Università degli Studi di Torino
Scuola di Dottorato in Scienza ed Alta Tecnologia

SM Higgs boson search in the $H \to ZZ \to 4\ell$ channel at $\sqrt{s} = 7$ TeV with the CMS detector

Alberto Graziano

Università degli Studi di Torino
Scuola di Dottorato in Scienza ed Alta Tecnologia

Indirizzo di Fisica ed Astrofisica

SM Higgs boson search in the $H \to ZZ \to 4\ell$ channel at $\sqrt{s} = 7$ TeV with the CMS detector

Candidate: **Alberto Graziano**

Tutor: **Prof. Marco Costa**

Thesis board: **Prof. Nicola De Filippis**
Prof. Ezio Maina
Prof. Reisaburo Tanaka (opponent)

Ph.D. coordinator: **Prof. Guido Boffetta**

Date of dissertation: March 19, 2012

Ciclo XXIV

On fait la science avec des faits,
comme on fait une maison avec des pierres :
mais une accumulation de faits
n'est pas plus une science
qu'un tas de pierres n'est une maison.

Henri Poincaré (1854 - 1912)

Contents

Introduction

The Standard Model (SM) of elementary particles has been repeatedly confirmed by experimental data since it was proposed in 1973. It is a very well-established theory, in which the Higgs boson is the only still undiscovered particle. The Higgs mechanism has been proposed to describe the ElectroWeak Symmetry Breaking (EWSB) in the framework of the Standard Model. It predicts the existence of a neutral scalar particle, the Higgs boson, whose mass is a free parameter. Nevertheless, both theoretical and experimental constraints can be put on the Higgs mass.

The search for the SM Higgs boson is one of the main goals of the CMS and ATLAS experiments at the Large Hadron Collider (LHC) at CERN. The LHC started providing proton-proton collisions on 23 November 2009, at the centre-of-mass energy of $\sqrt{s} = 900$ GeV. This energy was then raised to $\sqrt{s} = 2360$ GeV and later, on 30 March 2010, to $\sqrt{s} = 7$ TeV, which has been the value adopted until the end of 2011. The total integrated luminosity that has been certified as 'good quality data' for the purpose of physics analyses amounts to $L = 4.71 \pm 0.21$ fb^{-1} so far. The peak instantaneous luminosity reached by the LHC is $\mathcal{L} \sim 3.5 \times 10^{33}$ cm^{-2}s^{-1}.

During 2012 the LHC, which is going to resume its activity after the winter break just in these days (early March), will run at $\sqrt{s} = 8$ TeV and with an instantaneous luminosity as high as $\mathcal{L} \sim 6.8 \times 10^{33}$ cm^{-2}s^{-1}, reached after gradually increasing steps. The target integrated luminosity that is expected to be available by the end of 2012 is $L = 15$ fb^{-1}. The expected significance for SM Higgs searches that can be attained from analysing so large a data set is foreseen to be above 5σ over the whole range of Higgs mass values not yet excluded (to date, $114.4 < m_H < 127$ GeV at 95% CL). Also excluding (at 95% CL) the existence of the SM Higgs boson over the whole Higgs mass range is foreseen to be possible by the end of 2012. The upcoming months will therefore be crucial for SM Higgs searches, either in case of a discovery or of an exclusion.

At the same time, the operational conditions of the LHC are challenging the physics analyses under several points of view. The very high instantaneous luminosity values will result into a high number of primary vertices for each event, which is referred to as *pile-up*. Generally speaking, the very good vertex resolution achieved at CMS has been proven to cope well with this

feature, but in some cases it is necessary to introduce off-line corrections to properly account for the effects of operating in a very crowded environment.

This thesis describes the outcome of the search for the SM Higgs boson in the $H \to ZZ^{(*)} \to 4\ell$ channel, over the full data set recorded by CMS up to the end of 2011. This decay channel is often called the *'golden'* one, because of its very clean experimental signature and of the relatively small amount of background events surviving the selection sequence.

Chapter 1 provides a concise description of the CMS detector. The different sub-systems are presented, along with some details about their performances. An overview of the theoretical framework of the Higgs boson physics is presented in Chapter 2. Also the main characteristics of Higgs boson phenomenology at the LHC are outlined in this section.

In Chapter 4 the Higgs search in the $H \to ZZ^{(*)} \to 4\ell$ is thoroughly described. Details about the real data sets and the *MonteCarlo* (MC) simulated ones are given. The cuts defining the event selection sequence are then listed and the results of applying this selection on the data and MC samples are shown. The procedures adopted to estimate the event yield for each background are reviewed, with a particular emphasis on the usage of data-driven techniques and on the assessment of the systematic uncertainties affecting the results.

Chapter 3 focuses on my main contribution to the 4ℓ analysis, which is about lepton isolation. The cut on lepton isolation represents an important step of the selection sequence, because of its high discriminating power between signal and reducible backgrounds, mostly $t\bar{t}$ and $Zb\bar{b}$. The isolation variables used in the analysis and the various parameters on which they depend are described. The performances of isolation cuts in terms of signal efficiency and background rejection power are also discussed. Since isolation observables are very sensitive to pile-up, the algorithm implemented to reduce as much as possible their dependence on the number of event primary vertices is of paramount importance and it is presented in detail in this section. Moreover, a method for estimating the systematic uncertainty on lepton isolation efficiency (the so-called *Lepton Kinematic Template Cone* method) is described.

Finally, Sec. 4.7 *ff.* includes the statistical results of Higgs boson searches, as obtained from the full 2010-2011 data set, about both discovery and exclusion. The excluded Higgs mass range at 95% CL is shown and some details about local excesses of events passing the signal selection are given. An overview on expected results from future data taking campaigns with the CMS experiment is also presented.

Chapter 1

The Compact Muon Solenoid

The CMS detector has been built to identify precisely the properties of muons, electrons, photons and jets. Neutrinos and other very weakly interacting particles escape without leaving signals: they can only be measured indirectly through the determination of missing transverse energy, which requires a hermetic detector. Therefore CMS must cover as much of the solid angle as possible. For this purpose, new forward detectors have been added to the original CMS design.
In this chapter a general description of each sub-detector composing CMS is presented.

1.1 The CMS detector

The structure of the CMS detector is sketched in Fig. 1.1. Two endcaps close a cylindrical barrel part. The overall diameter is 14.6 m, the length 21.6 m, the weight about 12500 tons. The thickness of the electromagnetic calorimeter, expressed in radiation lengths, is larger than 25 X_0, while the one of the hadronic calorimeter, expressed in interaction lengths, ranges from 7 to 11 λ_I depending on η.

The CMS cartesian coordinate system is a right-handed reference frame with the x axis pointing towards the centre of the LHC ring, the y axis pointing upwards and the z axis parallel to the beam. The origin is located at the nominal interaction point. In this reference frame, let the quadri-momentum of a particle be (E, p_x, p_y, p_z). The *longitudinal momentum* is p_z, the *transverse momentum* is $p_T = \sqrt{p_x^2 + p_y^2}$. The *rapidity* is defined as:

$$y = \frac{1}{2} \cdot \ln\left(\frac{E + p_z}{E - p_z}\right). \tag{1.1}$$

Since y is Lorentz-invariant under a boost of the centre-of-mass along the z axis, it is used to describe angular distributions in an event. If a particle is

ultra-relativistic ($p \gg m$), its rapidity can be approximated by the *pseudo-rapidity* η, defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right), \tag{1.2}$$

θ being the angle between the $+z$ semi-axis and the particle momentum vector $\vec{p}$.

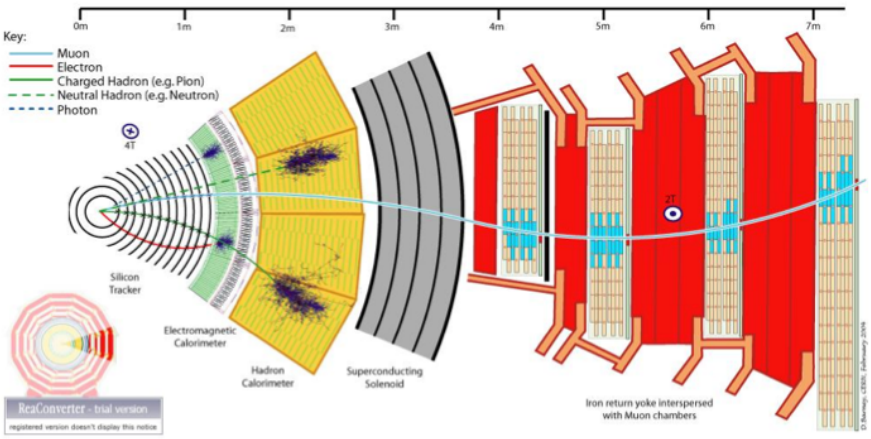


Figure 1.1: *The CMS detector transverse section.*

From the interaction point outwards, the CMS layout can be described as an 'onion-like' sequence of subdetectors. The innermost one is the Inner Tracking System, made up of a silicon pixel vertex detector and of a silicon strip tracker. The former has 3 layers in the barrel and 2 disks in each endcap, the latter consists of 10 layers in the barrel region and 12 disks in the endcaps. The main goal of the tracker is the precise reconstruction of the position of primary and secondary vertices and the accurate measurement of the impact parameter of charged tracks, This is possible even in the case of a high track multiplicity, because of the very fine segmentation of the sensors.

The electromagnetic calorimeter (ECAL) is located outside of the tracker volume. It is a homogeneous detector consisting of lead tungstate ($PbWO_4$) scintillating crystals that are read out by avalanche photo diodes or vacuum phototriodes. Its aim is the precise measurement of the energy and position of electrons and photons. The pseudo-rapidity coverage reaches $|\eta| < 3.0$. In the endcap region, a lead/silicon pre-shower detector is installed in front of ECAL in order to improve the resolution of electron and photon direction measurements and to help discriminate between photons and neutral pions.

The hadronic calorimeter (HCAL) surrounds ECAL. It is a sampling calorimeter consisting of brass absorber plates intersparsed with scintillator layers. It is designed to reconstruct energy and position of hadrons and jets.

The barrel (HB) and endcap (HE) parts have the same η coverage as ECAL, while a very forward calorimeter (HF) extends up to $|\eta| = 5.3$.

A superconducting solenoidal magnet (CB) encompasses all of these subdetectors. It is 13 m long, its diameter is 5.9 m and it provides a uniform magnet field of 3.8 T. It is cooled with liquid helium. The muon spectrometer is embedded in the iron return yoke (YB, YE) of the magnet. Drift tube (DT) detectors are installed in the barrel region, cathode strip chambers (CSC) in the endcaps, and resistive plate chambers (RPC) are placed both in the barrel and in the endcaps to increase the redundancy and robustness of the muon trigger system and to provide bunch-crossing timing measurements. The overall pseudo-rapidity coverage of the muon detectors is $|\eta| < 2.4$.

Fig. 1.1 describes the interactions of various kinds of particles with the CMS subdetectors. Muons with sufficient transverse momentum cross the entire detector and their tracks are reconstructed both in the inner tracker and in the muon chambers. They behave as minimum ionizing particles (MIPs) in the calorimeters, *i.e.* they deposit little energy (few hundreds of MeV) in them.

Electrons and photons loose all their energy in the ECAL crystals, producing electromagnetic showers. Electrons are charged particles, therefore their tracks are reconstructed in the tracker, unlike photons. Hadrons such as pions or neutrons produce large showers in the calorimeters. Unless they are very energetic, they release all their energy before reaching the muon chambers.

1.2 The Inner Tracking System

The inner tracker is designed to reconstruct tracks of charged particles with high efficiency and momentum resolution and to allow the reconstruction of secondary vertices. The trajectory of a particle with transverse momentum p_T and charge q moving through a magnetic field of strength B is a helix of radius R. These quantities are related by the following formula:

$$p_T[\text{GeV}] = 0.3 \cdot q \cdot B[\text{T}] \cdot R[\text{m}]. \tag{1.3}$$

The transverse momentum resolution is given by:

$$\frac{\delta_{p_T}}{p_T} = C_1 \cdot p_T \oplus C_2. \tag{1.4}$$

The term C_2 contains the multiple Coulomb scattering effects, whereas C_1 depends on the detector geometry, in particular on the number of points (n) used to reconstruct a track, on its length (L) and on the resolution on the

single point measurement (σ_x):

$$C_1 \propto \frac{\sigma_x}{\sqrt{n} \cdot B \cdot L^2}. \tag{1.5}$$

The C_2 term is dominant for low energy particles. C_1 is minimal for tracks made of many hits and travelling a long path within the tracker volume. The single-point resolution σ_x can be expressed as

$$\sigma_x = \sqrt{\sigma_{int}^2 + \sigma_{syst}^2}, \tag{1.6}$$

where σ_{int} is the intrinsic detector resolution and σ_{syst} is the systematic error associated to the uncertainty on the exact position of a hit module.

The tracker is composed by two subsystems: a silicon pixel detector in the innermost part, closest to the beam pipe, and a silicon strip detector of overall radius $r = 1.1$ m. The total coverage in pseudorapidity is $-2.5 < \eta < 2.5$. More details are given in the following paragraphs.

1.2.1 The pixel detector

The pixel tracker is composed of $n-$type silicon pixels laying on a $n-$type silicon substrate. The layout of the detector is shown in Fig.1.2. There are three layers in the barrel, at radii $r = 4.4$ cm, 7.3 cm, 10.2 cm, with a total length of 53.3 cm, and two disks in each endcap, at $|z| = 34.5$ cm, 46.5 cm, with radii ranging from 6 to 15 cm. The pixels are organized in modular

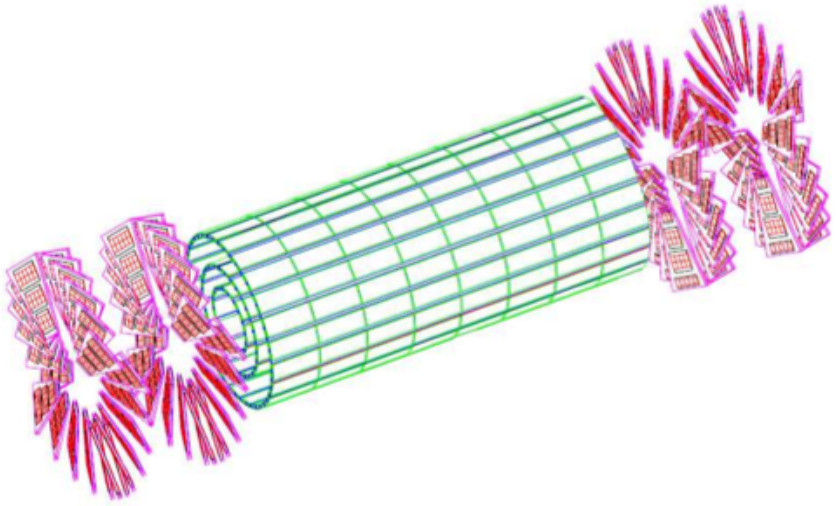


Figure 1.2: *Schematic view of the Pixel Tracker.*

detector units, called *modules*, and connected to read-out chips (ROC) by

Indium bump-bonds. The pixel size is 100 μm $\times$150 μm and the total number of pixels is 47923200 in the barrel, 17971200 in the endcaps. A spatial resolution of $\sim$ 10 μm along the $r-\phi$ coordinate and of $\sim$ 20 μm in the $r-z$ plane can be reached. It is enhanced by the fact that the charge deposited by a track is shared by several pixels, which allows to interpolate among them. Charge sharing is mostly due to Lorentz bending in the barrel and to a turbine-like layout of the modules in the endcaps.

1.2.2 The strip detector

The strip tracker consists of 10 layers of silicon microstrip detectors. Its layout is shown in Fig.1.3.

The strip detector is divided in four parts: in the barrel region, *i.e.*

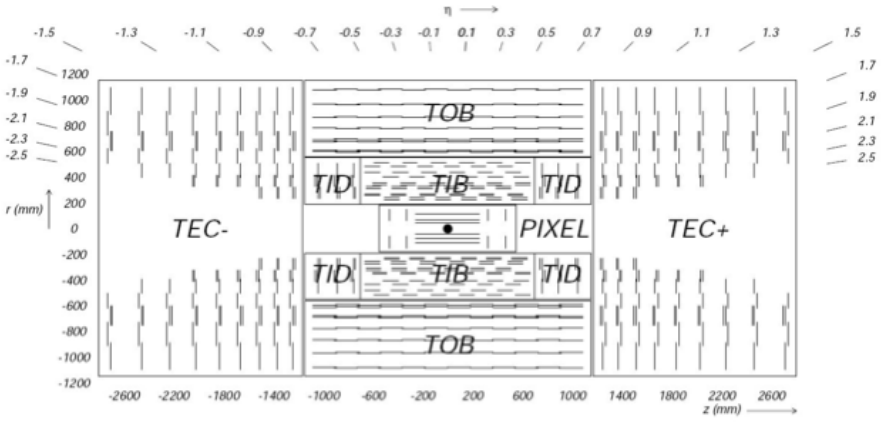


Figure 1.3: *Schematic section of the Inner Tracking System in the $r-z$ plane. The η ranges of the different subsystems are also shown.*

for $|\eta| < 1.6$, the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB), in the endcap region, *i.e.* for $1.6 < |\eta| < 2.5$, the Tracker Inner Disks (TID) and the Tracker End-Caps (TEC). The TIB is made of four cylindrical layers of *p-on-n* silicon sensors, with radii $20 < r < 60$ cm. The two innermost layers have double-sided modules: two detectors are mounted back-to-back and their strips are tilted by 100 mrad, to provide a measurement both in the $r-\phi$ and in the $r-z$ plane. The single-point resolution varies between $23-34$ μm $(r-\phi)$ and 230 μm $(r-z)$. The TOB consists of six layers. Since the radiation levels are lower in this region, thicker silicon sensors (500 μm) can be used to ensure a good signal-to-noise ratio for these longer strips. The strip pitch varies between 120 and 180 μm. The two innermost TOB layers are double-sided and provide a stereo measurement with $\sigma_x = 35 \div 52$ μm $(r-\phi)$ and $\sigma_x = 530$ μm $(r-z)$.

The TID is made of three vertical disks on each side of the barrel, filling the gap between the TIB and the TEC. The TEC consists of nine disks with z coordinates in the range 120 cm $< |z| <$ 280 cm. Both sub detectors have wedge shaped modules arranged in rings, centered on the beam line, with strips pointing towards the beam line.

1.3 The calorimeters

The calorimeters measure the energy of both neutral and charged particles. Electrons, positrons and photons deposit all their energy through electromagnetic interactions, by producing showers in the innermost calorimeter, the electromagnetic one (ECAL). On the contrary, the hadronic calorimeter (HCAL) measures the hadronic showers and the energy of jets and contributes to increase the hermeticity needed for a precise determination of missing energy. To achieve a large rapidity coverage, in addition to the barrel (HB) and the endcap (HE) parts two forward hadron calorimeters (HF) are located on each side of the detector. An outer hadron calorimeter (HO), consisting of an additional layer of scintillators, has been included in the barrel region, just outside the coil, to increase the number of interaction lengths and thus to reduce as much as possible the phenomenon of *punch-through* of hadronic showers into the muon system.

1.3.1 The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is a high-resolution, high-granularity detector composed of 75848 lead tungstate ($PbWO_4$) crystals. Lead tungstate is a fast, radiation-hard scintillator characterised by a small Molière radius (R_M = 1.9 mm) and by a short radiation length (X_0 = 8.9 mm), which allows to contain showers in the limited volume available for the calorimeter. Moreover, the scintillation decay time of the crystals is very short, which makes it possible for the electronic read-out to collect ∼80% of the light within a 25 ns time window. ECAL can achieve a good mass resolution: the detector design was optimised for standard-model Higgs boson searches at low masses ($m_H \simeq 120$ GeV), aiming for a 1% resolution on the di-photon invariant mass.

A good rejection of $\pi^0 \to \gamma\gamma$ decays, as well as the possibility of better estimating the direction of photons, is provided by a pre-shower detector in the endcap regions. It consists of two lead radiators and of two planes of silicon strip detectors, for a total radiation length of $3X_0$, and it covers the range $1.65 < |\eta| < 2.6$. The ECAL pseudo-rapidity coverage extends up to $|\eta| = 3.0$, as shown in Fig.1.4. The crystals are almost pointing to the interaction point: their axes are tilted by 3° in the barrel and by $2° \div 5°$

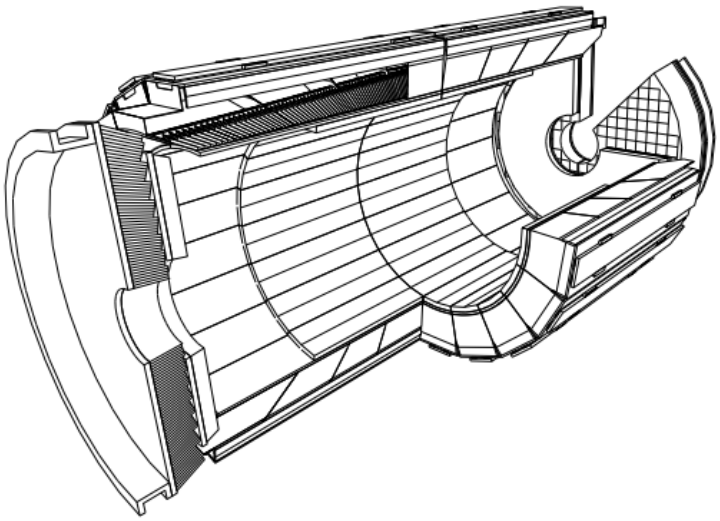


Figure 1.4: *Schematic view of the ECAL detector.*

in the endcaps with respect to straight lines originating from the nominal vertex.

The energy resolution can be parametrised as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2, \tag{1.7}$$

where a is a stochastic term including fluctuations of the number of photo-electrons and of the shower containment, b accounts for the noise due to electronics and pile-up and c is a constant term associated to energy leakage and to inter-calibration errors. The values of these three constants, measured with test beams, are listed in Table 1.1. The different contributions as a function of the energy are shown in Fig.1.5.

The crystals response depends strongly on their temperature, which must

Contribution	Barrel ($\eta = 0$)	Endcap ($\eta = 2$)
Stochastic term	2.7%	$5,7\%$
Constant term	0.55%	0.55%
Noise (low luminosity)	0.155 GeV	0.155 GeV
Noise (high luminosity)	0.210 GeV	0.245 GeV

Table 1.1: *Contributions to the energy resolution of ECAL.*

therefore be kept constant as a function of time. This is obtained with a cooling system capable of collecting the heat dissipated by the read-out

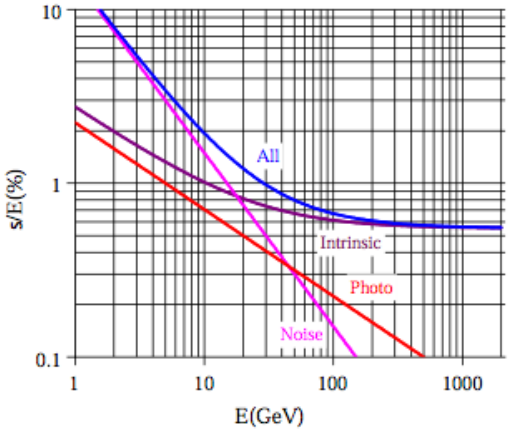


Figure 1.5: *Different contributions to the energy resolution of the ECAL as a function of the energy.*

electronics and of ensuring thermal stability. The nominal ECAL operating temperature is 18°.

1.3.2 The hadron calorimeter

The hadron calorimeter (HCAL) fulfills the task of measuring energy and direction of jets, the total transverse energy and the missing transverse energy in an event. HCAL is installed between ECAL (at a radius $r = 1.77$ m) and the magnet coil (whose inner side is at $r = 2.95$ m), as shown in Fig.1.6. The barrel and endcap parts (HB, HE) extend up to $|\eta| < 3.0$, but a total coverage of $|\eta| < 5.3$ is reached with a forward calorimeter (HF), which provides the required hermeticity. Since the full containment of a hadronic shower is not possible in the detector volume, an outer hadron calorimeter (HO), or *tail-catcher*, is placed outside of the solenoid in the barrel region. HCAL is a sampling calorimeter composed of plastic scintillators as active elements, interspersed with brass and stainless steel absorbers and read out by wavelength-shifting fibres. The dynamic energy range goes from 5 MeV to 3 TeV. The absorber material has been chosen because of its large hadronic interaction length and of its property of being non-magnetic. The HB is split into two half barrels, each containing 18 identical wedges. The HE is organised in 10 sectors, with eighteen 80 mm thick absorber layers. Both HB and HE scintillators have a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, except in the very high η-regions, where it matches the ECAL one. The HF calorimeters (situated about 11 m far from the interaction point) are useful to identify and reconstruct very forward jets. The forward region is char-

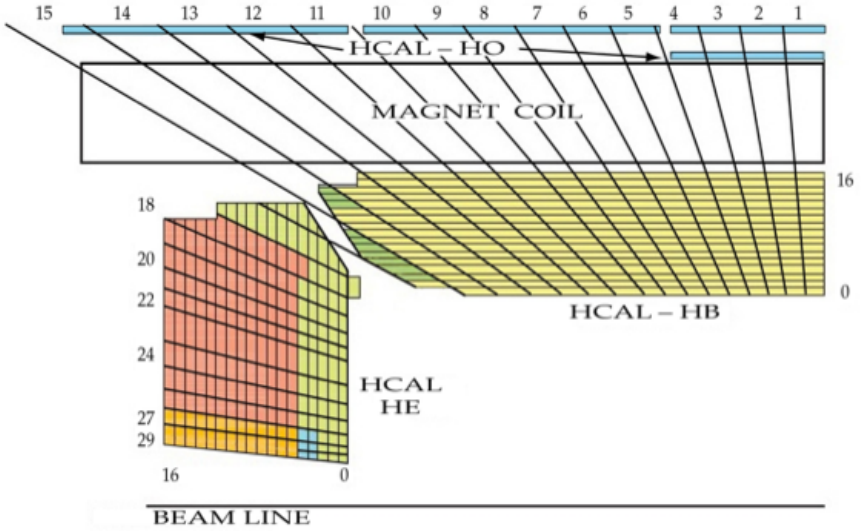


Figure 1.6: *The HCAL tower segmentation in the r,z plane for one-fourth of the HB, HO, and HE detectors. The shading represents the optical grouping of scintillator layers into different longitudinal readout.*

acterised by a high radiation field, which is best sustained by quartz fibres as active material. They emit Cherenkov light, detected by photodiodes, and they are placed between 5 mm thick steel absorber plates. The total absorber thickness of the hadronic calorimeter is summarized in the table 1.2. The HCAL depth, expressed in interaction lengths, ranges from $5.1\lambda_I$ at $\eta = 0$ to $9.1\lambda_I$ at $\eta = 1.3$, whereas it is $10.5\lambda_I$ in the endcaps. The design energy resolution is

$$\begin{aligned} \frac{\sigma_E}{E} &= 65\%\sqrt{E} \oplus 5\% \qquad \text{(HB)}, \\ \frac{\sigma_E}{E} &= 85\%\sqrt{E} \oplus 5\% \qquad \text{(HE)}, \\ \frac{\sigma_E}{E} &= 100\%\sqrt{E} \oplus 5\% \qquad \text{(HF)}, \end{aligned} \tag{1.8}$$

with E expressed in GeV.

1.4 The magnet

The CMS superconducting magnet [3] has a diameter of 6 m and a length of 12.5 m. It has been designed to generate a uniform 3.8 T field in the inner region, with a stored energy of 2.5 GJ at full current. The magnet is operated at the temperature of 4 K, as low as to make a flat $NiTb$ cable superconducting and to produce a 20 kA current not subject to any appreciable loss.

HCAL total absorber thickness	
Eta	Interaction length λ
$\eta = 0$	5.15
$\eta = 1.3$	9.1
Endcap	10.5

Table 1.2: *HCAL total absorber thickness.*

A vacuum cylinder isolates the magnet from the external environment. A 10000 t return yoke closes the B-field lines. It consists of 5 barrel layers and of 3 disks in each of the two endcaps. The ratio between stored energy and cold mass is large (11.6 KJ/kg), which causes a large mechanical deformation (0.15 %) during the ramp-up phase. The parameters of the CMS magnet are summarized in Table 1.3.

General Parameters	
Magnetic Length	12.5 m
Cold bore diameter	6.3 m
Central magnetic induction	4 T
Total Ampere-turns	41.7 MA - turns
Nominal current	19.14 kA
Inductance	14.2 H
Stored energy	2.6 GJ
Cold Mass	
Radial thickness of cold mass	312 mm
Radiation thickness of cold mass	3.9 X_0
Weight of cold mass	220 t
Maximum induction on conductor	4.6 T
Temperature margin wrt operating temperature	1.8 K
Stored energy/unit cold mass	11.6 kJ/kg
Iron Joke	
Outer diameter of the iron flats	14 m
Length of Barrel	13 m
Thickness of the iron layers in Barrel	300, 630, 630 mm
Mass of iron in Barrel	6000 t
Thickness of iron disks in Endcaps	250, 600, 600 mm
Mass of iron in each Endcap	2000 t
Total mass of iron in return yoke	10 000 t

Table 1.3: *A summary of the main features of the CMS magnet.*

1.5 The muon system

The muon spectrometer has been designed to provide an efficient muon trigger and a precise measurement of muon momentum and charge, even without relying on information from the tracking system. Muon detectors are embedded in the iron return yoke of the magnet, as shown in Fig.1.7. Other particles than muons, as well as muons with transverse momentum lower than $\simeq 5$ GeV, do not reach the muon chambers. Three subsystems compose the spectrometer.

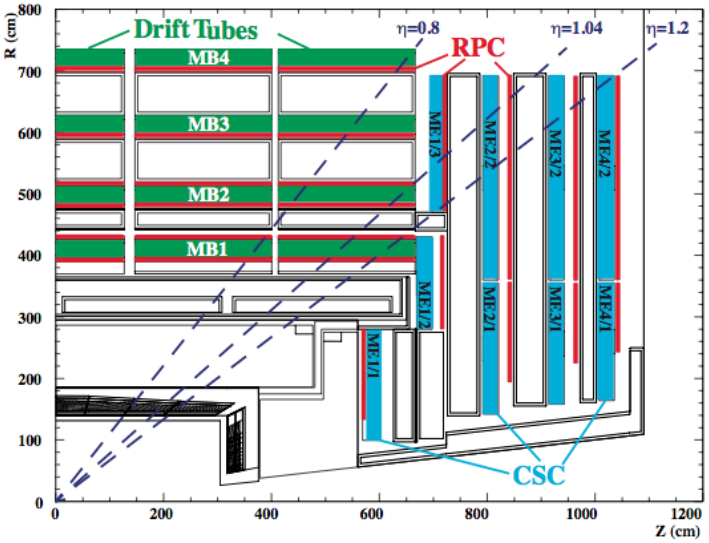


Figure 1.7: *The longitudinal view of the muon spectrometer.*

1.5.1 The Drift Tube Chambers

Drift tube (DT) chambers are located in the barrel region ($|\eta| < 1.2$), where the residual magnetic field and the track occupancy are low. The DTs are divided in five wheels along the z coordinate, each including 12 azimuthal sectors. Along the radial coordinate, four stations (MB1, MB2, MB3, MB4) are made of 12 chambers each, one per ϕ sector, except for MB4, which contains 14 chambers.

The basic constituent of a DT chamber is a *cell*, whose size is 42×13 mm^2. A cell is bounded by two parallel aluminium planes and by *'I-shaped'* aluminium beams serving as cathodes. The anodes are 50 μm stainless steel wires located in the centre of the cells. Muons passing through a cell ionise the gas mixture that fills the cell volume. The drift time of the resulting electrons is used to measure the distance between the muon track and the

wire. The linearity of relation between time and distance is enhanced by means of an additional field shaping, given by two positively biased insulated strips glued on the planes in correspondence to the wire. The gas mixture within a cell is composed of Ar (85%) and CO_2 (15%). It guarantees good quenching properties and the saturation of the drift velocity, which is ~ 5.4 cm/μs. This corresponds to a maximum drift time of ~ 390 ns, or 15 bunch crossings. The efficiency of a single cell is $\sim 99.8\%$, its spatial resolution is ~ 180 μm. Each chamber has a resolution of ~ 100 μm in the $r-\phi$ plane and of ~ 1 mrad along the ϕ coordinate.

1.5.2 The Cathode Strip Chambers

Cathode Strip Chambers (CSC) have been installed in the endcaps ($0.8 < |\eta| < 2.4$), where the residual magnetic field between the plates of the return yoke is intense and the particle rate is high. They are multi-wire proportional chambers made of two cathode planes, one of which is segmented into strips, and of an array of anode wires laying between these two planes, as shown in Fig.1.8. A charged track passing through a chamber generates an avalanche that induces a charge on several cathode strips. By interpolating among these strips one reaches a very fine spatial resolution of 50 μm. The resolution along the ϕ coordinate is 10 mrad. The CSCs consist of four stations (ME1 to ME4), the innermost one including three concentric rings, the other ones only two. The inner rings of stations ME2 to ME4 contain 18 chambers, all the other ones include 36 chambers.

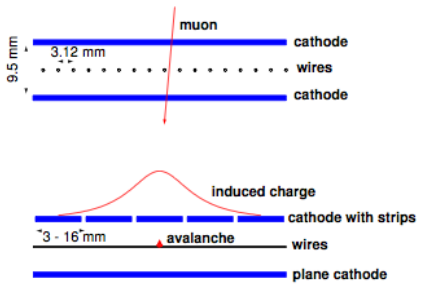


Figure 1.8: *Working principle of a CSC chamber.*

1.5.3 The Resistive Plate Chambers

Resistive Plate Chambers (RPC) are located both in the barrel and in the endcaps ($|\eta| < 2.1$), for redundancy purposes. Their spatial resolution is limited, but their time resolution is very good, about 1 ns, a shorter time than the 25 ns LHC bunch spacing. Therefore RPC detectors are used to identify unambiguously a bunch crossing and to provide prompt trigger

decisions. RPCs are made of two parallel plates of bakelite, a high-resistivity plastic material, with a few mm thick gas gap in between them and a graphite coat outside of them (see Fig.1.9). Alluminium strips, separated from the graphite layers by an insulating PET (polyethylene terephtalate) film, read out the signals. The gas mixture filling the gap consists of $C_2H_2F_4$ (95%) and of $i-C_4H_{10}$ (5%). The geometrical layout of the RPC chambers depends on their position. In the barrel region, six layers of RPCs are there: four of them are attached to each side of the MB1 and MB2 DT chambers, the other two to the inner side of MB3 and MB4. In the endcaps, four disks of trapezoidal RPC are attached to the CSCs.

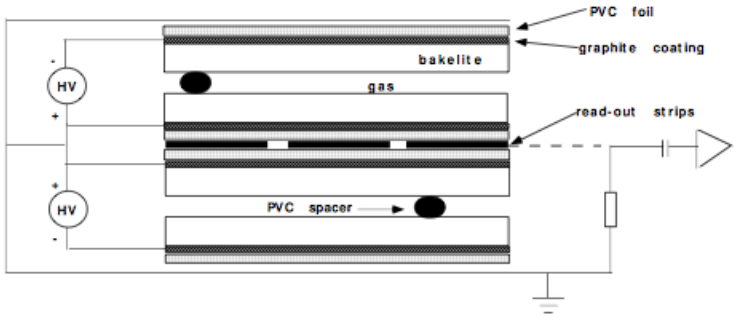


Figure 1.9: *Structure of a double-gap RPC detector.*

1.6 Forward detectors

1.6.1 CASTOR

The CASTOR [2] (Centauro And Strange Object Research) detector is a quartz-tungsten calorimeter placed 14.37 m far from the interaction point so as to extend the forward pseudo-rapidity coverage to the region $-6.6 < \eta < -5.2$. CASTOR is made of quartz plates serving as active material, interspersed with tungsten (W) absorber plates. It has a cylindrical shape with inner radius $r_{in} = 3.7$ cm and outer radius $r_{out} = 14$ cm. The inner and outer surfaces are octagonal. The working principle relies on the collection of Čerenkov photons by photomultiplier tubes, through air-core light guides. The detector is composed of a 20.12 X_0 thick electromagnetic section and of a 9.5 X_0 thick hadronic one. It is not segmented along the η direction, whereas it is divided in 16 sections along ϕ and in 14 sections along the z coordinate, the first two of which are electromagnetic, the other ones hadronic. CASTOR has been designed to allow the measurement of showers produced by particles impinging on it. The energy resolution in the acceptance range is better than 1%.

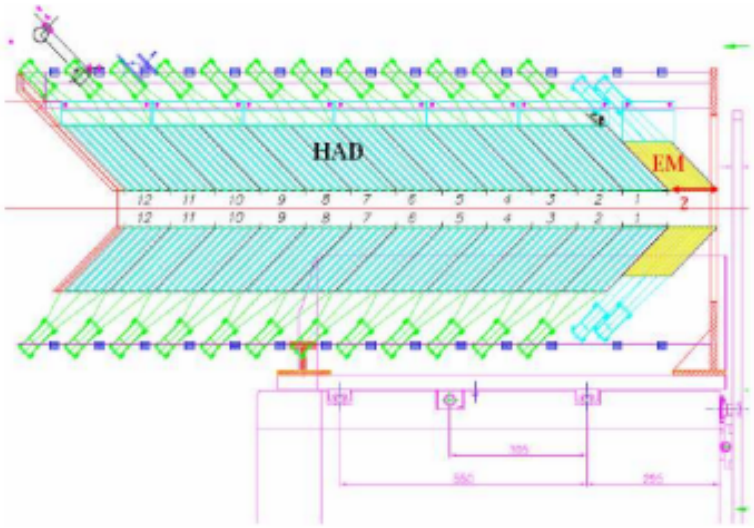


Figure 1.10: *The CASTOR detector.*

1.6.2 The Zero Degree Calorimeter

Two identical Zero Degree Calorimeters (ZDC) have been installed in the very forward pseudo-rapidity regions ($\eta \geq 8.3$), as shown in Fig.1.11. They are useful for performing diffractive analyses. The ZDCs are located between the two LHC beam pipes, on each side of the CMS interaction point, 140 m far from it. They are composed of two independent parts: an electromagnetic one (EM) and a hadronic one (HAD). The ZDC sampling calorimeters are made of tungsten and quartz fibres and they allow to reconstruct 2.75 TeV spectator neutrons with a resolution of $10 \div 15\%$.

1.7 Trigger and data acquisition

1.7.1 The Trigger System

The bunch crossing rate at the CMS interaction point is 40 MHz, but no more than 100 Hz of data can be written on tape, since a typical raw event size is 1 MB. The trigger system is designed to perform this rate reduction in an optimal way, *i.e.* by discarding low-energy processes while keeping as many interesting high-energy events as possible. Since the LHC bunch crossing time is 25 ns, the trigger decisions must be taken in a very short time. This is achieved by splitting the whole workflow in three steps or *'levels'*: L1, L2, L3. Each of them reads out and processes only a limited fraction of the available information. The level-one step [5] is totally hardware-based, whereas L2 and L3 are sets of software requirements and are usually referred

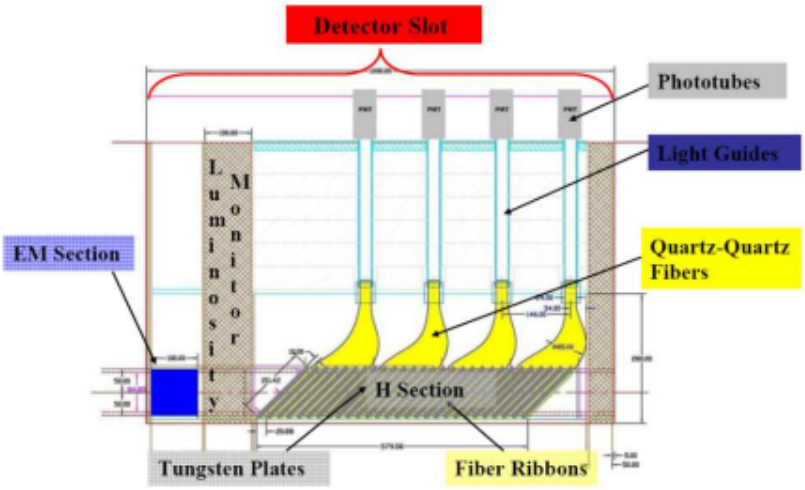


Figure 1.11: *The zero degree calorimeter.*

to as High Level Trigger (HLT) [6]. The HLT algorithm takes as input relatively few events, therefore it can analyse them in a more detailed way even if the available time is short.

1.7.1.1 The Level-1 Trigger

The L1 trigger consists of mostly custom-designed, programmable hardware capable of bringing down the event rate from the initial 40 MHz to 100 kHz. The full data content is stored in pipelines of processing elements until the trigger decision is taken. The maximum allowed latency is 3.2 μs, after which, if the L1 accepts the event, the data starts being processed by the High Level Trigger. Since it would not be possible to read out and analyse the whole information contained in an event, mostly because of the time needed by tracking algorithms, only calorimeters and muon chambers are involved in the L1 step, as shown in Fig. 1.12.

The Calorimeter Trigger finds out the four 'best' candidates of each of the following categories: electrons and photons, central jets, forward jets and jets identified on the basis of the shape of the deposited energy. These candidates are handed over to the Global Calorimeter Trigger (GCT), along with the measured missing E_T.

The Muon Trigger is performed independently by DTs, CSCs and RPCs. The DT and CSC triggers carry out a local muon reconstruction by comparing the slopes of track segments built in subsequent detector layers. The RPC trigger compares a given muon track with predefined hit patterns depending on the track p_T. The four best muon candidates are passed to the

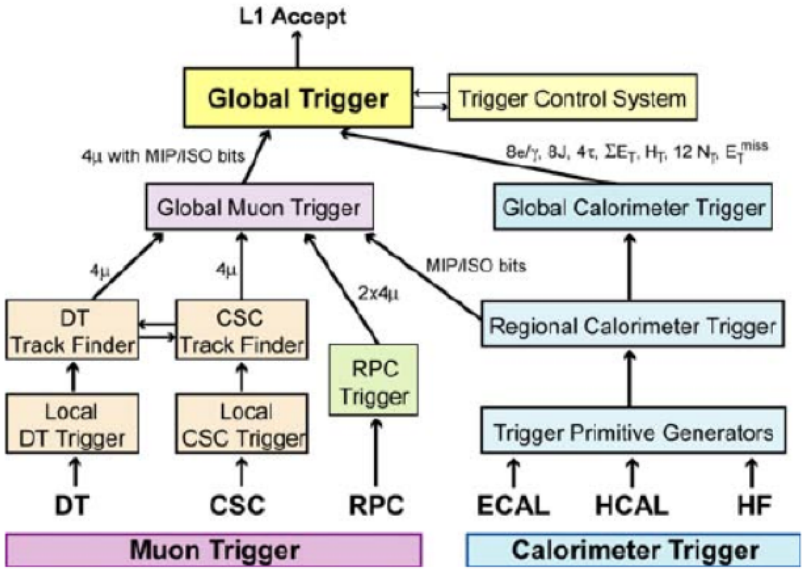


Figure 1.12: *Schematic representation of the CMS L1 trigger system.*

Global Muon Trigger system, which is in charge of matching those from DTs and CSCs with those from RPCs and of discarding low-quality tracks. The L1 electronics is installed partly directly on the detectors, partly in the underground control room about 20 m far from the experimental cavern.

1.7.1.2 The High Level Trigger

The HLT is a software system running on a farm of about 1000 commercial processors, designed to reduce the event rate down to the final output of ~ 100 Hz that can be written on tape.

The Level-2 trigger performs a stand-alone muon reconstruction in the muon chambers, using L1 tracks as seeds, and defines a region in the $\eta - \phi$ space in which a seed for a L3 muon is found. Trajectories are reconstructed with a Kalmán filter technique.

The Level-3 trigger matches a L2 muon with a tracker track to build a *Global Muon* object. Muon tracks are also associated to energy deposits in the calorimeters.

1.7.2 The Data Acquisition System

The CMS Data Acquisition (DAQ) [7] in responsible for conveying the data from about 650 read-out modules to the filter units that will process the

events. Each module provides event fragments with a size of $\sim$ 2 kB. The CMS DAQ structure is outlined in Fig. 1.13. The DAQ system of CMS is shown in Figure 1.13. The detector sensors are read out by the so-called

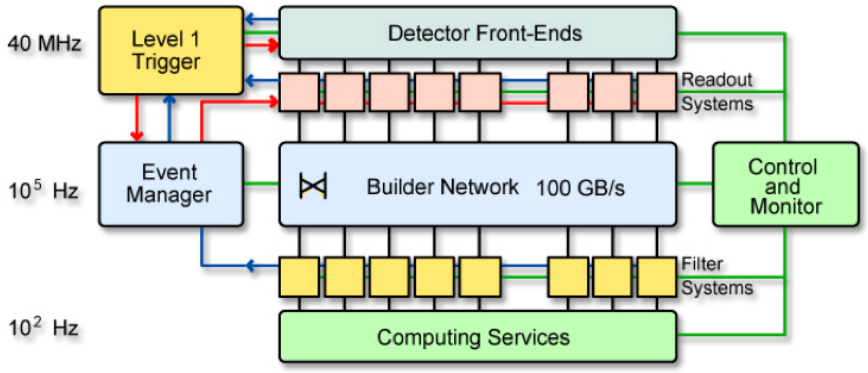


Figure 1.13: *The structure of the CMS DAQ system.*

Front-End Drivers (FEDs) through a builder network having a bisectional bandwidth of 100 GB/s. The FEDs are situated in the underground counting room, $\sim$ 70 m far from the detector. Events are passed to the event filter systems at a maximum rate of 100 kHz. This large rate, corresponding to the L1 one, is due to the design choice of building the full event already after the L1 trigger stage, unlike in the standard multi-level trigger systems.

Chapter 2

The Standard Model and the Higgs mechanism

The fundamental components of matter and their interactions are nowadays best described by the Standard Model of Particle Physics (SM), which is based on two separate quantum field theories, describing the *electroweak interaction* (Glashow-Weinberg-Salam model or GWS) and the *strong interaction* (Quantum Chromo-Dynamics or QCD). In this chapter, a short overview of the electroweak theory (Sec. 2.2) is given, focusing the attention on the *Electro Weak Symmetry Breaking* (EWSB), the Higgs mechanism and the Higgs boson (Sec. 2.3). Finally, in Sec. 2.4 the bases of the Higgs boson search are introduced.

Natural units are used in the following, unless otherwise specified, *i.e.* it will always be $\hbar = c = 1$.

2.1 The Standard Model of elementary particles

The SM [8] describes the matter as composed by twelve elementary particles, the *fermions*, all having half-integer spin. The fermions can be divided into two main groups, *leptons* and *quarks*, whose classification is given in Tab. 2.1. Quarks are subject to both strong and electroweak interactions and do not exist as free states, but only as constituents of a wide class of particles, the *hadrons*, such as protons and neutrons. Leptons, instead, only interact by electromagnetic and weak forces.

In the SM, the interactions between particles are described in terms of the exchange of *bosons*, integer-spin particles which are carriers of the fundamental interactions. The main characteristics of bosons and corresponding interactions are summarised in Tab. 2.2 (the *gravitational* interaction is not taken into account, as it is not relevant at the scales of mass and distance typical of the particle physics).

As previously mentioned, the SM describes these interactions by means

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

Table 2.1: *Classification of the three families of fundamental fermions.*

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

Table 2.2: *Fundamental interactions relevant in particle physics and corresponding carriers.*

of two gauge theories: the Quantum Chromo-Dynamics and the theory of the electroweak interaction (Glashow-Weinberg-Salam model), which unifies the electromagnetic and weak interactions. Since the present work deals with a purely electroweak decay, in the next sections only the latter theory will be described in some detail.

2.2 The electroweak theory

From a historical point of view, the starting point for the study of electroweak interactions is Fermi's theory of muon decay [9], which is based on an effective four-fermion Lagrangian[1]:

$$\mathcal{L} = -\frac{4G_F}{\sqrt{2}} \bar{\nu}_\mu \gamma^\alpha \frac{1-\gamma_5}{2} \mu \bar{e} \gamma_\alpha \frac{1-\gamma_5}{2} \nu_e \,, \tag{2.1}$$

with $G_F \simeq 1.16639 \times 10^{-5}\, GeV^{-2}$. Eq. 2.1 represents a "point like" interaction, with only one vertex and without any intermediate boson exchanged. It is usually referred to as a $V - A$ interaction, being formed by a *vectorial* and an *axial* component. The term $\frac{1}{2}(1 - \gamma_5)$ that appears in it is the negative helicity projector. Only negative helicity (*left-handed*) component of fermions takes part to this interaction.

[1] The same formalism can also be used to treat β decays, starting from a Lagrangian similar to Eq. 2.1.

Fermi's Lagrangian is not renormalisable and it results in a non-unitary $\mathcal{S}$ matrix. Both renormalisability and unitarity problems can be overcome by describing the weak interaction with a *gauge* theory, i.e. requiring its Lagrangian to be invariant under local transformations generated by the elements of some Lie group (*gauge transformations*). The specific group of local invariance (*gauge group*) is to be determined by the phenomenological properties of the interaction and of the particles involved. In particular, the resulting Lagrangian must reduce to Eq. 2.1 in the low energy limit. A detailed derivation of this Lagrangian is not provided here, but the results are summarised in the following (for details about the GWS model, see [10] [11] [12]).

A gauge theory for weak interactions is conceived as an extension of the theory of electromagnetic interaction, the Quantum Electro-Dynamics or QED, which is based on the gauge group $U(1)_{EM}$, associated to the conserved quantum number Q (*electric charge*). In this case, the condition of local invariance under the $U(1)_{EM}$ group leads to the existence of a massless vector field, the *photon*.

A theory reproducing both the electromagnetic and weak interaction phenomenology is achieved by extending the gauge symmetry to the group $SU(2)_I \otimes U(1)_Y$ (in this sense, the weak and electromagnetic interactions are said to be unified). The generators of $SU(2)_I$ are the three components of the *weak isospin* operator, $t^a = \frac{1}{2}\tau^a$, where τ^a are the Pauli matrices. The generator of $U(1)_Y$ is the *weak hypercharge* Y operator. The corresponding quantum numbers satisfy

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

where I_3 is the third component of the weak isospin (eigenvalue of t^3).

Fermions can be divided in doublets of negative-helicity (*left-handed*) particles and singlets of positive-helicity (*right-handed*) particles, as follows:

$$L_L = \begin{pmatrix} \nu_{\ell,L} \\ \ell_L \end{pmatrix}, \quad \ell_R\,, \quad Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad u_R\,, \quad d_R\,, \tag{2.2}$$

where $\ell = e, \mu, \tau$, $\quad u = u, c, t$ and $d = d, s, b$. Neutrinos have no *right* component, as their mass is ~ 0. In Tab. 2.3, I_3, Y and Q quantum numbers of all fermions are reported.

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

$$\mathcal{L} = \overline{\Psi}_L \gamma^\mu \left(i\partial_\mu + g t_a W_\mu^a - \tfrac{1}{2} g' Y B_\mu \right) \Psi_L + \overline{\psi}_R \gamma^\mu \left(i\partial_\mu - \tfrac{1}{2} g' Y B_\mu \right) \psi_R \tag{2.3}$$

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

Table 2.3: *Isospin (I_3), hypercharge (Y) and electric charge (Q) of all fermions.*

where

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

and where Ψ_L and ψ_R are summed over all the possibilities in Eq. 2.2.

As already stated, $W_\mu^{1,2,3}$ and B_μ do not represent physical fields, which are given instead by linear combinations of the four mentioned fields: the charged bosons W^+ and W^- correspond to[2]

$$W_\mu^\pm = \sqrt{\frac{1}{2}}(W_\mu^1 \mp i W_\mu^2), \tag{2.4}$$

while the neutral bosons γ and Z correspond to

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \tag{2.5}$$

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W, \tag{2.6}$$

obtained by mixing the neutral fields W_μ^3 and B_μ with a rotation defined by the *Weinberg angle* θ_W. In terms of the fields in Eqs. 2.4 and 2.6, the interaction term between gauge fields and fermions, taken from the Lagrangian in Eq. 2.3, becomes

$$\mathcal{L}_{int} = \frac{1}{2\sqrt{2}} g (J_\alpha^+ W^{(+)\alpha} + J_\alpha^- W^{(-)\alpha}) + \frac{1}{2}\sqrt{g'^2 + g^2} J_\alpha^Z Z^\alpha - e J_\alpha^{EM} A^\alpha, \tag{2.7}$$

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

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

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

$$g \sin\theta_W = g' \cos\theta_W = e \tag{2.8}$$

[2] In the following, a different notation will be also used: $W_\mu^{(-)} = W_\mu$, $W_\mu^{(+)} = W_\mu^\dagger$.

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

2.3 The Higgs mechanism

In order to correctly reproduce the phenomenology of weak interactions, both fermion and gauge boson fields must acquire mass, in agreement with experimental results. Up to this point, however, all particles are considered massless: in the electroweak Lagrangian, in fact, a mass term for the gauge bosons would violate gauge invariance[3], which is needed to ensure the renormalisability of the theory. Masses are thus introduced with the *Higgs mechanism* [13], which allows fermions and $W^{\pm}$, Z bosons to be massive[4], while keeping the photon massless. Such mechanism is accomplished by means of a doublet of complex scalar fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix}, \tag{2.9}$$

which is introduced in the electroweak Lagrangian within the term

$$\mathcal{L}_{EWSB} = (D^\mu \phi)^\dagger (D_\mu \phi) + V(\phi^\dagger \phi), \tag{2.10}$$

where $D_\mu = \partial_\mu - igt_a W_\mu^a + \frac{i}{2} g' Y B_\mu$ is the covariant derivative. The Lagrangian in Eq. 2.10 is invariant under $SU(2)_I \otimes U(1)_Y$ transformations, since the kinetic part is written in terms of covariant derivatives and the potential V only depends on the product $\phi^\dagger\phi$. The ϕ field is characterised by the following quantum numbers:

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

Writing the potential term as follows (see also Fig. 2.1)

$$V(\phi^\dagger\phi) = -\mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \tag{2.11}$$

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

$$\phi^\dagger \phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}. \tag{2.12}$$

[3] Explicit mass terms for fermions would not violate gauge invariance, but in the GWS model the Lagrangian is also required to preserve invariance under *chirality* transformations, and this is achieved only with massless fermions.

[4] Rigorously speaking, the Higgs mechanism is only needed to explain how $W^{\pm}$, Z acquire their mass. A *fermiophobic* Higgs boson, *i.e.* not coupling to fermions, is also looked for at the LHC [60, 61].

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

$$\langle\phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v^2 = -\frac{\mu^2}{\lambda}. \tag{2.13}$$

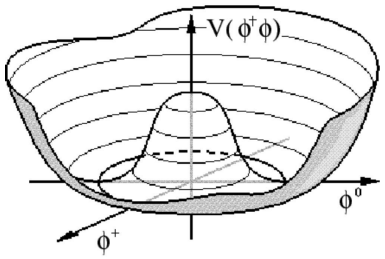


Figure 2.1: *Shape of the Higgs potential of Eq. 2.11.*

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

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

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

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

The remaining field, the *Higgs field*, has now a zero expectation value.

Rewriting the Lagrangian in Eq. 2.10 with the ϕ field in the unitary gauge, $\mathcal{L}_{EWSB}$ results from the sum of three terms:

$$\mathcal{L}_{EWSB} = \mathcal{L}_H + \mathcal{L}_{HW} + \mathcal{L}_{HZ}, \tag{2.14}$$

where the three terms can be written as follows, using the approximation

$V \sim \mu^2 H^2 + cost$ and neglecting higher order terms:

$$\begin{aligned} \mathcal{L}_H &= \frac{1}{2}\partial_\alpha H \partial^\alpha H + \mu^2 H^2 \\ \mathcal{L}_{HW} &= \frac{1}{4}v^2 g^2 W_\alpha W^{\dagger\alpha} + \frac{1}{2}v g^2 H W_\alpha W^{\dagger\alpha} \qquad (2.15) \\ &= m_W^2 W_\alpha W^{\dagger\alpha} + g_{HW} H W_\alpha W^{\dagger\alpha} \\ \mathcal{L}_{HZ} &= \frac{1}{8}v^2(g^2+g'^2) Z_\alpha Z^\alpha + \frac{1}{4}v(g^2+g'^2) H Z_\alpha Z^\alpha \qquad (2.16) \\ &= \frac{1}{2}m_Z^2 Z_\alpha Z^\alpha + \frac{1}{2}g_{HZ} H Z_\alpha Z^\alpha. \end{aligned}$$

Eqs. 2.15 and 2.16 now contain mass terms for $W^\pm$ and Z: each of the three gauge bosons has acquired mass and an additional degree of freedom, corresponding to the longitudinal polarisation. At the same time, three of the four Goldstone bosons have disappeared from the Lagrangian $\mathcal{L}_{EWSB}$, thus preserving the total number of degrees of freedom: the degrees linked to the missing Goldstone bosons have become the longitudinal degrees of the vector bosons. Only the H scalar field is still present and has acquired mass itself: it is the Higgs boson.

In summary, the Higgs mechanism is used to introduce the weak boson masses without explicitly breaking the gauge invariance and thus preserving the renormalisability of the theory. When a symmetry is "spontaneously" broken, in fact, it is not properly eliminated: it is rather "hidden" by the choice of the ground state. It can be shown that the minimum for the Higgs field is still invariant for the $U(1)_{EM}$ group: the electromagnetic symmetry is therefore unbroken and the photon remains massless.

2.3.1 Vector boson masses and couplings

Eqs. 2.15 and 2.16 show that the masses of vector bosons $W^\pm$ and Z are related to the parameter v, characteristic of the EWSB, and to the electroweak coupling constants:

$$\begin{cases} m_W = \frac{1}{2}vg \\ m_Z = \frac{1}{2}v\sqrt{g^2+g'^2} \end{cases} \quad \rightarrow \quad \frac{m_W}{m_Z} = \frac{g}{\sqrt{g^2+g'^2}} = \cos\theta_W. \qquad (2.17)$$

Also the couplings of vector bosons to the Higgs can be obtained from Eqs. 2.15 and 2.16 and are found to depend on the square of m_W and m_Z:

$$g_{HW} = \frac{1}{2}vg^2 = \frac{2}{v}m_W^2 \qquad (2.18)$$

$$g_{HZ} = \frac{1}{2}v(g^2+g'^2) = \frac{2}{v}m_Z^2. \qquad (2.19)$$

A relation between decay ratios of Higgs boson to a W pair and to a Z pair can be derived from Eqs. 2.18 and 2.19:

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

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

$$v = \left(\frac{1}{\sqrt{2}G_F}\right)^{\frac{1}{2}} \simeq 246 \text{ GeV}. \tag{2.20}$$

2.3.2 Fermion masses and couplings

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

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

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

$$\mathcal{L}_\ell = -\frac{G_{H\ell}}{\sqrt{2}} v \bar{\ell}\ell - \frac{G_{H\ell}}{\sqrt{2}} H \bar{\ell}\ell \,. \tag{2.21}$$

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

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

which becomes in the unitary gauge

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

The Yukawa Lagrangian will be therefore

$$\mathcal{L}_Y = -G_{H\ell} \overline{L}_L \phi \ell_R - G_{Hd} \overline{Q}_L \phi d_R - G_{Hu} \overline{Q}_L \phi^c u_R + h.c.\,. \tag{2.22}$$

From Eq. 2.21, the mass of a fermion (apart from neutrinos) and its coupling constant to the Higgs boson are found to be

$$m_f = \frac{G_{Hf}}{\sqrt{2}} v \tag{2.23}$$

$$g_{Hf} = \frac{G_{Hf}}{\sqrt{2}} = \frac{m_f}{v} \,. \tag{2.24}$$

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

2.3.3 Higgs boson mass

Among the 18 free parameters of the SM[5], the Higgs boson mass is the only still undetermined one. The Higgs in fact has never been observed experimentally and its mass cannot be predicted by the SM. It depends on the parameters v and λ, but while the former can be estimated by its relation with the G_F constant of Fermi's theory, the latter is characteristic of the field ϕ and cannot be determined other than measuring the Higgs mass itself. However, both theoretical and experimental constraints exist, including those from direct search at colliders, in particular LEP.

2.3.3.1 Theoretical constraints

Theoretical constraints to the Higgs mass value [14] can be found by imposing the energy scale Λ up to which the SM is valid, before the perturbation theory breaks down and non-SM phenomena emerge. The upper limit is obtained by requiring that the running quartic coupling of Higgs potential λ remains finite up to the scale Λ (*triviality*). A lower limit is found instead by requiring that λ remains positive after the inclusion of radiative corrections, at least up to Λ: this implies that the Higgs potential is bounded from below, i.e. the minimum of such potential is an absolute minimum (*vacuum stability*). A looser constraint is found by requiring such minimum to be local, instead of absolute (*metastability*). These theoretical bounds on the Higgs mass as a function of Λ are shown in Fig. 2.2.

If the validity of the SM is assumed up to the Planck scale ($\Lambda \sim 10^{19}$ GeV), the allowed Higgs mass range is between 130 and 190 GeV, while for $\Lambda \sim$ 1 TeV the Higgs mass can be up to 700 GeV. On the basis of these results, however, colliders should look for the Higgs boson up to masses of ~ 1 TeV. If the Higgs particle is not found in this mass range, then a more sophisticated explanation for the EWSB mechanism will be needed.

[5]They are: 9 fermion masses (+ 3 neutrino masses, if $m_\nu \neq 0$), 3 CKM mixing angles + 1 phase (+ 3 more angles + 1 additional phase for neutrinos), the electromagnetic coupling constant α_{EM}, the strong coupling constant α_S, the weak coupling constant G_F, the Z boson mass and the Higgs boson mass.

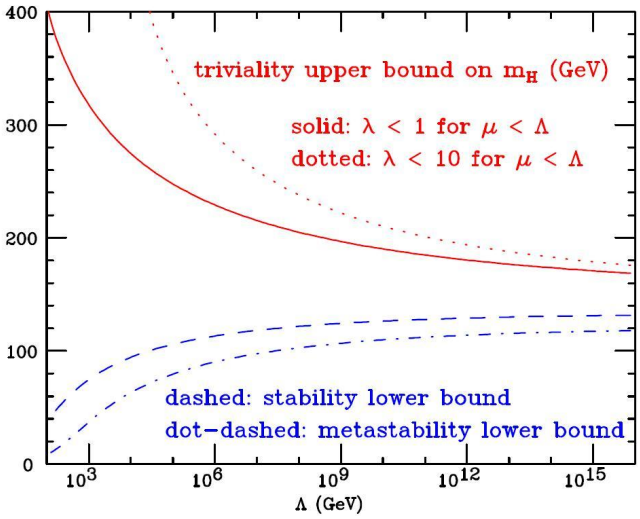


Figure 2.2: *Red line: triviality bound (for different upper limits to λ); blue line: vacuum stability (or metastability) bound on the Higgs boson mass as a function of the new physics (or cut-off) scale Λ [14].*

2.3.3.2 Experimental constraints

Bounds on the Higgs mass are also provided by measurement at LEP, SLC and Tevatron [15] (updated at July 2007). Direct searches at LEP-II have set the limit $m_H > 114.4$ GeV (95% C.L.) [16]. Moreover, since the Higgs boson contributes to radiative corrections, many electroweak observables are logarithmically sensitive to m_H and can thus be used to constrain its mass. All the precision electroweak measurements performed by the four LEP experiments and by SLD, CDF and DØ (nowadays updated with the LHC measurements) have been combined together and fitted [17], assuming the SM as the correct theory and using the Higgs mass as free parameter. The result of this procedure is summarised in Fig. 2.3, where $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ is plotted as a function of m_H. The solid curve is the result of the fit, while the shaded band represents the theoretical uncertainty due to unknown higher order corrections.

The indirectly measured value of the Higgs boson mass, corresponding to the minimum of the curve, is $m_H = 76^{+33}_{-24}$ GeV (68% C.L. for the black line in Fig. 2.3, thus not taking the theoretical uncertainty into account). An upper limit of 144 GeV can also be set (one-sided 95% C.L., including the theoretical uncertainty; this limit increases to 182 GeV when including the direct search limit of 114.4 GeV).

Such results are obviously model-dependent, as the loop corrections take into account only contributions from known physics. This result is thus well-grounded only within the SM theory and has always to be confirmed by the

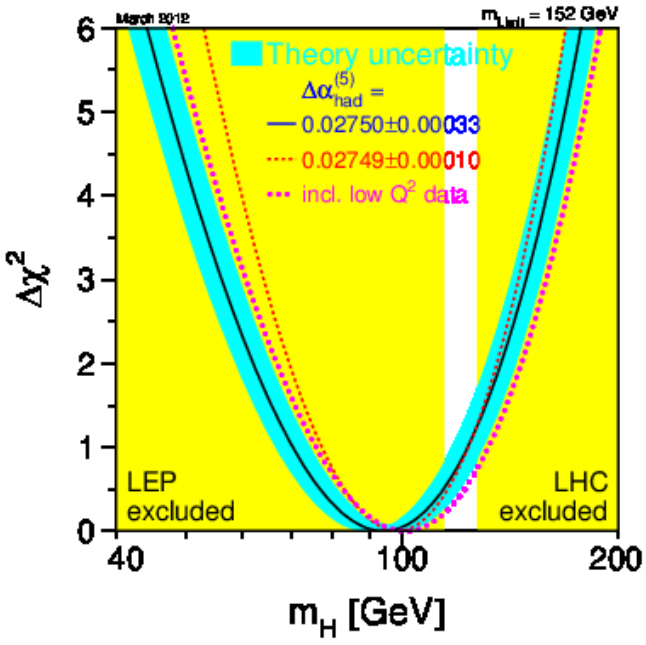


Figure 2.3: $\Delta\chi^2$ *of the fit to the electroweak precision measurements of LEP, SLC, Tevatron and LHC as a function of the Higgs mass (December 2011). The solid line represents the result of the fit and the blue shaded band is the theoretical error from unknown higher-order corrections. The yellow area represents the region excluded by direct search.*

direct observation of the Higgs boson.

2.4 Higgs boson search at the LHC

The experiments at the LHC are searching for the SM Higgs boson within the mass range from 100 to 600 GeV. The results obtained so far will be shown in Sec. 4.7 *ff.* In this section, the main Higgs production and decay processes are described, in order to discuss what are the most promising channels to look at for the Higgs discovery.

While the Higgs mass is not predicted by the theory, the Higgs couplings to the fermions or bosons are predicted to be proportional to the corresponding particle masses (for fermions) or squared masses (for bosons), as in Eqs. 2.18, 2.19 and 2.24. For this reason, the Higgs production and decay processes are dominated by channels involving the coupling of Higgs to heavy particles, mainly to $W^{\pm}$ and Z bosons and to the third generation fermions. For what concerns the remaining gauge bosons, the Higgs does not couple to photons and gluons at tree level, but only by one-loop graphs where the main contribution is given by $q\bar{q}$ and by W^+W^- loops.

2.4.1 Higgs production

The main processes contributing to the Higgs production at a hadron collider are represented by the Feynman diagrams in Fig. 2.4 and the corresponding cross sections for a centre of mass energy $\sqrt{s} = 7$ TeV, corresponding to the value currently[6] adopted by the LHC, are shown in Fig. 2.5.

2.4.1.1 Gluon-gluon fusion

The gg fusion is the dominating mechanism for the Higgs production at the LHC over the whole Higgs mass spectrum, because of the high luminosity of gluons at the nominal centre of mass energy. The process is shown in Fig. 2.4(a), with a t-quark loop as the main contribution. Next-to-leading order QCD corrections have been found to increase the cross section for this process by a factor of ~ 2. Next-to-next-to leading order calculations are also available and show a further increase of about 10% to 30%. Other sources of uncertainty are the higher order corrections ($10 \div 20\%$ estimated) and the choice of parton density function ($\sim 10\%$).

[6] It has recently (February 13, 2012) been decided that the LHC will be operated at $\sqrt{s} = 8$ TeV during 2012 data taking.

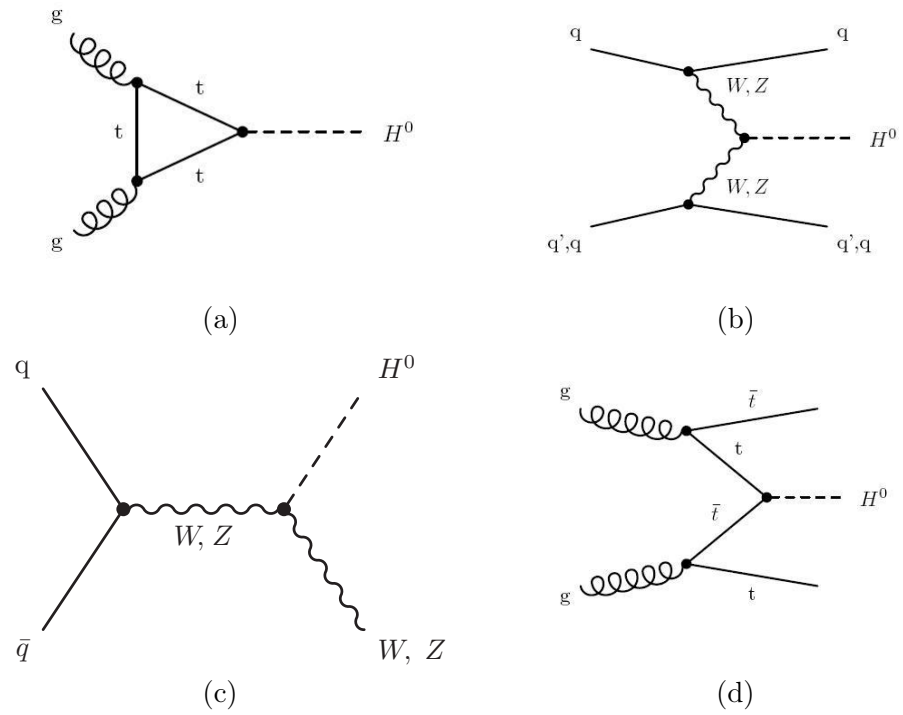


Figure 2.4: *Higgs production mechanisms at tree level in proton-proton collisions: (a) gluon-gluon fusion; (b) VV fusion; (c) W and Z associated production (or Higgsstrahlung); (d) $t\bar{t}$ associated production.*

2.4.1.2 Vector boson fusion

The VV fusion (Fig. 2.4(b)) is the second contribution to Higgs production cross section. It is about one order of magnitude lower than gg fusion for a large range of m_H values and the two processes become comparable only for very high Higgs masses ($\mathcal{O}$(1 TeV)). However, this channel is very interesting because of its clear experimental signature: the presence of two spectator jets with high invariant mass in the forward region provides a powerful tool to tag the signal events and discriminate the backgrounds, thus improving the signal to background ratio, despite the low cross section. Moreover, both leading order and next-to-leading order cross sections for this process are known with small uncertainties and the higher order QCD corrections are quite small.

2.4.1.3 Associated production

In the *Higgsstrahlung* process (Fig. 2.4(c)), the Higgs boson is produced in association with a $W^{\pm}$ or Z boson, which can be used to tag the event. The

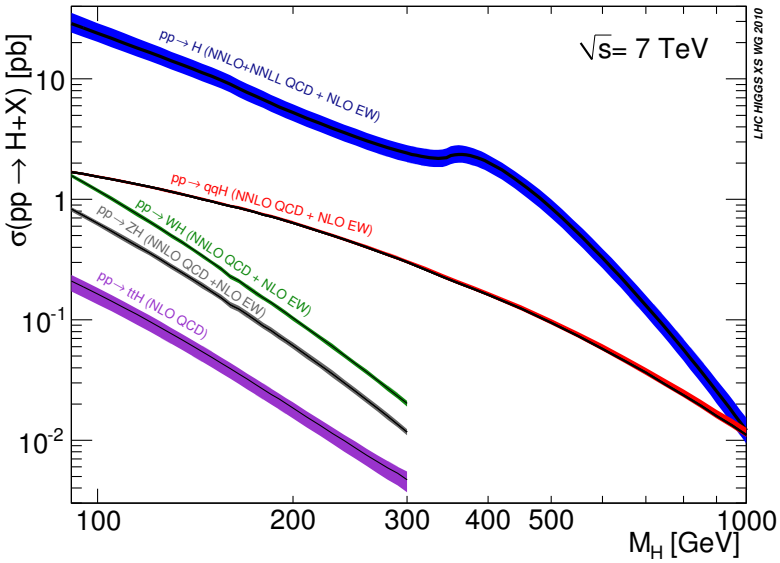


Figure 2.5: *Higgs production cross sections at* $\sqrt{s} = 7$ TeV *as a function of the Higgs mass.*

cross section for this process is several orders of magnitude lower than gg and VV fusion ones. The QCD corrections are quite large and the next-to-leading order cross section results to be increased by a factor of $1.2 \div 1.4$ with respect to the leading order one.

The last process, illustrated in Fig. 2.4(d), is the associated production of a Higgs boson with a $t\bar{t}$ pair. Also the cross section for this process is orders of magnitude lower than those of gg and VV fusion, but the presence of the $t\bar{t}$ pair in the final state can provide a good experimental signature. The higher order corrections increase the cross section of a factor of about 1.2.

2.4.2 Higgs decay

The branching ratios of the different Higgs decay channels are shown in Fig. 2.6 as a function of the Higgs mass. Fermion decay modes dominate the branching ratio in the low mass region (up to ~ 150 GeV). In particular, the $H \to b\bar{b}$ channel is the most important contribution, since the b quark is the heaviest available fermion. When the decay channels into vector boson pairs open up, they quickly dominate. A peak in the $H \to W^+W^-$ decay is visible around 160 GeV, when the production of two on-shell W's becomes possible and the production of a real ZZ pair is still not allowed. At high masses (~ 350 GeV), also $t\bar{t}$ pairs can be produced.

The most promising decay channels for the Higgs discovery do not only

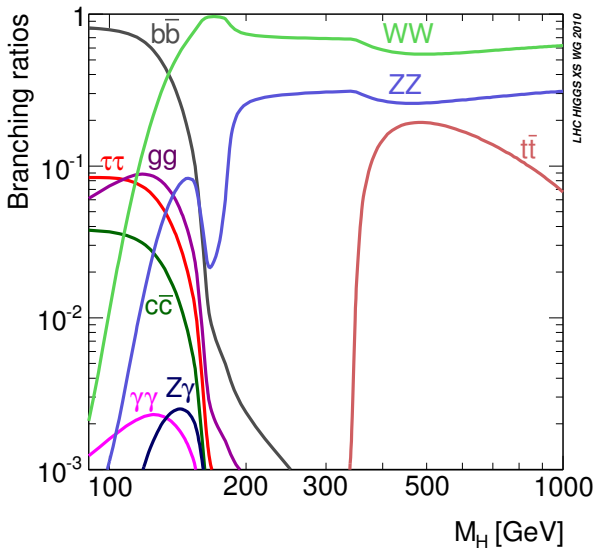


Figure 2.6: *Branching ratios for different Higgs decay channels as a function of the Higgs mass.*

depend on the corresponding branching ratios, but also on the capability of experimentally detecting the signal while rejecting the backgrounds. Such channels are illustrated in the following, depending on the Higgs mass range.

2.4.2.1 Low mass region

Though the branching ratio in this region is dominated by the Higgs decay into $b\bar{b}$, the di-jet background makes it quite difficult to use this channel for a Higgs discovery. However, the $b\bar{b}$ channel has been exploited in the *boosted* regime, in association with a vector boson decaying leptonically, by CMS [18], by ATLAS [19] and by CDF and D0 [20]. The final-state leptons allow to discriminate signal events from QCD backgrounds with only two jets.

For $m_H < 130$ GeV, instead, the channel $H \to \gamma\gamma$ seems to be the most promising. In spite of its lower branching ratio, the two high energy photons constitute a very clear signature, which only suffers from the $q\bar{q} \to \gamma\gamma$ and $Z \to e^+e^-$ backgrounds.

2.4.2.2 Intermediate mass region

For a Higgs mass value between 140 and 180 GeV, the Higgs decays into $WW^{(*)}$ and $ZZ^{(*)}$ open up and their branching ratios quickly increase, so the best channels in this mass region are $H \to WW^{(*)} \to 2\ell 2\nu$ and $H \to$

$ZZ^{(*)} \to 4\ell$.

The branching ratio of $H \to WW^{(*)}$ is higher, because of the higher coupling of the Higgs boson to charged current with respect to neutral current. Moreover, this decay mode becomes particularly important in the mass region between $2m_W$ and $2m_Z$, where the Higgs can decay into two real W's (and not yet into two real Z's): its branching ratio is ~ 1. Anyway, this channel is disfavoured because of the presence of the two ν's in the final state, which makes it impossible to reconstruct the Higgs mass. Such measurement can be performed instead when one W decays leptonically and the other one decays in two quarks. But, in this case, the final state suffers from the high hadronic background.

The decay $H \to ZZ^{(*)} \to 4\ell$, despite its lower branching ratio, offers a very clear experimental signature and high signal to background ratio. Furthermore, it allows to reconstruct the Higgs mass with high precision. Therefore, this channel has always played a major role for a Higgs discovery in this mass range.

2.4.2.3 High mass region

This region corresponds to Higgs mass values above the $2 \cdot m_Z$ threshold, where the Higgs boson can decay into a real ZZ pair. Though the $H \to ZZ$ width is still lower than $H \to WW$ one, a decay into four charged leptons (muons or electrons) has been the 'golden channel' for a high mass Higgs discovery.

2.4.2.4 Higgs total decay width

The total width of the Higgs boson resonance, which is given by the sum over all the possible decay channels, is shown in Fig. 2.7 as a function of m_H. Below the $2\, m_W$ threshold, the Higgs width is of the order of some MeV, then it rapidly increases, but remains lower than 1 GeV up to $m_H \simeq 200$ GeV: the low mass range is therefore the most challenging region, because the Higgs width is dominated by the experimental resolution.

In the high mass region ($m_H > 2\, m_Z$), the total Higgs width is dominated by the W^+W^- and ZZ partial widths, which can be written as follows:

$$\Gamma(H \to W^+W^-) = \frac{g^2}{64\pi}\frac{m_H^3}{m_W^2}\sqrt{1-x_W}\left(1 - x_W + \frac{3}{4}x_W^2\right) \quad (2.25)$$

$$\Gamma(H \to ZZ) = \frac{g^2}{128\pi}\frac{m_H^3}{m_W^2}\sqrt{1-x_Z}\left(1 - x_Z + \frac{3}{4}x_Z^2\right), \quad (2.26)$$

where

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

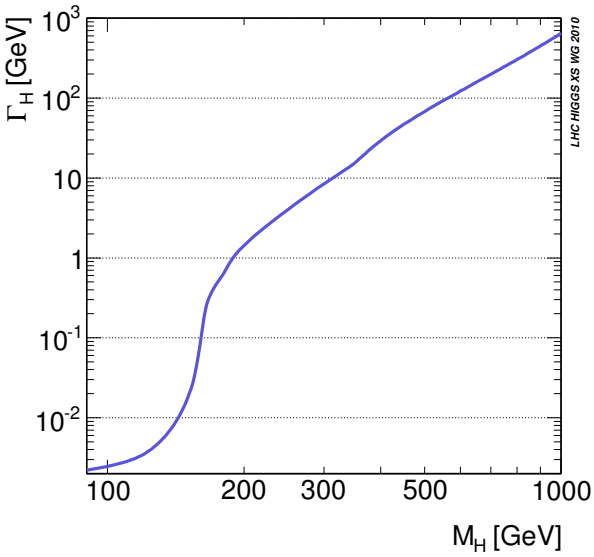


Figure 2.7: *Higgs total decay width as a function of the Higgs mass.*

As the Higgs mass grows, x_W, $x_Z \to 0$ and the leading term in Eqs. 2.25 and 2.26 grows proportional to m_H^3. Summing over the W^+W^- and ZZ channels, the Higgs width in the high mass region can be written as

$$\Gamma(H \to VV) = \frac{3}{32\pi} \frac{m_H^3}{v^2} \,. \tag{2.27}$$

From Eq. 2.27, it results that $\Gamma_H \simeq m_H$ for $m_H \simeq 1.4$ TeV. When m_H becomes larger than a TeV it becomes experimentally very problematic to separate the Higgs resonance from the VV continuum. Actually, being the resonance width larger than its own mass, the Higgs cannot be properly considered as a particle any more. In addition, if the Higgs mass is above 1 TeV, the SM predictions violate unitarity (see Sec. 2.2). All these considerations suggest the TeV as a limit for the Higgs boson mass: at the TeV scale at least, the Higgs boson must be observed, or new physics must emerge.

Chapter 3

Lepton isolation in the $H \to ZZ \to 4\ell$ analysis

3.1 Introduction to the detector-based lepton isolation

Lepton isolation is a powerful tool to determine whether a lepton is inside or very close to a jet, and therefore to discriminate between signal and background events. The underlying idea at the base of the isolation technique is that, if one considers a certain region around the track of a reconstructed lepton, one can define an observable that quantifies the amount of energy of the particles detected into this region. Then a cut is applied on this variable.

The leptons passing the isolation cut, i.e. whose isolation variable is under a given threshold, will be referred to as 'isolated' in the following.

A single-lepton isolation efficiency can be defined as the ratio of the number of isolated leptons to the total number of leptons:

$$\varepsilon_{iso} = \frac{N_\ell^{iso}}{N_\ell^{tot}}. \tag{3.1}$$

For the purposes of the $H \to ZZ \to 4\ell$ analysis it is convenient to combine the isolation variables of some leptons into a variable related to the whole event. A per-event isolation efficiency can thus be defined with respect to this observable, as the ratio of events whose isolation is less than a given threshold to the total number of events.

Since the isolation cut should discard at the same time as few signal events and as many background events as possible, the isolation algorithm must be optimised so as to achieve the best trade-off between a large signal efficiency and a great background rejection power.

The whole isolation procedure will be illustrated in the following paragraphs. The algorithm consists of several steps; a particular emphasis will

be put on the optimisation of the cut on the isolation variable, but many other parameters have been tuned to get the best performances.

3.1.1 Muon isolation

3.1.1.1 Definition of the isolation cone

The region in which lepton isolation is computed is a cone in the (η, ϕ) space[1]:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = \sqrt{(\eta - \eta_0)^2 + (\phi - \phi_0)^2}, \tag{3.2}$$

where η_0, ϕ_0 refer to the cone axis. For muons, it is the direction of the inner muon track at the vertex position, where the cone apex is set, as shown in Fig 3.1. The vertex is defined as the point of closest approach of the track to the beam line on a plane perpendicular to the beam. In the case of HCal isolation, the cone axis is defined as the direction of the centre of the tower pointed to by the muon direction at the vertex.

The cone size is a parameter that can be optimised in terms of signal efficiency and background rejection. See Section 3.2 for details. The optimal value has been found to be $\Delta R = 0.3$.

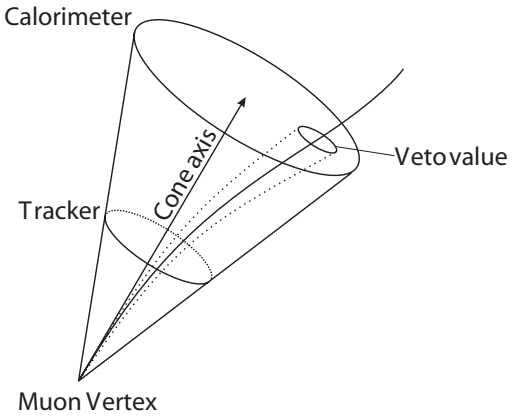


Figure 3.1: *A schematic representation of an isolation cone.*

[1] for the definition of the pseudo-rapidity η, see Eq. 1.2

3.1.1.2 Definition of the vetoes

To remove the muon footprint, a veto cone is also drawn around the muon track. The isolation variables computed in this veto cone must be subtracted from the total amount. The vetoes in the calorimeters, in the muon case, are performed after the propagation of the muon inner track to the inner calorimeter surface, to account for track bending due to the magnetic field. The optimised veto cone size values are listed in Table 3.1, together with the thresholds on the single deposits.

3.1.1.3 Muon isolation variables

- **TrkIso** The tracker isolation variable (also referred to as *sumPt* or Σp_T) is the sum of the transverse momenta of tracker tracks within the isolation cone, passing the cuts

$$p_T > 1.0 \text{ GeV}, \quad |\Delta z| < 0.2 \text{ cm}, \quad \Delta r < 0.1 \text{ cm}, \tag{3.3}$$

  where Δz, Δr are the minimum distances from a tracker track to the cone apex in the longitudinal and in the radial direction, respectively. These cuts help to reject poorly reconstructed tracks and the *loopers*, which are very soft tracks that produce many curls in the tracker volumes before stopping because of energy loss.

- **EcalIso** The ECAL isolation variable is the sum of the transverse energies deposited in the electromagnetic calorimeter by particles passing through it. The transverse energy is defined as

$$E_T = E \sin\theta, \tag{3.4}$$

  where θ is the angle between the beam pipe and the direction associated to the deposit.
  In order to cut off the electronic detector noise and to reduce the impact of pile-up, only the ECAL crystals with $E > 0.25$ GeV are taken into account in the sum over the deposits.

- **HcalIso** The HCAL isolation variable is defined like the ECAL one, but with respect to the hadronic calorimeter. No cuts are applied on single HCAL deposits.

- **CombIso** The 'combined' isolation variable is the linear sum of the tracker, ECAL and HCAL contributions for each lepton:

$$CombIso = TrkIso + EcalIso + HcalIso. \tag{3.5}$$

  This variable has been proven to have comparable, if not slightly better, performances in terms of signal efficiency and background rejection

with respect to more complicated definitions[2], as discussed in Section 3.2.

3.1.2 Electron isolation

3.1.2.1 Definition of the isolation cone

The electron isolation cone in the $H \to ZZ \to 4\ell$ analysis is defined in the same way, and with the same radius, as for muons, although for electrons the cone axis is the direction of the ECAL supercluster centroid as seen from the electron vertex, taken at $(0, 0, 0)$.

3.1.2.2 Definition of the vetoes

The veto cones in the electron case are defined by requirements on both $\Delta\eta$ and ΔR with respect to the direction of the electron. The $\Delta\eta$ band is used to account for *bremsstrahlung* photons emitted by electrons. The combination of a stripe and a cone in ECAL, as shown in Fig. 3.2, is sometimes referred to as *'Jurassic'* veto. Cuts are applied on the number of crystals: for a detailed discussion of this choice and of the related performances, see [22]. The tracker vetoes only consist of a ΔR region as for muons, while the HCAL vetoes are $\Delta\eta$ bands. See Table 3.1 for details.

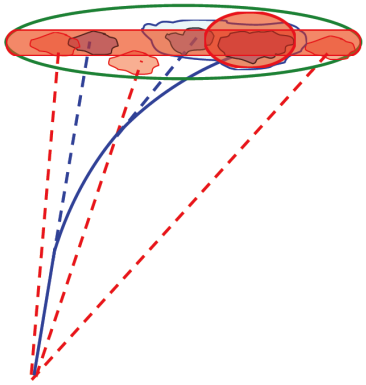


Figure 3.2: *The so-called 'Jurassic veto' used for electron isolation.*

3.1.2.3 Electron isolation variables

The isolation variables used for electrons are the same as for muons, but with slightly different cuts on the single contributions that are summed up within the cone.

[2] also combinations of TrkIso, EcalIso, HcalIso with weights different from (1, 1, 1) have been tried. See *e.g.* [21]

- **TrkIso** Only the tracks with $p_T > 0.7$ GeV and $|\Delta z| < 0.2$ cm are considered in the sum of transverse momenta.
- **EcalIso** The thresholds on single ECAL deposits are different for barrel and endcaps. In the barrel region, only RecHits with measured transverse energy $E_T > 0.08$ GeV are considered. In the endcaps, the required energy is $E > 0.1$ GeV.
- **HcalIso** No thresholds on HCAL single deposits are set, like in the muon case.

Electrons				
Type	ΔR	Deposits	Veto region	Thresholds
Tracker	0.3	CTF tracks	$\|\Delta\eta\| < 0.015$	$p_T > 0.7$ GeV, $\|\Delta z\| < 0.2$ cm
ECAL	0.3	RecHits	$\|\Delta\eta\| < 1.5$ crys., $\Delta R < 1.5$ crys.	$E_T > 0.08$ GeV (EB), $E > 0.1$ GeV (EE)
HCAL	0.3	Towers	$\Delta R < 1.5$ crys.	
Muons				
Type	ΔR	Deposits	Veto region	Thresholds
Tracker	0.3	CTF tracks	$\Delta R < 0.015$	$p_T > 1.0$ GeV, $\|\Delta z\| < 0.2$ cm, $\Delta R < 0.1$ cm
ECAL	0.3	RecHits	$\Delta R < 0.07$	$E > 0.25$ GeV
HCAL	0.3	Towers	$\Delta R < 0.1$	

Table 3.1: *List of parameters for electron and muon isolation. $\Delta\eta$ is the difference in pseudo-rapidity with respect to the direction of the cone axis, ΔR is the radius of the veto cone, p_T is the transverse momentum of the tracks in the cone, E is the energy deposited in each ECAL rechit within the cone, $E_T = E \cdot \sin(\theta)$ is the transverse energy , Δz, ΔR are the minimum distances from a track to the cone apex in the longitudinal and in the radial direction, respectively.*

3.1.3 Isolation variables for the $H \to ZZ^{(*)} \to 4\ell$ analysis

The isolation variables defined in Sec. 3.1.1.3 for muons and in Sec. 3.1.2.3 for electrons can also be normalised to the p_T of the lepton for which the isolation is measured. Dividing the isolation variables by p_T^ℓ accounts for the higher multiplicity of tracker tracks and for the increased energy deposition in the calorimeters at higher energies. The relative isolation variables, obtained after dividing the absolute ones by the lepton p_T, are often referred to as **TrkRelIso**, **EcalRelIso**, **HcalRelIso** and **CombRelIso**. The definition of *CombRelIso*, both for electrons and for muons, is:

$$CombRelIso = \frac{TrkIso + EcalIso + HcalIso}{p_T^\ell}. \tag{3.6}$$

A comparison between the performances of the absolute and relative isolation variables can be found in Sec. 3.2.

In the $H \to ZZ^{(*)} \to 4\ell$ analysis, four reconstructed leptons passing quality cuts are assigned to the '4ℓ best candidate'. This assignment is performed with an algorithm matching reconstructed leptons with the four signal leptons in the final state. These leptons are then sorted by isolation and the two ones with the largest isolation (with respect to the combined relative isolation variable) are considered. They are often referred to as the 'two least isolated' leptons. Their combined relative isolation variables are summed up, thus resulting in a variable which is often labelled as $CombRelIso_{2-least}$. This variable, as explained in the following, has been proven to be the most suitable one for the 4ℓ analysis in terms of signal-background discrimination.

3.2 Optimisation of the lepton isolation algorithm

3.2.1 Isolation efficiency vs threshold

From Eq. 3.1 one can see that the isolation efficiency varies as a function of the threshold value. The higher the threshold, the higher the efficiency, both for the signal and for the $Zb\bar{b}, t\bar{t}$ backgrounds, because more leptons will have their isolation value below that threshold. Figure 3.4 shows the trend of the isolation efficiency as a function of the threshold for signal ($m_H = 120$ GeV), $Zb\bar{b}$ and $t\bar{t}$. These plots show the integral of the distributions of Fig. 3.3 up to the threshold value divided by the integral of the distributions over their whole range.

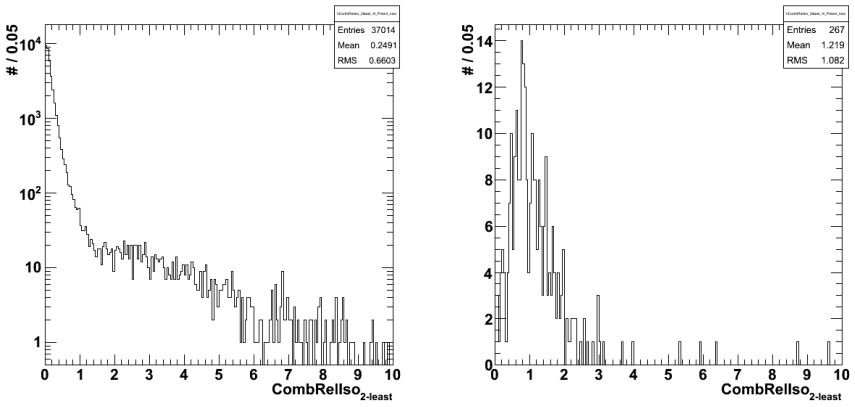
It is already clear from a comparison of these curves that an appropriate choice of the cut can provide a very high signal efficiency as well as a very high background rejection. However, the trade-off between these two needs may vary. Sometimes one can accept a tiny loss of signal efficiency, provided that it results into a considerably lower background one. More about that will be discussed in Sec. 3.2.2.

Since the efficiencies follow a binomial distribution, the statistical error on them is evaluated as:

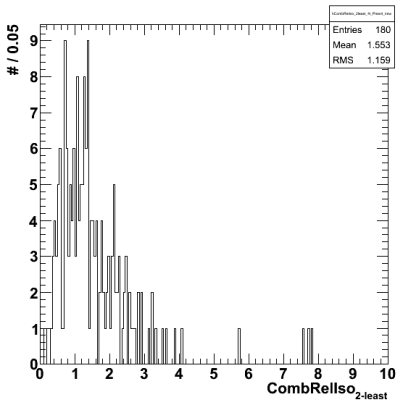
$$\sigma_\varepsilon = \frac{\sqrt{(\sum w_i^2)\varepsilon(1-\varepsilon)}}{\sum w_i}, \tag{3.7}$$

where w_i is the weight of the i-th event. If all of the events of a given data sample have the same weight $w_i = 1$, then $\sum w_i$ is the total number of events and Eq. 3.7 boils down to Eq. 3.8:

$$\sigma_\varepsilon = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N_{tot}^{evt}}}. \tag{3.8}$$



(a) $m_H = 120$ GeV (b) $Zb\bar{b}$



(c) $t\bar{t}$

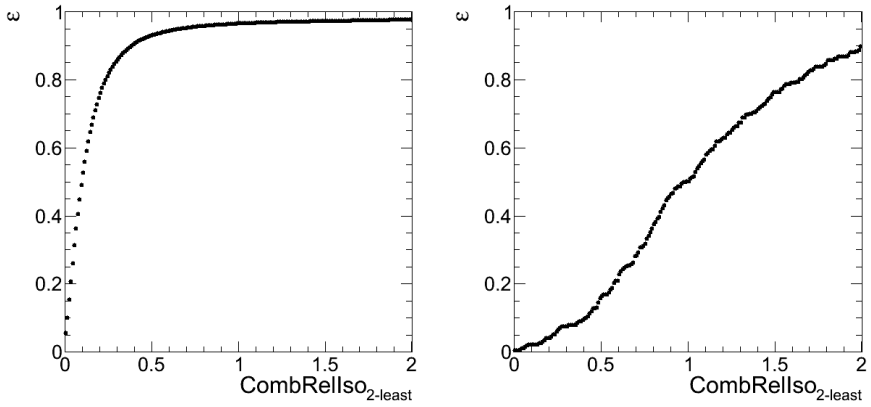
Figure 3.3: *Distributions of the $CombRelIso_{2-least}$ isolation variable, after the 'ρ correction' (see Section 3.3.2), as obtained from MC samples, for (a) signal, $m_H = 120$ GeV, (b) $Zb\bar{b}/Zc\bar{c}$ (+ 0, 1, 2, 3 jets) background, (c) $t\bar{t}$ background. The plots refer to the events surviving the $H \to ZZ \to 4\ell$ preselection, up to the choice of the best 4ℓ candidate included.*

3.2.2 The *'ROC curves'*

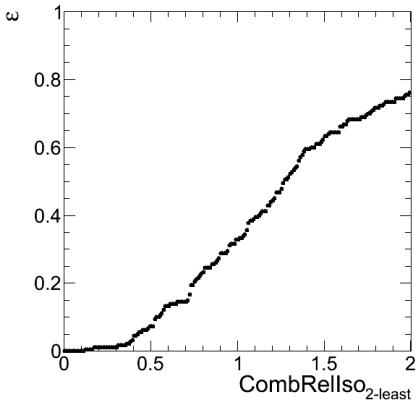
The best choice of the isolation threshold is the one that provides at the same time the highest signal efficiency and the lowest background one. A very clear and useful plot for this purpose is the so-called *'ROC curve'* [3], which shows the signal efficiency as a function of the background one, for a given choice of the signal and background samples (see Fig. 3.5).

The ideal working point on this plane would be ($\varepsilon_{SIG} = 1$, $\varepsilon_{BKG} = 0$). The optimisation is then performed by looking for the closest point to this one

[3] 'Receiver Operating Characteristic'



(a) $m_H = 120$ GeV (b) $Zb\bar{b}$



(c) $t\bar{t}$

Figure 3.4: *Isolation efficiency curves as a function of the threshold on the $CombRelIso_{2-least}$ isolation variable, after the 'ρ correction' (see Section 3.3.2), as obtained from MC samples, for (a) signal, $m_H = 120$ GeV, (b) $Zb\bar{b}/Zc\bar{c}$ (+ 0, 1, 2, 3 jets) background, (c) $t\bar{t}$ background. The plots refer to the events surviving the $H \to ZZ \to 4\ell$ preselection, up to the choice of the best 4ℓ candidate included.*

on the curve, provided that the signal efficiency remains higher than a given value. Since the signal event yield is low for the current integrated luminosity, the requirement is to have $\varepsilon_{SIG} \gtrsim 90\%$, which, for $CombRelIso_{2-least}$ at $m_H = 120$ GeV and after the choice of the best 4ℓ candidate, for the three final states considered together, corresponds to background efficiencies as low as $\varepsilon_{t\bar{t}} = 2\%$ for the $t\bar{t}$ background and $\varepsilon_{Zb\bar{b}} = 8\%$ for the $Zb\bar{b}$ one.

The decision about the cut optimisation should take into account not only the signal and background efficiency values, but also the shape of the ROC curves and the constraint on signal efficiency. In addition, the best cut

turns out to be slightly different for $Zb\bar{b}$ and $t\bar{t}$, as well as for different m_H values, so that a trade-off between these cases must be found. Therefore, no mathematical algorithm has been defined to optimise the choice of the threshold.

A detailed description of the techniques used to choose the optimal values of all the parameters involved in lepton isolation can be found in [21].

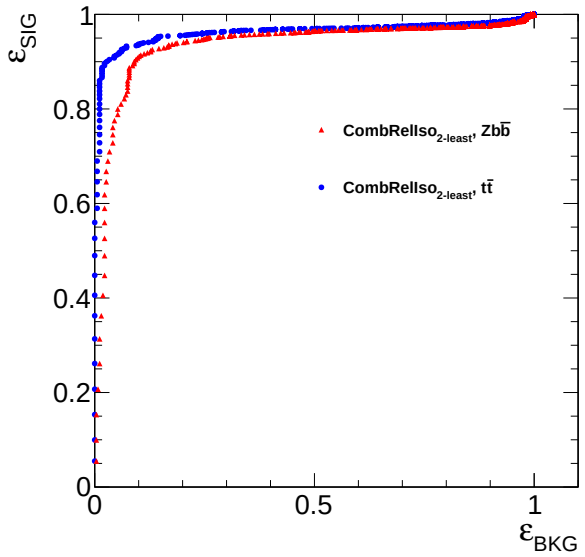


Figure 3.5: *A signal vs background efficiency curve, for the $CombRelIso_{2-least}$ variable, after the choice of the best 4ℓ candidate, for signal ($m_H = 120$ GeV) and for the $Zb\bar{b}/Zc\bar{c}$ (+ 0, 1, 2, 3 jets) and $t\bar{t}$ backgrounds, as obtained from MC samples.*

3.3 Impact of pile-up on lepton isolation

3.3.1 Pile-up conditions in CMS

Isolation variables are sensitive to contributions both from pile-up (PU) and from underlying event (UE) activity. The former consists of minimum-bias collisions that occur in the same bunch crossing as the main event (*in-time pile-up*) or in the neighbouring bunch crossings (*out-of-time pile-up* [4]). The latter refers to soft processes originating from interactions among spectator partons in the same *p-p* collision as the main event. PU events are entirely uncorrelated with the one from the hard scatter, whereas UE activity is not. However, it is a good approximation to treat both PU and UE as independent from the hard interaction [24].

[4] often referred to as 'OOT PU'.

Data	Period	Run Range	average $\mathcal{L}_{inst}$ $[cm^{-2}s^{-1}]$	$< N_{vxt} >$
2011A	1	[160404-163869]	$5 \cdot 10^{32}$	5.16
2011A	2	[165088-167913]	$1 \cdot 10^{33}$	5.15
2011A	3	[170269-173198]	$2 \cdot 10^{33}$	5.96
2011A	4	[173236-173692]	$3 \cdot 10^{33}$	6.95
2011B	5	[175832-177791]	$3 \cdot 10^{33}$	9.10

Table 3.2: *The five data-taking periods defined for Tag-and-Probe studies. The average instantaneous luminosity and number of pile-up vertices are listed for each period.*

The number of pile-up vertices in a given bunch crossing has been continuously increasing during the 2011 and 2012 data taking campaigns, along with the increasing instantaneous luminosity. Table 3.2 lists the average value of the number of reconstructed vertices in five periods, numbered from 1 to 5, in chronological order. The same information is plotted in Fig. 3.6. It is clear visible that the mean number of reconstructed vertices becomes larger in later periods, in which the average instantaneous luminosity was also higher.

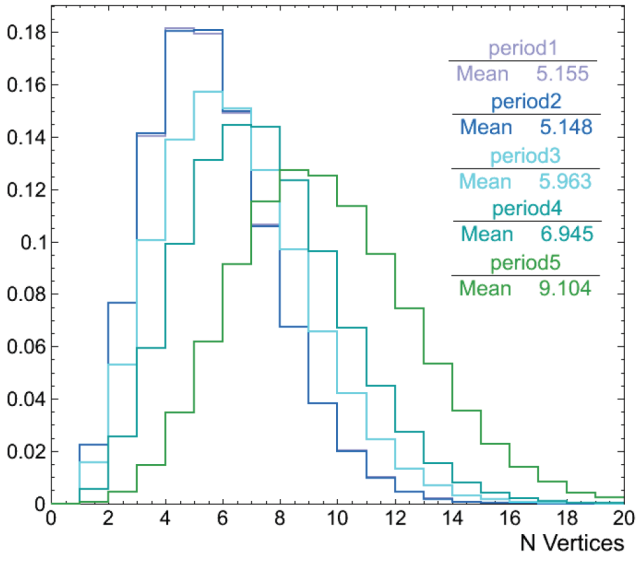


Figure 3.6: *The distribution of mean number of reconstructed primary vertices in the five data-taking periods listed in Table 3.2.*

3.3.2 The 'ρ correction'

The contribution of pile-up to the tracker-based isolation variable (*TrkIso*) can be removed, because of the very good vertex resolution of the CMS tracking system [27] and of the longitudinal spread of the beam collision region. This allows to associate very precisely a track to the vertex it comes from and thus to reject charged tracks originating from pile-up vertices. On the contrary, in the case of neutral particles, only calorimetric deposits are available. Because of the poor angular resolution of the calorimeters, the matching between energy deposits and vertices is not as efficient as for tracker tracks. The *EcalIso* and *HcalIso* variables, however, can be corrected for PU and UE contributions on an event-by-event basis, by means of the *'ρ correction'* method [5].

The ρ variable, as explained in the next paragraph, represents the jet energy surface density in a given event. This can be converted into an average jet energy value by multiplying it by the jet area. It is then possible to subtract the average contribution of jets from PU and UE from the total amount of the calorimetric isolation variables, according to the following formulae:

$$\begin{aligned} \text{EcalIso}_{corr} &= \text{EcalIso} - \rho \cdot A^{\mu,e}_{EcalIso}(\eta) \\ \text{HcalIso}_{corr} &= \text{HcalIso} - \rho \cdot A^{\mu,e}_{HcalIso}(\eta). \end{aligned} \tag{3.9}$$

The jet areas depend on the lepton being a muon or an electron and on its position in η. The η dependency in Eq. 3.9 is currently taken into account with a coarse division in barrel and end caps, with the transition between the two regions being set at $\eta = 1.479$ because of the ECAL barrel/end cap separation[6].

In analogy to Eq. 3.6, the combined relative isolation variable after the ρ correction is defined as:

$$\text{CombRelIso}_{corr} = \frac{\text{TrkIso} + \text{EcalIso}_{corr} + \text{HcalIso}_{corr}}{p^{\ell}_{T}}. \tag{3.10}$$

The geometrical definition of jet area as πR^2, where R is the radius of the jet cone in the (η, ϕ) space, has been proven not to be optimal because

[5] A correction relying directly on the reconstructed number of event vertices was also envisaged, but it was proven from MC studies that the vertex reconstruction algorithm used at that time suffered from an efficiency loss in events with a high number of generated PU vertices. Nowadays, after the *Deterministic Annealing* (DA) algorithm [23] has been deployed, N^{RECO}_{vtx} follows linearly N^{GEN}_{PU} and this proposal could be reconsidered. However, as shown later on (see Fig. 3.9), the ρ variable is itself directly proportional to N_{vtx}, which is now determined with the DA method.

[6] since only calorimetric isolation undergoes the ρ correction, the choice of this η value has been driven by the ECAL layout. Defining two eta regions has been chosen as a trade-off between having a large enough statistical population in all bins and checking whether there is any different trend of isolation as a function of η.

it does not properly take into account veto regions within the isolation cone. In some cases, *e.g.* like in the *'jurassic veto'* used in electron isolation (see Sec. 3.1.2.2), the veto region has a complicated shape that is not accounted for by the geometrical jet area. Moreover, in the muon case the granularity of vetoes is driven by the CaloTower one, which is rather coarse (see Sec. 3.1.1.2). This makes it difficult to correctly account for the vetoed area. For these reasons, *effective* areas are currently used instead of the geometrical ones. Effective areas are defined as the ratio of the fit slope of the isolation variable as a function of the number of reconstructed vertices N_{vtx} (see Figs. 3.11, 3.12) to the fit slope of ρ as a function of N_{vtx} (see Fig. 3.9). This definition of effective areas flattens by construction the distributions of isolation as a function of N_{vtx} from which they are evaluated.

From Eq. 3.9 one can see that, if the isolation variables are smaller than $\rho \cdot A_{eff}$, they can become negative. No protection against this case has been introduced in the analysis, although negative isolation variables lose their physical meaning. The main reason for not requiring isolation to be non-negative is that this would entail a loss of discriminating power between perfectly isolated leptons and leptons with small (non-corrected) isolation values. For instance, this would increase the number of times the choice of the two least isolated leptons, as required in the $H \to ZZ \to$ analysis, is ambiguous[7].

3.3.2.1 Definition of ρ

Let p_{Tj} be the transverse momentum of the $j-$th jet in a given event and A_j its area[8]. Then the jet energy density variable ρ can be defined for each event as the median of the distribution of p_T/A for all jets within the detector acceptance (which is $|\eta| < 2.5$ for the CMS tracker) and with $p_T^{jet} > 3$ GeV:

$$\rho = \text{median}\left[\left\{\frac{p_{Tj}}{A_j}\right\}\right]. \tag{3.11}$$

The jet areas, and therefore ρ, are well defined provided that the jet reconstruction algorithm used to build up the jets is infrared-safe. The definition itself of jet area relies on jets being infrared-safe, since A_j is the area of the region in which a very large number of very soft particles, called *ghosts*, would be merged into the $j-$th jet. For a detailed discussion about this requirement, see *e.g.* [24]. In many Higgs analyses, including the 4ℓ one, also ghosts are only allowed to be generated within the detector acceptance ($|\eta_{ghost}| < 2.5$). Moreover, the current implementation of the ρ correction algorithm works under the assumption that pile-up is uniformly distributed

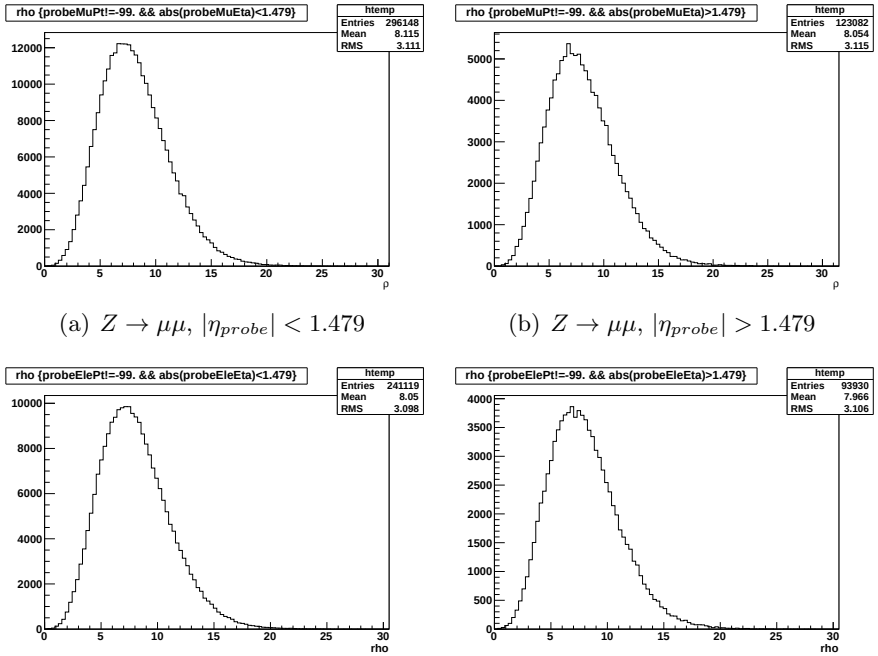
[7] imagine *e.g.* a 4ℓ event with $CombRelIso^{uncorr}_{(\ell_1,\ell_2,\ell_3,\ell_4)} = (0., 0., 0.10, 0.15)$ that after applying the ρ correction turns into $CombRelIso^{corr}_{(\ell_1,\ell_2,\ell_3,\ell_4)} = (0., 0., 0., 0.09)$. The choice of the next-to-least isolated lepton would become ambiguous.

[8] further details about the definition of *'jet area'* can be found e.g. in [24] and in [26].

as a function of η and ϕ. Although this is not exactly true, it is a realistic approximation for $|\eta| < 2.5$, as demonstrated in [25], Sec. 6.2.

The median has been chosen for Eq. 3.11 because it is a more robust statistical estimator than the mean of the same distribution. The jet clustering algorithm that has been used to obtain the results described in the following is the k_T one, with $R = 0.6$. Details about the k_T algorithm can be found in [28].

Some ρ distributions obtained from data with a *'Tag-and-Probe'* $Z \to \ell^+\ell^-$ event selection are shown in Fig. 3.7.



(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$ (b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

(c) $Z \to ee$, $|\eta_{probe}| < 1.479$ (d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.7: *ρ distributions from data, as obtained from the Tag-and-Probe event selection described in Sec. 3.3.2.2. The ρ values are plotted for a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps.*

3.3.2.2 Evaluation of the effective areas

The latest evaluation of effective areas for the $H \to ZZ \to 4\ell$ channel has been performed during the preparatory period towards publishing CMS PAS-HIG-11-025 [29], but not in time for incorporating it into the analysis workflow. This will be done for the upcoming reload.

The latest set of effective areas has been calculated from a data sample

Selection type	Cut	
Event selection	*'Particle Flow'* [30] $MET < 25$ GeV	
Tag and Probe selection	60 GeV $< m_{\ell^+\ell^-} <$ 120 GeV trigger matching for both leptons	
Tag selection	*muons*:	$p_T > 20$ GeV & muon Tight ID
	electrons:	$p_T > 20$ GeV & WP60 (2011)
Probe selection	*muons*:	$p_T > 7$ GeV & GlobalMuon
	electrons:	$p_T > 7$ GeV

Table 3.3: *The selection cuts used to evaluate the effective areas.*

of $L = 1.385$ fb^{-1}, corresponding to certified runs[9] between run n. 175860 and run n. 177878. These runs were taken during the so-called *Run2011B* data taking period, with the highest beam intensity and instantaneous luminosity, and therefore number of pile-up events, so far. The input datasets are two Primary Datasets containing events that pass the *DoubleMu* or the *DoubleElectron* trigger selection.

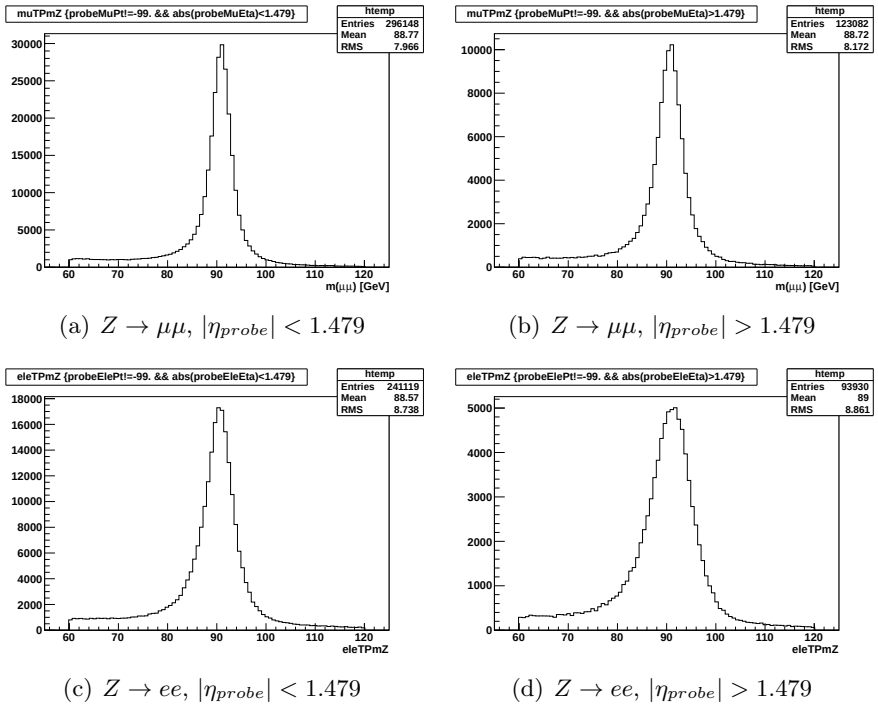
The selection cuts used for this study are listed in Table 3.3. If multiple Z candidates pass the selection, only the one with $m_{\ell^+\ell^-}$ closest to $m_Z^{PDG} = 91.1876$ GeV is considered. No background subtraction is performed by fitting the sidebands of the $m_{\ell^+\ell^-}$ distribution (see Fig. 3.8).

The distributions of ρ *vs* N_{vtx} are shown in Fig. 3.9, along with the straight-line fits whose angular coefficients are used as denominators in the definition of effective areas. The fit range is $4 \leq N_{vtx} \leq 15$ and it has been chosen such that it contains most of the events and that the trends of ρ and of isolation are most linear with respect to N_{vtx}. It has been checked that this range also contains most of the events from all data-taking periods (see Figs. 3.6 and 3.10), which makes the effective areas calculated only from Run2011B data applicable to the whole CMS data set.

The distributions of *EcalIso* and *HcalIso vs* N_{vtx} can be found in Figs. 3.11 and 3.12. The straight-line fits interpolate the data points with a generally good agreement. The angular coefficients of these fit lines enter the A_{eff} calculation in the numerator. The intercept of the fit line with the y axis can be physically interpreted as the residual energy due to UE only. The ρ method is sensitive to both PU and UE, but in the limit case of zero reconstructed vertices only UE activity is present.

The values resulting from the calculation of effective areas are listed in Table 3.4, whereas Figs. 3.13, 3.14, 3.15 show the distributions of the *EcalIso*, *HcalIso* and *CombRelIso* variables of the probe lepton as a function of N_{vtx}

[9] only runs flagged as 'good' by the Data Quality Monitoring team and included in the JSON file Cert_160404-178078_7TeV_PromptReco_Collisions11_JSON.txt have been analysed.



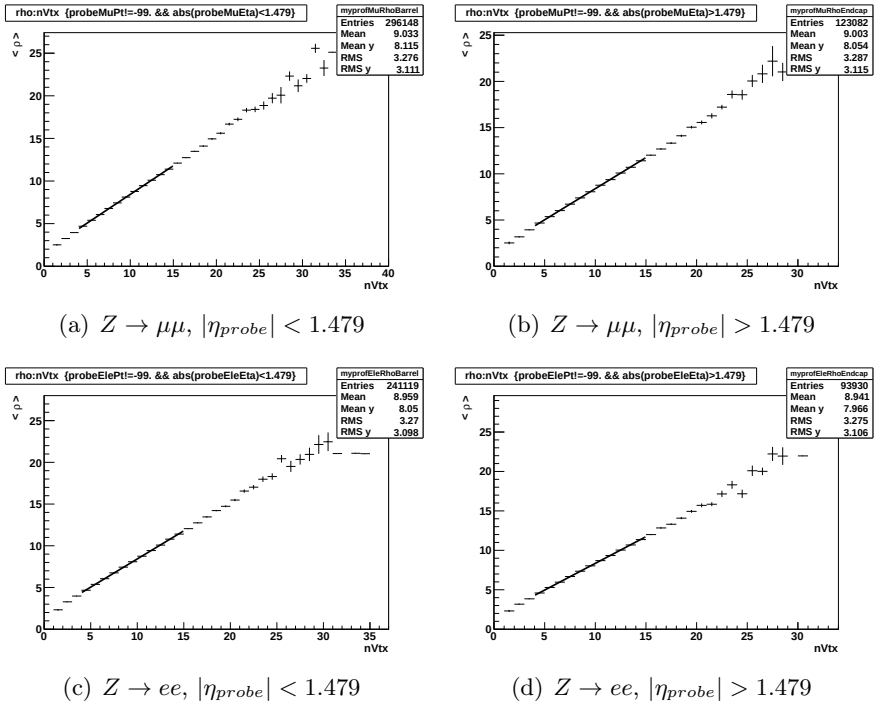
(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$ (b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

(c) $Z \to ee$, $|\eta_{probe}| < 1.479$ (d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.8: *Distribution of $m_{\ell^+\ell^-}$ as obtained from the Tag-and-Probe event selection described in Sec. 3.3.2.2, in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps.*

after the ρ correction has been applied. The data set on which this check has been performed is the same as the one from *Run2011B* used to calculate the effective areas. The new values are shown together with the old ones, which have been obtained from a statistically small MC sample and without the requirement of leptons matched to online trigger objects. This can explain some discrepancies (up to a factor of 2) between the two sets of numbers. The distributions prove that in most cases the recently calculated set of effective areas makes the distributions more uniform with respect to N_{vtx}.

It can also be observed from Figs. 3.11, 3.13 that the ECAL isolation variable for barrel electrons from $Z \to ee$, both before and after the ρ correction, has an offset a factor of 3 larger than the one of the other isolation variables. The explanation has been found in the way vetoes are applied when isolating electrons in ECAL. As listed in Table 3.1, the $\Delta\eta$, ΔR cuts for electron vetoes are expressed in numbers of ECAL crystals. A given crystal, however, covers a larger η range in the end caps than in the barrel, therefore larger energy deposits, on average, are vetoed in the end caps than in the barrel. This causes the isolation efficiencies as a function of η to be larger in the end

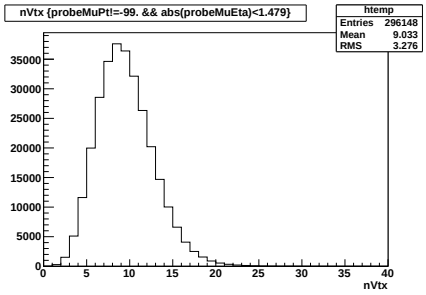


(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$ (b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

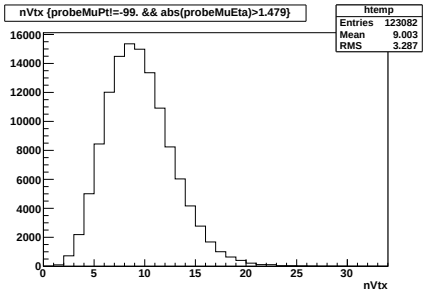
(c) $Z \to ee$, $|\eta_{probe}| < 1.479$ (d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.9: *Dependency of the ρ variable on the mean number of reconstructed primary vertices, in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps. The fits are performed in the range $4 \leq N_{vtx} \leq 15$.*

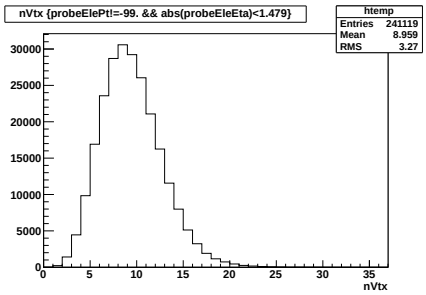
caps than in the barrel. Moreover, slightly different thresholds are applied to ECAL isolation for barrel (where transverse energy is taken into account) and end caps (where energy is considered). See again Table 3.1. Further investigations are being carried out in the $H \to ZZ \to 4\ell$ group, to fully understand the differences between barrel and end cap electron isolation.



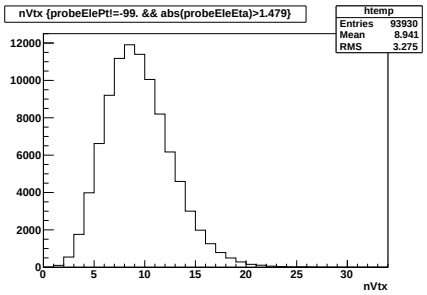
(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$



(b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$



(c) $Z \to ee$, $|\eta_{probe}| < 1.479$

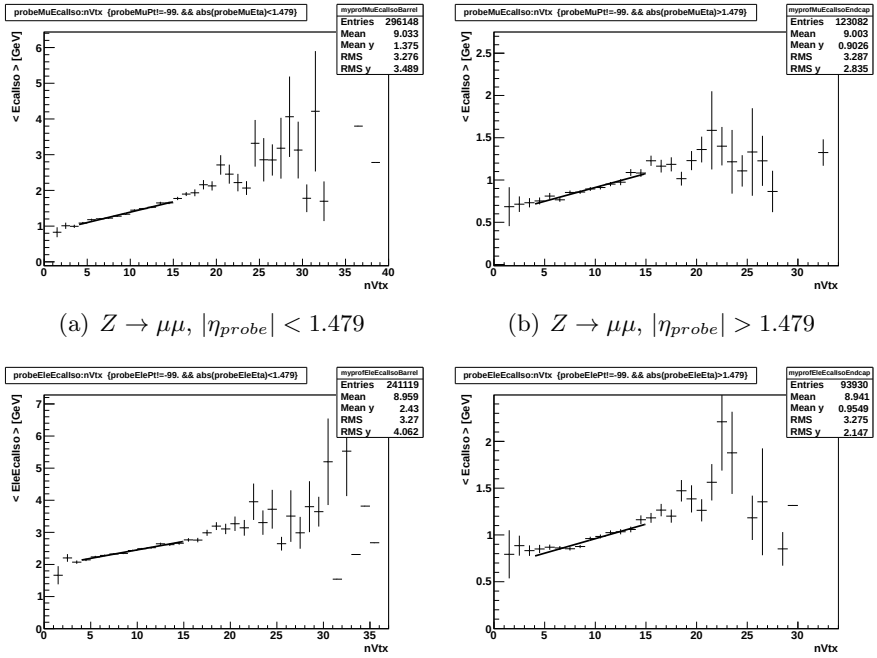


(d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.10: *Distribution of N_{vtx} for the selected events in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps.*

Electrons		
	$A_{EcalIso}$	$A_{HcalIso}$
Barrel	0.078 *(0.101)*	0.026 *(0.021)*
End caps	0.046 *(0.046)*	0.072 *(0.040)*
Muons		
	$A_{EcalIso}$	$A_{HcalIso}$
Barrel	0.087 *(0.074)*	0.042 *(0.022)*
End caps	0.049 *(0.045)*	0.059 *(0.030)*

Table 3.4: *The values of effective areas as evaluated from Run2011B data, from $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ tag-and-probe. The corresponding values obtained from MC (without the requirement of leptons matched to the online trigger objects) are listed in parentheses. These are the ones used so far.*



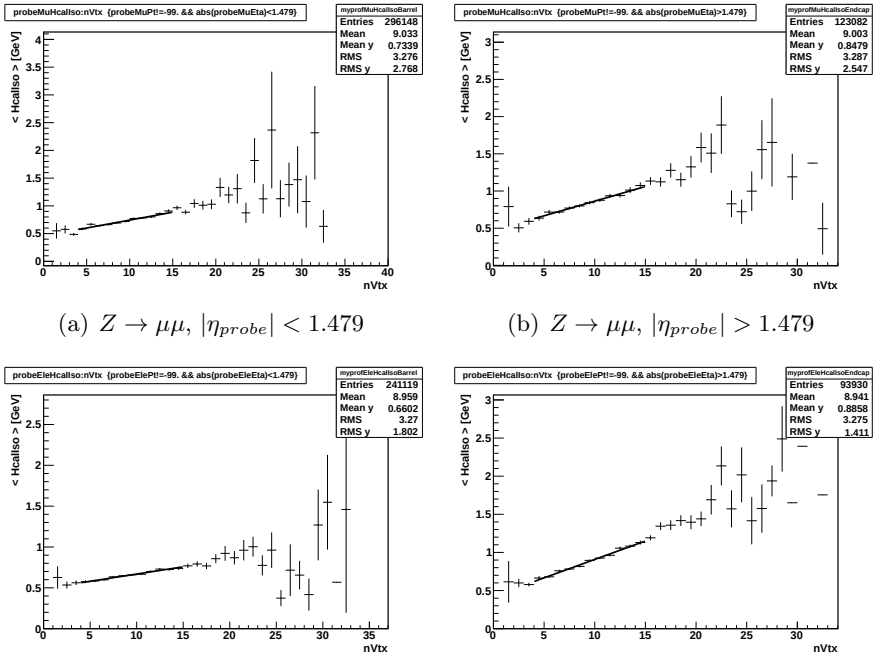
(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$

(b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

(c) $Z \to ee$, $|\eta_{probe}| < 1.479$

(d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.11: *Dependency of the EcalIso variable on the mean number of reconstructed primary vertices, in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps. The fits are performed in the range $4 \leq N_{vtx} \leq 15$.*



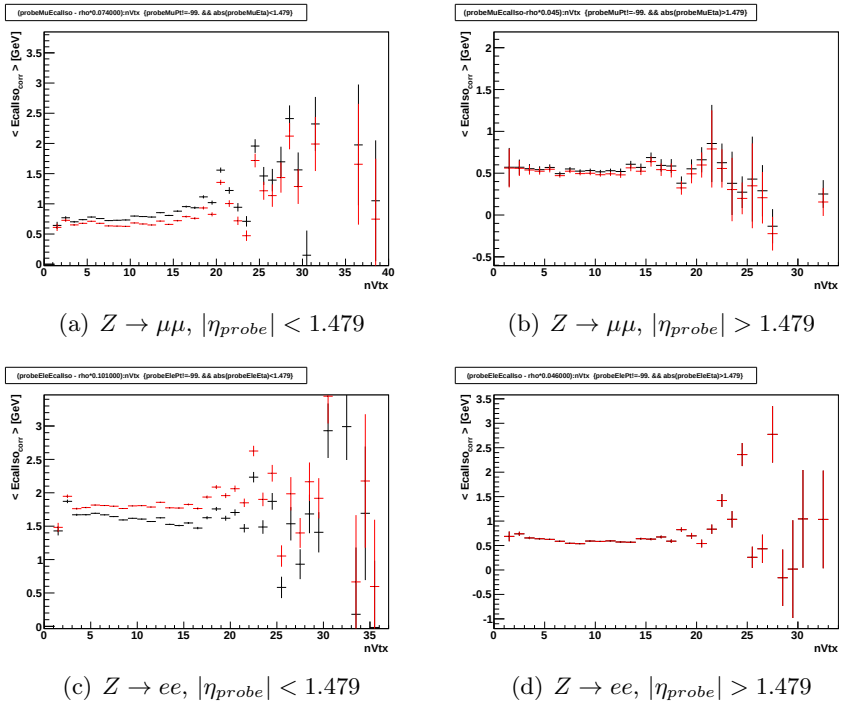
(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$

(b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

(c) $Z \to ee$, $|\eta_{probe}| < 1.479$

(d) $Z \to ee$, $|\eta_{probe}| > 1.479$

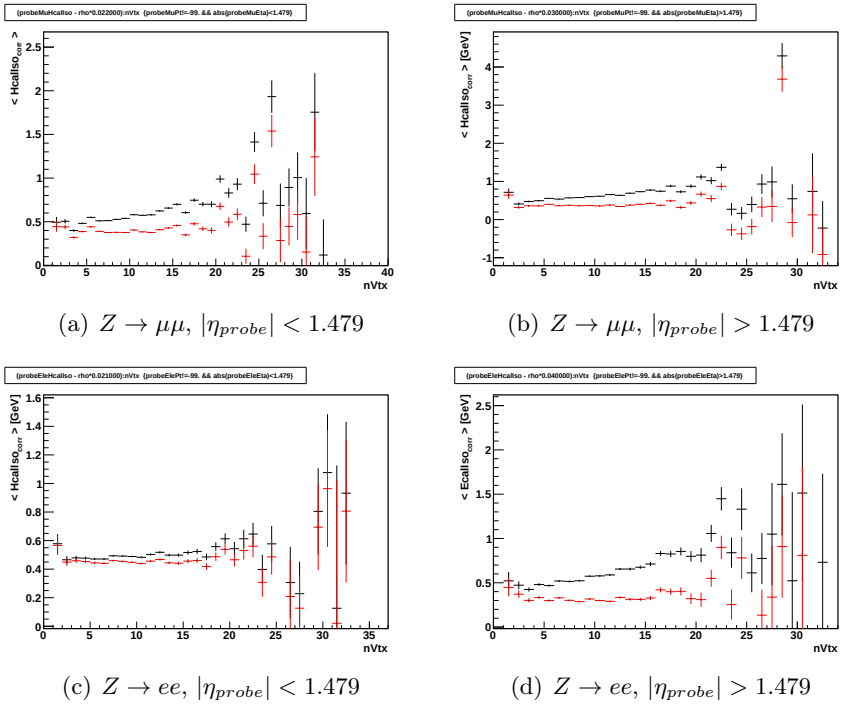
Figure 3.12: *Dependency of the HcalIso variable on the mean number of reconstructed primary vertices, in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps. The fits are performed in the range $4 \leq N_{vtx} \leq 15$.*



(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$ (b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

(c) $Z \to ee$, $|\eta_{probe}| < 1.479$ (d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.13: *Comparison between the distributions of the ρ-corrected EcalIso variable as a function of the number of reconstructed vertices with the 'old' (in black) and the 'new' (in red) sets of effective areas, in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps.*



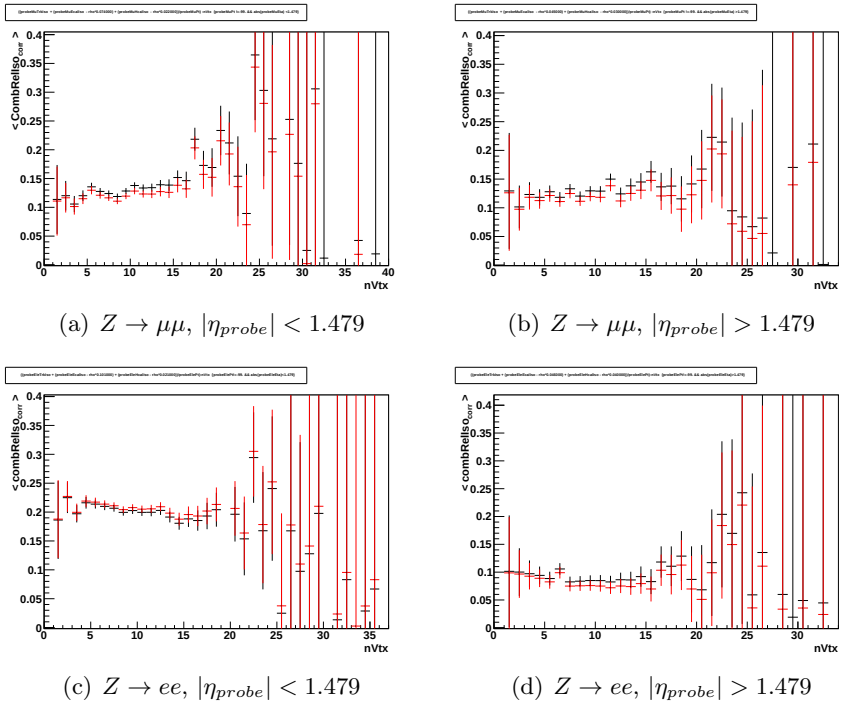
(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$

(b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

(c) $Z \to ee$, $|\eta_{probe}| < 1.479$

(d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.14: *Comparison between the distributions of the ρ-corrected HcalIso variable as a function of the number of reconstructed vertices with the 'old' (in black) and the 'new' (in red) sets of effective areas, in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps.*



(a) $Z \to \mu\mu$, $|\eta_{probe}| < 1.479$

(b) $Z \to \mu\mu$, $|\eta_{probe}| > 1.479$

(c) $Z \to ee$, $|\eta_{probe}| < 1.479$

(d) $Z \to ee$, $|\eta_{probe}| > 1.479$

Figure 3.15: *Comparison between the distributions of the ρ-corrected CombRelIso variable (see Eq. 3.10) as a function of the number of reconstructed vertices with the 'old' (in black) and the 'new' (in red) sets of effective areas, in the four cases: a) $Z \to \mu\mu$ events with the probe muon in the barrel, b) in the end caps, c) $Z \to ee$ events with the probe electron in the barrel, d) in the end caps.*

3.3.2.3 Cross-check of the evaluation of effective areas

A completely independent analysis performed by the University of Bologna group has produced results [31] that can provide a cross-check for the evaluation of effective areas in the context of the ρ correction. The primary goal of this study was to monitor the $Z \to \ell^+\ell^-$ ($\ell = e, \mu$) event yield, as well as the trend of other observables related to the Z or to the lepton pair, as a function of data-taking time. Data has been split in subsets (called *lumi slots*) of size $L = 20$ pb^{-1} and for each of them an average value of the observables has been calculated.

The total luminosity analysed for this study amounts to $L = 4.64$ fb^{-1}. *DoubleElectron* and *DoubleMu* primary datasets have been used as input. The event selection cuts are listed in Table 3.5. The ρ-correction has been applied and the most recent set of values of the effective areas has been used, as reported in Table 3.4.

variable	cut	
	electrons	*muons*
p_T^ℓ	$p_T^e > 7$ GeV	$p_T^\mu > 5$ GeV
η^ℓ	$\lvert\eta^e\rvert < 2.5$	$\lvert\eta^\mu\rvert < 2.4$
ID	eID tight	GlobalMuon
number of track hits	N(expected inner hits) ≤ 1	N(strip + pixel hits) > 10
isolation	TrkRelIso < 0.7	
3D IP significance	$\mathrm{SIP}_{3D} < 100$	
Z candidate mass	$m_Z > 50$ GeV	
p_T^ℓ	$p_T^{\ell_1} > 20$ GeV & $p_T^{\ell_2} > 10$ GeV	
isolation	$\mathrm{CombRelIso}_{\ell_1} + \mathrm{CombRelIso}_{\ell_2} < 0.35$	
3D IP significance	$\max(\mathrm{SIP}_{3\mathrm{D}}^{\ell_1}, \mathrm{SIP}_{3\mathrm{D}}^{\ell_2}) < 4$	

Table 3.5: *Selection cuts defining the data sample for $Z \to \ell\ell$ ($\ell = e, \mu$) yield studies.*

The isolation variable R_{iso} is defined as the sum of the ρ-corrected CombRelIso variables (see Sec. 3.1.1.3) of the two leptons that reconstruct a candidate Z. The R_{iso} trend for $Z \to \mu\mu$ events after the rho correction (see Fig. 3.16) is fairly stable as a function of the *lumi slot* number. A slight residual increase in time, along with the higher and higher instantaneous luminosity provided by the LHC, may indicate that the effective areas are actually slightly underestimated. However, the normalised chi-square $\tilde{\chi}^2 = 332.9/230 = 1.45$ obtained from a fit with a horizontal straight line reflects a generally good agreement with the hypothesis of constancy as a function of time.

The same plot, but in the di-electron case (see Fig. 3.17), clearly shows a large discontinuity around *lumi slot* 94, as well as a dip in the *lumi slot*

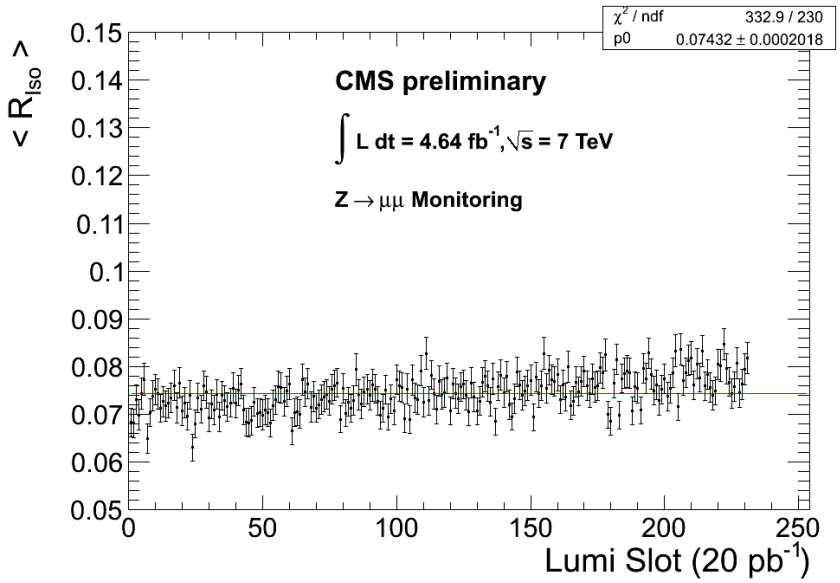


Figure 3.16: *Trend of the sum of the ρ-corrected $CombRelIso$ variables of the two muons associated to a $Z \to \mu\mu$ candidate as a function of the lumi slot number. For each lumi slot the average value of the isolation variable is plotted. The red straight line is the best-fit line in the hypothesis of a constant trend with respect to lumi slot number.*

range $\sim 55 \div 70$. This latter one has been traced back to the fact that two di-electron trigger paths[10] that were meant to remain unprescaled were actually prescaled by mistake at the time of the deployment of the $2E33$ online trigger menu. The amount of data affected by this issue is the run range $171050 \div 171578$, for an integrated luminosity of $L = 58$ pb^{-1}. Although a recipe for correcting for this unwanted prescale has been released, it has not been implemented in this analysis. Also the discontinuity around *lumi slot* 94 is due to trigger requirements. It coincides with the deployment of the $3E33$ online trigger menu, which introduced new working points for the *EGamma* HLT paths. Since these paths include tighter isolation cuts than those applied in the offline selection, the discontinuity caused by the introduction of new trigger paths with higher isolation cuts is still visible in Fig. 3.17. This also explains why a similar gap is not observed in the $Z \to \mu\mu$ channel. If a set of cuts emulating the online trigger is applied to this data sample, the discontinuity disappears, as shown in Fig. 3.18. What is interesting for the purpose of cross-checking the effective area calculation is that the trend of the isolation variable before and after this discontinuity

[10] The two affected trigger paths are:
$HLT\_Ele17\_CaloIdT\_CaloIsoVL\_TrkIdVL\_TrkIsoVL\_Ele8\_CaloIdT\_CaloIsoVL\_TrkIdVL\_TrkIsoVL\_v7$
$HLT\_Ele17\_CaloIdVT\_CaloIsoVT\_TrkIdT\_TrkIsoVT\_Ele8\_Mass30\_v6$

is rather flat as a function of time, apart from a slight rise compatible with the different event selections in the two analyses.

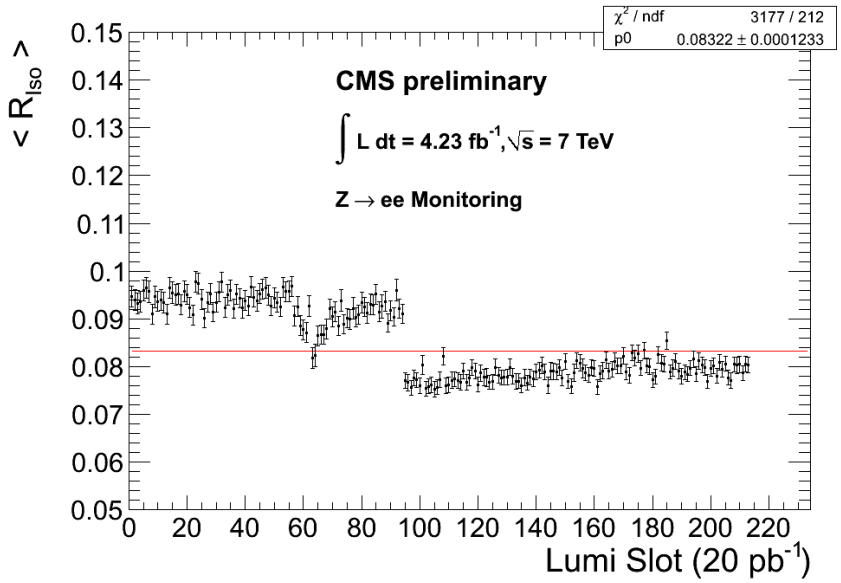


Figure 3.17: *Trend of R_{iso}, defined as the sum of the ρ-corrected $CombRelIso$ variables of the two electrons associated to a $Z \rightarrow ee$ candidate, as a function of the lumi slot number. For each lumi slot the average value of the isolation variable is plotted. The red straight line is the best-fit line in the hypothesis of a constant trend with respect to lumi slot number.*

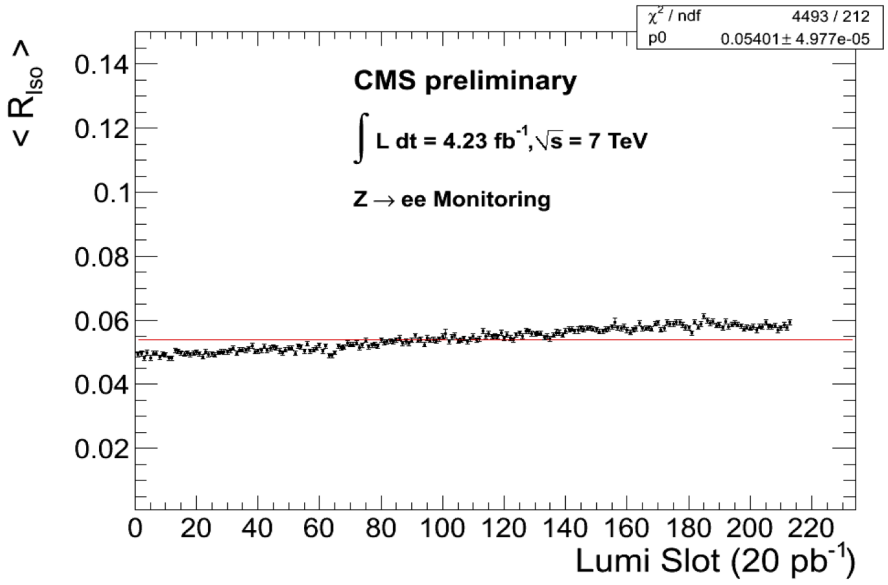


Figure 3.18: *Trend of R_{iso}, defined as the sum of the ρ-corrected CombRelIso variables of the two electrons associated to a $Z \to ee$ candidate, as a function of the lumi slot number, for data passing the trigger emulation cuts. For each lumi slot the average value of the isolation variable is plotted. The red straight line is the best-fit line in the hypothesis of a constant trend with respect to lumi slot number.*

3.4 Systematic uncertainties on lepton isolation efficiencies

3.4.1 Evaluation with the Lepton Kinematic Template Cone method

3.4.1.1 The LKTC method

The Lepton Kinematic Template Cone ($LKTC$) method [32] can be used to estimate the systematical uncertainty on lepton isolation efficiency. The method consists in running on two data sets of processes having a similar underlying event. One of them is a MonteCarlo sample from which lepton (p_T, η, ϕ) distributions are extracted. These can be used as probability density functions for generating hundreds of sets of pseudo-tracks, which are then plugged into real data events of the other sample. For instance, if kinematic templates are extracted from a simulated sample of $Z \to \ell^+\ell^-$ events, hundreds of pseudo-track pairs of same-flavour, opposite-charge leptons are generated in data events. These tracks are then treated as real leptons and their isolation is computed. The difference between the isolation efficiency extracted from pseudo-tracks and from real-data ones can be used as an estimate of the systematic uncertainty on the efficiency.

The LKTC method relies on the assumption that the kinematics of leptons produced in $W \to \ell\nu$, $Z \to \ell\ell$ events from hard parton scattering in $p - p$ collisions is unrelated to the underlying event activity. It can also be applied to efficiencies with respect to relative isolation variables, because the information about lepton p_T is included. Moreover, it is a powerful tool to properly study multiple-lepton isolation, since kinematic templates allow to take into account correlations among leptons belonging to the same event. It is also applicable in case of very limited data samples, because of the high number of pseudo-tracks that can be plugged into an event.

3.4.1.2 Results

All the results reported in the following only refer to muons. No attempt to apply the LKTC technique to decay processes with final-state electrons has been made recently in the $H \to 4\ell$ group. The MC samples used to extract the kinematic templates are part of the so-called *Fall10* production campaign. They are $gg \to H \to ZZ \to 4\ell$ ($m_H = 150$ GeV), $W \to \mu\nu$, $Z^{(*)}/\gamma^* \to \mu\mu$. The data sample is a subset of events from the so-called *Nov4* reprocessing that pass a selection for W/Z candidates[11]. The integrated

[11] These are the exact data set names for MC:

- /GluGluToHToZZTo4L_M-150_7TeV-powheg-pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO
- /DYToMuMu_M-20_CT10_TuneZ2_7TeV-powheg-pythia/Fall10-START38_V12-v1/GEN-SIM-RECO
- /WToMuNu_TuneZ2_7TeV-pythia6/Fall10-START38_V12-v1/GEN-SIM-RECO

and for data:

luminosity for this sample is $L = 36$ pb^{-1}. The selection used to search for W/Z candidates is the VBTF one[12] [33], except for two relaxed requirements (both for data and for MC): the matching of a reconstructed muon to an online muon trigger object and the muon isolation cuts are not applied. Isolation cuts, however, are included in the definition of the real-data subset, to increase its purity without biassing the results obtained from pseudo-track isolation. As for MC samples, the other cuts are tight enough to ensure that the samples from which the templates are extracted are highly pure.

The comparison shown in Fig. 3.19 is a sanity check performed with W and Z MC samples only. Kinematic templates are injected into the same samples from which they are extracted, after being reshuffled so as to lose the event-by-event correspondence. Isolation efficiencies for LKTC pseudo-tracks and for real MC tracks are compatible within 1% for values close to the reference threshold $CombRelIso_{\rm thr} = 0.15$.

The main result about LKTC evaluation of systematic uncertainty on muon isolation efficiency is presented in Fig. 3.20. Isolation curves have been derived from $W \to \mu\nu$ templates injected into $W \to \mu\nu$ MC and data samples, from $Z \to \mu\mu$ templates injected into $Z \to \mu\mu$ MC and data samples and from $H \to ZZ \to 4\mu$ templates injected into the $Z \to \mu\mu$ MC sample. The spread of the efficiency values for a given isolation threshold ranges between 1 and 2% over the whole *CombRelIso* range and it is $\sim 1\%$ for $CombRelIso_{\rm thr} = 0.15$. According to these results, a conservative systematic uncertainty of 2% is therefore assigned to muon isolation efficiency.

- /Mu/Run2010B-WZMu-Nov4Skim_v1/RAW-RECO

[12]Details can be found at the following link:
`http://cmssw.cvs.cern.ch/cgi-bin/cmssw.cgi/CMSSW/ElectroWeakAnalysis/ZMuMu/python/ZMuMuGolden_cfi.py`

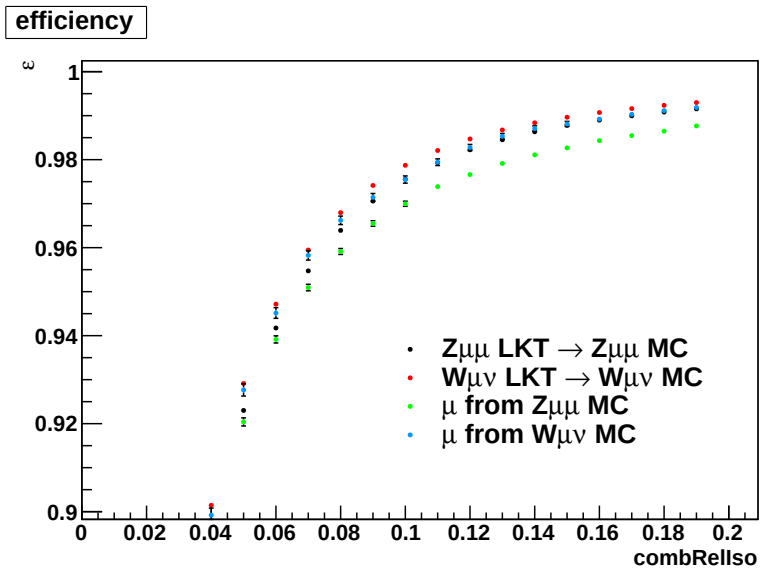


Figure 3.19: *Isolation efficiency as a function of the CombRelIso variable of each muon for events passing the LKTC selection. The plot compares the efficiency curves obtained from LKTC pseudo-track from W (red) and Z (black) decays with those obtained from final-state muons in the W (blue) and Z (green) MC samples.*

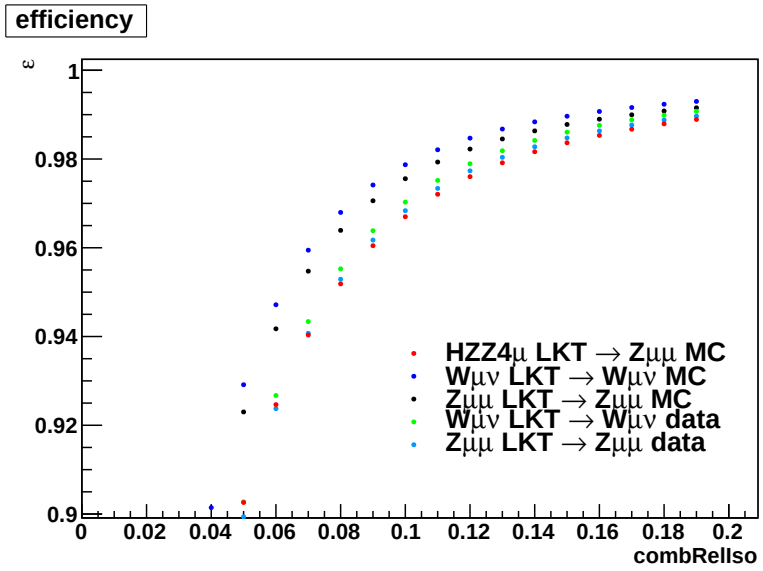


Figure 3.20: *Isolation efficiency as a function of the CombRelIso variable of each muon for events passing the LKTC selection. The legenda lists, for each curve, the sample from which the kinematic templates have been extracted and the one in which the muon pseudo-track have been generated.*

Chapter 4

Search for a SM-like Higgs boson in the $H \to ZZ^{(*)} \to 4\ell$ decay channel

This chapter describes the search for a Standard Model-like Higgs boson in the decay channel $H \to ZZ^{(*)} \to \ell^+\ell^-\ell'^+\ell'^-$, with $\ell, \ell' = e, \mu$, with a data set corresponding to a total integrated luminosity of $\mathcal{L} = 4.71 \pm 0.21$ fb^{-1} recorded by the CMS detector during 2010 and 2011 proton-proton data taking campaigns at $\sqrt{s} = 7$ TeV. Three final states ($4e$, 4μ, $2e2\mu$) are considered and treated in the same way across the whole analysis, except for different selection cuts on the muon and electron objects and for the interference among identical leptons described in Sec. 4.1. The search is performed in the mass range $100 < m_H < 600$ GeV, by means of a *'cut-and-count'* selection. The 4ℓ final state is the *'golden'* channel for Higgs searches, because of:

- a very clean signature: two pairs of same-flavour, opposite-charge, high-p_T isolated leptons compatible with the same reconstructed vertex
- the absence of physical (*i.e.* not instrumental) missing energy, which is an advantage with respect to *e.g.* the $H \to WW^{(*)} \to 2\ell2\nu$ channel, although the branching ratio is lower
- a clean reconstructed 4ℓ invariant mass peak. Moreover, at least one of these lepton pairs comes from an on-shell Z decay, hence it peaks around m_Z^{PDG}.
- the relatively small amount of background with respect to other Higgs decay channel involving *e.g.* jets

The main background sources are $ZZ \to 4\ell$, which is often referred to as *irreducible* because of its signal-like kinematics, $t\bar{t}$, $Zb\bar{b}, Zc\bar{c}$ with leptons

originating from semi-leptonic b, c decays and $Z+$jets, $W+$jets, QCD with jets misidentified as leptons. The strategy for suppressing these backgrounds while keeping as high a signal efficiency as possible is detailed in the following. Here and henceforward Z stands for Z, Z^* and, when possible, γ^*.

4.1 Signal samples

The Higgs signal samples have been generated with `POWHEG` [34] with both NLO gluon fusion ($gg \to$ H) and weak-boson fusion ($q\bar{q} \to q\bar{q}$H). The Higgs boson widths listed in [35] and the CTEQ6M parton distribution function (PDF) set are used. `PYTHIA` has been run to generate other samples of ZH, WH, $t\bar{t}H$ processes. In all samples both Z bosons can be off-mass-shell and are forced to decay into a lepton pair.

The Higgs mass values considered in *Summer11* Monte Carlo production span the range $115 < m_H < 600$ GeV, with mass points every 10 GeV up to $m_H = 230$ GeV and every 25 GeV up to $m_H = 600$ GeV.

In the current analysis only $gg \to$ H samples are used. They are rescaled to the total Higgs production cross section, which includes all other production processes (weak-boson fusion, ZH, WH, $t\bar{t}H$). The gluon fusion contribution to $\sigma(pp \to \mathrm{H})$ is calculated up to NNLO and NNLL, whereas the weak-boson fusion one is evaluated at NNLO.

The total cross section is scaled by the $BR(\mathrm{H} \to 4\ell)$ branching ratio [35, 36, 37, 38, 39]. The $\mathrm{H} \to 4\ell$ cross-section as a function of the Higgs mass m_H for $\sqrt{s} = 7$ TeV is shown in Fig. 4.1 (a). The 4μ and $4e$ final states have an enhanced $\sigma \cdot BR$ with respect to the $2e2\mu$ one because of the interference of amplitudes with permutation of identical leptons coming from the decay of different Z bosons. The enhancement is shown in Fig. 4.1(b) and it is properly taken into account by `Prophecy4f` [35, 36, 37].

4.2 Background samples

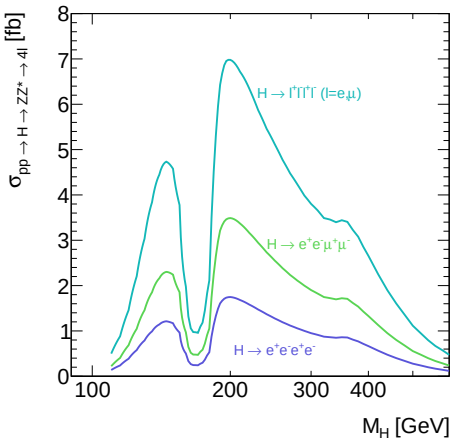
4.2.1 Dibosons

4.2.1.1 $q\bar{q} \to ZZ^{(*)} \to 4\ell$

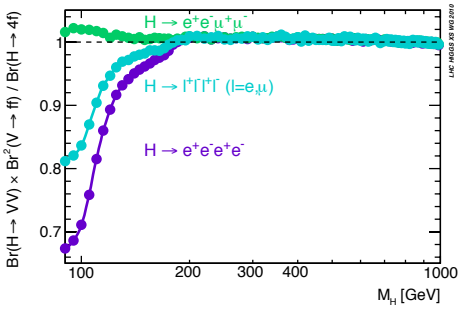
The $q\bar{q} \to \mathrm{ZZ}^{(*)} \to 4\ell$ samples have been produced with `POWHEG`. They take into account the complete NLO simulation, interfaced to `PYTHIA` for showering, hadronisation, decays and for the underlying event.

4.2.1.2 $gg \to ZZ^{(*)} \to 4\ell$

The `gg2ZZ` tool [40] has been chosen to generate $gg \to$ ZZ events at LO, whereas showering and hadronisation are taken care of by `PYTHIA`. `gg2ZZ` includes a cut on the dilepton invariant mass at generator level, $m_{\ell^+\ell^-} > 10$ GeV.



(a)



(b)

Figure 4.1: *(a) Cross-section for SM Higgs in* $\mathrm{H} \to 4\ell$, $\mathrm{H} \to 2e2\mu$ *and* $\mathrm{H} \to 4e$ *(or* 4μ*) as a function of* m_{H} *in* $p-p$ *collisions at* $\sqrt{s} = 7$ *TeV (b) Cross-section enhancement due to the interference of amplitudes with permutations of identical leptons originating from different* Z *bosons, as a function of* m_{H}.

The LO PDF set CTEQ6L1 and the central renormalisation and factorisation scales $\mu_{\mathrm{R}} = \mu_{\mathrm{F}} = m_{\mathrm{Z}} = 91.188$ GeV are used. The total cross section is 3.48 fb for events with different-flavour final-state lepton pairs only. The differential cross-section for $gg \to \mathrm{ZZ}^{(*)}$ as a function of $m_{4\ell}$ for different-flavour lepton pairs is reported in Ref. [41].

4.2.2 W/Z + jets

A sample of Z+jets $\to \ell^+\ell^-$+jets has been generated with `MadGraph`. A cut on the dilepton invariant mass at generator level, $m_{\ell\ell} > 50$ GeV is applied. The jets can originate from the hadronisation of both light (u, d, s) and

heavy (c, b) quarks, but a filter has been run to separate the events with heavy-flavour quarks (henceforth referred to as $Zb\bar{b}$ events) from Z+light jet events. The total NNLO cross section for this sample (including all jet flavours) is 3048 pb.

`MadGraph` has also been used to generate a sample of W+jets events, whose cross section is 31314 pb.

4.2.3 $t\bar{t} \to 2\ell 2\nu 2b$

A $t\bar{t} \to 2\ell 2\nu 2b$ sample has been generated with `POWHEG` using CTEQ6M. The theoretical NLO cross section is $\sigma_{NLO}(pp \to t\bar{t} \to 2\ell 2\nu 2b) = 17.32$ pb [42].

4.3 Data samples

The $H \to ZZ^{(*)} \to 4\ell$ analysis includes results from the whole data sample recorded during 2010 and 2011 data taking periods. The total integrated luminosity amounts to (4.71 ± 0.21) fb^{-1}, measured with an uncertainty of 4.5% [43]. Further information about these samples can be found in Table 4.1.

period	L_{int} (pb^{-1})	$\mathcal{L}_{instant}$ (cm^{-2}s^{-1})	run range	dataset
2010	36	$1 \cdot 10^{29} \div 6 \cdot 10^{31}$	136033 ÷ 149442	April11 ReReco
2011A	2125	$2 \cdot 10^{32} \div 2 \cdot 10^{33}$	160329 ÷ 168437	July5 ReReco
			170053 ÷ 172619	Aug5 ReReco
			172620 ÷ 173692	Oct3 ReReco
2011B	2549	$3 \cdot 10^{33}$	175860 ÷ 177051	Prompt Reco

Table 4.1: *Data samples used in the $H \to ZZ^{(*)} \to 4\ell$ analysis.*

Data samples are centrally organised in Primary Datasets (PDs) whose definition in terms of HLT paths (see Sec. 1.7.1.2) has been changing along with the increasing instantaneous luminosity delivered by the LHC. During 2010 data taking, single-electron or single-muon trigger requirements were applied on data. From the 2011A period on, symmetric or asymmetric trigger thresholds have been set on muon and electron pairs, often including additional identification and isolation requirements[1]. Further details about the PD definition can be found in [44].

4.3.1 The 4ℓ data *'skim'*

A pre-filtered version of the data sample has been produced for the benefit of all $H \to 4\ell$ analysis groups. The presence of at least two reconstructed lepton candidates is required, with $p_{T,1} > 20$ GeV, $p_{T,2} > 10$ GeV for electrons, $p_{T,1} > 20$ GeV, $p_{T,2} > 7$ GeV for muons, and a dilepton invariant mass

[1] Some HLT paths for 2012 data taking will also include p_T requirements on three leptons.

$m_{\ell^+\ell^-} > 40$ GeV. All these cuts are looser than those of the selection sequence, therefore applying this 'skimming' does not result in any loss of efficiency. The selection efficiencies with respect to the total number of 4ℓ events in data are listed in Table 4.2.

	$m_H = 120$ GeV	$m_H = 150$ GeV	$m_H > 200$ GeV
$4e$	88.5%	98.1%	99.6%
4μ	94.5%	98.2%	100.0%
$2e2\mu$	96.0%	99.0%	100.0%

Table 4.2: *'Skimming' efficiencies with respect to the total number of 4ℓ events in data.*

4.4 The event selection sequence

4.4.1 Definition of the *electron* and *muon* objects

It is of paramount importance for the $H \to ZZ \to 4\ell$ analysis to benefit from a very high lepton reconstruction efficiency ε_ℓ. The overall reconstruction efficiency for a 4ℓ final state is $\varepsilon_{4\ell} \propto \varepsilon_\ell^4$, therefore the single-lepton efficiency is required to be as high as possible. It is actually very close to 100% for the two leptons coming from on-shell Z boson decays, whereas it is lower for low-p_T leptons from off-shell Z decays. A good capability of reconstructing and identifying leptons also helps to keep the event yield from instrumental backgrounds low, by means of a low fake rate.

In the currently not excluded Higgs mass range, and more generally for $m_H < 2 \cdot m_Z$, at least one lepton pair comes from an off-shell Z boson and it has in most cases $p_T < 10$ GeV. At such very low p_T values it is particularly difficult to keep the lepton reconstruction and identification efficiency high without losing discrimination power from hadronic jets.

4.4.1.1 Electron reconstruction and identification

Electron reconstruction [46] begins in the ECAL, from energy deposits in a cluster of clusters (*supercluster*) collecting also Bremsstrahlung photons radiated off the electron track in the tracker volume. The supercluster is then used to look for matching tracker hits, from which the electron track is seeded. The electron energy loss in the tracker is correctly taken into account and a fit with a Gaussian Sum Filter[2] is performed. In case several tracks are reconstructed, because of photon conversions, a cleaning algorithm resolves the ambiguities. A set of loose cuts on track-cluster matching variables is implemented. Then, the angular information coming from the GSF tracker

[2]hence the name *'GSF electron'* for this kind of electron tracks

track and the energy value obtained by combining tracker and ECAL measurements are exploited to assign the electron a four-momentum vector [45]. All the observables attached to the electron track refer are evaluated at its point of closest approach to the primary vertex in the plane perpendicular to the beam line.

The $H \to ZZ^{(*)} \to 4\ell$ analysis requires four electrons candidates with transverse momentum $p_T^e > 7$ GeV and $|\eta^e| < 2.5$, passing a loose isolation cut, tracker relative isolation below 0.7 (see Sec. 3.1.2.3 for the definition of the isolation variable). On top of these acceptance cuts, a selection called *Cuts-in-Categories* (in short, *CiC*) is introduced [47]. It consists of nine categories[3] (see Table 4.3) in which cuts are optimised so as to provide the largest signal-to-background ratio.

The cut variables are the following ones:

- $|\Delta\eta_{in}| = |\eta_{sc} - \eta_{in}^{extrap.}|$: the difference between the energy-weighted η position of the supercluster and the one extrapolated from the inner track to the position of closest approach to the supercluster
- $|\Delta\phi_{in}| = |\phi_{sc} - \phi_{in}^{extrap.}|$: the same as $|\Delta\eta_{in}|$, but with respect to the azimuthal coordinate
- E_{seed}/p_{in}: the seed cluster energy divided by the track momentum at the track vertex
- H/E, the ratio of the energy deposited in HCAL behind the ECAL supercluster to the energy deposited in the ECAL supercluster
- $\sigma_{i\eta i\eta}$: the supercluster width in η, as derived from the cluster shape covariance.

In order to reject electrons from converted photons, the electron track is also required to have no more than one expected missing hit in the tracker.

4.4.1.2 Muon reconstruction and identification

The muon candidates used in the $H \to ZZ^{(*)} \to 4\ell$ analysis are *Global Muons* [48]. The track reconstruction begins in the muon chambers, where track segments are built up in DT and CSC subdetectors, whereas 3D hits are defined in the RPCs. A track candidate is defined in the muon system, then it is propagated inwards to the tracker volume, where a match with a tracker track is looked for. If it is found, the whole set of hits and segments is fit and a Global Muon is produced.

[3] The *'brem'* class refers to a phase-space region with little contamination from fakes, the *'lowbrem'* one to a region populated both by real and by fake electrons, the *'badtrack'* one to a region with few real electrons, but nevertheless too many to reject all of them.

number	electron type	region
1	'brem' electron	barrel
2	'lowbrem' electron	barrel
3	'badtrack' electron	barrel
4	'brem' electron	end caps
5	'lowbrem' electron	end caps
6	'badtrack' electron	end caps
7	'crack' electron	barrel
8	'crack' electron	end caps
9	'pure tracker-driven' electron	-

Table 4.3: *The categories used in the CiC algorithm*

The muon four-momentum is the one of the tracker track associated to it. The observables relative to a muon track refer to its point of closest approach to the primary vertex position in the transverse plane. In the 4ℓ analysis muons are required to be within the acceptance cuts $p_T^{\mu} > 5$ GeV, $|\eta^{\mu}| < 2.4$ and loosely isolated (tracker relative isolation below 0.7, see Sec. 3.1.1.3 for the definition of this variable). Moreover, the fit has to be based on at least 10 tracker hits, so that the muon momentum is precisely measured.

4.4.2 The selection sequence

The selection sequence is run on the events included in the 4ℓ 'skims' described in Sec. 4.3.1. They include acceptance cuts and a loose selection based on lepton p_T and isolation, as well as the requirement of leptons matched to online high-level trigger objects, both for data and for MC samples[4].

4.4.2.1 The first Z candidate

The first Z candidate, often referred to as Z_1, is defined as the one with the dilepton invariant mass closest to the nominal Z mass, in the mass window $50 < m_{\ell^+\ell^-} < 120$ GeV, after a selection including cuts on the p_T of both leptons ($p_{T,1} > 20$ GeV, $p_{T,2} > 10$ GeV), on their isolation ($CombRelIso_1 + CombRelIso_2 < 0.35$) and on the significance of their 3D impact parameter with respect to the primary event vertex (both leptons are required to have $|SIP_{3D}| = |IP_{3D}/\sigma_{IP_{3D}}| < 4$).

As an example, some distributions obtained from data and MC samples after this selection step are shown for the four-muon final state and for an

[4] in the case of MC samples, in which the trigger paths can not reflect those defined for data, a trigger *emulation* is implemented, which is equivalent to the cuts included in the trigger path definition.

integrated luminosity of $L = 4.71$ fb^{-1}. The p_T distribution of the largest-p_T muon is plotted in Fig. 4.2, whereas the m_{Z1} distribution is shown in Fig. 4.3.

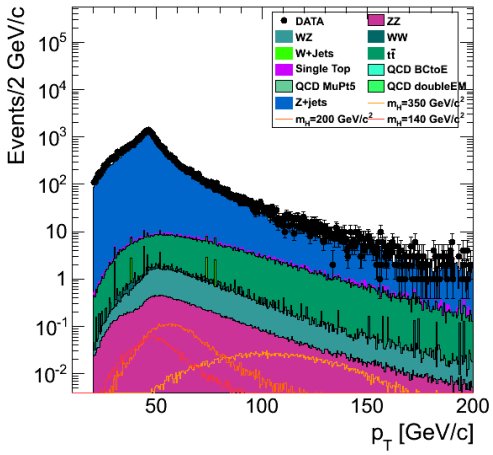


Figure 4.2: *Transverse momentum distribution related to the muons with largest p_T in the 4μ final state, for events passing the selection of the first Z candidate. Black dots refer to data, coloured regions to the different MC backgrounds. Also three signal mass values have been considered (solid lines): $m_H = 140$ GeV, $m_H = 200$ GeV, $m_H = 350$ GeV. The effects of requiring $p_{T,1} > 20$ GeV are well visible. The overall MC shape well reproduces the one obtained from real data.*

4.4.2.2 Adding the third and the fourth lepton

The presence of a third high-quality lepton of any flavour and charge is required. At this stage of the selection the phase space of the main reducible backgrounds is preserved for data-driven background estimation and control. Also the requirement of a fourth lepton with matching flavour and opposite charge with respect to the third one is added.

4.4.2.3 Choosing the 'best' 4ℓ candidate

The second Z candidate is reconstructed from the two highest-p_T leptons not associated to Z_1 and passing $m_{Z_2} > 12$ GeV. At this stage the ambiguity due to combinatorics in events with extra fake leptons is limited and the *'best 4ℓ candidate'* is chosen. The 4ℓ candidate must satisfy $m_{4\ell} > 100$ GeV. Moreover, in the $4e$ and 4μ final states, at least three out of the four possible $\ell^+\ell^-$ combinations must have $m_{\ell^+\ell^-} > 12$ GeV, so that background events with leptonic decays of low-mass resonances (*e.g.* J/ψ, Υ) are rejected.

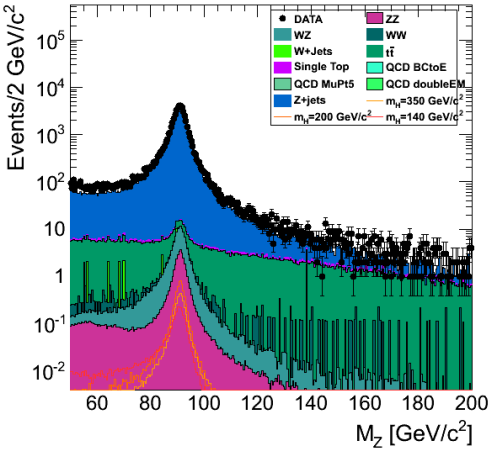


Figure 4.3: *Invariant mass distribution obtained from the two muons associated to the Z_1 candidate in the 4μ final state, for events passing the selection of the first Z candidate. Black dots refer to data, coloured regions to the different MC backgrounds. Also three signal mass values have been considered (solid lines): $m_H = 140$ GeV, $m_H = 200$ GeV, $m_H = 350$ GeV. The MC distributions for backgrounds containing a Z boson, such as ZZ and Z+jets, show a peak around m_Z^{PDG}. The overall MC shape well reproduces the one obtained from real data.*

As an example, some distributions obtained from the 4μ-final-state events passing this 4ℓ candidate selection can be found in Fig. 4.4.

4.4.2.4 Cut on relative isolation

The two leptons with the largest isolation variable out of the four associated to the 'best candidate' are then considered. They must satisfy the requirement[5] $CombRelIso_3 + CombRelIso_4 < 0.35$. This coincides with requiring that all possible lepton pairs $\ell_i^+\ell_j^-$ have $CombRelIso_i + CombRelIso_j < 0.35$.

4.4.2.5 Cut on the significance of the 3D impact parameter

The displaced vertex of leptons originating from b-quark decays can be a handle for further background rejection. A cut is therefore applied on the significance of the 3D impact parameter of the lepton track with respect to the reconstructed primary vertex. The same cut as in Sec. 4.4.2.1, *i.e.* $|SIP_{3D}| < 4$, is now extended to all four leptons associated to the 'best candidate'.

[5] the indices refer to the list obtained after sorting leptons by increasing $CombRelIso$ isolation variable

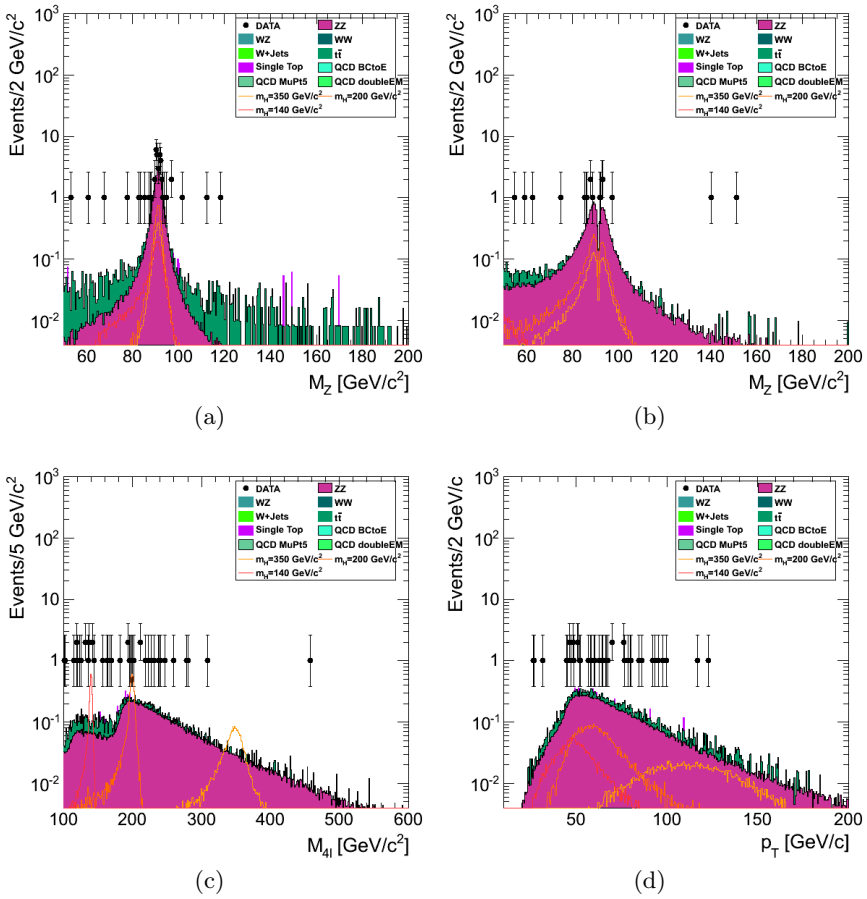


Figure 4.4: *Some distributions obtained from events passing the 'best 4ℓ candidate' selection in the 4μ final state, for data and MC samples, with an integrated luminosity of $L = 4.71$ fb^{-1}. The plotted variables are: (a) m_{Z1}, the invariant mass of the first Z candidate (b) m_{Z2}, the invariant mass of the second Z candidate (c) $m_{4\ell}$, the four-lepton invariant mass assigned to the best candidate, (d) $p_{T,1}$, the p_T of the stiffest lepton from the 'best 4ℓ candidate'.*

4.4.2.6 Cuts on Z_1, Z_2 kinematics

Finally, additional constraints are imposed on the p_T of the selected leptons ($p_T^{\ell_1,\ell_2,\mu_3,\mu_4} > 20, 10, 5, 5$ GeV for muons, $p_T^{\ell_1,\ell_2,e_3,e_4} > 20, 10, 7, 7$ GeV for electrons) and on the invariant mass of the Z candidates. There are three possibilities:

- the *baseline* selection:
  $50 < m_{Z_1} < 120$ GeV, $12 < m_{Z_2} < 120$ GeV

- the *intermediate-mass* selection:
  $60 < m_{Z_1} < 120$ GeV, $20 < m_{Z_2} < 120$ GeV
- the *high-mass* selection:
  $60 < m_{Z_1} < 120$ GeV, $60 < m_{Z_2} < 120$ GeV

The high-mass selection is used to measure the $ZZ \to 4\ell$ cross section [49], which is an interesting byproduct of the $H \to ZZ^{(*)} \to 4\ell$ analysis.

The gain that can be obtained by relaxing the cuts on m_{Z_1}, m_{Z_2} in terms of signal efficiency at low Higgs mass values is well visible from Fig. 4.5.

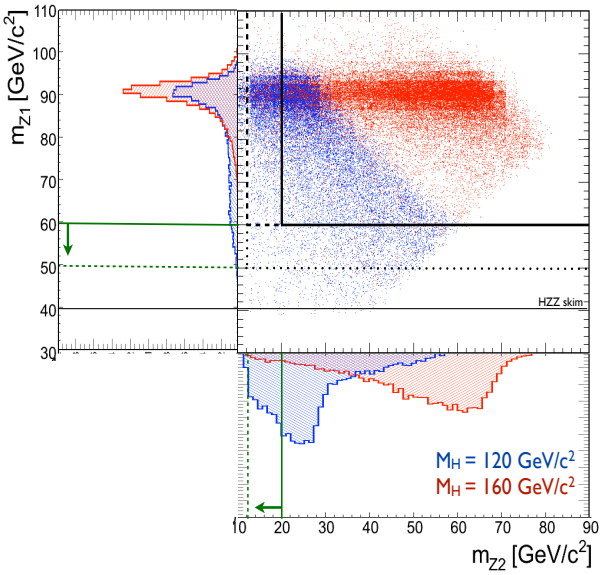


Figure 4.5: *Scatter plot showing the correlation between the m_{Z_1} and the m_{Z_2} variables for MC signal events generated in the mass hypotheses $m_H = 120$ GeV (blue dots) and $m_H = 160$ GeV (red dots). The arrows show what changes when moving from the intermediate-mass selection to the baseline one. In case of low m_H values, a significant fraction of signal events that would be discarded by requiring $m_{Z_2} > 20$ GeV passes the baseline selection by virtue of the relaxed cut on m_{Z_2}.*

4.4.3 Results

The event yields for each channel ($4e$, 4μ, $2e2\mu$), for several data and MC sample and for each step of the selection sequence are listed in Table 4.4. The same results are plotted in Fig. 4.6. The signal event yield decreases by about one order of magnitude across the whole selection, but the backgrounds are considerably suppressed. The most selective requirements against the reducible backgrounds are those up to the 'best 4ℓ candidate' choice.

(a)

Cut	QCD	tt	Z+jets	Zbb/cc	WZ	ZZ	$m_H = 200$	Total	Data
HLT	1.12×10^6	7.14×10^3	1.42×10^6	6.56×10^5	652	161	30.9	$(3.25 \pm 0.02) \times10^6$	4.0×10^6
Z_1	1.77×10^4	4.65×10^3	1.24×10^6	5.76×10^5	510	119	25.4	$(1.84\pm0.002)\times10^6$	1.9×10^6
Z_1+ℓ	0	126	2.7×10^3	2.1×10^3	136	36.4	8.09	$(5.12\pm0.04)\times10^3$	6.1×10^3
'best 4ℓ'	0	0.703	0.793	1.98	0.235	13.5	3.62	17.3±1.05	26
Isolation	0	0.0553	0	0.397	0.107	12.9	3.43	13.4±0.399	14
SIP_{3D}	0	0.0158	0	0	0.0843	12.3	3.30	12.4±0.0413	13
baseline	0	0.0158	0	0	0.0588	11.7	3.30	11.8±0.0397	13
intermediate	0	0.0158	0	0	0.0537	11.3	3.28	11.4±0.0389	12
high-mass	0	0.0079	0	0	0.0153	9.91	3.05	9.93±0.0343	9

(b)

Cut	QCD	tt	Z+jets	Zbb/cc	WZ	ZZ	$m_H = 200$	Total	Data
HLT	1.09×10^6	1.17×10^4	1.52×10^6	7.09×10^5	688	173	33.7	$(3.37\pm0.01)\times10^6$	3.6×10^6
Z_1	908	5.25×10^3	1.43×10^6	6.66×10^5	582	136	28.8	$(2.11\pm0.001)\times10^6$	2.1×10^6
Z_1+ℓ	0	290	1.31×10^3	3.9×10^3	176	45.5	9.69	$(5.76\pm0.05)\times10^3$	7.0×10^3
'best 4ℓ'	0	3.31	0.793	13.1	0.197	20.8	5.26	38.4±2.35	49
Isolation	0	0.079	0	0.793	0.0639	19.9	5.02	20.9±0.567	32
SIP_{3D}	0	0.0237	0	0	0.0383	19.5	4.93	19.7±0.0674	26
baseline	0	0.0158	0	0	0.0332	18.7	4.93	18.8±0.0661	23
intermediate	0	0.0158	0	0	0.0307	17.2	4.91	17.3±0.0647	21
high-mass	0	0	0	0	0.00511	14.5	4.55	14.5±0.0398	14

(c)

Cut	QCD	tt	Z+jets	Zbb/cc	WZ	ZZ	$m_H = 200$	Total	Data
HLT	1.43×10^6	1.97×10^4	2.96×10^6	1.37×10^6	1.35×10^3	312	56.3	$(5.87\pm0.02)\times10^6$	8×10^6
Z_1	1.8×10^4	9.97×10^3	2.69×10^6	1.25×10^6	1.1×10^3	233	48.1	$(3.97\pm0.002)\times10^6$	4.0×10^6
Z_1+ℓ	40.7	637	5×10^3	6.27×10^3	298	81.4	18.0	$(1.24\pm0.008)\times10^4$	1.4×10^4
'best 4ℓ'	0	5.7	1.19	18.6	0.457	35.1	8.76	61.6±2.82	66
Isolation	0	0.142	0.397	3.57	0.233	33.0	8.09	37.5±1.26	43
SIP_{3D}	0	0.0869	0	1.59	0.194	32.1	7.84	34±0.8	39
baseline	0	0.0553	0	1.59	0.148	30.9	7.84	32.6±0.798	37
intermediate	0	0.0553	0	1.19	0.125	29.0	7.82	30.4±0.692	33
high-mass	0	0.0237	0	0.397	0.0562	25.4	7.42	25.8±0.404	30

Table 4.4: *Event yields in the (a) 4e, (b) 4μ and (c) 2e2μ channel for the trigger and the seven event selection steps, with the third step (see Sec.* **??***) and the fourth one (see Sec. 4.4.2.3) merged under the 'best 4ℓ' category. The samples correspond to an integrated luminosity of $\mathcal{L} = 4.71$ fb^{-1}. The MC yields are not corrected for background expectation.*

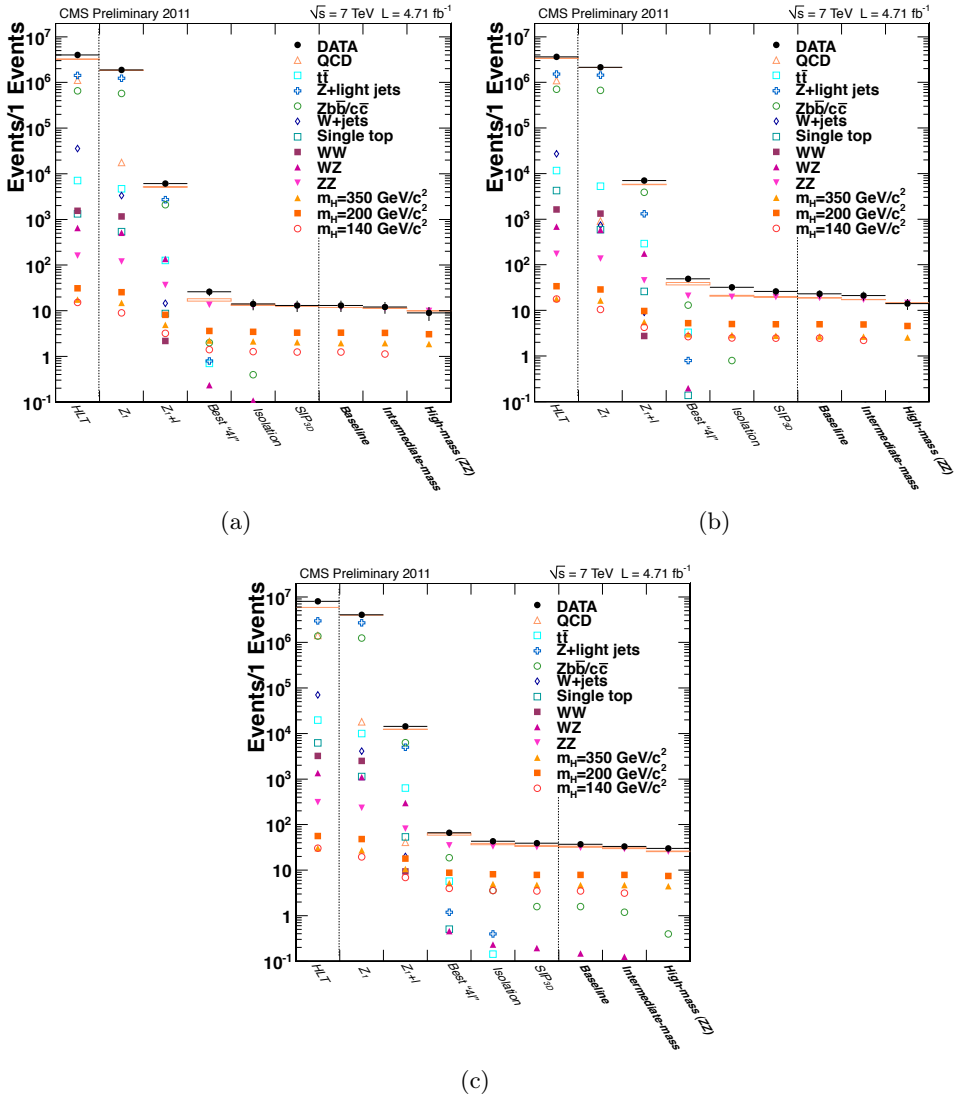


Figure 4.6: *Event yields in the (a) 4e, (b) 4μ and (c) 2e2μ channel as a function of the selection steps. The samples correspond to an integrated luminosity of* $\mathcal{L} = 4.71$ fb^{-1}. *The MC yields are not corrected for background expectation.*

The reconstructed four-lepton invariant mass distributions for the $4e$, 4μ and $2e2\mu$ channels, as well as for the overall 4ℓ case, are presented in Fig. 4.7(a, b, c, d, respectively). The distributions from data are compared to the MC expected ones.

With the current amount of data, the combination of the three channels does not show any particular clusterisation around any given mass.

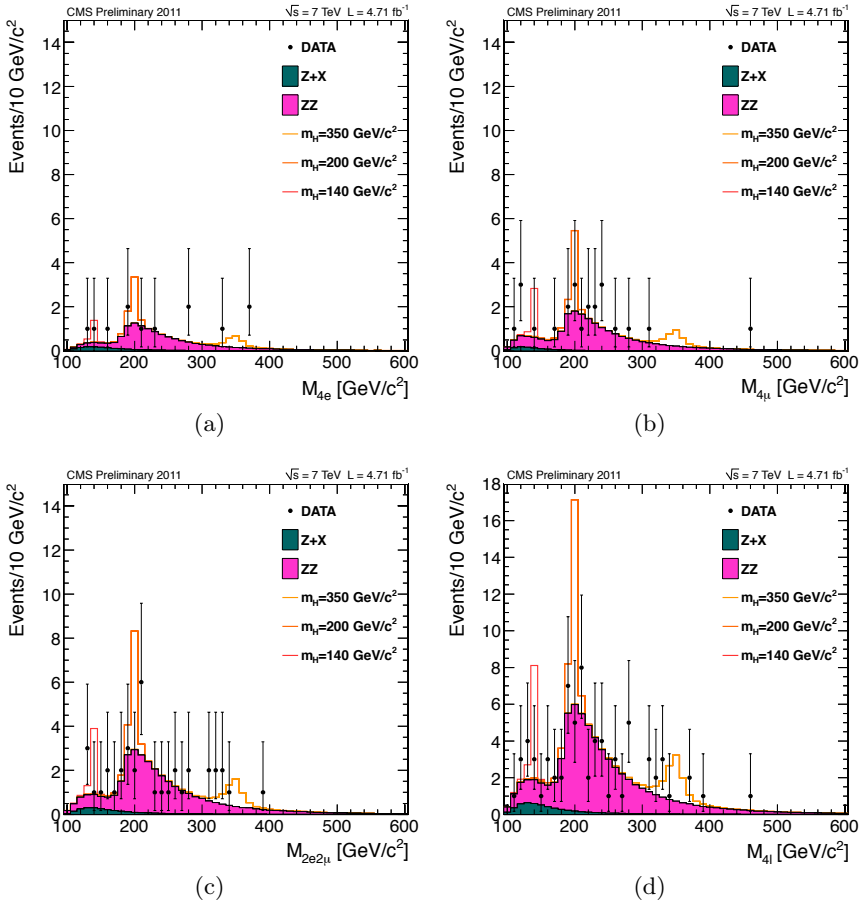


Figure 4.7: *Distribution of the four-lepton reconstructed mass for the baseline selection in the (a) 4e, (b) 4μ, (c) 2e2μ, and (d) the sum of the 4ℓ channels. These results refer to an integrated luminosity of* $\mathcal{L} = 4.71$ fb^{-1}.

The number of observed events from data samples and the background event yields in the signal region within the mass range $100 < m_{4\ell} < 600$ GeV after the baseline selection are reported for each final state in Table 4.5.

A zoom of the four-lepton invariant mass distribution on the low mass range ($m_H < 160$ GeV) is shown in Fig. 4.8 for the combination of the three final states. The reducible and instrumental backgrounds turn out to contribute very little or negligibly.

Table 4.5: *Number of observed event candidates passing the baseline selection and background and signal yields for each final state, for* $\mathcal{L} = 4.71$ fb^{-1} *and in the mass range* $100 < m_{4\ell} < 600$ GeV. *For the Z+X background, the data driven estimations are listed.*

Baseline	$4e$	4μ	$2e2\mu$
ZZ	12.27 ± 1.03	19.11 ± 1.54	30.25 ± 2.46
Z+X	1.67 ± 0.55	1.13 ± 0.55	2.71 ± 0.96
All background	13.94 ± 1.17	20.24 ± 1.64	32.96 ± 2.64
$m_{\mathrm{H}} = 120$ GeV	0.25	0.62	0.68
$m_{\mathrm{H}} = 140$ GeV	1.32	2.48	3.37
$m_{\mathrm{H}} = 350$ GeV	1.95	2.61	4.64
Observed	12	23	37

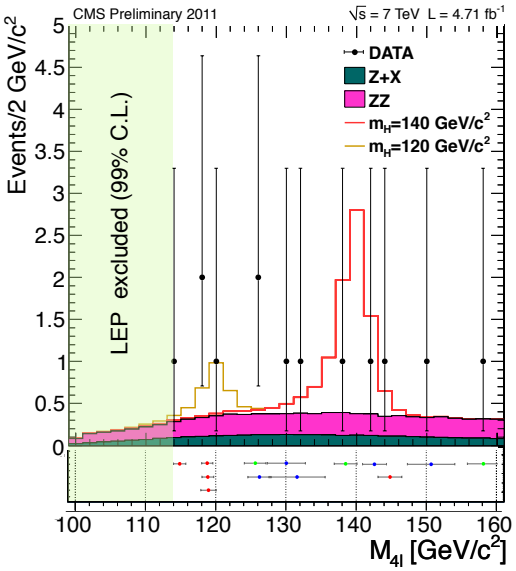


Figure 4.8: Four-lepton invariant mass distribution for the sum of the 4ℓ channels in the low-mass domain in the region $m_H < 160$ GeV. Points refer to data, shaded histograms represent the signal and background expectations. Central values and event-by-event mass uncertainties are shown. In the bottom frame, the $m_{4\ell}$ distribution for the final 4ℓ candidates from data is plotted irrespective of binning. The results are presented for an integrated luminosity of 4.71 fb^{-1}.

4.5 Background estimation and control

With the current amount of data recorded by CMS, there are not enough events for backgrounds to be precisely estimated from the sidebands of a given distribution, therefore other methods are considered. The basic principle of 4ℓ background estimation is that a *control region* can be defined by relaxing some cuts, in such a way that (ideally) only the backgrounds that one wants to measure populate it, with enough events. Then the event yield for the background B in the *signal region* spanning the 4ℓ invariant mass range $[m_1, m_2]$ is extrapolated from the one in the control region by means of the following formula:

$$N^B_{exp}[m_1, m_2] = N^B_{control} \times \left(\frac{A^B_{signal}}{A^B_{control}} \right) \times \int_{m_1}^{m_2} \rho^B(m)\, dm, \tag{4.1}$$

where $N^B_{control}$ is the number of background events in the control region, A^B_{signal} and $A^B_{control}$ are the background acceptances in the signal and in the control region, respectively, and $\rho^B(m)$ is the background density per unit interval of $m_{4\ell}$. The acceptances are derived from MC samples, with the same definition of signal and control region as for the data sample.

4.5.1 $ZZ^{(*)}$ background

The $ZZ^{(*)}$ background is the most important one surviving the selection described in Sec. 4.4. Its event yield is measured from the $Z \to \ell^+\ell^-$ one, which is then used to perform an extrapolation to the 4ℓ signal region. The control region is defined by the selection cuts up the stage described in Sec. 4.4.2.1, *i.e.* up to the selection of the first Z candidate, but with the addition of a veto on a third lepton. Three Z_1 mass windows are considered: $50 < m_{Z_1} < 120$ GeV for the baseline selection, $60 < m_{Z_1} < 120$ GeV for the intermediate- and high-mass ones. The $m_{\ell\ell}$ distribution for events in the control region is fit with the convolution of two functions, one of which accounts for contributions from backgrounds different from the ZZ one. The signal region is the one defined by the whole selection sequences, up to the kinematical cuts detailed in Sec. 4.4.2.6.

With reference to Eq. 4.1, the ratio of acceptances is evaluated as:

$$\frac{A^{ZZ}_{signal}}{A^{ZZ}_{control}} = \frac{A^{ZZ\to 4\ell}_{signal}}{A^{Z\to\ell\ell}_{control}} = R^{\sigma}_{theory} \times R^{\varepsilon}_{MC}, \tag{4.2}$$

where R^{σ}_{theory} is the ratio of the $ZZ \to 4\ell$ cross section to the $Z \to 2\ell$ one, summed over the $q\bar{q}$ and the gg contribution:

$$R^{\sigma}_{theory} = R^{\sigma}_{theory,q\bar{q}} + R^{\sigma}_{theory,gg} = \frac{\sigma^{q\bar{q}\to ZZ\to 4\ell}_{NLO}}{\sigma^{pp\to Z\to 2\ell}_{NNLO}} + \frac{\sigma^{gg\to ZZ\to 4\ell}_{LO}}{\sigma^{pp\to Z\to 2\ell}_{NNLO}} \tag{4.3}$$

and R^{ε}_{MC} is the ratio of selection efficiencies, which has been calculated from MC samples:

$$R^{\varepsilon}_{MC} = \frac{\varepsilon^{ZZ\to 4\ell}_{MC}}{\varepsilon^{Z\to 2\ell}_{MC}}. \tag{4.4}$$

Further details about the calculation of the cross sections entering Eq. 4.3 can be found in Table 4.6.

process	pert. order	calculated with	$R^{\sigma} = \frac{\sigma(ZZ\to 4\ell)}{\sigma(Z\to 2\ell)}$	$\frac{\delta R^{\sigma}_{4\ell}}{R^{\sigma}_{4\ell}}$
$q\bar{q} \to ZZ \to 4\ell$	NLO	`MCFM 5.8` [51]	$1.270 \cdot 10^{-5}$ $(4e)$ $1.270 \cdot 10^{-5}$ (4μ) $2.404 \cdot 10^{-5}$ $(2e2\mu)$	8.3%
$gg \to ZZ \to 4\ell$	LO	`GG2ZZ` [52]	$5.709 \cdot 10^{-7}$ $(4e)$ $5.709 \cdot 10^{-7}$ (4μ) $1.142 \cdot 10^{-6}$ $(2e2\mu)$	30%
$pp \to Z \to 2\ell$	NNLO	`FEWZ 2.0` [53]	-	-

Table 4.6: *Details about cross section calculation for ZZ background estimation.*

Always referring to Eq. 4.1, the event density $\rho^{ZZ}(m)$ is obtained by fitting the $m_{4\ell}$ distribution obtained from a $ZZ^{(*)}$ MC sample, both after the baseline and after the high-mass selection, with an empirical fit function[6].

The numerical values of all the quantities listed above can be found in [54]. Here the final results for the number of $ZZ^{(*)} \to 4\ell$ events in the signal region are reported, in Table 4.7. The table shows, for comparison, also the results of a cross-check based only on MC events. The expected number of $ZZ^{(*)}$ events in the signal region can be evaluated from MC by simply counting the number of events passing the selection:

[6] The functional form used for the fit (see [50] for details) is:

$$\frac{dN}{dm} = f_1 \times F(m, \hat{a}) + (1 - f_1) \times F(m, \hat{b}), \tag{4.5}$$

where $\hat{a}$, $\hat{b}$ are vectors of four parameters and

$$F(m, \hat{a}) = (0.5 + 0.5 \times sign(m - a_1) \times erf\left(\frac{|m - a_1|}{a_2}\right) \times exp\left(\frac{m}{a_3}\right). \tag{4.6}$$

$$N^{ZZ}_{expected}[m_1, m_2] = \int_{m_1}^{m_2} \left(\frac{dN}{dm_{4\ell}}\right)_{MC} dm_{4\ell}. \tag{4.7}$$

The two methods are in good agreement within the uncertainties.

baseline			
	channel	Normalization to Z rate	MC model simulation
$q\bar{q}$	$N^{ZZ\to 4e}$	11.0 ± 1.0	10.9 ± 1.0
	$N^{ZZ\to 4\mu}$	17.3 ± 1.5	17.1 ± 1.6
	$N^{ZZ\to 2e2\mu}$	29.1 ± 2.6	28.6 ± 2.9
gg	$N^{ZZ\to 4e}$	0.78 ± 0.24	0.77 ± 0.23
	$N^{ZZ\to 4\mu}$	1.12 ± 0.34	1.12 ± 0.34
	$N^{ZZ\to 2e2\mu}$	1.88 ± 0.56	1.84 ± 0.57
intermediate-mass			
	channel	Normalization to Z rate	MC model simulation
$q\bar{q}$	$N^{ZZ\to 4e}$	10.6 ± 0.9	10.4 ± 1.0
	$N^{ZZ\to 4\mu}$	15.7 ± 1.4	15.6 ± 1.5
	$N^{ZZ\to 2e2\mu}$	27.2 ± 2.5	26.8 ± 2.7
gg	$N^{ZZ\to 4e}$	0.77 ± 0.23	0.76 ± 0.23
	$N^{ZZ\to 4\mu}$	1.10 ± 0.33	1.10 ± 0.34
	$N^{ZZ\to 2e2\mu}$	1.85 ± 0.55	1.82 ± 0.56
high-mass			
	channel	Normalization to Z rate	MC model simulation
$q\bar{q}$	$N^{ZZ\to 4e}$	9.30 ± 0.8	9.10 ± 0.8
	$N^{ZZ\to 4\mu}$	13.3 ± 1.2	13.1 ± 1.2
	$N^{ZZ\to 2e2\mu}$	23.7 ± 2.1	23.3 ± 2.3
gg	$N^{ZZ\to 4e}$	0.72 ± 0.22	0.71 ± 0.22
	$N^{ZZ\to 4\mu}$	0.98 ± 0.29	0.97 ± 0.30
	$N^{ZZ\to 2e2\mu}$	1.71 ± 0.51	1.68 ± 0.53

Table 4.7: *ZZ background event yield, with the corresponding uncertainties, in the signal region in the mass range $100 < m_{4\ell} < 600$ GeV, estimated both from data by normalising to the measured single-Z rate and from MC samples, for baseline, intermediate and high-mass event selections.*

4.5.2 $Z + X$ background

This method allows for the simultaneous measurement of all the reducible and instrumental background yields. Another strategy is adopted to estimate only the contribution from backgrounds with semileptonic decays of heavy-flavour quarks (see Sec. 4.5.3). The basic principle, also in this case, consists

in extrapolating the number of background events in the signal region from a properly defined control region.

4.5.2.1 Definition of the control region

The selection defining the $Z + X$ control region branches off from the stage requiring the Z_1 candidate to be reconstructed (see Sec. 4.4.2.1). An additional $\ell^{\pm}\ell^{\pm}$ pair of same-sign, same-flavour (SS-SF) leptons is looked for. Since the charge misassignment probability is very small for high-p_T leptons ($\mathcal{O}(0.1\%)$), the charge constraint rejects signal events. Neither identification nor isolation requirements are imposed on the SS-SF leptons, but they must have $|SIP_{3D}| < 4$, their invariant mass $m_{\ell^{\pm}\ell^{\pm}}$ must pass the baseline, intermediate- or high-mass selection and the 4ℓ invariant mass must be $m_{4\ell} > 100$ GeV. The propagation to the signal region needs for a *single-lepton fake rate* estimation first.

4.5.2.2 Single-lepton fake rate measurement

The event selection used to estimate the fake rate from data is defined by the presence of a Z_1 candidate (see Sec. 4.4.2.1) and of exactly one additional lepton, called *'fakeable object'*, which can be either a reconstructed electron or a global or tracker muon passing the cut $|SIP_{3D}| < 4$. An additional cut on missing transverse energy, $MET < 25$ GeV, is used to suppress contributions from the WZ background.

Identification (see Sec. 4.4.1) and isolation requirements are imposed on fakeable objects. The single-lepton fake rate is defined as the ratio of the number of fakeable object passing these cuts to the total number of fakeable objects. This ratio is computed as a function of the lepton p_T, in two η regions (barrel and end caps), for muons and electrons separately and with two isolation cuts ($CombRelIso < 0.175$ and $0.175 < CombRelIso < 0.35$). Results are shown in Fig. 4.9 for electrons and in Fig. 4.10 for muons.

4.5.2.3 Extrapolation to the signal region

The expected number of $Z + jets$ events in the signal region, which spans the mass range $100 < m_{4\ell} < 600$ GeV, is derived from the following formula:

$$N_{\text{expect}}^{\text{Z+X}} = N_{control}^{\text{DATA}} \times (\frac{\text{OS}}{\text{SS}})^{\text{MC}} \times \Big[\epsilon_1(\text{p}_\text{T},\eta)_{|\text{Iso}<0.175} \times \epsilon_2(p_T,\eta)_{|Iso<0.175} + \frac{1}{2} \times \epsilon_1(p_T,\eta)_{|Iso<0.175} \times \epsilon_2(p_T,\eta)_{|0.175<Iso<0.35} + \frac{1}{2} \times \epsilon_1(p_T,\eta)_{|0.175<Iso<0.35} \times \epsilon_2(p_T,\eta)_{|Iso<0.175}\Big], \tag{4.8}$$

where:

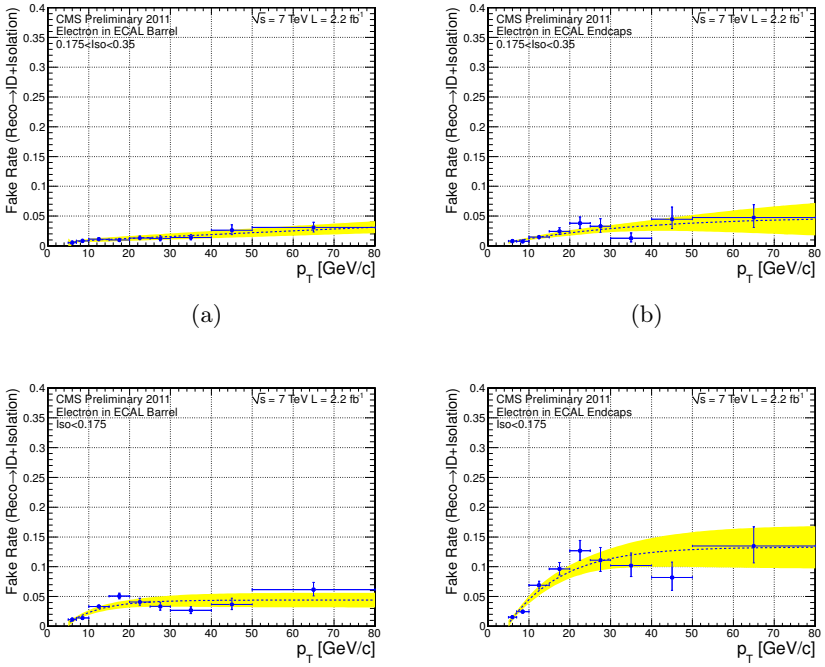


Figure 4.9: *Single-lepton fake rate distribution for electrons as a function of p_T^e, in the case $0.175 < CombRelIso < 0.35$ for electrons in the barrel (a) and in the end caps (b), and in the case $CombRelIso < 0.175$ for electrons in the barrel (c) and in the end caps (d). The dashed blue line refer to the parameterisations, the yellow bands to their statistical uncertainties.*

- $N_{control}^{\mathrm{DATA}}$ is the number of events in the control region
- $(\frac{\mathrm{OS}}{\mathrm{SS}})^{\mathrm{MC}}$ is a correction factor accounting for the slightly different event numbers obtained from a same-sign MC control sample and from an opposite-sign one from which the ZZ contribution has been subtracted. The values of this correction factor for the three final states are $(OS/SS)_{MC}^{4e} = 0.93$, $(OS/SS)_{MC}^{4\mu} = 1.28$, $(OS/SS)_{MC}^{2e2\mu} = 0.94$
- $\epsilon_i(p_T, \eta)_{|\mathrm{iso\_cut}}$ is the fake rate probability for each of the two leptons forming the $\ell_1^{\pm}\ell_2^{\pm}$ pair ($i = 1, 2$) as a function of p_T^{ℓ} and η^{ℓ} and for different isolation cuts. The three regions (A, B_1, B_2) corresponding to these isolation cuts in the ($CombRelIso_1$, $CombRelIso_2$) plane are shown in Fig. 4.11. The triangle defined by $CombRelIso_1 + CombRelIso_2 < 0.35$ is populated by the events passing the isolation cut included in the signal selection sequence (see Sec. 4.4.2.4).

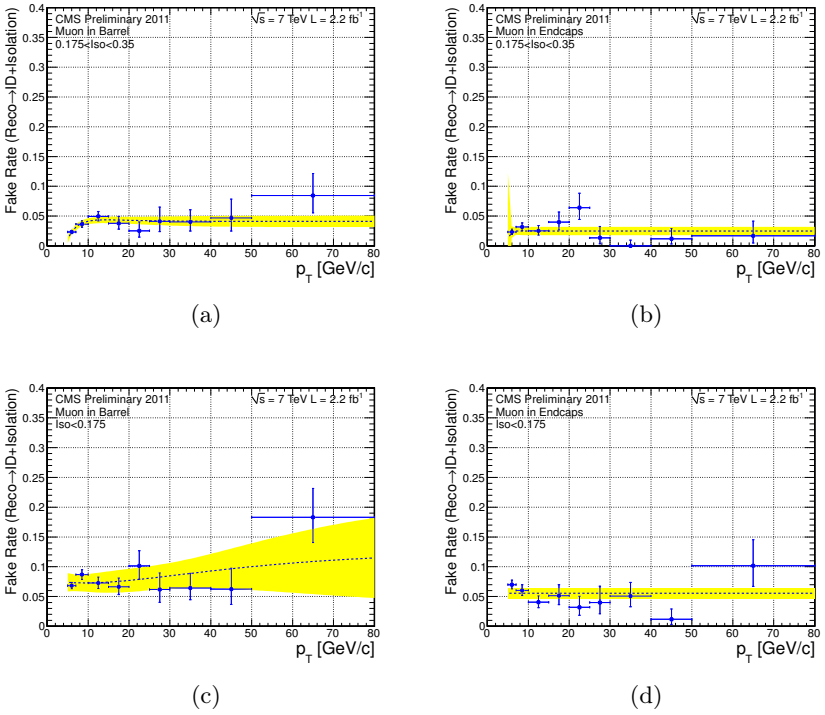


Figure 4.10: *Single-lepton fake rate distribution for muons as a function of p_T^{μ}, in the case $0.175 < CombRelIso < 0.35$ for muons in the barrel (a) and in the end caps (b), and in the case $CombRelIso < 0.175$ for muons in the barrel (c) and in the end caps (d). The dashed blue line refer to the parameterisations, the yellow bands to their statistical uncertainties.*

Each of the three terms summed up in Eq. 4.8 represents the joint probability of both leptons from the $\ell_1^{\pm}\ell_2^{\pm}$ pair passing the isolation cut. Since the two leptons are supposed to be independent of each other, this probability is the product of the two single-lepton fake rates. The factor 1/2 is introduced to account for the triangular shape of the B_1, B_2 regions, under the assumption (which has been proven to be true) that the event density in the two halves of each square is approximately the same.

The number of $Z + X$ events in the signal region as obtained from Eq. 4.8 is reported in Table 4.8 for each final state and for the baseline, the intermediate- and the high-mass selection.

4.5.3 $Zb\bar{b}$, $Zc\bar{c}$, $t\bar{t}$ backgrounds

The $Zb\bar{b}$, $Zc\bar{c}$ and $t\bar{t}$ backgrounds can be estimated in a data-driven way with the following method. A control region is defined by the presence of a Z_1

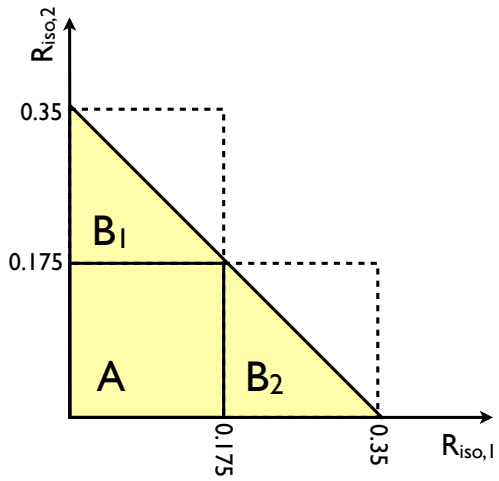


Figure 4.11: *Sketch of the signal region in the ($CombRelIso_1$, $CombRelIso_2$) plane, where the indices 1, 2 refer to the two least isolated leptons.*

	baseline
$N^{Z+X\to 4e}$	1.67 ± 0.05 (3.2%) (stat., 952 events)$\pm$0.50(30.2%)(syst.)
$N^{Z+X\to 4\mu}$	1.13 ± 0.09 (8.3%) (stat., 143 events)$\pm$0.46(40.6%)(syst.)
$N^{Z+X\to 2e2\mu}$	2.71 ± 0.08 (2.9%) (stat., 1215 events)$\pm$0.88(32.6%)(syst.)
	intermediate-mass
$N^{Z+X\to 4e}$	1.45 ± 0.05 (3.7%) (stat., 746 events)$\pm$0.3(20.4%)(syst.)
$N^{Z+X\to 4\mu}$	0.81 ± 0.08 (9.8%) (stat., 103 events)$\pm$0.26(31.7%)(syst.)
$N^{Z+X\to 2e2\mu}$	2.22 ± 0.08 (3.3%) (stat., 934 events)$\pm$0.54(22.5%)(syst.)
	high-mass
$N^{Z+X\to 4e}$	0.47 ± 0.04 (8.4%) (stat., 143 events)$\pm$0.11(22%)(syst.)
$N^{Z+X\to 4\mu}$	0.22 ± 0.03 (20.8%)(stat., 23 events)$\pm$0.06(35.7%)(syst.)
$N^{Z+X\to 2e2\mu}$	0.65 ± 0.05 (7.6%) (stat., 175 events)$\pm$0.16(23.5%)(syst.)

Table 4.8: *The expected number of $Z + X$ events and the corresponding systematics and statistical errors in the signal region in a mass range between $m_1 = 100$ GeV and $m_2 = 600$ GeV, for the baseline, the intermediate- and the high-mass selection.*

candidate and of an additional lepton pair of any flavour and charge, without any vertex or isolation requirement. In order to discard contributions from signal and $Z+$ light-jet events, the cut on impact parameter significance is reversed: $|SIP_{3D}| > 5$ for both leptons not associated to Z_1. The number of events in this control region is sufficiently large to allow for a fit to the

data (see Fig. 4.12). The convolution of a Crystal Ball with a Breit-Wigner is used to describe the Z_1 peak in the m_{Z_1} distribution, whereas the tails are fit with a Chebychev polynomial. The outcome of the fit is 151 ± 17 $Zb\bar{b}/Zc\bar{c}$ events and 84 ± 15 $t\bar{t}$ events.

The composition of the control region in terms of number of events for each sub-channel is listed in Table 4.9. The fit results have been split by final state according to the relative percentages listed in the table.

final state	event number	percentage
$Z_1 + \mu\mu$	167	71.0%
$Z_1 + e\mu$	58	24.7%
$Z_1 + ee$	10	4.3%

Table 4.9: *Break-down by final state of the number of events passing the cuts that define the $Zb\bar{b}/Zc\bar{c}, t\bar{t}$ control region. The proportion between the three sub-channels have been applied to the fit results.*

The expected number of events for a given background (*e.g.* $Zb\bar{b}/Zc\bar{c}$) and for a given final state (*e.g.* 4μ) can be obtained with the following formula:

$$N_{expected}^{Zb\bar{b}\to 4\mu} = N_{control}^{Zb\bar{b}\to 4\ell} \times \alpha_{4\mu} \times \alpha_{kin_1} \times \alpha_{SIP_{3D}} \times \alpha_{CombRelIso} \times \alpha_{kin_2}. \quad (4.9)$$

Let's see in detail how $\alpha_{4\mu}$ is calculated. It represents the scale factor for going from the $Z_1 + \mu\mu$ control-region phase space to the $\mu^+\mu^-\mu^+\mu^-$ signal-like one. The Z_1 candidate is required to decay into a couple of opposite-sign leptons, *i.e.* $\mu^+\mu^-$, e^+e^-, in either case with the same probability, therefore the $Z_1 \to \mu^+\mu^-$ decay gets a factor 1/2. The other muon pair can be either a same-sign (SS) or an opposite-sign (OS) one. The ratio (OS/SS) has been measured from data and found equal to 2.5. The value of $\alpha_{4\mu}$, expressed in terms of probabilities, is then:

$$\alpha_{4\mu} = \frac{\mathcal{P}(\mu^+\mu^-\mu^+\mu^-)}{\mathcal{P}(Z_1 + \mu\mu)} = \frac{\mathcal{P}(\mu^+\mu^-)}{\mathcal{P}(Z_1)} \times \frac{\mathcal{P}(\mu^+\mu^-)}{\mathcal{P}(\mu\mu)} = \frac{1}{2} \times \frac{2.5}{1 + 2.5}. \quad (4.10)$$

The other scale factors can be computed in a similar way. The general idea is to evaluate ratios between two event numbers derived from data: one related to the signal region, after the signal selection sequence, the other to the control region, after the 'relaxed' selection defining it. In particular,

- α_{kin_1} takes into account the kinematical cuts on m_{Z_2} and on $m_{4\ell}$ included in the signal selection

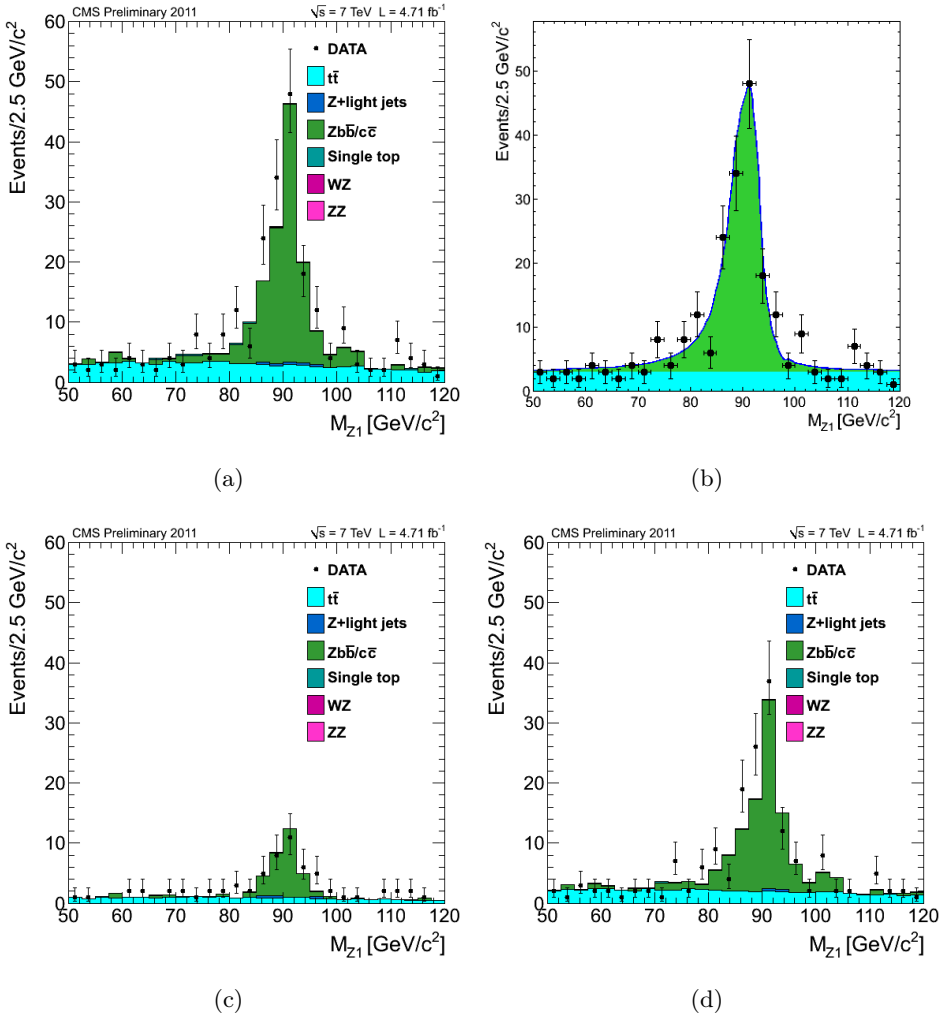


Figure 4.12: *Measurement of the Z1 mass in the Z1+2ℓ background control region for the $Zb\bar{b}$ and $t\bar{t}$. The results are shown for data (points with statistical uncertainties) and stacked Monte Carlo expectation for the backgrounds (shaded histograms) for the (a) the full statistics corresponding to $\mathcal{L} = 4.71$ fb^{-1} of integrated luminosity. The same distribution is shown for the Z1+SS (same sign leptons) events (c) and events and for the Z1+OS (opposite sign leptons) (d). A fit to the data is performed (b) using a Breit-Wigner function convoluted with a Crystal Ball function for the Z1 peak and a Chebychev Polynomials for the $t\bar{t}$ spectrum.*

- $\alpha_{SIP_{3D}} = \frac{N_{evt}(|SIP_{3D}|<4)}{N_{evt}(|SIP_{3D}|>5)}$ accounts for the reversal of the SIP_{3D} cut for the two leptons not associated to Z_1

- $\alpha_{CombRelIso} = \frac{N_{evt}(CombRelIso_i + CombRelIso_j < 0.35)}{N_{evt}}$ rescales the number of events in the control region by the reduction factor due the isolation cut on the $Z_2 \to \ell_i^+ \ell_j^-$ lepton pair

- α_{kin_2} varies depending on whether the selection is the *baseline*, the *intermediate-* or the *high-mass* one.

The event yields in the signal region as obtained from Eq. 4.9 are reported in Table 4.10.

	baseline	intermediate	high-mass
$N^{\mathrm{Z}b\bar{b}\to 4e}$	0.04 ± 0.03	0.04 ± 0.03	0.01 ± 0.01
$N^{\mathrm{Z}b\bar{b}\to 4\mu}$	0.70 ± 0.30	0.50 ± 0.20	0.08 ± 0.04
$N^{\mathrm{Z}b\bar{b}\to 2e2\mu}$	0.70 ± 0.30	0.50 ± 0.20	0.09 ± 0.04
$N^{t\bar{t}\to 4e}$	0.03 ± 0.01	0.02 ± 0.01	-
$N^{t\bar{t}\to 4\mu}$	0.02 ± 0.00	0.01 ± 0.00	-
$N^{t\bar{t}\to 2e2\mu}$	0.04 ± 0.01	0.04 ± 0.01	-

Table 4.10: *Number of events for the $Zb\bar{b}/Zc\bar{c}$ and $t\bar{t}$ backgrounds estimated from a control region with reversed $|\mathrm{SIP_{3D}}|$ cut, relaxed isolation, charge and flavour requirements for two leptons. They corresponds to an integrated luminosity of $\mathcal{L} = 4.71$ fb^{-1}.*

4.6 Systematic uncertainties

A detailed discussion about all sources of systematic uncertainties can be found in [54]. Only a short summary is reported here for completeness.

- *systematic uncertainty on the signal cross section*: it depends on PDF+α_s systematics and on QCD scale uncertainties. These two sources of uncertainties are treated as uncorrelated [55].

- *systematic uncertainty on* $BR(H \to 4\ell)$: it is estimated equal to 2% [56] and it is considered independent of m_H.

- *theoretical uncertainties at high* m_H *values*: they are related to the way running Higgs width is treated in the theory, in the limit of a very large total Higgs boson width Γ_H, and to the interference between $gg \to H \to ZZ$ signal diagrams and $gg \to ZZ$ background ones. All the uncertainties concerning high-mass Higgs boson cross sections are estimated [57] as $150\% \times (m_H/\mathrm{TeV})^3$.

- *systematic uncertainties on the signal acceptance*, which is measured to be $\mathcal{A} \sim 0.6 - 0.9$ after the kinematical cuts [58].

The contribution due to QCD scale uncertainty is estimated by varying the QCD renormalisation (μ_R) and factorisation (μ_R) scales by a factor of 2, both upwards and downwards[7]. The resulting values of $\delta\mathcal{A}/\mathcal{A} \sim 0.1 - 0.2\%$ are very small and they can be neglected. The contribution coming from PDF+α_s uncertainties is estimated by following the PDF4LHC prescription and by using the three PDF sets CT10, MSTW08, NNPDF. It results in a 2% uncertainty on the acceptance, independent of m_H.

- *systematic uncertainties on the ZZ background cross section*: both the uncertainties due to PDF+α_s systematics and those related to QCD scale uncertainties have been evaluated with the same method as in the signal case (see above).

- *systematic uncertainty on the integrated luminosity measurement*: it is 4.5% [59].

- *systematic uncertainty on trigger efficiency*: the trigger efficiency is close to 100% for the signal and the ZZ background within the acceptance considered in the 4ℓ analysis. A systematic uncertainty of 1.5% on trigger efficiency is assigned. It has been evaluated from data.

- *systematic uncertainty on lepton reconstruction and identification efficiency*: it has been measured in a data-driven way, with a *Tag-and-Probe* technique, from data/MC single lepton efficiency ratios which are then propagated to the 4ℓ Higgs boson candidates and used to correct the final MC event yield on an event-by-event basis. The value of this uncertainty is $2 - 3\%$.

- *systematic uncertainty on lepton isolation efficiency*: it has been determined as a function of p_T^ℓ and η^ℓ for a given isolation cut, with a *Tag-and-Probe* technique, from data/MC single lepton efficiency ratios. Since the isolation cut included in the event selection sequence involves two leptons, while the systematic uncertainty is estimated for a single lepton, the following procedure is adopted. The cut on the two least isolated leptons is seen as a variable cut on the least isolated one: $CombRelIso_{least} = 0.35 - CombRelIso_{next-to-least}$. The largest uncertainty on single-lepton isolation efficiency in the whole range $[0., 0.35]$, given by the largest data/MC efficiency ratio, is then taken as systematic uncertainty on isolation. Its value is about 2%.

- *systematic uncertainty on muon and electron momentum and energy scale*: it is 0.5% and it has been propagated through the selection cuts,

[7] namely, starting from the default value $\mathcal{A}_0(\mu_R = \mu_F = m_H/2)$, the acceptances in the two cases $\mathcal{A}_{up}(\mu_R = \mu_F = m_H)$ and $\mathcal{A}_{down}(\mu_R = \mu_F = m_H/4)$ are measured and the relative uncertainty is defined as $\delta\mathcal{A}/\mathcal{A} = \max|\Delta\mathcal{A}/\mathcal{A}_0|$.

resulting into an uncertainty on the 4ℓ invariant mass distributions for signal and backgrounds.

- *systematic uncertainty on 4ℓ invariant mass resolution*: an uncertainty of 30% is assigned to the width of the convolution of a Breit-Wigner and of a Crystal Ball functions used to fit the signal shape.

4.7 Exclusion limits in the $H \to ZZ \to 4\ell$ channel

The statistical procedure followed to derive upper limits on the SM Higgs boson cross section times branching ratio is the so-called CL_s method, described in Appendix A.1.

The expected and observed upper limits at 95% confidence level on the signal strength modifier $\mu = \sigma/\sigma_{SM}$ are shown in Fig. 4.13. In other words, the plot shows the factor by which the SM Higgs boson cross section times branching ratio must be scaled to be excluded at 95% CL, as a function of the Higgs mass.

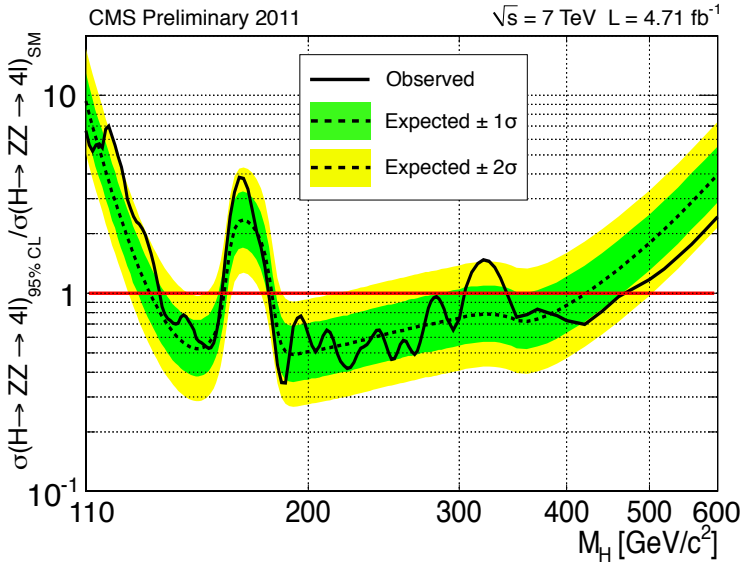


Figure 4.13: *The 95% CL upper limits on the signal strength modifier $\mu = \sigma/\sigma_{SM}$, as a function of the SM Higgs boson mass in the range 110 ÷ 600 GeV, for the $H \to ZZ^{(*)} \to 4\ell$ channel alone, with $L = 4.71$ fb^{-1}. The black line corresponds to the observed limits. The dashed line indicates the median expected limit on μ for the background-only hypothesis, while the green (yellow) bands indicate the ranges that are expected to contain 68% (95%) of all observed limit excursions from the median.*

The dashed line has been obtained by tossing toy Monte Carlo pseudo-experiments, whereas the solid line shows the results obtained from real data.

The green and yellow bands indicate the μ ranges in which an additional measurement would lie with a probability of 68% and 95%, respectively. The Higgs mass range corresponding to $\mu < 1$ is excluded at 95% CL. For the $H \to ZZ^{(*)} \to 4\ell$ channel alone, these excluded Higgs mass intervals are $134 < m_H < 158$ GeV, $180 < m_H < 305$ GeV, $340 < m_H < 460$ GeV.

4.8 Exclusion limits from the combination of all decay channels

In a similar way as for results described in Sec. 4.7, 95% CL upper limits on $\mu = \sigma/\sigma_{SM}$ have been derived for the combination of all the considered decay channels, for a total integrated luminosity of $L = 4.71$ fb^{-1}. They are shown in Fig. 4.14.

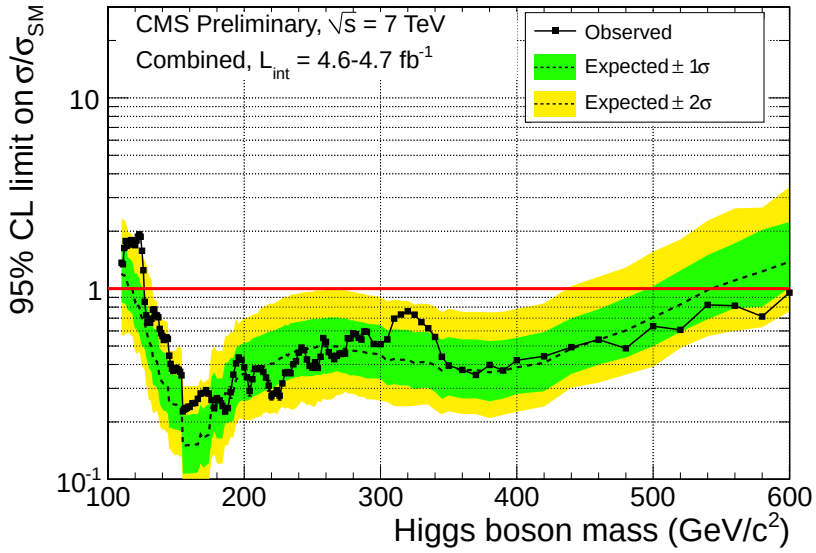


Figure 4.14: *The combined 95% C.L. upper limits on the signal strength modifier $\mu = \sigma/\sigma_{SM}$, as a function of the SM Higgs boson mass in the range $110 \div 600$ GeV, for all the channels listed in Table 4.11, with $L = 4.71$ fb^{-1}. The black line corresponds to the observed limits. The dashed line indicates the median expected limit on μ for the background-only hypothesis, while the green (yellow) bands indicate the ranges that are expected to contain 68% (95%) of all observed limit excursions from the median.*

The decay channels included in the combination are summarised in Table 4.11. The expected excluded Higgs mass range at 95% CL is $117 \div 543$ GeV, whereas the observed one is $127 \div 600$ GeV, which becomes $128 \div 525$ GeV at 99% CL.

Channel	m_H range (GeV)	Lumi (fb^{-1})	sub-channels	m_H resolution
$H \to \gamma\gamma$	$110 - 150$	4.7	4	$1 - 3\%$
$H \to \tau\tau$	$110 - 145$	4.6	9	20%
$H \to b\bar{b}$	$110 - 135$	4.7	5	10%
$H \to WW \to \ell\nu\ell\nu$	$110 - 600$	4.6	5	20%
$H \to ZZ \to 4\ell$	$110 - 600$	4.7	3	$1 - 2\%$
$H \to ZZ \to 2\ell 2\tau$	$190 - 600$	4.7	8	$10 - 15\%$
$H \to ZZ \to 2\ell 2\nu$	$250 - 600$	4.6	2	7%
$H \to ZZ \to 2\ell 2q$	$130 - 164$ $200 - 600$	4.6	6	3%

Table 4.11: *Some details about the channels included in the statistical combination for the SM Higgs boson search. Here ℓ stands for either an electron or a muon.*

4.9 The low mass region

A zoomed version of Fig. 4.13 on the low Higgs mass range is shown in Fig. 4.15. One can see that the $H \to ZZ^{(*)} \to 4\ell$ alone allows to exclude at 95% CL the Higgs mass range $134 < m_H < 158$ GeV. Data points indicate a $\sim 2\sigma$-wide deviation from the expected limit in a mass range around $118 \div 130$ GeV.

The low-mass results obtained by combining all channels are presented in Fig. 4.16. An excess of events around $m_H = 124$ GeV makes the observed exclusion limit in that region weaker than the expected one. Further details about this excess are given in Sec. 4.10.

4.10 The local and global p-values

The compatibility of the observed excesses of data events with the background-only hypothesis can be quantified by the $p-$value, as defined in Eq. A.4. Its trend as a function of the SM Higgs boson mass is plotted in Fig. 4.17 (top frame). The p_0 value is the probability of observing an upwards fluctuation of the background event yield at least as large as the one measured in data. It is estimated by tossing a large number of toy MC pseudo-datasets. One can see two dips. One of them, at $m_H = 119$ GeV, is due to three Higgs candidates passing the $H \to ZZ^{(*)} \to 4\ell$ selection. The other one, at $m_H = 124$ GeV, is mostly due to an excess in the $H \to \gamma\gamma$ channel. It has the minimum value, $p_{min} = 0.005$, and the largest local significance, 2.6σ.

The compatibility of the observed excess with the presence of a signal can be assessed with the plot of Fig. 4.17 (bottom frame). It shows the best-fit

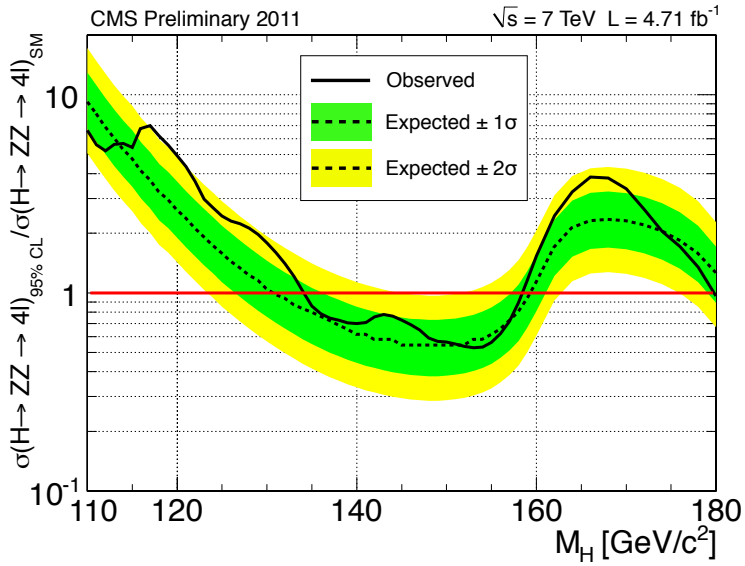


Figure 4.15: *The 95% CL upper limits on the signal strength modifier $\mu = \sigma/\sigma_{SM}$, as a function of the SM Higgs boson mass in the range 110 ÷ 180 GeV, for the $H \to ZZ^{(*)} \to 4\ell$ channel alone, with $L = 4.71$ fb^{-1}. The black line corresponds to the observed limits. The dashed line indicates the median expected limit on μ for the background-only hypothesis, while the green (yellow) bands indicate the ranges that are expected to contain 68% (95%) of all observed limit excursions from the median.*

$\hat{\mu}$ value of the signal strength modifier μ as a function of the Higgs mass. This value is the factor by which the SM Higgs boson cross section should be rescaled to best agree with the experimental data. It turns out that the whole mass region not excluded at 95% CL is within the $\pm 1\sigma$ uncertainty band from the best-fit points, therefore it is not incompatible with the signal hypothesis.

A correct physical interpretation of these results is actually based on the *global* $p-$value, not on the *local* one. The former can be obtained from the latter by calculating a *trial factor* [66] that accounts for the *'look-elsewhere effect'*. The 2.6σ excess at $m_H = 124$ GeV has a local $p-value$ of 0.005, which turns into a global $p-$value of 0.026 if one considers the range $110 < m_H < 145$ GeV. The corresponding global significance would be 1.9σ (see Fig. 4.18). If the look-elsewhere effect is considered over the whole mass range, $110 < m_H < 600$ GeV, the global significance of the same excess goes down to 0.6σ.

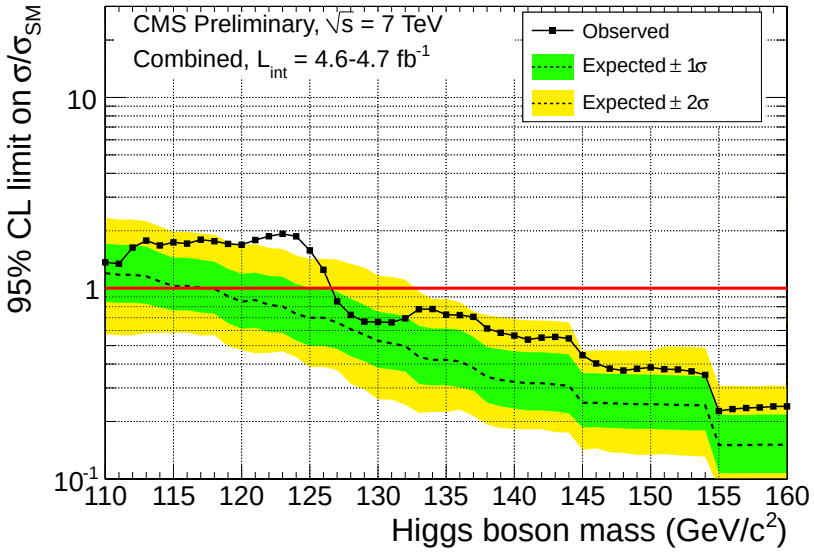


Figure 4.16: *The 95% CL upper limits on the signal strength modifier $\mu = \sigma/\sigma_{SM}$, as a function of the SM Higgs boson mass in the range 110÷160 GeV, for all the considered channels, with $L = 4.71$ fb^{-1}. The black line corresponds to the observed limits. The dashed line indicates the median expected limit on μ for the background-only hypothesis, while the green (yellow) bands indicate the ranges that are expected to contain 68% (95%) of all observed limit excursions from the median.*

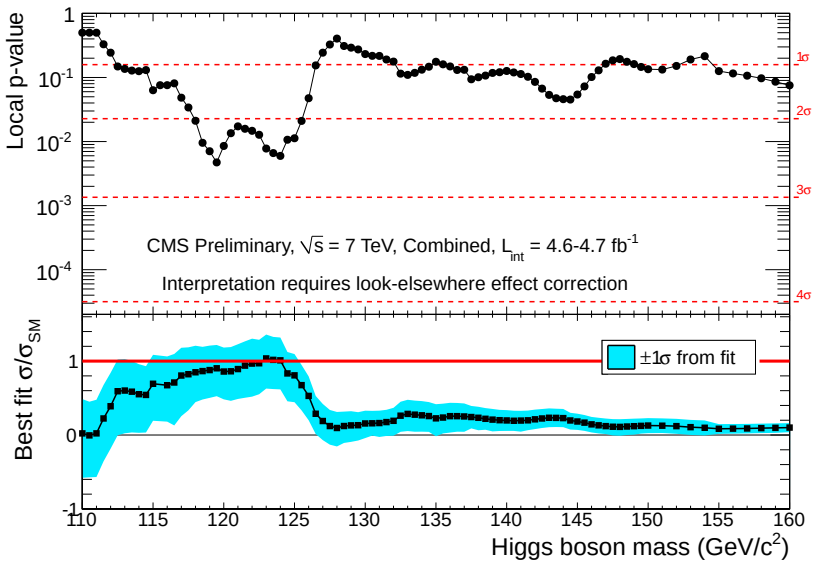


Figure 4.17: *Top panel: the observed local p-value p_0 as a function of m_H. The horizontal red lines correspond to the number of standard deviations (σ) of a one-tail Gaussian distribution for which the area of the tail would equal p_0. As specified in the figure, a correct physical interpretation of this plot can be done only after the look-elsewhere effect has been taken into account. Bottom panel: the best-fit value $\hat{\mu}$ of the signal strength modifier $\mu = \sigma/\sigma_{SM}$, as a function of m_H. The blue band corresponds to $\pm 1\sigma$ (statistical + systematical) fluctuations around the central value.*

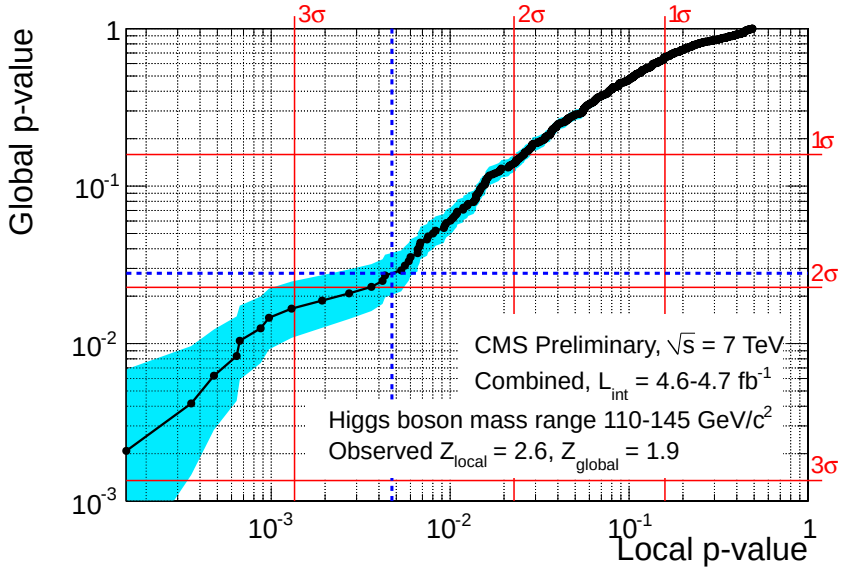


Figure 4.18: *The global p−value as a function of the local p−value in the restricted mass range* $110 < m_H < 145$ GeV. *The global p-value is computed by generating several sets of pseudo-data. The crossing point of the two blue dashed lines corresponds to the most significant excess ov events in data* ($m_H = 124$ GeV).

4.11 Perspectives for SM Higgs boson searches in 2012

The LHC is scheduled to run at the centre-of-mass energy of $\sqrt{s} = 8$ TeV [67] between March and November 2012, before a 20-month long shut-down that will prepare the machine for operations at even higher energies. The target integrated luminosity by the end of 2012, both for CMS and for ATLAS, is 15 fb^{-1}. The bunch spacing time will remain 50 ns as it is now.

The advantage of running at 8 TeV is twofold. From Fig. 4.19 [68] one can see that the gain in luminosity obtained by moving from $\sqrt{s} = 7$ TeV to 8 TeV is a factor ~ 1.2 for mass values $\mathcal{O}(10^2$ GeV). Also the Higgs production cross section is larger at 8 TeV than at 7 TeV, by $20 - 30\%$, depending on the mass and on the production mode, as shown in Fig. 4.20.

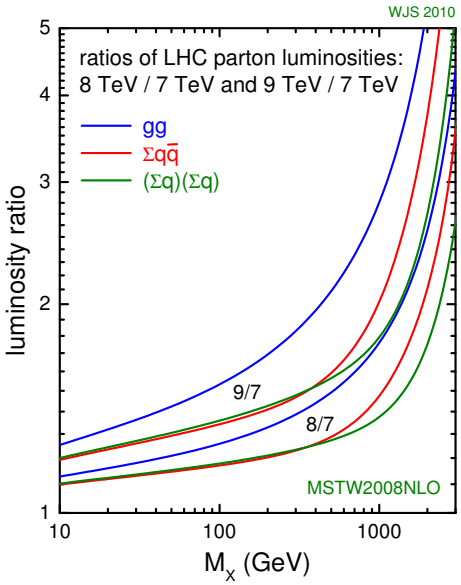


Figure 4.19: *Ratio of parton luminosities for different centre-of-mass energies (dashed lines: 9 to 7 TeV, solid lines: 8 to 7 TeV) at the LHC, for gluon pairs (in blue), $q\bar{q}$ pairs (in red) and qq pairs (in green), as a function of the mass of the system produced in the hard scattering process. The LHC is going to be operated at $\sqrt{s} = 8$ TeV in 2012.*

The expected significance that can be reached with CMS for a SM-like Higgs boson discovery is shown in Fig. 4.21, for several centre-of-mass energy and total integrated luminosity scenarios. These projections have been obtained by rescaling the event yields by appropriate factors accounting for the different luminosity and cross section values. The procedure for calculating the signal significance has then be rerun on the projected numbers. A 5σ-discovery is foreseen to be possible with $L = 15$ fb^{-1} in the whole Higgs

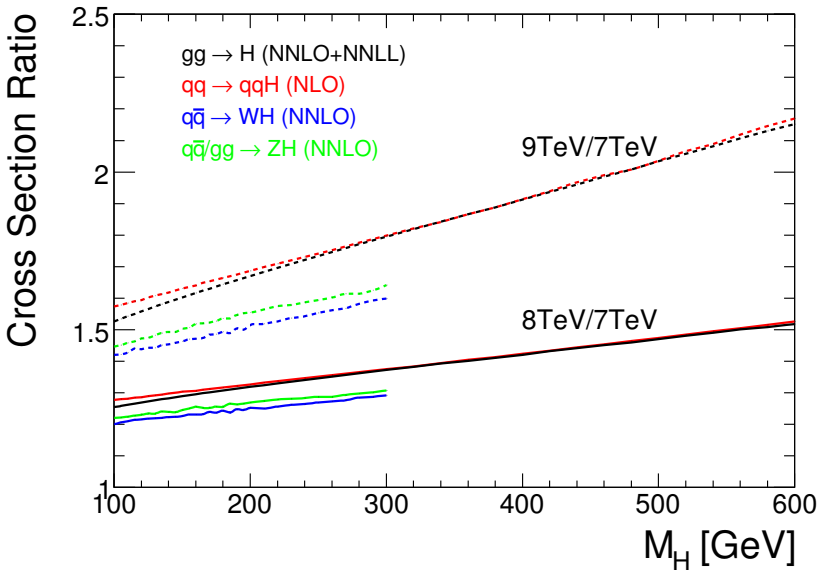


Figure 4.20: *Ratio of SM Higgs production cross sections for different centre-of-mass energies (9 to 7 TeV, 8 to 7 TeV) as a function of the Higgs boson mass. Four different production modes are considered: gluon-gluon fusion (black), vector boson fusion (red), 'Higgs-strahlung' off W boson (blue), 'Higgs-strahlung' off Z boson (green).*

mass range that has not been ruled out with the current amount of data ($114.4 < m_H < 127$ GeV).

If the Higgs boson is found, then its properties (mass, width, couplings, spin, branching ratios into bosons and fermions) need to be precisely measured. Also its production and decay modes will have to be studied in detail. If, on the contrary, the SM Higgs boson is not found by the end of 2012, several theories beyond the Standard Model will attract more and more interest both from the theoretical and from the experimental communities. In both cases, interesting studies about vector boson scattering and its properties can be carried out.

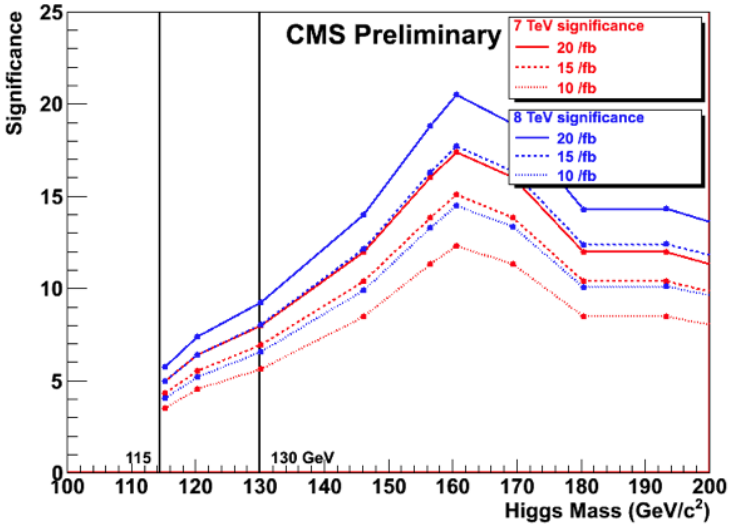


Figure 4.21: *Expected significance that can be attained by SM Higgs boson searches with the CMS detector, as a function of the Higgs mass, in several centre-of-mass energy and total integrated luminosity scenarios:* $\sqrt{s} = 7$ TeV *(in red),* $\sqrt{s} = 8$ TeV *(in blue);* $L = 20$ fb^{-1} *(solid line),* $L = 15$ fb^{-1} *(coarsely dashed line),* $L = 10$ fb^{-1} *(finely dashed line). The two black vertical lines roughly indicate the mass range in which the SM Higgs boson has not been excluded at* 95% *CL with the current amount of data.*

Conclusions

The SM Higgs boson search in the $H \to ZZ^{(*)} \to 4\ell$ channel takes advantage of the very clean signature of this final state, with two pairs of same-flavour, opposite-charge, high-p_T isolated leptons compatible with the same reconstructed vertex. Lepton isolation is a useful handle to discriminate between signal and background events, because leptons originating from $b-$quark decays in backgrounds such as $Zb\bar{b}$, $t\bar{t}$, Z+jets, W+jets tend to be close or within a jet. The isolation variable provides an estimate of the energy flow in the vicinity of a lepton track, which is, on average, larger for background events than for signal ones.

This thesis has intended to demonstrate the importance of lepton isolation in the context of the 4ℓ final state analysis. The implementation of lepton isolation algorithms in this analysis has undergone frequent changes during last years (as least since when I started working on this topic in 2008) but what has remained unchanged over this period is the procedure adopted to optimise the performances of the algorithm as a function of signal efficiency and background rejection. The results of this optimisation, as well as the rationale behind this procedure, have been discussed in Chapter 3.

The number of pile-up vertices for each event has been increasing gradually but by a significant amount and it is foreseen to continue its growth. During 2012 run, its peak value for instantaneous luminosities $\mathcal{L} = 5.1(6.2, 6.8) \times 10^{33}$ cm^{-2}s^{-1} is expected to be 26 (31, 35). Isolation variables are affected by high pile-up conditions insofar as calorimeters do not allow to associate a given energy deposit to the correct event vertex. I have implemented in the 4ℓ analysis framework and optimised the algorithm (usually referred to as *ρ-correction*) that makes the lepton isolation efficiency distribution uniform as a function of the number of reconstructed primary vertices. The performances of this algorithm and a cross-check derived from a completely independent study have been discussed in this thesis.

The $H \to ZZ^{(*)} \to 4\ell$ analysis has been described in Chapter 4. After a brief overview of the considered data and MonteCarlo samples, the selection sequence has been presented, starting from the requirements on electron and muon reconstruction and identification, moving then to the 'pre-selection' cuts defining the set of four leptons associated to the 'best 4ℓ candidate', up to kinematical cuts delineating the signal phase-space. Discussions are

ongoing about restructuring the selection sequence and relaxing some cuts, in order to allow a more detailed exploration of the low Higgs mass region.

My personal contribution to the 4ℓ analysis is mostly related to cut optimisation, which had to be repeated several times in the past to ensure that the trade-off between signal efficiency and background rejection was performing at its best. This effort pays off, since the overall signal efficiency is several orders of magnitude larger than the background one, as demonstrated in Chapter 4.

The techniques for estimating the background contributions to the final event yield have been described and a summary of the systematic uncertainties affecting the 4ℓ analysis has been provided.

Last but not least, the statistical conclusions that can be drawn from the current amount of data have been presented in Sec. 4.7 *ff.* both for SM Higgs boson exclusion and for its discovery. The SM Higgs mass range that has not been excluded at 95% CL has shrunk to $114.4 < m_H < 127$ GeV. It is foreseen that with the data that will be taken in 2012 (hopefully $L = 15$ fb^{-1}) it will be possible either to exclude the existence of the SM Higgs boson over the whole allowed mass range, if it does not exist, or to discover it with a $\sim 5\sigma$ significance for any mass value, if it exists. The currently available data show two interesting excesses of event, around $m_H = 119$ GeV and $m_H = 124$ GeV. This latter one is not incompatible with an excess observed by the ATLAS collaboration around $m_H = 126$ GeV. Only a larger data set will allow us to state the ultimate answer to the question of the existence of the Higgs boson.

Ringraziamenti

Denken ist Danken
('thinking is thanking')

M. Heidegger

Per tutto quello che ho vissuto in questi tre anni, e soprattutto per l'esperienza di quasi due anni passati continuativamente al CERN (e non due anni qualsiasi, ma un periodo senza precedenti, in particolare per le ricerche sull'Higgs...), mi considero un privilegiato. Il primo sentimento che provo quando penso di essere un privilegiato è un grandissimo *grazie*.

Grazie a tutti i membri del gruppo CMS di Torino. Marco C. mi conosce bene ormai: grazie per avermi voluto approfondire come persona, e per avere sempre insistito su *come* si devono fare le cose, oltre e prima che su *cosa* fare. Grazie a Chiara e a Nicola A. per avermi insegnato gran parte di quello che so con una passione fuori dal comune. Grazie a Nadia per essersi sempre interessata a me e a quello che faccio. Ringrazio Ernesto, Alessandra, Lino, Stefano A., Niccolò, Michele, Silvia, Amedeo, Riccardo e Sara, perché negli anni mi hanno fatto capire come funzioni un gruppo grande e unito.

Ringrazio Nicola D.F. per avermi seguito con cura sempre da quando faccio parte di CMS, ma soprattutto in questi ultimi mesi. Non dimenticherò quest'attenzione particolare.

Un grazie speciale va ai miei compagni di ufficio, di avventure e di casa di questi anni. Matteo, Marco Mu., Marco Ma.: promettetemi che organizzeremo ancora qualche *briscolino* ogni tanto, anche se le nostre strade dovessero dividersi, geograficamente e non solo. Grazie per la bella amicizia che ci lega, nata tra discussioni fisiche in pausa caffè e racconti poco fisici in pranzi e cene vari. Un ringraziamento speciale a Teddy per avermi ripetuto in piú occasioni che tra me e il successo non c'è nessun ostacolo! Grazie a Cris, a Dani, a Rob, a Giorgia, a Susy, a Vale, a Gogo, a Serena, per tutto quello che avete condiviso con me in questi anni. Ringrazio anche Mario, Alessio, Pietro, Maja, Tiina, per aver reso *casa Avouzon* la dimora di gran lunga più divertente ed accogliente del circondario (nel raggio di diversi kilometri!).

Grazie anche alle giovani *'new entries'* che vengono a dare man forte al gruppo CMS ora che i vecchi come me si dirigono verso nuovi lidi. Martina, Alberto, Giulia, Alessandro, Livia, Daniele, Federico, Daniela, grazie

per avermi fatto passare una delle piú belle estati degli ultimi tempi. Grazie anche a Fabrizio, Filippo, Marco T., Simone, Stefano C. e Luca per i bei momenti, anche se pochi, passati insieme in ufficio. In bocca al lupo, comunque decidiate di proseguire la vostra carriera da fisici!

Grazie di cuore a Roko per essere stato molto piú di un collega e per non avermi mai fatto mancare il suo sostegno e aiuto. Grazie a Marco Me. per i bei momenti in ufficio e fuori, lassú in alto! *Many thanks to Matija and to Somnath for joining me in my everyday's life. I also owe a lot to Ivica, Yves, Christophe, Stéphanie, Roberto, for the pleasure of working in close contact with them and of celebrating together the brilliant outcome of our work. Thanks to Kevin, Suchandra, Laura, Burt, Regina, who have been sharing with me one of the most beautiful trackers ever. I also want to thank all my friends from the Cern Dancing Club, for the good time we have spent together in a joyful and lively environment.*

I miei genitori, mia zia Lena, mia nonna Amalia, mio fratello Paolo e Lorena mi hanno sempre incoraggiato, sia nei momenti sereni sia in quelli difficili. Forse non basterà un *grazie* per tutto quello che hanno fatto.

Non può mancare un grazie sincero a Eleonora, per mille motivi, ma soprattutto perché mi ha fatto riflettere sulla forza di cambiare le cose che posso cambiare, sulla pazienza di accettare le cose che non posso cambiare, e sull'intelligenza di saperle distinguere. Forse cambierò particella, ma se ti interesserà ancora continuerò a raccontarti. Manterrò aggiornati anche Pibì, Barnaba, Giulia, Miriam, Manu, Vipe, Peppi, Chiara e Fil, Mariangela, Totò e Zigghi. *Obrigado* Ugo, *takk* Jacopo e Ilaria per la vostra curiosità nei confronti dell'affascinante mondo della fisica!

Un grazie va anche a Chiara M., per avermi dato la possibilità di distrarmi quando meno avrei dovuto farlo e per avermi fatto riscoprire insieme con lei la bellezza di tante piccole cose. Ringrazio Jasna, Chiara C., Melinda, Stefania e tutti coloro che ho conosciuto nei frequenti viaggi tra Torino e Ginevra.

Σας ευχαριστώ πολύ Silvia: quale sarà la nostra prossima scommessa? Grazie per avermi fatto viaggiare su tappeti volanti in mondi fantastici, armato di una *μπουγάτσα* nella mano destra e di una *ρετσίνα* nella sinistra!

Ringrazio Alessia, Sara e Francesco, Alessandra e Daniele, Emanuele, Giurgia, Luigi e Angela, Roby, Claudia, Reby, Ruzzi, Michela, Federica ed Elena, per aver mantenuto vive le belle amicizie che ci legano malgrado la mia vita nomade di persona non facile da raggiungere. Un altro grazie va al gruppo dei miei ex compagni di fisica (piú alcuni fisici adottivi), per le belle risate che ci scambiamo sul *Resto del Neutrino*: fisica in senso lato!

Spero di mantenermi in contatto con il maggior numero possibile di queste persone anche negli anni futuri. Resterò nell'esperimento CMS, ma proseguirò su argomenti diversi da quelli trattati in questa tesi e, dal punto di vista geografico, mi dividerò tra Meyrin e Santander. Sono sicuro che il futuro continuerà a riservarmi belle sorprese.

Appendix A

Statistical tools

A.1 The CL_s statistical method

The statistical technique used in the ATLAS and CMS Higgs groups is the modified frequentist one, also known as CL_s method [62, 63].

In the following, the expected signal event yield will be referred to as $s(\theta)$, the background one as $b(\theta)$. They are functions of a vector of nuisance parameters θ_i, which account for the considered sources of uncertainty. A *signal strength modifier* μ is defined by $\sigma = \mu \cdot \sigma_{SM}$ and it scales the signal SM cross section of all production channels by the same factor.

The likelihood function $\mathcal{L}(data|\mu, \theta)$ is defined as:

$$\mathcal{L}(data|\mu, \theta) = \text{Poisson}(data|\mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta) \tag{A.1}$$

where the probability density function $p(\tilde{\theta}|\theta)$ is the probability of measuring a set of nuisance parameters $\tilde{\theta}$, given its true value θ. Here *'data'* stands either for real measured data or for pseudo-data from toy MC simulated experiments. The explicit form of the poissonian term in Eq. A.1 for the product of probabilities of observing n_i events in the $i-$th bin is

$$\text{Poisson}(data|\mu \cdot s(\theta) + b(\theta)) = \prod_i \frac{(\mu \cdot s_i(\theta) + b_i(\theta))^{n_i}}{n_i!} \cdot e^{-\mu \cdot s_i(\theta) - b_i(\theta)} \tag{A.2}$$

A.1.1 The case of discovery

The test statistic q_0 used to quantify the statistical significance of an excess of events over the background-only expectation is defined as:

$$q_0 = -2 \ \ln \frac{\mathcal{L}(data|b(\hat{\theta}_{\mu=0}))}{\mathcal{L}(data|\hat{\mu} \cdot s(\hat{\theta}) + b(\hat{\theta}))} \qquad (\hat{\mu} \geq 0) \tag{A.3}$$

where $\hat{\mu}$, $\hat{\theta}$ are the μ, θ values that maximise the likelihoods in the numerator and in the denominator and $\hat{\theta}_{\mu=0}$ is the θ value that maximises the likelihood in the numerator in the background-only hypothesis ($\mu = 0$).

The values that q_0 can take on can be either null (if $\hat{\mu} = 0$, *i.e.* if no signal excess is measured) or positive. Given the outcome q_0^{obs} of some measurement with data, let $Z = \sqrt{q_0^{obs}}$. The probability of getting a q_0 value as large as q_0^{obs} or larger in the background-only hypothesis is called $p-$value:

$$p_0 = P(q_0 > q_0^{obs}|b) = \int_Z^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx \tag{A.4}$$

This *local* $p-$value is used to search for a signal excess in the background-only hypothesis across the whole explored m_H range. In order to account for the larger probability of finding any excess over a broad range than over a narrow one, the *look-elsewhere effect* [64, 65] must be properly evaluated.

A.1.2 The case of exclusion

As the $\mu = 0$ subscript indicates, q_0 is used to look for a signal under the hypothesis of pure background. On the contrary, if one wants to exclude the presence of a given signal one can assume that the signal strength is μ and define q_μ as:

$$q_\mu = -2\ \ln \frac{\mathcal{L}(data|\hat{\mu}\cdot s(\hat{\theta_\mu}) + b(\hat{\theta}_\mu))}{\mathcal{L}(data|\hat{\mu}\cdot s(\hat{\theta}) + b(\hat{\theta}))} \qquad (0 \leq \hat{\mu} \leq \mu) \tag{A.5}$$

Similarly to Eq. A.4, two $p-$values can be defined under the hypothesis of signal and background ($s + b$) and of background only (b):

$$\begin{aligned} CL_{s+b} &= P(q_\mu > q_\mu^{obs}|s+b) \\ CL_b &= P(q_\mu > q_\mu^{obs}|b) \end{aligned} \tag{A.6}$$

A quantity called CL_s is then defined as:

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{A.7}$$

An upper limit on the signal strength μ can then be set by finding the μ value that corresponds *e.g.* to $CL_s = 0.05$ for a 95% confidence level.

Bibliography

[1] The CMS-ECAL group, P. Adzic et al., *Results of the first performance tests of the CMS electromagnetic calorimeter*, Eur. Phys. J. C44 (2006) S1.1

[2] The CMS Collaboration, CMS, *the Compact Muon Solenoid: Technical proposal*, CERN-LHCC-94-38.

[3] G. Acquistapace et al., CMS, *The Magnet Project: Technical Design Report*, CERN-LHCC-97-10.

[4] C. Eck et al., *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical Design Report LCG. CERN, Geneva, 2005.

[5] The CMS Collaboration, *The TriDAS Project Technical Design Report, Vol. 1.*, CERN/LHCC 2000-038, 2000.

[6] The CMS Collaboration, *The TriDAS Project Technical Design Report, Vol. 2.*, CERN/LHCC 2002-026, 2002.

[7] The CMS Collaboration, *Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger*, CERN LHCC 2002-36, 2002

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

[9] E. Fermi, *Trends to a theory of Beta radiation*, *Nuovo Cimento* 11, (1952) 1.

[10] S.L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* 22, (1961) 579.

[11] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* 19, (1967) 1264.

[12] A. Salam, *Elementary Particle Theory*, Almquist & Wisksells, Stockholm (1968) p. 367.

[13] S. Dawson, *Introduction to Electroweak Symmetry Breaking*, hep-ph/9901280.

[14] G. Ridolfi, *Search for the Higgs boson: theoretical perspective*, hep-ph/0106300.

[15] The LEP Collaborations, the LEP Electroweak Working Group and the SLD Heavy Flavour Group, *A combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, LEP-EWWG 2003-01.

[16] The LEP Working Group for Higgs Boson Searches, *Search for the Standard Model Higgs boson at LEP*, CERN/EP 2003-11.

[17] `http://lepewwg.web.cern.ch/LEPEWWG/`

[18] The CMS Collaboration, *Search for Higgs Boson in VH Production with H to $b\bar{b}$*, CMS PAS HIG-11-031 (2011)

[19] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a $b-$quark pair using up to* 4.7 fb^{-1} *of pp collision data at* $\sqrt{s} = 7$ TeV *with the ATLAS detector at the LHC*, ATLAS-CONF-2012-015

[20] The Tevatron New-Phenomena and Higgs Working Group, for the CDF and D0 Collaborations, *Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to* 10 fb^{-1} *of Data*, FERMILAB-CONF-12-065-E, CDF Note 10806, D0 Note 6303

[21] A. Graziano, *Muon isolation studies at CMS in the $H \to ZZ \to 4\mu$ discovery channel*, Master Thesis, University of Turin, 2003

[22] `https://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=slides&confId=55582`

[23] E. Chabanat, N. Estre, *Deterministic Annealing for Vertex Finding at CMS*, prepared for *Computing in High-Energy Physics (CHEP $\tilde{O}$04)*, Interlaken, Switzerland, 27 Sep - 1 Oct 2004, `http://cdsweb.cern.ch/record/865587`

[24] M. Cacciari, G. P. Salam and G. Soyez, *The catchment area of jets*, JHEP 0804 (2008) 005 [arXiv:0802.1188]

[25] The CMS Collaboration, *Determination of jet energy calibration and transverse momentum resolution in CMS*, CMS PAS JME-10-011 (2010), 2011 JINST 6 P11002

[26] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet user manual* (2011) [arXiv:1111.6097]

[27] Bornheim, A. et al., *CMS tracking performance results from early LHC operation*, European Physical Journal C, 70 (4). pp. 1165-1192. ISSN 1434-6044 (2010)

[28] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. B 406 (1993) 187 and refs. therein; S. D. Ellis and D. E. Soper, Phys. Rev. D 48 (1993) 3160 [hep-ph/9305266].

[29] The CMS Collaboration, *Search for a Standard Model Higgs boson produced in the decay channel* $H \to ZZ^{(*)} \to 4\ell$, CMS PAS HIG-11-025 (2011)

[30] The CMS Collaboration, *Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV*, CMS PAS PFT-10-002 (2010).

[31] `https://indico.cern.ch/getFile.py/access?contribId=13&sessionId=3&resId=0&materialId=slides&confId=175489`

[32] The CMS Collaboration, *Performance of muon identification in pp collisions at* $\sqrt{s} = 7$ *TeV*, CMS PAS MUO-10-002 (2010)

[33] `https://twiki.cern.ch/twiki/bin/viewauth/CMS/VbtfZMuMuBaselineSelection`

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

[35] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, CERN-2011-002 (CERN, Geneva, 2011), arXiv:1101.0593

[36] A. Bredenstein, A. Denner, S. Dittmaier et al., *Precise predictions for the Higgs-boson decay* $H \to WW/ZZ \to 4$ *leptons*, Phys. Rev. D74 (2006) 013004, arXiv:hep-ph/0604011, doi:10.1103/PhysRevD.74.013004

[37] A. Bredenstein, A. Denner, S. Dittmaier et al., *Radiative corrections to the semileptonic and hadronic Higgs-boson decays* $H \to WW/ZZ \to 4$ *fermions*, JHEP 0702 (2007) 080, arXiv:hep-ph/0611234

[38] A. Djouadi, J. Kalinowski, M. Muhlleitner et al., *An update of the program HDECAY*, in *The Les Houches 2009 workshop on TeV colliders: The tools and Monte Carlo working group summary report.* 2010. arXiv:1003.1643

[39] S. Actis, G. Passarino, C. Sturm et al., *NNLO Computational Techniques: the Cases $H \to \gamma\gamma$ and $H \to gg$*, Nucl. Phys. B811 (2009) 182Ð273, arXiv:0809.3667, doi:10.1016/j.nuclphysb.2008.11.024

[40] T. Binoth, N. Kauer, P. Mertsch, *Gluon-induced QCD corrections to $pp \to ZZ \to \ell\bar{\ell}\ell'\bar{\ell}'$* , arXiv:0807.0024, doi:10.3360/dis.2008.142

[41] S. Baffioni et al., *Search strategy for the Higgs boson in the $ZZ^{(*)}$ decay channel at $\sqrt{s} = 10$ TeV with the CMS experiment*, CMS AN 2010/237 (2010)

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

[43] D. Marlow, talk at the CMS General Weekly Meeting, `https://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=slides&confId=123724` (2011)

[44] https://twiki.cern.ch/twiki/bin/view/CMS/PhysicsPrimaryDatasets

[45] S. Baffioni et al., *Electron reconstruction in CMS*, CMS AN 2009/164 (2009).

[46] The CMS Collaboration, *Electron reconstruction and identification at $\sqrt{s} = 7$ TeV*, CMS PAS EGM-10-004 (2010).

[47] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCategoryBasedElectronID`

[48] The CMS Collaboration, *Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV*, CMS PAS MUO-10-002 (2010)

[49] The CMS Collaboration, *Measurement of the WW, WZ and ZZ cross sections at CMS*, CMS PAS EWK-11-010 (2011)

[50] S. Baffioni et al., *Search strategy for the Higgs boson in the ZZ decay channel with the CMS experiment*, CMS AN 2008/050 (2008)

[51] C.W. John Campbell, Keith Ellis, *MCFM - Monte Carlo for FeMtobarn processes*, http://mcfm.fnal.gov/

[52] T. Binoth, N. Kauer, P. Mertsch, *Gluon-induced QCD corrections to $pp \to ZZ \to \ell\bar{\ell}\ell'\bar{\ell}'$*, arXiv:0807.0024

[53] R. Gavin, Y. Li, F. Petriello, S. Quackenbush, *FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order*, arXiv:1011.3540

[54] N. Amapane et al., *Search for a Standard Model Higgs boson in the decay channel $H \to ZZ^* \to 4l$*, CMS AN 2011/387 (2011).

[55] The CMS Collaboration, *Combined Standard Model Higgs boson searches with up to* 2.3 *inverse femtobarns of pp collision data at* $\sqrt{s} = 7$ TeV *at the LHC*, CMS PAS HIG-11-023 (2011)

[56] A. Denner, S. Heinemeyer, I. Puljak et al., *Standard Model Higgs-Boson Branching Ratios with Uncertainties*, Eur. Phys. J. C71 (2011) 1753, arXiv:1107.5909

[57] The LHC Higgs Cross Section Working Group, `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HeavyHiggs`

[58] P. Avery et al., *A complete model of the four-lepton invariant mass distributions for* $H \to ZZ \to 4\ell$ *and* $ZZ \to 4\ell$ *events*, CMS AN 2011/202 (2011)

[59] The CMS Collaboration, *Absolute Calibration of the Luminosity Measurement at CMS*, CMS Detector Performance Summary CMS-DP-2011-002 (2011)

[60] The CMS Collaboration, *Search for the fermiophobic model Higgs boson decaying into two photons*, CMS PAS HIG-12-002 (2012)

[61] `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/Fermiophobic`

[62] A. L. Read, *Modified frequentist analysis of search results (the* CL_s *method)*, CERN Yellow Report CERN-2000-005 (2000) 81

[63] The CMS Collaboration, *Combination of CMS searches for a Standard Model Higgs boson*, CMS PAS HIG-11-032 (2011)

[64] L. Lyons, *Open statistical issues in particle physics*, The Annals of Applied Statistics 2008, Vol. 2, No. 3, 887-915

[65] L. Demortier, *P Values and Nuisance Parameters*, Proceedings of PHYSTAT-LHC Workshop, 2007, CERN-2008-001

[66] E. Gross, O. Vitells, *Trial factors for the look elsewhere effect in high energy physics*, [arXiv:1005.1891], Eur. Phys. J. C 70:525-530, 2010

[67] `http://press.web.cern.ch/press/PressReleases/Releases2012/PR01.12E.html`

[68] W.J. Stirling, private communication. `http://www.hep.phy.cam.ac.uk/~wjs/plots/plots.html`