UNIVERSITÀ
DEGLI STUDI
FIRENZE

Dottorato di Ricerca in Fisica e Astronomia
Ciclo XXVII

Search for a high mass Higgs-like boson produced via Vector Boson Fusion, in the decay channel $H \to ZZ \to l^+l^-q\bar{q}$ $(l = e, \mu)$ with the CMS detector

Settore Scientifico Disciplinare FIS/04

Dottoranda
Dott.ssa Valentina Gori

Tutore
Prof. Vitaliano Ciulli

Coordinatore
Prof. Roberto Livi

Anni 2012/2014

"Un anno al CERN equivale ad un viaggio interstellare" (cit.)

A tutti i viaggi interstellari,
e a tutte le persone che mi hanno insegnato qualcosa.

Contents

Introduction

The Higgs mechanism is an essential element of the Standard Model (SM) of particles and their interactions, explaining the origin of the mass of elementary particles and playing a key role in the physics of electroweak symmetry breaking. A suitable Higgs boson candidate, compatible with the one predicted by the Higgs mechanism, has recently been found with a mass around 125 GeV[1]. However, many models predict more than a single boson, so further searches for Higgs-like particles using the SM Higgs as a benchmark model are being pursued.

The Higgs boson can be produced via four production mechanisms at the LHC: gluon-fusion (gg, the main mode at the LHC, in the whole mass range between 100 and 1000 GeV), Vector Boson Fusion (VBF, where the Higgs is produced via the fusion of two vector bosons which are emitted by two quarks) and associate VH (with a vector boson) and $t\bar{t}H$ (with a top-antitop couple) production. All these production modes, together with the resonance properties, are currently under investigation.

The VBF topology is characterized by the presence of two additional jets in the final state, with high η separation and high invariant mass. In the Standard Model this production mechanism has a lower cross section than the gg one (about 10 times smaller) and thus its observation is more challenging. The $H \to ZZ \to 2l2q$ channel is one of the best suited to observe such a topology, because of the large branching ratio of the decay $Z \to 2q$. The main background affecting this decay channel is given by Drell-Yan plus jets events; the cross section of this process decreases by a factor of 100 when requesting the presence of four jets instead of two. Thus, the VBF topology benefits from a highly reduced background with respect to the gg production mode, such that even if the VBF production mechanism has a cross section smaller than the gluon-fusion, a ten times higher signal-to-noise ratio is expected. Furthermore, in beyond-SM models, VBF production could be enhanced with respect to gg.

If the considered Higgs mass is more than twice the Z mass, both Z bosons are on shell, and the couple of jets from the hadronic Z can be easily identified among the jets in the event by requiring their invariant mass to be close to the Z nominal mass. Moreover, the VBF jets can be easily identified thanks to their peculiar topology.

In 2013 results have been published [1] about the search for a high mass Higgs-like boson in this decay channel, for the gluon fusion production mechanism, in the mass range between 230 and 600 GeV. The VBF category was not considered, not even the possibility to reconstruct the couple of jets in the final state as a single merged jet in case of boosted topology. The analysis, investigating a mass range between 230 and 600 GeV, showed no

[1]Natural units, where $c = 1$, have been chosen.

evidence for such a SM-like Higgs boson and had the sensitivity to set upper limits on its production cross section and to exclude this resonance at 95% confidence level in the mass range between 275 and 600 GeV.
The analysis which is presented here has been performed to study this decay channel in the VBF production mode. Adding the VBF contribution to the main gluon-fusion analysis helps improving the overall sensitivity in the search for a high mass SM-like Higgs boson, especially at very high masses, allowing to expand the mass range where an exclusion is feasible. In fact, the new analysis, accounting for the VBF production also with the merged-jet topology, allows to set upper limits on the production cross section of a Higgs-like particle in the mass range between 230 and 1000 GeV and has enough sensitivity to exclude a SM-like Higgs boson at 95% confidence level in the mass range between 305 and 744 GeV.
In addition, the results have been re-interpreted in terms of an electroweak singlet scalar mixing with the 125 GeV Higgs boson, and exclusion limits have been set.

Chapter 1

Physics at the LHC

1.1 The Standard Model of particle physics: the electroweak Lagrangian

The Standard Model is the physical theory which is currently adopted to provide a quantitative description of three of the four interactions in nature: electromagnetism, weak interactions and the strong nuclear force. It has been elaborated at the end of the 1960's by Glashow, Weinberg and Salam [2–4]. It is a renormalizable quantum field theory, compatible with special relativity. Its Lagrangian presents a non-Abelian gauge symmetry $SU(3) \otimes SU(2) \otimes U(1)$. During the past decades its predictions have been confirmed by a large number of experiments with outstanding precision. The Standard Model (SM) can be divided into two sectors: the strong sector, known as Quantum Chromodynamics (QCD), and the electroweak sector. Then, a Higgs Lagrangian needs to be added, in order to give mass to fermions and bosons. Hence, the SM Lagrangian may be written as the sum of three parts:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{H} \tag{1.1}$$

Quantum Chromodynamics describes the interactions of quarks and gluons, mediated by the strong force through the colour charge. Its Lagrangian satisfies the $SU(3)_C$ colour symmetry, and will be described later. The Lagrangian which governs the electroweak sector is instead invariant under gauge transformations of the symmetry group $SU(2)_L \otimes U(1)_Y$. The $SU(2)_L$ group refers to the weak isospin charge (I), and $U(1)_Y$ to the weak hypercharge (Y). Left-handed (L) fermions are paired in $I = 1/2$ isospin doublets, whereas right-handed (R) fermions in $I = 0$ singlets. The presence of these local gauge symmetries introduces four vector bosons: three for the $SU(2)$ group, the W^i fields ($i = 1,\ 2,\ 3$), and one for $U(1)$, the B field.
This gives rise to a quantum field theory, invariant under local gauge symmetries, whose Lagrangian is expressed as:

$$\mathcal{L}_{EW} = \sum_f \bar{\psi} i\gamma^\mu \mathcal{D}_\mu \psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - \frac{1}{4} \vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu} \ , \tag{1.2}$$

where the sum is extended over all the fermions f and where covariant derivatives which preserve the local $SU(2)_L \otimes U(1)_Y$ gauge invariance have the following form ($\vec{\tau}$ indicates

the three Pauli matrices and g, g' are the coupling constants which correspond respectively to $SU(2)_L$ and $U(1)_Y$):

$$\mathcal{D}_\mu = \partial_\mu + ig\vec{W}_\mu \cdot \frac{\vec{\tau}}{2} + i\frac{g'}{2} Y B_\mu \tag{1.3}$$

and

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ E^\alpha_{\mu\nu} &= \partial_\mu W^\alpha_\nu - \partial_\nu W^\alpha_\mu - g\epsilon^{\alpha\beta\gamma} W^\beta_\mu W^\gamma_\nu \ . \end{aligned} \tag{1.4}$$

The physical fields are obtained as linear combinations of these fields:

$$\begin{aligned} A_\mu &= \sin\theta_W W^3_\mu + \cos\theta_W B_\mu \\ Z_\mu &= \cos\theta_W W^3_\mu - \sin\theta_W B_\mu \\ W^\pm_\mu &= \frac{W^1_\mu \mp W^2_\mu}{\sqrt{2}} \end{aligned} \tag{1.5}$$

The above equations represent two neutral particles (the photon, described by the A_μ field, and the Z boson) and two charged particles (the W^+ and W^- bosons). We have further introduced the angle θ_W, which is known as the weak mixing angle or Weinberg angle. Up to here, the theory is necessarily incomplete: all particles it describes are massless, contradicting experimental evidence. The Lagrangian symmetries, on the other hand, seem to forbid the introduction of mass terms without spoiling its gauge invariance. Higgs' proposal solves this problem by spontaneously breaking the Lagrangian symmetry.

1.2 The Higgs mechanism

A Lagrangian is symmetrical if it is invariant under a group of transformations. If a symmetrical Lagrangian presents a degenerate ground state, there is no univocal state which describes the system's fundamental configuration: one of the degenerate states must be chosen, but any of the chosen states will not share the Lagrangian symmetry. This procedure of obtaining an asymmetric ground state is known as *spontaneous symmetry breaking*. The simplest way of spontaneously breaking the $SU(2) \otimes U(1)$ symmetry group is that of introducing a scalar field Φ (the *Higgs field*) which is an isospin ($SU(2)_L$) doublet:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \begin{pmatrix} (\Phi_1 + i\Phi_2)/\sqrt{2} \\ (\Phi_3 + i\Phi_4)/\sqrt{2} \end{pmatrix} \tag{1.6}$$

where we have introduced four real fields Φ_a ($a = 1, 2, 3, 4$) to manifest the complexity of the Φ^+ and Φ^0 fields. The simplest Lagrangian of a self-interacting scalar field has the form:

$$\mathcal{L}_H = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi) \tag{1.7}$$

where

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{1.8}$$

The potential $V(\Phi)$ depends on two parameters, μ and λ. The requirement $\lambda > 0$ ensures that the energy spectrum has a lower bound, and therefore the existence of a ground

state. If the μ parameter is chosen so that $\mu^2 < 0$, the symmetry of $V(\Phi)$ may be broken, as it has a minimal value in correspondence of:

$$\Phi^\dagger\Phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.9}$$

Perturbation theory requires an expansion of Φ around its ground state, where the latter must be chosen among the set of states which satisfy Eq. 1.9, all of which break the rotational symmetry of the Lagrangian. When breaking the electroweak symmetry group $SU(2)_L \otimes U(1)_Y$, some care is needed in order to avoid that the photon is assigned a spurious mass. Therefore, we need to choose a ground state Φ_0 which conserves the electric charge symmetry group $U(1)$, i.e. with null electric charge Q. From the Gell-Mann–Nishijima relation [5, 6]

$$Q = I_3 + \frac{Y}{2} \tag{1.10}$$

which connects the electric charge to weak isospin and hypercharge, we see that this condition is satisfied if we choose Φ_0 with weak isospin $I = 1/2$, $I_3 = -1/2$ and hypercharge $Y = 1$:

$$\Phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v\end{pmatrix} \tag{1.11}$$

Therefore the Φ field will be expressed, in the unitary gauge, as:

$$\Phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\v+h(x)\end{pmatrix} \tag{1.12}$$

In this way the bosonic fields $W^\pm$ and Z, connected to the broken symmetry group $SU(2)$, will acquire a mass

$$m_W = \frac{v}{2}g \qquad m_Z = \frac{v}{2}\sqrt{g^2+g'^2} \tag{1.13}$$

We have furthermore introduced a physical particle, the Higgs boson, described by the h(x) field, with mass equal to

$$m_H = \sqrt{2}\mu = v\sqrt{2\lambda} \tag{1.14}$$

Mass terms for the fermions also arise through the interaction with the Higgs field, without breaking the $SU(2)_L \otimes U(1)_Y$ symmetry. Introducing an interaction term of the fermions with the Higgs field, in the context of spontaneous symmetry breaking, makes it possible to assign masses to the fermions. This can be done via a Yukawa coupling. For the down-type quarks, for example, the lagrangian can be written

$$\mathcal{L}_{\text{mass}}^{\text{quark}} = -G_q^{ij}\Big(\bar{Q}_L^i\ \Phi\ d_R^j + \bar{d}_R^j\ \Phi^\dagger\ Q_L^i\Big), \tag{1.15}$$

where $Q_L = \binom{u}{d}_L$, the indices i, j refer to the generation and G_q are the matrices which contain the Yukawa couplings for quarks.
For leptons there is a perfect analogy with what stated for quarks, except for the fact that no mixing is observed among flavours because of the negligible mass of neutrinos.
For example for the electrons:

$$\mathcal{L}_{\text{mass}}^{\text{electron}} = -g_e\ (\bar{\chi}_L\ \Phi\ e_R + \bar{e}_R\ \Phi^\dagger\ \chi_L)\ , \tag{1.16}$$

which can be divided into two terms by using Eq. 1.12 for the Higgs field and $\chi_L = \binom{\nu}{e^-}_L$:

$$\mathcal{L}_{\text{mass}}^{\text{electron}} = -g_e \frac{v}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) - \frac{g_e}{\sqrt{2}}(\bar{e}_L e_R + \bar{e_R} e_L) h(x) \ , \tag{1.17}$$

g_e being the Yukawa coupling constant for the electron. The mass term for the electron is therefore:

$$m_e = g_e \frac{v}{\sqrt{2}} \ . \tag{1.18}$$

It is worth noting that the coupling constant between the Higgs boson and a fermion is proportional to the fermion mass.

1.2.1 Theoretical constraints on the SM Higgs boson mass before its observation

As seen in Eq. 1.14, the Higgs boson mass depends not only on the vacuum expectation value v but also on the Higgs fields self-coupling constant λ. The value of v is determined by the Fermi constant G_F by the following relation:

$$v = \frac{2m_W}{g} = (\sqrt{2}G_F)^{-1/2} \tag{1.19}$$

The current estimate of G_F, which comes from precise muon lifetime measurements [7,8], allows to determine $v \simeq 246$ GeV.
On the other hand, the model is not predictive on the value of the λ parameter, therefore the mass of the Higgs particle is a free parameter of the theory.
Nevertheless, we can exploit the perturbative nature of the theory to impose approximate theoretical boundaries on m_H. Besides the upper bound on the Higgs boson mass from unitarity constraints [9], additional theoretical arguments place approximate upper and lower bounds on m_H [10]. A first limit is obtained by requiring that the breaking of symmetry actually takes place:

$$V(v) < V(0) \tag{1.20}$$

This argument is known as the *vacuum state stability*, and leads to requiring λ to be positive at all energies[1]. For small values of λ, the top quark contribution could determine a negative value of $\lambda(Q^2)$; therefore, m_H is constrained to be larger than a certain value which is proportional to squared top mass and depends on the cut-off scale value Λ_C, above which we assume that the Standard Model is not valid. This gives a lower bound on the Higgs mass.
Viceversa, if m_H is too large, then the Higgs self-coupling λ diverges at some scale Λ below the Planck scale. That would mean that there is new physics before the Planck scale.
Given a value of Λ_C, one can compute the minimum and maximum allowed Higgs boson masses. Conversely, the value of m_H itself can provide an important constraint on the scale up to which the SM remains successful as an effective theory. In particular, a Higgs boson with mass in the range 130 GeV $\lesssim m_H \lesssim$ 180 GeV would be consistent with

[1] Like the other coupling constants of renormalizable theories, also λ runs with energy.

an effective SM description that survives all the way to the Planck scale. For smaller Higgs mass values, the stability of our universe prefers new physics at lower scale. The lower bound on m_H can be reduced to about 115 GeV if one allows for the electroweak vacuum to be metastable, with a lifetime greater than the age of the universe. The main uncertainties in the stability and perturbativity bounds come from the uncertainties in the value of α_s and the top quark mass.

1.2.2 Experimental constraints on the SM Higgs boson mass before its observation

Before the observation of a scalar boson having a mass around 125 GeV by the ATLAS and CMS collaborations, which led to the claim of the Standard Model Higgs boson discovery (see Sec. 1.7), bounds on the Higgs mass were provided by both direct searches of a Higgs bosons and by indirect precision measurements. Precision electroweak measurements are sensitive to radiative corrections which depend logarithmically on m_H. This allows to estimate a confidence interval for m_H from a combined fit of the electroweak parameters where the SM is assumed to be correct and the Higgs boson mass is left as a free parameter. For example, radiative corrections to the $W^{\pm}$ mass depend only on m_H (after the measurement of the top quark mass at Tevatron). Such indirect measurements were performed by the four LEP collaborations, ALEPH, DELPHI, L3 and OPAL and by the SLD experiment at the Stanford Linear Collider [11].
At LEP, the direct search for the Higgs bosons at centre of mass energies up to 209 GeV has been focused on the analysis of the Higgs-stralhung production process, where the electron-positron pair annihilates into a virtual Z which then splits into a Higgs particle and a real Z. The combination of results from all four collaborations at LEP provides a lower limit at 95% of confidence level to the Higgs mass [12]: $m_H > 114$ GeV.
Results on direct searches for a SM Higgs Boson in $p\bar{p}$ collisions at the Fermilab Tevatron accelerator, with 10.0 fb^{-1} of data collected at $\sqrt{s} = 1.96$ TeV, presented on July 2012, exclude at 95% C.L. the presence of a SM-like Higgs boson in the mass range m_H between 100 and 103 GeV and between 147 and 180 GeV [13]. CDF and D$\emptyset$ experiments performed a search for the Higgs boson in the following channels: H$\rightarrow b\bar{b}$, H$\rightarrow$ W$^+$W$^-$, H$\rightarrow$ ZZ, H$\rightarrow \tau^-\tau^+$, H$\rightarrow \gamma\gamma$.

1.3 Beyond the Standard Model: the Electroweak Singlet

The recent discovery of a new particle which is in accordance with the scalar boson from the Higgs mechanism by the LHC experiments (see Sec. 1.7) is one of the biggest breakthroughs in contemporary particle physics. If the discovered particle is indeed the Higgs Boson predicted from a SM-like Higgs-doublet sector, all its properties are completely determined by theory. Therefore, the current quest of the theoretical and experimental community is to establish whether the properties of such particle are in accordance with standard predictions, or it is only a component of a more involved Higgs sector. For this reason, all couplings as well as the spin structure of the new particle need to be severely tested.

We are now considering the simplest extension of the SM Higgs sector [14–16], adding an additional singlet which is neutral under all quantum number of the SM gauge groups.

The idea is to enlarge the SM Higgs sector with a further real Higgs singlet χ, which is invariant under each gauge group of the SM.

The most general gauge-invariant and renormalisable scalar Lagrangian is then:

$$\mathcal{L} = (\mathcal{D}^\mu H)^\dagger (\mathcal{D}_\mu H) + (\mathcal{D}^\mu \chi)^\dagger (\mathcal{D}_\mu \chi) - V(H,\chi) \ , \tag{1.21}$$

where $\mathcal{D}_\mu$ is defined in Eq. 1.3 when applied to H, while it represents simply the derivative ∂_μ when applied to the singlet χ, and with the scalar potential given by

$$\begin{aligned} V(H,\chi) &= -m^2 H^\dagger H - \mu^2 |\chi|^2 - (H^\dagger H \ |\chi|^2) \begin{pmatrix} \lambda_1 & \frac{\lambda_3}{2} \\ \frac{\lambda_3}{2} & \lambda_2 \end{pmatrix} \begin{pmatrix} H^\dagger H \\ |\chi|^2 \end{pmatrix} \\ &= -m^2 H^\dagger H - \mu^2 |\chi|^2 + \lambda_1 (H^\dagger H)^2 + \lambda_2 \ |\chi|^4 + \lambda_3 H^\dagger H \ |\chi|^2 \end{aligned} \tag{1.22}$$

We here implicitely impose a Z_2 symmetry ($\chi \to -\chi$) which forbids additional terms in the potential. To determine the condition for $V(H,\chi)$ to be bounded from below, it is sufficient to study its behaviour for large field values, controlled by the matrix in the first line of Eq. 1.22. Since the minimisation procedure is not affected by the choice of the gauge, it is not restrictive to define the two vacuum expectation values (VEVs) in the following way:

$$\langle H \rangle \equiv \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad \langle \chi \rangle \equiv \frac{x}{\sqrt{2}} \ , \tag{1.23}$$

with v and x real and non-negative.

Then, the search for extrema of V is made by means of the following differential set of equations:

$$\begin{cases} \frac{\partial V}{\partial v}(v,x) = v \ \left(-m^2 + \lambda_1 v^2 + \frac{\lambda_3}{2} x^2\right) = 0 \\ \frac{\partial V}{\partial x}(v,x) = x \ \left(-\mu^2 + \lambda_2 x^2 + \frac{\lambda_3}{2} v^2\right) = 0 \end{cases} \tag{1.24}$$

The physically interesting solutions are the ones obtained for $v,\, x > 0$:

$$\begin{aligned} v^2 &= \frac{\lambda_2 m^2 - \frac{\lambda_3}{2}\mu^2}{\lambda_1\lambda_2 - \frac{\lambda_3^2}{4}} \\ x^2 &= \frac{\lambda_1 \mu^2 - \frac{\lambda_3}{2} m^2}{\lambda_1\lambda_2 - \frac{\lambda_3^2}{4}} \end{aligned} \tag{1.25}$$

In order to identify the extrema, we need to evaluate the Hessian matrix:

$$\mathcal{H}(v,x) = \begin{pmatrix} \frac{\partial^2 V}{\partial v^2} & \frac{\partial^2 V}{\partial v \partial x} \\ \frac{\partial^2 V}{\partial v \partial x} & \frac{\partial^2 V}{\partial x^2} \end{pmatrix} = \begin{pmatrix} 2\lambda_1 v^2 & \lambda_3 v x \\ \lambda_3 v x & 2\lambda_2 x^2 \end{pmatrix} \tag{1.26}$$

It is straightforward to verify that the solutions in Eq. 1.25 are minima if and only if

$$\begin{aligned} 4\lambda_1\lambda_2 - \lambda_3^2 > 0 \\ \lambda_1,\ \lambda_2 > 0 \end{aligned} \tag{1.27}$$

where the condition given by the second line in Eq. 1.27 corresponds to the requirement that the potential is bounded from below for large field values, while the first line comes directly from imposing that the Hessian determinant is positive, in other words it guarantees that the extremum is indeed a local minimum. Since the physical mass eigenvalues are gauge invariant, we define the Higgs field following the unitary-gauge prescription:

$$H \equiv \begin{pmatrix} 0 \\ \frac{\tilde{h}+v}{\sqrt{2}} \end{pmatrix}, \qquad \chi \equiv \frac{h' + x}{\sqrt{2}} \tag{1.28}$$

The explicit expressions for the scalar mass eigenvalues are:

$$\begin{aligned} m_h^2 =& \lambda_1 v^2 + \lambda_2 x^2 - \sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + (\lambda_3 x v)^2}\ , \\ m_H^2 =& \lambda_1 v^2 + \lambda_2 x^2 + \sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + (\lambda_3 x v)^2}\ , \end{aligned} \tag{1.29}$$

where h and H are the scalar fields of definite masses m_h and m_H respectively, with $m_h^2 < m_H^2$.
These eigenvalues are related to the following eigenvectors:

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \tilde{h} \\ h' \end{pmatrix}\ , \tag{1.30}$$

where $-\pi/2 \leq \alpha \leq \pi/2$ fulfills[2]:

$$\begin{aligned} \sin 2\alpha &= \frac{\lambda_3 x v}{\sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + (\lambda_3 x v)^2}}\ , \\ \cos 2\alpha &= \frac{\lambda_2 x^2 - \lambda_1 v^2}{\sqrt{(\lambda_1 v^2 - \lambda_2 x^2)^2 + (\lambda_3 x v)^2}}\ . \end{aligned} \tag{1.31}$$

[2]The whole interval $0 \leq \alpha < 2\pi$ is halved because an orthogonal transformation is invariant under $\alpha \to \alpha + \pi$

From Eq. 1.30, it is clear that the light (heavy) Higgs couplings to SM particles are now suppressed by $\cos\alpha$ ($\sin\alpha$).
From equations 1.29 and 1.31 it is possible to express λ_1, λ_2, λ_3 as a function of α, m_h, m_H, v, x.
In summary, the heavy Higgs is a "twin" version of the light Higgs with rescaled couplings to the matter contents and to the gauge fields of the SM. In fact, the only novel channel with respect to the light Higgs case is $H \to hh$. The decay width Γ and coupling strength μ' of the $H \to hh$ decay are:

$$\begin{aligned}\Gamma(H \to hh) &= \frac{|\mu'|^2}{8\pi m_H}\sqrt{1-\frac{4m_h^2}{m_H^2}}\ , \\ \mu' &= -\frac{\lambda_3}{2}(x\cos^3\alpha + v\sin^3\alpha) + (\lambda_3 - 3\lambda_1)v\cos^2\alpha\sin\alpha + (\lambda_3 - 3\lambda_2)x\cos\alpha\sin^2\alpha\ .\end{aligned} \tag{1.32}$$

Our simple singlet extension model has in principle 5 free parameters on the Lagrangian level (as opposed to 2 for the SM):

$$\lambda_1, \lambda_2, \lambda_3, v, x. \tag{1.33}$$

The coupling parameters λ_i are related to the masses and the effective mixing, so an equivalent set of independent parameters is:

$$m_h, m_H, \alpha, v, x. \tag{1.34}$$

Moreover, it is possible to express x by $\tan\beta$ according to

$$\tan\beta = \frac{v}{x} \tag{1.35}$$

to accomodate for standard notation in models with extended Higgs sectors. If we assume the vacuum expectation value of the Higgs doublet to be Standard Model-like such that $v \sim 246$ GeV, and equally set the Higgs mass of the light Higgs to $m_h = 125$ GeV, we are left with three independent parameters m_H, α, and one parameter to be chosen between x and $\tan\beta$.
Theoretical and experimental limits can be put on the singlet extension model:

- limits from perturbative unitarity: tree-level perturbative unitarity [9] puts a constraint on the Higgs masses of our theory via a relation on the partial wave amplitudes of all possible $2 \to 2$ scattering processes;
- limits from electroweak precision measurements;
- perturbativity constraints on the couplings, as well as conditions on a potential which is bounded from below;
- limits from measurements of the light Higgs signal strength;

If we define the signal strength modifier μ as the ratio

$$\mu \equiv \frac{\sigma_{\mathrm{BSM}}(m_h)}{\sigma_{\mathrm{SM}}(m_h)} \tag{1.36}$$

then we have

$$\mu = \frac{\cos^4\alpha\Gamma_{\mathrm{SM}}(m_h)}{\cos^2\alpha\Gamma_{\mathrm{SM}}(m_h) + \sin^2\alpha\Gamma_{\mathrm{hid}}(m_h)} \ , \tag{1.37}$$

where $\Gamma_{\mathrm{SM}}(m_h)$ denotes the decay width of the light Higgs in the SM while $\Gamma_{\mathrm{hid}}(m_h)$ indicates the decay width of the light Higgs in a possible hidden sector. If we here constraint ourselves to cases where $\Gamma_{\mathrm{hid}} = 0$, the above equation simply leads to $\mu = \cos^2\alpha$. Then we can give a simple and rough estimate for the α parameter. Given the ATLAS and CMS measurements for the signal strength μ, a constraint can be put on α, which is then expected to be a small angle.
By the way, being α small, we can see that using the correct expression for μ would not have led to a significantly different estimate.
These limits on the measurement of the 125 GeV Higgs boson coupling strength are much more stringent than EW precision observables.

Translating to collider observables, we have:

- the generic suppression of the production of the heavy Higgs, given by a factor $\sin^2\alpha$;
- a small suppression of the production of the light Higgs, by a factor $\cos^2\alpha$;
- the suppression of the SM decay modes of the heavy Higgs, again by a factor $\sin^2\alpha$;
- we have to take into account the additional mode $H \to hh$, not suppressed.

The total width of the heavy Higgs is then modified to

$$\Gamma' = \Gamma_{\mathrm{TOT}}(m_H) = \sin^2\alpha\ \Gamma_{\mathrm{SM}}(m_H) + \Gamma_{H\to hh} \tag{1.38}$$

which can also be rewritten as

$$\Gamma' = \sin^2\alpha\ \Gamma_{\mathrm{SM}}(m_H) + \Gamma'\ BR_{\mathrm{new}} \tag{1.39}$$

where BR_{new} is the branching ratio of the new decay $H \to hh$. We can rename

$$\begin{cases} C = \cos\alpha \\ C' = \sin\alpha \end{cases} \tag{1.40}$$

so that C and C' are respectively the scale factor of the couplings of the low and high mass boson with respect to the SM. It is thus easy to show that Γ' can be written as:

$$\Gamma' = \Gamma_{\mathrm{SM}}(m_H) \cdot \frac{C'^2}{1 - BR_{\mathrm{new}}} \tag{1.41}$$

where $\Gamma_{\rm SM}(m_H)$ will be indicated simply by $\Gamma_{\rm SM}$ in the following (see Sec. 7.3).
The signal strength modifier μ', defined in complete analogy to Eq. 1.36 but referring to the new resonance of mass m_H:

$$\mu' \equiv \frac{\sigma_{\rm BSM}(m_H)}{\sigma_{\rm SM}(m_H)} \tag{1.42}$$

will be written as:

$$\mu' = \frac{\sin^4\alpha\ \Gamma_{\rm SM}}{\sin^2\alpha\ \Gamma_{\rm SM} + \Gamma(H \to hh)} \tag{1.43}$$

which, according to Eq. 1.40 and 1.41, gives

$$\mu' = C'^2(1 - BR_{\rm new}). \tag{1.44}$$

1.4 Strong interaction

Out of the quark model proposed by Gell-Mann [17] in 1964, the idea of the "colour" quantum number was proposed by Han and Nambu [18] in 1965 to avoid the apparent paradox that the quark model seemed to require a violation of the Pauli exclusion principle to describe hadron spectroscopy. Quantum Chromo Dynamics (QCD) was then quantized as a gauge theory with $SU(3)_c$ symmetry in 1973 by Fritzsch [19], Gross and Wilczec [20] and Weinberg [21].

1.4.1 QCD Lagrangian

QCD is an $SU(3)_c$ gauge theory whose Lagrangian is written in the following form:

$$\mathcal{L}^{QCD} = \sum_f \bar{\psi}_f \big(i\gamma_\mu \mathcal{D}^\mu - m_f\big)\psi_f - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{1.45}$$

where f runs over the six quark fields, $\mathcal{D}^\mu$ is the covariant derivative:

$$\mathcal{D}_\mu = \partial_\mu + igA_{\mu a} T_a \tag{1.46}$$

and

$$F_{\mu\nu} = \partial_\mu A_{\nu a} - \partial_\nu A_{\mu a} - g C_{abc} A_{\mu b} A_{\nu c} \tag{1.47}$$

where $A_{\mu a}$ are the fields of the eight coloured gluons, T_a are the eight generators of $SU(3)$, C_{abc} are the structure constants that define the commutation rules of the $SU(3)$ generators.

1.4.2 Running α_s: asymptotic freedom and confinement

Let's consider a physical observable $\mathcal{O}$, whose value depends on a scale Q, e.g. the momentum transfer in a scattering process. When we calculate $\mathcal{O}$ as a series in the coupling constant α_s we need to introduce a renormalization scale μ, to make integrals calculable. In fact, QCD suffers from divergent loop integrals, and the way to compute these kind of integrals is to replace them by finite ones by limiting the momentum in the loop to a cutoff μ. In general, $\mathcal{O}$ is a function of Q^2/μ^2, and of α_s itself. To require that $\mathcal{O}$ does not depend on our choice of μ is equivalent to the following equation:

$$\mu^2 \frac{d\mathcal{O}(Q^2/\mu^2, \alpha_s(\mu^2))}{d\mu^2} = 0. \tag{1.48}$$

It can be shown [22] that solutions of Eq. 1.48 have form $\mathcal{O}(1, \alpha_s(Q^2))$, which means that the whole scale dependence of the observable $\mathcal{O}$ is embedded in the running value of the coupling constant.

This means that $\mathcal{O}$ can be calculated at a fixed order in perturbation theory and the running of α_s predicts how $\mathcal{O}$ changes with the scale.

The perturbative expansion of α_s is such that

$$\frac{d\alpha_s(Q^2)}{dQ^2} < 0, \tag{1.49}$$

which means that the value of the coupling constant decreases at higher scales. This is one of the most peculiar characteristics of QCD with respect to QED, where exactly the opposite holds. In QED the coupling decreases with decreasing scales and increasing distances. This is intuitively interpreted as a screening of the bare electric charge due to fermion loops, and is anyway quite a little effect. On the contrary in QCD the value of the coupling constant is tiny at high energies (small distances) and becomes very big as the scale decreases.

The fact that the coupling constant is small at high energies is often referred to as *asymptotic freedom*, and it justifies the success of the perturbative approach (perturbative QCD or pQCD) in describing hard processes.

On the other hand the fact that the coupling becomes so strong as the distance grows justifies the *confinement*, that is to say the fact that in our experiments we always observe bound quark states, and never free quarks or gluons; qualitatively, this happens because as the partons produced in the hard interactions go apart from each other, it becomes easier to produce quark-antiquark pairs, that recombine into hadrons, than to keep pulling against a growing force. In this regime, the theory behaves in a non-perturbative way.

The expression of $\alpha_s(Q^2)$ follows [23]:

$$\frac{\alpha_s(Q^2)}{4\pi} = \frac{1}{\beta_1 \ln(Q^2/\Lambda^2_{QCD})} - \frac{\beta_2 \ln(\ln(Q^2/\Lambda^2_{QCD}))}{\beta_1^3 \ln^2(\ln(Q^2/\Lambda^2_{QCD}))} + O\Big(\frac{1}{\ln^3(\ln(Q^2/\Lambda^2_{QCD}))}\Big) \qquad (1.50)$$

with β functions defined in [23]. Λ_{QCD} is often regarded as the parameter at which the interaction becomes strong.

Experiments usually measure α_s at a certain scale, usually the mass of the Z boson; Eq. 1.50 then predicts the value of α_s at any other scale.

1.5 Proton-proton interaction

The fundamental difference between the hadronic and leptonic collisions lies in the fact that the hadrons are not elementary particles such as leptons but have an internal structure, which can be described in terms of the QCD-improved *parton model*. Hadrons, along with the valence quarks that contribute to their quantum numbers (*uud* for protons), contain virtual quark-antiquark ($q\bar{q}$) pairs known as *sea quarks*. Sea quarks form when a gluon of the hadron colour field splits; this process also works in reverse so that the annihilation of two sea quarks produces a gluon. Sea quarks are much less stable than their valence counterparts, and they typically annihilate each other within the hadrons. Despite this, sea quarks can hadronize into baryonic or mesonic particles under certain circumstances. In proton-proton collisions, the interaction generally involves a pair of partons and it is therefore likely that any of these partons (quarks q and q' or gluon g) interact, making a qq, $q\bar{q}$, qq', $q\bar{q}'$, gq, $g\bar{q}$ or gg interaction possible.

1.5.1 Parton Distribution Functions (PDFs)

At an hadronic collider, the cross section for a $2 \to F + X$ process, where F is the sought final state, can be conveniently approximated as

$$d\sigma_{pp\to F} = \int_0^1 dx_1 \int_0^1 dx_2 \; f_1(x_1, \mu_F^2) \; f_2(x_2, \mu_F^2) \; d\hat{\sigma}_{pp\to F}(\mu_F^2) \tag{1.51}$$

In this expression $\hat{\sigma}$ is the parton level cross section, x_1 and x_2 are the momentum fraction of the proton momentum carried by the two colliding partons (also called "Bjorken's scaling variables"); $f_{1,2}$ are the parton distribution functions (PDFs), that describe the probability that a parton of flavour 1,2 carries momentum fraction $x_{1,2}$; μ_F is the factorization scale, that is the scale at which the separation between the hard perturbative interaction and the long distance, non-perturbative, evolution of the produced partons takes place.

PDFs evolution with scale is governed by the DGLAP equation [24], as long as $\alpha_s(Q)$ remains in the perturbative validity region. DGLAP equation allows global fits of a variety of data taken from different experiments, at different scales. Three collaborations are the main provider of global PDFs fits, CTEQ [25], MSTW 2008 [26] and NNPDF [27].

1.5.2 Boost-invariant variables

Since the center-of-mass of the parton-parton scattering is normally boosted in the beam direction with respect to that of the two incoming hadrons, it is therefore useful to classify the final state in terms of variables that are invariant under Lorentz transformations in that direction. For this purpose, the variables rapidity y, transverse momentum p_T, and azimuth angle φ are introduced. In term of these variables, the four-momentum of a particle of mass m may be written as

$$p^\mu = (E, p_x, p_y, p_z) = (m_T \cosh y, p_T \sin\varphi, p_T \cos\varphi, m_T \sinh y) \; , \tag{1.52}$$

where p_x, p_y and p_z are the Cartesian coordinates of the momentum $\vec{p}$ (p_z is directed along the beam direction) and the transverse mass is defined as $m_T = \sqrt{p_T^2 + m^2}$. The rapidity y is defined by the relation:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \; , \tag{1.53}$$

and is not invariant for relativistic transformations but since it transforms according to the law $y = y + \beta$ (where β is the relative velocity between two frames), the rapidity differences Δy are boost invariant. In the approximation of ultra-relativistic particles ($p \gg m$, $E \simeq |\vec{p}|$), rapidity can be approximated by the pseudorapidity η, defined as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \simeq y \; , \tag{1.54}$$

where θ is the angle between the particle and the beam direction, ($\cos\theta = p_z/|\vec{p}|$) and is therefore directly measurable in the detector. The transverse energy $E_T = E \sin\theta$ is another useful variable.

1.6 Monte Carlo simulations

The structure of events produced at high energy colliders is extremely complex, and complex numeric simulations are necessary to effectively simulate realistic events. Monte Carlo event generators are programs that subdivide the problem of producing realistic events into a sequence of tasks that can be handled separately with the help of both analytic and numeric computation.
Different event generators exist that implement computations at different levels of precision and with different techniques. Typically, the highest precision calculations, that take into account several orders in perturbation theory, are only available for a limited number of processes, thus making it hard to derive predictions on inclusive quantities. On the other hand these quantities can often be described with reasonable precision with programs that implement lower order calculations. A schematic representation of the different components (and calculation steps) that are implemented in event generators is shown in Fig. 1.1. The production of hadron-hadron collision events is the result of the following chain of calculations:

- the first step is the calculation of the elements of the scattering matrix for the selected process considering the partons (quarks or gluons) that participate in it as free particles. Cross sections are calculated for a pair of incoming partons extracted from the colliding hadrons. This step is performed by using the *Matrix Element* (ME) method, which calculates the matrix element associated to the Feynman diagrams of the process;

- the event production starts with two colliding hadrons with given momenta. One parton out of each hadron is selected to enter the scattering process we are interested in. This step is often referred to as *hard scattering* generation. Final state partons and leptons are produced according to the calculated differential cross sections;

- resonances produced in the hard event are allowed to decay;

- when two partons take part in the hard event, accelerated colour charges are present, thus Bremsstrahlung can occur. This effect is called *Initial State Radiation* (ISR) and is simulated by the event generator with the *Parton Shower* (PS) algorithm;

- also the final state partons can produce further radiation, called *Final State Radiation* (FSR). Such radiation is also simulated by the Parton Shower algorithm;

- in addition to the partons taking part in the hard interaction, several other parton pairs can interact during a hadron-hadron collision, giving rise to interactions with smaller transferred momentum. These Multiple Parton Interactions (MPI) contribute to the so called underlying structure of the event or "*underlying event*" (UE). Such interactions need to be simulated too if we want to produce realistic events, and also ISR and FSR need to be simulated for these collisions;

- leftovers of the interacting hadrons need to be simulated to balance the colour charge and four-momentum conservation. The beam remnant handling is thus another step in the event generation;

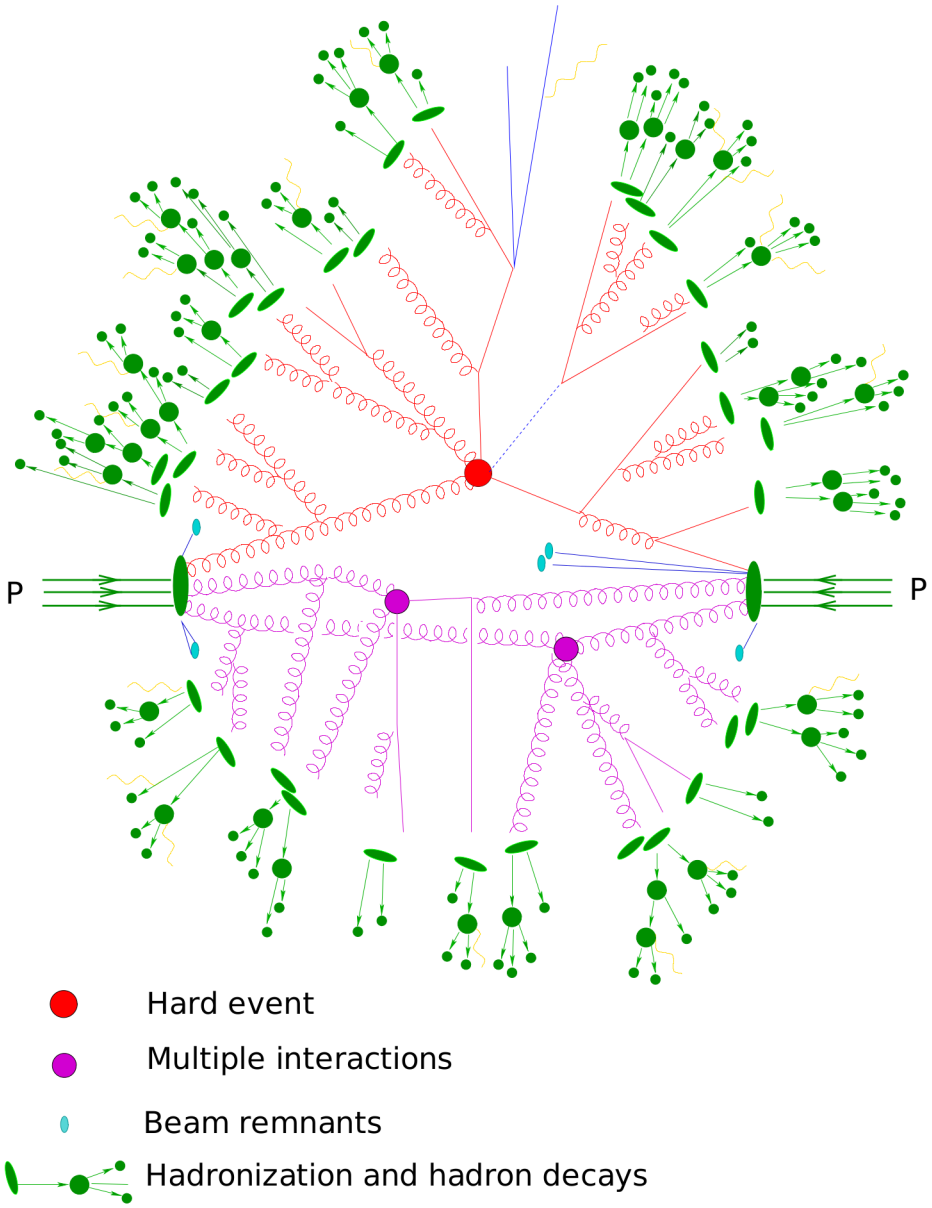


Figure 1.1: A schematic representation of the generation of an event in a typical event generator [28]. Partons from the two incoming hadrons participate in the hard scattering and in softer multiple interactions. Hadron remnants are properly treated. Quarks and gluons are turned into hadrons by hadronization and then hadrons decay.

- the resonances and the partons that are produced in the final state after the hard scattering (including processes ISR and FSR) are not observed free but are subjected to the *hadronization* process, that cannot be described in perturbative QCD. This process is simulated with some empirical models;
- finally the event generator takes care of decaying τ leptons and B-hadrons; in general particles with very short lifetime are allowed to decay by the generator itself. Those that live enough to reach the detector are left undecayed.

The main steps of the event generation, i.e. Matrix Element, Parton Shower and hadronization, will be briefly described in the following sections.

1.6.1 Matrix Element (ME)

The first step in the generation of an event is the calculation of the hard processes cross sections. General purpose event generators can perform such calculations for a wide variety of processes. The Matrix Element (ME) method is based on the exact calculation of the matrix elements and corresponds to the order-by-order calculation of Feynman diagrams in perturbative QCD. This approach is in principle more correct and it can analytically provide the complete solution of the problem.
Tree-level cross section calculations can be performed up to several (in order of eight) partons in the final state [29].
Loop calculations are complex and they are available for a limited number of processes.
The ME method presents two complications: first, the presence of a process with a high number of partons with low transverse momentum makes the calculation complicated (*soft* divergences); secondly, the method fails to describe situations in which the emitted parton is collinear with respect to the radiating parton (*collinear* divergences). Since at tree-level the loop corrections that would cancel these divergences are omitted, the phase space has to be carefully tailored to avoid the problematic regions. This means that the matrix element cross section calculations are performed away from soft and collinear divergences. In order to produce realistic events, phase-space regions omitted from the Matrix Element calculations need therefore to be recovered, with care to avoid divergences (using only the ME method, therefore, it is not possible to describe the internal structure of a jet). This is done in a quite effective way by using Parton Shower calculations.

1.6.2 Parton Shower (PS)

When treating $2 \to n$ processes, tree-level Matrix Elements suffer from divergences in the soft and collinear regions. The tree-level divergences would be removed including also virtual corrections. Such calculations, however, are extremely complex and are available only for a limited set of processes.
Parton Shower (PS) algorithm offers an alternative way both to handle the complexity of several successive branchings and to remove soft and collinear divergences. The parton showers are described by the algorithm as a sequence of elementary events $a \to bc$, where each event can happen with a certain probability. The introduction of a threshold value and the application of an angular sorting procedure in the emission of partons allows to

eliminate soft and collinear divergences typical of the ME method. The parton cascade is evolved down to a certain virtuality, of the order of 1 GeV^2. After that, non perturbative effects take place and the hadronization is applied. It should be noticed that the parton shower machinery relies on a collinear approximation of the matrix element, thus it should perform well in the description of the evolution of jets, but one cannot expect it to give a precise answer for the description of well separated parton configurations.

1.6.3 Merging Matrix Element and Parton Shower

Matrix Element and Parton Shower calculations have different virtues and different applicability limits. It is possible to summarize some of the main features of the ME calculations as follows:

- as long as tree-level is concerned, these calculations can be performed up to several (order of eight) partons in the final state;
- ME well describes separated parton configurations;
- ME calculations are exact to a given order in perturbation theory.

However:

- ME cross sections have divergences in the soft and collinear regions, thus they cannot describe the internal structure of a jet;
- since hadrons is what is observed in experiments, fragmentation models need to be applied to partons. To use bare ME partons would imply the need to tune these models for each center of mass energy; this fact limits the applicability of bare ME calculations.

On the other hand, Parton Showers:

- are universal; given the basic hard process, the PS technique will produce realistic parton configurations;
- are derived in the collinear limit, and handle divergences by requiring conservation of total probability. This makes them particularly suited to describe the evolution of jets;
- can be used to evolve partons down to a common scale; this removes the need of tuning fragmentation models at different scales.

However, since they are derived in the collinear approximation, they may fail in efficiently filling the phase space for well separated parton configurations.
The two above-mentioned techniques are therefore complementary and their combined application in the intermediate cases allows to exploit the characteristics of the two algorithms in the respective limits of validity [30–32]. Several prescriptions exist to perform ME-PS matching avoiding double-counting or holes in the phase space.
In this thesis work, the POWHEG [33] and MadGraph [34] generators are used for matrix element calculation, interfaced to PYTHIA [35] for PS and hadronization.

1.7 Experimental results in the SM Higgs boson search

The Large Hadron Collider (LHC), operating at CERN since 2009, has been designed having in mind the Higgs boson search as one of the main goals and highlights.
The Higgs search at the LHC can be performed in a range of m_H values that goes from 115 GeV up to 1 TeV.
In 2012, ATLAS [36] and CMS [37] experiments observed a boson with a mass of about 125 GeV. Results shown in the cited papers were consistent with the expectations for a Standard Model Higgs boson. The boson was consistent with the hypothesis of being a scalar (in fact it was also observed in the decay channel $H \to \gamma\gamma$) and its signal strenght parameter, which is the ratio $\mu = \sigma/\sigma_{SM}$ between the observed production cross section σ and the production cross section predicted by the Standard Model σ_{SM}, was shown to be consistent with the SM hypothesis $\mu = 1$.
Further studies performed by the CMS and ATLAS collaborations to investigate the properties of the new particle will be presented in Sec. 1.7.3, after we have examined the main production mechanisms and the most relevant decay channels of the Higgs boson at the LHC.

1.7.1 Higgs production mechanisms at the LHC

In a proton-proton collider like the LHC, several different mechanisms are possible for the Higgs boson production (Fig. 1.2). From the most to the less probable, they are:

- gluon-gluon fusion ($gg \to H$), where a couple of gluons interact via a quark loop (mainly, a top loop because of the strong coupling of the Higgs boson to massive particles) and give rise to a Higgs boson, see Fig. 1.2a;

- vector-boson fusion (VBF) ($qq \to qqH$), where each of the interacting quarks q_1, q_2 emits a virtual W or Z boson, which eventually interact generating a Higgs boson, see Fig. 1.2b. Quarks q_1 (q_1'), q_2 (q_2'), deriving from q_1, q_2 after the vector boson emission, proceed according to the original direction and are present in the final state, giving a peculiar signature of this production mechanism, as will be shown in the following;

- the vector-boson associated production $q\bar{q} \to WH/ZH$ (also known as *Higgsstrahlung*, see Fig. 1.2c), where the Higgs boson is emitted from a W or a Z boson which has been produced by a pair of quark-antiquark;

- the top-quark associated production ($q\bar{q}/gg \to t\bar{t}H$) where a pair of top/antitop is present in the final state and whose diagram is shown un Fig. 1.2d.

At LHC, the gluon fusion is the dominating Higgs production mechanism over the entire Higgs mass spectrum. The second contribution to the Higgs boson production is the vector-boson fusion mechanism. The cross section for this process is about one order of magnitude lower than for the gluon fusion one in a wide range of m_H values, even if they become almost comparable for very high Higgs boson mass values (Fig. 1.3). The cross sections of these processes depend on the Higgs boson mass (see again Fig. 1.3) as

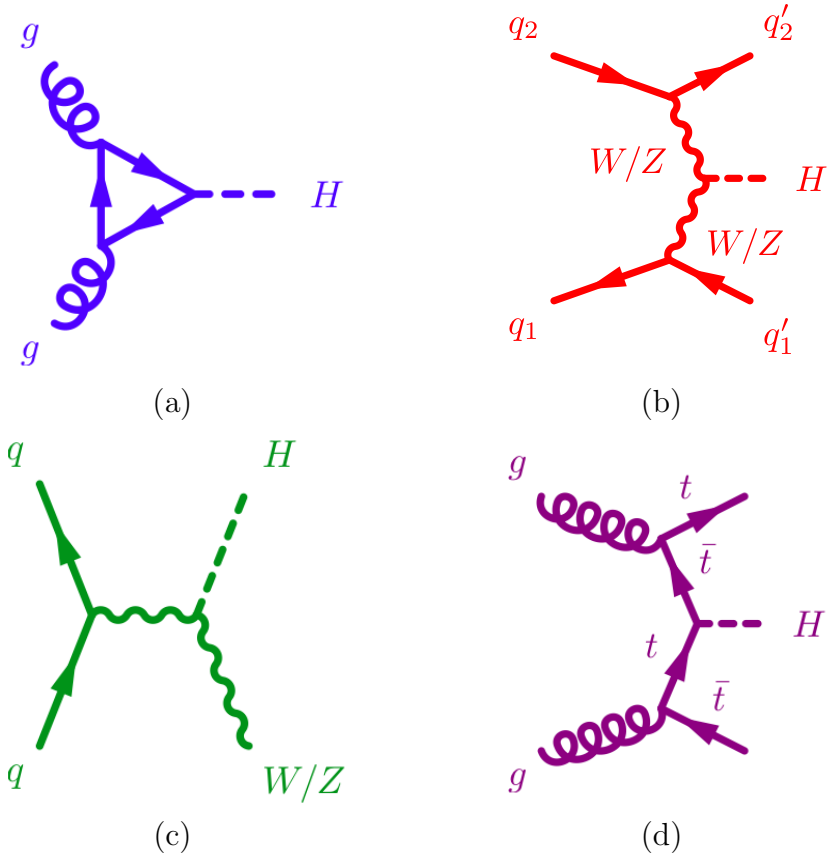


Figure 1.2: Higgs boson production mechanisms at tree level in proton-proton collisions: (a) gluon-gluon fusion; (b) vector boson fusion (VBF); (c) W and Z associated production (*Higgsstralhung*); (d) production associated to a top/antitop pair.

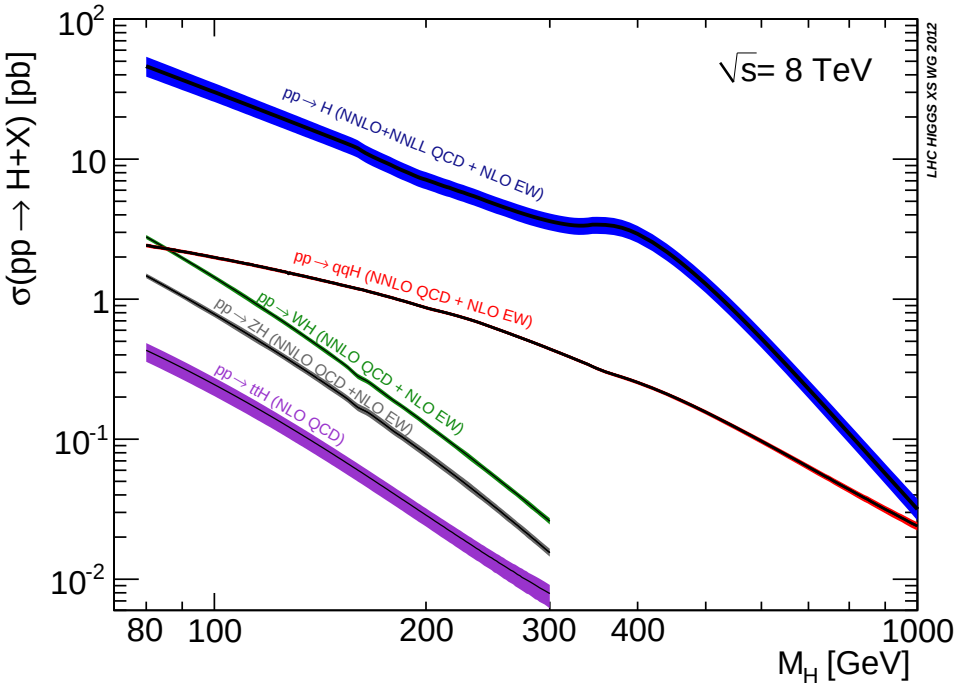


Figure 1.3: SM Higgs production cross section for p-p collisions at $\sqrt{s} = 8$ TeV, for the different production mechanisms, as a function of the Higgs boson mass, including theoretical uncertainties [38].

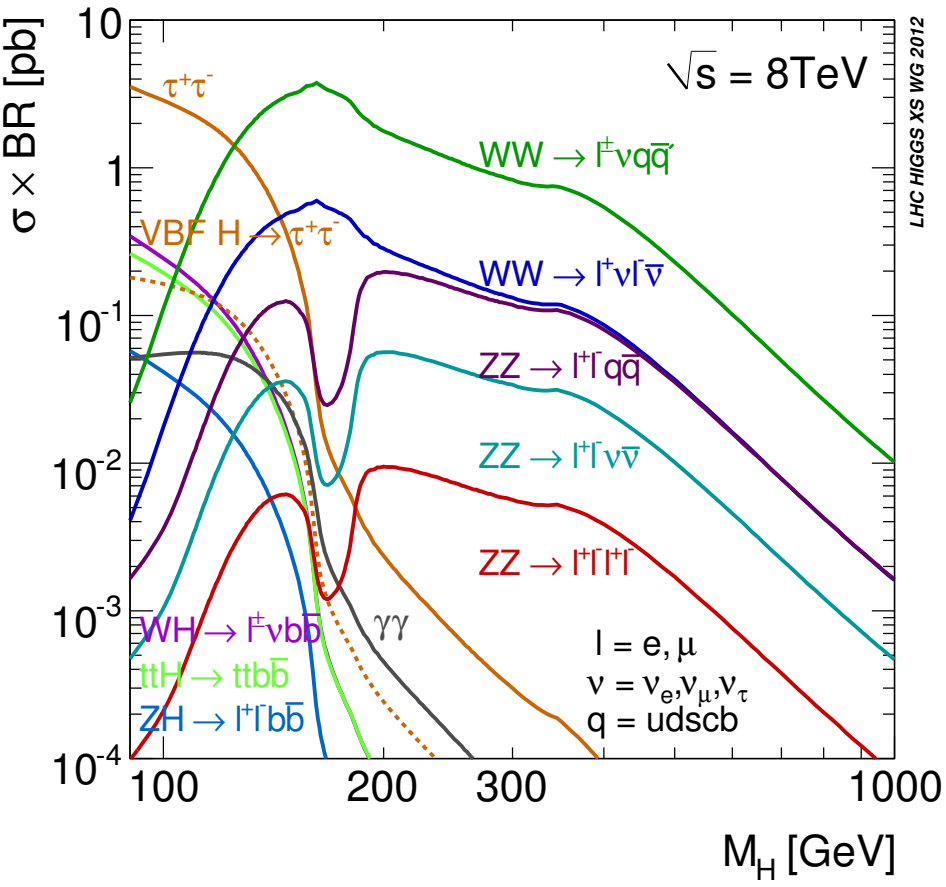


Figure 1.4: Standard Model Higgs boson production cross section times branching ratio at $\sqrt{s} = 8$ TeV, as a function of the Higgs mass [38].

well as on the pp collision center-of-mass energy ($\sqrt{s}$). In general, all production cross sections decrease with increasing the Higgs boson mass. Raising the center of mass energy, instead, reflects in an increase in Higgs production cross sections over all the mass range.

1.7.2 Higgs decay channels at the LHC

The branching ratios (BRs) for the most relevant decay modes of the SM Higgs boson, as a function of its mass, are shown in Fig. 1.4. For low masses, the Higgs boson mainly decays to a bottom quark pair. As m_H increases, decays to vector bosons pairs become energetically allowed, and constitute the dominant decay channels for masses larger than 150 GeV.

1.7.3 Measured properties of the observed Higgs boson

After the claim, on the 4th of July 2012, of the observation of a new boson by the ATLAS and the CMS collaborations, a huge effort has been addressed to go further into the determination of the boson properties. Latest results from both the experiments confirm that all the properties of the discovered particle are consistent with the hypothesis of being the Higgs boson predicted by the Standard Model. Such properties are, mainly:

- the signal strength modifier μ, which is the ratio between the observed production cross section and the one predicted by the Standard Model at that particular mass value;
- the couplings to fermions and bosons;
- the spin and the behaviour under a parity transformation;
- the width of the resonance.

The CMS experiment, which investigated the Higgs decay channels to ZZ, WW, $\gamma\gamma$, $\tau\tau$, $b\bar{b}$ using 2011 and 2012 data, has recently published combined results [39] about the Higgs properties.
The combined best-fit signal strength, relative to the SM expectation, is found to be $\mu = 1.00 \pm 0.09(\text{stat.})^{+0.08}_{-0.07}(\text{theo.}) \pm 0.07(\text{syst.})$ at the measured mass, i.e., combining all the sources of uncertainties, $\mu = 1.00 \pm 0.13$, which is consistent with the SM prediction of $\mu = 1$.
Fig. 1.5 shows the μ values obtained in different subcombinations of channels for $m_H = 125.0$ GeV, grouped by predominant decay mode or by additional tags targeting events from particular production mechanisms.

For the couplings of the Higgs boson to SM particles, the combination of all measurements in all decay channel is used to extract ratios between the observed coupling strengths and those predicted by the SM. The formalism used to test for deviations in the magnitudes of the Higgs boson scalar couplings is that set forth by the LHC Higgs Cross Section Working Group in [40]. This formalism assumes, among other things, that the observed state has $J^P = 0^+$ and that the narrow width approximation holds, leading to a factorization of the couplings in the production and decay of the boson. The four main Higgs boson production mechanisms can be associated with either fermion couplings (ggH and ttH) or vector-boson couplings (VBF and VH) of the Higgs boson. Therefore, a combination of channels associated with a particular decay mode tag, but explicitly targeting different production mechanisms, can be used to test the relative strengths of the couplings to the vector bosons and fermions (mainly the top quark). Fig. 1.6 shows the 68% confidence level (CL) regions for the signal strength modifiers associated with the ggH and ttH, and with the VBF and VH production mechanisms, $\mu_{ggH,ttH}$ and $\mu_{VBF,VH}$, respectively. The five sets of contours correspond to the five combinations by predominant decay mode. The SM Higgs boson expectation of (1, 1) is within the 68% CL regions in all cases.

About spin and parity properties, $H \to ZZ \to 4l$, $H \to WW$ and $H \to \gamma\gamma$ channels confirm the hypothesis of a scalar boson ($J^P = 0^+$). The pseudo-scalar hypothesis ($J^P = 0^-$) is disfavoured by results in the ZZ and WW decay channels; in particular, $H \to ZZ \to$

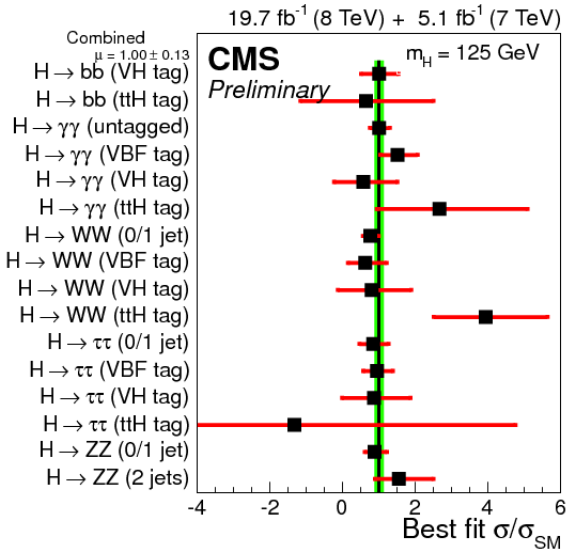


Figure 1.5: Values of the best-fit σ/σ_{SM} for the combination (solid vertical line) and for subcombinations. The vertical band shows the overall σ/σ_{SM} uncertainty. The σ/σ_{SM} ratio denotes the production cross section times the relevant branching fractions, relative to the SM expectation. The horizontal bars indicate the ± 1 standard deviation uncertainties in the best-fit σ/σ_{SM} values for the individual modes; they include both statistical and systematic uncertainties.

$4l$ excludes this hypotesis at 99% CL or higher. The hypothesis of mixture of vector (1^+) and pseudo-vector (1^-) states is also highly disfavoured by the results in $H \to ZZ \to 4l$ channel [3]. The hypothesis of a $J^P = 2^+$ particle has been tested in $H \to ZZ \to 4l$, $H \to \gamma\gamma$ and $H \to WW$ decay channels, and also this hypothesis is excluded.
About the Higgs width, the Standard Model predictes the value of the resonance width for each value of its mass (see Fig. 1.7). The predition for a 125.6 GeV Higgs[4] is $\Gamma_H^{SM} = 4.15$ MeV. A direct measurements can be performed in the $H \to ZZ \to 4l$ decay channel, but the experimental resolution does not allow to give a useful constraint ($\Gamma_H < 3.4$ GeV at 95% CL). An indirect method based on the measurement of both the on-shell and the off-shell production cross section, and exploiting both the $H \to ZZ \to 4l$ and $H \to ZZ \to 2l2\nu$ channels (the latter to increase the statistics for the off-shell measurement) [41] gives a much tighter constraint on the Higgs width: $\Gamma_H < 17.4$ MeV.
Similar results for the Higgs boson signal strength and couplings have been obtained by the ATLAS collaboration, from the combination of the $H \to \gamma\gamma$, $H \to ZZ \to 4l$, $H \to WW \to l\nu l\nu$, $H \to b\bar{b}$ and $H \to \tau\tau$, always confirming the Standard Model compatibility [42]. Also the spin-parity measurements performed by ATLAS [43] in the $H \to \gamma\gamma$, $H \to WW \to l\nu l\nu$ and $H \to ZZ \to 4l$ favour the $J^P = 0^+$ hypothesis, while the $J^P = 2^+$ hypothesis is excluded with a confidence level above 99.9%.

[3] We are thus assuming that this resonance is not the same observed in the $\gamma\gamma$ decay.
[4] This is the value measured in the $H \to ZZ \to 4l$ decay channel only.

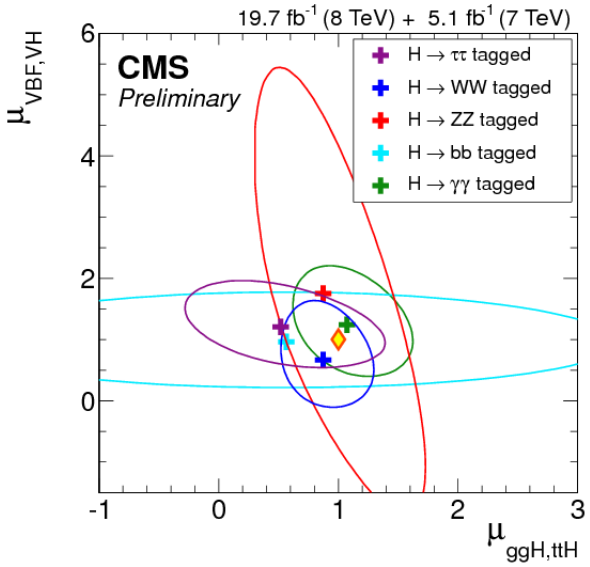


Figure 1.6: The 68% CL regions (bounded by the solid curves) for the signal strength of the ggH and ttH, and of the VBF and VH production mechanisms, $\mu_{ggH,ttH}$ and $\mu_{VBF,VH}$, respectively. The different colours show the results obtained by combining data from each group of predominant decay modes: $\gamma\gamma$ (green), WW (blue), ZZ (red), $\tau\tau$ (violet), $b\bar{b}$ (cyan). The crosses indicate the best-fit values. The diamond at (1, 1) indicates the expected values for the SM Higgs boson.

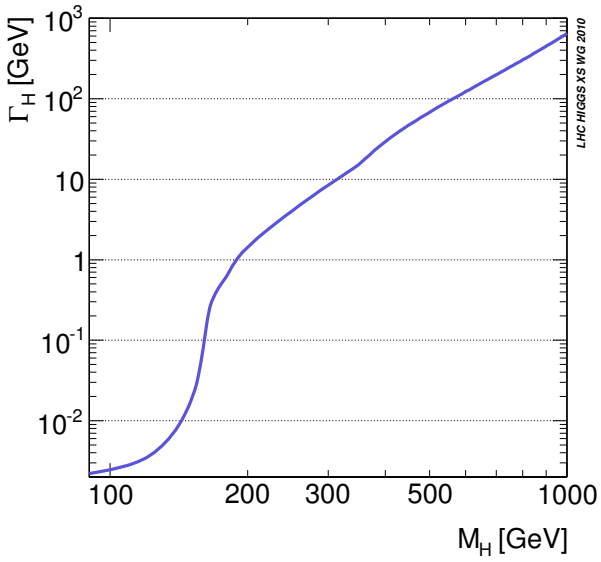


Figure 1.7: Higgs boson width as a function of its mass, according to the Standard Model prediction.

About the boson mass, which however is not predicted by the SM and consequently cannot be taken as an index of the compatibility with the model, the latest measurements are:

- by the CMS experiment, from the combination of the $H \to ZZ \to 4l$ and the $H \to \gamma\gamma$ decay channels, a value $m_H = 125.03^{+0.26}_{-0.27}(\text{stat.})^{+0.13}_{-0.15}(\text{syst.})$ has been obtained [39]
- by the ATLAS experiment, from the combination of the same channels[5], a value $m_H = 125.36 \pm 0.37(\text{stat.}) \pm 0.18(\text{syst.})$ has been obtained [44].

[5]Using the pp collision data sample recorded at $\sqrt{s} = 7$ TeV and 8 TeV, corresponding to an integrated luminosity of 25 fb^{-1}, see Chap. 2 for the definition of integrated luminosity.

1.7.4 The $H \to ZZ \to 2l2q$ decay channel

Let's now concentrate on the $H \to ZZ$ channel, the analysis described in the following being focused on a particular final state of this decay process. The $H \to ZZ$ cross section has a maximum around $\sim 2\ m_Z$, when the production of two real Z bosons becomes possible. The product of the $H \to ZZ$ cross section times its branchig ratio reaches highest values for $m_H > 200$ GeV, being the decay into a pair of W bosons the one with the highest cross section times branching ratio in the mass interval after $\sim 2\ m_W$.
The process $H \to ZZ \to 4l$ has a very clean final state and allows a precise reconstruction of the Higgs boson thanks to the presence of isolated charged leptons with high transverse momentum. For this reason, it is a very powerful discovery channel (the "golden channel"). Other final states of $H \to ZZ$ decay such as $H \to ZZ \to 2l2q$ and $H \to ZZ \to 2l2\nu$ are mostly relevant for a Higgs boson mass above the ZZ doubly resonant peak, where the background contribution can be significantly reduced and the production is very abundant. In particular, the $ZZ \to l^+l^-q\bar{q}$ has a $\sigma\times$BR product about 20 times the "golden channel" $ZZ \to 4l$ decay BR.
The yield for the decay channels started by the $H \to WW$ decay is even higher, but the presence of neutrinos in the final state does not allow to have a completely reconstructed final state. Viceversa, $H \to ZZ \to 2l2q$ decay allows to have a high signal yield and, at the same time, a completely reconstructed final state, with two leptons and two jets coming from the hadronization of the quark-antiquark pair.
The production mechanism which has been investigated in this analysis work is the Vector Boson Fusion (VBF). As stated above (see also Fig. 1.3), its cross section is between one and two orders of magnitude lower than that of gluon-gluon fusion for a wide range of Higgs masses (~ 10 times smaller for $m_H \simeq 125$ GeV, almost ~ 100 times for $m_H \simeq 400$ GeV), thus its observation is more challenging. Nevertheless, it becomes competitive with the gluon-gluon fusion mechanism for Higgs masses reaching very high values; for m_H approaching 1 TeV, the two production cross sections become comparable.
The cross section of the main background affecting this decay channel, that is Drell-Yan lepton pairs plus jets events, decreases by about two orders of magnitude[6] when requesting the presence of four jets instead of two. So the VBF production, providing two more jets (the *VBF jets* coming from the hadronization quarks from production) to the final state, benefits from a highly reduced background with respect to the gg production mode, such that even if the VBF production mechanism has a branching ratio smaller than the gluon-gluon fusion, a higher signal-to-noise ratio is expected. Furthermore, the peculiar topology of the two VBF or "*tagging*" jets in the final state is a good handle to discriminate signal from background events. VBF jets, in fact, tend to be emitted "forward", that is, with a high angular separation[7] and with high invariant mass, as they are highly energetic.
In the following, the search for a high-mass Standard Model-like Higgs boson produced with the VBF mechanism and decaying into $ZZ \to 2l2q$ is described. A possible theoretical frame for such a Higgs boson is the Electroweak Singlet interpretation described

[6] If assuming the jet energy of the order of the Z mass.

[7] Here we refer to the polar angle which is measured with respect to the beam line; see Cap. 2.2 for the definition of the coordinate system.

above. Other extensions of the SM Higgs sector (i.e., the Two Higgs Doublet Model) would be possible interpretations as well, but will not be considered in this work.

Chapter 2

The CMS experiment at the LHC

2.1 The *Large Hadron Collider*

The *Large Hadron Collider* (LHC) [45–48] at CERN, officially inaugurated on 21st October 2008, is the largest and most powerful hadron collider ever built. Installed in the underground tunnel which housed the *Large Electron Positron Collider* (LEP) [49–51], the leptonic accelerator in operation until 2nd November 2000, the LHC accelerator has the shape of a circle with a length of about 27 km and is located at a depht varying between 50 m to 175 m, straddling the Franco-Swiss border near Geneva. It is designed to collide two 7 TeV counter-circulating beams of protons resulting in a center-of-mass energy of 14 TeV, or two beams of heavy ions, in particular lead nuclei at an energy of 2.76 TeV/nucleon in the center-of-mass frame.

The transition from a leptonic collider to a hadronic collider entailed the following advantages: first, it has been possible to build a machine that having the same size of the previous one (and therefore accommodated in the same LEP tunnel, substantially reducing the cost and time of construction), could reach a higher energy in the center-of-mass frame. This is due to the much lower amount of energy loss through synchrotron radiation emitted by the accelerated particles, that is proportional to the fourth power of the ratio E/m between their energy and their mass. Secondly, the composite structure of protons compared to the elementary structure of electrons allows LHC to be able to access simultaneously a wider energy spectrum, despite the production of many low energies particles in a complex environment. This feature is particularly important for a machine dedicated to the discovery of "new" physics.

In Fig. 2.1 a schematic description of the accelerator complex installed at CERN is shown. The acceleration is performed in several stages [48]. The protons source is a *Duoplasmatron*: the protons are obtained by removing electrons from a source of hydrogen gas and then sent to the LINAC2, a 36 m long linear accelerator which generates a pulsed beam with an energy of 50 MeV using Radio Frequency Quadrupoles (RFQ) and focusing quadrupole magnets. The beam is subsequently sent to the *Proton Synchrotron Booster* (PSB), a circular accelerator consisting of four superimposed synchrotron rings with a circumference of about 160 m, which increases the proton energy up to 1.4 GeV. Then, protons are injected into the *Proton Synchrotron* (PS), a single synchrotron ring with a circumference of about 600 m where the energy is increased to 25 GeV. The sequential

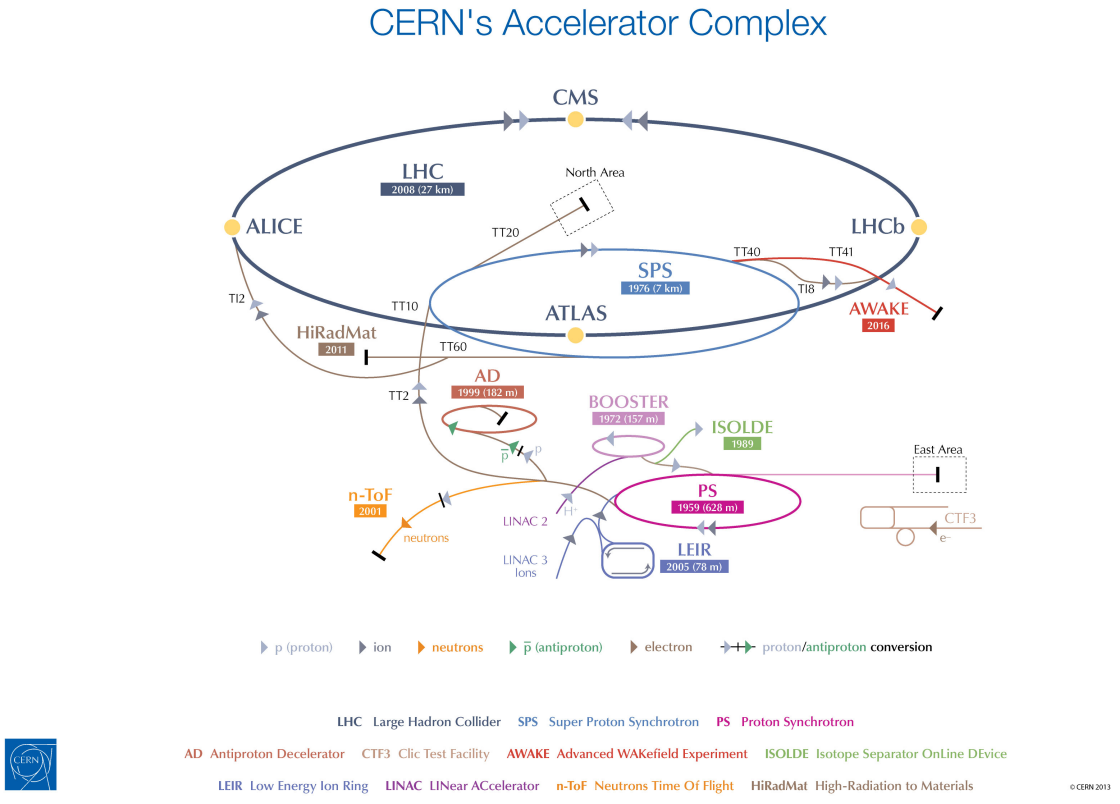


Figure 2.1: Schematic description of the accelerator complex installed at CERN.

combination of these two synchrotrons also allows to create a series of protons bunches interspersed by 25 ns (i.e. at the frequency of 40 MHz) as required for the final correct operation of LHC. The final proton injection stage is the *Super Proton Synchrotron* (SPS), a synchrotron with a circumference of approximately 7 km where protons reach an energy value of 450 GeV. Subsequently, protons are extracted and injected into the LHC ring via two transmission lines, to generate two beams running in opposite directions in two parallel pipes and which are accelerated up to the energy of interest. The beams collide at four interaction points where the four main experiments (ALICE, ATLAS, CMS and LHCb) are placed:

- CMS (*Compact Muon Solenoid*) [52] and ATLAS (*A Toroidal LHC ApparatuS*) [53] are two general-purpose detectors designed to investigate the largest possible spectrum of physics. In particular, they have been devoted to the detection of particles produced by a Higgs boson decay and to look for any evidence of possible new physics. The use of two detectors chasing the same objectives but designed independently is crucial for a cross-check of any possible new discovery;

- LHCb (*LHC beauty*) [54] is an experiment primarily designed to study CP (combined Charge conjugation and Parity symmetry) violation in electroweak interactions and to study asymmetries between matter and antimatter through the analysis of rare decays of hadrons containing b quarks. The detector is also able to perform measurements in the forward region, at small polar angles with respect to the beam line;

- ALICE (*A Large Ion Collider Experiment*) [55] is an experiment studying heavy ions collisions, through the production of a new state of matter called *quark-gluon plasma*.

Two small experiments, TOTEM and LHCf, which focus on forward particles, have also been built. TOTEM (*TOTal Elastic and diffractive cross section Measurement*) [56] measures proton-proton interaction cross section and accurately monitors the luminosity of the LHC using detectors positioned on either side of the CMS interaction point. LHCf (*LHC forward*) [57] is made up of two detectors which sit along the LHC beamline, at 140 meters either side of the ATLAS collision point. It uses particles thrown forward by LHC collisions as a source to simulate the interaction with the atmosphere of very high energy cosmic rays (between 10^{17} TeV and 10^{20} TeV) in laboratory conditions.

At the LHC, a series of about 1200 magnetic dipoles bend the beam along the accelerator ring. They are located along the "arc" structures of the circumference. The ring, in fact, can be subdivided into octants, with eight curve regions (the "arcs") separated by rectilinear regions. In these straight regions, instead, almost 400 focusing and defocusing quadrupoles are located, which maintain the beam stable along the orbit, and some other small multipolar magnets (sextupoles and octupoles) are used to make additional minor corrections to the beam direction. A radio frequency acceleration system, consisting of 16 superconducting radio-frequency resonant cavities, is used to increase the proton energy by 0.5 MeV with each beam revolution. The 7 TeV per-beam-energy limit on the LHC is not determined by the electric field generated by the radiofrequency cavity but by the

magnetic field necessary to maintain the protons in orbit, given the current technology for the superconducting magnets.

In Tab. 2.1 the main parameters for the LHC machine are listed.
One of the most important parameter of an accelerator is the *instantaneous luminosity* $\mathcal{L}$. For a process having a cross section σ and producing N particles per time unit, the instantaneous luminosity $\mathcal{L}$ is defined by the relation

$$N = \sigma \mathcal{L} \ . \tag{2.1}$$

It is important to note that the definition is independent of the process considered.
The design luminosity for LHC is $\mathcal{L} = 10^{34}\ cm^{-2}s^{-1}$, leading to around 1 bilion proton-proton interaction per second.
Instantaneous luminosity can be expressed as a function of the characteristic parameters of the beam using the following approximate formula [58]:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} \tag{2.2}$$

where n_1 and n_2 represent the number of particles contained in the two colliding bunches at the frequency f and σ_x, σ_y are the transverse dimensions of the beams. At the LHC nominal luminosity, the proton bunches will collide at a frequency of about 40 MHz. The transverse dimensions of the beam can be squeezed down to 15 μm.
Then the *integrated luminosity* L can be defined as

$$L = \int \mathcal{L}\, dt \ . \tag{2.3}$$

The LHC startup was in September 2008, but due to an accident caused by a failure in an interconnection between two magnets, the operation was stopped and restarted in March 2010. LHC has been running at $\sqrt{s} = 7$ TeV in 2010 and 2011 and at $\sqrt{s} = 8$ TeV in 2012. In particular, the LHC ran from 30th March 2010 to 31st November 2010 and from 13th March 2011 to 30th November 2011 producing proton-proton collisions at a center-of-mass energy $\sqrt{s} = 7$ TeV and delivering an integrated luminosity of about, respectively, 44 pb^{-1} and 6.1 fb^{-1} of data. The encouraging results in the Higgs boson search provided by the ATLAS and CMS collaborations led to the decision to modify the LHC running program and to extend the data taking period up to the end of 2012 by raising the energy of the beams up to 4 TeV per beam, reaching the new record of 8 TeV in the center-of-mass frame. The LHC machine produced therefore proton-proton collisions from 4th April 2012 to 16th December 2012. The total integrated luminosity delivered by LHC to the experiments at the end of 2012 was 23.3 fb^{-1}.
With the constant increase in luminosity, a significant probability that multiple proton-proton interactions are produced in a single bunch crossing has to be taken into account. In this instance, one is typically concerned about identifying and reconstructing a single primary collision where a physics event of interest occurs amongst the background of the additional proton-proton collisions. Such backgrounds are due to processes that occur with high likelihood, like low-p_T jet production. These additional collisions are known as pileup (PU). The rate of pileup is quoted in units of the number of additional collisions.

Table 2.1: LHC technical parameters for proton-proton collisions.

Parameter	**Value**
Maximum magnetic dipole field	8.33 T
Dipole operating temperature	1.9 K
Beam energy at injection	450 GeV
Beam energy at collision (nominal)	7 TeV
Beam energy at collision (2010-2011)	3.5 TeV
Beam energy at collision (2012)	4 TeV
Maximum instantaneous luminosity (nominal)	10^{34} cm^{-2} s^{-1}
Maximum instantaneous luminosity (2012)	$7.7 \cdot 10^{33}$ cm^{-2} s^{-1}
Number of bunches per proton beam (nominal)	2808
Maximum number of bunches per proton beam (2012)	1380
Maximum number of protons per bunch	$1.69 \cdot 10^{11}$
Bunch separation in time (nominal)	25 ns
Bunch separation in time (2012)	50 ns
Collision frequency (nominal)	40 MHz
Collision frequency (2012)	20 MHz
Transverse beam size (collision point)	15 μm
Energy loss per turn (at 14 TeV)	7 keV

The 2012 LHC run at $\sqrt{s} = 8$ TeV had an average pileup rate of 21 additional collisions, with some events exhibiting well over 40 pileup collisions (see Fig. 2.2).

The LHC is now experiencing its first long shutdown, LS1. Over the coming months major consolidation and maintenance work will be carried out across the whole of CERN accelerator chain. The LHC will be prepared for higher energy running, and the experiments will undergo essential maintenance. LHC running is scheduled to resume in the first months of 2015, while the rest of the CERN complex is already starting up again.

2.2 The *Compact Muon Solenoid* detector

The Compact Muon Solenoid (CMS) [52] is a general purpose detector optimized for the proton-proton interactions analysis with the expected energy and luminosity of the LHC particle accelerator design, identifying with precision muons, electrons and photons. It has been designed to investigate a wide range of physics, with the search for the Higgs boson as one of the main highlights. Search for new physics is also an important goal of the experiment, as well as top physics and, of course, Standard Model precision measurements. Although it has the same scientific goals as the ATLAS experiment, it uses different technical solutions and a different magnet-system design. The CMS experiment is one of the largest international scientific collaborations in history, involving more than 4000 people (particle physicists, engineers, technicians, students and support staff) from about 180 universities and institutes in more than 40 countries.

The experiment is placed in a cavern 100 m underground in the area called Point 5 (an

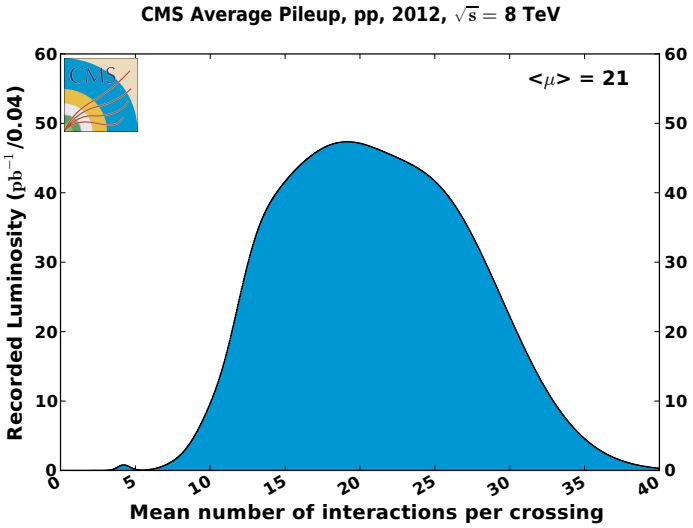


Figure 2.2: Average pileup distribution in the recorded data at 8 TeV.

old LEP access point) near the village of Cessy, in France.
The coordinate system used in CMS is a right-handed Cartesian system, having the origin in the nominal beam collision point inside the detector. The x-axis points radially towards the center of the LHC ring, the y-axis is directed upwards along the vertical and the z-axis is oriented along the direction of the beams, along the counter-clockwise direction of the ring if seen from above. The cylindrical symmetry of CMS design and the invariant description of proton-proton physics suggest to define a new coordinate system based on pseudo-angular coordinates, given by the triplet (r, φ, η), where r is the distance from the z-axis, φ is the azimuth angle measured on the $x - y$ plane starting from the x-axis and η is the pseudorapidity defined in Eq. 1.54.

The CMS detector (shown in Fig. 2.3) is 21.6 m long, has a diameter of 15 m and a weight of about 12, 500 tons. The constructive element that characterizes the experiment is a solenoidal superconducting magnet which produces an internal constant magnetic field of 3.8 T along the direction of the beams. The CMS detector is designed as a dodecagonal base prism. The central part of the prism, named *barrel*, contains several layers of detectors with cylindrical symmetry, coaxial with respect to the direction of the beams. A set of detector disks, called *endcaps*, close the detector at its ends, to ensure its tightness. From the inner region to the outer one, the various components of CMS are:

- **Silicon Tracker**: it is placed in the region $r < 1.2$ m and $|\eta| < 2.5$. It consists of a silicon pixel vertex detector and a surrounding silicon microstrip detector, with a total active area of about 215 m^2. It is used to reconstruct charged particle tracks and vertices;

- **Electromagnetic Calorimeter (ECAL)**: it is placed in the region 1.2 m $< r <$ 1.8 m and $|\eta| < 3$. It consists of scintillating crystals of lead tungstate ($PbWO_4$) and it is used to measure the trajectory and the energy released by photons and electrons;

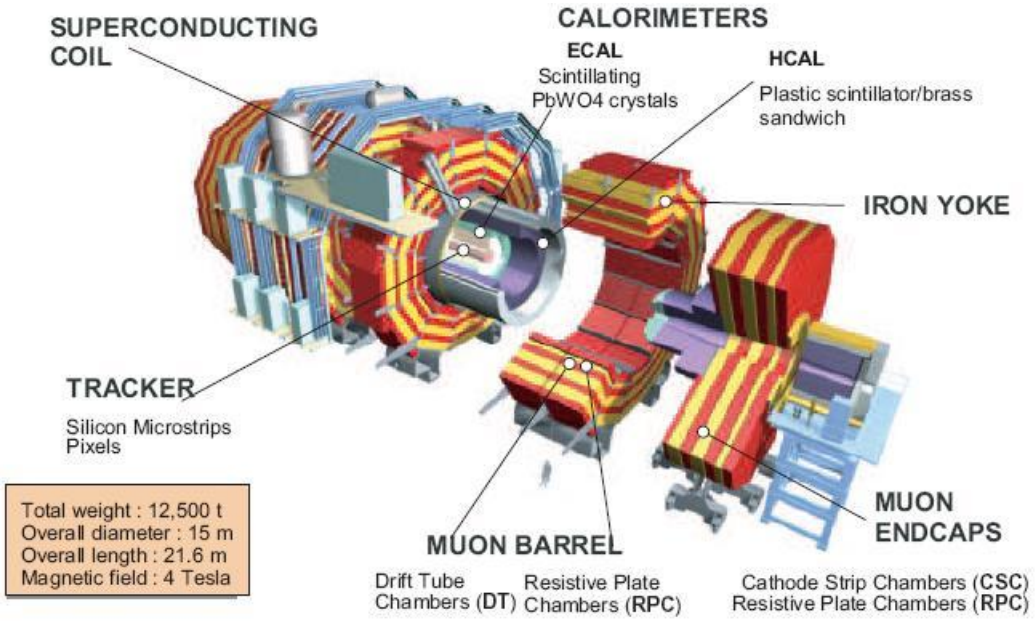


Figure 2.3: A view of the CMS detector with its subdetectors labeled.

- **Hadron Calorimeter (HCAL)**: it is placed in the region 1.8 m $< r <$ 2.9 m and $|\eta| < 5$. It consists of brass layers alternated with plastic scintillators and it is used to measure the direction and the energy released by the hadrons produced in the interactions;

- **Superconducting Solenoidal Magnet**: it is placed in the region 2.9 m $< r <$ 3.8 m and $|\eta| < 1.5$. It generates an internal uniform magnetic field of 3.8 T along the direction of the beams, necessary to deflect the charged particles in order to allow a measurement of their momentum through the curvature observed in the tracking system. The magnetic field is closed with an iron yoke 21.6 m long with a diameter of 14 m, where a residual magnetic field of 1.8 T is present, in the opposite direction with respect to the 3.8 T field;

- **Muon System**: it is placed in the region 4 m $< r <$ 7.4 m and $|\eta| < 2.4$. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and in the endcaps. This composite tracking system for muons is used to reconstruct tracks released by muons that pass through it. The muons chambers are housed inside the iron structure of the return yoke that encloses the magnetic field.

Fig. 2.4 shows how the various CMS detectors contribute to the recognition of particles passing through them. The Silicon Tracker, ECAL and HCAL are located inside the magnetic coil. Muon Chambers are located in the magnet return yoke. In the following sections a brief description of each component is given.

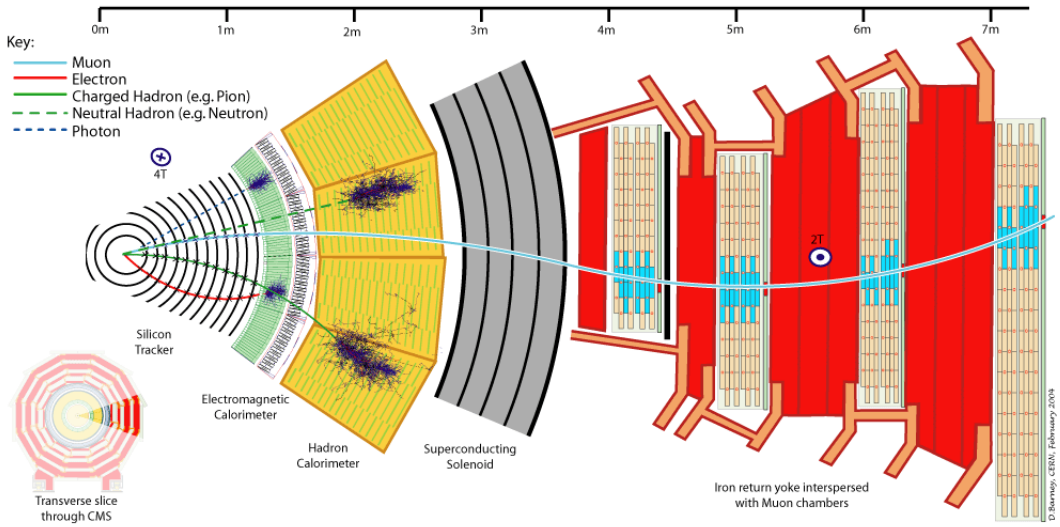


Figure 2.4: CMS detector response to the transit of different particles.

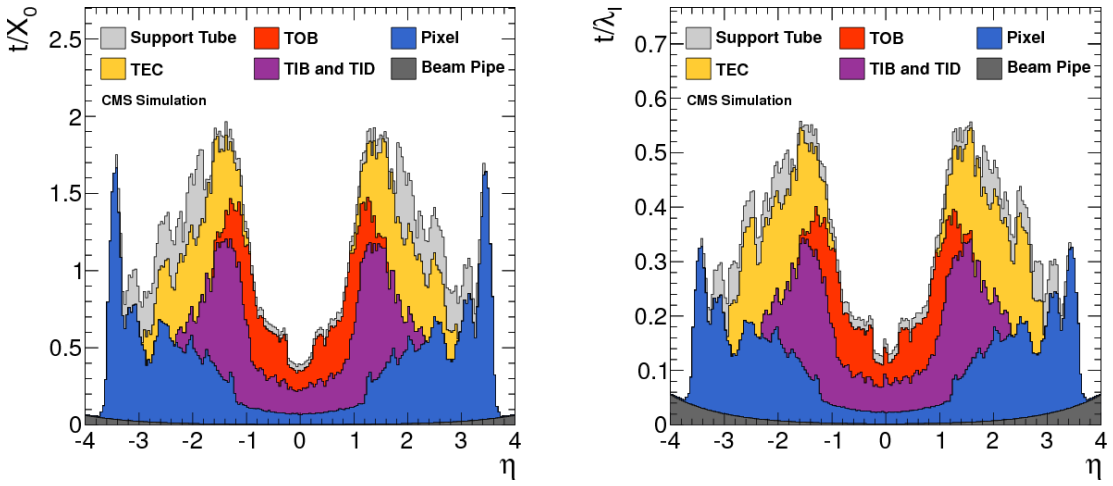
2.2.1 The Solenoid

The CMS magnet [59], which houses the tracker, the electromagnetic and the hadronic calorimeters, is the biggest superconducting solenoid ever built in the world. The solenoid achieves a magnetic field of 3.8 T in the free bore of 6 m in diameter and 12.5 m in length. The energy stored in the magnet is about 2.6 GJ at full current. The superconductor is made of four Niobium-Titanium layers. In case of a quench, when the magnet loses its superconducting property, the energy is dumped to resistors within 200 ms. The magnet return yoke of the barrel is composed with three sections along the z-axis; each one is split into 4 layers (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly saturated, and the field in the yoke is about the half (1.8 T) of the field in the central volume.

2.2.2 The Tracker

The silicon tracker [60, 61] is the detector closest to the beams collision point. Its goal is to reconstruct, with the highest possible resolution, the trajectory of charged particles originating from the interaction point and to identify the position of secondary vertices in events containing particles with short mean life (in particular, hadrons containing the quark b that decay after a few hundred μm). At the nominal luminosity of operation, 10^{34} cm^{-2} s^{-1}, 40 minimum bias events are expected for each collision. They will overlap with the event of interest and lead to about 1000 charged tracks every 25 ns. In order to make pattern recognition easier, two requirements are fundamental:

- a low occupancy detector;
- a large redundancy of measured points per track.



(a) t/X_0 as a function of η.

(b) t/λ_I as a function of η.

Figure 2.5: Material budget plots of the Tracker for each subdetector in units of radiation length t/X_0 and in units of nuclear interaction length t/λ_I as a function of η, for the nominal geometry of the detector [62].

The low occupancy is obtained building a detector with high *granularity*[1]. The redundancy of measured points (*hits*) per track is necessary to reduce the ambiguity in assigning hits to the track itself. Nevertheless, the amount of tracker material has to be as low as possible in order to avoid compromising the measures themselves: an excessive amount of material, in fact, would deteriorate the accuracy of the measurements of the tracker itself, because of particle multiple scattering, and of surroundings detector elements, especially ECAL because of an increased probability for the creation of an electron-positron pair from a photon. Therefore, the detection active layers have been limited in number and thickness. In Fig. 2.5, the material budget of the tracker is reported in units of radiation length[2] t/X_0 (Fig 2.5a) and in units of nuclear interaction length[3] t/λ_I (Figure 2.5b) as a function of η as obtained from the full simulation of the tracker. The material budget is higher in the region $1 < |\eta| < 2$ due to the presence of cables and services in this region. The tracker is composed of two subdetectors: a silicon pixel vertex detector and a silicon microstrip detector. A three-dimentional representation of the subdetectors can be found in Fig. 2.6, while in Fig. 2.7 a scheme of the transverse $(r-z)$ section of the detector layers is shown. The tracker has a cylindrical shape, is 6 m long and has an outer diameter of 2.6 m. It consists of a central barrel and two endcaps in order to maximize tightness. The structure is optimized to have on average 12-14 hits per track, in order to have a high reconstruction efficiency and a low rate of false tracks.

[1]The *granularity* of a detector is the angular range that the individual detector elements manage to resolve and it is given by the product $\Delta\eta \times \Delta\varphi$.

[2]The radiation length X_0 is defined as the path that a highly energetic electron must traverse in a material to reduce its energy to a fraction equal to $1/e$ of the initial value.

[3]The nuclear interaction length λ_I is defined as the mean free path for nuclear interactions of a hadron in a given material.

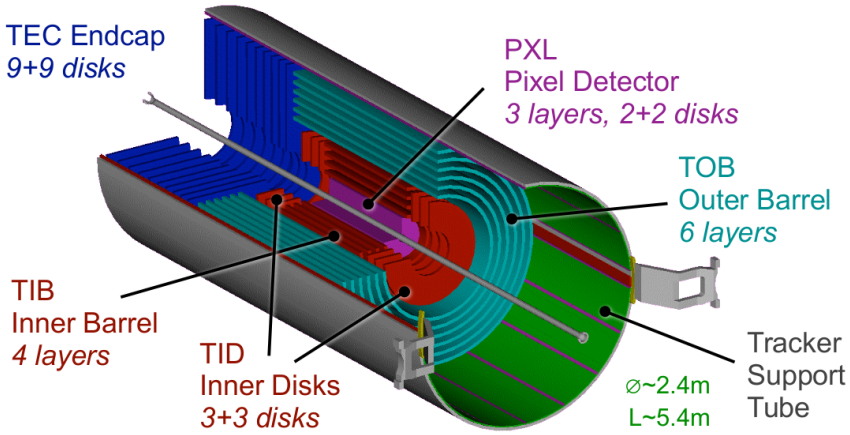


Figure 2.6: A schematic view of the CMS silicon tracker.

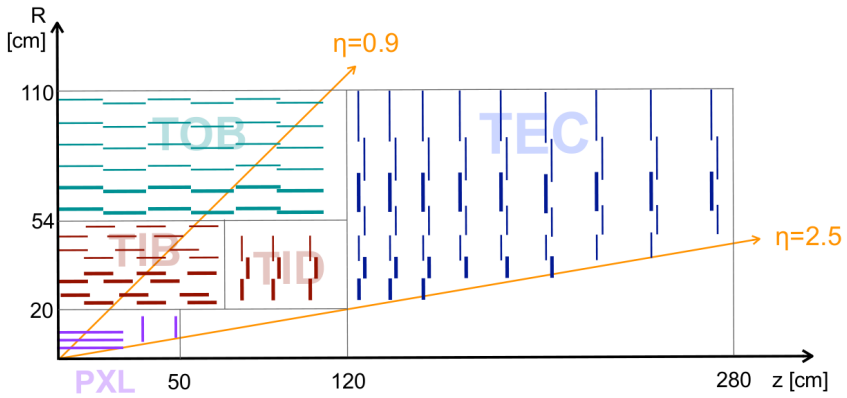


Figure 2.7: A $r-z$ section of a quarter of the tracker system. Each line-element represents a detector module. Closely spaced double line-elements indicate back-to-back silicon strip modules, in which one module is rotated through a "stereo" angle, so as to permit reconstruction of the hit positions in 3-D. Within a given layer, each module is shifted slightly in r or z with respect to its neighbouring modules, which allows them to overlap, thereby avoiding gaps in the acceptance.

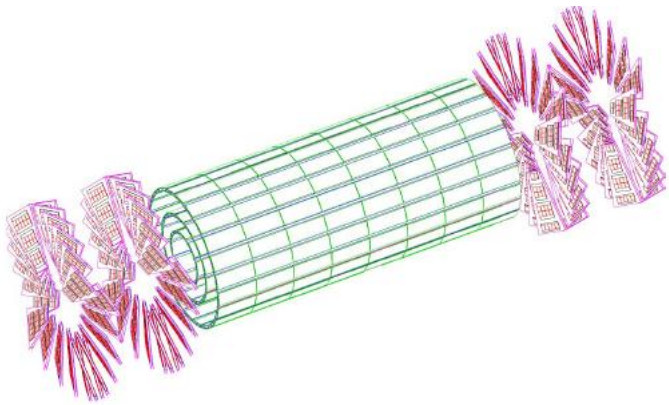


Figure 2.8: A schematic view of the pixel vertex detector.

2.2.2.1 The Pixel Vertex Detector

The pixel vertex detector, shown in Fig. 2.8, is mainly used in CMS as a starting point for the reconstruction of tracks and is essential for the reconstruction of the primary vertex (PV) and any possible secondary vertices. It is placed in the region closest to the collision point, where the particle flow is maximum. It covers the region $|\eta| < 2.5$ and is composed of a central part (barrel) and by two forward parts (endcaps). The barrel is composed of three concentric cylindrical sectors 53 cm long, located at an average distance r of 4.4 cm, 7.3 cm and 10.2 cm. Each half-cylinder is composed of *ladders* and *half ladders* that serve as support and cooling structure for the modules of pixels, with each ladder containing 8 modules. In total, the barrel is composed of 768 modules. Each endcap is composed of two disks placed at a distance of 34.5 cm and 46.5 cm from the nominal beams impact point. They cover a radius r in a range between 6 cm and 15 cm in such a way that each track included in the detector acceptance passes through at least two layers. Each disk is divided into 24 segments, on each of which 7 modules of different sizes are mounted, for a total of 672 modules on all the endcaps. Each module is composed of several units that contain a highly integrated and segmented silicon sensor with a thickness of 250 μm. In order to optimize the reconstruction of vertices and the track parameters near the vertex, a set of rectangular pixels with a size of 150×100 μm^2 are used, with the 100 μm side oriented along the $r\varphi$ direction in the barrel region and along the z direction in the endcap region. The resolution in the hit reconstruction is about $10 \div 15$ μm in the barrel and about 15 μm in the endcaps.

2.2.2.2 The Silicon Microstrip Detector

In the region of the detector that is more than 20 cm far from the beam, the flow of particles is sufficiently limited to allow the use of a silicon microstrip detector (*Silicon Strip Tracker*, SST). Overall, this detector consists of 15400 units (*modules*), composed of one or two sensors sticked on a support of carbon fiber, together with the readout electronics. In case of a "doubled" sensor, the second detector is rotated with respect to the first one in order to have strips forming an angle of 100 mrad between them. This "stereo" combination, although of lower resolution, is preferable compared to a pixel

segmentation because it has a lower number of readout channels. The ambiguities due to the hit recognition are resolved in the process of reconstruction of the entire track. The silicon microstrip tracker is 5.4 m long, extending up to a distance of 1.1 m from the axis of the beams. It consists of a barrel and two endcaps and it is divided into four distinct parts, as shown in Fig. 2.7. The detectors that constitute the barrel are:

- TIB (*Tracker Inner Barrel*): it consists of 4 cylindrical layers of detectors, covering the region up to $|z| < 65$ cm. The detectors in this region have a thickness of 300 μm and a *pitch*[4] between the strips (117 mm long) variable in the range $80 - 120$ μm. The first two layers contain stereo modules. The other modules have stripes oriented along the z direction. The measurement of the crossing point of the particle is therefore more precise in the $r - \varphi$ coordinates (spatial resolution in the range $23 - 34$ μm) rather than in the z coordinate (spatial resolution of 230 μm);

- TOB (*Tracker Outer Barrel*): it consists of 6 layers of detectors placed in the region $55 < r < 110$ cm. Double-sided modules are mounted on the two inner layers. The density of particles going through this region is lower with respect to the inner part, therefore the pitch between the strips is greater with respect to the TIB modules (in the range $120 - 180$ μm) and the strips are longer (190 mm). The spatial resolution varies in the range $35 - 52$ μm in the $r - \varphi$ coordinates and is 530 μm in the z coordinate.

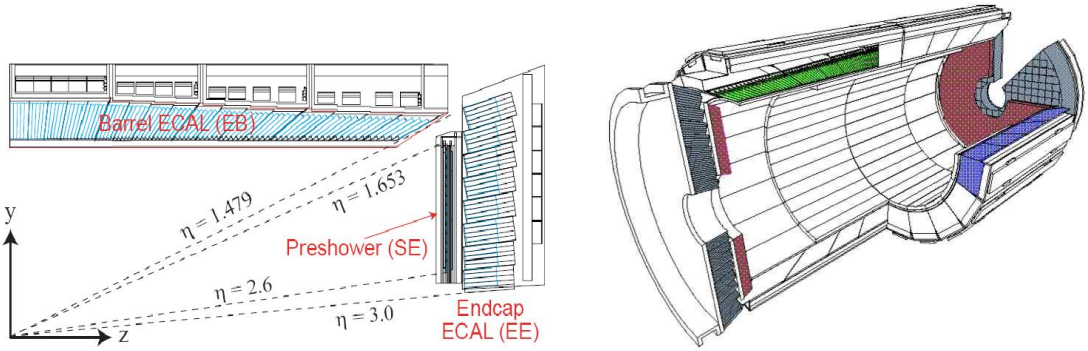
The detectors that constitute the endcaps are:

- TID (*Tracker Inner Disks*): it consists of 6 disks (3 per side), placed orthogonally to the beam direction in the region between the TIB and TOB. The TID modules are placed in a ring disposition, with the strips oriented in the radial direction, and they are alternately positioned on the internal and on the external side of the disk. The two innermost rings of the TID are equipped with stereo modules. The thickness of the silicon is 300 μm;

- TEC (*Tracker EndCap*): each one of these detectors is made of 18 disks (9 per side) which extend in the region 120 cm $< |z| <$ 280 cm. Each disk is divided into 8 slices in each of which a number ranging from 4 to 7 modules are mounted in a ring shape, depending on the position along z. Also in this case the modules are alternately mounted on the internal and on the external side of the disk, with the strips radially oriented. On the two innermost rings and on the fifth one the modules are installed in a stereo way to measure the z coordinate. The thickness of the sensors is 300 μm for the three innermost rings and 500 μm for the others.

The alignment of the tracker modules is very important to obtain a high spatial resolution. Deviations are caused by assembly inaccuracy, deformations due to cooling and stress from the magnetic field. Therefore, three methods are used for the tracker alignment. The geometry was determined during the assembly to an accuracy of 80 to 150 μm. An infrared laser system is used for continuous monitoring of the position of selected tracker modules. The final alignment is done with tracks from well known physics processes, e.g.

[4]The *pitch* is defined as the distance between two consecutive strips.



(a) Projection of an ECAL sector on the $r - z$ plane.

(b) ECAL 3D view.

Figure 2.9: Schematic representation of the electromagnetic calorimeter ECAL.

cosmic muons, or di-muon pairs from J/Ψ, Y and Z particles. The p_T of the tracks is reconstructed with a resolution of 1% at 100 GeV.

2.2.3 The Electromagnetic Calorimeter (ECAL)

The main function of an electromagnetic calorimeter is to identify electrons and photons and to measure accurately their energy. The electromagnetic calorimeter (Fig. 2.9) of CMS (ECAL, *Electromagnetic CALorimeter*) [63, 64] is a homogeneous calorimeter with cylindrical geometry, whose elements are scintillating crystals of lead tungstate ($PbWO_4$) with a truncated pyramidal shape. It consists of an *ECAL Barrel* (EB) with 61200 crystals and two *ECAL Endcaps* (EE) containing 7324 crystals each one.
Crystals are grouped into 5×5 matrices called *towers*. The barrel has an inner radius of 129 cm, a length of 630 cm and it extends in the region $|\eta| < 1.479$. Crystals in the ECAL barrel have the following dimensions: 22×22 mm^2 at the front face, 26×26 mm^2 at the rear face, and a length of 23 cm, corresponding to 25.8 X_0. Each submodule, consisting in a 5×2 crystals arrays mounted on a glass fiber structure, forms the elementar unit of EB. The granularity of a single crystal is about 1°. To avoid that cracks might align with particle trajectories, the crystal axes are tilted by 3° with respect to the direction from the interaction point, both in the η and φ direction.
Each endcap covers the region $1.479 < |\eta| < 3$ and is formed by two semicircular halves of aluminum called *dees*. Crystals in endcaps have a length of 22 cm and frontal area equal to 28.6×28.6 mm^2. They are arranged in *supercrystals* with 5×5 elementary unity. Unlike the crystals in the barrel, arranged in a $\eta - \varphi$ symmetry, the endcap crystals are arranged according to a $x - y$ geometry.
Two *preshower* detectors are placed in front of the endcaps in order to separate the showers produced by a primary γ from those produced by a primary π^0. This detector, which covers the region $1.653 < |\eta| < 2.6$, is a sampling calorimeter and it consists of two disks of lead converters (2 X_0 and 1 X_0 thick respectively) that start the electromagnetic shower of the incident photon/electron, alternating with two layers of silicon microstrip detectors in which a measurement of the released energy and the identification of the

shower profile are performed. The strips are arranged orthogonally in the two planes, according to a $x - y$ configuration.

The choice of the $PbWO_4$ crystals as scintillating material for ECAL is due to several reasons. First, the high-density ($\rho = 8.3$ g/cm^3), the short radiation length ($X_0 = 0.89$ cm) and the reduced *Molière radius*[5] ($R_M = 2.2$ cm) allow to build a compact and high granularity calorimeter. Furthermore, the 15 ns decay scintillation time allows to collect about 80% of the emitted light during the 25 ns that exist between two consecutive beam interactions in the LHC design performance. Finally, the $PbWO_4$ crystals have a good intrinsic radiation hardness and therefore they can operate for years in the hard LHC environment, with a modest deterioration in performance. The main disadvantage of these crystals is the low light yield (~ 10 photoelectrons/MeV) which makes an internal amplification for the photodetectors necessary. This is achieved through the use of silicon avalanche photodiodes (APD [65]) in barrel and single stage photomultipliers (*Vacuum Photo-Triode*, VPT [66]) in the endcaps, both resistant to the radiation and to the strong magnetic field of CMS.

The energy resolution of a homogeneous calorimeter is usually expressed by the sum in quadrature of three terms, according to the formula

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \ . \tag{2.4}$$

The *stochastic term* a is the dominant term at low energies: it includes the contribution of statistical fluctuations in the number of photoelectrons generated and collected. The fluctuations are Poissonian and this term takes into account the crystal light emission, the light collection efficiency and the *quantum efficiency*[6] of the photodetector.
The *noise term* b includes contributions from the electronic noise, both due to the photodetector and to the preamplifier, and from pileup events. The contributions vary at different pseudorapidities and with the operational luminosity of the LHC.
The *constant term* c, dominant at high energies, takes into account several contributions: the stability of the operating conditions (in particular of temperature and voltage), the presence of dead material in front of the crystals and the rear leakage of the electromagnetic shower, the longitudinal non uniformity of the crystal light yield, the intercalibration errors and the radiation damage of the crystals.
The ECAL barrel energy resolution for electrons in beam tests has been measured to be [67]:

$$\frac{\sigma_E}{E} = \frac{2.8\% \text{ GeV}^{1/2}}{\sqrt{E}} \oplus \frac{12\% \text{ GeV}}{E} \oplus 0.3\% \ , \tag{2.5}$$

where the energy E is measured in GeV.

[5]The *Molière radius* R_M gives an estimate of the transverse development of an electromagnetic shower in a calorimeter: on average, 90% of the energy released by the electromagnetic shower lies in a cylinder with radius R_M.

[6]The *quantum efficiency* is the ratio between the number of collected electron-hole pairs and the number of photons incident on the photodetector.

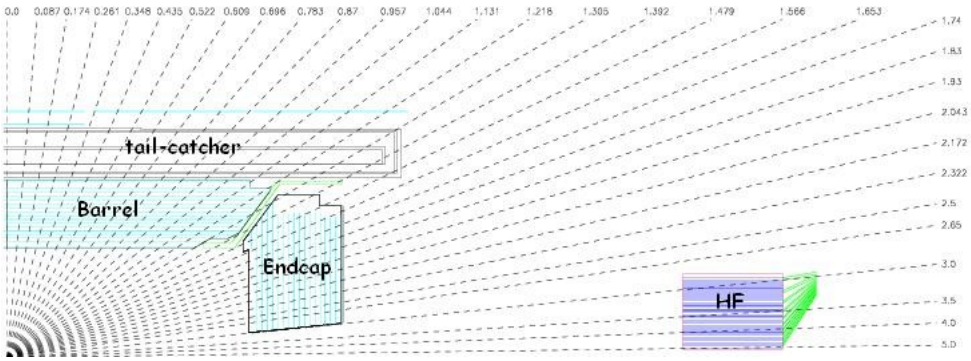


Figure 2.10: A schematic $r - z$ view of a quadrant of the CMS hadronic calorimeter HCAL.

2.2.4 The Hadronic Calorimeter (HCAL)

The hadronic calorimeter HCAL (*Hadronic CALorimeter*) [68], together with the electromagnetic calorimeter, makes a complete calorimetric system for the jet energy and direction measurement. Furthermore, thanks to its tightness, it can provide a measurement of the features of non-interacting particles, such as neutrinos, by measuring the missing energy deposited in the transverse plane E_T^{miss}. The CMS hadronic calorimeter is a hermetic sampling calorimeter that covers the region $|\eta| < 5$. As shown in Figure 2.10, it is divided into four subdetectors: HB (*Barrel Hadronic Calorimeter*), located in the barrel region inside the magnet, extending up to pseudorapidities $|\eta| < 1.4$; HE (*Endcap Hadronic Calorimeter*), situated in the endcaps region inside the magnet, extends in the pseudorapidity region $1.3 < |\eta| < 3$, partially overlapping the HB coverage; HO (*Outer Hadronic Calorimeter*, also called *tail-catcher*, placed along the inner wall of the magnetic field return yoke, just outside of the magnet; HF (*Forward Hadronic Calorimeter*), a sampling calorimeter consisting of quartz fibers sandwiched between iron absorbers, consisting of two units placed in the very forward region ($3 < |\eta| < 5$) outside the magnetic coil. The quartz fibers emit Cherenkov light with the passage of charged particles and this light is detected by radiation resistant photomultipliers.

In order to maximize particle containment for a precise missing transverse energy measurement, the amount of absorber material was maximized, reducing therefore the amount of the active material. Since HCAL is mostly placed inside the magnetic coil, a non-magnetic material like brass was chosen as absorber. HB and HE are therefore made with brass absorber layers interleaved with plastic scintillators (*wavelength shifters*, WLS) coupled to transparent optical fibers, which transmit the light to the HPD (*Hybrid Photodiodes*) photodetectors. HCAL granularity is $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$.

The energy resolution in the different geometrical regions of HCAL may be parametrized as follows, by using one stochastic term and one constant term:

$$barrel/endcaps: \ \frac{\sigma_E}{E} = \frac{90\% \ \mathrm{GeV}^{1/2}}{\sqrt{E}} \oplus 4.5\% \ , \tag{2.6}$$

$$forward: \ \frac{\sigma_E}{E} = \frac{172\% \ \mathrm{GeV}^{1/2}}{\sqrt{E}} \oplus 9\% \ , \tag{2.7}$$

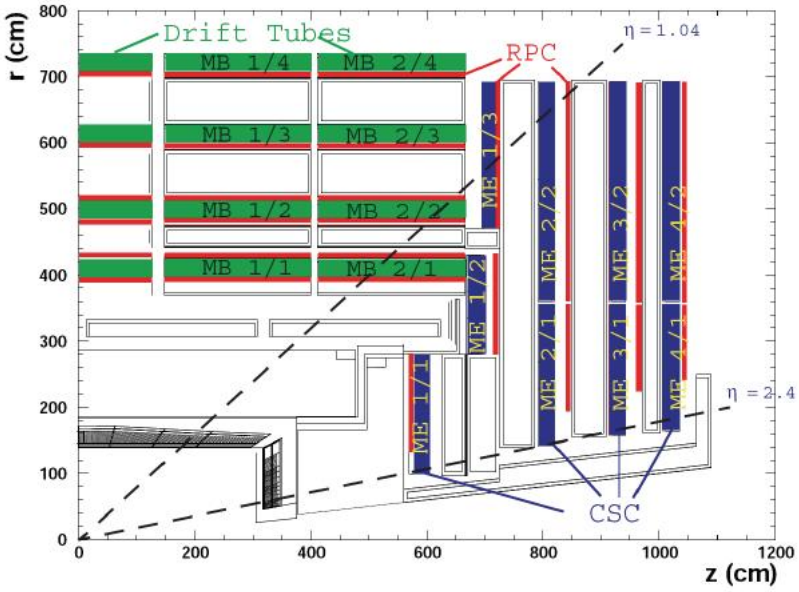


Figure 2.11: A schematic view of a quadrant of the CMS muon system.

where the energy E is expressed in GeV.

2.2.5 The Muon System

The CMS Muon System [69] is dedicated to the identification and measurement of high p_T muons, in combination with the tracker. Furthermore, it provides a time measurement of the bunch-crossing and also works as trigger for events involving muons. Momentum measurement, in the muon system, is determined by the muon bending angle at the exit of the 3.8 T coil, considering the interaction point as the origin of the muon. Up to p_T values of 200 GeV, the resolution of the muon system is dominated by multiple scattering and the best resolution is rather given by the silicon tracker. The system is placed outside the magnetic coil, embedded in the return yoke, to fully exploit the 1.8 T return flux. It consists of three independent subsystems, as shown in Fig. 2.11:

- *Drift Tubes* (DT) are placed in the barrel region, where the occupancy is relatively low (< 10 Hz/m^2);
- *Cathode Strip Chambers* (CSC) are in the endcaps, where the occupancy is higher (> 100 Hz/m^2);
- *Resistive Plate Chambers* (RPC) are both in the barrel and in the endcaps.

The Drift Tube system is placed in the barrel region $|\eta| < 1.2$, where the particle flux is low enough (~ 10 Hz/cm^2) and the magnetic field is sufficiently weak and homogeneous (1.8 T).
Along the longitudinal direction, the muon system barrel is divided into 5 wheels which are subdivided into 12 sectors each covering a 30° azimuthal angle.
The device is organized into 4 stations (MB1, MB2, MB3 and MB4) interspersed with

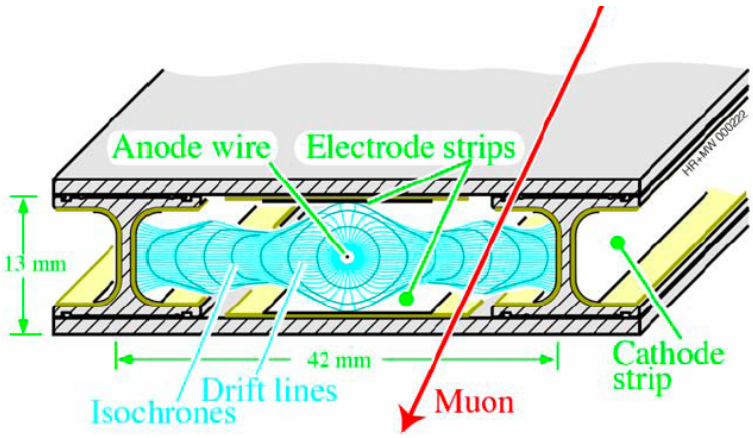


Figure 2.12: Schematic view of a drift cell. The electrons, coming from the gas ionization due to the passage of a muon, drift towards the wire anode. The voltages applied to the electrodes are +3600 V for wires, +1800 V for strips, and -1200 V for cathodes.

layers of the flux return plates. The stations form concentric cylinders around the beam line; the three inner cylinders have 60 drift chambers each and the outer cylinder has 70. Chambers are filled with a gas mixture of Ar (85%) and CO_2 (15%).
A DT chamber is made of 3 (or 2, in the fourth station) superlayers (SL), each made of 4 layers of rectangular drift cells staggered by half a cell. The SL is the smallest independent unit of the design. The wires in the first and third SLs are parallel to the beam line and provide a track measurement in the magnetic bending plane ($r - \varphi$). In the middle (the second) SL, the wires are orthogonal to the beam line and measure the z position along the beam. This third, z-measuring, SL is not present in the fourth station, which therefore measures only the φ coordinate. The efficiency of single chamber lies around 99.8% with a spatial resolution of ~ 100 μm both in $r - \varphi$ and z directions.
The drift cells (Fig. 2.12) consist of a stainless steel anode wire placed between two parallel aluminum layers.

In the endcaps, the high and non-uniform magnetic field (up to 3.5 T) and the particle rate expected in the muon system do not allow to use drift tubes detectors to perform measurements. Therefore, a solution based on Cathode Strip Chamber detector has been adopted. CSC are multiwire proportional chambers filled with a gas mixture of Ar (40%), CO_2 (50%), CF_4 (10%) in which the plane cathode is segmented into stripes oriented radially and transversely with respect to the anode wires (see Fig. 2.13). This allows the simultaneous measurement of two coordinates (r through the wires, φ using the strips). The drift path of the charge carriers is shorter with respect to the drift tubes, therefore these detectors can be placed in regions with higher flows of charged particles and less homogeneous magnetic fields. The strip spatial resolution is about $80 - 85$ μm while r can be determined with a precision of 0.5 cm. The CSC coverage is $0.8 < |\eta| < 2.4$.
Resistive Plates Chambers are used both in barrel and endcaps, complementing DT and CSC systems, in order to ensure robustness and redundancy to the muon spectrometer. RPCs are gaseous detectors characterized by a coarse spatial resolution, but are able to perform precise time measurements, comparable with the ones provided by scintillators.

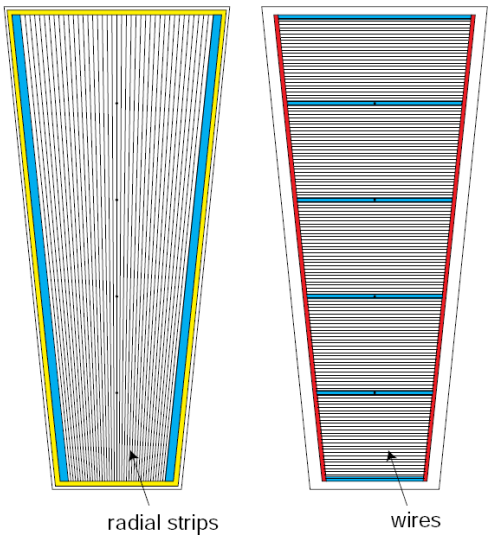


Figure 2.13: A schematic representation of CSC cathode panel (left) and anode panel (right).

This ensures precise bunch crossing identification to the muon trigger system.
CMS uses double-gap RPC chambers composed of 4 bakelite planes alternated to form two gas gaps 2 mm thick. The planes are coated by graphite to make electrodes that are set at 9.5 kV of potential difference. The central part of the chamber is equipped with insulated aluminum strips, used to collect the signal generated by crossing particles. In the barrel the strips are rectangularly segmented and run along the beam axis, whereas the endcaps are equipped with trapezoidal shaped strips. The gap between the plates is filled with a mixture of $C_2H_2F_4$ (94.5%) and Isobutane. In order to sustain higher rates, the detector operates in avalanche mode. The RPC coverage is $|\eta| < 2.1$.

2.3 Trigger and Data Acquisition

LHC can produce interactions at 40 MHz frequency, but only a small fraction of these events can be written on disk. On one hand the speed at which data can be written to mass storage is limited, on the other hand the vast majority of events produced is not interesting, because it involves low transverse momentum interactions (minimum bias events). Thus, a trigger system is needed to save interesting events at the highest possible rate. The maximum rate of events written on disk is about 800 Hz. CMS has chosen a two-level trigger system, consisting of a *Level-1* Trigger (L1) [70] and a *High Level Trigger* (HLT) [71].
Level-1 trigger runs on dedicated processors, and accesses coarse level granularity information from calorimetry and muon system. A L1 Trigger decision has to be taken for each bunch crossing within 3.2 μs. Its task is to reduce the data flow from 40 MHz to about 100 kHz.
The High Level Trigger is responsible for reducing the L1 output rate down to a maximum rate of the order of 1 kHz. The HLT code runs on a farm of commercial processors and

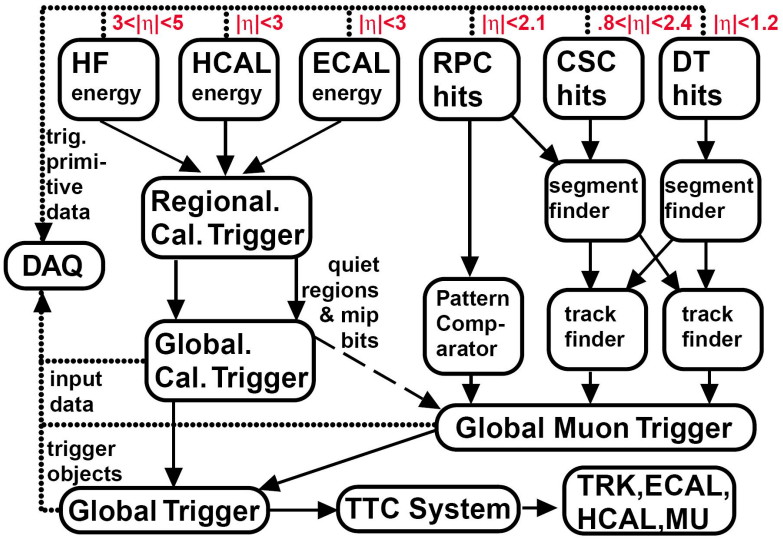


Figure 2.14: Level-1 Trigger components.

can access the full granularity information of all the subdetectors.
The main characteristics of the CMS trigger system will be described in the following.

2.3.1 Level-1 Trigger

The Level-1 Trigger is responsible for the identification of electrons, muons, photons, jets and missing transverse energy. It is required to have high efficiency. Its output rate and speed are limited by the readout electronics and by the performances of the data acquisition (DAQ) [71] system.
It consists of three main subsystems:

- L1 Calorimeter Trigger;
- L1 Muon Trigger;
- L1 Global Trigger.

The L1 Global Trigger is responsible for combining the output of L1 Calorimeter Trigger and L1 Muon Trigger and for making the decision whether to retain the event or not. The organization of CMS Level-1 Trigger is schematically summarized in Figure 2.14.

2.3.1.1 L1 Calorimeter Trigger

The input for L1 Calorimeter Trigger are the calorimeter towers, which are clusters of signals collected both from ECAL and HCAL (including HF). Towers are calculated by calorimeter high level readout circuits, called Trigger Primitive Generators. The Regional Calorimeter Trigger finds out electron, photon, τ and jet candidates along with their transverse energy and sends them to the Global Calorimeter Trigger. The Global Calorimeter Trigger sorts the candidates according to their transverse energy and sends the first four objects to the L1 Global Trigger.

2.3.1.2 L1 Muon Trigger

L1 Muon Trigger is actually a composed system itself: information from RPC, CSC and DT specific triggers are combined in the so called L1 Global Muon Trigger.
The RPC trigger electronics builds Track Segments, gives an estimate of the p_T and sends these segments to the Global Muon Trigger.
It also provides the CSC logic unit with information to solve hit position ambiguities in case two or more muon tracks cross the same CSC chamber.
The CSC trigger builds Local Charged Tracks (LCT), that is track segments made out of the cathode strips only. A p_T value and a quality flag are assigned to the LCTs. The best three LCTs in each sector of nine CSC chambers are passed to the CSC Track Finder, that uses the full CSC information to build tracks, assigns them a p_T and a quality flag and sends them to the Global Muon Trigger.
DTs are equipped with Track Identifier electronics, which is able to find groups of aligned hits in the four chambers of a superlayer. Those Track Segments are sent to the DT Track Correlator that tries to combine segments from two superlayers, measuring the φ coordinate. The best two segments are sent to the DT Track Finder that builds tracks and sends them to the Global Muon Trigger.
The Global Muon Trigger sorts the RPC, CSC and DT muon tracks and tries to combine them. The final set of muons is sorted according to the quality, and the best four tracks are passed to the L1 Global Trigger.

2.3.1.3 L1 Global Trigger

The L1 Global Trigger is responsible for collecting objects created from the Calorimeter and Muon Triggers and for making a decision whether to retain the event or not. If the event is accepted, the decision in sent to the Timing Trigger and Control System, that commands the readout of the remaining subsystems.
In order to make the decision, the L1 Global Trigger sorts the ranked objects produced by calorimetry and muon system and checks if at least one of the thresholds in the L1 trigger table is passed.

2.3.2 High Level Trigger

The High Level Trigger (see diagram in Fig. 2.15) is designed to reduce the L1 output rate to about 800 events/s[7] that is going to be written to mass storage. Events passing the HLT are stored on local disk or in CMS Tier-0[8].
The HLT code runs on commercial processors and performs reconstruction using the

[7] In 2012, the output rate for offline data storage on local disk or CMS Tier-0 was, on average, of about 400 Hz for prompt reconstruction within 48 hours ("core"), and of about 400 Hz for a later reconstruction which was performed in 2013 ("*parking*").

[8] The Worldwide LHC Computing Grid (WLCG) is composed of four levels, or "Tiers", called 0, 1, 2 and 3. Each Tier is made up of several computer centres and provides a specific set of services; they process, store and analyse all the data from the Large Hadron Collider (LHC). Tier 0 is the CERN Data Centre. All of the data from the LHC pass through this central hub. Tier 0 distributes the raw data and the reconstructed output to Tier 1's, and reprocesses data when the LHC is not running.

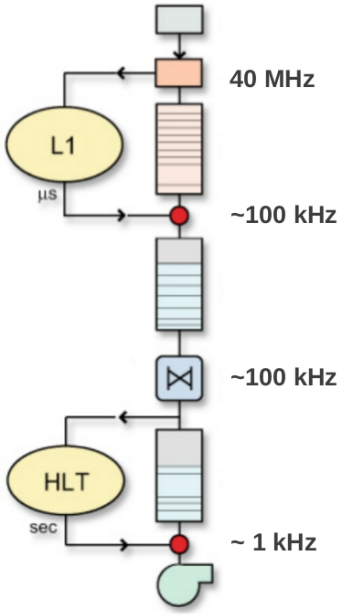


Figure 2.15: High Level Trigger diagram.

information from all subdetectors. Data read from subdetectors are assembled by a builder unit and then assigned to a switching network that dispatches events to the processor farm. The CMS switching network has a bandwidth of 1 Tbit/s.
This simple design ensures maximum flexibility to the system, the only limitation being the total bandwidth and the number of processors. The system can be easily upgraded adding new processors or replacing the existing ones with faster ones as they become available. Since the algorithm implementation is fully software, improvements to the algorithms can be easily implemented and do not require any hardware intervention.
Event by event, the HLT code is run on a single processor, and the time available to make a decision is about 200 ms (at an input rate of 100 kHz). It is important to note that HLT algorithms use the same software framework and most of the same reconstruction code used for offline reconstruction and analyses. However, the real-time nature of this selection imposes several constraints on the resources the algorithms can use. The reliability of HLT algorithms is of capital importance, because events not selected by the HLT are lost. HLT algorithms must also be flexible to adapt to changes in data-taking conditions, like changes in luminosity or special conditions occurring during the CMS commissioning or dedicated LHC fills. Its performance should be robust with respect to changes in alignment and calibration constants and stable with respect to pileup (PU).
Moreover, it must be able to provide on-line detector monitoring by means of specific trigger paths for calibration and alignment.

In order to efficiently process events, the HLT code has to be able to reject not interesting events as soon as possible; computationally expensive algorithms must be run only on good candidates for interesting events. In order to cope with this requirement the HLT code is organized in a virtually layered structure:

- Level 2: uses only complete muon and calorimetry information;

- Level 2.5: uses also the pixel information;
- Level 3: makes use of the full information from all the tracking detectors.

Each step reduces the number of events to be processed in the following step. The most computationally expensive tasks are executed in the Level 3; time consuming algorithms such as track reconstruction are only executed in the region of interest. Besides, since the ultimate precision is not required at a HLT level, track reconstruction is performed on a limited set of hits, and is stopped once the required resolution is achieved (see Appendix C).

Chapter 3

Event reconstruction and particle identification

In the analysis I am going to discuss in Chap. 4 and 5, two same-flavour opposite-sign leptons (specifically, electrons or muons) and four jets (two from Z and two from the VBF production mechanism) are searched for to select signal events.
In CMS, the object reconstruction follows some standard algorithms which have been developed by the collaboration to ensure a high quality and standardized performance. In this section, the reconstruction techniques for the objects of our interest are described.

3.1 The Particle Flow technique

The Particle Flow (PF) event reconstruction algorithm [72] aims at reconstructing all stable particles in a given pp collision - muons, electrons, photons, charged hadrons and neutral hadrons - combining information from all CMS sub-detectors to determine their direction, energy and type. The design of the CMS detector is well suited for the Particle Flow technique. The presence of an internal large silicon tracker immersed in an uniform 3.8 T magnetic field allows the reconstruction of charged particles with high efficiency and provides a precise measurement of the transverse momentum down to about 150 MeV. Together with the excellent tracking system and the large magnetic field, the high granularity of the ECAL is a key element for the PF technique, allowing the reconstruction of photons and electrons with high energy resolution.
Firstly the PF algorithm proceeds with the identification of fundamental elements as reconstructed in the sub-detectors: charged-particles, calorimeter clusters and muon tracks. These elements are then connected to each other by making use of link algorithms identifying *blocks* of elements which are topologically compatible (e.g. a charged-particle track is linked to a calorimeter cluster if the extrapolated position from the track to the calorimeter is within the cluster boundaries). From the blocks, particle flow particle candidates (PF Candidates) are fully reconstructed and identified in the following order and according to the described procedures:

- Muons: a *global muon*, reconstructed from the combination of a track in the tracker and a track in the muon system, gives rise to a *PF muon*. After the identification,

the corresponding track is removed from the block;

- Electrons: the link between a charged-particle track (refitted with the Gaussian-Sum Filter [73, 74]) and one or more ECAL clusters identifies *PF electrons*. The corresponding track and ECAL clusters are removed from further processing;
- Charged hadrons: the remaining tracks give rise to *PF charged hadrons*. Tracks can be linked to ECAL and HCAL clusters, and the energy is determined taking into account information from calorimeters;
- Photons and Neutral hadrons: ECAL clusters not compatible with charged-tracks give rise to *PF photons*, while unaccounted HCAL deposits are interpreted as *PF neutral hadrons*.

Once the list of PF Candidates is defined, *PF Jets* are clustered as described in Sec. 3.3.2. *PF Missing Transverse Energy* (PF MET) is reconstructed at the end of event reconstruction: it consists in forming the transverse momentum-vector sum over all reconstructed PF Candidates in the event and then taking the opposite of this vectorial sum. The missing transverse energy is the modulus of this vector.

3.2 Lepton reconstruction and identification

3.2.1 Muon reconstruction and identification

Muons cross the detector and interact with the silicon tracker through ionization. Their loss of energy is negligible, so they manage to reach the outer part of the CMS detector, where the muon chambers are located (see Chap. 2.2). Muons are thus reconstructed both in the silicon tracker and in the external muon chambers. Their reconstruction [75] begins from the muon spectrometer, collecting position measurements from DT, CSC and RPC sub-detectors. Hits in DT and CSC chambers are matched to form segments, which are then collected and matched to generate seeds used as a starting point for the track fit of DT, CSC and RPC hits. The reconstructed track in the muon spectrometer is called *Stand-alone Muon*.
Muon track reconstruction is performed in the inner silicon tracker as well. Seeds are built using two or three consecutive hits in the pixel and/or in the strip detector. The *pattern recognition* is performed starting from these seeds and proceeding layer by layer, with an iterative technique based on the Kalman Filter technique [76]. At the end of this algorithm, a track fit is performed and the track parameters are updated. The identified tracker track is then combined with a given *Stand-alone Muon* track in order to construct a global track, which defines a *Global Muon*. A global fit is performed for each pair of tracks reconstructed in the inner tracker and in the muon system. If more than one track matching the stand-alone track is found, then the one giving the best χ^2 in the global fit is chosen.
In the analysis, only muons reconstructed as *Global Muons* are considered. Moreover, they are required to be recognized as muons also by the Particle Flow (PF) algorithm, to ensure a good quality of the reconstruction. Further quality requirements are applied:

Observable	Cut
Is global muon	true
Is PF muon	true
Tracker layers with measurements	> 5
Number of valid pixel hits	> 0
Number of valid muon hits	> 0
Number of matched muon stations	> 1
$\chi^2/ndof$	< 10
$d_{xy}(PV)$	< 0.2 cm
$d_z(PV)$	< 0.5 cm

Table 3.1: Muon identification and isolation requirements.

- the tracker track has to be reconstructed from at least 5 tracker layers with hits;
- at least one hit must be present in the pixel detector;
- at least one hit must be present in the muon detector;
- at least one muon chamber hit should be included in the *Global Muon* track fit;
- the normalized χ^2 of the *Global Muon* track fit should be less than 10;
- the muon track reconstructed in the tracker must have a distance to the primary vertex (PV) smaller than 2 mm in the transverse plane and smaller than 5 mm in the longitudinal direction.

These requirements are summarized in Tab. 3.1.

3.2.1.1 Muon isolation

In the analysis, muons in signal events are expected to be isolated, because they origin from the decay of a Z boson. A source of background are, instead, muons produced in jets. It is possible to distinguish between muons produced in jets and muons coming from heavy particle decays (Z or W) selecting isolated muons, i.e. muons that do not have nearby particles, and whose energy deposit in its vicinity is below a certain threshold. Thus, muons we are going to use to build the Higgs candidate are requested to pass an *isolation* criterion which also includes a PU mitigation correction called "$\Delta\beta$ correction". In fact, pileup (PU) causes a mean energy deposited in the detector which is not due to particles of the event and that contaminates the energy deposition measured in a given cone. As the instantaneous luminosity rises and the pileup increases, the mean value of the mean deposited energy increases as well. The isolation variable is therefore sensitive to the pileup and a correction is needed in order to ensure its independence and robustness on the number of simultaneous interactions. The isolation variable used (with $\Delta\beta$ correction) is defined as follows:

$$RelIso_{\Delta\beta} = [\sum_{ChH}(p_T) + max(0, \sum_{NH}(p_T) + \sum_{Ph}(p_T) - 0.5\sum_{ChHPU}(p_T))]/p_T^{muon} \qquad (3.1)$$

The sums in Eq. 3.1 are performed in a cone of 0.4 units in ΔR around the muon. ChH means charged hadrons, NH is neutral hadrons, Ph is photons, $ChHPU$ is charged hadrons not coming from the primary vertex. $RelIso_{\Delta\beta}$ on selected muons is requested to be less than 0.12.

3.2.1.2 Muon scale corrections

The measurement of muon momentum is affected by our reconstruction capability and our limited knowledge of the physical configuration of the detector [77]. We want to correct these effects and estimate the remaining systematics after these corrections. The strategy relies on exploiting well known resonances (J/ψ, Y, Z) to correct the muon momentum scale:

- discover all the possible biases and understand their source;
- compute the bias with a multivariate likelihood approach;
- release a function of the muon kinematic variables to correct the muon momentum scale.

By product, we obtain the muon resolution as a function of the muon kinematic variables. For this purpose, the "*MuScleFit*" package is used. For the scale corrections, *MuScleFit* provides momentum scale corrections for muons, which are meant to recover bias in the mass of the Z with respect to the η, ϕ of the muon and the $\Delta\eta$ of the two muons.
Moreover, *MuScleFit* corrections can recover the global scale shift observed between data and MC in the fitted peak of the di-muon mass ($\sim 0.14\%$ both in 2011 and 2012 samples). Muon momentum corrections are based on an *ansatz* function which corrects the curvature of the muon taking into account alignment effects, like shifts and deformations in the transverse plane, "twists" and magnetic field effects. It includes:

- an overall multiplicative scale term;
- harmonic terms in 5 η bins;
- η-linear terms in 5 η bins;
- an overall shift term for charge-dependent bias.

After applying corrections, there is still a residual mismatch in the resolution between data and MC. *MuScleFit* correction package provides a recipe to add an extra smearing for 2012 and 2011 MC samples, based on the difference between the fitted resolution in data and MC.

3.2.2 Electron reconstruction and identification

Electron reconstruction [78] is based on the combination of ECAL and tracker information. Electrons traversing the silicon layers of the tracker radiate Bremsstrahlung photons and the energy reaches the ECAL with a significant spread in the azimuthal direction ϕ.

The reconstruction algorithm starts by measuring the energy deposited in ECAL as one or more associated clusters. The ECAL clustering, and in particular the building of "superclusters" (clusters of clusters), is designed to take into account the spread and collect the Bremsstrahlung energy. After superclusters reconstruction, the algorithm proceeds matching them to track seeds identified as pairs or triplets of hits in the inner tracker layers. The Bremsstrahlung emission introduces, in general, non-Gaussian contributions to the event-by-event fluctuations of the calorimetry and tracking measurements. Electron trajectories are thus reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian Sum Filter (GSF) [73]. In order to solve ambiguous cases where several tracks are reconstructed due to the conversion of radiated photons in the tracker material, a preselection is applied. This preselection is based on the matching between the GSF track and the supercluster in η and ϕ.
Together with "ECAL driven" reconstruction, in CMS, electron reconstruction is also performed in a "tracker driven" approach, seeded by electron tracks reconstructed in the inner tracker layers. A preselection is applied on these tracks, based on a multivariate analysis. More detailed information about the electron reconstruction and results from the commissioning with the first collision data can be found in [79].
In order to ensure a good electron reconstruction, the $|\eta|$ of the electron supercluster must be inside the ECAL acceptance volume, $|\eta| < 2.5$, and outside the ECAL barrel-endcap overlap region, $1.4442 < |\eta| < 1.566$.

The electron identification variables found to be the most discriminating are:

- $\Delta\eta_{trk,SC}$ and $\Delta\phi_{trk,SC}$: the measurements of the spacial matching between the track and the supercluster in η and ϕ respectively;
- $\sigma_{i\eta,i\eta}$: a variable related to the calorimeter shower shape, measuring the width of the ECAL supercluster along the η direction computed for all the crystals in the 5×5 block of crystals centred on the highest energy crystal of the seed supercluster;
- H/E: the ratio between the energy deposit recorded in the HCAL tower behind the electromagnetic seed supercluster and the energy of the seed supercluster.

The rate of high-energy photons converting to electrons ($\gamma \rightarrow e^+e^-$) and crossing the material in front of the electromagnetic calorimeter is high and, in order to avoid selecting electrons from photon conversion, a dedicated set of cuts is used:

- $|1/E - 1/p|$, the absolute value of the difference of the inverse of energy E measured in the ECAL and inverse of momentum p measured in the tracker;
- the number of missing hits in the back-propagation of the track to the beam line;
- d_{xy} and d_z, which are the transverse and longitudinal impact parameters computed with respect to the reconstructed primary vertex (PV);
- a conversion veto based on a full vertex fit, including three dimensional collinearity constraint in the fit.

In our analysis, electrons are requested to satisfy *loose* identification criteria which are listed in Tab. 3.2.

Observable	Barrel cut	Endcap cut
$\Delta\eta_{SC,trk}$	< 0.007	< 0.009
$\Delta\phi_{SC,trk}$	< 0.150	< 0.100
$\sigma_{i\eta i\eta}$	< 0.010	0.030
H/E	< 0.120	< 0.100
$\lvert 1/E - 1/p \rvert$	< 0.050	< 0.050
Missing hits	< 2	< 2
$d_{xy}(PV)$	< 0.02cm	< 0.02cm
$d_z(PV)$	< 0.2cm	< 0.2cm
Conversion veto	true	true

Table 3.2: Electron ID requirements, for the *loose* working point.

3.2.2.1 Electron isolation

In the analysis described in the following, electrons are required to pass isolation criteria including a PU mitigation correction based on the electron effective catchment area, defined according to a jet-inspired effective area algorithm. Seven slices have been defined in electron $|\eta|$ and effective areas have been calculated for those slices, as reported in Tab. 3.3. Based on these values, a corrected relative isolation is computed for each electron as follows:

$$RelIso_{EA\ corrected} = [\sum_{ChH}(p_T) + max(0, \sum_{Ph}(p_T) + \sum_{NH}(p_T) - \rho EA)]/p_T^{electron} \qquad (3.2)$$

where ChH is charged hadrons, Ph is photons, NH is neutral hadrons, ρ is the PU energy density calculated with the FastJet algorithm [80, 81] for k_T jets with a radius parameter of 0.6. The isolation cone is 0.4 units in ΔR with a veto cone of 0.08 around the electron. In the analysis, $RelIso_{EA\ corrected}$ on selected electrons is requested to be less than 0.15.

Rapidity interval	Effective area
$0 < \lvert\eta\rvert < 1$	0.21
$1 < \lvert\eta\rvert < 1.479$	0.21
$1.479 < \lvert\eta\rvert < 2$	0.11
$2 < \lvert\eta\rvert < 2.2$	0.14
$2.2 < \lvert\eta\rvert < 2.3$	0.18
$2.3 < \lvert\eta\rvert < 2.4$	0.19
$2.4 < \lvert\eta\rvert < 999$	0.26

Table 3.3: Effective areas for electron isolation correction.

3.2.3 Lepton identification efficiency

Lepton identification efficiencies are evaluated using the *tag and probe* (T&P) technique [82]. This method uses a pure sample of $Z \to l^+l^-$ events and requires the reconstruction of a dilepton system compatible with a Z boson decay hypothesis and,

therefore, with invariant mass falling in the range [60, 120] GeV. One lepton candidate, called the "tag", satisfies tight identification and isolation requirements. The other lepton candidate, called the "probe", is selected with criteria that depend on the efficiency being measured. The selected dilepton systems are categorized in two exclusive samples according to whether the probe has passed or failed the selection criteria under investigation. Due to the presence of background, a fit is performed to the invariant mass distribution of the dilepton system to obtain the signal yields in the two categories. The efficiency is measured as the relative signal yield in subsamples with passing or failing probes. This procedure is performed separately in different bins of probe pseudo-rapidity (η) and transverse momentum (p_T). Thus, the efficiency is obtained as function of p_T and η of the probe lepton. Identification efficiency is not perfectly reproduced in simulation, therefore, data/simulation scale factors are deduced by dividing efficiencies in data to the ones obtained from simulation using exactly the same procedure described above. Simulated distributions are corrected by these scale factors.

3.3 Jet reconstruction

Jets are the experimental signature of quarks and gluons produced in high energy physics processes. They are the result of hadronization of the parent parton and play a predominant role in a hadronic collider like the LHC, where their production cross section is very large. A correct reconstruction and calibration of jets is crucial to perform an analysis with jets in the final state. In this section, the jets reconstruction adopted in CMS and in this analysis is described.

3.3.1 Jet reconstruction in CMS

In the analysis described, Particle Flow jets have been used. The PF jets are reconstructed by clustering the four-momentum vectors of particle-flow candidates, according to the particle-flow algorithm described in Sec. 3.1 which combines the information from all relevant CMS sub-detectors to identify and reconstruct all visible particles in the event, namely muons, electrons, photons, charged hadrons, and neutral hadrons. The PF jet momentum and spatial resolutions are greatly improved with respect to calorimeter jets, which are instead reconstructed using only the calorimeter information, as the use of the tracking detectors allows a better p_T resolution for the charged component. For what concerns simulation, the generated jets are reconstructed by clustering the four-momentum vectors of all stable (decay length$>$ 1 cm) particles generated in the simulation, except neutrinos.

3.3.2 Jet clustering algorithms

Jets are defined through sequential, iterative jet clustering algorithms that combine four-vectors of input pairs of particles until certain criteria are satisfied and jets are formed [83]. Different algorithms are available for jet clustering. From a theoretical point of view, an ideal jet clustering algorithm should have these features [84]:

- *infrared safety*: any infrared singularity should not appear in the perturbative calculation; furthermore, it should also have solutions which are insensitive to soft radiation in the event;
- *collinear safety*: collinear singularities must not appear in the perturbative calculations; furthermore, it should also find jets that are insensitive to any collinear radiation in the event;
- *invariance under boosts*: the algorithm should find the same solutions independently by boosts in the longitudinal direction. This is particularly important for pp collisions where the center-of- mass of the individual parton-parton collisions is typically boosted with respect to the *pp* center-of-mass;
- *order independence*: the algorithm should find the same jets at parton, particle, and detector level;
- *straightforward implementation*: the algorithm should be straightforward to implement in perturbative calculations.

From an experimental point of view, a jet algorithm should follow these criteria [84]:

- *detector independence*: the performance of the algorithm should be as independent as possible of the detector that provides the data. For example, the algorithm should not be strongly dependent on detector segmentation, energy response, or resolution;
- *minimization of resolution smearing and angle biases*: the algorithm should not amplify the inevitable effects of resolution smearing and angle biases;
- *stability with luminosity*: jet finding should not be strongly affected by pileup at high beam luminosities;
- *efficient use of computing resources*: the jet algorithm should provide jet identification with a minimum of computer time consumption;
- *maximal reconstruction efficiency*: the jet algorithm should efficiently identify all physically interesting jets;
- *ease of calibration*: the algorithm should not obstruct to the reliable calibration of the final kinematic properties of the jet;
- *ease of use*: the algorithm should be straightforward to implement with typical experimental detectors and data;
- *fully specified*: the algorithm must include specifications for clustering, energy and angle definition, and all details of jet splitting and merging.

Since a parton is not a well-defined object, a jet definition is not unique and several approaches are therefore available for jet clustering, each of them with different characteristics in order to satisfy the previous requirements. Mainly, two broad classes of jet

clustering algorithms have been developed. The first one consists of the *conical recombination* [84], where jets are defined as dominant directions of energy flow. One introduces the concept of *stable cone* as a circle of fixed radius R in the $\eta - \varphi$ plane such that the sum of all the momenta of the particles within the cone points to the same direction as the center of the circle. Cone algorithms attempt to identify all the stable cones. Most implementations use a seeded approach: starting from one seed for the center of the cone, one iterates until the cone is found to be stable. The set of seeds can be taken as the set of initial particles (sometimes over a p_T threshold) or as the midpoints between previously-found stable cones.
The second class is called *sequential recombination* and it works by defining a distance between pairs of particles, performing subsequent recombinations of the pair of closest particles and stopping when all resulting objects are too far apart. Algorithms within that class differ in the definition of the distance.
The main requirement for a jet clustering algorithm is that it is *InfraRed* and *Collinear safe* (IRC-safe), i.e. the algorithm must be independent of infrared and collinear corrections in perturbative QCD. The infrared (collinear) corrections concern the emission of a gluon with an infinitely low energy (angle), a phenomenon which happens with infinite probability in perturbative QCD. If the algorithm is IRC-safe, this soft (low-angle) emission must not affect the number of reconstructed jets. An IRC-safe algorithm presents also the advantage of being less sensitive to the calorimeter noise (small energy deposits do not affect the number of reconstructed jets). The standard IRC-safe algorithms adopted by CMS are the SISCone [85–87] in the conical recombination class, and the k_t, *Cambridge-Aachen* (CA) and *anti-*k_t [88] algorithms in the sequential recombination class. In the following, the k_t, the anti-k_t and the CA algorithms are described. For the analysis presented here, we used the anti-k_t and the CA algorithms.

3.3.2.1 The k_t, anti-k_t and Cambridge-Aachen algorithms

The k_t, the Cambridge-Aachen and anti-k_t are IRC-safe algorithms which belong to the sequential recombination class. They introduce two definitions of distance: d_{ij}, the distance between the objects (particles or pseudojets) i and j, and d_{iB}, the distance between the object i and the beam. These distances are defined as follows:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \, \frac{\Delta_{ij}^2}{R^2} \, , \tag{3.3}$$

$$d_{iB} = k_{ti}^{2p} \, , \tag{3.4}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\varphi_i - \varphi_j)^2$ and k_{ti}, y_i and φ_i are, respectively, the transverse momentum, rapidity and azimuth of particle i. In the two expressions, R is the radial parameter and p is a parameter which rules the momentum relative power versus the geometrical scale Δ_{ij}. The algorithm therefore proceeds as follows:

- the distances d_{ij} and d_{iB} are calculated, respectively for each pair of particles i, j and for each particle i, and the smallest quantity (which could be either of kind d_{ij} or d_{iB}) is identified;

- if the smallest quantity is of kind d_{ij}, the objects i and j are recombined summing the four-vectors;
- if instead the smallest quantity is of kind d_{iB}, i is considered as a jet and removed from the list of objects;
- the distances are recalculated and the procedure is repeated until no objects are left.

Changing the value of the p parameter in the distance definition, a different jet clustering algorithm can be obtained:

- the value $p = 1$ defines the k_t algorithm [89]. The general behaviour of the algorithm for $p > 0$ with respect to soft radiation is rather similar to the one observed for the k_t algorithm, because what matters is the ordering between particles, and for finite Δ_{ij} this is maintained for all positive values of p;
- the value $p = 0$ defines the inclusive Cambridge-Aachen algorithm [90];
- the value $p = -1$ finally defines the so-called anti-k_t algorithm [88].

Physically, the differences between the three algorithms are contained in the momentum weighting. For the k_t algorithm, the weighting $(\min(k_{ti}^2, k_{tj}^2))$ is done so as to preferentially merge constituents with low transverse momentum with respect to their nearest neighbours. For the anti-k_t algorithm, the weighting $(\min(1/k_{ti}^2, 1/k_{tj}^2))$ is done so as to preferentially merge constituents with high transverse momentum with respect to their nearest neighbours. The C-A algorithm relies only on distance weighting with no k_t weighting at all.

The functionality of the IRC-safe anti-k_t algorithm can be understood by considering an event with a few well separated hard particles with transverse energy k_{t1}, $k_{t2}, \cdots$ and many soft particles. The $d_{1i} = \min(1/k_{t1}^2, 1/k_{ti}^2)\ \frac{\Delta_{1i}^2}{R^2}$ distance between a hard particle 1 and a soft particle i is exclusively determined by the transverse momentum of the hard particle and by the Δ_{ij} separation. The d_{ij} between similar soft particles will instead be greater. Therefore the algorithm will tend to cluster soft particles with the hard ones instead of along themselves. If a hard particle has no hard neighbours within a distance $2R$, then it will simply accumulate all the soft particles within a circle of radius R, resulting in a perfectly conical jet. If another hard particle 2 is present such that $R < \Delta_{12} < 2R$, then there will be two hard jets and it will not be possible for both to be perfectly conical. If $k_{t1} \gg k_{t2}$, jet 1 will be conical, while jet 2 will be partly conical, since it will miss the part overlapping with jet 1. Instead, if $k_{t1} = k_{t2}$, neither jet will be conical and the overlapping part will simply be equally divided by a straight line between the two; in general, if $k_{t1} \sim k_{t2}$ both cones will be clipped. In the case where $\Delta_{12} < R$, the two hard particles will cluster to form a single jet, which will be conical and centered on k_1 if $k_{t1} \gg k_{t2}$, and will be instead more complex if $k_{t1} \sim k_{t2}$. The key feature above is that the soft particles do not modify the shape of the jet, while hard particles do. It results in a regular shape: anti-k_t generates a circular hard jet, which clips a lens-shaped region out of the soft one. The anti-k_t is therefore a robust algorithm as far as non-perturbative effects like

hadronization and underlying event (UE) contamination are concerned, improving in this way the momentum resolution and the calorimeter performance. Its regular shape also allows to subtract in an easy way the contribution of pileup, as explained below. Finally, another feature that justifies the choice of this algorithm is its operating speed. For all these reasons, the anti-kt has been chosen as the main clustering algorithm by the CMS experiment and it has been used in the present work to reconstruct the jets coming from the decay of the hadronic Z[1] and the VBF jets associated to the Higgs boson production. The distance parameter was fixed at the value $R = 0.5$.

3.3.3 The merged-jet topology

At large Higgs boson masses, the Z bosons have a large momentum, which results in the hadronic decay products being collimated to the point that a standard jet algorithm would reconstruct them as a single "merged" jet. Algorithmic methods use information from the jet clustering procedure to extract the internal structure of jets and are able to successfully distinguish between jets originating from highly boosted, hadronically decaying bosons (having a two-subjets structure) and the "QCD jets" originating from light quarks or gluons (having a single-subjet substructure) [91]. For this purpose, the ratio of the N-subjettiness jet shape observables τ_2/τ_1 has been found to be especially effective [92]. The N-subjettiness jet shape observable τ_N gives a measure of how a jet is compatible with having a N-subjets substructure, and is defined as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \, \min \{\Delta \mathrm{R}_{1,\mathrm{k}}, \Delta \mathrm{R}_{2,\mathrm{k}}, ..., \Delta \mathrm{R}_{\mathrm{N},\mathrm{k}}\} \tag{3.5}$$

where k runs over the constituent particles in a given jet, $p_{T,k}$ are their transverse momenta and $\Delta R_{J,k} = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$ is the distance in the rapidity-azimuth plane between a candidate subjet J and a constituent particle k. The normalization factor d_0 is taken as

$$d_0 = \sum_k p_{T,k} R_0 \tag{3.6}$$

where R_0 is the characteristic jet radius used in the original jet clustering algorithm. It is straightforward to see that τ_N quantifies how N-subjetty a particular jet is, or in other words, to what degree it can be regarded as a jet composed of N subjets. Jets with $\tau_N \simeq 0$ have all their radiation aligned with the candidate subjet directions and therefore have N (or fewer) subjets. Jets with $\tau_N >> 0$ have a large fraction of their energy distributed away from the candidate subjet directions and therefore have at least $N+1$ subjets. Less obvious is how best to use τ_N for identifying boosted Z bosons. Studies show [92] that the ratio τ_2/τ_1 has a very good discriminating power and that jets from Z have smaller τ_2/τ_1 values than QCD jets.

In addition, there are "jet grooming" techniques such as *filtering*, *pruning*, *trimming* and their combinations which aid in the identification of boosted objects by reducing the

[1]The anti-kt is used to reconstruct jets from Z when two distinct jets can be reconstructed; when, instead, the Z from a heavy Higgs is highly boosted and its decay jets are emitted at small angular distance, the CA algorithm is used, as better described in the following.

smearing effects of jet contamination from initial state radiation, underlying event activity and pileup. The use of *jet pruning* [93], for example, improves discrimination by pushing the QCD jets towards lower mass while maintaining the mass of the merged jets from Z.

3.3.4 Jet grooming techniques

Jet *grooming* techniques [83] that reduce the impact of contributions from the underlying event, pileup and low-p_T gluon radiation can be useful irrespective of the specific nature of the analysis. These kinds of contributions to jets are typically soft and diffuse, and hence contribute energy to the jet proportional to the area of the jets themselves. Because grooming techniques reduce the areas of jets without affecting the core components, the resulting jets are less sensitive to contributions from UE and PU, while still reflecting the kinematics of the hard original process. Three forms of grooming have been developed, referred to as *filtering*, *trimming* and *pruning*. We are going to consider the *pruning* technique, which is the one used in our analysis for the merged jets.

3.3.4.1 Jet pruning

A recombination algorithm naturally defines a substructure for the jet. The sequence of recombinations tells us how to construct the jet in step-by-step $2 \to 1$ mergings, and we can unfold the jet into two, three, or more subjets by undoing the last recombinations. Because the jet algorithm begins and ends with physically meaningful information (starting at calorimeter cells, for example, and ending at jets), the intermediate (subjet) information generated by the k_t and CA (but not the anti-k_t[2]) recombination algorithms is expected to have physical significance as well. In particular, we expect the earliest recombinations to approximately reconstruct the QCD shower, while the last recombinations in the algorithm, those involving the largest-p_T degrees of freedom, may indicate whether the jet was produced by QCD alone or a heavy particle decay plus QCD showering.
Jet *pruning* is a technique that modifies the jet substructure to reduce the systematic effects that "obscure" heavy particle reconstruction. It is expected to reduce the impact of uncorrelated soft radiation from the underlying event and pileup, and to improve the mass resolution of reconstructed decays. We also expect that this procedure will systematically shift the QCD mass distribution lower, reducing the background in the signal mass window.

Let's focus on the top branching (the last merging) of the jet splitting into two daughter subjets, which we will label $J \to 1, 2$. If ΔR_{12} is the angle between the daughter particles, and z is the ratio of the minimum daughter p_T to the parent p_T:

$$z \equiv \frac{min(p_{T1}, p_{T2})}{p_{TJ}} \tag{3.7}$$

the following pruning procedure can be defined:

[2]The anti-k_t algorithm, because of its metrics (see 3.3.2.1), tends to cluster protojets with the hardest protojet, resulting in cone-like jets with uninteresting substructure.

(a) start with a jet found by any jet algorithm, and collect the objects[3] in the jet into a list L. Define parameters D_{cut} and z_{cut} for the pruning procedure;

(b) rerun a jet algorithm on the list L, checking for the following conditions in each recombination $i, j \to p$: $z < z_{cut}$ and $\Delta R_{ij} > D_{cut}$. This algorithm must be a recombination algorithm such as the CA or k_t algorithms, and should give a "useful" jet substructure (one where we can meaningfully interpret recombinations in terms of the physics of the jet);

(c) if the conditions in (b) are met, do not merge the two branches 1 and 2 into p. Instead, discard the softer branch, i.e., veto on the merging. Proceed with the algorithm;

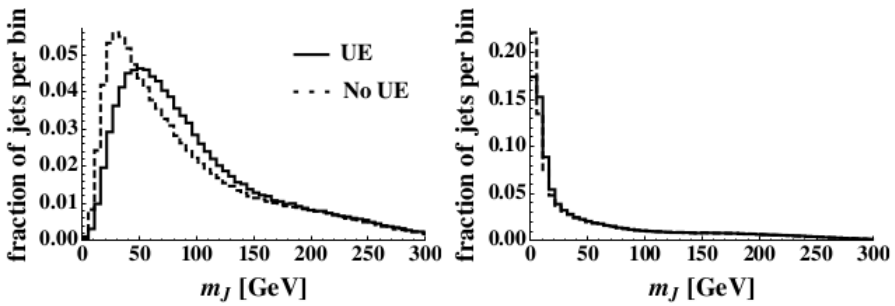
(d) the resulting jet is the pruned jet, and can be compared with the jet found in step (a).

The results of pruning are rather insensitive to the z_{cut} and D_{cut} parameters. For both the k_T and CA algorithms, a $z_{cut} = 0.10$ is set. The parameter D_{cut} can be determined on a jet-by-jet basis, allowing pruning to be more adaptive than a fixed parameter procedure. D_{cut} essentially determines how much of the jet substructure can be pruned, with smaller values allowing for more pruning. A natural scale for D_{cut} is the opening angle of the jet. In Fig. 3.1 the mass distribution of jets with and without the UE in both a QCD and a $t\bar{t}$ samples for the CA algorithm is shown, with and without pruning. Three distinctions between pruned and unpruned jets are clear. First, the distributions with and without the UE are very similar for pruned jets, while they noticeably differ for unpruned jets. This shows that pruning has drastically reduced the contribution of the underlying event. Second, the mass peak of jets near the top quark mass in the $t\bar{t}$ sample is significantly narrowed by the introduction of pruning (especially when the UE is included). This is evidence of the improved mass resolution of pruning, and will contribute to the improvement in heavy particle identification with pruning. And finally, the mass distribution of QCD jets is pushed significantly downward by pruning. The QCD jet mass is dominantly built from the soft, large-angle recombinations. Removing these by pruning the jets reduces the QCD mass distribution in the large mass range and will contribute to the reduction of the QCD background.

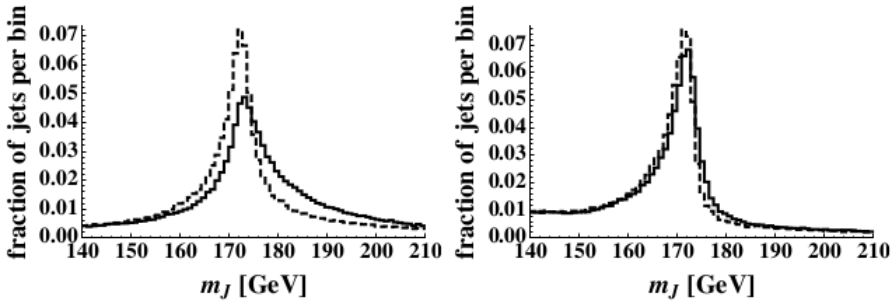
3.3.5 Jet identification

As already mentioned, the 2012 LHC run at $\sqrt{s} = 8$ TeV had an average pileup rate of 23 additional collisions, with some events exhibiting well over 40 pileup collisions. The typical p_T density of pileup is roughly 0.7 GeV per unit area (in the η, ϕ plane) per reconstructed primary vertex. For the 2012 running of CMS, this gives a total pileup p_T of 10 GeV for a typical anti-k_T jet with radius parameter $R = 0.5$. The origin of pileup deposits are varied, however most pileup jets are built from low p_T QCD jet production resulting from pileup collisions. This implies that the pileup itself is clustered. Additionally, it is known from extrapolations of the inclusive jet cross sections down to low p_T that a single

[3] In our case the PF objects, but they could be any objects that were clustered into jets, e.g. calorimeter towers.



(a) unpruned QCD jets (b) pruned QCD jets



(c) unpruned top jets (d) pruned top jets

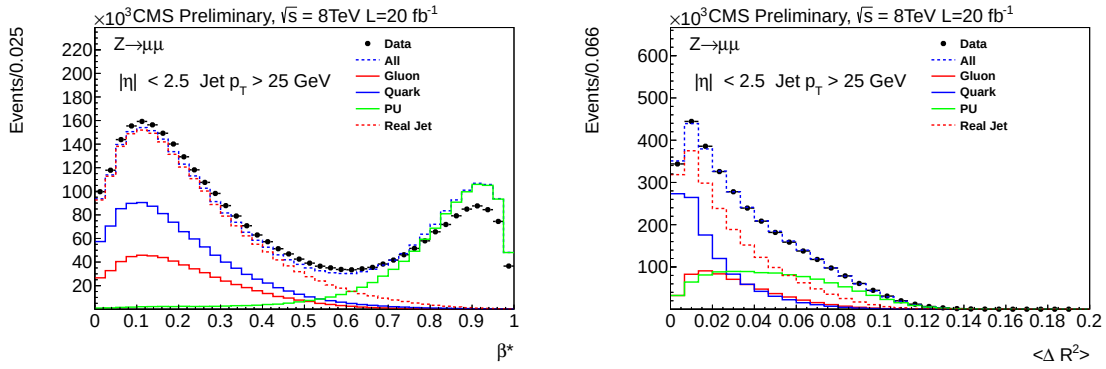
Figure 3.1: Distributions in m_J with and without underlying event, for QCD and top jets, using the CA algorithm, with and without pruning. The jets have p_T between 500 and 700 GeV, and $D_{cut} = 1.0$.

jet with a $p_T > 5$ GeV occurs with nearly every collision. Such a large incidence of low p_T jets induces a phenomenon whereby the low p_T jets combine to form one single high p_T jet. The resulting jet formed from overlapping jets is known as a "pileup jet". Due to the fact that pileup jets primarily originate from overlapping jets incurred during pileup interactions, pileup jets exhibit two characteristic features: they are both diffuse and, when charged particle identification is possible, some fraction of the charged particles will not point to the primary vertex (PV) [4]. These characteristics allow for the identification of pileup jets in both regions where charged particle tracking is available and regions where jet shape identification is possible. Both vertex and shape information are combined through a multivariate analysis technique, to give a single discriminator targeting the identification (ID) of pileup jets. This technique is known as the *pileup jet ID*. This is how different variables could allow for PU jet identification [94]:

- *vertexing related variables*: charged PF candidates, with associated tracks, contribute to roughly half of the total pileup. Inside or near the tracker volume, a distinct enhancement in the ability to discriminate against pileup is possible by exploiting the compatibility of the jet tracks with the PV. To perform the jet identification, a variable β^* is defined as the sum of the p_T of all PF charged candidates associated to other vertices divided by the sum of the p_T of all charged candidates in the jet:
$$\beta^* = \frac{\sum_{i \notin \text{PV}} p_{T_i}}{\sum_i p_{T_i}} \tag{3.8}$$
- *shape related variables*: outside the tracker volume the use of vertexing is not possible; thus jet shower shapes are the only handle to distinguish pileup jets. Since characteristically overlapping pileup jets tend to result in wider jets, shape related variables are precisely designed to target the diffuseness of a jet. To perform the jet identification, a single radial variable is defined as:
$$\langle \Delta R^2 \rangle = \frac{\sum_i \Delta R_i^2 \cdot p_{T_i}^2}{\sum_i p_{T_i}^2} \ , \tag{3.9}$$
where the sum runs over all PF candidates inside the jet and $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}$ is the distance of the PF candidate with respect to the jet axis.

Shape and vertex information can be combined through a multivariate BDT to give the pileup jet ID available for all jets used in CMS.
In our analysis, this multivariate PU jet ID has been used, because VBF jets can be often emitted outside the tracker acceptance (they are searched for up to $|\eta| = 4.7$).
In Fig. 3.2 the behaviour of the jet ID vertexing related variable β^* and the jet ID shape related variable $\langle \Delta R^2 \rangle$ are shown in a study performed on $Z(\rightarrow \mu\mu)$ + jets events for jets with $p_T > 25$ GeV in the tracker acceptance $|\eta| < 2.5$. Values are based on 20 fb^{-1} data collected at a center-of-mass energy $\sqrt{s} = 8$ TeV [94]. As shown in the figure, PU jets tend to be at high values of β^*, close to 1, and they have smaller values and a flatter

[4]The primary vertex is defined as the vertex for which $\sum p_{T_i}^2$ is largest, where i runs over the tracks pointing to the considered vertex.



(a) Jet ID vertexing related variable β^*.

(b) Jet ID shape related variable $\langle \Delta R^2 \rangle$.

Figure 3.2: Jet ID vertexing related variable β^* **(a)** and Jet ID shape related variable $\langle \Delta R^2 \rangle$ **(b)** in $Z(\rightarrow \mu\mu)$ + jets events for jets with $p_T > 25$ GeV in the Tracker acceptance $|\eta| < 2.5$. Values are based on 20 fb^{-1} of 8 TeV data samples [94].

profile in the $\langle \Delta R^2 \rangle$ distribution: therefore, they can be well separated from QCD jets. The PU jet ID is the technique chosen for the analysis described in the following.
Another technique commonly used in CMS and orthogonal to the pileup jet ID is known as *charged hadron subtraction*. In this technique charged particle flow candidates pointing at another vertex are removed and the jets are allowed to recluster.

3.3.6 Jet energy calibrations/corrections

The purpose of the jet energy calibration is to relate, on average, the energy measured in the detector to the true energy of the corresponding final state particle jet or parton jet. A true particle jet results from the clustering (with the same clustering algorithm applied to detector jets) of all stable particles originating from the fragmenting parton, as well as of the particles from the underlying event (UE) activity. The correction is applied as a multiplicative factor $\mathcal{C}$ to each component of the raw jet four-momentum vector p_μ^{raw} (components are indexed by μ in the following):

$$p_\mu^{cor} = \mathcal{C} \cdot p_\mu^{raw} \tag{3.10}$$

The correction factor $\mathcal{C}$ is composed of the offset correction C_{offset}, the MC calibration factor C_{MC}, and the residual calibrations C_{rel} and C_{abs} for the relative and absolute energy scales, respectively. The offset correction removes the extra energy due to noise and pileup, and the MC correction removes the bulk of the non-uniformity in η and the non-linearity in p_T. Finally, the residual corrections account for the small differences between data and simulation. The various components are applied in sequence:

$$\mathcal{C} = C_{offset}(p_T^{raw}) \cdot C_{MC}(p_T', \eta) \cdot C_{rel}(\eta) \cdot C_{abs}(p_T''), \tag{3.11}$$

where p_T' is the transverse momentum of the jet after applying the offset correction and p_T'' is the p_T of the jet after all previous corrections.

3.3.6.1 Offset correction

Pileup of multiple proton-proton collisions and electronic noise in the detector produce an energy offset. The goal of the offset correction is to subtract, on average, the unwanted energy from the jet. The offset correction is the first step in the chain of the factorized corrections. For its estimation, the *Jet Area Method* has been used [95]. For each event, an average p_T-density ρ per unit area is estimated, which characterizes the soft jet activity and is a combination of the underlying event, the electronics noise, and the pileup. The two latter components contaminate the hard jet energy measurement and need to be corrected for with the offset correction. The key element for this approach is the jet area A_j. A very large number of infinitely soft four-momentum vectors (soft enough not to change the properties of the true jets) are artificially added in the event and clustered by the jet algorithm together with the true jet components. The extent of the region in the $y - \phi$ space occupied by the soft particles clustered in each jet defines the active jet area. The other important quantity for the pileup subtraction is the p_T density ρ, which is calculated with the k_T jet clustering algorithm with a distance parameter $R = 0.6$. The k_T algorithm naturally clusters a large number of soft jets in each event, which effectively cover the entire $y - \phi$ space, and can be used to estimate an average p_T-density. The quantity ρ is defined on an event-by-event basis as the median of the distribution of the variable p_{Tj}/A_j, where j runs over all jets in the event, and is not sensitive to the presence of hard jets. At the detector level, the measured density ρ is the convolution of the true particle-level activity (underlying event, pileup) with the detector response to the various particle types. Based on the knowledge of the jet area and the event density ρ, an event-by-event and jet-by-jet pileup correction factor can be defined:

$$C_{area}(p_T^{raw}, A_j, \rho) = 1 - \frac{(\rho - \langle\rho_{UE}\rangle)\ A_j}{p_T^{raw}}. \tag{3.12}$$

In the formula above, ρ_{UE} is the p_T-density component due to the UE and electronics noise, and is measured in events with exactly one reconstructed primary vertex (no pileup). Fig. 3.3 shows the PF p_T-density ρ, as a function of the leading jet p_T in QCD events and for various pileup conditions. The fact that ρ does not depend on the hard scale of the event confirms that it is really a measure of the soft jet activity. Finally, the density ρ shows linear scaling properties with respect to the amount of pileup. In Fig. 3.4 the offset transverse momentum correction for the CMS reconstructed jets is shown as a function of jet pseudorapidity in data and Monte Carlo for different intervals of reconstructed primary vertices (N_{PV}). Values are based on 11 fb^{-1} of 8 TeV data collection [96].

3.3.6.2 MC calibration correction

The MC calibration is based on the simulation and corrects the energy of the reconstructed jets such that it is equal on average to the energy of the generated jets (GenJets). The GenJets reconstruction algorithm is identical to the one applied to the data. Each reconstructed jet is spatially matched in the $\eta - \phi$ space with a GenJet by requiring $\Delta R < 0.25$. In each bin of the GenJet transverse momentum p_T^{gen}, the response variable $\mathcal{R} = \frac{p_T^{\text{reco}}}{p_T^{\text{gen}}}$ and

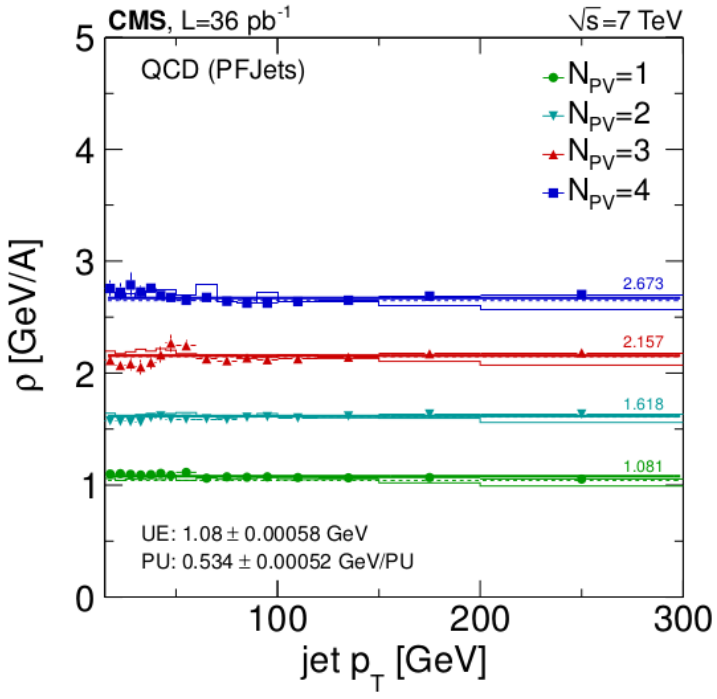


Figure 3.3: Pileup and underlying event PF p_T-density ρ, as a function of the leading jet p_T in the QCD multijet sample for various pileup conditions. Here NPV denotes the number of reconstructed vertices, and A denotes the unit area in the $y - \phi$ space.

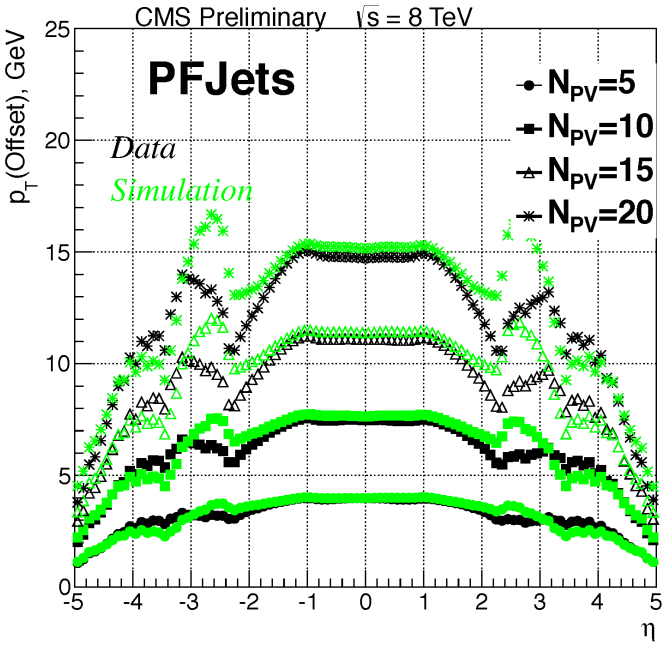


Figure 3.4: Offset correction as a function of jet pseudorapidity in data and Monte Carlo varying the number of reconstructed primary vertices (N_{PV}). Values are based on 11 fb^{-1} of 8 TeV data collection.

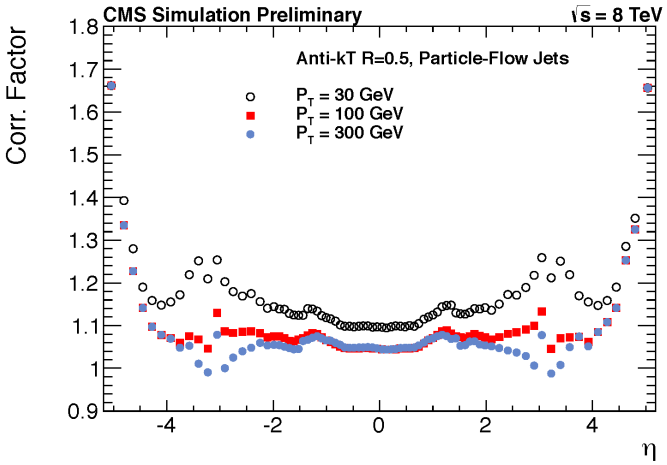


Figure 3.5: MC calibration corrections as a function of jet pseudorapidity for 30 GeV, 100 GeV and 300 GeV tranverse momenta, on 8 TeV simulated samples.

the detector jet p_T^{reco} are recorded. The average correction in each bin is defined as:

$$C_{\text{MC}}(p_T^{\text{reco}}) = \frac{1}{\langle \mathcal{R} \rangle} \ , \tag{3.13}$$

and is expressed as a function of the average detector jet p_T, $\langle p_T^{\text{reco}} \rangle$. In Fig. 3.5 the MC calibration corrections for the CMS reconstructed jets as a function of jet pseudorapidity for three reference transverse momentum values (30 GeV, 100 GeV and 300 GeV) are shown on MC samples at 8 TeV [96].

3.3.6.3 Relative jet energy scale

The goal of the relative jet energy scale correction is to make the jet response flat versus η. The uniformity in pseudorapidity is achieved by employing a data-driven method, the di-jet p_T-balance technique. Starting from an event with at least two jets in the final state, one jet (barrel jet, chosen as the "tag" jet) is required to lie in the central region of the detector ($|\eta| < 1.3$) and the other jet ("probe" jet) at arbitrary η. The central region is chosen as a reference because of the uniformity of the detector, the small variation of the jet energy response, and because it provides the highest jet p_T-reach. It is also the easiest region to calibrate in absolute terms, using γ + jet and Z + jet events. The two leading jets in the event must be azimuthally separated by $\Delta\varphi > 2.7$ rad. Fig. 3.6 shows the residual relative corrections $C_{\text{rel}}(\eta)$ as a function of jet pseudorapidity for a class of PF jets built with the *Charged Hadron Subtraction* (CHS) algorithm[5]. The Jet Energy Scale (JES) uncertainty, taken at the representative transverse momentum value of 100 GeV, is shown in with the statistical uncertainty. Values are based on 11 fb^{-1} of 8 TeV data samples [96].

[5]The CHS algorithm uses the CMS detector excellent tracking capabilities to identify and remove jet constituents (charged hadrons) which are known to have originated from pileup vertices. The algorithm uses a particle-by-particle pileup subtractions technique.

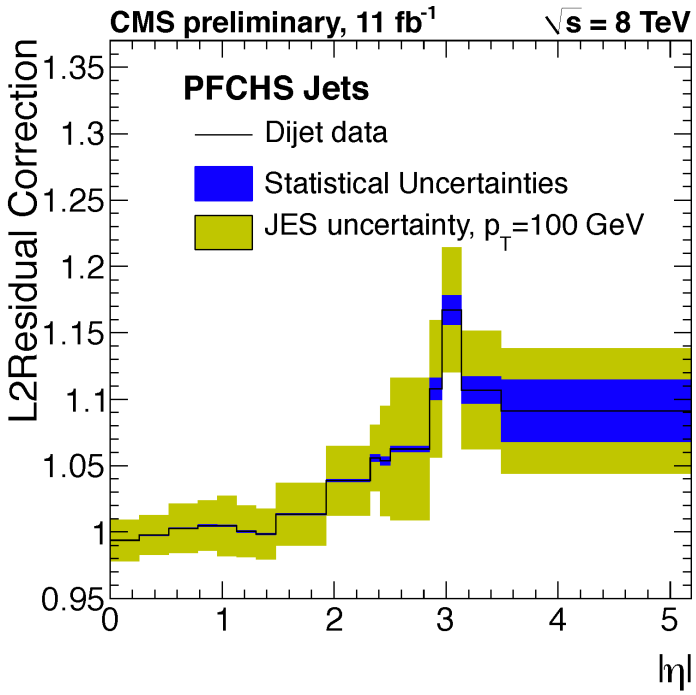


Figure 3.6: Residual relative corrections as a function of jet pseudorapidity for a class of PF jets built with the CHS algorithm. Values are based on 11 fb^{-1} of 8 TeV data samples.

3.3.6.4 Absolute jet energy scale

The goal of the absolute jet energy scale correction is to make the jet response flat versus p_T. Once a jet has been corrected for η dependence, it is corrected back to particle level. The absolute jet energy response is measured in the reference region $|\eta| < 1.3$ with the *Missing Transverse Energy Projection Fraction* (MPF) method [97] using γ, Z + jets events. The method is based on the fact that the γ, Z + jets events have no intrinsic missing transverse energy $\vec{E}_T^{\text{miss}}$ and that, at parton level, the γ or Z is perfectly balanced by the hadronic recoil in the transverse plane:

$$\vec{p}_T^{\,\gamma,Z} + \vec{p}_T^{\,\text{recoil}} = 0 \ . \tag{3.14}$$

For reconstructed objects, this equation can be rewritten as:

$$R_{\gamma,Z} \cdot \vec{p}_T^{\,\gamma,Z} + R_{recoil} \cdot \vec{p}_T^{\,\text{recoil}} = -\vec{E}_T^{\,\text{miss}} \ , \tag{3.15}$$

where $R_{\gamma,Z}$ and R_{recoil} are the detector responses to the γ or Z and the hadronic recoil, respectively. Solving the two above equations for R_{recoil} gives:

$$R_{recoil} = R_{\gamma,Z} + \frac{\vec{E}_T^{\,\text{miss}} \cdot \vec{p}_T^{\,\gamma,Z}}{\left(p_T^{\gamma,Z}\right)^2} \equiv R_{MPF} \ . \tag{3.16}$$

This equation provides the definition of the MPF response R_{MPF}. The additional step needed is to extract the jet energy response from the measured MPF response. In general,

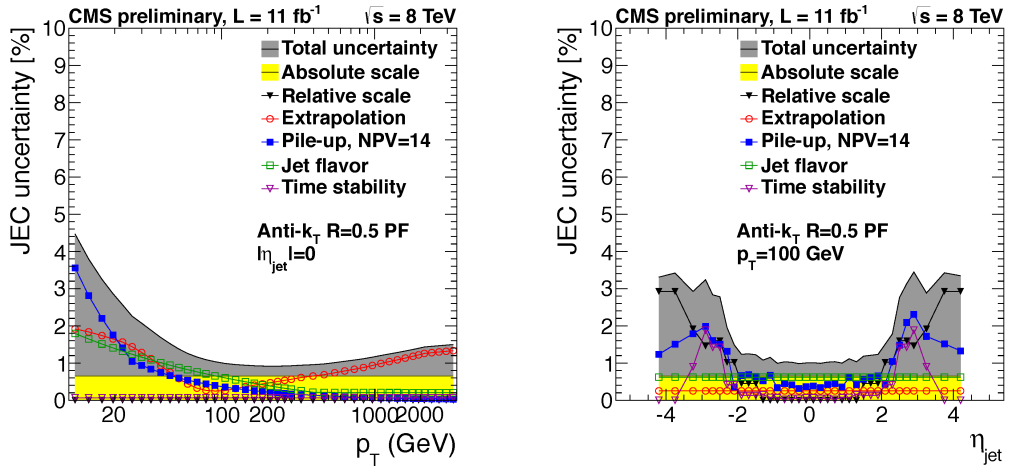
the recoil consists of additional jets, beyond the leading one, soft particles and unclustered energy. The relation $R_{\text{leadjet}} = R_{\text{recoil}}$ holds to a good approximation if the particles, that are not clustered into the leading jet, have a response similar to the ones inside the jet, or if these particles are in a direction perpendicular to the photon axis. The γ or the Z are used as reference objects because their energy is accurately measured in ECAL (photon, $Z \to e^+e^-$) or in the tracker and muon detectors ($Z \to \mu^+\mu^-$).

3.3.6.5 Jet energy uncertainties at CMS

Each type of correction has uncertainties arising from many different sources. These sources can be categorized as it follows [98]:

- physics modeling in MC such as showering, modeling of underlying event, etc.;
- MC modeling of true detector response and properties such as noise, zero suppression, etc.;
- potential biases in the methodologies used to estimate the corrections.

In CMS more than 16 such sources of uncertainties have been identified. Several are related and can be combined into groups that are relative to the absolute scale, relative scale, extrapolation in p_T, pileup, jet flavor and time stability. In Fig. 3.7 the contribution of the jet energy correction uncertainty from the combined sources is shown as a function of the jet p_T and jet η. The total uncertainty on the jet energy correction is computed as the quadrature sum of the uncertainty of each different contribution.



(a) (b)

Figure 3.7: Jet energy correction uncertainties as a function of p_T for CMS PF jets reconstructed around $|\eta| = 0$ (a) and as a function of jet η for jets with $p_T = 100$ GeV (b), based on 11 fb^{-1} of 8 TeV data collected by CMS [96]. Different contributions are shown with markers of different colors, and the total uncertainty is shown as a gray band. All jets are reconstructed with the anti-k_t algorithm with $R = 0.5$.

3.3.7 Jet resolution smearing

Measurements show that the jet energy resolution (JER) in data is worse than in the simulation, so the jet energy in MC need to be smeared to better describe data. [99] The method chosen to smear the jet energy resolution in this analysis is the following. The reconstructed jet transverse momentum is scaled based on the p_T difference between the reconstructed jet and the corresponding (matching) generated jet:

$$p_T \rightarrow max[0.,\ p_T^{gen} + c \cdot (p_T – p_T^{gen})] \tag{3.17}$$

where c is the "core" resolution scaling factor, i.e. the measured data/MC resolution ratio. This method only works for jets which are well matched to a generated jet, otherwise the method can lead to large response shifts.

3.3.8 Jet b-tagging

In the analysis which will be described in the following, some jets in signal events come from Z decay, where the quark flavour is - in first approximation - equally distributed among quarks u, d, s, c, b, with down flavours slightly favoured because of the electroweak coupling of the Z boson. Instead, the main background process is given by Z+jets events, where jets are originated by both quarks and gluons (the latter gives about 45% of the total rate). Furthermore, quark jets in Z+jets background are mainly originated by light quarks u and d. Then, jet flavour can be in principle exploited to have important information about the origin of jets and consequently on the signal- or background-like nature of the event.
In CMS, several algorithms dedicated to the tagging of jets from b quarks (*b-tagging* algorithms) are available. Hence, a *b-tagging* technique can be easily applied; events with jets which are tagged as coming from b will be more likely signal. However, in the VBF analysis, which is the main subject of this document, the result of the *b-tagging* algorithm will be only exploited at the "Higgs candidate selection" stage (see Sec. 4.4), that is, no event classification nor signal over background discrimination is made on the basis of *b-tagging*, which is only used as a criterion to preferentially choose a candidate among the available ones, when more than one is found.
The properties of the bottom hadrons can be used to identify b-jets [100]. These hadrons have relatively large masses, long lifetimes (typical impact parameter values for tracks from b-hadron decays are at the level of a few 100 μm) and daughter particles with hard momentum spectra. Different reconstructed objects - tracks, vertices and identified leptons - can be used to build observables that discriminate between b and light-parton jets. Several simple and robust algorithms use just a single observable, while others combine several of these objects to achieve a higher discrimination power. Each of these algorithms yields a single discriminator value for each jet. The minimum thresholds on these discriminators define loose ("L"), medium ("M"), and tight ("T") operating points with a misidentification probability for light-parton jets of close to 10%, 1%, and 0.1%, respectively, at an average jet p_T of about 80 GeV. In the present analysis, the Jet Probability *b-tagging* algorithm [100] is used to identify jets originating from heavy-flavour quarks. The Jet Probability (JP) algorithm uses an estimate of the likelihood

that all tracks associated to the jet come from the primary vertex. The estimate for the likelihood, P_{jet}, is defined as

$$P_{jet} = \Pi \cdot \sum_{i=0}^{N-1} \frac{(-\ln \Pi)^i}{i!} \quad \text{with} \quad \Pi = \sum_{i=0}^{N} \max(P_i, 0.005), \tag{3.18}$$

where N is the number of tracks under consideration and P_i is the estimated probability for track i to come from the primary vertex. The P_i are based on the probability density functions for the impact parameter (IP) significance of prompt tracks. The cut-off parameter for P_i at 0.5% limits the effect of single, poorly reconstructed tracks on the global estimate. The distribution of the JP discriminator in data and simulation is shown in Fig. 3.8.

In the analysis I'm going to describe, *b-tagging* is applied to both the jets coming from the Z boson. Each of the two jets can be tagged as *b-tagged* in a tight, medium or loose way, as introduced above, according to the JP algorithm working points which are defined by the selection cuts in Tab. 3.4. Three categories based on *b-tag* can then be defined:

- 2-btag category, including events with one jet tagged using the medium working point of the JP algorithm (JPM) and another jet tagged using the loose working point (JPL);
- 1-btag category, including events with at least one jet tagged using JPL;
- 0-btag category, including the untagged events.

Working point	Cut
Tight	0.790
Medium	0.545
Loose	0.275

Table 3.4: Working points for the Jet Probability b-tag algorithm.

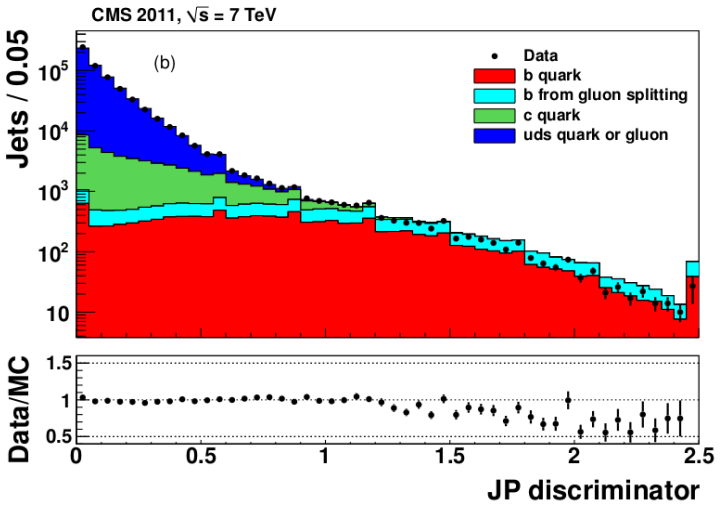


Figure 3.8: Discriminator values for the JP algorithm. The small discontinuities in the distributions are due to the single track probabilities which are required to be greater than 0.5%.

Chapter 4

Event selection for the $H \to ZZ \to l^+l^-q\bar{q}$ analysis in the VBF production mechanism

The goal of the analysis which will be described in the following is the search for a high-mass Higgs boson decaying into a pair of Z bosons, one furtherly decaying to a pair of leptons, the other one to a pair of jets, with the Higgs boson produced via Vector Boson Fusion. This analysis has to be framed-up into the more general analysis aiming at the search for the high-mass Higgs boson in this final state, taking into account all the relevant production mechanisms. So, our VBF sub-analysis is then combined with two additional sub-analyses focusing on the gluon-fusion production mechanism:

- one searching for a couple of separate jets from Z in the final state;
- one searching for a single merged-jet coming from the hadronic Z, with a two-subjets substructure, aimed at the analysis of a boosted topology expected in events with a very massive Higgs.

These two gluon-fusion sub-analyses will be described in some detail in Appendix A.

Let's now focus on the VBF analysis. The analysis aims at the selection of signal events characterized by the presence, in the final state, of two same-flavour opposite-sign leptons (specifically, electrons or muons) from the decay of the "leptonic" Z from the Higgs boson, and four jets, two from the decay of the "hadronic" Z from the Higgs and two from the VBF production mechanism. In Fig. 4.1 a Feynman diagram of the signal process which has been studied is shown.

The jets coming from the VBF production (in the following, the *VBF jets*) are expected to have a very peculiar topology, which will be exploited in the analysis to discriminate the signal from the background and from the gluon-fusion signal.
The main background sources are given by Z+jets events (Drell-Yan Z+jets), dibosons (WZ, ZZ) and $t\bar{t}$ production. In Fig. 4.2, 4.3 and 4.4 examples of Feynman diagrams for these background sources are shown.

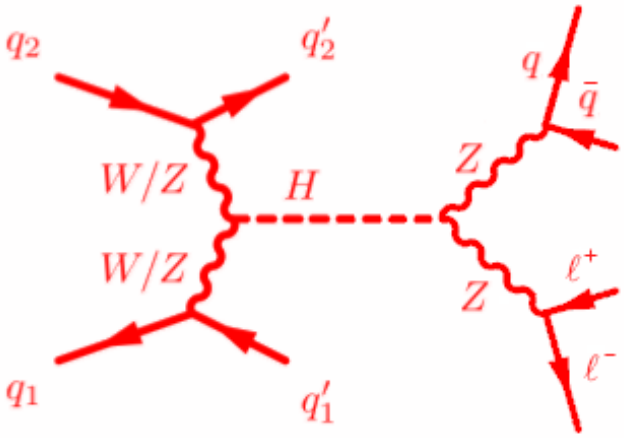


Figure 4.1: Feynman diagram of the signal process. Each interacting quark (q_1, q_2) emits a vector boson (W or Z). The vector bosons then merge into a Higgs boson, which subsequently decays into a pair of Z bosons. One of them decays into a pair of quark-antiquark, the other one into a pair of leptons. Only electrons and muons have been considered.

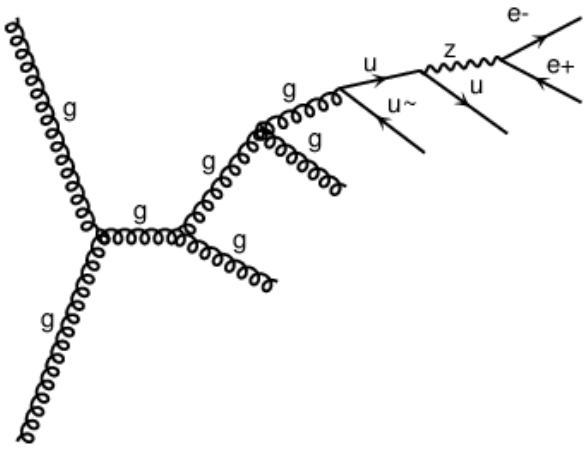


Figure 4.2: Example of a Feynman diagram for the Drell-Yan Z + 4 jets process.

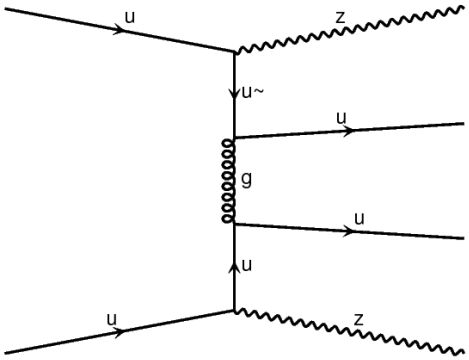


Figure 4.3: Example of a Feynman diagram for the ZZ + 2 jets process.

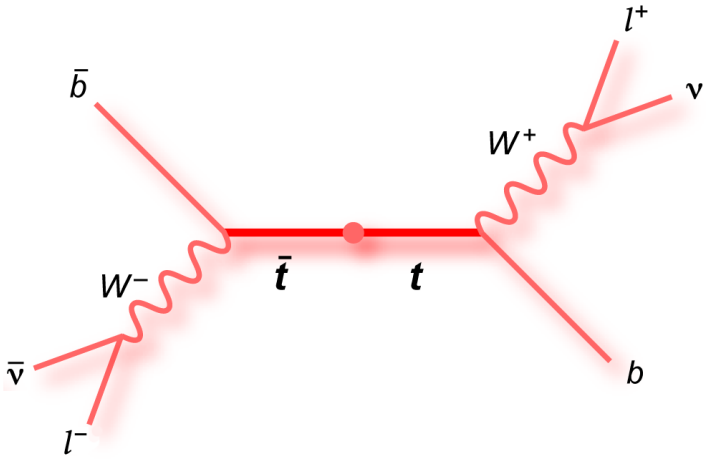


Figure 4.4: Feynman diagram for the decay of the top quark. Production of $t\bar{t}$ + 2 jets can mimic our signal because the final state differs from the one we are searching only for the additional presence of missing transverse energy due to undetectable neutrinos. An uppper cut on the missing transverse energy significance can thus allow to reject this background source.

Drell-Yan (DY) production of a pair of leptons can mimic the signal if four jets, coming from initial or final state radiation, are also reconstructed in the final state.
WZ events can mimic the signal if the Z decays into leptons and the W into a pair quark-antiquark; an additional pair of jets is then required, emitted as initial or final state radiation, to give a total amount of four jets in the final state. ZZ events signature is obviously similar to the signal one if the process is associated to the production of two additional jets.
Finally, $t\bar{t}$ events can give a signal-like signature if the decay chain is $t \to Wb$ and subsequently $W \to l\nu$. In such a case, in fact, there are two leptons and two jets (from the hadronization of the b quark) in the final state; other two jets can be given, as usual, by initial or final state radiation. This process, however, is characterized by the presence of missing energy in the final state, given by the presence of undetectable neutrinos, which is not expected in signal events. So this process can be easily rejected by putting requirements on the measured missing transverse energy.

4.1 Data and Monte Carlo samples

The analysis is entirely based on standard CMS simulation and reconstruction software. The name of the data samples used is listed in Tab. 4.1.
Data analyzed in this work was collected by the CMS experiment from 4th April, 2012 to 16th December, 2012 in proton-proton collisions at a center-of-mass energy $\sqrt{s} = 8$ TeV, corresponding to an integrated luminosity $L = 19.7$ fb^{-1}.
All data collected by the experiment are split into Primary Datasets (PD) by trigger selections and by physics interest. The datasets used in this analysis fall into the DoubleMu(Parked) and DoubleElectron PDs, given that they have been collected using a di-muon and a di-electron trigger. Di-muon events are required to fire the trigger path `HLT_Mu13_Mu8` which requires two global muons with $p_T > 13$ GeV and 8 GeV, respectively. The double electron trigger is the `HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_-Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL_v*` path, which requires the presence of two HLT electron objects in the event with transverse energy, E_T, greater than 17 and 8 GeV, respectively. Then, in order to reduce the rate of fake electrons, candidates are selected only if they pass loose requirements on the electromagnetic calorimeter shower shape and very loose isolation requirements. The "Parked" label refers to the fact that these data have been normally collected and stored by the HLT, but they have not been promptly reconstructed.
The "Run2012*" label in the listed samples refer to the data-taking period when these samples have been collected. In fact, the 2012 data collection has been performed in four different run periods, named A, B, C and D that differ for detector and beam conditions. The date in the sample name refers to when the reprocessing has been performed. In fact, these data have been reprocessed in January 2012 with the best detector calibration conditions (labelled as "22Jan2013"). This is the latest reprocessing and the related dataset is often referred to as "8 TeV Legacy dataset". "AOD" means *Analysis-Oriented Data* and is a "light" event format (about 120 kB in size) which is usable directly by physics

Sample	luminosity
/DoubleMu/Run2012A-22Jan2013-v1/AOD	876/pb
/DoubleMuParked/Run2012B-22Jan2013-v1/AOD	4412/pb
/DoubleMuParked/Run2012C-22Jan2013-v1/AOD	7054/pb
/DoubleMuParked/Run2012D-22Jan2013-v1/AOD	7370/pb
/DoubleElectron/Run2012A-22Jan2013-v1/AOD	876/pb
/DoubleElectron/Run2012B-22Jan2013-v1/AOD	4412/pb
/DoubleElectron/Run2012C-22Jan2013-v1/AOD	7054/pb
/DoubleElectron/Run2012D-22Jan2013-v1/AOD	7370/pb

Table 4.1: Data samples used.

analyses [101].

The Monte Carlo samples used to simulate VBF signal events have been produced using POWHEG and PYTHIA 6 for Higgs mass values ranging from 200 to 1000 GeV at 50 GeV steps.
For the simulation of backgrounds, samples produced with MadGraph 1.3.32 [34], using PYTHIA 6.426 [35] for the parton shower and hadronization, were mostly used. This choice ensure a state-of-the-art description of multijet final states, which is crucial for the analysis. The main background contribution for this analysis, the Drell-Yan Z+jets production, has been simulated using several samples that cover exclusive regions of the phase space and that have been optimized to allow coverage of the high p_T and high jet multiplicity tails with good statistical power. Both samples binned in parton multiplicity and samples binned in partonic H_T (namely, the scalar sum of the partonic transverse momenta) were used. Samples binned in jet multiplicity have been made explicitly exclusive with respect to the samples binned in the H_T by cutting away any event with parton level H_T higher than 200 GeV, which is the lower edge of the lowest H_T bin. VBF signal samples were simulated with POWHEG [102] and include the NLO $q\bar{q} \to Hq\bar{q}$ production mechanism matched with the parton shower (implemented with PYTHIA 6.426). For completeness, the list of samples is reported in Tab. B.1 for background processes and in Tab. B.2 for signal (see Appendix B).

The POWHEG samples were used for gluon fusion H+2 jets at NLO, produced with the so called MINLO (Multi-scale Improved NLO) technique [103].

4.2 Skimming

Jets were reconstructed with the anti-k_t reconstruction algorithm (see Sec. 3.3) with 0.5 radius parameter (AK5) using Particle Flow candidates as input (see Sec. 3.1). Standard pileup subtraction is applied (see Sec. 3.3.6.1).
Jets reconstructed with the Cambridge-Aachen (CA) algorithm were also used, with a radius parameter of 0.8 (CA8). CA8 jets were built using Particle Flow candidates as input, after the charged hadron subtraction algorithm.
Electrons and muons were reconstructed using the standard CMS algorithms (see Sec. 3.2).

Higgs candidates consist of a $Z \to ll$ candidate, a $Z \to jj$ candidate and two additional

jets, representing the VBF jets. The two leptons that make the $Z \to ll$ candidate can be either electrons or muons, with opposite charge, and are required to have p_T greater than 18 GeV. A $Z \to jj$ candidate is built for any pair of AK5 jets that has an invariant mass of at least 40 GeV, and in which the two jets have a transverse momentum of at least 20 GeV, with no pseudorapidity cut at this stage.
For high Higgs masses (above ∼600 GeV) the two Z bosons are significantly boosted and in particular the two jets from $Z \to jj$ decay are close and tend to be reconstructed as a single object. In order to regain efficiency in those cases, a $Z \to jj$ candidate is also reconstructed for any CA8 jet that has a substructure compatible with the presence of two subjets. To reconstruct the subjets, a pruning algorithm is applied to these CA8 jets (see Sec. 3.3.4.1). Subjets are required to have a p_T greater than 20 GeV and an invariant mass of at least 40 GeV. The jet pair regarded as the VBF pair is required to have a mass of at least 50 GeV and a $\Delta\eta > 2$, the jets are also required to have p_T greater than 20 GeV with no pseudorapidity cut. At this stage, no quality requirement is applied to leptons or jets and as many Higgs candidates are built as the combinatorial allows. The six objects that form a Higgs candidate (four jets and two leptons) are requested to be separated with $\Delta R > 0.5$.

Events are retained if they contain at least one Higgs candidate built as described above. This represents our skimming requirement. Further selection requirements are applied later on in the analysis and described in the next section.

4.3 Event selection

Events passing the skimming requirement described above are selected further as described in this section.
In Tab. 4.2 a summary of the main selections applied is shown. In order to suppress the dominant Drell-Yan Z+jets background and contamination from $t\bar{t}$ events, candidates are selected in the region where the invariant mass of the jets from Z is close to the Z boson nominal mass.
Also the lepton invariant mass is required to be in a range around the Z mass; this helps rejecting $t\bar{t}$ events.
Such events, however, are mostly rejected by a cut on the Missing Transverse Energy (MET) significance[1] because neutrinos (and consequently missing energy) are expected in the final state, from the decay chain $t \to bW^+$ with $W^+ \to l^+\nu$ (and charge conjugate).

4.3.1 Jet selection

AK5 Jets are required to have p_T higher than 30 GeV, and they need to pass loose jet identification requirements listed in Tab. 4.3 (see also Sec. 3.3). In addition, jets are required to pass an MVA-based pileup jet identification 3.3.5.

[1]The MET significance assesses, on an event-by-event basis, the likelihood that the observed MET is consistent with a fluctuation from zero because of detector-related limitations, like measurement resolution. The determination of the significance requires evaluation of the uncertainty on the total measured transverse energy.

Observable	Cut	Cut in boosted category
First (second) lepton p_T	> 40 (20) GeV	idem
Jets from Z p_T	> 30 GeV	-
Subjets from Z p_T	-	> 30 GeV
Pruned CA8 jet from Z p_T	-	> 100 GeV
VBF jets p_T	> 30 GeV	idem
Jets from Z $\|\eta\|$	< 2.4	-
Subjets from Z $\|\eta\|$	-	< 2.4
VBF jets $\|\eta\|$	< 4.7	idem
$(Z \to ll)$ mass	$\in$ [76 GeV, 106 GeV]	idem
$(Z \to ll)$ p_T	-	> 200 GeV
τ_2/τ_1	-	< 0.5
MET significance	< 10	idem

Table 4.2: Event selection for the standard category and for the boosted category.

Observable	Cut
Neutral hadron fraction	< 0.99
Gamma fraction	< 0.99
Number of constituents	> 1
Hadronic energy fraction	> 0 (applied if $\|\eta < 2.4\|$)
Electromagnetic energy fraction	< 0.99 (applied if $\|\eta < 2.4\|$)
Charged multiplicity	> 0 (applied if $\|\eta < 2.4\|$)

Table 4.3: Loose Jet Id requirements.

The two AK5 jets that form a $Z \to jj$ candidate are also required to have $|\eta| < 2.4$ (in order to be within the tracker acceptance), while this cut does not apply to the AK5 jets that form the VBF tagging pair, which are allowed to be reconstructed within a larger acceptance ($|\eta| < 4.7$) in order not to lose events, given the peculiar and well known large pseudorapidity separation between them.
The sub-jet candidates from a CA8 jet are also required to have p_T higher than 30 GeV and $|\eta| < 2.4$, as for the AK5 jets from Z. They also need to pass the loose jet identification requirements listed in Tab. 4.3. As further quality requirements, the pruned CA8 jet is required to have p_T higher than 100 GeV and the ratio of the 2-subjettiness over the 1-subjettiness (τ_{21}) (see Sec. 3.3.4.1) is required to be less than 0.5.

Due to the relatively large branching fraction of the Z-bosons decaying into a pair of bottom-antibottom quarks, compared to the abundance of light-quark or gluon jets in Z+jets background events, b-tagging (see Sec. 3.3.8) is used to identify jets originating from heavy-flavour quarks. No selection of candidates is performed based on b-tagging probabilities. Nevertheless, the b-tag category is accounted for when chosing the Higgs candidate (see Sec. 4.4).

4.3.2 Electron selection

When building leptonic Z candidates, electrons are selected with p_T greater than 40 GeV for the leading electron in the pair and p_T greater than 20 GeV for the second electron in the pair. They are requested to satisfy loose electron identification criteria (see Sec. 3.2.2). Selected electrons also have to pass relative isolation criteria as described in Sec. 3.2.2.1. As already seen in the dedicated section, $RelIso_{EA\ corrected}$ on selected electrons is requested to be less than 0.15.

4.3.3 Muon selection

Kinematic cuts on muons are identical to those for electrons (leading muon in the pair forming a Z candidate is requested to have p_T greater than 40 GeV, subleading is requested to have a p_T greater than 20 GeV). Muons are requested to satisfy the tight muon identification requirements described in Sec. 3.2.1. In addition, muons are requested to pass a relative isolation criterion including a $\Delta\beta$ correction (see Sec. 3.2.1.1).

4.4 Higgs candidate selection

As already mentioned in Sec. 4.2, a Higgs candidate is built out of any two opposite sign leptons plus four AK5 jet combination, or plus two AK5 jets and one CA8 jet with substructure. Only events with at least one of those Higgs candidates are retained to go through the actual selection phase described in this section. Higgs candidates built as described in Sec. 4.2 are selected such that:

- the two leptons satisfy the requirements described in Sec. 4.3.2 or Sec. 4.3.3, have opposite charge, and have an invariant mass between 76 GeV and 106 GeV.
- the two jets from the hadronic Z decay satisfy the requirements described in Sec. 4.3.1 and have an invariant mass between 71 GeV and 111 GeV.
- the two VBF jets satisfy the requirements described in Sec. 4.3.1.

If at least one Higgs candidate passes this selection, the event is selected. If more than one Higgs candidate passes this selection, they are ranked according to the following criteria.

1. Candidates are ranked in three b-tag categories and the ones with the best ranking are selected. The ranking, from best to worst, is the following (see Sec. 3.3.8 for the definition of the b-tag working points):
   - 1 medium b-tag and 1 loose b-tag.
   - 1 loose b-tag.
   - No b-tag.
2. If still more than one candidate survives, the candidate with the hadronic mass closest to 91.2 GeV is chosen.

VBF mass point (GeV)	Only di-jet	Both	Only merged
200	1	0	0
300	0.993	0.0067	0.0024
400	0.973	0.026	0.001
500	0.81	0.18	0.008
600	0.62	0.33	0.044
700	0.53	0.36	0.11
800	0.51	0.33	0.16
900	0.53	0.28	0.19
1000	0.55	0.26	0.18

Table 4.4: Fraction of events with only di-jet, both or only merged-jet candidates, for different masses of the VBF signal.

3. If still more than one candidate survives, the candidate with the leptonic mass closest to 91.2 GeV is chosen.

4. If still more than one candidate survives, the candidate with the highest VBF invariant mass is chosen.

The ranking described above is applied separately to candidates containing a pair of AK5 jets (di-jet candidate) and those containing a pair of CA8 subjets (merged-jet candidate). If both a di-jet and a merged-jet candidate are present, the merged-jet candidate is chosen if it has a $Z \to jj$ mass less than 130 GeV (edge of the sideband region described below), and if the $Z \to ll$ candidate has a p_T greater than 200 GeV. The motivations for this di-jet versus merged-jet arbitration criterion are discussed in Sec. 4.4.1 below.

4.4.1 Di-jet Vs. merged-jet arbitration

Since the CA8 jets have a significantly larger radius with respect to the AK5 jets, a not negligible fraction of events has both a di-jet and a merged-jet reconstructed $Z \to jj$ candidate, which may actually correspond to the same physical object.
Tab. 4.4 shows the fraction of events with only a di-jet candidate, both a di-jet and a merged candidate and only a merged candidate. Events are selected according to the criteria described above except the final di-jet merged-jet arbitration. The number of merged-jet candidates has a strong increase around a mass of 500 GeV.

In order to decide which of the two candidates is the best choice, mass resolution and fake rate were studied, on simulated samples. The fake rate here is defined as the fraction of events with a reconstructed $Z \to jj$ candidate in which at least one of the two legs is not matched to the generated parton from Z decay in a cone radius of 0.3. Fake rate and mass resolution have been studied both separately for any best di-jet and best merged-jet candidate and in events in which both candidates were present.

Fake rates computed on the VBF signal samples for different masses are shown in Tab. 4.5 for di-jet candidates and merged jet-candidates.

Mass resolution plots are shown in Fig. 4.5 for di-jet and merged-jet candidates at different Higgs masses. Candidates entering these plots are all those that pass the full

VBF mass point (GeV)	fake rate di-jet	fake rate merged-jet
200	0.54	-
300	0.12	0.62
400	0.075	0.13
500	0.067	0.024
600	0.056	0.016
700	0.068	0.014
800	0.070	0.011
900	0.082	0.009
1000	0.080	0.009

Table 4.5: Fake rate in VBF signal events for mis-reconstructed di-jet (in red) and merged-jet (in blue) candidates.

VBF mass point (GeV)	fake rate di-jet	fake rate merged-jet
200	-	-
300	0.37	0.62
400	0.10	0.15
500	0.048	0.032
600	0.044	0.020
700	0.050	0.016
800	0.043	0.016
900	0.063	0.020
1000	0.052	0.015

Table 4.6: Fake rate in VBF signal events for mis-reconstructed di-jet/merged-jet candidates in events with both a di-jet and a merged-jet candidate.

selection except the final di-jet/merged-jet arbitration. A gaussian fit is performed in order to give an estimate of the central value and of the width of the distribution.

When selecting events with both a di-jet and a merged-jet candidate we observe the fake rates shown in Tab. 4.6.

Mass resolution plots are shown in Fig. 4.6 for events with both a merged-jet and a di-jet candidate. It can be observed that in events with both a di-jet and a merged-jet candidate the merged-jet candidates show a smaller fake-rate and a slightly better mass resolution, except at low mass, where the fraction of events with a merged-jet is anyway small. For these reasons it has been decided to take the merged-jet candidate as the best candidate in events in which both a merged and a di-jet candidate were present.

4.5 Background determination

The following SM processes are considered as background in this analysis: Z+jets, di-bosons (WZ, ZZ) and $t\bar{t}$ production. They are simulated using the MC samples discussed in Sec. 4.1 and summarized in Tab. B.1.
Most of the events in the selected $l^+l^-q\bar{q}$ sample come from Drell-Yan Z+jets processes

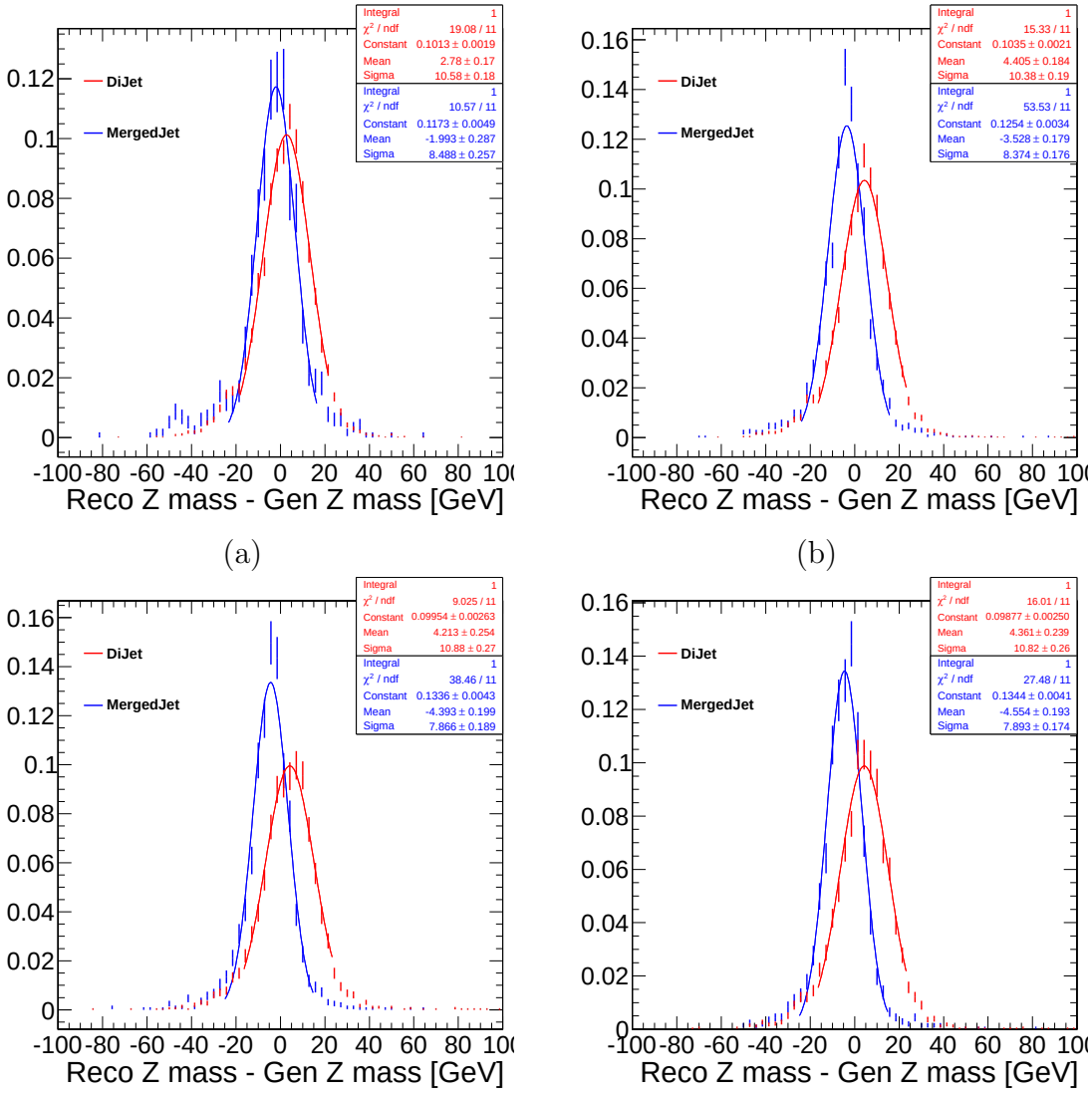


Figure 4.5: Difference between the reconstructed $Z \to jj$ mass and the generated one for different H mass points (a) 500 GeV, (b) 700 GeV, (c) 900 GeV, (d) 1000 GeV.

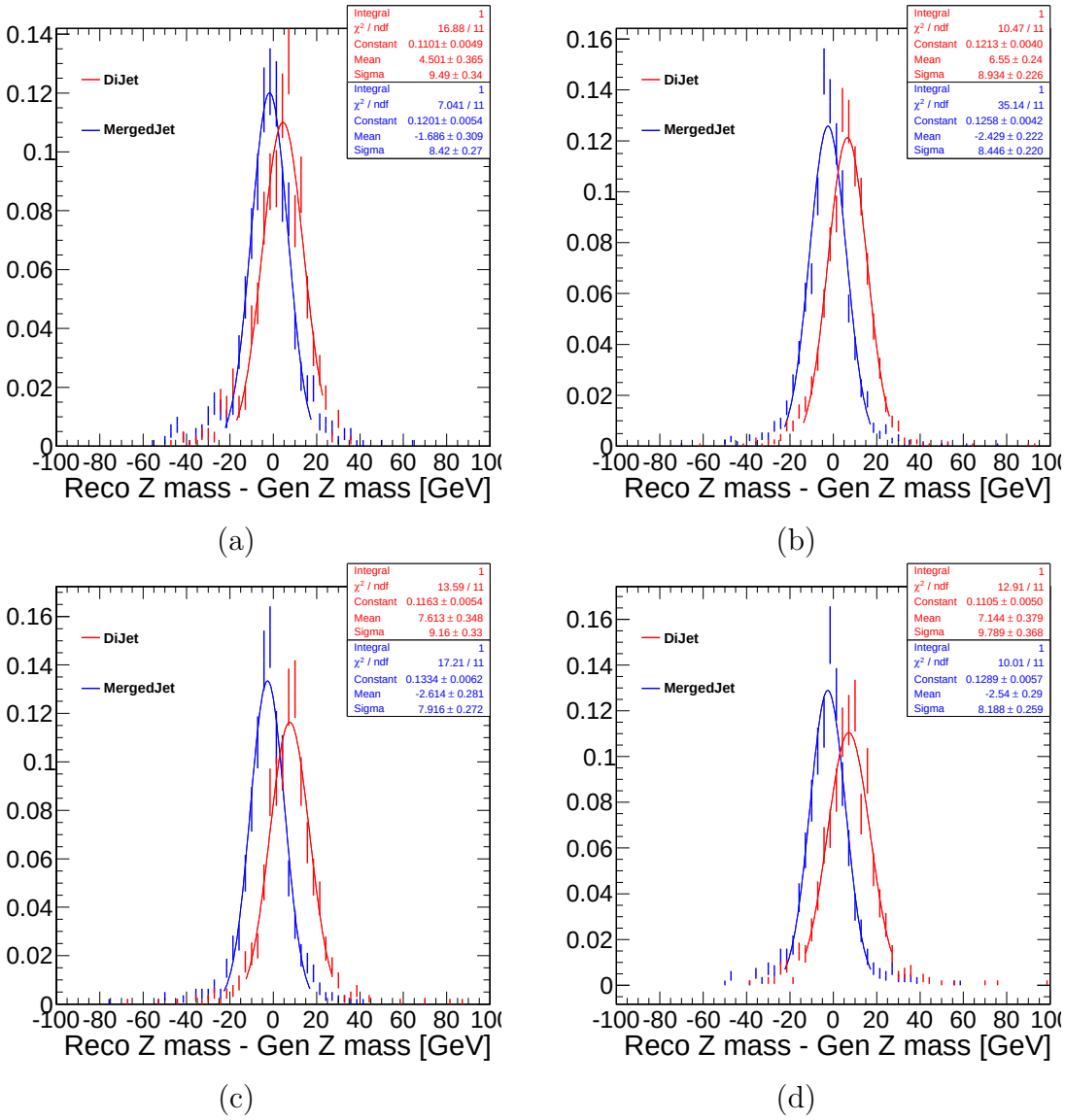


Figure 4.6: Difference between the reconstructed $Z \to jj$ mass and the generated one for different H mass points in events with both a merged (in blue) and a di-jet (in red) candidate. (a) 500 GeV, (b) 700 GeV, (c) 900 GeV, (d) 1000 GeV.

(about 93% of the total number of events, after the kinematic selection, for a simulated Higgs mass of 400 GeV). A real Z boson decaying into l^+l^- and a number of high p_T jets may easily yield signal-like combinations, even though jets do not stem from a Z boson. The m_{jj} distribution of these events is essentially a falling exponential above 220 GeV. The dibosons and $t\bar{t}$ backgrounds give respectively a contribution of about 2% and 3% of the total number of events (again, for m_H=400 GeV) after the selection described in Sec. 4.3 and summarized in Tab. 4.2.

4.6 The *blind analysis*

To avoid a possible bias of the experimenter in the process of developing the analysis strategy and optimizing the selection requirements, firstly a *blind analysis* was performed by defining a background-only *sideband* control region and only looking at data events falling in such a region. Precisely, the control region is defined according to all selection criteria described above, except that the invariant mass of the two jets contributing to the Higgs candidate four-momentum is required to be either in the window between 50 GeV and 71 GeV or in the window between 111 GeV and 130 GeV. Data in the signal region, i.e. with m_{jj} within the range [71, 111] GeV, were not examined until the event selection criteria were settled, only relying on MC simulations for what concerns signal. In this way, the analysis strategy was decided looking at the simulated events. Only when data distributions in signal-depleted regions confirmed the robustness of the analysis, data in signal region were inspected. For these reasons, the agreement between data and MC was checked on a sideband region.

4.7 Data-MC agreement

4.7.1 Recommended procedures for data/MC agreement

Several recommended reweighting factors have been applied to improve the agreement between data and MC. A pileup reweighting is applied to Monte Carlo events to make the simulated pileup distribution match the pileup distribution in data. The distribution of the true number of PU interactions in data is divided by the simulated number of PU interactions, resulting in a weight that is applied to each Monte Carlo event. The true PU distribution in data is obtained using the per-bunch instantaneous luminosity and the total inelastic cross section. The comparison of the number of valid primary vertices in data and simulation is shown in Fig. 4.7. This plot is made on events selected according to the "side-band" control region described below in Sec. 4.6.

A Complex Pole Scheme (CPS) lineshape reweighting is also applied [104–106]. This effect takes into account the non-zero width of the resonance and is important for high Higgs boson masses ($>$ 400 GeV). In the analysis, it is included in the samples via a reweighting of the simulated samples.

A jet resolution smearing is applied to the MC to match the jet resolution observed in data. This procedure is described in Sec. 3.3.7.

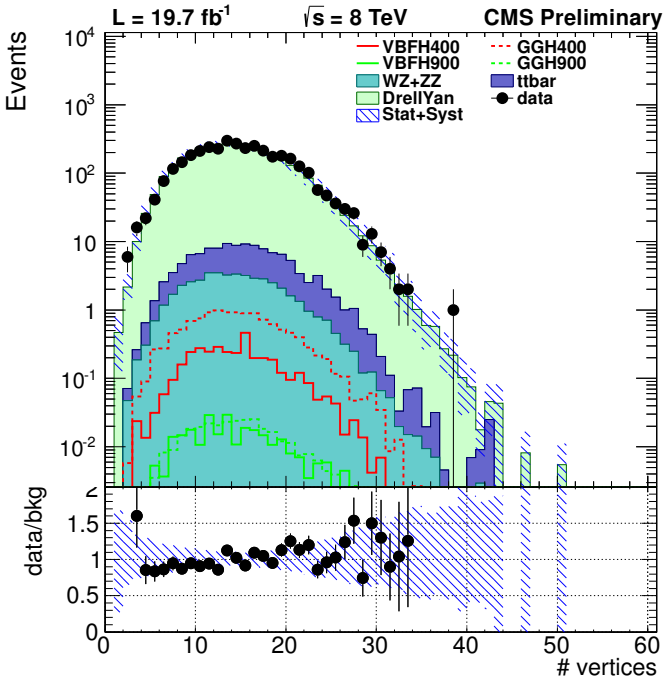


Figure 4.7: Number of reconstructed valid primary vertices in data and MC.

A boosted Z tagging efficiency data-to-MC scale factor of 0.91±0.10 taken from [91, 107] is applied to diboson events (signal and background dibosons) satisfying the final selection cuts.
Muon scale corrections are applied to data and MC as described in Sec. 3.2.1.2. On MC events an additional smearing is applied to match the resolution in data.
Electron regression corrections are applied to data and MC events to improve the determination of the electron momentum. An additional MC smearing is also applied.
Muon and electron efficiency scale factors are applied to MC events, in bins of muon pseudorapidity according to the prescriptions given in [108] and in bins of electron p_T and pseudorapidity, as described in [109].

4.7.2 Global scale factor

A global normalization correction is finally applied to the electron and muon channels separately by defining a background-only sideband control region and taking the ratio of the number of events in data and MC in that region.
The scale factors calculated in this way are summarized in Tab. 4.7. They vary depending on the stage of the selection at which they are evaluated.

4.7.3 Checks in control region

In order to validate the data-Monte Carlo agreement, several control plots were checked in the signal-free sideband region defined above. Full set of control plots is available

Selection stage	k-factor (electronic channel)	k-factor (muonic channel)
Kinematic selection	1.112 ± 0.028	1.041 ± 0.025
Kinematic sel. and LD > 0.2	1.112 ± 0.035	1.030 ± 0.031
Kin. sel. and LD> 0.2 and BDT> 0.4	1.09 ± 0.11	1.22 ± 0.11

Table 4.7: Electron and muon global scale factors computed in the $Z \to jj$ sideband region, for the different selection stages. LD is the likelihood discriminant which will be better described in Sec. 5.1.

in [110]. The distributions are built using events passing the kinematic selection described in Sec. 4.3 and summarized in Tab. 4.2. No cut on the discriminators described in Chap. 5 is applied at this stage.
Hadronic Z control plots are shown in Fig. 4.8. For each observable, Monte Carlo backgrounds are stacked and two representative signal mass points are shown (400 GeV and 900 GeV Higgs, in both the gluon fusion and VBF production mechanisms). An overflow bin is defined for every plot and shown as the rightmost bin. The hatched blue band represents the combination of statistical and systematic error on the MC prediction. Data and Monte Carlo agree roughly within 10%, except for the leading jet p_T where MC tends to overshoot the tail.

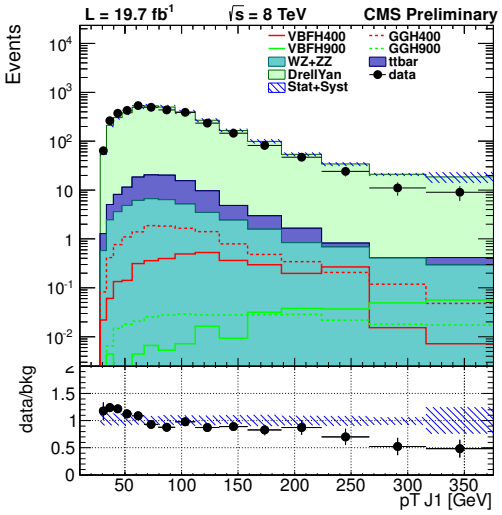
Leptonic Z control plots are shown in Fig. 4.9 for Z to di-electrons and Z to di-muons candidates. The agreement is good.

Kinematic variables of the leading and subleading VBF jet candidates are shown in Fig. 4.10. A tendency of the Monte Carlo to predict a harder spectrum is observed in the p_T distribution of both jets. Jet pseudorapidity and energy are in good agreement with expectation.
Fig. 4.11 shows some other VBF related observables, namely the invariant mass of the VBF pair, the $\Delta\eta$ and the $\Delta\varphi$ between the two VBF jets, for events with either a reconstructed Z to di-leptons or a reconstructed Z to di-muons. The agreement is reasonably good.

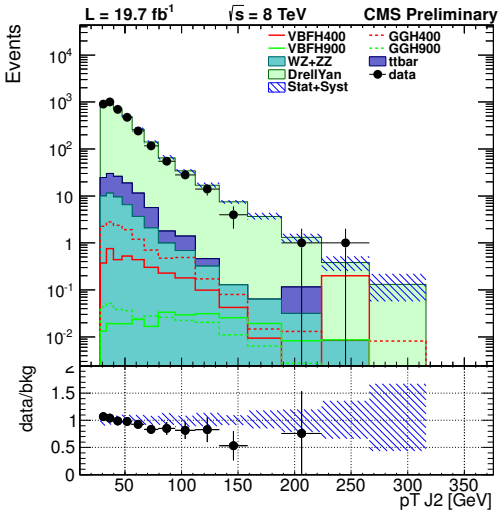
More Higgs-related variables are shown in Fig. 4.12 for events with a reconstructed di-electron di-jet candidate and for a reconstructed di-muon di-jet candidate. The observables shown are the reconstructed Higgs candidate mass and transverse momentum.

4.7.4 Checks in the signal region

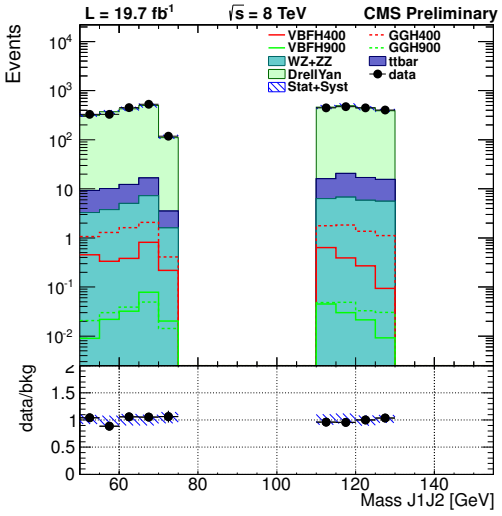
Similarly to Sec. 4.6, in this section we review the data to MC comparison in the signal region, i.e. in events with the hadronic Z candidate in the signal region, after the *unblinding* procedure. In the interest of making the document more readable we show only the hadronic Z candidate plots in Fig. 4.13. Full set of plots is available at [111]. In addition to the cuts used for the plots in Sec. 4.6, an LD>0.2 cut (see 5.1) is applied.



(a)



(b)



(c)

Figure 4.8: Hadronic Z control plots. (a) Logarithm of the leading jet p_T, (b) logarithm of the second jet p_T, (c) di-jet mass. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.

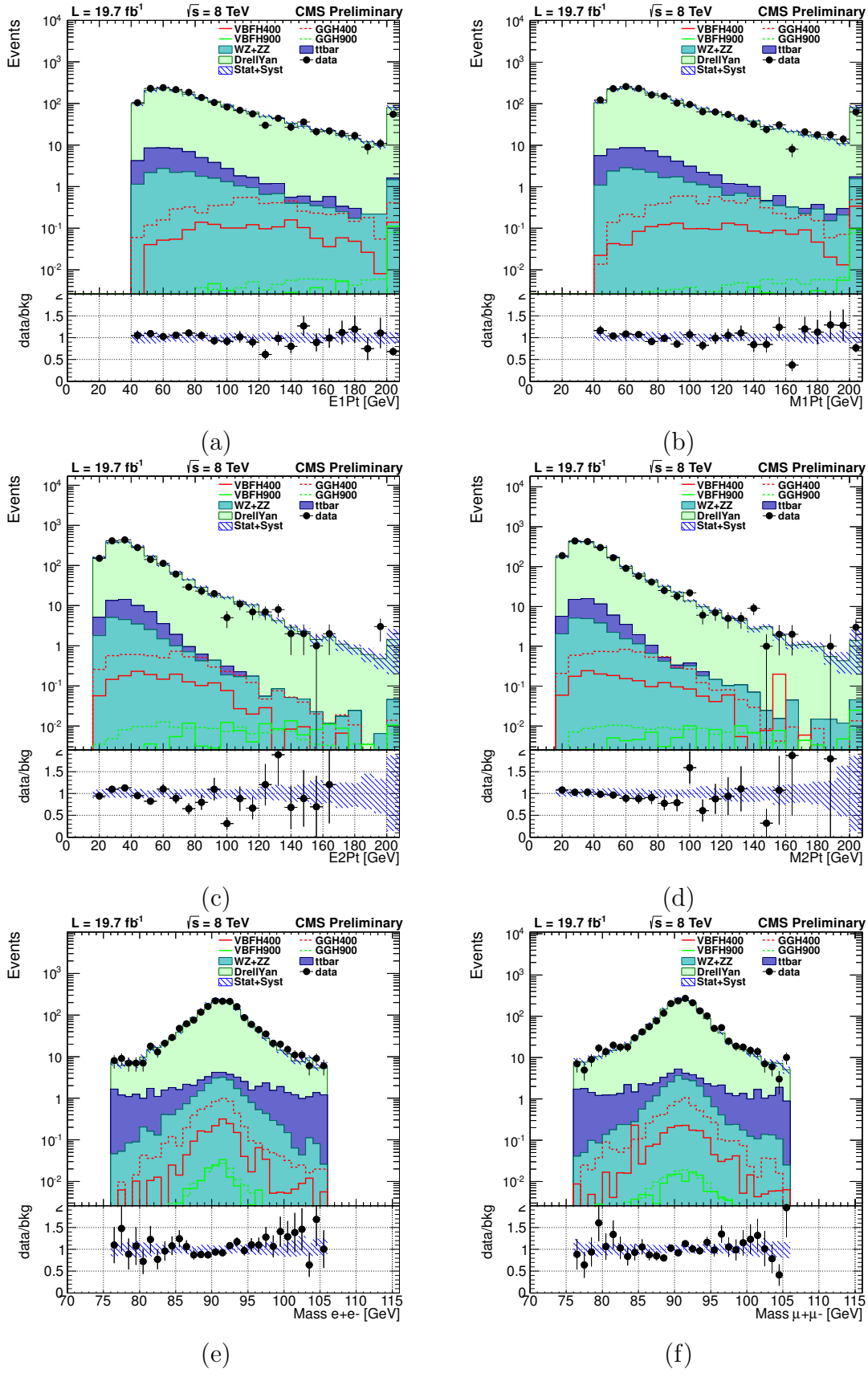


Figure 4.9: Z to di-lepton control plots. (a), (b) Leading lepton p_T, (c), (d) second lepton p_T, (e), (f) di-lepton mass in the sidebands. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.

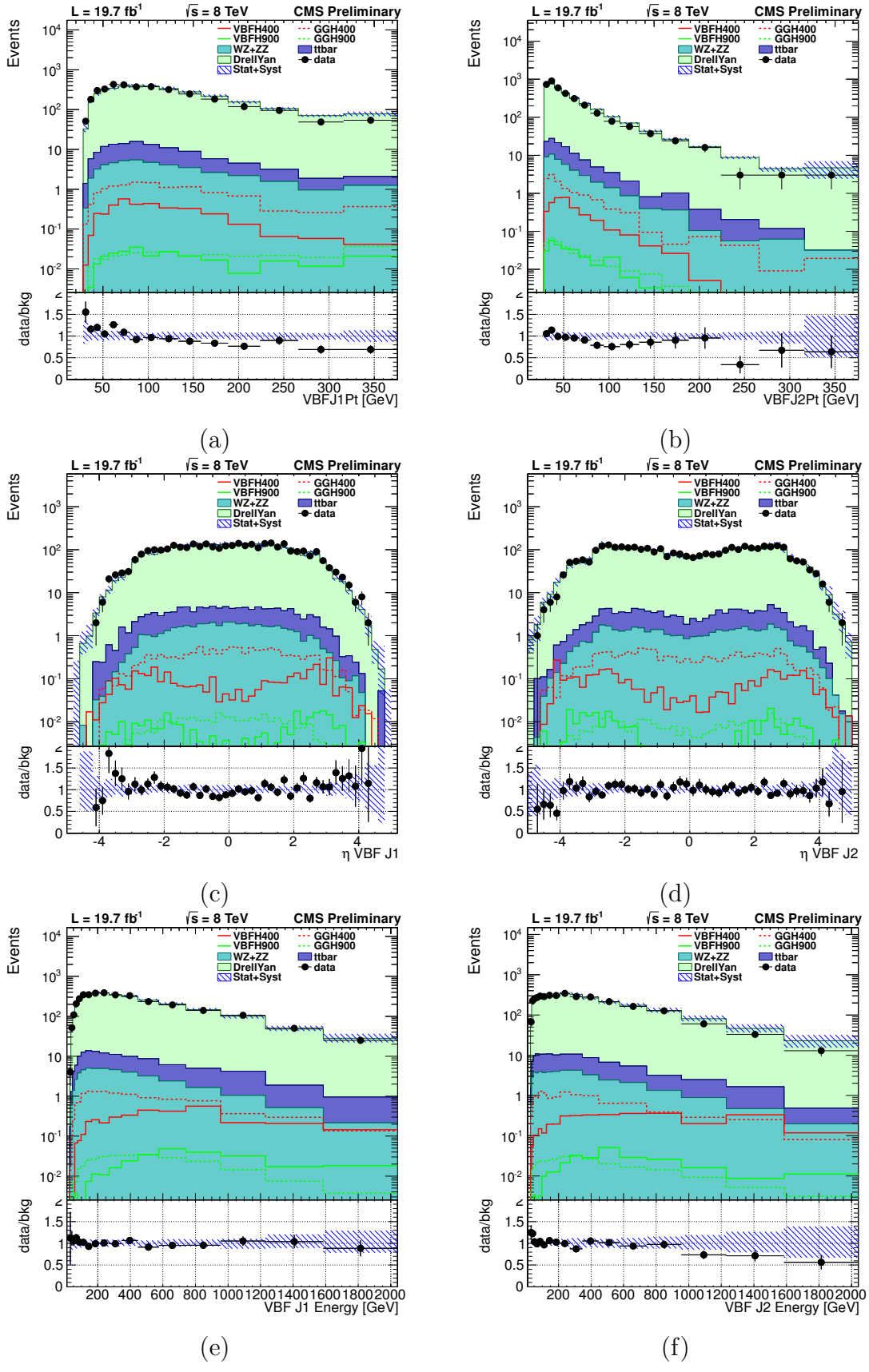
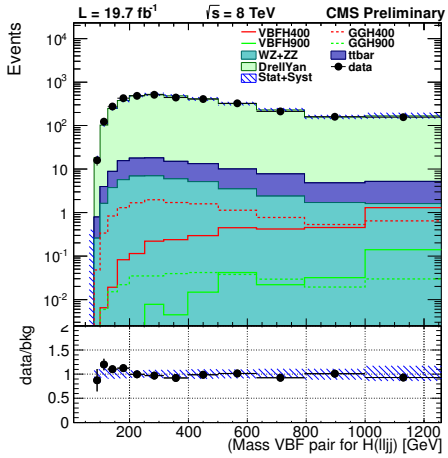
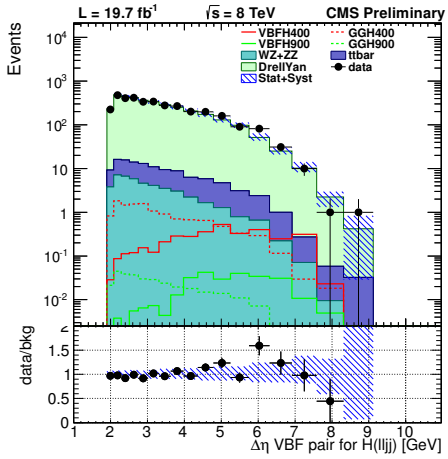


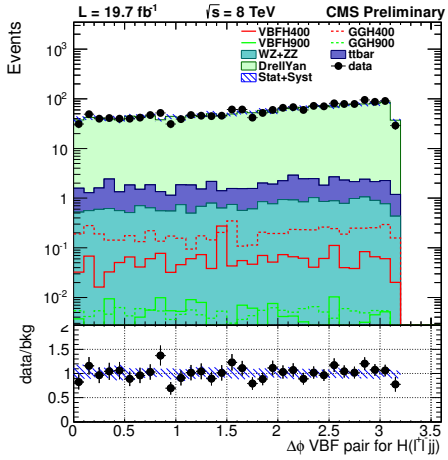
Figure 4.10: (a, c, e) Kinematic variables of the leading selected VBF jet. (b, d, f) kinematic variables of the second selected VBF jet, in the sidebands.



(a)



(b)



(c)

Figure 4.11: (a) Mass, (b) $|\Delta\eta|$ and (c) $|\Delta\varphi|$ of the VBF jet pair in the sidebands.

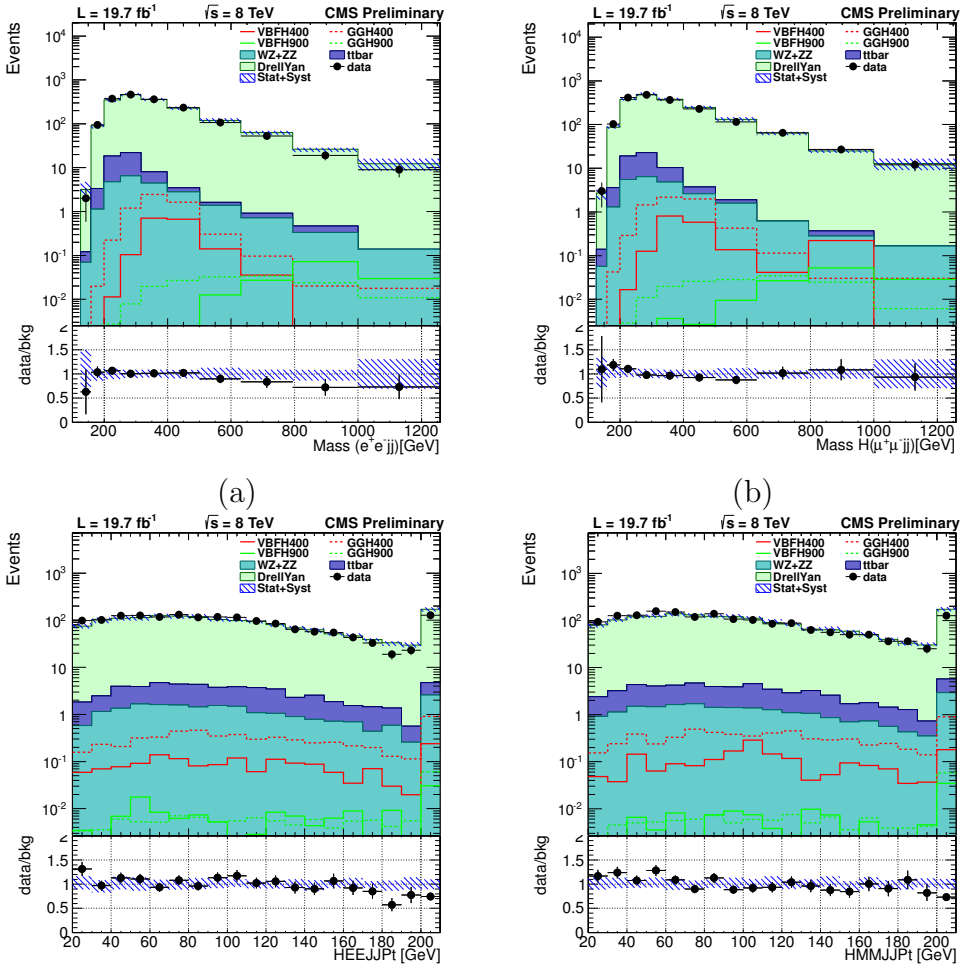
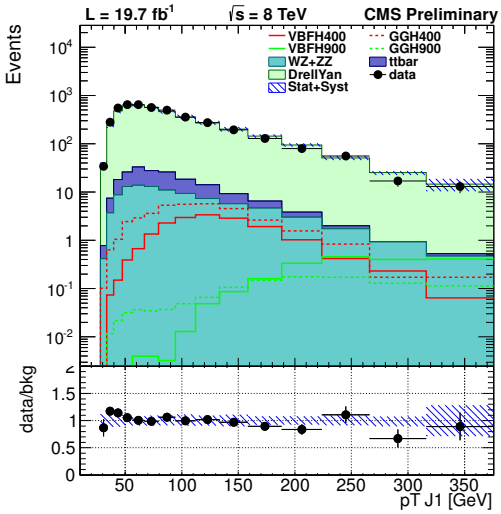
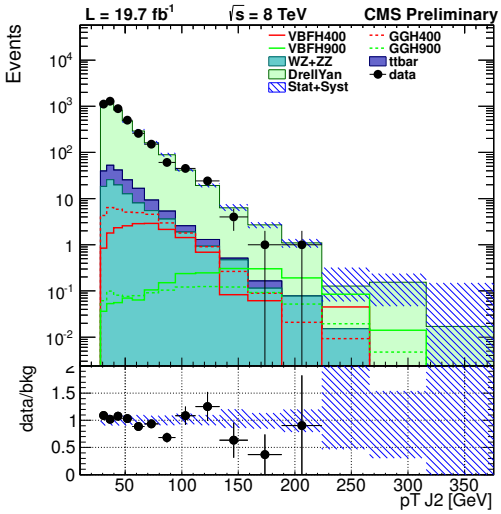


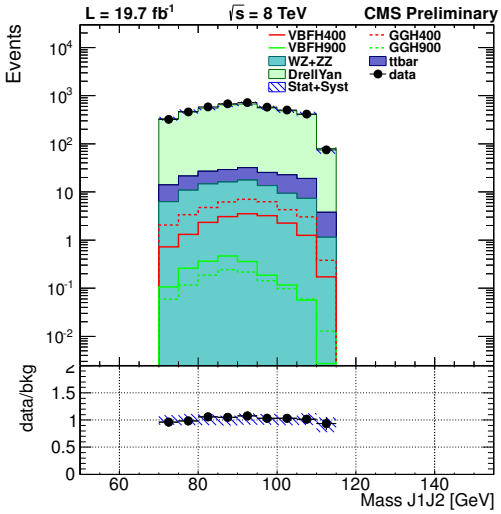
Figure 4.12: (a,b) Higgs mass and (c,d) Higgs p_T for events with a di-electron or a di-muon in the sidebands.



(a)



(b)



(c)

Figure 4.13: Hadronic Z plots in the signal region. (a) Leading jet p_T, (b) second jet p_T, (c) di-jet mass. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.

Chapter 5

$H \to ZZ \to l^+l^-q\bar{q}$ signal extraction

5.1 The Likelihood Discriminant (LD)

In the decay chain where a spin-0 boson (i.e. the Higgs boson) decays into a pair of spin-1 bosons (i.e. the Z bosons), each of them furtherly decaying into a pair of fermions, a precise spin correlation is observed in the angular distributions of the final state objects. Conversely, such a correlation is not present in background events, and a different angular distribution is expected. This topological difference can be exploited to build an angular discriminant capable to distinguish between signal and background.
Three helicity angles (θ_1, θ_2 and Φ) and two production angles (θ^* and Φ_1, see Fig. 5.1) can be used to fully describe the kinematics of the $H \to ZZ \to l^+l^-q\bar{q}$ process in the Higgs boson rest frame. Furthermore, since a Higgs-like boson carries no spin, the angular distribution of its decay products is independent of the production mechanism.
Kinematic selections exploiting these five angular observables, as described in [112, 113], increase the discriminating power between signal and background. The variables are only weakly correlated with the Higgs mass and with the mass of the two Z bosons, and also with the longitudinal and transverse momenta of the boson candidate, and they provide a good discriminating power between signal and background.
The angular likelihood discriminant (LD) is built based on the probability ratio of the signal and background hypotheses, as described in [113].
Event by event, the probability densities of the five helicity angles for signal and background, $\mathcal{P}_{sig}$ and $\mathcal{P}_{bkg}$ are built, as the matrix elements for the hypotheses of signal or background calculated at leading order. Then, LD is built as the ratio

$$\mathrm{LD} = \frac{\mathcal{P}_{sig}}{\mathcal{P}_{sig} + P_{bkg}} \tag{5.1}$$

By definition, signal events are most likely to have LD values close to one while background events are most likely to have values closer to zero. Distributions of the angular LD for signal and background are shown in Fig. 5.2. The signal probability distribution is a correlated five-dimensional angular parameterization multiplied by empirically determined polynomial acceptance functions from simulation that describe non-uniform reconstruction efficiencies in the detector. The background distribution is extracted from simulations as a product of five one-dimensional distributions, i.e. neglecting correlations.

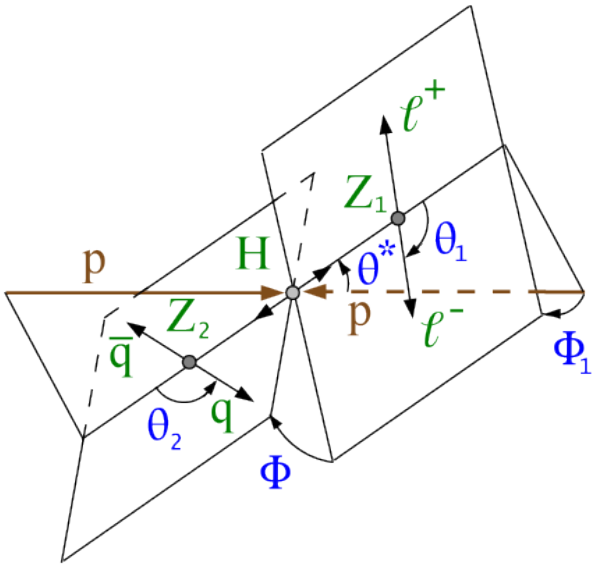


Figure 5.1: The five angles used to build the Likelihood Discriminant (LD).

Both signal and background parameterizations are given as functions of m_{ZZ}.
More in detail, the product of the ideal, fully correlated distribution which is derived in [113] and a set of four one-dimensional acceptance functions is calculated:

$$\mathcal{P}_{sig} = \mathcal{P}_{ideal}(\theta^*, \theta_1, \theta_2, \Phi, \Phi_1; m_{ZZ}) \cdot \mathcal{G}_{\theta^*}(\theta^*; m_{ZZ}) \cdot \mathcal{G}_{\theta_1}(\theta_1; m_{ZZ}) \cdot \mathcal{G}_{\theta_2}(\theta_2; m_{ZZ}) \cdot \mathcal{G}_{\Phi_1}(\Phi_1; m_{ZZ}) \tag{5.2}$$

where the four acceptance functions $\mathcal{G}_{\theta^*}$, $\mathcal{G}_{\theta_1}$, $\mathcal{G}_{\theta_2}$ and $\mathcal{G}_{\Phi_1}$ have been obtained empirically from fits to 2011 simulation at $\sqrt{s} = 7$ TeV.
For the analyzed mass range, not significant difference is expected in the simulation at 8 TeV, therefore, the study performed on 2011 MC is used for 2012 analysis as well.
The probability distribution function for the background was approximated with a product of five one-dimensional functions:

$$\begin{aligned} &\mathcal{P}_{bkg}(\theta^*, \theta_1, \theta_2, \Phi, \Phi_1; m_{ZZ}) = \\ &\quad \mathcal{P}_{\theta^*}(\theta^*; m_{ZZ}) \cdot \mathcal{P}_{\theta_1}(\theta_1; m_{ZZ}) \cdot \mathcal{P}_{\theta_2}(\theta_2; m_{ZZ}) \cdot \mathcal{P}_{\Phi}(\Phi; m_{ZZ}) \cdot \mathcal{P}_{\Phi_1}(\Phi_1; m_{ZZ}) \end{aligned} \tag{5.3}$$

All functions were obtained empirically from fits to simulations in 2011. As for the signal, the background shapes are not expected to change from 2011 to 2012, therefore the ones obtained from 2011 simulation have been used for 2012 analysis as well.

All selected events are required to have a value of the likelihood discriminant larger than 0.2. The efficiency of this cut with respect to the kinematic cuts is between 80% and 90% for both VBF and gluon-gluon signals, while it is 60% for the background.
We have chosen an LD cut looser than the one in the central gluon-fusion analysis (0.5, see Appendix A) to gain statistics in signal, in particular for very low and very high mass values, where the LD distribution flattens, as shown in Fig. 5.3. For this analysis the final number of selected events is very low, as it will be shown later; this looser cut allows us to regain a bit of statistics, while we use the additional VBF discriminator to tame the background.

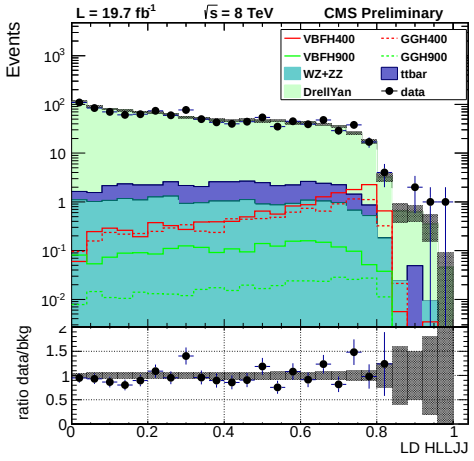


Figure 5.2: Likelihood Discriminant (LD) for events passing the kinematic cuts, with final state in electrons or muons. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.

5.2 The VBF discriminant

In order to suppress background and to extract signal events, the idea is to heavily exploit the peculiar VBF topology.
Simulated events from VBF signal and from background samples having at least one Higgs candidate (as described in Sec. 4.4) and passing a selection cut on the LD discriminant LD$>$ 0.2 have been used to train a multivariate analysis (MVA) in which the shape of VBF-related variables is used to discriminate between signal and background events. Multivariate classification methods based on automatic learning are very powerful tools when few signal events have to be discriminated among a huge background. A *Boosted Decision Tree with Gradient boost* (BDTG) was chosen, and TMVA[1] [114] was used for the BDTG training and testing.

5.2.1 Multivariate techniques: the Boosted Decision Tree

A *decision tree* is a binar-tree-structured classifier similar to the one sketched in Fig. 5.4. Repeated left/right (yes/no) decisions are taken on one single variable at a time until a stop criterion is fulfilled. In fact, starting from the root node, a sequence of binary splits using the discriminating variables x_i is applied to the data. Each split uses the variable that at this node gives the best separation between signal and background when being cut on. The same variable may thus be used at several nodes, while others might not be used at all. The phase space is split this way into many regions that are eventually classified as signal or background, depending on the majority of training events that end up in the final leaf node. The *boosting* of a decision tree [115] extends this concept from one tree to several trees which form a *forest*. The trees are derived from the same training

[1] A Toolkit for a Multivariate Analysis. It's a package for multivariate analysis developed for the ROOT framework.

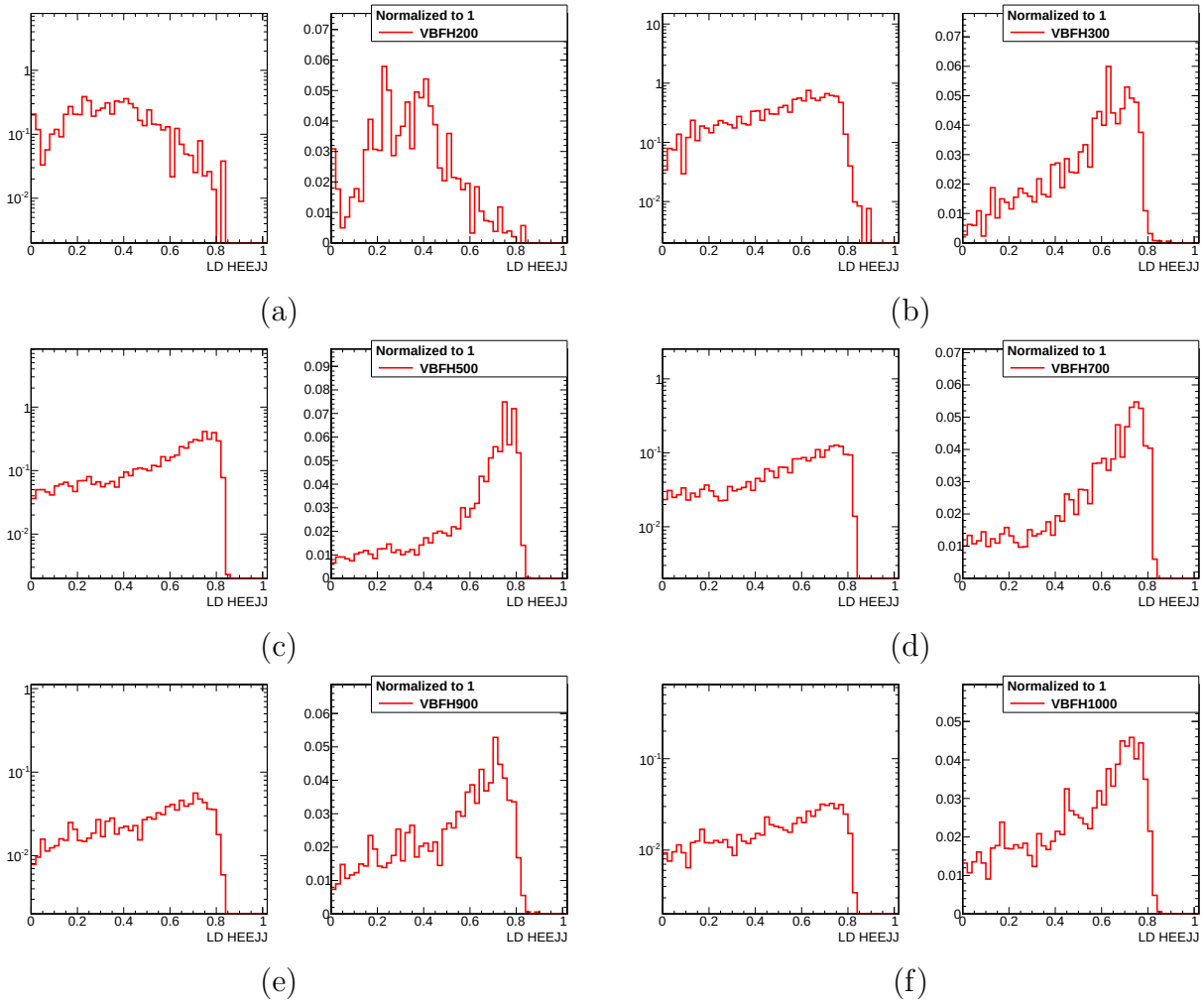


Figure 5.3: Likelihood discriminant in signal samples, after kinematic cuts, in logarithmic (left) and linear (right) scale, for different Higgs mass points (a) 200 GeV, (b) 300 GeV, (c) 500 GeV, (d) 700 GeV, (e) 900 GeV, (f) 1000 GeV.

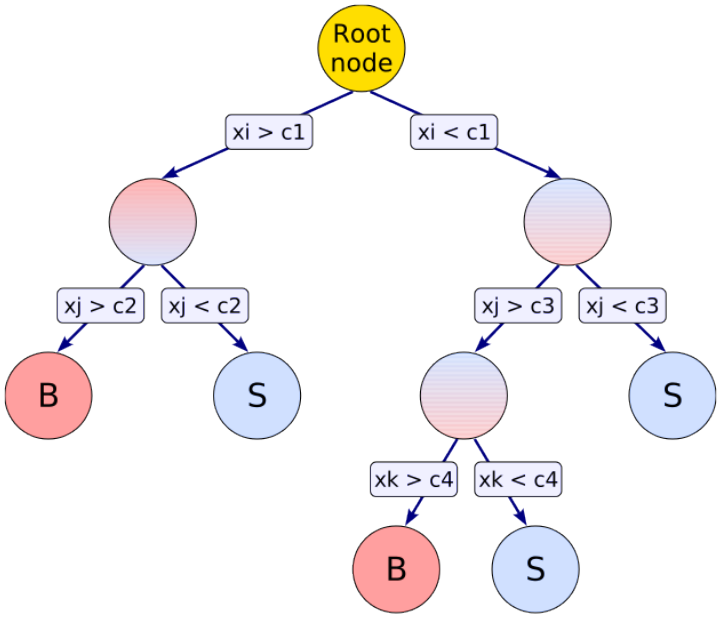


Figure 5.4: Schematic view of a decision tree. The leaf nodes at the bottom end of the tree are labeled "S" for signal and "B" for background depending on the majority of events that end up in the respective nodes.

ensemble by reweighting events, and are finally combined into a single classifier which is given by an average of the individual decision trees. *Boosting* stabilizes the response of the decision trees with respect to fluctuations in the training sample and is able to considerably enhance the performance with respect to a single tree. Formally, if $h_i(\mathbf{x})$ is the response of a single decision tree to the vector of input observables $\mathbf{x}$, such that $h(\mathbf{x}) = +1$ or -1 respectively for signal and background, the response of the *boosted decision tree* $y_{BDT}(\mathbf{x})$ is given by:

$$y_{BDT}(\mathbf{x}) = \sum_{i}^{N_{\text{forest}}} w_i \cdot h_i(\mathbf{x}) \tag{5.4}$$

which is the weighted sum on all the N trees in the forest, w_i being some appropriate weights. The different *boosting* procedures available provide different methods to determine these weights w_i. All the trees in the forest are produced from the same *training* sample, after a procedure of reweighting of single events. The *boosting* procedure is iterative, each tree is built taking the previous as input, by differently weighting the events correctly classified from those "badly" classified. The boosting algorithm is based on the minimization of the so-called *loss function*, which is a function giving a measure of the "deviation" for badly classified events. Starting from the $k-th$ tree, the $(k+1)-th$ tree of the forest is built in such a way that the loss function is minimized.
The boosting procedure is fully determined by the choice of the loss function. For the *Gradient boosting*, the loss function is defined as:

$$L(y_{BDT}) = \ln\left(1 + e^{-2y_{BDT}(\mathbf{x})\ y}\right).$$

where y is a number describing the real signal-like or background-like nature of the single event in the training sample (+1 for true signal events, −1 for true background events).

A typical multivariate classification analysis consists of independent phases:

- the *training*, which is the process that defines the splitting criteria for each node. In other words, it's the phase when the automatic learning takes place. A training sample is used for this scope;
- the *testing* phase, when the results of the automatic learning are tested on an independent sample;
- the *application* phase, when the chosen methods are used for the concrete classification or regression problem they have been trained for.

Overtraining occurs when a machine learning problem has too few degrees of freedom, because too many model parameters of an algorithm were adjusted to too few data points. Boosted decision trees usually suffer from at least partial overtraining, owing to their large number of nodes. Overtraining leads to a seeming increase in the classification performance over the objectively achievable one, if measured on the training sample, and to an effective performance decrease when measured with an independent test sample. Various method-specific solutions to counteract overtraining exist. For example, the number of nodes in boosted decision trees can be reduced by removing insignificant ones ("tree pruning"). Also the number of trees in the forest can be reduced, and so on. With no overtrainig, the decision tree output evaluated on a test sample and the one evaluated on the training sample should have the same distribution. The Kolmogorov-Smirnov test[2] can be used to give an estimate of the consistency of these two distributions.

5.2.2 A multivariate discriminant for the VBF topology

The variables used to build the multivariate discriminant were (Fig. 5.5):

- $\Delta\eta$ between the VBF jets, VBF jets invariant mass, $\Delta\varphi$ between the VBF jets;
- Leading VBF jet energy, p_T and pseudorapidity;
- Second VBF jet energy, p_T and pseudorapidity;

The idea behind this choice of variables is that they are able to put into evidence the topological peculiarity of VBF events and at the same time, for signals, are relatively independent of the mass of the Higgs boson. The signal used in the training stage is the 400 GeV VBF signal. The training was performed at one fixed mass point because the variables used in the discriminator are only weakly correlated with the Higgs signal mass, so it is not expected to have a different behaviour depending on the mass point chosen;

[2]The Kolmogorov-Smirnov (K-S) test [116] is used to decide if a sample comes from a population with a specific distribution. It has some advantages with respect to the χ^2 goodness-of-fit test. For example, the distribution of the K-S test statistic itself does not depend on the underlying cumulative distribution function being tested. Furthermore, it does not depend on an adequate sample size to be valid.

moreover, the selection was meant to be independent of the hypothesized signal mass. All the backgrounds have been used. One half of the signal and background samples is used for the training and the remaining half for the test. The gluon fusion signal was not used in the training as either signal or background. This choice is motivated by the fact that, while this analysis aims at putting into evidence the VBF topology, the gluon fusion production is still a signal for us. On the other hand we don't want to spoil the peculiarity of the VBF signature by using the gluon fusion as a signal in the training.

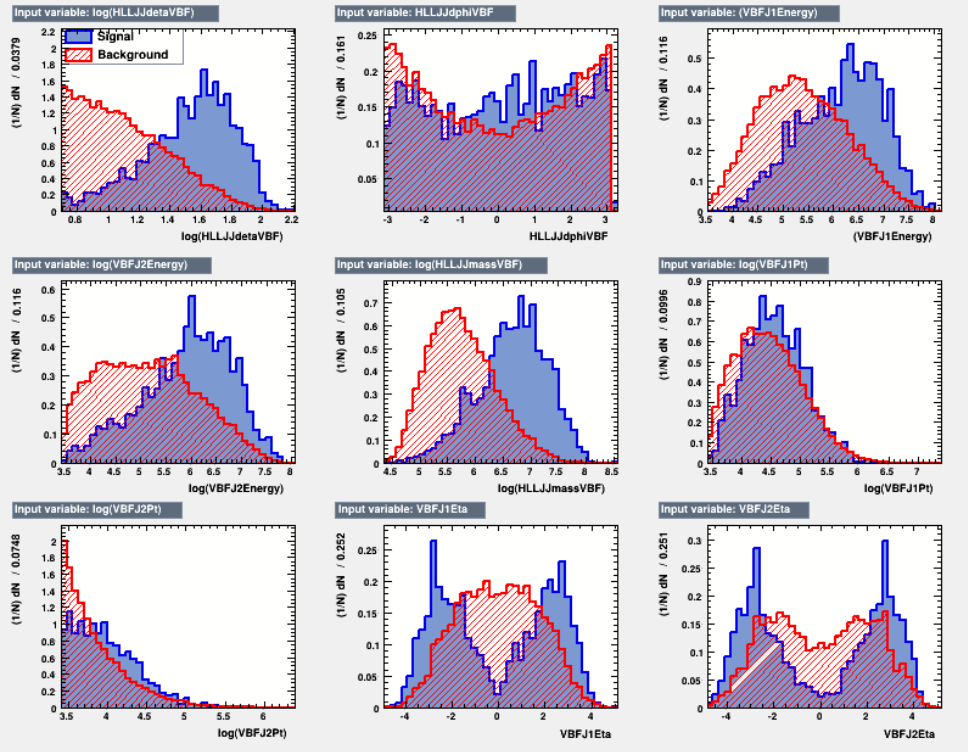
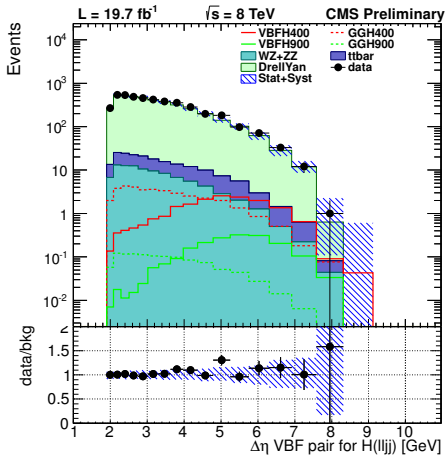


Figure 5.5: Distribution of the nine variables used to build the multivariate discriminant, in signal and background.

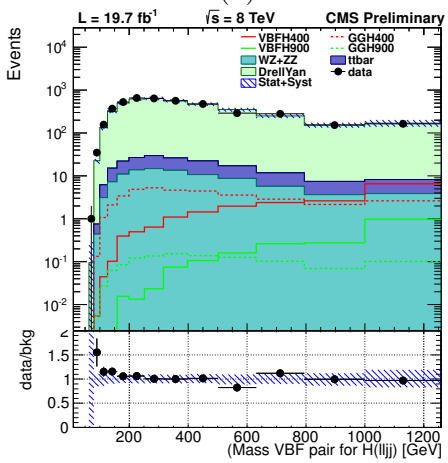
In Fig. 5.6, 5.7, 5.8 the variables are shown for data and MC samples after the LD cut. A good agreement between data and simulation can be observed.

In Fig. 5.9 the distribution of the BDTG output discriminator is shown for both signal and background test samples. A good discriminating power between signal and background can be seen. A cut on this BDTG discriminator is applied, at 0.4. This cut has an efficiency between 70% and 80% for VBF signals and around 20% for gluon fusion signals. The efficiency on backgrounds is 10%. These efficiencies are relative to the number of events that pass the kinematic selection and the LD cut.

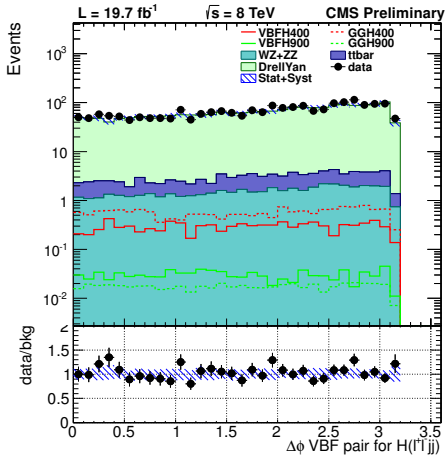
The discriminator is also shown in Fig. 5.10, where the shape for VBF and gluon fusion signals at 400 GeV and 900 GeV are compared with the background. The plot shows that gluon fusion signals are similar in shape to the background, and also that the shape of the discriminator depends relatively weakly on the mass of the signal.



(a)

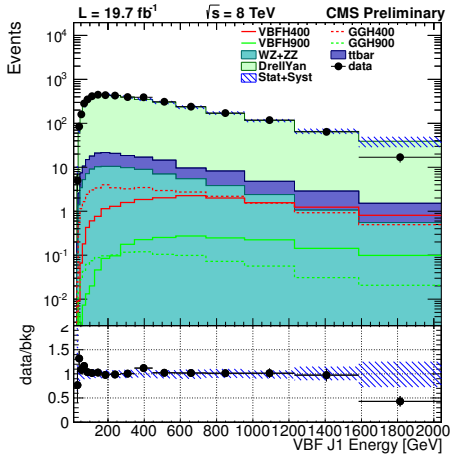


(b)

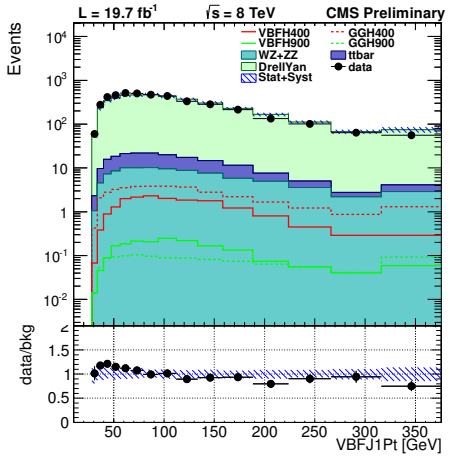


(c)

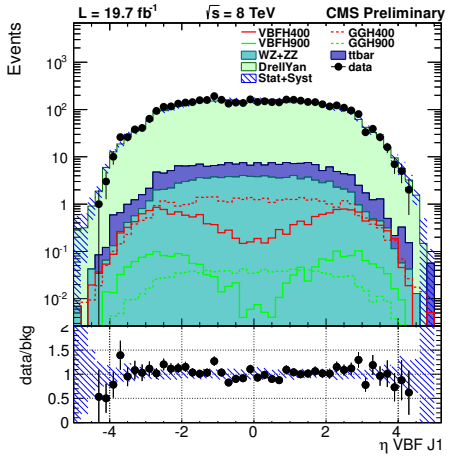
Figure 5.6: VBF pair variables for LD>0.2. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.



(a)

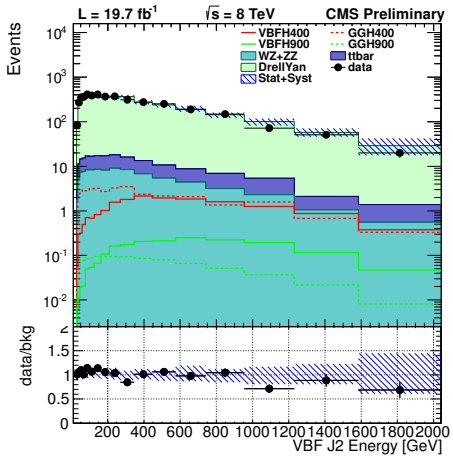


(b)

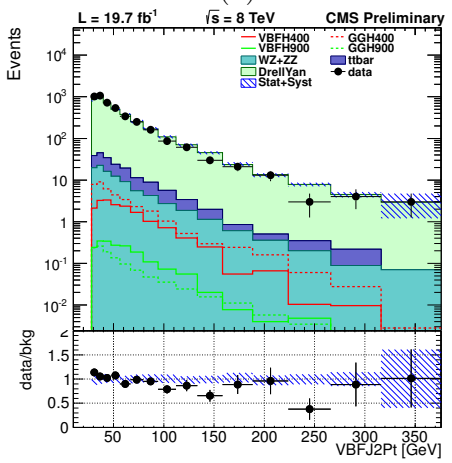


(c)

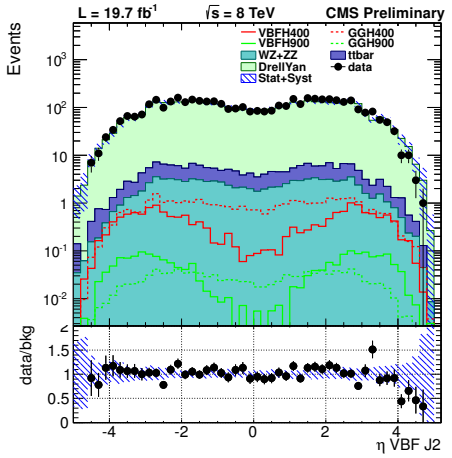
Figure 5.7: Leading VBF jet variables for LD>0.2. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.



(a)

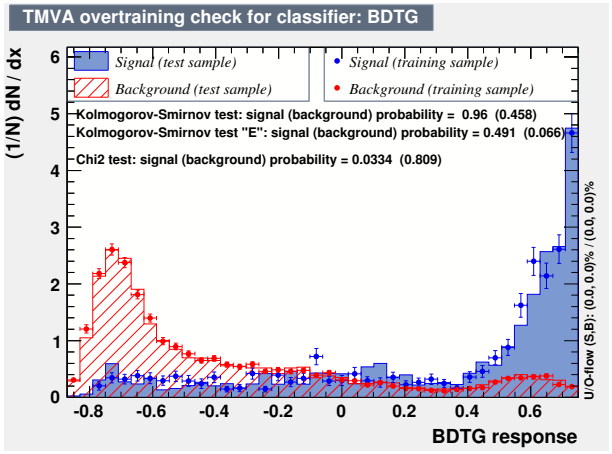


(b)

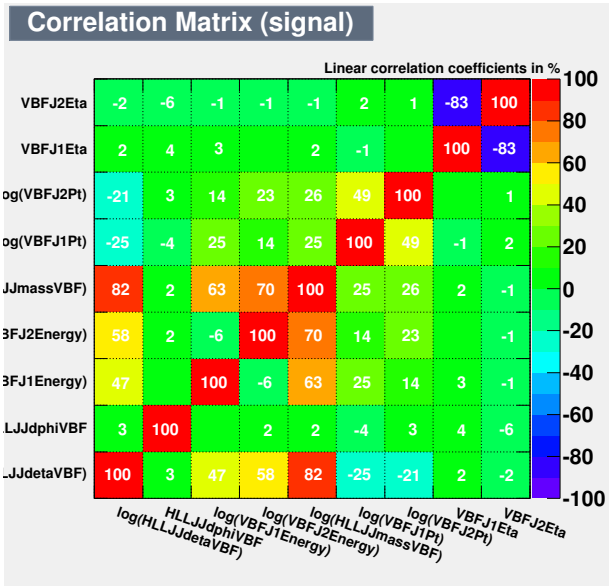


(c)

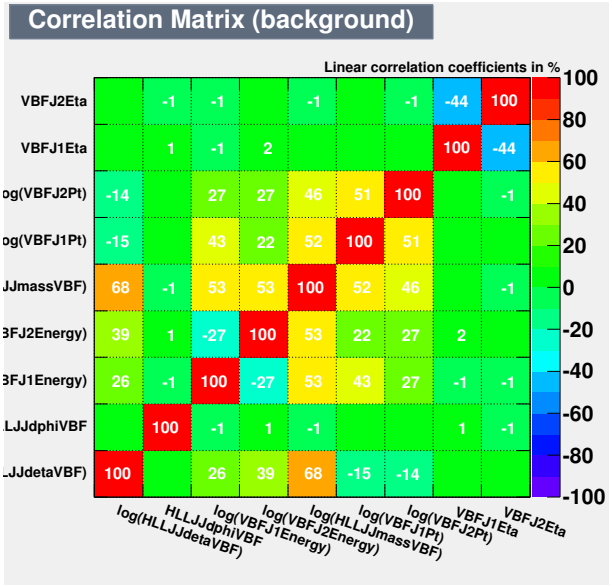
Figure 5.8: Second VBF jet variables for LD>0.2. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.



(a)



(b)



(c)

Figure 5.9: (a) BDTG output discriminant for signal and background test samples. Results for the Kolmogorov-Smirnov test and χ^2 test for signal and background are reported. In (b) and (c) the correlation coefficients between input variables are shown, for signal and background respectively.

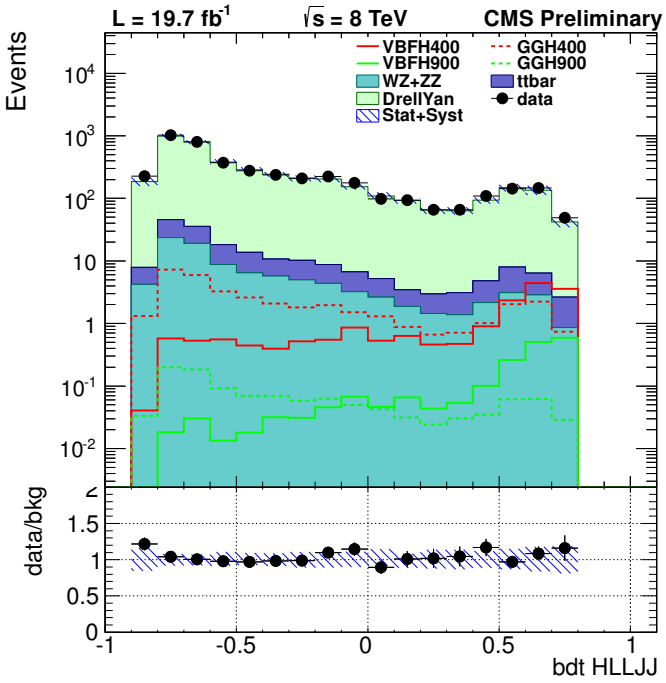


Figure 5.10: BDTG output discriminant for VBF and gluon fusion signals at 400 GeV and 900 GeV compared with the background. Events are in the signal region, passing the LD cut. The hatched blue band represents the combination of statistical and systematic error on the MC prediction.

Chapter 6

Systematic uncertainties

In this section the main sources of systematic uncertainties are described and the event yields for signal and background passing the full selection are reviewed. Some of the systematics are accounted for as global normalization factors, others produce alternative template shapes to be used in the shape analysis described in 7.1.

6.1 Theoretical uncertainties on signal

The main sources of theoretical uncertainty on the VBF and gluon fusion signals are the following:

- *Scale and PDF uncertainties on the total rate and PDF uncertainty on acceptance.* Both the VBF and inclusive gluon fusion cross sections are known at Next to Leading Order (NLO) with their perturbative accuracy [117].
  The scale uncertainties account for both the renormalization scale and the factorization scale and give a normalization uncertainty on the cross section, which has been calculated by the Higgs Cross Section Working Group. They give an effect of 6%.
  Then, uncertainties linked to the parameterization of the PDFs as a function of $x_{Bjorken}$ and Q (see Sec. 1.5.1) has also been taken into account, and give rise to normalization and acceptance effects. The effect is of the order of 3%. The PDF4LHC prescriptions have been followed. Parton distribution functions (PDFs) are obtained from global fits to data from deep-inelastic scattering, Drell-Yan and jet data. Some different working groups (for our purpose: MSTW08, CTEQ6.6 and NNPDF2.0 PDFs) produce publicly available PDFs using different data sets and analysis frameworks. They also use different parameterizations for the PDFs. For each group (parameterization), the correlation matrix for the set of parameters used (of the order of 30-40) is made diagonal, in order to obtain uncorrelated parameters; this matrix contains central values for the transformed parameters. Associated to these central values are the corresponding uncertainties, estimated from the fit procedure. Then, when evaluating the PDF as a function of these parameters, for a single event, one can give a central value for the PDF at the specific $x_{Bjorken}$ and Q values in the event, but also two (upper and lower) variations, evaluated as the

combination (i.e. the squared sum) of the uncertainties on the single parameters. All this holds for the single working group. When combining the results for the three different groups, event by event, the envelope is taken among the three sets of values plus errors, computed this way:

$$\begin{aligned} U &= \max_i \left(PDF_i + \sigma(PDF)_i^{(+)} \right) \\ L &= \min_i \left(PDF_i - \sigma(PDF)_i^{(-)} \right) \end{aligned} \tag{6.1}$$

with i =CTEQ, MSTW, NNPDF. As a central value, the midpoint of this envelope is used. This definition of an envelope has been proposed because the deviations between the predictions are currently greater than their uncertainties would strictly suggest. Then, on an event-by-event basis, two weights can be evaluated, as the ratio between the varied (up and down) and the central value for the PDF. By reweighting each event, a normalization and shape uncertainty can be evaluated, as the systematics related to the choice of the functional form for the PDF's.
The gluon fusion contribution in this analysis poses additional problems because what contributes to our final state is gluon-gluon Higgs plus two jets. For this reason we use special inclusive gluon fusion Higgs samples produced with the POWHEG-MINLO program, where H+2 jets is simulated at NLO accuracy. This procedure takes the same kind of systematic uncertainties as the event generation with POWHEG.

- *Uncertainties on the $H \to ZZ$ branching ratio*, varying the values of the branching ratio according to the error given by the Cern Yellow Report [117]. It gives an effect of 2%.

6.2 Theoretical uncertainties on background

The following theoretical uncertainties on the backgrounds are considered:

- *Cross Section Uncertainty.* The Z+jets normalization is left floating in the limit, in such a way that the limit tool will choose the best value that fits both the signal and sideband regions.

- *Residual Shape Uncertainties.* This uncertainty is meant to cover residual data-MC disagreement that might be due either to limited MC statistics or to residual mismodeling on the background. The contribution of this uncertainty on the mass template is evaluated in the sidebands by taking the relative difference of data and MC and fitting it to smoothen out the statistical fluctuations. We have fitted the data-MC relative difference with a straight line, which is shown as the red line in Fig. 6.1. It can be observed that the two parameters of the fit (p_0, p_1, respectively the coefficients for the zero and first degree terms) are negatively correlated. The steepest variation has been taken, obtained by adding 1 σ to p_0 and by subtracting 1 σ to p_1, where σ here refers to the error from the fit. This corresponds to the steep blue line in Fig. 6.1. Finally, given that the data points do not show a strong

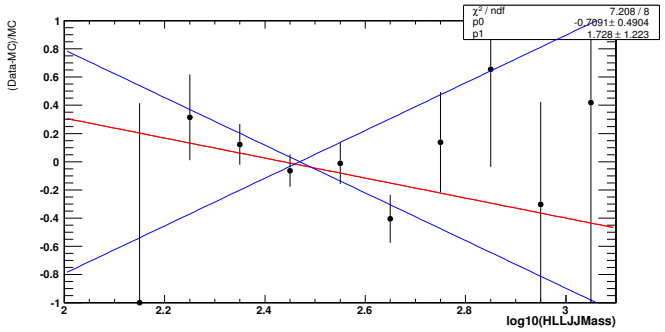


Figure 6.1: Relative data-MC difference in the sidebands. The red line is a straight line fit. The two blue lines are the shape used to derive a residual up and down variation in the signal region.

tendency, and to have a conservative estimate of the uncertainty, the curve obtained as a specular reflection around the x axis has been taken as the second extreme of the systematic variation. The two blue lines in Fig. 6.1 are thus used to estimate two alternative shapes on the mass distribution in the signal region. It would finally be noted that this uncertainty is purely a shape uncertainty, the normalization of the two variations is forced to be the same as the central value. The two central mass distribution and the two alternatives are shown in Fig. 6.2.

6.3 Jet energy scale and jet energy resolution

Jet energy scale (JES) and jet energy resolution (JER) uncertainties affect both signal and background and they are evaluated by varying the 4-momentum of each jet according to standard recipes described respectively in [118] and [99]. Also the angular variables like LD are recomputed when changing the JES and the JER. These uncertainty sources result in alternative template shapes for the up and down JES variations and for the up and down JER variations.
The effect on the normalization is $O(10\%)$ for JES systematics and $O(1\%)$ for JER ones.

6.4 Minimum Bias cross section

The pileup reweighting procedure described in Sec. 4.7 is affected by the uncertainties on the knowledge of the Minimum Bias cross section. The reweighting procedure has been performed using the central value of the Minimum Bias cross section (69.30 mb) and then repeated using two variations (65.84 mb and 72.77 mb). This uncertainty source results in two alternative template shapes. The effect on the normalization is $O(1\%)$.

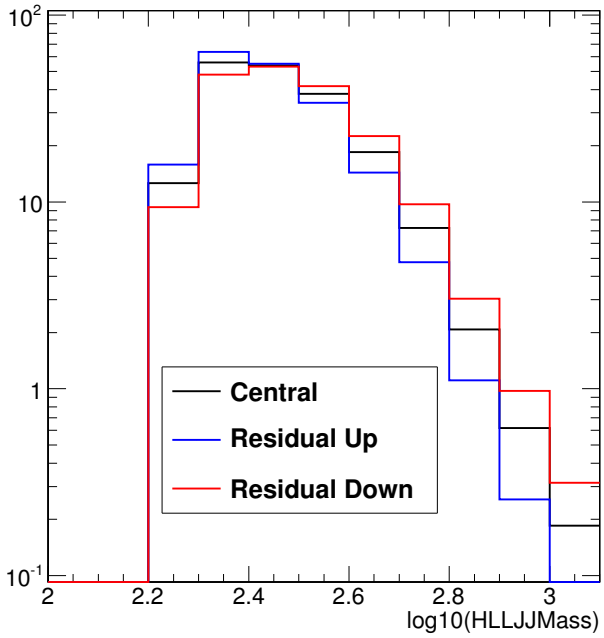


Figure 6.2: Invariant mass plot in the signal region used as template in the limit. Central and residual variation up and down are shown.

6.5 Electron and muon trigger, identification and isolation efficiencies

Lepton trigger, identification and isolation efficiencies are the same as the ones in the gluon fusion main analysis, as well as the systematic uncertainties affecting these values. Efficiencies are computed using the tag-and-probe technique[1] , where one lepton from an inclusive sample of Z decays serves as a tag and efficiency for the other lepton is calculated. For muons the total normalization uncertainty is 1.8%, with contributions from the trigger (1.7%), identification (0.5%), and isolation (0.2%). For electrons the total normalization uncertainty is 2%, dominated by identification ($\sim$ 2%) and a much smaller contibution from the dielectron trigger efficiency.

6.6 Boosted Z-tagging

The selection of boosted jets from Z decays through the cut on the N-subjettiness observable τ_{21} and through jet pruning techniques is referred to as "*boosted Z tagging*". The

[1]The tag-and-probe (or Tag & Probe, T&P) technique is a data driven technique used to calculate efficiencies exploiting a very well-known mass resonance (i.e. J/ψ, Υ or Z). The "tag" is i.e. a muon or electron that pass very tight selection criteria, while the "probe" is requested to pass looser criteria. The known resonance is exploited to have a constraint on the invariant mass of the lepton pair to be chosen as T&P pair. The efficiency is calculated on the probe lepton. See [119] for a more detailed description of the technique.

boosted Z tagging efficiency has been estimated on simulated samples. A data-to-MC scale factor for boosted W tagging of 0.91 ± 0.10 had been determined by previous analyses using top-enriched control samples [91, 107]. Given that the cuts and kinematics involved are almost equivalent, this scale factor is assumed to be appropriate also for correcting the boosted Z tagging efficiency in our analysis and is applied to the signal and diboson background predictions. So, the estimated uncertainty on the data/MC scale factor for boosted Z tagging is about 10% of the number of events with a boosted Z candidate.

6.7 Luminosity uncertainty

As discussed in Sec. 2.1, the instantaneous luminosity $\mathcal{L}$ measures the mean number of collisions per second, whereas the integrated luminosity L measures the integral over time, and it is representative of the amount of data collected. The luminosity is determined from the counting rates as measured by the luminosity detectors. In CMS two detectors are exploited for the luminosity measurement: the forward hadronic calorimeter HF (Hadron Forward), and the silicon pixel detector [120], characterized by very low occupancy and excellent stability over time. The pixel cluster counting (PCC) method [121] used exploits the very large number of pixels in the inner part of the CMS detector. The number of pixels is about 7×10^7 , which means that the probability of a given pixel being hit by two different tracks from the same bunch crossing is exceedingly small. The average number of clusters per minimum bias interaction is about 200 and the average number of pixels per cluster is about 5. Thus, even at pile-up values of 100, the fraction of occupied pixels is at the per-mil level. As a result, the number of hit pixel clusters per crossing is expected to be a very linear function of the number of interactions per crossing and therefore a very good measure of luminosity.
The data luminosity uncertainty results in a 2.6% normalization uncertainty on all templates for signal and background.

6.8 Monte Carlo statistics

One important source of uncertainty is the limited statistics of the MC that we use to derive the mass templates (see 7.1). This results in a global normalization uncertainty of the order of $2 - 14\%$.

6.9 Event yields and systematics contribution

Expected signal event yields and global normalization uncertainties for the different sources of uncertainty are listed in Tab. 6.1. The global luminosity uncertainty is not shown in the table. This table should just give an idea of the importance of the different sistematics. As mentioned above not all the systematics are treated as mere normalization uncertainties, some are shape uncertainty, event though also for these ones only the integrated normalization uncertainty is shown in the table.

Expected event yields and normalization uncertainties for background are listed in Tab. 6.2.

Process	Exp. ev.	JES	JER	σ MinBias	MCstat	PDF+Scale Rate	PDF Acc.	Boosted-Z tag
200 GeV								
ggH($\to ee$)	0.8	15.1%/-9.0%	2.2%/-0.4%	-1.8%/1.4%	13.9%/-13.9%	13.4%/-14.5%	4.2%	0.0%
VBF H($\to ee$)	2.4	7.0%/-9.0%	-1.9%/0.6%	0.2%/-1.5%	7.2%/-7.2%	2.8%/-2.8%	2.2	0.0%
ggH($\to \mu\mu$)	1.0	15.0%/-20.0%	-1.9%/-2.6%	-2.4%/3.3%	13.9%/-13.9%	13.4%/-14.5%	3.1%	0.0%
VBF H($\to \mu\mu$)	3.0	16.7%/-6.9%	3.1%/3.7%	-1.1%/1.8%	7.2%/-7.2	2.8%/-2.8%%	2.8%	0.0%
250 GeV								
ggH($\to ee$)	1.6	18.0%/-11.9%	2.2%/-1.9%	-0.3%/-0.8%	7.9%/-7.9%	13.2%/-14.1%	4.3%	0.0%
VBF H($\to ee$)	6.2	7.3%/-7.1%	1.8%/-0.7%	-1.1%/0.2%	4.1%/-4.1%	2.8%/-2.7%	1.7	0.0%
ggH($\to \mu\mu$)	2.1	9.2%/-13.4%	0.1%/-2.9%	0.6%/0.1%	7.9%/-7.9%	13.2%/-14.1%	3.5%	0.0%
VBF H($\to \mu\mu$)	8.5	6.8%/-7.8%	-1.2%/-0.9%	-0.5%/1.2%	4.1%/-4.1	2.8%/-2.7%%	1.3%	0.0%
300 GeV								
ggH($\to ee$)	2.1	13.2%/-7.4%	9.5%/-2.8%	-1.5%/0.4%	6.8%/-6.8%	13.4%/-14.0%	2.6%	0.0%
VBF H($\to ee$)	6.8	6.2%/-3.4%	-0.2%/0.8%	-1.3%/0.3%	3.4%/-3.4%	2.8%/-2.8%	1.5	0.0%
ggH($\to \mu\mu$)	2.9	9.5%/-5.6%	5.5%/-0.9%	0.5%/0.3%	6.8%/-6.8%	13.4%/-14.0%	9.0%	0.1%
VBF H($\to \mu\mu$)	8.6	5.0%/-6.3%	-1.7%/0.7%	0.4%/0.6%	3.4%/-3.4	2.8%/-2.8%%	1.2%	0.0%
350 GeV								
ggH($\to ee$)	2.8	10.9%/-6.3%	1.4%/-0.8%	-0.9%/0.2%	5.8%/-5.8%	13.7%/-14.1%	2.8%	0.2%
VBF H($\to ee$)	6.2	4.6%/-7.0%	-1.8%/-0.6%	-0.6%/-0.8%	3.0%/-3.0%	2.8%/-2.9%	1.7	0.0%
ggH($\to \mu\mu$)	3.6	10.3%/-12.6%	4.5%/-2.7%	-1.1%/2.5%	5.8%/-5.8%	13.7%/-14.1%	6.0%	0.2%
VBF H($\to \mu\mu$)	8.2	5.1%/-6.1%	-1.3%/0.8%	0.0%/1.1%	3.0%/-3.0	2.8%/-2.9%%	1.4%	0.1%
400 GeV								
ggH($\to ee$)	2.9	13.4%/-11.2%	5.0%/-2.4%	-1.5%/0.0%	5.4%/-5.4%	14.0%/-13.6%	3.1%	0.4%
VBF H($\to ee$)	5.4	5.0%/-5.1%	-1.3%/-0.5%	-1.0%/-0.5%	2.9%/-2.9%	2.9%/-3.2%	1.7	0.6%
ggH($\to \mu\mu$)	3.6	11.6%/-7.8%	5.0%/-4.3%	0.2%/0.6%	5.4%/-5.4%	14.0%/-13.6%	7.2%	0.6%
VBF H($\to \mu\mu$)	6.7	5.5%/-4.8%	-0.5%/0.8%	0.0%/1.1%	2.9%/-2.9	2.9%/-3.2%%	1.5%	0.5%
450 GeV								
ggH($\to ee$)	2.0	8.5%/-7.6%	2.6%/-3.8%	-2.1%/1.0%	5.2%/-5.2%	14.4%/-13.6%	3.6%	1.0%
VBF H($\to ee$)	4.2	4.4%/-3.4%	0.3%/0.8%	-1.2%/-0.4%	2.6%/-2.6%	3.1%/-3.6%	1.3	1.3%
ggH($\to \mu\mu$)	2.5	11.2%/-9.9%	3.2%/-1.0%	0.2%/1.0%	5.2%/-5.2%	14.4%/-13.6%	6.8%	0.6%
VBF H($\to \mu\mu$)	5.5	5.0%/-4.7%	-0.8%/-0.2%	0.7%/0.9%	2.6%/-2.6	3.1%/-3.6%%	1.8%	1.2%

Process	Exp. ev.	JES	JER	σ MinBias	MCstat	PDF+Scale Rate	PDF Acc.	Boosted-Z tag
500 GeV								
ggH($\rightarrow ee$)	1.4	14.3%/-6.0%	4.7%/-0.9%	-1.1%/-0.3%	4.3%/-4.3%	14.9%/-13.6%	3.6%	1.8%
VBF H($\rightarrow ee$)	3.6	4.5%/-3.5%	-0.4%/1.5%	-1.0%/-0.7%	2.6%/-2.6%	3.4%/-3.7%	0.8	3.0%
ggH($\rightarrow \mu\mu$)	1.8	8.9%/-10.1%	2.6%/-2.6%	-0.3%/2.1%	4.3%/-4.3%	14.9%/-13.6%	4.6%	2.0%
VBF H($\rightarrow \mu\mu$)	4.5	3.3%/-4.4%	-1.0%/0.6%	0.5%/1.1%	2.6%/-2.6	3.4%/-3.7%%	1.9%	2.9%
550 GeV								
ggH($\rightarrow ee$)	0.89	23.9%/4.3%	19.1%/8.7%	-1.2%/-0.3%	4.1%/-4.1%	15.2%/-13.8%	9.0%	3.6%
VBF H($\rightarrow ee$)	2.9	3.3%/-4.8%	-1.0%/0.4%	-0.3%/-1.0%	2.5%/-2.5%	3.5%/-4.1%	0.9	4.6%
ggH($\rightarrow \mu\mu$)	1.2	22.6%/-1.4%	15.5%/7.3%	-1.2%/3.1%	4.1%/-4.1%	15.2%/-13.8%	8.0%	3.1%
VBF H($\rightarrow \mu\mu$)	3.6	4.2%/-5.6%	-0.2%/-0.0%	0.1%/1.6%	2.5%/-2.5	3.5%/-4.1%%	1.1%	4.7%
600 GeV								
ggH($\rightarrow ee$)	0.75	8.8%/-7.1%	5.5%/-2.2%	-1.2%/-0.3%	4.2%/-4.2%	15.4%/-13.8%	3.2%	4.3%
VBF H($\rightarrow ee$)	2.0	4.4%/-3.7%	0.4%/1.1%	-1.3%/-0.4%	2.9%/-2.9%	3.9%/-4.3%	1.5	6.1%
ggH($\rightarrow \mu\mu$)	0.84	6.8%/-6.7%	4.0%/-2.4%	0.1%/1.4%	4.2%/-4.2%	15.4%/-13.8%	3.9%	4.3%
VBF H($\rightarrow \mu\mu$)	2.6	4.6%/-4.6%	-0.3%/-0.3%	1.1%/0.5%	2.9%/-2.9	3.9%/-4.3%%	3.2%	5.8%
650 GeV								
ggH($\rightarrow ee$)	0.52	10.3%/-9.0%	2.5%/-1.1%	-2.3%/0.5%	4.5%/-4.5%	15.6%/-14.0%	2.7%	5.1%
VBF H($\rightarrow ee$)	1.9	4.7%/-5.0%	-1.5%/1.1%	-1.3%/-0.4%	2.6%/-2.6%	4.0%/-4.8%	2.8	6.9%
ggH($\rightarrow \mu\mu$)	0.60	6.8%/-11.2%	1.0%/-2.5%	0.2%/1.3%	4.5%/-4.5%	15.6%/-14.0%	4.3%	4.8%
VBF H($\rightarrow \mu\mu$)	2.2	4.4%/-5.3%	0.1%/-0.7%	0.8%/0.7%	2.6%/-2.6	4.0%/-4.8%%	1.6%	6.7%
700 GeV								
ggH($\rightarrow ee$)	0.36	7.6%/-6.8%	2.4%/-3.3%	-2.2%/0.8%	3.8%/-3.8%	16.1%/-14.4%	4.3%	5.6%
VBF H($\rightarrow ee$)	1.5	5.7%/-3.1%	0.1%/2.0%	-1.1%/-0.7%	2.6%/-2.6%	4.4%/-5.0%	1.7	7.4%
ggH($\rightarrow \mu\mu$)	0.44	8.5%/-9.4%	4.1%/-2.9%	0.5%/1.1%	3.8%/-3.8%	16.1%/-14.4%	6.7%	4.9%
VBF H($\rightarrow \mu\mu$)	1.9	4.2%/-5.2%	-0.3%/-0.2%	0.1%/1.7%	2.6%/-2.6	4.4%/-5.0%%	1.5%	7.4%
750 GeV								
ggH($\rightarrow ee$)	0.23	7.4%/-12.1%	2.1%/-3.9%	-1.2%/-0.7%	4.0%/-4.0%	16.5%/-14.8%	2.7%	5.8%
VBF H($\rightarrow ee$)	1.3	5.2%/-2.8%	2.1%/0.7%	-1.3%/-0.8%	2.7%/-2.7%	4.7%/-5.2%	2.1	8.4%
ggH($\rightarrow \mu\mu$)	0.29	7.3%/-9.9%	2.8%/-2.0%	0.8%/0.9%	4.0%/-4.0%	16.5%/-14.8%	2.5%	6.0%
VBF H($\rightarrow \mu\mu$)	1.4	4.9%/-3.9%	0.5%/0.3%	0.7%/1.3%	2.7%/-2.7	4.7%/-5.2%%	1.9%	8.2%

Process	Exp. ev.	JES	JER	σ MinBias	MCstat	PDF+Scale Rate	PDF Acc.	Boosted-Z tag
800 GeV								
ggH($\rightarrow ee$)	0.19	7.7%/-8.8%	2.0%/-3.4%	-1.8%/-0.2%	3.6%/-3.6%	16.7%/-15.0%	2.9%	6.2%
VBF H($\rightarrow ee$)	1.0	4.3%/-6.3%	0.8%/-2.0%	-1.1%/-0.9%	2.7%/-2.7%	5.0%/-5.7%	1.0	8.3%
ggH($\rightarrow \mu\mu$)	0.20	11.6%/-7.3%	5.3%/-0.6%	0.9%/0.7%	3.6%/-3.6%	16.7%/-15.0%	3.7%	6.3%
VBF H($\rightarrow \mu\mu$)	1.2	4.8%/-6.4%	-0.2%/-0.5%	0.8%/0.8%	2.7%/-2.7	5.0%/-5.7%%	1.3%	8.2%
850 GeV								
ggH($\rightarrow ee$)	0.13	6.2%/-8.0%	2.4%/-3.5%	-0.9%/-0.9%	3.8%/-3.8%	17.1%/-15.2%	2.6%	6.4%
VBF H($\rightarrow ee$)	0.88	2.1%/-6.8%	0.1%/-3.1%	-2.2%/0.6%	2.8%/-2.8%	5.4%/-6.0%	2.1	8.6%
ggH($\rightarrow \mu\mu$)	0.17	7.8%/-7.4%	4.1%/-2.7%	1.0%/0.5%	3.8%/-3.8%	17.1%/-15.2%	3.5%	6.4%
VBF H($\rightarrow \mu\mu$)	0.97	5.2%/-5.2%	0.5%/1.6%	1.1%/0.9%	2.8%/-2.8	5.4%/-6.0%%	2.5%	8.7%
900 GeV								
ggH($\rightarrow ee$)	0.091	8.6%/-7.1%	4.4%/-2.4%	-1.4%/-0.2%	4.1%/-4.1%	17.4%/-15.5%	4.2%	6.5%
VBF H($\rightarrow ee$)	0.74	5.7%/-4.5%	-0.6%/0.4%	-1.8%/-0.3%	2.8%/-2.8%	5.6%/-6.5%	2.2	8.9%
ggH($\rightarrow \mu\mu$)	0.11	8.5%/-10.1%	3.8%/-2.9%	1.1%/0.6%	4.1%/-4.1%	17.4%/-15.5%	3.8%	6.2%
VBF H($\rightarrow \mu\mu$)	0.83	3.9%/-6.1%	1.0%/-1.0%	0.9%/0.9%	2.8%/-2.8	5.6%/-6.5%%	2.6%	8.6%
950 GeV								
ggH($\rightarrow ee$)	0.072	5.9%/-10.7%	0.7%/-3.8%	-1.5%/-0.4%	4.0%/-4.0%	17.9%/-16.1%	3.9%	6.6%
VBF H($\rightarrow ee$)	0.60	4.2%/-4.7%	-0.5%/-0.7%	-0.7%/-1.2%	2.9%/-2.9%	6.0%/-6.8%	7.0	9.0%
ggH($\rightarrow \mu\mu$)	0.080	13.4%/-7.2%	3.2%/-0.1%	0.2%/1.4%	4.0%/-4.0%	17.9%/-16.1%	3.5%	6.5%
VBF H($\rightarrow \mu\mu$)	0.67	1.4%/-6.9%	-0.2%/-0.3%	-0.3%/2.4%	2.9%/-2.9	6.0%/-6.8%%	40.3%	8.8%
1000 GeV								
ggH($\rightarrow ee$)	0.051	10.2%/-11.9%	2.2%/-3.9%	-1.3%/-0.6%	4.2%/-4.2%	18.6%/-16.7%	2.3%	6.5%
VBF H($\rightarrow ee$)	0.43	4.0%/-5.4%	0.2%/-0.1%	-2.1%/0.5%	3.1%/-3.1%	6.3%/-7.2%	8.7	9.0%
ggH($\rightarrow \mu\mu$)	0.069	11.4%/-6.4%	5.3%/-2.2%	0.2%/1.4%	4.2%/-4.2%	18.6%/-16.7%	4.7%	6.6%
VBF H($\rightarrow \mu\mu$)	0.54	5.1%/-2.6%	0.1%/2.4%	1.2%/1.2%	3.1%/-3.1	6.3%/-7.2%%	6.0%	9.3%

Table 6.1: Expected event yields for different Higgs masses after all cuts.

Process	Exp. ev.	JES	JER	σ MinBias	MCstat	PDF+Scale Rate	Boosted-Z tag
DY ($\rightarrow ee$)	193.2	12.4%/-12.6%	5.8%/-3.2%	0.1%/-0.3%	2.5%/-2.5%	-	0.2%
tt ($\rightarrow ee$)	5.8	18.4%/-11.1%	6.7%/-6.1%	-1.1%/2.1%	8.4%/-8.4%	-	0.0%
VV ($\rightarrow ee$)	4.4	11.6%/-10.2%	3.8%/-3.7%	-0.9%/0.1%	4.1%/-4.1%	-	0.3%
DY ($\rightarrow \mu\mu$)	236.8	14.8%/-13.4%	6.6%/-6.3%	-0.3%/0.7%	2.4%/-2.4%	-	0.2%
tt ($\rightarrow \mu\mu$)	8.0	12.7%/-8.7%	-2.4%/0.3%	-0.3%/1.0%	8.4%/-8.4	-	0.1%
VV ($\rightarrow \mu\mu$)	5.4	13.1%/-10.5%	5.4%/-3.4%	0.5%/0.0%	4.1%/-4.1	-	0.4%

Table 6.2: Expected event yields for backgrounds after all cuts.

Chapter 7

Statistical analysis and results for the VBF analysis

The final variable used in this analysis to distinguish the signal on top of background is the invariant mass of the two lepton two jet system m_{lljj} coming from the Higgs decay[1]. The m_{lljj} variable is significantly different for events containing a real Higgs (signal events) or just coming from random combination of leptons and jets produced in the event (background events).
Simulated events passing the selection described in Chap. 4 and Chap. 5 are used to define Monte Carlo based templates for the $lljj$ (Higgs) invariant mass. Such templates are used to perform a *shape analysis* on data, separately for the di-electron and di-muon channels.

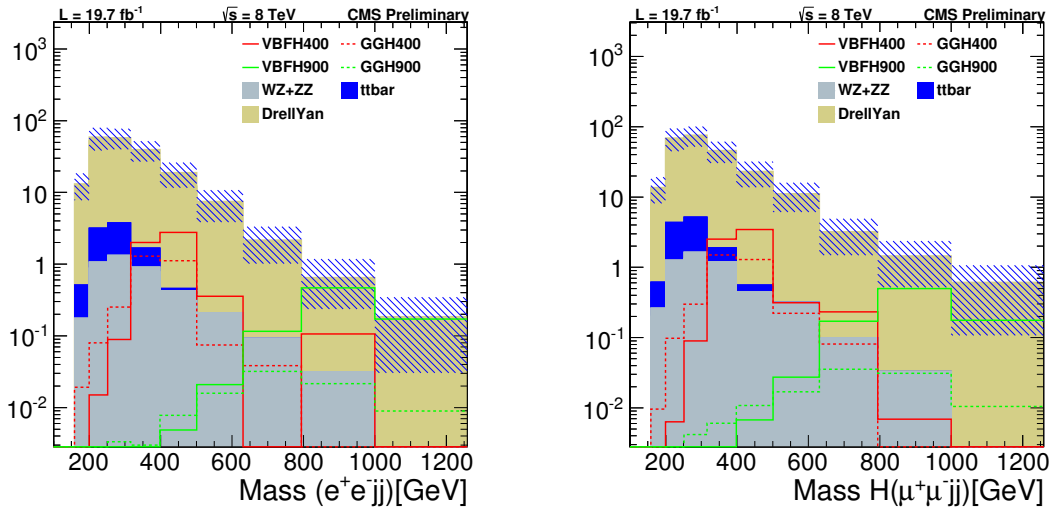
7.1 The shape analysis

In Fig. 7.1 the $2l2q$ mass distributions for simulated events only are shown, in the signal region, for the two different flavour channels. Mass distributions in data and simulated events in the signal region - after the *unblinding* procedure - are shown in Fig. 7.2. These distributions are the ones used as input for the shape analysis.

7.2 Statistical interpretation of results using the SM Higgs boson as a benchmark

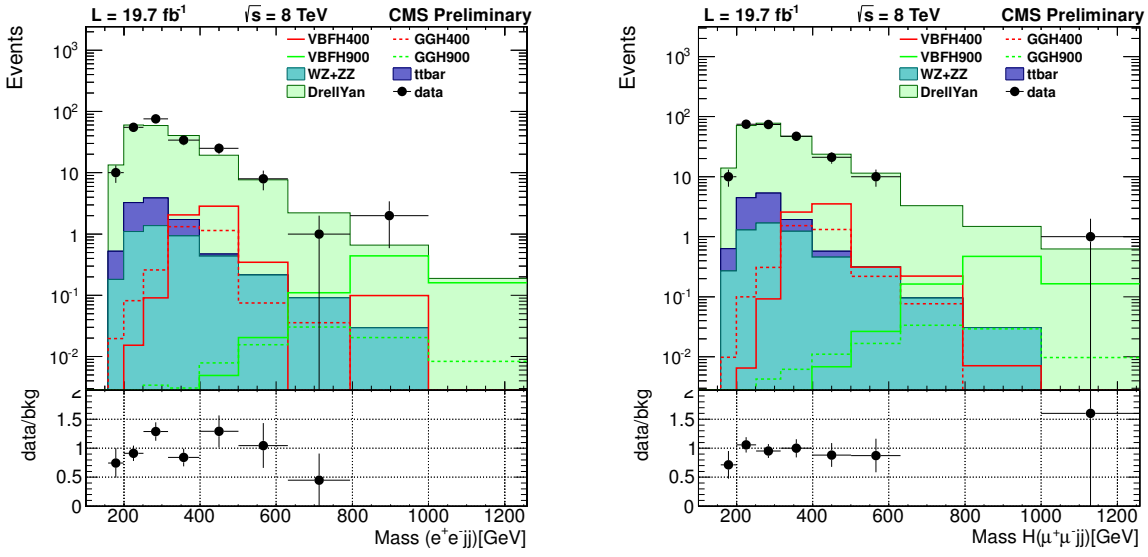
The expected and observed upper limits to the SM production cross section are determined as function of the Higgs boson mass m_H. For each mass point between 200 GeV and 1 TeV, in 50 GeV intervals, the expected limit at 95% confidence level and the one and two sigma bands have been computed using the modified frequentist method CL_s [122].

[1] In the following, the expressions "two lepton two jet system", "two jets" or the symbol m_{lljj} are used for the sake of simplicity to indicate both the case in which the hadronic Z has been reconstructed as two separate jets and the case in which it has been reconstructed as a single merged jet with two sub-jets.



(a) (b)

Figure 7.1: Higgs candidate mass, in events with the leptonic Z reconstructed in two electrons (a) and in two muons (b), in the signal region. All the selection cuts are applied including loose LD cut and the cut on the MVA discriminator. The hatched blue band represents the systematic uncertainty.



(a) (b)

Figure 7.2: Mass distributions in the signal region after all selections, including unblinded data.

Systematic uncertainties affecting both signal and background predictions are incorporated in the analysis via nuisance parameters $\boldsymbol{\theta}$. Signal and background expectations then become functions of this set of parameters: $s(\boldsymbol{\theta})$ and $b(\boldsymbol{\theta})$. Upper limits are set on the signal strength modifier $\mu = \sigma/\sigma_{SM}$. As said above, the so-called "CL_s", which is the ratio of the confidences in the signal plus background to the background hypotheses is used to set these upper limits.
The first step is to define a "test-statistic" or function of the observables and the model parameters of the known background and hypothetical signal which ranks experiments from the least to most signal-like (most to least background-like). Then, the rules for exclusion have to be defined, i.e. specify ranges of values of the test-statistic in which observations will lead to one conclusion or the other. In other words, the significance of the exclusion has to be specified, thus a confidence level for the exclusion will be quoted. The test statistic used to discriminate signal plus background from the background-only hypothesis is a likelihood ratio Q, that according to the Neyman-Pearson lemma is the most powerful discriminator. The profile likelihood ratio used is defined as:

$$Q_\mu = -2 \ \ln \frac{\mathcal{L}(obs \mid \mu \cdot s + b, \ \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(obs \mid \hat{\mu} \cdot s + b, \ \hat{\boldsymbol{\theta}})} \tag{7.1}$$

Let's now focus on the ratio which is the argument of the logarithm. At the numerator there is the profile likelihood function in which $\hat{\boldsymbol{\theta}}_\mu$ is the value of $\boldsymbol{\theta}$ maximizing $\mathcal{L}$ for the assumed μ; in other words, $\hat{\boldsymbol{\theta}}_\mu$ is the conditional maximum-likelihood estimator of $\boldsymbol{\theta}$ and consequently is a function of μ itself. The denominator, instead, is maximized in an unconstrained way (with both μ and $\boldsymbol{\theta}$ free to vary), thus $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the true maximum likelihood estimators. Hence, by definition, the ratio has values between 1, when the hypothesized μ coincides with $\hat{\mu}$, showing thus great compatibility between the data and the hypothesis, and 0 when, instead, the assumed μ is at odd with $\hat{\mu}$, denoting in this way a high degree of incompatibility between the data and the hypothesis.
Viceversa, higher values of Q_μ correspond to increasing incompatibility between the data and the μ hypothesis. The presence of the nuisance parameters broadens the profile likelihood as a function of μ relative to what one would have if their values were fixed. This reflects the loss of information about μ due to the systematic uncertainties [123].
It is also to note that, in the asymptotic regime, according to Wilks' theorem [124], the probability density functions of the test statistic Q_μ is expected to have a χ^2 distribution for one degree of freedom.
The chosen test-statistic Q is constructed to decrease (because of the minus sign in front of the logatithm) monotonically for increasingly signal-like experiments, so that the confidence in the signal plus background hypothesis is given by the probability that the test-statistic is greater than or equal to the value Q_μ^{obs} observed in the experiment:

$$\mathrm{CL}_{s+b} = P_{s+b}(Q \geq Q_\mu^{obs}) \tag{7.2}$$

where:

$$P_{s+b}(Q \geq Q_\mu^{obs}) = \int_{Q_\mu^{obs}}^{+\infty} \frac{\mathrm{d}P_{s+b}}{\mathrm{d}Q} \, dQ \tag{7.3}$$

and where $\mathrm{d}P_{s+b}/\mathrm{d}Q$ is the probability distribution function (p.d.f.) of the test-statistic for signal plus background experiments. Small values of CL_{s+b} indicate poor compatibility with the signal plus background hypothesis and favor the background-only hypothesis. Similarly, the confidence in the background hypothesis is given by the probability that the test-statistic is greater than or equal to the value Q_{μ}^{obs} observed in the experiment:

$$\mathrm{CL}_b = P_b(Q \geq Q_{\mu}^{obs}) \tag{7.4}$$

where:

$$P_b(Q \geq Q_{\mu}^{obs}) = \int_{Q_{\mu}^{obs}}^{+\infty} \frac{\mathrm{d}P_b}{\mathrm{d}Q}\, dQ \tag{7.5}$$

and where $\mathrm{d}P_b/\mathrm{d}Q$ is the p.d.f. of the test-statistic for background-only experiments. Values of CL_b very close to 1 indicate poor compatibility with the background hypothesis and favor the signal plus background hypothesis. One possible technique in case of small expected signal in the presence of backgrounds is to normalize the confidence level observed for the signal+background hypothesis, CL_{s+b}, to the confidence level observed for the background-only hypothesis, CL_b:

$$\mathrm{CL}_s \equiv \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} \tag{7.6}$$

This makes it possible to obtain sensible exclusion limits on the signal even when the observed rate is so low that the background hypothesis is called into question. This procedure is known as *modified frequentist* or CL_s method. It gives an approximation to the confidence in the signal hypothesis, CL_s, one might have obtained if the experiment had been performed in the complete absence of background, or in other words, if it had been possible to discard with absolute certainty the selected events due to background processes. Although CL_s is not, strictly speaking, a confidence (it is a ratio of confidences), the signal hypothesis will be considered excluded at the confidence level CL when

$$1 - \mathrm{CL}_s \geq \mathrm{CL} \tag{7.7}$$

The definition of the confidence interval obtained is useful: it indicates the boundary (or boundaries) of a region where one would not have expected to observe equally or less signal-like results than the actual observation in case the signal hypothesis were true (at or below a specified rate).

The tool used to extract the limits uses the *RooStats* [125] package[2]. The input to the tool used to compute the limit consists of four channels:

- an electron and muon channel in the signal region, for which template shapes are used;
- an electron and muon channel counting experiment in the sideband region.

[2] *RooStats* is a project to create advanced statistical tools required for the analysis of LHC data, with emphasis on discoveries, confidence intervals, and combined measurements.

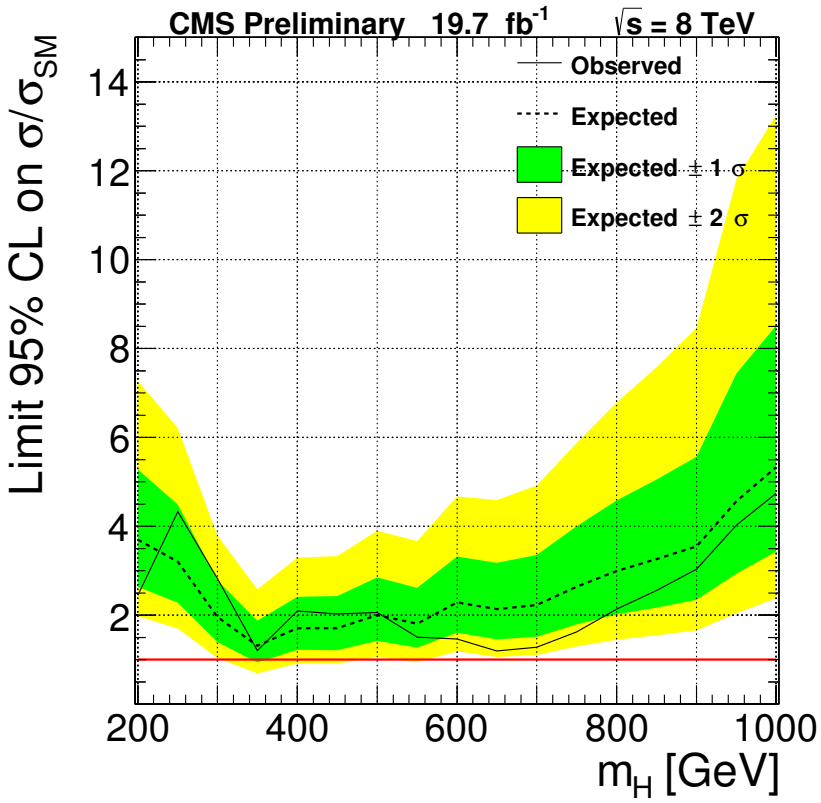


Figure 7.3: Expected and observed asymptotic limit as a function of the Higgs mass.

The uncertainties of each process are correlated between the signal and background regions. The input for the signal region channels is the two-lepton two-jet invariant mass template, for signal and background. For the sideband channels, instead, we just use the expected event yields. Both gluon fusion Higgs and VBF Higgs are regarded as signals, while the backgrounds are Z+jets, dibosons and top. All the systematics mentioned in Sec. 6 are included as nuisances, either in the form of normalization uncertainties or in the form of normalization plus shape uncertainties. For the normalization uncertainties a *log-normal* distribution[3] is assumed. The Z+jets normalization is left floating in the limit, with signal region and sideband floated in a correlated way. The limit tool extracts from the Maximum Likelihood fit of data the best estimate of the Z+jets normalization. The diboson cross section and top cross section are assigned a large systematic uncertainty (12% and 15% respectively) so that the limit can adjust their normalization if needed. They are however not completely free to fluctuate like the Z+jets.

After unblinding we obtain the limit plot shown in Fig. 7.3. The observed limit shows an evident ripple because of the limited statistics in the data sample. Moreover, no interpolation between mass points has been applied. The observed limit is higher than the expected in the mass region between 250 and 350 GeV because of an overfluctuation of data in that range. Conversely, the observed limit is lower than the observed in the

[3]A *log-normal* distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed.

high mass region (beyond 500 GeV) because of a downfluctuation of data with respect to the MC expectation (see Fig. 7.2).
This search alone does not allow to esclude the SM-like Higgs boson, but helps gaining sensitivity in the combination with the gluon fusion analysis, as discussed in Chap. 8.

7.3 Statistical interpretation of results with the Electroweak Singlet model

When taking into account the recently discovered Higgs boson with a mass around 125 GeV, the search for a similar scalar particle is put in the proper context with models adding a second state that mixes with the boson at 125 GeV, as it is the case of the Electroweak (EWK) Singlet model discussed in Sec. 1.3.
The phenomenology within such a model is done by modifying the couplings of the two gauge eigenstates (SM and singlet), which should be related by unitarity. The original coupling strength of the light Higgs boson is therefore reduced with respect to the SM case. If we define C (C') as the scale factor of the couplings of the low (high) mass boson with respect to the SM, one can write $C^2 + C'^2 = 1$ as the unitarity condition to be preserved. The EWK Singlet cross-section is also modified by a factor, μ', and the modified width is Γ'; they are defined as:

$$\mu' = C'^2 \cdot (1 - BR_{\text{new}}) \tag{7.8}$$

$$\Gamma' = \Gamma_{\text{SM}} \cdot \frac{C'^2}{1 - BR_{\text{new}}} \tag{7.9}$$

where BR_{new} is the branching ratio of the EWK Singlet to non-SM-like decay modes.

In our analysis we have focused on the case where $C'^2 \leq (1 - BR_{\text{new}})$. In this regime the new state is expected to have an equal or narrower width with respect to the SM case. Most of the analysis is also focused on the $BR_{\text{new}} = 0$ benchmark: the high mass of the singlet and the indirect limits from the H(125) observation are expected to constrain significantly this parameter. The extension with $BR_{\text{new}} \neq 0$ is exemplified in the end. Small differences in the final results are expected with respect to the limits obtained for the simple $BR_{\text{new}} = 0$ case. Under the BSM hypothesis summarised above we use the C'^2 definition (as well as $BR_{\text{new}} \neq 0$) to generate different widths of the singlet scalar boson. The generated SM signal mass line shape generated by POWHEG is re-weighted in order to simulate the singlet scalar line shape. In this re-weighting procedure we set as target line shape a relativistic Breit-Wigner with a narrower signal width with respect to the SM Higgs one. Notice that due to the poor resolution of this channel, departures from the Breit-Wigner approximation are not expected to modify significantly the final results. The results of our search have therefore being reinterpreted within this model and the results are summarized in Fig. 7.4 in which the observed limits are shown in comparison with the expectations as a function of m_H for different values of C' and BR_{new}.

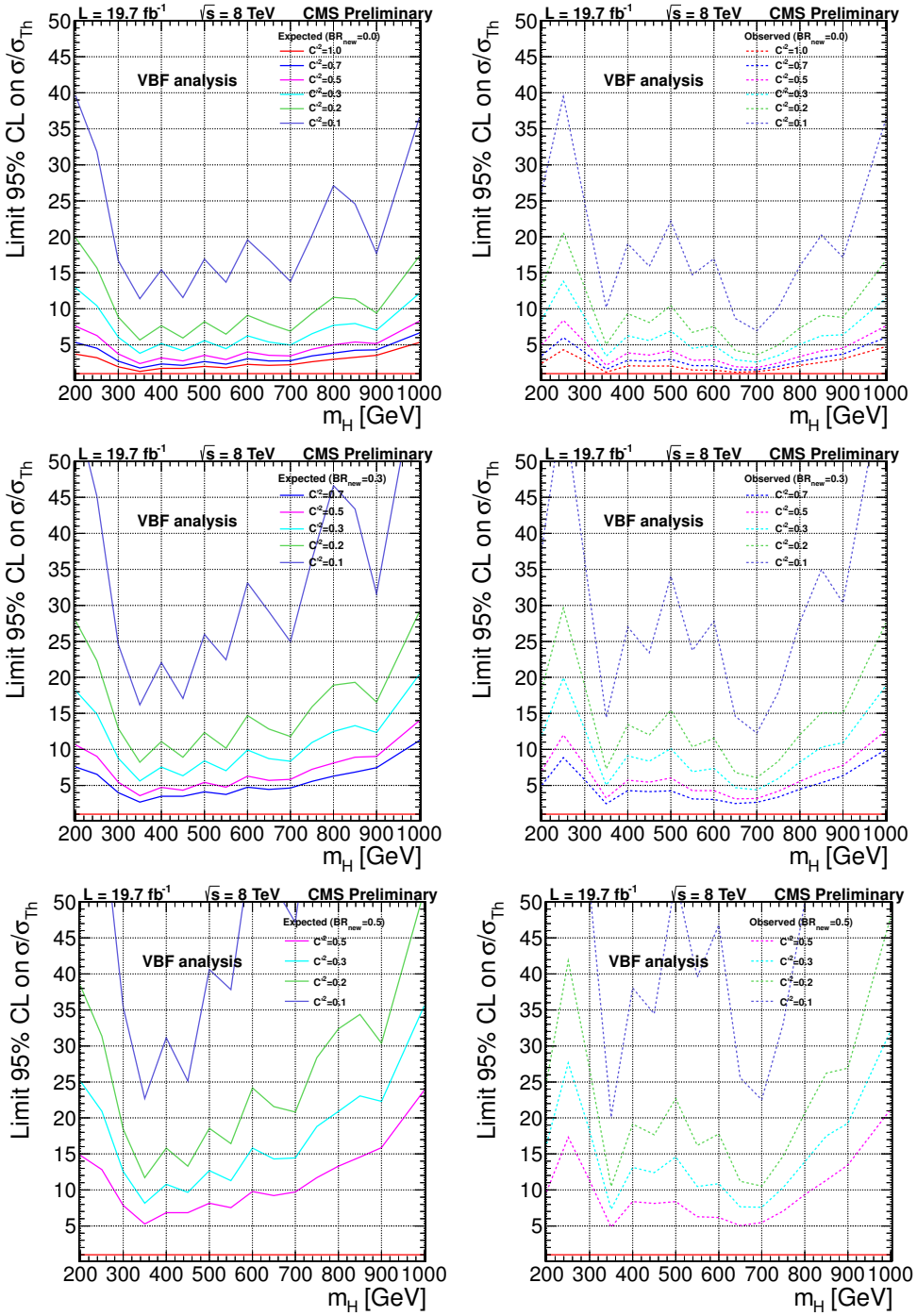


Figure 7.4: Expected (left) and observed (right) limits on the Electroweak Singlet interference model for three values of BR_{new} in the VBF category.

Chapter 8

Statistical analysis and results for the combined analysis

All the different sub-analyses (see Appendix A) has been combined to obtain the final results and produce the limit plots for the analysis. The sub-analyses have been made orthogonal one to each other in order not to double-count events. In order to achieve this, events are tagged as VBF and excluded from the gluon fusion analysis if they contain a pair of VBF jet candidates separated by $\Delta\eta > 3.5$ and with an invariant mass larger than 500 GeV.

8.1 Statistical interpretation of results using the SM Higgs boson as a benchmark

The m_{lljj} spectra are studied for 16 Higgs boson mass hypotheses in the range between 230 and 1000 GeV, where the mass steps are optimized to account for the expected width and resolution for measurement of m_H [126].
Based on the normalization and shape of the m_{lljj} distribution for signal and background, we extract exclusion limits using the formalism developed by the CMS and ATLAS Collaborations in the context of the LHC Higgs Combination Group [127]. Fourteen data categories[1] are used simultaneously in a maximum likelihood fit of the binned histograms for each mass hypothesis with a background-only and a signal-plus-background models, and the likelihood ratio of the two fits serves as test statistic for the modified frequentist criterion CL_s [128] [129]. Systematic uncertainties are incorporated as nuisance parameters and are treated according to the frequentist paradigm. The results presented here are obtained using asymptotic formulae [123], including a few updates recently introduced in the RooStats package [130]. In Fig. 8.1 (left) the overall exclusion limit, obtained combining the VBF analysis discussed in this thesis and the gluon-fusion analyses described in Appendix A, is shown. Limits on the SM Higgs boson production cross section times branching fraction of $H \to ZZ$ are presented in Fig. 8.1 (right). For comparison, expectations are shown for a SM-like Higgs boson. This analysis excludes the existence of a

[1] Two categories (lepton flavour) for the VBF analysis and six categories (three b-tag categories for each lepton flavour) for both the standard and the merged-jet gluon fusion analyses, see Appendix A.

resonance with properties of the SM Higgs boson in the mass range between 305 and 593 GeV as well as between 613 and 693 GeV. The merged jet category has a higher sensitivity when the Z bosons are boosted so strongly that both decay quarks hadronize into one jet, at high m_H masses. Also the VBF analysis gives a contribution at high mass values, improving the limit sensitivity.

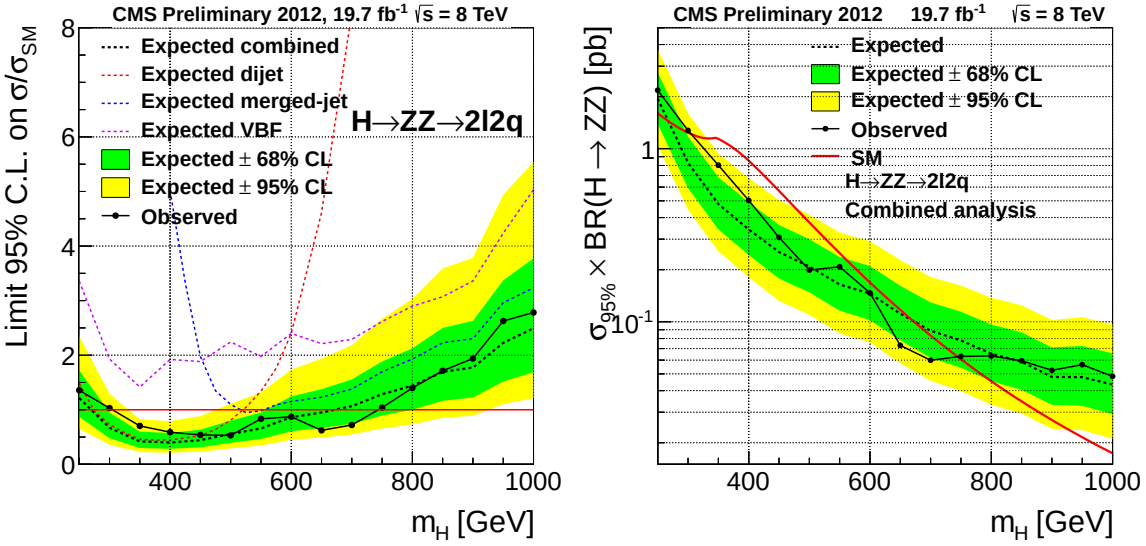


Figure 8.1: Observed (solid) and expected (dashed) 95% CL upper limits, obtained with the CL_s technique, on the ratio of the production cross section to the SM expectation for the Higgs boson (left), and on the production cross section times branching fraction of $H \to ZZ$ (right). The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. In the left graph, the solid horizontal line at unity indicates the expectation for a SM-Higgs-like boson.

8.2 Statistical interpretation of results with the Electroweak Singlet model

Also the combined results can be interpreted in terms of an Electroweak Singlet scalar mixing with the 125 GeV Higgs boson, as described in Sec. 7.3, where also symbols and formalism have been defined. Indirectly we can set an upper limit at 95% CL on $C'^2 < 0.446$ using the signal strength fits to the H(125) candidate as obtained in [131]. The analysis is performed re-weighting the generated signal shapes to a relativistic Breit-Wigner function of width Γ' in the VBF channel and to the corresponding Complex Pole Scheme (CPS) shape (including interference effects) for the gluon fusion channels. Residual differences are expected to be small compared to the experimental resolution. Expected and observed exclusion limits in the $C'^2 - m_{\mathrm{H}}$ plane, for $BR_{\mathrm{new}} = 0$, are displayed in Fig. 8.2 for the gluon fusion and VBF analyses, and for their combination. Similarly, expected and observed exclusion limits in the $BR_{\mathrm{new}} - C'^2$ plane are depicted in Fig. 8.3,

for $m_H = 400$ and 900 GeV. CMS data disfavor EW Singlet models with high C'^2 and low BR_{new} values, excluding some regions in the parameter space for m_H in the range 300 to 500 GeV and around 700 GeV.

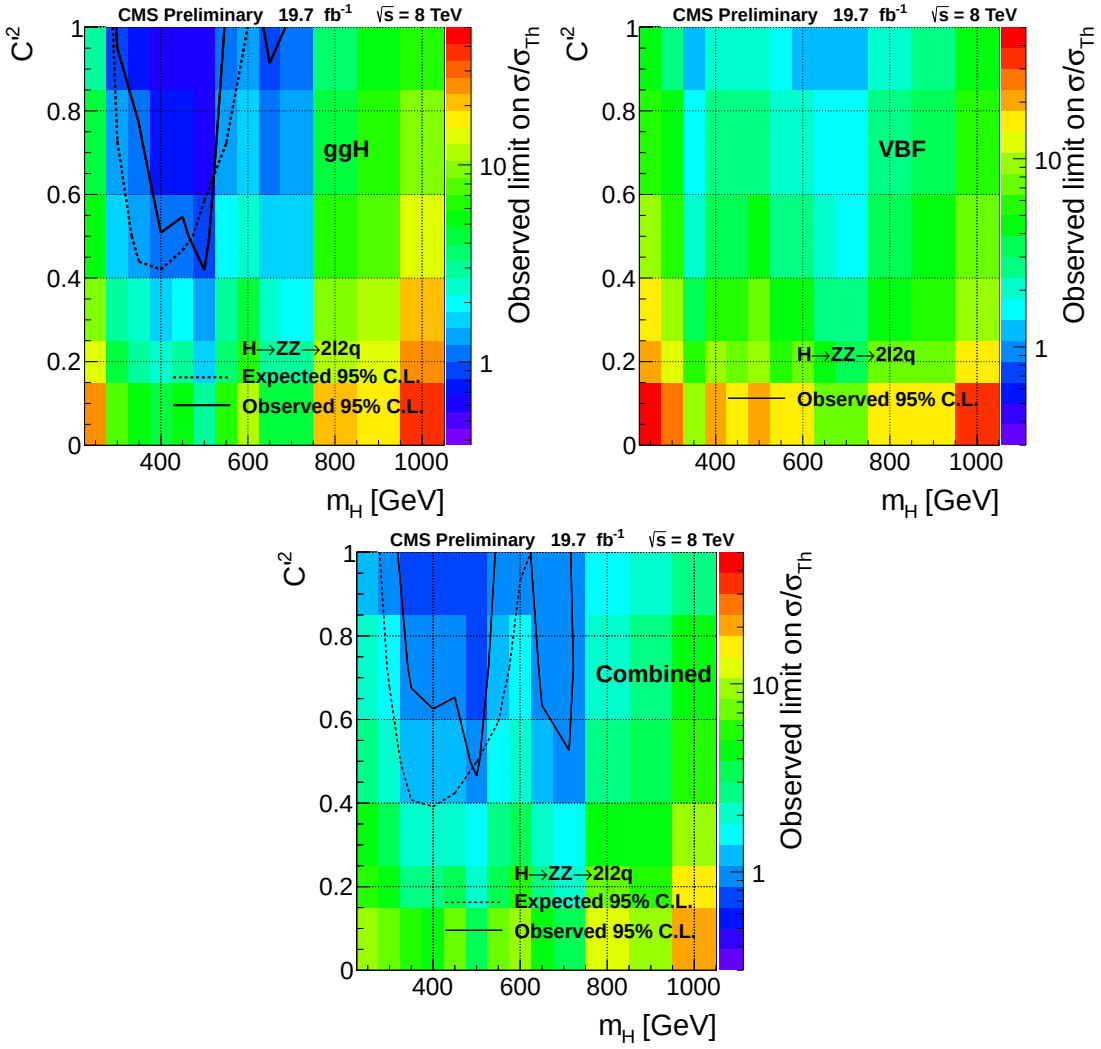


Figure 8.2: Contours of the observed limits at the 95% CL in the $C'^2 - m_{\text{H}}$ plane ($BR_{\text{new}} = 0$) for the gluon fusion and VBF analyses combined. The solid (dashed) lines depict the observed (expected) exclusion limits, corresponding to $\sigma/\sigma_{\text{th}} = 1$. No region is excluded in the $C'^2 - m_{\text{H}}$ parameter space for the VBF analysis alone.

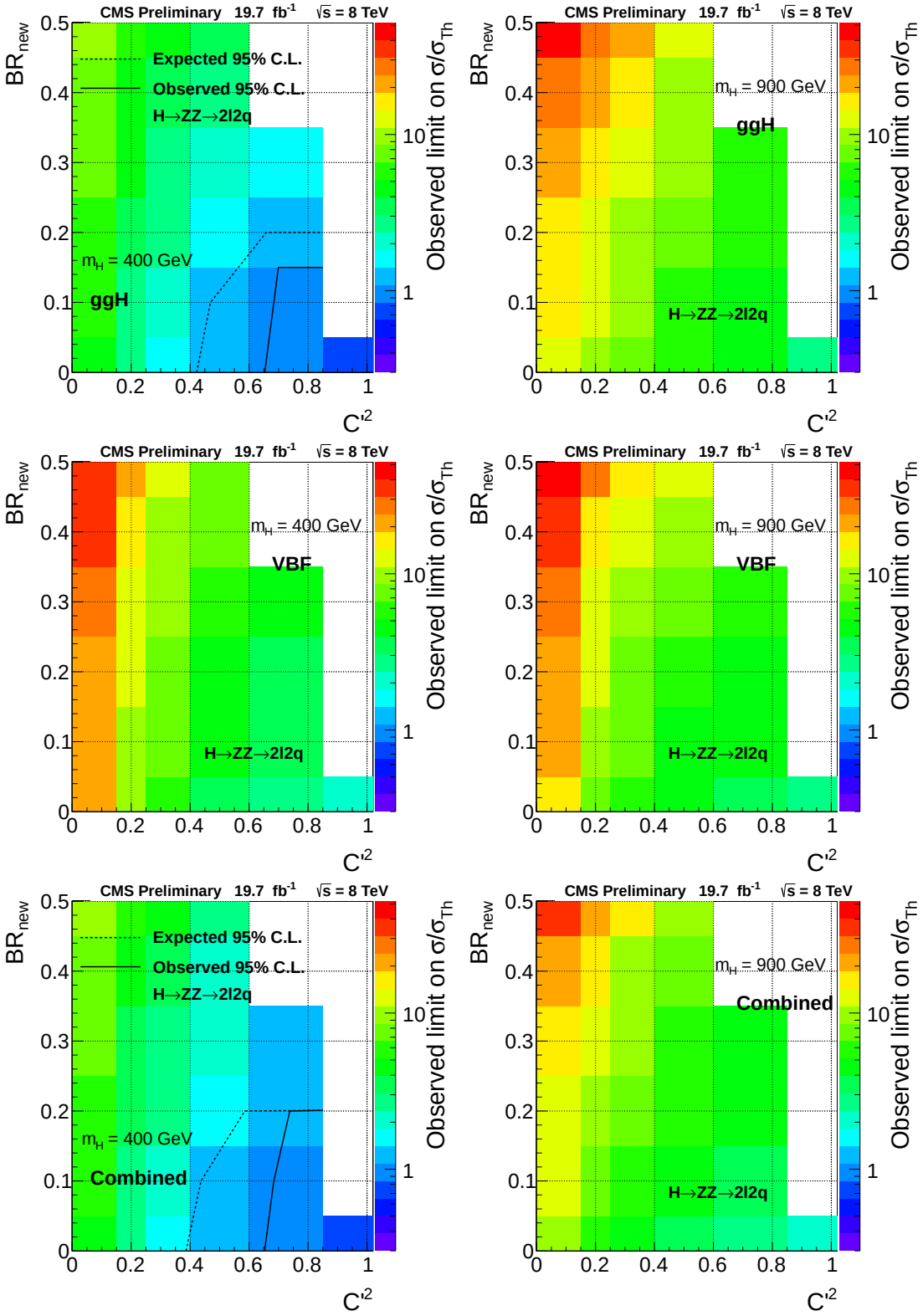


Figure 8.3: Contour of the observed limits at the 95% CL in the $BR_{\text{new}} - C'^2$ plane, for $m_H = 400$ GeV (left) and $m_H = 900$ GeV (right), for the combination of the gluon fusion and VBF analyses. The solid (dashed) lines depict the observed (expected) exclusion limits, corresponding to $\sigma/\sigma_{\text{th}} = 1$. The white areas in the graphs correspond to model parameters not probed in these analyses.

Conclusions

A search for a Standard Model-like Higgs boson in the $H \to ZZ \to l^+l^-q\bar{q}$ decay channel ($l = e,\ \mu$), with the Higgs boson produced via the Vector Boson Fusion (VBF) mechanism, has been performed in the mass range between 200 and 1000 GeV.
Data samples collected by the CMS experiment during 2012 at $\sqrt{s} = 8$ TeV, for a total luminosity of 19.7 fb^{-1}, have been used.
This channel has a high branching ratio and a completely reconstructed final state. The presence of two additional jets from the production mechanism can be exploited as a signature of signal events. The main sources of background in this channel are represented by Drell-Yan Z plus jets, dibosons (ZZ, WZ) plus jets, $t\bar{t}$ plus jets events.
Some standard kinematic cuts are applied, in order to roughly reduce background, by requiring the invariant mass of the lepton pair and of the jet pair from Z to be close to the nominal Z mass, and by vetoing the presence of missing energy in the event. A further background suppression is achieved by exploiting the characteristic angular distribution of the Higgs decay products, in the hypothesis that the sought Higgs boson is still a spin-0 particle. Then, a multivariate discriminant is built, exploiting the peculiar VBF topology, i.e. the large angular separation and large invariant mass of the VBF jets. Throughout all the analysis, the boosted topology in which the jets from the hadronic Z are reconstructed as a single merged jet has also been taken into account. That happpens when the Higgs is so massive that the jets from the Z are emitted with small angular separation and thus reconstructed as a single large jet with a substructure.
The final statistical analysis is a shape analysis performed on the reconstructed invariant mass of the Higgs decay products, m_{lljj}, which is expected to be peaked for signal events around the Higgs boson mass and has a broader distribution for background processes.
The modified frequentist method CL_s has been used to determine upper limits on the Higgs boson production cross section, taking the Standard Model Higgs boson as a benchmark. Then, a signal reweighting has been applied, in order to re-interpret the results according to the Electroweak Singlet model, which represents the simplest extension of the Standard Model Higgs sector where the heavy resonance also contributes to electroweak symmetry breaking and has similar properties to the SM Higgs boson with modifications to its width and allowing for non-SM-like Higgs decay modes.
The analysis described, considering the VBF production mechanism, has then been combined to the analysis in the gluon-fusion production, thus leading to an overall exclusion limit, both in the Standard Model scenario and in the Electroweak Singlet frame.
The VBF analysis alone does not allow to exclude any region in the mass range explored. Nevertheless, the VBF analysis contribution is important to improve the overall sensitivity, leading to a widening of the excluded mass region with respect to the gluon-fusion

analysis alone, especially in the high mass range. The mass range excluded by the combined analysis is 305-744 GeV.
Even in the Electroweak Singlet frame, the addition of the VBF analysis contributes in improving the sensitivity for the exclusion of some regions in the parameters space.
Results from the $H \to ZZ \to 2l2q$ analysis are being combined with the ones obtained from the high-mass analyses of the other $H \to ZZ$ and $H \to WW$ decay channels. The other channels considered are $H \to ZZ \to 2l2\nu$, $H \to ZZ \to 4l$, $H \to WW \to l\nu l\nu$, $H \to WW \to l\nu qq$. The aim is the search for a SM-like Higgs boson at high mass, assuming the properties predicted by the SM; then the results will be reinterpreted in the Electroweak Singlet scenario, as done in this work.
Results are expected to be submitted for the publication by the end of the year.

Appendix A

Study of the process $H \to ZZ \to l^+l^-q\bar{q}$ in the gluon fusion production mechanism

A brief description of the two sub-analyses dealing with a gluon fusion produced Higgs decaying in two leptons and two quarks is here presented. For further details, see [132]. In Sec. A.1) the sub-analysis in which jets from the Z decay are reconstructed as two separate jets is discussed. In Sec. A.2 the sub-analysis focusing on the boosted topology where jets from Z are reconstructed as a single merged jet is presented.

A.1 Analysis for ggH production with di-jets

A.1.1 Data and MC samples

Data samples are the same used for the VBF analysis (see 4.1). They have been collected using the same dilepton triggers as for the VBF analysis. In addition to the e^+e^- and $\mu^+\mu^-$ signal sample, events with a muon and an electron of opposite charge are selected in order to estimate the background from $t\bar{t}$ production.
Monte Carlo samples for $gg \to H \to ZZ$ have been generated using Powheg, interfaced to Pythia6 for parton shower, for masses between 200 and 600 GeV.

Simulated samples for backgrounds are listed in Tab. A.1.

A.1.2 Event selection

The event selection is basically the same as the one used for the VBF analysis described in detail. Obviously, only two jets (the ones from Z) are expected in the final state.

Jets originating from pile-up interactions are removed by requiring jets to be taggable (see later) and requiring $\beta > 0.2$, where β is the sum of transverse momenta of the charged particles in the jet coming from the primary vertex normalized to the sum of transverse momenta of all the charged particles in the jet.

The kinematic event selection is summarized in Tab. A.2

Process	σ [pb]	lumi [fb^{-1}]
Z+jets	3503.71	8.7
Z+1 jet	660.6	36.4
Z+2 jet	215.1	101.9
Z+3 jet	65.79	167.4
Z+4 jet	27.59	232.0
$t\bar{t}$	225.197	29.9
ZZ	17.654	555
WZ	22.88	437
WW	57.1097	175

Table A.1: Background samples for the gluon fusion analysis using di-jets.

Observable	Cut
First (second) lepton p_T	$>$ 40 (20) GeV
Jets p_T	$>$ 30 GeV
m_{ll}	$\in$ [76 GeV, 106 GeV]
m_{jj}	$\in$ [71 GeV, 111 GeV]
MET significance	$<$ 10

Table A.2: Event selection for the gluon fusion analysis using di-jets.

A.1.3 Higgs candidate selection

The Higgs boson candidate is selected among those in the M_{jj} signal region that pass all the selection criteria. When no candidate survives the selection cuts in the signal region, the Higgs boson candidate is chosen from the candidates in the sideband. The event is rejected whenever no candidates are found in either region. When there is more than one Higgs boson candidate per event, the candidate in the highest b-tag category is selected. If more than one candidate remains, the one with minimum $|M_{ll} - m_Z| + |M_{jj} - m_Z|$ is chosen.

A.1.4 Signal discrimination

The three b-tag categories (2-, 1-, and 0-btag) already introduced for the VBF analysis are used to classify events into three exclusive categories. The different signal-to-background ratio of each category helps improving the sensitivity of the analysis. Note that the event classification into b-tag categories represents a difference with respect to the VBF analysis, where instead the b-tagging information is used only for the candidate selection. Here, viceversa, events belonging to the three exclusive categories are analyzed separately, and then a combination is performed.
The linear likelihood discriminant (LD) described in Sec. 5.1 is constructed from the signal and background probabilities. Signal sensitivity is improved requiring events with LD$>$ 0.5, a selection cut almost optimal for the three btag categories.

The $t\bar{t}$ background is an important source of contamination in the 2-btag category. It is estimated from a data sample using $e^{\pm}\mu^{\mp}$ events passing the same cuts as the signal. To reduce this background source, a cut on the MET significane at 10 is applied.
Tha analysis is then performed exploiting the shape of the mass distribution of the $lljj$ system.

A.1.5 Background determination

The following SM processes are considered as background in this analysis: diboson, Z+jets and $t\bar{t}$ production. The diboson production (mainly ZZ and WZ) is simulated using MC while the other two contributions are estimated either using MC simulation corrected to data in control regions (Z+jets), or extracted directly from data in control regions ($t\bar{t}$). To estimate the background from Z+jets and the shape of its m_{lljj} distribution, large exclusive simulated samples of Z + n jets (n = 1 to 4) events are used. These samples have been thoroughly verified to reproduce the data in the m_{jj} sideband control region. The small discrepancies found are attributed to a bad modeling of the p_T spectrum of the l^+l^-jj system, the so-called p_T^H observable. The mismodeling of the p_T spectrum in the simulation is the same in the signal and sideband regions. A weight calculated as the ratio of the p_T distributions in data over simulations in the sideband region, and fitted to the function

$$f(p_T) = \left(1 + \frac{1}{a + bp_T^2}\right) \frac{1}{e^{-p_T/c} + 1}$$

as function of p_T, is used to correct the Z+jets distributions in an event-by-event basis.
The $t\bar{t}$ background is an important source of contamination in the 2-btag category. It is estimated from the data using $e^{\pm}\mu^{\mp}$ events passing the same cuts as the signal. This method accounts for other small backgrounds (as WW + jets, Z$\rightarrow \tau^+\tau^-$ + jets, single top, fakes) where the lepton flavor symmetry can be invoked as well.

A.1.6 Systematic uncertainties

In Tab. A.3 are reported the main systematic uncertainties on signal normalization. In Tab. A.4 the summary of systematic uncertainties on the normalization and shape of the background determination is reported.

A.1.7 Results

Expected upper limits on the SM Higgs boson production cross section for the dijet analysis alone are determined as function of the Higgs boson mass hypothesis, taking as input the m_{lljj} distribution for data, and for the background and signal expectations. The statistical procedure, based on the profile likelihood method, uses the asymptotic CL_s approach. Systematic uncertainties are treated as nuisance parameters. The results are expressed as upper limits on the ratio of the cross section times branching fraction of the process $H \rightarrow ZZ \rightarrow l^+l^-q\bar{q}$ divided by the expectation for the Standard Model Higgs boson, σ/σ_{SM}. A particular Higgs boson mass hypothesis is excluded whenever $\sigma/\sigma_{SM} < 1$. Figure A.1 shows the expected and observed limits for the di-jet analysis

Source	0 b-tag	1 b-tag	2 b-tag
Muon trigger & ID	1.8%		
Electron trigger & ID	2%		
Electron energy scale	0.2%		
Muon momentum scale	0.1%		
Jet reconstruction	1-4%		
b-tagging eff. and mistag rate	1-2%	1-3%	5-6%
MET	0.6%		
Pile-up	0.1-1%		
Production mechanism (PDF)	1.5%		
Production mechanism (lineshape)	0-7%		
Luminosity	2.6%		
Higgs cross-section (for R)	13-18%		

Table A.3: Summary of systematic uncertainties on signal normalization. Most sources are multiplicative errors on the cross-section measurement, except for expected Higgs cross-section (which is relevant for the measurement of the ratio to SM expectation R).

Source	Normalization	Shape
Muon trigger & ID	1.8%	
Muon momentum scale	0.1%	
Electron trigger & ID	2.0%	
Electron energy scale	0.1%	
Jet energy scale	3.6%	0-4%
b-tagging efficiency SF 0-tag	+0.4%	
b-tagging efficiency SF 1-tag	-0.8%	
b-tagging efficiency SF 2-tag	-4.6%	
Mistag SF 0-tag	-1.1%	
Mistag SF 1-tag	+3.7%	
Mistag SF 2-tag	+3.6%	
MET	0.3%	
Pile-up	0.1%	
p_T^{lljj} weighting	0.8%	0-3%
Diboson cross section	12%	
Luminosity	2.6%	
Residual difference		0-20% (0-btag)
data-background in		0-25% (1-btag)
control region		0-40% (2-btag)

Table A.4: Summary of systematic uncertainties on the normalization and shape of the background determination.

using the full dataset recorded during 2012 at 8 TeV, corresponding to a luminosity of 19.7 fb^{-1}. Taking into account this channel only, the presence of a Standard Model-like Higgs boson can be excluded at 95% CL in the mass range from ~ 300 to ~ 520 GeV.

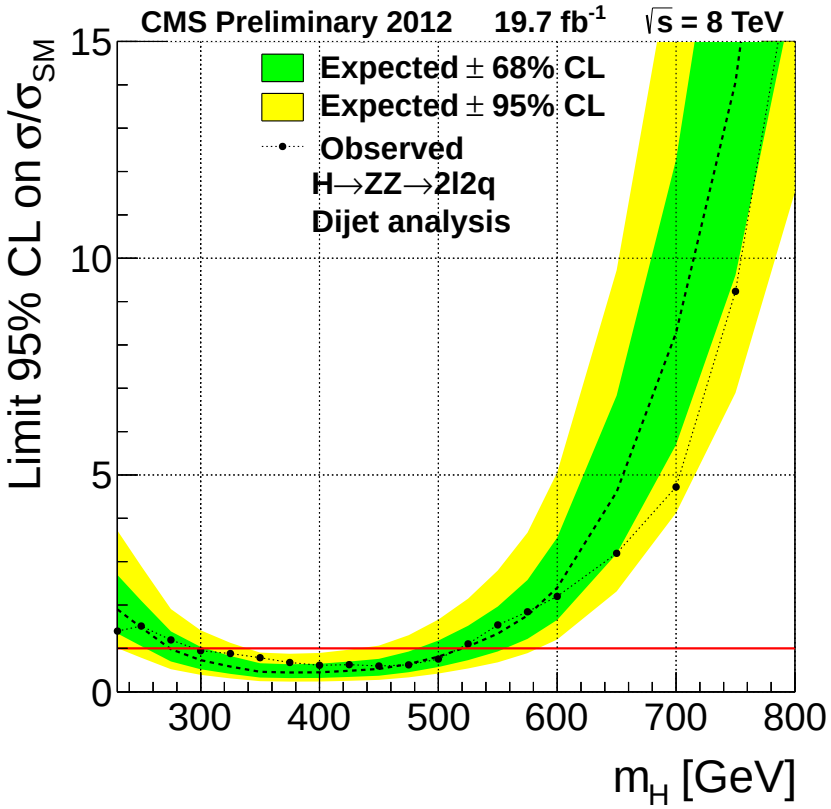


Figure A.1: Observed (dots) and expected (dashed line) 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson obtained using the CL_s technique for the dijet analysis. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the expectation for a SM-Higgs-like boson.

A.2 Analysis for ggH production with merged jets

An outline of the merged-jet analysis is now given. As already stated in Sec. 3.3.3, at large Higgs masses the Z bosons have a large transverse momentum, so that hadronic decay products are collimated to the point of being reconstructed as a single merged jet with a two-subjets structure. In this analysis the kinematic and flavour information of the subjets are exploited for effective signal isolation. Using merged jets, the signal efficiency for Higgs masses above 600 GeV is increased with respect to the di-jet approach.

A.2.1 Data and MC samples

Data and MC samples are the same used for the gluon fusion analysis using di-jets. For signal, mass now ranges from 400 to 1000 GeV.

A.2.2 Event selection

For this analysis the signature of a Higgs boson signal event is a lepton pair, selected as described previously, and a single merged jet containing the entire hadronic Z decay. The Cambridge-Aachen algorithm with radius parameter $R = 0.8$ (CA8) is adopted for reconstructing the hadronic Z decay. CA8 merged jets are required to be inside the tracker acceptance, $|\eta(j)| < 2.4$. To avoid double counting of the same object reconstructed in different collections, merged jets are required to be separated from high-p_T isolated leptons by $\Delta R(j) = \sqrt{\Delta\phi^2 + \Delta\eta^2} > 0.8$. Jet pruning improves the resolution on the jet mass and helps reducing the effect of pile-up. The AK7-CHS jet energy corrections are applied to data and simulated merged jets [133, 134]. Merged jets are required to have pruned jet mass $M_{pruned}(j) > 50$ GeV and $p_T(j) > 100$ GeV. These minimal cuts reject fake candidates from QCD. Each pruned jet has two associated subjets which are required to be taggable[1]. A cut at $\tau_2/\tau_1 \leq 0.5$ is implemented (see Sec. 3.3.3). The combination of the cut on the pruned mass and the τ_2/τ_1 cut is referred to as *boosted Z tagging*. Then, the same kinematic selection as in the di-jet analysis is applied (see Tab. A.2) , with the only difference that the cuts and requirements applied to jets are here applied to the sub-jets. The subjets are tagged using the loose (JPL) and medium (JPM) working points of the JP algorithm, with events classified into three exclusive b-tag categories as for the di-jet analysis.

A.2.3 Higgs candidate selection

After applying the cuts, the events generally contain a single merged Z candidate; in the small fraction of events with more than one candidate, the one with highest transverse momentum is chosen. In order to suppress the dominant Drell-Yan Z+jets background, Higgs candidates are selected in the pruned jet mass region $71 \leq M_{pruned}(j) \leq 111$ GeV, called the "signal region". Outside of this signal region, candidates with pruned jet mass within $60 \leq M_{pruned}(j) \leq 130$ GeV are used for background determination ("sideband region").

[1] It means that it has to satisfy the loose Jet ID requirements, see Sec. 4.3.1.

A.2.4 Signal discrimination

This analysis also adopts the angular likelihood discriminant LD from the di-jet analysis; the variables related to jets are now related to subjets.
Then, a shape analysis is performed, exploiting the m_{llj} distribution.

A.2.5 Background determination

The following SM processes are considered as background in this analysis: dilepton events with no real Z decays, diboson production with Z$\rightarrow l^+l^-$ and Drell-Yan Z+jets production. Dilepton backgrounds that do not contain a real leptonic Z decay are estimated from the data using $e^{\pm}\mu^{\mp}$ events passing the same cuts as the signal. This method accounts for $t\bar{t}$ production, WW + jets, Z$\rightarrow \tau^+\tau^-$+jets, single top and fakes. The size of this contamination is expected to be very small, and is indeed found to be negligible. The diboson production background (ZZ and WZ, with Z$\rightarrow l^+l^-$) is estimated from simulation. A boosted Z tagging efficiency data-to-MC scale factor of 0.91 ± 0.10 is applied to diboson events satisfying the final selection cuts. The diboson contamination represents 7%, 10%, and 22% of the total background in the 0-, 1- and 2-btag categories, respectively. Most background events come from Drell-Yan Z+jets production. To estimate this background source, MadGraph simulated samples (see above) are used. These samples have been thoroughly verified to reproduce the data after preselection cuts, and after selection in the sideband control region. To tune the Madgraph MC description to the data, two p_T-related corrections were implemented. The p_T spectrum of the llj system was corrected reweighting the simulated events with the same function calculated for the di-jet analysis. A reweighting factor was also applied to the p_T spectrum of the ll system. This factor was calculated as the ratio of two exponentials fit to the data and MC leptonic Z p_T distributions, with the one-sigma statistical uncertainty taken as a systematic. The shapes of the Z+jets background m_{llj} signal distributions are extracted from simulation; their normalization is only constrained, independently for each b-tag category, by the normalization of the diboson background-subtracted data.

A.2.6 Systematic uncertainties

In Tab. A.5 a summary of the systematic uncertainties on the normalization of signal and backgrounds in the merged-jet analysis is presented.

A.2.7 Results

The observed and expected limit on σ/σ_{SM} is determined for m_H hypotheses between 400 GeV and 1000 GeV using m_{llj} distributions. Fig. A.2 shows the limits for the full dataset recorded during 2012 at 8 TeV, corresponding to a luminosity of 19.7 fb^{-1}. This subanalysis, uncombined, is not able to exclude any mass region, but it helps improving the exclusion sensitivity, especially for high mass values, when combined to the di-jet and VBF analyses.

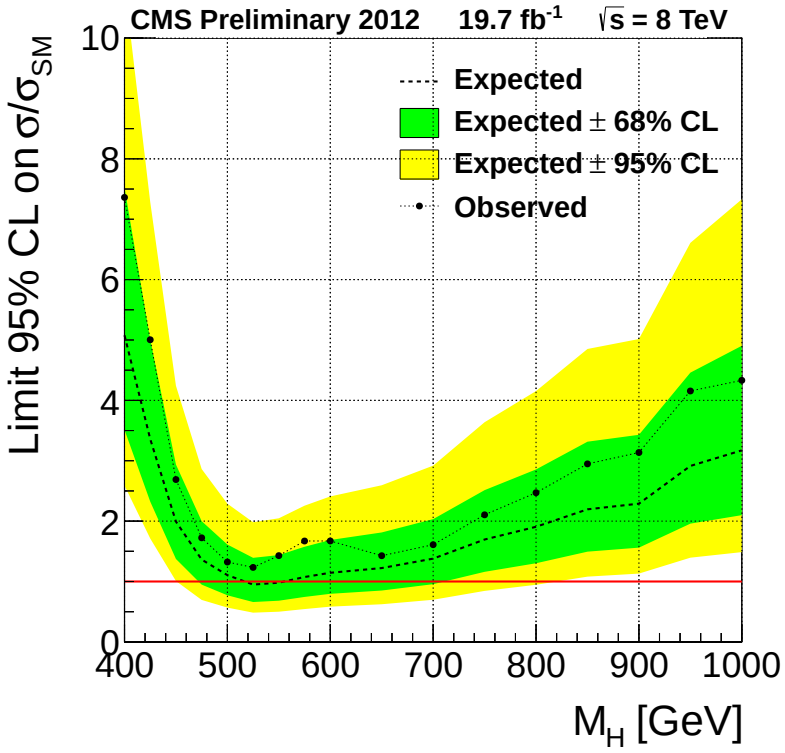


Figure A.2: Observed and expected 95% CL upper limit on the ratio of the production cross section to the SM expectation for the Higgs boson obtained using the CL_s technique for the merged jet analysis. The 68% and 95% ranges of expectation for the background-only model are also shown with green and yellow bands, respectively. The solid line at 1 indicates the expectation for a SM-Higgs-like boson.

Source	Background	H 500 GeV	H 600 GeV	H 700 GeV	H 1000 GeV
Luminosity	2.6%				
Electron trigger & ID	2.0%				
Muon trigger & ID	1.7%				
Electron energy scale	0.8%	0.9%	0.2%	0.1%	0.04%
Muon momentum scale	0.2%	0.1%	0.04%	0.02%	0.01%
Jet energy scale	0.9%	0.2%	0.5%	0.6%	0.6%
b-tagging efficiency SF 0-tag	+0.3%	+0.8%	+0.7%	+0.7%	+1.1%
b-tagging efficiency SF 1-tag	-0.7%	-0.1%	-0.1%	-0.0%	+0.2%
b-tagging efficiency SF 2-tag	-4.6%	-5.5%	-5.6%	-5.8%	-7.5%
Mistag SF 0-tag	-1.4%	-1.3%	-1.4%	-1.5%	-2.1%
Mistag SF 1-tag	+5.0%	+2.8%	+3.0%	+3.2%	+4.8%
Mistag SF 2-tag	+6.7%	+0.5%	+0.3%	+0.3%	+1.3%
Pile-up	0.7%	0.8%	0.7%	0.6%	0.5%
ZZ/WZ normalization	12%				
Boosted Z tagging	10% (ZZ/WZ)	10%			
Higgs lineshape reweighting		6.9%	6.3%	4.7%	1.1%
Higgs cross-section		14%	15%	15%	18%

Table A.5: Summary of systematic uncertainties on the normalization of signal and backgrounds in the merged jet analysis.

Appendix B

Tables

Process	Sample	σ (pb)
Z+jets	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	668
	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7C-v1/AODSIM	215
	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	61
	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27.4
	/DYJetsToLL_HT-200To400_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	23.43
	/DYJetsToLL_HT-400ToInf_TuneZ2Star_8TeV-madgraph/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	3.35
Di-bosons	/WZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	1.144
	/ZZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola//Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	0.57
$t\bar{t}$	/TTJets_FullLeptMGDecays_8TeV-madgraph-tauola/Summer12_DR53X-PU_S10_START53_V7C-v2/AODSIM	25.8

Table B.1: Background samples.

Process	Sample
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-200_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-250_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-300_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-350_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-400_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-450_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-500_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-550_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-600_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-650_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-700_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-750_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-800_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-850_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-900_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-950_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
VBF H($\to l^+l^-q\bar{q}$)	/VBF_HToZZTo2L2Q_M-1000_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM

Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-200_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-250_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-300_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-350_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-400_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-450_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-500_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-550_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-600_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-650_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-700_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-750_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-800_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-850_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-900_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-950_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM
Gluon Gluon H($\rightarrow l^+l^-q\bar{q}$)	/GluGluToHToZZTo2L2Q_M-1000_8TeV-minloHJJ-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V19-v1/AODSIM

Table B.2: Signal samples.

Appendix C

Track reconstruction at the High Level Trigger

In the past years, the CMS Florence collaboration has been heavily involved in the huge effort made to project and build the tracker detector. In particular, a strong contribution has been given in the strip detectors design, building and testing phase. As a "legacy" of this effort, some responsibilities related to the tracker detector, and in a broader way, to the *tracking* task, have been assigned to the Florence collaboration, in order to ideally carry on this effort.
By *tracking* we mean the set of algorithms aimed at the reconstruction of tracks starting from hits in the pixel or strip detectors. In particular, I have been involved in the *HLT tracking* task, that is to say the tracking performed at the High Level Trigger of the experiment. The peculiarity of this kind of tracking is that it has to be much faster than the offline one, at the same time aiming to reach similar performance in efficiency and resolution.
Before describing the peculiarities of the HLT tracking algorithms, it is mandatory to firstly describe the CMS tracking strategy more generally.

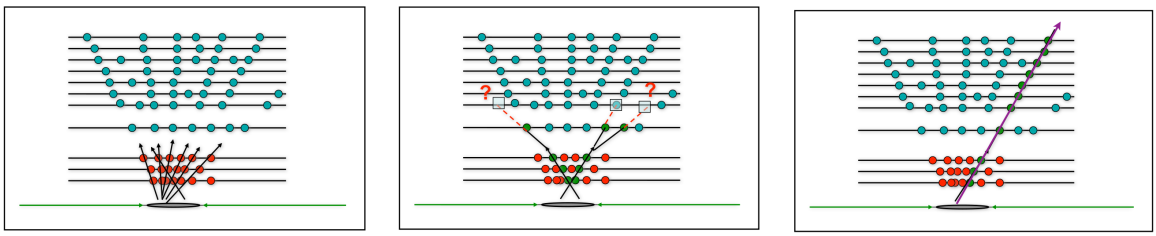
C.1 Track reconstruction: the *Iterative Tracking*

What follows is valid for both online (HLT) and offline track reconstruction. The CMS tracking software [62] is known as the Combinatorial Track Finder (CTF) [135] based on the Kalman filter method [76].
The collection of reconstructed tracks is produced by multiple iterations of the CTF track reconstruction sequence, in a process called *Iterative Tracking*.
In the early iterations, tracks with relatively high p_T, produced near the interaction region, are reconstructed. After each iteration, hits associated with tracks already found are removed, reducing the combinatorial complexity and thus allowing later iterations to search for lower p_T or highly displaced tracks.
For 2012 data reconstruction, the offline iterative tracking consisted of 6 iterations. Iteration 0 is the source of most tracks and is designed to reconstruct prompt tracks with $p_T > 0.8$ GeV and which have three pixel hits. Iteration 1 is used to recover prompt tracks

which only have two pixel hits or slightly lower p_T . Iteration 2 is configured to find low p_T prompt tracks. Iterations 3 to 5 are intended to find tracks which originate outside the beamspot[1] and to recover tracks not found by the previous iterations. At the beginning of each iteration, hits associated with high purity tracks found in previous iterations are discarded.
Each iteration proceeds in four steps (see also Fig. C.1):

- The seed generation provides initial track candidates using only a few (2 or 3) hits. A seed defines the initial estimate of the trajectory parameters and their uncertainties.
- The track finding is based on a global Kalman filter. It extrapolates the seed trajectories along the expected flight path of a charged particle, searching for additional hits that can be assigned to the track candidate.
- The track fitting is used to provide the best possible estimate of the parameters of each trajectory by means of a Kalman filter and *smoother* (see in the following).
- The track selection sets quality flags and discards tracks that fail certain criteria.

The main differences between the 6 iterations lie in the configuration of the seed generation and final track selection steps.



(a) (b) (c)

Figure C.1: (a) The track seeding stage; for simplicity, only the case in which seeding is performed using a triplet of pixel hits is shown here. (b) The trajectory building stage: compatible hits are searched in silicon layers found going outwards in the navigation. (c) The track fitting stage; only the first fitting procedure is shown. The procedure is then repeated, proceeding inwards.

C.1.1 Seed generation

The trajectory *seeds* define the starting trajectory parameters and associated uncertainties of potential tracks. Because of the 3.8 T magnetic field inside the tracker, charged particles follow helices and therefore five parameters are needed to define a starting trajectory. To obtain these five parameters, at least 3 hits are required, or 2 hits and a beam constraint. To limit the number of hits combinations, seeds are required to satisfy loose criteria, for

[1]The beamspot represents a three-dimensional profile of the luminous region, where the LHC beams collide at CMS. A rough estimate of the beamspot can be given by using pixel vertices, which are recontructed using only Pixel Tracks.

example on their minimum transverse momentum and consistency with originating from the proton-proton interaction region.
Seeds are built in the inner part of the tracker and the track candidates are reconstructed outwards. The reason for this approach is that, although the track density is much higher in the inner region of the tracker, the high granularity of the pixel detector ensures that the average occupancy of the inner pixel layer is much lower than the average occupancy of the outer strip layer. Secondly, the pixel layers provide three-dimensional space-point measurements which allow for a tighter constraint and better parameter estimation.

C.1.2 Track finding

The track finding module of the CTF algorithm is based on the Kalman filter method. The filter begins with a coarse estimate of the track parameters provided by the trajectory seed and then builds track candidates by adding hits from successive layers one by one. The information provided at each layer includes the location and uncertainty of any found hit as well as the amount of material crossed, which is used to estimate the uncertainty arising from multiple Coulomb scattering. The Kalman filter method is implemented in four steps.
The first step, *navigation*, uses the parameters of the track candidate, evaluated at the current layer, to determine which adjacent layer(s) of the tracking detector are intersected by the extrapolated trajectory.
The second step is a search for compatible detectors in the layers returned by the navigation step. A detector is considered compatible with the trajectory if the position at which the latter intercepts the detector surface is no more than a given number (currently three) of standard deviations outside the detector boundary.
The third step forms groups of hits, obtained by collecting all the hits compatible with a certain trajectory. A configurable parameter also allows the addition of an invalid hit to represent the possibility that the particle failed to produce a hit in the detector group, for example, due to detector inefficiency. A χ^2 test is used to check which of the hits are compatible with the extrapolated track trajectory.
The fourth and last step is to update the trajectory state. New track candidates are formed from each of the original ones, by adding to them exactly one of the compatible hits from each detector grouping (where this hit may be an invalid hit). The candidate trajectory parameters are then updated at the new detector surface, by combining the information from the hit with the extrapolated track trajectory of the original candidate.

C.1.3 Track fitting

For each trajectory, the track finding stage results in a collection of hits and an estimate of the track parameters. However, the full information about the trajectory is only available at the last hit of the trajectory and the estimate can be biased by constraints applied during the seeding stage. Therefore the trajectory is refitted using a Kalman filter and a *smoother*.
The Kalman filter is initialized at the location of the innermost hit with the trajectory estimate obtained during seeding. The fit then proceeds in an iterative way through the

full list of hits, updating the track trajectory estimate sequentially with each hit. For each valid hit, the hit position estimate is re-evaluated using the current values of the track parameters.
This first filter is complemented by the *smoothing* stage: a second filter is initialized with the result of the first one and is run backwards from the outermost layer towards the beam line. The track parameters at the surface associated with any of its hits can then be obtained from the weighted average of the track parameters of these two filters, evaluated on the surface itself.

C.2 Tracking at the HLT

The HLT uses the same reconstruction software which is used for offline event reconstruction. However, it has to run much faster, so a dedicated version of the software is achieved by modifying several configuration parameters controlling the tracking algorithms.
First of all, the number of iterations in the iterative tracking sequence is reduced.
At the HLT, tracking starts with the reconstruction of *Pixel Tracks* from triplets of pixel hits. Then, after Pixel Tracks building, only 5 iterations (instead of 6), from Iter0 to Iter4, are performed. The main configuration differences between offline and online iterations are about the seeding and fitting stages. Different constraints on p_T and on the transverse and longitudinal impact parameters are used during the seed, and different criteria are used in the final fitting and filtering stage, in order to decide to keep the track or to discard it. For example, HLT uses a higher p_T requirement when forming the seeds (usually at least 1 GeV) than is used for offline reconstruction. These tight requirements on impact parameter and p_T reduce the number of seeds found, and thus the amount of time spent building track candidates. In Fig. C.2, differences in seeding constraints are shown.

HLT (offline) cuts

Iteration	Seeds	pT cut [GeV]	d0 cut [cm]	dz cut
0	pixel triplets	0.8(0.6)	0.1(0.02)	0.3cm(4.0σ)
1	pixel triplets	0.5(0.2)	0.05(0.02)	0.1cm(4.0σ)
2	pixel pairs	1.2(0.6)	0.025(0.015)	0.05cm(4.0σ)
3 (3-4)	pixel,TIB,TID,TEC	0.8(0.3)	0.05(1.5)	0.05cm(10.0cm)
4 (5)	TIB,TID,TEC	0.8(0.7)	0.5(2.5)	0.5cm(15.0cm)
- (6)	TOB,TEC	NA(0.6)	NA(6.0)	NA(30.0cm)

Figure C.2: Different constraints for the seeding at different iterations, both online and offline.

Other possible strategies adopted to make the HLT tracking algorithm faster than the offline one are:

- performing a *regional* track reconstruction: only the tracks lying within a specified $\eta - \varphi$ region around an object of interest (muon, electron or jet candidate reconstructed using the calorimeters, ...) are reconstructed. It is accomplished by using *regional seeding*: only seeds from combinations of hits which would be consistent with a track heading into the desired $\eta - \varphi$ region are built;
- using a single iteration in the iterative tracking;
- performing a *partial* track reconstruction: the building of each track candidate is stopped once a specified number of hits have been assigned to it; this way, the hits in the outermost few layers of the tracker tend not to be used;
- changing other configurable parameters in the tracking configuration. For example, the number of best tracks retained for each track candidate when extrapolating the tracks to the next layer, in the track building.

All these changes with respect to the offline algorithm allow to save a large amount of CPU time.

C.2.1 Vertexing at the HLT

Pixel tracks are used to reconstruct the position of the interaction point. For vertex reconstruction, a simple gap clustering algorithm is used. All tracks are ordered by the z coordinate of their point of closest approach to the beamspot. Then, wherever two neighbouring elements in this ordered set had a gap between them exceeding a distance cut z_{sep}, this point is used to split the tracks on either side of it into separate groups corresponding to different vertices. In 2012 data taking, where up to 30 interactions per bunch crossing were registered, the number of reconstructed vertices still showed a linear dependence on the number of interactions without saturating (Fig. C.3). In this Figure the real number of interactions is measured using the information from the forward calorimeter (HF), which covers the pseudorapidity range $3 < |\eta| < 5$.

C.3 Tracking for 2015

In 2015, after the Long Shutdown 1 (LS1), a very high pileup scenario is expected: with a luminosity of about $10^{34}\ cm^{-2}s^{-1}$ and a bunch spacing of 25 ns, $\langle \text{PU} \rangle > 20$ is foreseen. A very important issue is then to keep the tracking timing low even in case of very high PU, when track reconstruction becomes really challenging because of the huge number of tracks for a single collision event. Several studies are ongoing, in order to try to reduce the tracking running time and to make it increase only linearly – not quadratically neither exponentially – with PU. Some modifications have been tried, both on CMS software (`CMSSW`), both on the configuration codes that are used to perform tracking. For example, in addition to the methods already mentioned, it is possible, for example, to neglect the reconstruction of all the tracks coming from PU vertices and thus to build only the ones which are compatible with the hypothesis of originating from the primary vertex, by

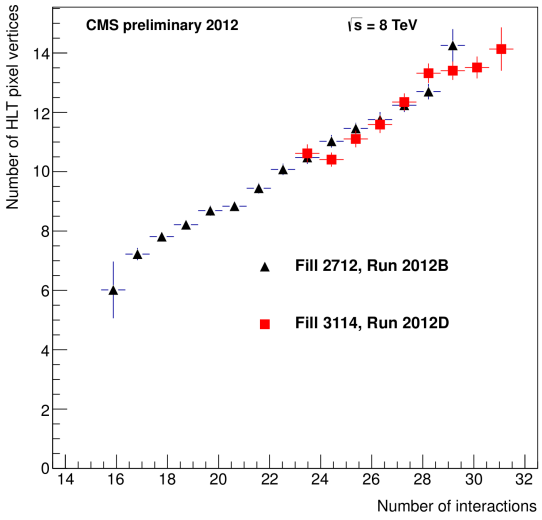


Figure C.3: Number of pixel vertices reconstructed at HLT. The number of interactions is calculated from the bunch luminosity as measured by the forward calorimeters (HF).

applying a so-called "Primary Vertex (PV) constraint".
In Fig. C.4 a plot summarizing the Iterative Tracking timing for several average PUs, for different versions of `CMSSW` or different tracking code configurations, is shown. These are the different code versions/configurations used:

- first step (in black): when using `CMSSW_6_1_0` (an old code version);
- second step (in red): with an improved `CMSSW` release (`CMSSW_6_2_0`);
- third step (in blue): after applying the PV constraint;
- fourth step (in green): after applying both the PV constraint and a tighter p_T cut on tracks reconstructed at Iter1 and Iter3.

It is worth to be noted that only in the last stage (PV constraint and tighter p_T cuts applied) an approximately linear behaviour of timing as a function of pile-up interactions is obtained.

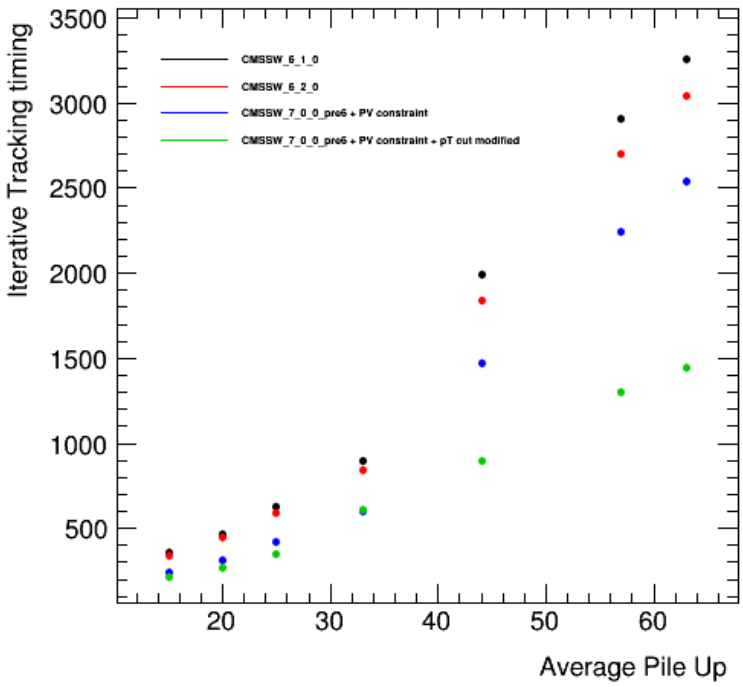


Figure C.4: Iterative Tracking timing at different average PUs. The average PU is calculated from the instantaneous luminosity, which in turn is measured by the Hadron Forward calorimeters.

Bibliography

[1] **CMS** Collaboration, "Search for a standard model like Higgs boson in the decay channel $H \to ZZ \to l^+l^-q\bar{q}$ at CMS", Tech. Rep. CMS-PAS-HIG-12-024, Geneva, 2013.

[2] F. Halzen and A. Martin, *QUARKS AND LEPTONS: An Introductory Course in Modern Particle Physics*. Wiley, New York USA, 1984.

[3] S. Glashow, "Partial Symmetries of Weak Interactions", *Nucl.Phys.* **22** (1961) 579–588.

[4] S. Weinberg, "A model of leptons", *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266. http://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[5] M. Gell-Mann, "The interpretation of the new particles as displaced charge multiplets", *Il Nuovo Cimento* **4** no. 2, (1956) 848–866. http://dx.doi.org/10.1007/BF02748000.

[6] K. Nishijima, "Charge Independence Theory of V Particles", *Progress of Theoretical Physics* **13** no. 3, (1955) 285–304, http://ptp.oxfordjournals.org/content/13/3/285.full.pdf+html. http://ptp.oxfordjournals.org/content/13/3/285.abstract.

[7] W. J. Marciano and A. Sirlin, "Electroweak radiative corrections to τ decay", *Phys. Rev. Lett.* **61** (Oct, 1988) 1815–1818. http://link.aps.org/doi/10.1103/PhysRevLett.61.1815.

[8] T. van Ritbergen and R. G. Stuart, "On the precise determination of the Fermi coupling constant from the muon lifetime", *Nucl.Phys.* **B564** (2000) 343–390, arXiv:hep-ph/9904240 [hep-ph].

[9] B. W. Lee, C. Quigg and H. Thacker, "Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass", *Phys.Rev.* **D16** (1977) 1519.

[10] **Particle Data Group** Collaboration, K. Olive *et al.*, "Review of Particle Physics", *Chin.Phys.* **C38** (2014) 090001.

[11] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD, Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, the SLD Electroweak and H. F. Groups, "Precision electroweak measurements and

constraints on the Standard Model", 2010. http://lepewwg.web.cern.ch/LEPEWWG.

[12] **LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL** Collaboration, R. Barate *et al.*, "Search for the standard model Higgs boson at LEP", *Phys.Lett.* **B565** (2003) 61–75, arXiv:hep-ex/0306033 [hep-ex].

[13] **TEVNPH (Tevatron New Phenomena and Higgs Working Group), CDF Collaboration, D0** Collaboration, "Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data", arXiv:1107.5518 [hep-ex].

[14] G. M. Pruna and T. Robens, "The Higgs Singlet extension parameter space in the light of the LHC discovery", *Phys.Rev.* **D88** (2013) 115012, arXiv:1303.1150 [hep-ph].

[15] R. Schabinger and J. D. Wells, "A Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the large hadron collider", *Phys.Rev.* **D72** (2005) 093007, arXiv:hep-ph/0509209 [hep-ph].

[16] B. Patt and F. Wilczek, "Higgs-field portal into hidden sectors", arXiv:hep-ph/0605188 [hep-ph].

[17] M. Gell-Mann, "A Schematic Model of Baryons and Mesons", *Phys. Lett.* **8** (1964) 214–215.

[18] *Phys. Rev.* **139** (1965) B1006–B1010.

[19] H. Fritzsch, M. Gell-Mann and H. Leutwyler, "Advantages of the Color Octet Gluon Picture", *Phys. Lett.* **B47** (1973) 365–368.

[20] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. 1", *Phys. Rev.* **D8** (1973) 3633–3652.

[21] S. Weinberg, "Nonabelian Gauge Theories of the Strong Interactions", *Phys. Rev. Lett.* **31** (1973) 494–497.

[22] R. K. Ellis, W. J. Stirling and B. R. Webber, *QCD and Collider Physics*. Cambridge University Press, N. Svartholm, Stockholm, 1996.

[23] **CTEQ** Collaboration, R. Brock *et al.*, "Handbook of perturbative QCD: Version 1.0", *Rev. Mod. Phys.* **67** (1995) 157–248.

[24] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl. Phys.* **B126** (1977) 298.

[25] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin *et al.*, "Implications of CTEQ global analysis for collider observables", *Phys.Rev.* **D78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[26] G. Watt, "Parton distribution function dependence of benchmark standard model total cross sections at the 7 tev lhc", *Journal of High Energy Physics* **2011** no. 9, (2011) . http://dx.doi.org/10.1007/JHEP09%282011%29069.

[27] **NNPDF** Collaboration, R. D. Ball *et al.*, "Parton distributions with QED corrections", *Nucl.Phys.* **B877** (2013) 290–320, arXiv:1308.0598 [hep-ph].

[28] T. Gleisberg, S. Hoeche, F. Krauss, A. Schalicke, S. Schumann *et al.*, "SHERPA 1. alpha: A Proof of concept version", *JHEP* **0402** (2004) 056, arXiv:hep-ph/0311263 [hep-ph].

[29] T. Stelzer and W. Long, "Automatic generation of tree level helicity amplitudes", *Nucl.Phys.Proc.Suppl.* **37B** (1994) 158–162.

[30] M. Bengtsson and T. Sjostrand, "Coherent Parton Showers Versus Matrix Elements: Implications of PETRA - PEP Data", *Phys.Lett.* **B185** (1987) 435.

[31] M. H. Seymour, "Matrix element corrections to parton shower algorithms", *Comput.Phys.Commun.* **90** (1995) 95–101, arXiv:hep-ph/9410414 [hep-ph].

[32] G. Miu and T. Sjostrand, "W production in an improved parton shower approach", *Phys.Lett.* **B449** (1999) 313–320, arXiv:hep-ph/9812455 [hep-ph].

[33] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **0411** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].

[34] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128, arXiv:1106.0522 [hep-ph].

[35] T. Sjöstrand, S. Mrenna and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[36] **ATLAS** Collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys.Lett.* **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[37] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys.Lett.* **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[38] L. H. C. S. W. Group, "Higgs cross sections at 7, 8 and 14 TeV", 2014. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections#Higgs_cross_sections_and_decay_b.

[39] **CMS** Collaboration, "Precise determination of the mass of the Higgs boson and studies of the compatibility of its couplings with the standard model", Tech. Rep. CMS-PAS-HIG-14-009, CERN, Geneva, 2014.

[40] **LHC Higgs Cross Section Working Group** Collaboration, S. Heinemeyer *et al.*, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", arXiv:1307.1347 [hep-ph].

[41] **CMS** Collaboration, V. Khachatryan *et al.*, "Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs", *Phys.Lett.* **B736** (2014) 64, arXiv:1405.3455 [hep-ex].

[42] "Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data", Tech. Rep. ATLAS-CONF-2014-009, CERN, Geneva, Mar, 2014.

[43] **ATLAS** Collaboration, "Study of the spin of the new boson with up to 25 fb^{-1} of ATLAS data", Tech. Rep. ATLAS-CONF-2013-040, CERN, Geneva, Apr, 2013.

[44] **ATLAS** Collaboration, G. Aad *et al.*, "Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data", *Phys.Rev.* **D90** (2014) 052004, arXiv:1406.3827 [hep-ex].

[45] **LHC Study Group** Collaboration, "The Large Hadron Collider: Conceptual design", CERN-AC-95-05-LHC, 1995. http://cds.cern.ch/record/291782?ln=it.

[46] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic *et al.*, "LHC Design Report, Vol. 1: The LHC Main Ring", CERN-2004-003-V-1, 2004. http://cds.cern.ch/record/782076?ln=it.

[47] O. Buning, P. Collier, P. Lebrun, S. Myers, R. Ostojic *et al.*, "LHC Design Report, Vol. 2: The LHC Infrastructure and General Services", CERN-2004-003-V-2, 2004. http://cds.cern.ch/record/815187?ln=it.

[48] M. Benedikt, P. Collier, V. Mertens, J. Poole and K. Schindl, "LHC Design Report, Vol. 3: The LHC Injector Chain", CERN-2004-003-V-3, 2004. http://cds.cern.ch/record/823808?ln=it.

[49] **Lep Injector Study Group** Collaboration, "LEP Design Report, Vol. 1: The LEP Injector Chain", CERN-LEP-TH-83-29, CERN-PS-DL-83-81, CERN-SPS-83-26, LAL-RT-83-09, 1983. http://cds.cern.ch/record/98881?ln=it.

[50] **Lep Injector Study Group** Collaboration, "LEP Design Report, Vol. 2: The LEP Main Ring", CERN-LEP/8-01, 1984. http://cds.cern.ch/record/102083?ln=it.

[51] **Lep Injector Study Group** Collaboration, "LEP Design Report, Vol. 3: LEP2", CERN-AC-96-01-LEP2, 1996. http://cds.cern.ch/record/314187?ln=it.

[52] **CMS** Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004.

[53] **ATLAS** Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003.

[54] **LHCb** Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005.

[55] **ALICE** Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08002.

[56] **TOTEM** Collaboration, "The TOTEM experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08007.

[57] **LHCf** Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *JINST* **3** (2008) S08006.

[58] **Particle Data Group** Collaboration, J. Beringer *et al.*, "Review of Particle Physics (RPP)", *Phys. Rev. D* **86** (2012) 010001.

[59] **CMS** Collaboration, "The CMS magnet project: Technical Design Report", CMS-TDR-1, 1997. `http://cds.cern.ch/record/331056?ln=it`.

[60] **CMS** Collaboration, V. Karimäki, M. Mannelli, P. Siegrist, H. Breuker, A. Caner, R. Castaldi, K. Freudenreich, G. Hall, R. Horisberger, M. Huhtinen and A. Cattai, "The CMS tracker system project: Technical Design Report", CMS-TDR-5, 1997. `http://cds.cern.ch/record/368412?ln=it`.

[61] **CMS** Collaboration, "The CMS tracker: addendum to the Technical Design Report", CMS-TDR-5-add-1, 2000. `http://cds.cern.ch/record/490194?ln=it`.

[62] **CMS** Collaboration, S. Chatrchyan *et al.*, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", `arXiv:1405.6569 [physics.ins-det]`.

[63] **CMS** Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report", CMS-TDR-4, 1997. `http://cds.cern.ch/record/349375?ln=it`.

[64] **CMS** Collaboration, P. Bloch, R. Brown, P. Lecoq and H. Rykaczewski, "Changes to CMS ECAL electronics: addendum to the Technical Design Report", CMS-TDR-4-add-1, 2002. `http://cds.cern.ch/record/581342?ln=it`.

[65] K. Deiters, M. Diemoz, N. Godinovic, C. Q. Ingram, E. Longo *et al.*, "Investigation of the avalanche photodiodes for the CMS electromagnetic calorimeter operated at high gain", *Nucl.Instrum.Meth.* **A461** (2001) 574–576.

[66] K. W. Bell, R. M. Brown, D. Cockerill, P. Flower, B. Kennedy *et al.*, "Vacuum phototriodes for the CMS electromagnetic calorimeter endcap", *IEEE Trans.Nucl.Sci.* **51** (2004) 2284–2287.

[67] **CMS** Collaboration, "Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV", *JINST* **8** (2013) P09009, arXiv:1306.2016 [hep-ex].

[68] **CMS** Collaboration, "The CMS hadron calorimeter project: Technical Design Report", CMS-TDR-2, 1997. http://cds.cern.ch/record/357153?ln=it.

[69] **CMS** Collaboration, "The CMS muon project: Technical Design Report", CMS-TDR-3, 1997. http://cds.cern.ch/record/343814?ln=it.

[70] **CMS** Collaboration, "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems", CMS-TDR-6-1, 2000. http://cds.cern.ch/record/706847?ln=it.

[71] **CMS** Collaboration, "CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger", CMS-TDR-6, 2002. http://cds.cern.ch/record/578006?ln=it.

[72] **CMS** Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET",.

[73] W. Adam, R. Frühwirth, A. Strandlie and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC",.

[74] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, "Reconstruction of Electron Tracks With the Gaussian-Sum Filter",.

[75] **CMS** Collaboration, S. Chatrchyan *et al.*, "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, arXiv:1206.4071 [physics.ins-det].

[76] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl.Instrum.Meth.* **A262** (1987) 444–450.

[77] https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuScleFitMuonMomentumCorrections.

[78] **CMS** Collaboration, "Electron reconstruction and identification at sqrt(s) = 7 TeV",.

[79] **CMS** Collaboration, "Electron commissioning results at $\sqrt{s}$=7 TeV",.

[80] M. Cacciari, G. P. Salam and G. Soyez, "FastJet User Manual", *Eur.Phys.J.* **C72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[81] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys.Lett.* **B659** (2008) 119–126, arXiv:0707.1378 [hep-ph].

[82] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of the Inclusive W and Z Production Cross Sections in *pp* Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1110** (2011) 132, arXiv:1107.4789 [hep-ex].

[83] **CMS** Collaboration, S. Chatrchyan *et al.*, "Studies of jet mass in dijet and W/Z + jet events", *JHEP* **1305** (2013) 090, arXiv:1303.4811 [hep-ex].

[84] G. C. Blazey, J. R. Dittmann, S. D. Ellis, V. D. Elvira, K. Frame *et al.*, "Run II jet physics", arXiv:hep-ex/0005012 [hep-ex].

[85] G. Soyez, "The SISCone and anti-k_t jet algorithms", DIS2008, 2008. arXiv:0807.0021 [hep-ph].

[86] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone algorithm", *JHEP* **0705** (2007) 086, arXiv:0704.0292 [hep-ph].

[87] G. Soyez and G. Salam, "The SISCone Jet Algorithm", http://siscone.hepforge.org/.

[88] M. Cacciari, G. P. Salam and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[89] S. Catani, Y. L. Dokshitzer, M. Seymour and B. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions", *Nucl.Phys.* **B406** (1993) 187–224.

[90] Y. L. Dokshitzer, G. Leder, S. Moretti and B. Webber, "Better jet clustering algorithms", *JHEP* **9708** (1997) 001, arXiv:hep-ph/9707323 [hep-ph].

[91] **CMS** Collaboration, "Identifying Hadronically Decaying Vector Bosons Merged into a Single Jet", Tech. Rep. CMS-PAS-JME-13-006, CERN, Geneva, 2013.

[92] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **1103** (2011) 015.

[93] S. Ellis, et al., "Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches", *Phys. Rev.* **D81** (2010) 094023.

[94] **CMS** Collaboration, "Pileup Jet Identification", CMS PAS JME-13-005, 2013. http://cds.cern.ch/record/1581583?ln=it.

[95] **CMS** Collaboration, S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **6** (2011) P11002, arXiv:1107.4277 [physics.ins-det].

[96] **CMS** Collaboration, "Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11 fb^{-1} of data in CMS", CMS DP 2013/011, May, 2013. http://cds.cern.ch/record/1545350?ln=it.

[97] **D0** Collaboration, B. Abbott *et al.*, "Determination of the absolute jet energy scale in the D0 calorimeters", *Nucl.Instrum.Meth.* **A424** (1999) 352–394, arXiv:hep-ex/9805009 [hep-ex].

[98] **CMS** Collaboration, R. Eusebi, "Jet energy corrections and uncertainties in CMS: Reducing their impact on physics measurements", *J.Phys.Conf.Ser.* **404** (2012) 012014.

[99] https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#JER_Uncertainty.

[100] **CMS** Collaboration, S. Chatrchyan *et al.*, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, arXiv:1211.4462 [hep-ex].

[101] 2008. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookDataFormats#AoD.

[102] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG", *JHEP* **1002** (2010) 037, arXiv:0911.5299 [hep-ph].

[103] J. M. Campbell, R. K. Ellis, P. Nason and G. Zanderighi, "W and Z bosons in association with two jets using the POWHEG method", *JHEP* **1308** (2013) 005, arXiv:1303.5447 [hep-ph].

[104] G. Passarino, C. Sturm and S. Uccirati, "Higgs Pseudo-Observables, Second Riemann Sheet and All That", *Nucl. Phys.* **B834** (2010) 77–115, arXiv:1001.3360 [hep-ph].

[105] S. Goria, G. Passarino and D. Rosco, "The Higgs Boson Lineshape", *Nucl.Phys.* **B864** (2012) 530–579, arXiv:1112.5517 [hep-ph].

[106] N. Kauer and G. Passarino, "Inadequacy of zero-width approximation for a light Higgs boson signal", *JHEP* **08** (2012) 116, arXiv:1206.4803 [hep-ph].

[107] **CMS** Collaboration, "Search for a Standard Model-like Higgs boson decaying into $WW \to l\nu q\bar{q}$ in pp collisions at $\sqrt{s} = 8$ TeV", Tech. Rep. CMS-PAS-HIG-13-008, CERN, Geneva, 2013.

[108] https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs#22Jan2013_ReReco_of_2012_data_re.

[109] https://twiki.cern.ch/twiki/bin/view/Main/EGammaScaleFactors2012.

[110] http://cmsdoc.cern.ch/~vagori/plots_newLumi_KinCuts.

[111] http://lenzip.web.cern.ch/lenzip/H2l2q/20140602/plots_LDcut_SR_AN-13-369_V9/.

[112] N. Cabibbo and A. Maksymowicz, "Angular Correlations in Ke-4 Decays and Determination of Low-Energy pi-pi Phase Shifts", *Phys.Rev.* **137** (1965) B438–B443.

[113] Y. Gao, A. V. Gritsan, Z. Guo, K. Melnikov, M. Schulze *et al.*, "Spin determination of single-produced resonances at hadron colliders", *Phys.Rev.* **D81** (2010) 075022, arXiv:1001.3396 [hep-ph].

[114] **TMVA Core Developer Team** Collaboration, J. Therhaag, "TMVA: Toolkit for multivariate data analysis", *AIP Conf.Proc.* **1504** (2009) 1013–1016.

[115] J. R. Quinlan, "Bagging, Boosting, and C4.5", in *In Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pp. 725–730. 1996. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.2457.

[116] I. M. Chakravarti, R. G. Laha and J. Roy, *Handbook of Methods of Applied Statistics*, vol. I. John Wiley and Sons, 1967.

[117] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR3#Higgs_2_gauge_bosons.

[118] https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources#2012_JEC.

[119] **CMS** Collaboration, "Measuring Electron Efficiencies at CMS with Early Data",.

[120] **CMS** Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", Tech. Rep. CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[121] **CMS** Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2012 Update", Tech. Rep. CMS-PAS-LUM-12-001, CERN, Geneva, 2012.

[122] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)",.

[123] G. Cowan, K. Cranmer, E. Gross and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur.Phys.J.* **C71** (2011) 1554, arXiv:1007.1727 [physics.data-an].

[124] S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", *Annals Math.Statist.* **9** no. 1, (1938) 60–62.

[125] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro *et al.*, "The RooStats Project", *PoS* **ACAT2010** (2010) 057, arXiv:1009.1003 [physics.data-an].

[126] **CMS** Collaboration, "Standard model Higgs boson search in pp-collisions at $\sqrt{s} = 7$ TeV and integrated luminosity 4.X fb^{-1}", Tech. Rep. CMS-HIG-11-032, Geneva, 2011.

[127] ATLAS and CMS Collaborations, "Procedure for the LHC Higgs boson search combination in Summer 2011", *ATL-PHYS-PUB-2011-11, CMS NOTE-2011/005* (2011) .

[128] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl.Instrum.Meth.* **A434** (1999) 435–443, arXiv:hep-ex/9902006 [hep-ex].

[129] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys.* **G28** (2002) 2693–2704.

[130] L. Moneta, K. Belasco, K. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke and M. Wolf, "The RooStats Project", in *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. SISSA, 2010. 1009.1003 [physics.data-an]. http://pos.sissa.it/archive/conferences/093/057/ACAT2010_057.pdf. PoS(ACAT2010)057.

[131] **CMS** Collaboration, "Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV", CMS Physics Analysis Summary CMS-PAS-HIG-13-005, CERN, Geneva, 2013. http://cds.cern.ch/record/1542387.

[132] **CMS** Collaboration, "Search for a standard model like Higgs boson in the $H \to ZZ \to l^+l^-q\bar{q}$ decay channel at $\sqrt{s} = 8$ TeV", Tech. Rep. CMS-PAS-HIG-14-007, CERN, Geneva, 2014.

[133] 2013. https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetWtagging.

[134] 2013. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections.

[135] W. Adam, B. Mangano, T. Speer and T. Todorov, "Track reconstruction in the CMS tracker",.

Acknowledgements

I would like to thank all the CMS people in Florence, with a hearty thanks to Vitaliano Ciulli and Elisabetta Gallo, who made all this possible, and to Giacomo Sguazzoni.
Special thanks go to Antonio Tropiano and Piergiulio Lenzi, the best teachers I could wish I had.
Then, I would like to express my gratitude to Oscar Gonzalez Lopez, Matthias Mozer and all the people from the 2l2q analysis working group, and also to Andrea Bocci, Mia Tosi, Simone Gennai, Roberto Carlin, Roberta Arcidiacono, Giuseppe Cerati and all the trigger/tracking people I had the opportunity to work with.
Many many thanks to Francesco Coradeschi for the very precious help he offered me during the draft of the first chapter of this thesis and for the several conversations about QFT, QCD, Susy and particle physics theory in general.

Ringraziamenti

Ringraziare, v. tr. [der. di *grazie*, col pref. *rin-*] (io *ringràzio*, ecc.). – Esprimere con parole o con altro segno esteriore il proprio sentimento di gratitudine verso qualcuno [...].

Tre anni, o forse un'era geologica.
Ho scritto molte pagine di ringraziamenti, nella mia testa, in questo periodo. Ogni volta allungando la lista. Ogni volta qualcosa in più per cui "essere grata".
Tutto è cominciato quasi per caso, ed è diventato il "futuro inverosimile" che tante volte avevo detto di volere.

Tante persone hanno reso possibile questo futuro inverosimile.
Vitaliano ed Elisabetta, prima di ogni altro, mi hanno dato piena fiducia e mi hanno messo a disposizione ogni mezzo necessario, incoraggiandomi sempre, fin dall'inizio. La loro passione e il loro impegno sono stati uno stimolo e un modello da seguire.
Antonio, che per primo mi ha insegnato tutto l'indispensabile per sopravvivere in CMS (e non solo). Credendo in me e supportandomi (e sopportandomi!) in ogni situazione, punto di riferimento e guida in una giungla terrificante di nomi, persone, codice e quant'altro. Sempre con il sorriso sulle labbra e una serenità incredibile. Non lo ringrazierò mai abbastanza.
Giulio, che ringrazio di cuore. Per la pazienza, la passione che trasmette in ogni cosa che fa e che insegna, per l'estrema competenza, la calma serafica, la saggezza. Fonte inesauribile di conoscenza per la Fisica e il Codice. L'unico che sia mai riuscito a smorzare le mie ansie planetarie con la sua pacatezza e il senso di sicurezza che riesce a trasmettere.

Le persone con cui ho condiviso questa era geologica e il "viaggio interstellare" non cadranno facilmente nell'oblio. Sono quelle che mi hanno fatto sentire a casa ogni volta che tornavo o che restavo, e che hanno fatto di tutto per non farmi sentire "sola" mai.
L'elenco è lungo e incompleto, e l'ordine sparso: Andrea, Rosanna & Marcello, Sabrina, Sandro (con le sue liste e i suoi consigli che troppo spesso non ho seguito, ma qualche volta sì, per fortuna), Antonio, Giulio, il Pugge (quanto supporto... come avrei fatto a sopravvivere senza?), Silvio, Filippo, Giulio, Massimo, Sergey (per i mille casini che mi ha aiutato a risolvere, per i film, la Vodka a Pasqua e quant'altro), Roberta, Maddalena (che una volta mi ha salvato la vita), Giuseppe (per l'Umanità in un ambiente talvolta disumano), Sara, Lucio e l'autostrada per arrivare alla Meyrinoise, Francesco Coradeschi, Silvia Lanfranchi, Ilaria & Viola, Gina & Gabriele, Ilaria, Francesco, Giuseppe e Lydie, Vieri, Alessandra Pastore, Damiano Tommasini, Zerocalcare (soprattutto per lo splendido pomeriggio a Ginevra), Luca Perniè, Martina Machet, Linda Finco, Livia Soffi, Alberto

Mengarelli, Giacomo Fedi, Mia (che energia! e che coraggio, e quanta stima!) e il suo Sandro, Piet, Andrea Latina, Tommaso Olmastroni, Roberto Mattei, Maurizio Pierini, Eduardo Navarro, Luca Malgeri, Roberto Carlin e il suo propaganda plot, Andrea Bocci & Simone Gennai (per avermi offerto la possibilità di fare conferenze in posti che non avrei mai immaginato di vedere), Oscar, Eduardo e Pablo Garcia, Britta & Diether, Vincenzo e molti altri.

Il tango, e quindi di nuovo Sandro e Roberta, ma soprattutto Daniele (per sempre il Maestro) e Miwa, e poi Joelle e Jerome, Yago, Maria, di nuovo Andrea, Attilio, Kay, Nancy e tutti gli altri della baracchina del Tango Club, consolazione e rifugio sicuro da ogni male del mondo ogni agognato giovedì.

E poi le guardie notturne, i luoghi non luoghi, la disperazione e la rabbia, il building 40 nei fine settimana o nelle sere più buie, il bisogno di sentirsi vivi, la fatica e la soddisfazione, e poi ancora il tango, e i tunnel interminabili e i tergicristalli per gli occhi, e l'hopital de la tour, le distese di fiori gialli, i tavoli da biliardo al cern, zauker e luana, gli aeroporti, le felpe di lindsey gray, il ronzio dell'alta tensione, il cineclub, le pizze d'asporto a plainpalais, grisù, la stanza 183, norma, la salle de la calame, e ancora prima la neve, milano centrale, gli openspace, e hanno ammazzato pablo, e poi l'emigrazione 2.0, la mia punto rossa carica di roba, l'ostello a saint genis, la magione in contea, brocard, l'AMR, la perle du lac, disorder, gli escargot e troppe, troppe altre cose.