.1	Higgs searches production mechanisms	143
9.2	Significance in different channels	145